www.ijemr.net ISSN (ONLINE): 2250-0758, ISSN (PRINT): 2394-6962

Volume-7, Issue-4, July-August 2017

International Journal of Engineering and Management Research

Page Number: 121-127

Data Mining for Business Intelligence

Abdussalm. P .K

Assistant Professor and Head of the Department of Commerce, NWC, Manjeri, Affiliated to University of Calicut, Kerala, INDIA

ABSTRACT

Data mining is the process of mining data. Data mining (DM) is used to search for patterns and correlations within a database of information. Business intelligence (BI) focuses on detail integration and organization. DM aids BI's objectives. DM and BI work together to process data and analyze it in a way that eases the workload for the users and aids with the understanding of the materials/findings. This is accomplished through recognizing relationships in the data and identifying opportunities and risks of the company. It also allows users to manipulate the data to fulfill their specific user-oriented objectives. Business Intelligence (BI) solutions have for many years been a hot topic among companies due to their optimization and decision making capabilities in business processes. The demand for yet more sophisticated and intelligent BI solutions is constantly growing due to the fact that storage capacity grows with twice the speed of processor power. This unbalanced growth relationship will over time make data processing tasks more time consuming when using traditional BI solutions.

Keywords-- *Data warehouse, Data mining, Business intelligence, Mining tasks*

I. INTRODUCTION

Many present Business Intelligence (BI) analysis solutions are manually operated making it both time consuming and difficult for users to extract useful information from a multidimensional set of data. By applying advanced Data Mining (DM) algorithms for BI it is possible to automate this analysis process, thus making the algorithms able to extract patterns and other important information from the data set.

Using technology to gain an edge in business is not a new idea. Whenever there is something new, entrepreneurs will be quick to try to find an application for it in the business world to make money. Data mining (DM) and business intelligence (BI) are among the information technology applications that have business value.

Data mining in various forms is becoming a major component of business operations. Al most every business process today involves some forms of data mining. Customer relationship management, supply chain management, optimization, demand forecasting, business intelligence and knowledge management are just some examples of business functions that have been impacted by data mining techniques. Data Mining emerged in late 1980s and is multidisciplinary field drawing tools from the statistics, database technology, artificial intelligence, pattern recognition, machine learning, information theory, knowledge acquisition, information retrieval and data visualization. Data mining refers to the technique of searching useful and relevant information from the data warehouse. It is very clear that the technology has something to do with discovering knowledge. Data mining is used in applications such as marketing, sales, and credit analysis and fraud detection. The primary goal of data mining is to deliver information. It is the extraction of hidden predictive information from large database. Data mining is a powerful technology with great potential to analyze important information in the data warehouse. The aim of data mining is to extract implicit, previously unknown and potentially useful patterns from data. It can be defined as the process of discovering meaningful new correlation, patterns and trends by digging in to large amount of data stored in warehouse, using statistical, machine learning, artificial intelligence and data visualization.

II. DATA MINING: A CONCEPTUAL CLARITY

Data mining is concerned with finding hidden information on relationship present in business data to allow businesses to make and initiate predictions for future use. It is the process of not so obvious but useful information from large data base.

Data mining- The definition

- Data mining is the efficient discovery of valuable, non obvious information from a large collection of data.
- It is the process of extracting previously unknown, valid and actionable information from large databases and then using the information to make crucial decisions.
- Data mining is a powerful technology with great potential to analyses important information in the data warehouse

Data mining helps the end users to extract useful information from large database. These large data bases are known as 'Data warehouses'. Data mining consists of five major elements.

- Extract, transform and load transaction data in to the data warehouse system
- Store and manage the data in a multidimensional data base system.
- Provide data access to business analyst and information technology professionals.
- Analyze the data by application software
- Present the data in a useful format, such as graphs or table.

III. FOUNDATION OF DATA MINING

Data mining techniques are the outcomes of rigorous research and product development. The evaluation data mining process began when business data was first stored on computers. , continued with improvements in data access, and generated technologies that allows users to navigate through their data in real time. Data mining takes this evolutionary process beyond data access and data navigation to information delivery. Data mining is applied in business community because it is supported by three technologies that are now sufficiently mature.

