RESEARCH PAPER ON ROLE OF DATA WAREHOUSING AND DATA MINING

Mohit Rathi, Manish Lohia

Abstract- **This paper includes the combined study of Business Intelligence (BI), Data Warehouse and Data Mining technologies. BI techniques are generally computer based that provide past, current and future trends of the enterprise. Its applications include activities of decision support system, query and reporting, complex event processing, online analytical processing, process mining, business performance management and statistical and predictive analysis. A data warehouse is a repository of relational databases designed for query and analysis. It separates analysis workload from transaction workload and enables an organization to consolidate data from different varying sources. This data warehouse is analyzed by data mining, which is a latest technique allowing enormous data sets to be explored so as to yield hidden and unknown predictions that can be used in future for efficient decision making.**

I. INTRODUCTION

Information technology is now required in every aspect of our lives which helps business and enterprise to make use of applications like decision support system, query and reporting, complex event processing, online analytical processing, statistical and predictive analysis and business performance management. In today's business environment, where companies face global competition, survival and defeat depends on the efficiency, timeliness, knowledge and better decision making ability. Business Intelligence (BI) refers to technologies and applications for collecting, storing and analyzing business data that ultimately helps the enterprise to make better decisions . BI techniques are generally computer based that provide past, current and future trends of the enterprise. Its applications include activities of decision support system, query and reporting, data mining, complex event processing, online analytical processing, process mining, business performance management, text mining, statistical and predictive analysis

II. DATA WAREHOUSE

Data Warehouse is actually a repository of business or enterprise databases which gives a picture of historical and current organization's operations [C. Date, 2003]. It focuses on internal sources as quality control, sales, inventory and production. Data warehouse is designed to enable efficient management decision making as it presents a coherent picture of business conditions at a single point of time. It involves development of systems that enable extraction of data from operating system and installation of a warehouse database system that helps managers to access data in flexible ways.

III. DATA MINING- A BOON IN BI TO GAIN COMPETITIVE ADVANTAGE

Data mining or knowledge discovery helps businesses in discovering new trends or patterns of behaviour that were previously not found or unnoticed. These patterns in turn, can be used in a predictive manner for variety of applications. It implies extraction of implicit, previously unknown and useful information from data. Small changes in the strategy of data mining process can result into profits or better scientific decisions. Data mining is data driven in contrast to statistical, online analytical processing and query and reporting tools which are usually user driven. It is primarily used by companies with a strong consumer focus like financial, marketing, retail and communication organizations and so enables them to find relationships among "internal factors" like price, staff skills, and product line and "external factors" like competition, customer choice, economic indicators and market segmentation.

3.1 Data Mining Process

Data mining is a pure application dependent stage that provides extraction of useful, valid, understandable patterns from database, texts and web. It provides ways to make best use of data through rapid computerization . Data mining software analyzes relationships and patterns in stored transaction's data based on open-ended user queries. It uses modelling techniques to make a model that is a set of examples or a mathematical relationship based on data from situations where the answer is known and then applying the same model to other situations where answers are hidden.

The process of data mining involves following three stages-

Exploration- This stage involves data preparation, data cleaning, and data transformations, selecting subsets of records and then performing feature selection to reduce number of variables to a manageable range.

Model building and validation - This stage involves data preparation, data cleaning, and data transformations, selecting subsets of records and then performing feature selection to reduce number of variables to a manageable range. This reduction and choice of variables depends on the complexity of analysis varying from simple predictions for regression model to exploratory analyses of graphical and statistical data.

Deployment- This stage involves data preparation, data cleaning, and data transformations, selecting subsets of records and then performing feature selection to reduce number of variables to a manageable range. This reduction and choice of variables depends on the complexity of analysis varying from simple predictions for regression model to exploratory analyses of graphical and statistical data.

Classification- It allows stored data to be divided into classes so as to locate data into pre-determined groups. For example, a company while marketing its products could mine stored customer's purchase data to determine when customers use their products in more numbers, in which area and what they purchase. This information could be now used to focus on such potential customers and market segments and increase sales by having discounts or free schemes.

Seuential pattern matching- It is based on the sequential rule A->B which implies \that event B will always be followed by event A. It allows data mining to predict behaviour patterns and trends.

Association- It is a rule X->Y such that X and Y are data item sets.

Clustering- It involves finding clusters of related or similar traits in groups. Data items are grouped in accordance with consumer preferences or logical relationships to identify market segmentation or consumer affinities. It may be of hierarchical order or non-hierarchical.

Deviation detection- It analyzes and finds significant changes in data.

Data visualisation- It enables usage of graphical ways to show hidden patterns in data.

IV. APPLICATIONS OF DATA WAREHOUSE AND DATA MINING IN BUSINESS INTELLIGENCE

Data warehousing and data mining has gained its popularity in almost all traits as it helps to quickly analyze large databases which otherwise would be too complex and time consuming.

