APPLICATIONS AND ISSUES OF DATA MINING

Reena Hooda*

Nasib S. Gill**

ABSTRACT

Data mining is a technique of data warehousing and is employed to extract important information or knowledge that may possibly be unknown otherwise. It offers a support to take appropriate decisions on certain facts and alternatives. There are various applications of data mining like analysis, predictions, marketing, trend analysis & forecasting and researches etc. Applications of this concept raises several issues including social factors like privacy, applicability, cost and integrity. The present paper exhibits an introduction to the data mining, the various applications of data mining together with the issues required to be taken care of.

Keywords: *Data Mining, Applications, Analysis, Decision, Knowledge.*

*Assistant Professor, Dept. of Computer Sc. & Applications, PDM College of Engineering, Bahadurgarh, Haryana, India.

**Head, Dept. of Computer Sc. & Applications, Maharshi Dayanand University, Rohtak Haryana, India.

INTRODUCTION

Data Mining may be viewed as automated search procedures for discovering credible and actionable insights from large volumes of high dimensional data [1]. Data mining or knowledge discovery is the process of analyzing data from different perspectives and summarizing it into useful information that can be used to increase revenue, cuts costs, or both [2]. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases [2]. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analysis offered by data mining move beyond the analysis of past events provided by retrospective tools typical of decision support systems [3]. Data mining tools can answer business questions that traditionally were too time-consuming to resolve. They scour databases for the hidden patterns, finding predictive information that experts can miss because it lies outside their expectations [3]. Data mining techniques can be applied on a wide variety of data types including databases, text, spatial data, temporal data, images, and other complex data [4].

Data mining is a technology to enable data exploration, data analysis as well as data visualization of large databases at a high level of abstraction, without a specific hypothesis in mind [5]. It can be defined as the non-trivial extraction of novel, implicit, and actionable knowledge from large datasets that includes [5]:

- Extremely large datasets.
- Discovery of the non-obvious i.e. hidden and unknown facts.
- Useful knowledge that can improve processes.
- Can not be done manually.

Data Mining employs techniques from statistics, pattern recognition, and machine learning, high performance computers, parallel algorithms, visualizations, database, etc. Many of these methods are also frequently used in vision, speech recognition, image processing, handwriting recognition, and natural language understanding. However, the issues of scalability and automated business intelligence solutions differentiate data mining from the other applications of machine learning and statistical modeling [1] [5].

APPLICATIONS OF DATA MINING

Data mining is used to discover patterns and relationships in the data in order to help Managers to take appropriate decisions, serve customers and gain the competitive edge. The various utilizations of data mining are as follow:

Applicable to Various Organizations: Data mining can be used by any organization-profitable, non-profitable, educational institution and even for diverse fields like: retail, financial, communication, and marketing organizations etc. for strong consumer focus, their retention, employee satisfaction & development, and even for the modernization of certain organization for their long survival and success.

Automated Prediction of Trends and Behaviors: Data mining automates the process of finding predictive information in large databases. A typical example of a predictive problem is targeted marketing. For instance, data mining uses data on past promotional mailings to identify the targets most likely to maximize return on investment in future mailings. Other predictive problems include forecasting bankruptcy & other forms of default, and identifying segments of a population likely to respond similarly to given events [3].

Automated Discovery of Previously Unknown Patterns: Data mining tools sweep through databases and identify previously hidden patterns in one step. An example of pattern discovery is the analysis of retail sales data to identify seemingly unrelated products that are often purchased together. Other pattern discovery problems include detecting fraudulent credit card transactions and identifying anomalous data that could represent data entry keying errors [3].

Analysis of Various Factors in the Market: It enables companies to determine relationships among "internal" factors such as price, product positioning, staff skills, and "external" factors such as economic indicators, competition, and customer demographics. It assists business houses to determine the impact on sales, customer satisfaction, and corporate profits [2].

Marketing: uses of data mining in marketing can be divided into following parts:

- *Market segmentation*: Identify the common characteristics of customers who buy the same products from your company [6].

- *Target marketing*: Identify the segment to target [6].

- *Customer churn*: Predict which customers are likely to leave the company and go to a competitor.

- *Direct marketing*: Identify which prospects should be included in a mailing list to obtain the highest response rate.

- *Interactive marketing*: Predicting the interest of each individual accessing a web site.

- *Market basket analysis*: Understand the products or services that are commonly purchased together; e.g., beer and diapers.

Pattern Regognition: The actual data-mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection) and dependencies (association rule mining). [7]

Extensibility: Data mining techniques can yield the benefits of automation on existing software and hardware platforms, can be implemented on new systems as existing platforms are upgraded and new products developed. When data mining tools are implemented on high performance parallel processing systems, they can analyze massive databases in minutes. Faster processing means that users can automatically experiment with more models to understand complex data. High speed makes it practical for users to analyze huge quantities of data. Larger databases, in turn, yield improved predictions [3].

Data mining takes this evolutionary process beyond retrospective data access and navigation to prospective and proactive information delivery. Data mining is ready for application in the business community because it is supported by three technologies that are now sufficiently mature [3]: massive data collection, powerful multiprocessor and data mining algorithms.