- Techniques to collect massive amount of data
- Availability of powerful multiprocessor computers
- Application of data mining algorithms

IV. STEPS IN THE EVOLUTION OF DATA MINING

Evolutionary Steps	Business, Question	Enabling Technologies	Product Providers	Characteristics
Data collection (1960's)	What were the total sales in the last five year	Computers, tapes, disks	IBM CDC	Static data delivery
Data Access (1980's)	What were unit sales in India last month	RDBMS, SQL , ODBC	Oracle, Sybase, Informix, IBM	Dynamic data delivery at record Microsoft level
Data warehousing and decision support (1990's)	What were unit sales in India last month? Drill down to Delhi	OLAP, multi dimensional data base, data ware houses	Pilot, Comshare, Arbor, Cogons, Microstrategy	Dynamic data delivery at multiple levels
Data mining (still manufacturing)	What's likely to happen to Delhi unit sales next month? Why?	Advanced algorithms, multiprocessors computers, massive databases	Pilot, Lockheed, IBM, SGI, numerous startups(nascent industry)	Proactive information delivery

V. WHAT ARE DATA MINING AND BUSINESS INTELLIGENCE?

The term Business Intelligence (BI) is popularized by Howard Denser in 1989 and it describes "a set of concepts and methods to improve business decision- making by using fact-based support systems" . In BI is referred to as "a process for increasing the competitive advantage of a business by intelligent use of available data in decision making." Both quotations provide a good general understanding of the BI concept and make it pretty clear why BI is so popular among a large group of modern companies. This group includes business segments such as banking, insurance, trade, software development, intelligence services to name a few. Data mining is the process of searching through data using various algorithms to discover patterns and correlations

within a database of information. Business intelligence, on the other hand, focuses more on data integration and organization. It will combine data analyze to help managers make operational, tactical, or strategic business decisions. Data mining can be used to aid the objectives of a business intelligence system.

VI. WHY ARE DM AND BI IMPORTANT?

BI tools, in conjunction with DM, make the process of getting data and analyzing it less onerous for users. BI software is usually flexible enough so that analysts can "slice and dice" the data any way they want. Since the information comes from a centralized set of data (possibly combining data from multiple databases), data extracts from the system are consistent with each other, seeing as the analysis is done on one set of data instead of on individual desktop computers that may have their own data set or analysis tools (Burns).

In addition, DM and BI can give users the ability to spot patterns by putting the data in a visual form. They can further enhance the usefulness of the information by enabling models to identify or confirm relationships, and providing the tools to the user to drill-down and focus on particular areas of interest. If the system is used regularly, comprehensive and timely information can be utilized to spot technical, organizational, and behavioral problems within the entity in time and with sufficient detail to correct the problem (Froelich). These techniques are available due to the maturity of the necessary technologies (massive data collection, powerful multiprocessor computers, and data mining algorithms) (Chaterjee). Effective and beneficial usage of DM and BI include, but not limited to:

1. Marketing – relationships are discovered between certain customer characteristics and buying patterns.
2. Corporate analysis and risk management – data is gathered and analyzed to aid in financial planning both internally and externally.
3. Fraud detection and management – patterns or irregularities are investigated, with past successes acting as heuristics to increase future successful detection.
4. E-Commerce – tracking customer preferences to provide customized content, products, and services.
5. Bioinformatics – finding sequencing and other relationships to help further scientific research.
6. Customer Relationship Management – finding, reaching, selling, satisfying, and retaining customers through understanding their wants and need (Hsu).

VII. APPLIED BUSINESS INTELLIGENCE

A huge variety of BI solutions and techniques are currently available. Some of them are listed below.

1. **AQL (Associative Query Logic**) - Analytical data processing tool that compared to OLAP is less time consuming and more machine driven.
2. **Score carding, Dash boarding and Information visualization** – Score carding is a method that allows managers to get a broad view of the performance of a business while Dash-boarding/Information Visualization deals with visual representation of abstract data.
3. **Business Performance Management** A tool for analyzing the current state of a business and for improving future strategies.
4. **DM (Data mining)** - Numerous methods for automatically searching large amounts of data for patterns and other interesting relations.
5. **Data warehouses** - Logical collections of information with structures that favor efficient data analysis (such as OLAP).
6. **DSS (Decision Support Systems**) - Machine driven system that aids the decision making process in a business.
7. **Document warehouses** - Instead of informing the business what things have happened (like the data warehouse does) the document warehouse is able to state why things have happened.
8. **EIS (Executive Information Systems)** - These systems are often considered as a spe- cialized form of DSS with the purpose of facilitating the information and decision making needs of senior executives.
9. **MIS (Management Information Systems)** - A machine driven system for processing data and providing analysis reports for decision making and planning. In order to retrieve data the system has access to all communication channels in a business.
10. **GIS (Geographic Information Systems)** - A computer system for working with geo- graphical data (e.g. Satellite images) with editing, analyzing and displaying functionality.
11. **OLAP (Online Analytical Processing)** - OLAP is a tool for doing quick analytical processing of multidimensional data by running queries against structured OLAP cubes that is build from a set of data sources.
12. **Text mining** - This task is generally referred to as the process of extracting interesting and nontrivial information/knowledge from unstructured text.