The following list includes some possible applications-

1 Market segmentation- to identify customer's common characteristics and behaviour that purchase same products of a company [Doug Alexander, 2000].

2 Banking- to learn underwriting, mortgage approval etc.

3 Web marketing- for targeted banner advertisements, personalization and cross sell/ upsell opportunities.

4 Customer Churn- to predict customers who are likely to leave the company and go to a competitor.

5 Direct marketing- for identifying prospects that are included in mailing list so as to obtain highest response time. It includes churn models, response models and next to buy analysis.

6 Fraud detection- to identify fraudulent transactions such as in credit card usage.

7 Finance- for stock and bond analysis, analysis and forecasting of business performance.

8 Manufacturing- for quality control, improvement and preventive maintenance.

9 Trend Analysis- to reveal the difference between a customer's behaviour over consecutive months.

10 Medicine- for diagnosis, epidemiological studies, drug analysis and quality control.

11 Government- for threat assessment and searching terrorist profiles.

V. FUTURE AND SCOPE OF DATA WAREHOUSING AND DATA MINING

Data mining technology has bright future in business domains as it helps to generate new opportunities by automated prediction of behaviours and trends in a large database. For example, targeted marketing to get better Return On Investment(ROI) can be done by data mining past promotional mailings and identifying population segments that respond similarly to given events. Data mining techniques help to automatically discover previously unknown patterns such as identifying anomalous data that could highlight errors generated during keying data entry.Data mining is not only a hit with sales and marketing companies but also with financial

institutions as it allows analysts to fast search through financial records and make best investment decisions. Even healthcare organizations are using data mining techniques to understand past trends and reduce future costs. So, future of data mining can be analyzed in three phases as following-

1. In short term- here data mining is profitable for micro-marketing campaigns which advertise for target potential customers.

2. In medium term- here data mining is as easy as to use internet and emails like finding best prices.

3. In long term- here prospects of data mining are too fruitful, enabling new decisions and new insights from the database.

Though data mining has lots of benefits in varying fields, it yet poses privacy concerns where one's data stored at one database could be accessed by others either in same physical location or across internet leading to eavesdropping, frauds and security issues.

VI. DATA MINING PRODUCTS

Data mining is now welcomed and used aggressively in industries . The major database vendors are already using data mining techniques in their platforms such as-

1 Darwin- It is an Oracle Data Mining Suite which implements classification and regression trees, k-nearest neighbours, regression analysis, neural networks and clustering algorithms.

2 SQL Server- It is Microsoft database platform which allows data mining functionality through the use of clustering algorithms and classification trees.

3 SPSS, SAS and S-Plus- These are advanced statistical packages allowing the implementation of data mining algorithms.

VII. CONCLUSION

Thus data warehouse and data mining is an essential component in business operations to gain competitive intelligence as it helps in quick and efficient analysis of volumes of data stored in data warehouse and data marts from different perspectives and suggests hidden and unknown predictions that ultimately enhance future decision

making process. Its techniques allow statistical, analytical and multidimensional analysis of data to evaluate relationships, correlations and trends. It is a powerful new technology helping companies to focus on the most important information in the data collected so as to evaluate and understand the behaviour of potential customers and in turn capture the market.

REFERENCES

[1] Alexander Doug, "Data Mining", (2000),http://www.laits.utexas.edu/norman/BUS.F OR /course.mat/Alex/, electronic article.

[2] B. de Ville, (2001), "Microsoft Data Mining: Integrated Business Intelligence for e-Commerce and Knowledge Management", Boston: Digital Press

[3] Berson Alex, Smith J. Stephen, Thearling Kurt, (1999), "Building Data Mining Applications for CRM", McGraw-Hill Companies

[4] Chapple Mike, "Data Mining: An Introduction", (2011), http:/ /databases.about. com/od/datamining/a/datamining.htm

[5] Chen, S. H, (2002), "Genetic Algorithms and Genetic Programming in Computational Finance", Boston, A: Kluwer

[6] D. Pyle, (2003), "Business Modeling and Data Mining", Morgan Kaufmann, San Francisco, CA.

[7] Frank, E., Paynter, G., Witten, I.H., Gutwin, C. and Nevill-Manning, C, (1999), "Domain-specific keyphrase extraction." Proc Int Joint Conf on Artificial Intelligence IJCAI-99. Stockholm, Sweden, pp. 668-673

[8] Gilman Michael, (2004), "Nuggets and Data Mining", Data Mining Technologies Inc. Melville, NY 11714, (631) 692-4400

[9] Hill.T, Lewicki.P, (2007), "STATISTICS: Methods and Applications", Statsoft, Tulsa, OK

[10] Kelly Chuck, (2002), "What is the role of Genetic Algorithms in Data Mining", Information Management: How your Business Works, Electronic Newsletter, http://www.information-management.com/news/5755-1.html