Data mining also facilitates to "drill down" into summary information to view detailed transactional data [2].

Science and Engineering: In recent years, data mining has been used widely in the areas of science and engineering, such as bioinformatics, genetics, medicine, education and electrical power engineering. Data mining methods have also been applied for dissolved gas analysis (DGA) on power transformers and traffic analysis as well. [7]

Spatial Data Mining: The immense explosion in geographically referenced data occasioned by developments in IT, digital mapping, remote sensing, and the global diffusion of GIS emphasises the importance of developing data driven inductive approaches to geographical analysis and modeling. [7]

Decision Support System and Business Intelligence: This amalgamation of data mining and the business decision making is called Business Intelligent System. Then knowledge can be acquired from thriving implementation of this automated system (Business Intelligence System) which may include intelligent exploration, integration, aggregation and a multidimensional analysis of data. Such kind of intelligent systems are meant to provide adequate and reliable up-to-date information on different aspects of enterprise activities. Business Intelligence System can be obliged for developing different comparative reports, e.g. on historical results, profitability of particular offers, effectiveness of distribution channels along with carrying out simulations of development or forecasting future results.

The systems also allowed for optimizing future proceedings and for modifying organizational, financial or technological aspects of company performance appropriately in turn to lucratively implementation of the budgets of an enterprise and confiscate the deviations and thus provides an aid to decision making. [8]

DATA MINING ISSUES

Although data mining aids business man, individuals to analyze the data from different perspective and facilitate them for knowledge discovery and superior decision making but there are also various social issues that may arise with the use of the data mining as stated below:

Privacy: It is the issue of individual privacy. Data mining makes it possible to analyze routine business transactions and gives a significant amount of information about individuals buying habits and preferences [2]. Therefore, the private data may be revealed that may result in a high personal loss.

Data Integrity: Clearly, data analysis can only be as good as the data that is being analyzed. Integrating conflicting or redundant data from different sources is a key challenge for implementation. For example, a bank may maintain credit cards accounts on several different databases. The addresses (or even the names) of a single cardholder may be different in each. Software must translate data from one system to another and select the address most recently entered [2].

Application Suitability: A hotly debated technical issue is whether it is better to set up a relational database structure or a multidimensional one. In a relational structure, data is stored in tables, permitting ad hoc queries. In a multidimensional structure, on the other hand, sets of cubes are arranged in arrays, with subsets created according to category. While multidimensional structure facilitates multidimensional data mining, the relational structure performs better in client/server environment. And, with the explosion of the Internet, the world is becoming one big client/server environment [2].

Cost: While system hardware costs have dropped dramatically within the past five years, data mining and data warehousing tend to be self-reinforcing. The more powerful the data mining queries, the greater the utility of the information being gleaned from the data, and the greater the pressure to increase the amount of data being collected and maintained, which increases the pressure for faster, more powerful data mining queries. This increases pressure for larger and faster systems which are more expensive [2].

CONCLUSIONS

Data mining is a core technique of data warehousing to facilitates diverse users in different areas. The applications of data mining are vary wide and covers almost every field from most general information extraction to more critical scientific applications having complex data and large proecedures. It provides a wide scope for the researchers also. Implementation of data mining inherents cost and various other issues, besides this the techniques required some training that may be fundamental or crucial, depending upon the type of application. Therefore, before stepping ahead, all the potential proc and cons should carefully be examined.

REFERENCES

[1] Apte, C. (2003). *Data Mining Analytics for Business Intelligence and Decision Support*. Retrieved from: http://www.research.ibm.com/dar/papers/pdf/orms2.pdf, http:// www.research.ibm.com/dar/publications.html. in OR/MS Today, February 2003.

[2] Palace, B. (1996). Data Mining. *Spring 1996,* Retrieved from: http://www.anderson.ucla. edu/ faculty/ jason. frand/teacher/technologies/palace/ datamining.htm

[3] Thearling, K. (2009). *Information about data mining and analytic technologies: An Introduction to Data Mining-Discovering hidden value in your data warehouse.* Retrieved from: http://www.thearling.com /text/ dmwhite /dmwhite.htm Copyright © 2009 Kurt Thearling.

[4] Frawley, W.J., Piatetsky-Shapiro, G., & Matheus, C.J. (1991). Knowledge Discovery in Databases: An Overview. In Piatetsky-Shapiro, G. & Frawley, W.J. (Eds.), Knowledge discovery in databases (pp. 1-27). Cambridge, MA: AAAI Press/MIT Press.

[5] Williams, G., Hegland, M., & Roberts, S. (1998). A Data Mining Tutorial. *In Proceedings of the Second IASTED International Conference on Parallel and Distributed Computing and Networks (PDCN'98),* 14 December 1998, Copyright c 1998.

[6] Gill, N. S. & Hooda, R. (2009c). Clustering in Management. *Sandesh- Masthnath journal 2009.*

[7] Wikipedia, the free encyclopedia (25 November, 2011). Data mining. Retrieved from: http://en.wikipedia.org/wiki/Data_mining

[8] Gill, N. S. and Hooda, R. (2009). *"Predictive Data Mining: Rolling from Data to Intelligence"*. Proceedings of International Conference on Innovative Technologies (ICIT-09), P-93.