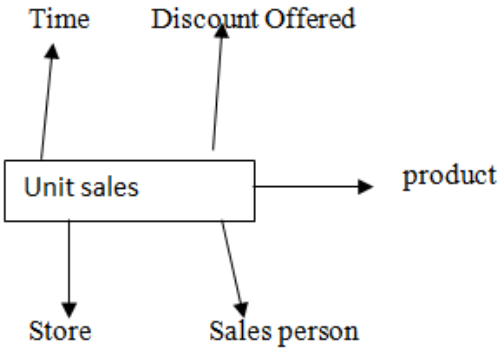
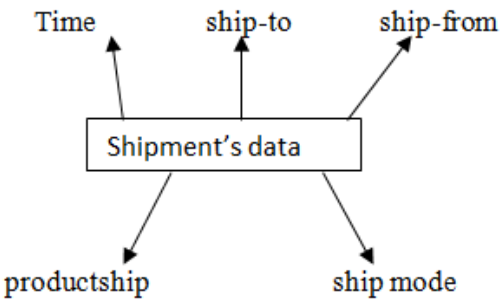
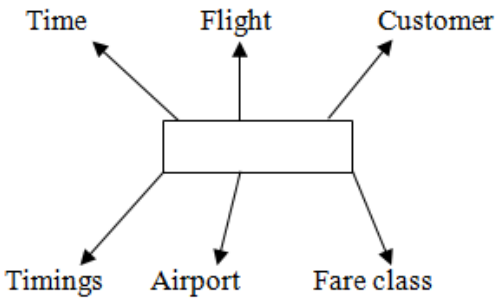
VIII. BUSINESS DIMENSION

Business dimensions are nothing but synthetically categories that allow the users to specify multiple ways to look the same information, according to natural business perspective under which its analysis can be performed. The concept of business dimensions is a key to the requirement

ISSN (ONLINE): 2250-0758, ISSN (PRINT): 2394-6962

definition of data warehouse.

Business dimension in various industries

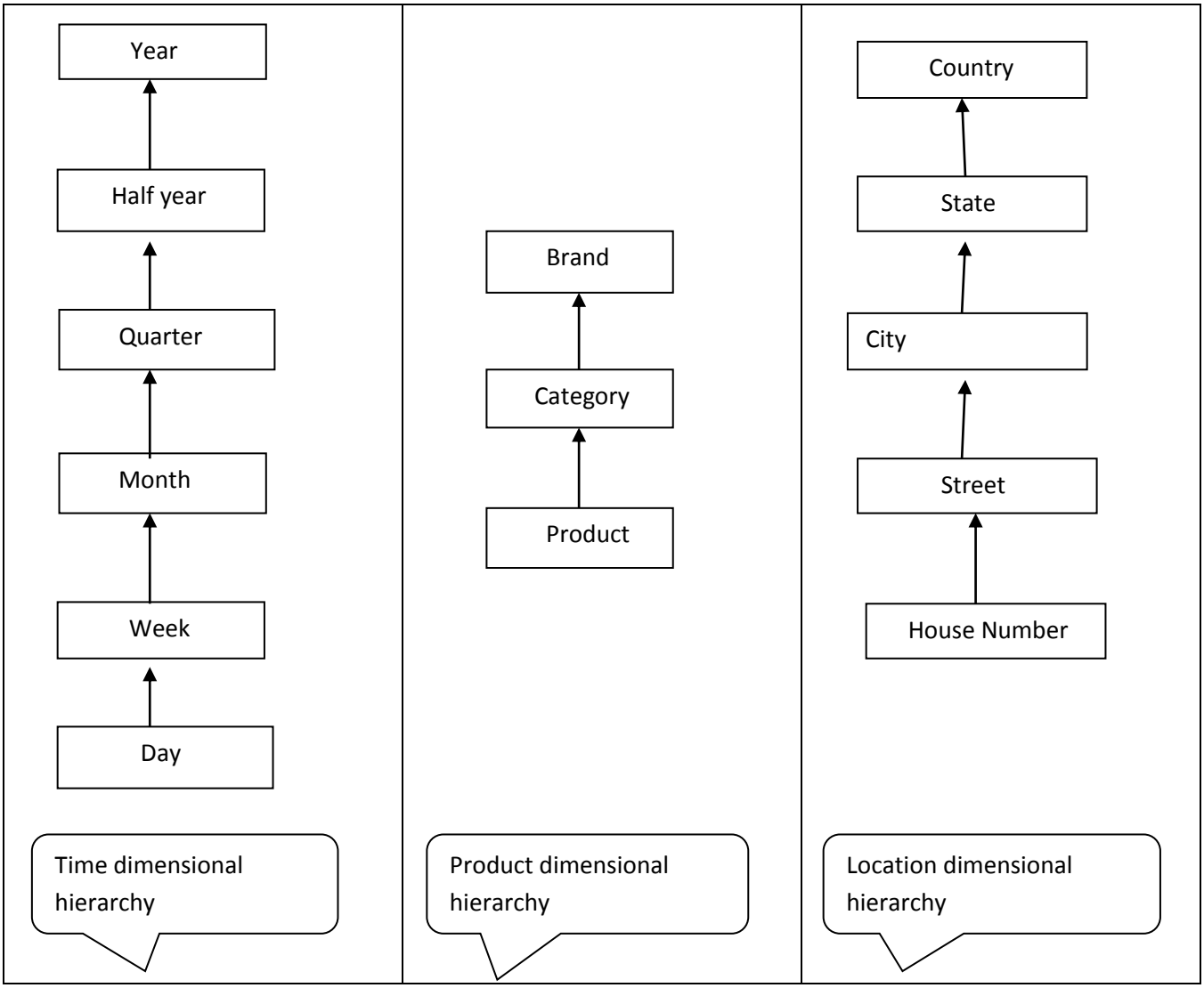
	Retail chain: in a retail chain the business analysis view the sales of a product according to time dimension.
	In the manufacturing company, shipment data is analyzed along the dimensions like, time, ship from, ship to, ship mode, product and dimensions. Time dimension is common to every business under analysis.
	In the airline business, frequent flyers data is analyzed along the dimension like time, flight, and customer, timings, airport and fare class. Time dimension is used to help track of flight timings. flight dimension store the details of different flights

From the example given above we conclude that the business dimension varies from one industry to another and these dimensions depend on the subject for analysis. We can see that the time dimension is common dimension in all examples, the reason being that all business analyses are performed over time.

IX. DIMENSIONAL HIERARCHY

Business dimension hierarchies are nothing but the path for drilling or rolling up while analyzing data at different levels of granularity. When level L1<L2 in the hierarchy, we say that L1 rolls up L2. It is important to note that a level may have descriptions associated with it. Within every individual dimension, values at different levels are related trough a family of roll up functions. If roll up function associates a value v1 of certain level to a value v2 of an upper level of hierarchy, we say that v1 rolls up v2.

ISSN (ONLINE): 2250-0758, ISSN (PRINT): 2394-6962



X. DATA MINING TASKS

Classification

Classification is supposedly the most popular Data Mining tasks considering its broad ap- placation domain. Its main purpose is to classify one or more data samples that may consist of few or many features (dimensions). The latter case makes the classification task more complex due to the large number of dimensions.

The actual number of classes is not always given or obvious in a classification task. There- fore, it is possible to distinguish between supervised and unsupervised classification. For supervised classification the number of classes is known along with the properties of each class. Neither of these is given in unsupervised classification which makes this task the more challenging one of the two.

The list below further exemplifies the use of the classification task.

1. Is a given credit card transaction fraudulent?
2. What type of subscription should be offered a given customer?
3. What type of structure does a specific protein have?
4. Is this customer likely to buy a bicycle?
5. Why is my system failing?

Estimation

Estimation is somewhat similar to classification algorithm-wise. However, estimation does not deal with determining a class for a particular data sample. Instead, it tries to predict a certain measure for a given data sample.

The list below further exemplifies the use of the estimation task.

1. What is the turnover of a company going to be?
2. What is the density of a given fluid?
3. When will a pregnant woman give birth?
4. For how long will this product work before failing?
5. How much is a specific project going to cost?

Segmentation

Segmentation basically deals with the task of grouping a given data set into a few main groups (clusters). The task of describing a large multidimensional data set (say customers) will therefore benefit from the use of segmentation. Moreover, many algorithm types can be used in segmentation systems.

The list below further exemplifies the use of the segmentation task.

1. How can a given buyer/supplier group be differentiated?
2. Which types of ground does a given satellite image contain?
3. is a specific transaction an outlier?
4. Which segments is a market based on?
5. Which groups of visitors are using a given search engine?

Forecasting

Forecasting is another important Data Mining task that is used for predicting future data values given a time series of prior data. Forecasting is a popular task often performed using simple statistical methods. However, forecasting done in the Data Mining domain uses advanced (learning) methods (e.g. Neural Networks, Hidden Markov Models) that in many cases is more accurate and informative than the standard statistical methods (e.g. Moving averages).

The list below further exemplifies the use of the forecasting task.

1. What will the weather be like tomorrow?
2. Will a particular stock price rise over the next couple of days?
3. What are the inventory levels next month?
4. How many sunspots will occur next year?
5. How will the average temperature on earth evolve throughout the next 10 years?

Association

Association deals with task of locating events that are frequently occurring together and benefiting from this knowledge. One of the most popular examples of association is probably Amazon.com's web shop that is able to recommend related products to customers.

The list below further exemplifies the use of the association task.

1. Which products should I recommend to my customers?
2. Which services are used together?
3. Which products are highly likely to be purchased together in a supermarket?
4. Which books are highly likely to be borrowed together in a library?
5. Which dishes from a cookbook go well together?

Text Analysis

Another key Data Mining task is text analysis. Text analysis has several purposes and is often used for finding key terms and phrases in text bits. In this way, text analysis can convert unstructured text into useful structured data that can be further processed by other Data Mining tasks (e.g. Classification, segmentation, association).

The list below further exemplifies the use of the text analysis task.

1. Which segments does a given mailbox contain?
2. How is a document classified?
3. Which subjects does a specific web page contain?
4. How is a quick overview of multiple lecture notes from a classmate gained?
5. Which terms are likely to occur together?

XI. CONCLUSION

The current information age is characterized by an extra ordinary growth of data that are generated and stored about all kinds of human endeavor. Data mining can be viewed as a result of natural evolution of information technology. It is the extraction of hidden predictive information from large data base. The importance of collecting data that reflect your business and scientific activities to achieve competitive advantage is widely recognized now. Powerful system for collecting data and managing it in large database are already in place in most large and mid- range companies. The data mining technologies has shown explosive growth in an increasing number of different areas of business, government and science. Data mining techniques have been applied in many areas of business like marketing management, risk management, fraud management, identification of customer spending pattern, purchase pattern, arrangements of products in the stores etc. Data mining in various forms is becoming a major component of business operations.

REFERENCES

[1] Adeyemi Adejuwon and AmirMosavi (2010) 'Domain Driven Data Mining- Application to Business, International Journal of Computer Science, Vol.7. Issue. 4, No.2, July. Pp 41-44.

[2] Berry M. and Linoff G. (1997) Data Mining Techniques for Marketing, Sales and Customer Support. John wiley and Sons.

[3] Glymour C, Madigam D, Pregibon D, (1997) 'Statistical Themes and Lesson for Data Mining', Data Mining and Knowledge Discovery.

[4] Z. Zhang and R. Zhang. *Multimedia Data Mining: A Systematic Introduction to Concepts and Theory. Chapman & Hall, 2009*

[5] L. Ye and E. Keogh. Time series shapelets: A new primitive for data mining. In *Proc. 2009 ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD'09), Paris, France, June2009*

[6] Weill P (1992) 'The relationship between Investment is Information Technology and firms performance", Information System Research, Vol.3, No.4, pp 307-333.

[7] Wikipedia. Business intelligence. http://en.wikipedia.org/wiki/Business Intelligence.
[8] Center for Mathematical and Information Sciences (CMIS), CSIRO. What is Business? Intelligence?] http://www.cmis.csiro.au/bi/what-is-BI.html
[9] Wikipedia. Information. http://en.wikipedia.org/wiki/Information.