ISSN 2229-5518

Data Mining

Anas Alarfaj

Introduction

Data mining is a technique to process and analyse a huge amount of data from different sources to get a meaningful result. It utilizes the past information to generate possible solutions to the future problems. It helps in finding out relationship between different variables of a large field database. It is largely used for decision making in companies and government establishment (Doug, Alexander, 2011). It helps the companies to learn more about the customer behaviour, to generate business strategies and to forecast about the possible changes in business environments. It is cost effective and helps the companies to take critical decisions with minimumul efforts.

There are many tools and techniques available for carrying out the data mining activities. Each of the tools has its own benefits and demerits. The user must be aware to select the appropriate tool to perform his task. Different kinds of the data mining tools are: 1) traditional data mining tools 2) dashboards. 3) text-mining tools. Each of these tools has different working style's and updates the results at different times. The techniques used for carrying out data mining are: 1) artificial neural networks. 2) decision trees 3) the nearest neighbour method 4) genetic algorithms 5) rule induction 6) data visualization techniques. The data mining techniques and tools are elaborately discussed in the coming sections.

Data mining techniques have many benefits and few demerits. The benefits involve 1) decision making capability with minimum strain 2) better forecasting 3) cost effective 4) fraud detection. The demerits of data mining are 1) privacy risks for user data 2) misuse of data 3) may not give practical results 4) the accuracy of the results may not be very accurate. All these advantages and disadvantages of the data mining technique's have to be kept in mind before using data mining.

Data mining is largely used by companies, banks, statistician, auditors, government establishments and other agencies for their specifictive purposes. These concerns use data mining for decision making, forecasting, and planning (Microsoft, 2012).
Results obtained through data mining techniques are normally bar graphs, percentage of some variable, ratio of some variable and accuracy percentage (Doug, Alexander, 2011). Different technique provides different type of results. The user has to select the tools based on the required result format.

- **What are the data mining tools and techniques?**
- **What is data mining?**

Data mining is a method to process and analyze a large amount of data to arrive at a meaningful result. Processing and analyzing of large amount of data visually difficult and nearly impossible task. It is normally used by statisticians and business corporate for decision making. Data mining makes use of different tools and techniques to process large data sets (Anderson, 2012). The data mining tasks can explained using an example. A retail shop can use the customer purchase records to determine the future trends and to forecast the customer behavior in future. Data mining technique finds the relationships between various variables of a large database and also displays the pattern in which a variable is varying with respect to time. For example, in case of a retail shop, the data mining method will find out the relationships between the customer standard for the purchase of a particular product and can display the pattern in which the customers a tending to purchase a product with times (Oracle, 2012). It helps in determining the customer purchase power and helps the retail shop to keep stock of the product according to the demands in the coming days. The relationship among the variables as required in data mining is classes, clusters, association and sequential pattern. The major activities taken out during data mining are extraction, transformation loading of the data to the available tool, provide the access capability to the user, and provision of all analysis features to the user and presentation of data in useful format to the user (Anderson, 2012). The challenges associated with the data mining techniques are the complexity of the large data set and the ability of the software to handle the large data set between missing records. While handling large data sets, there are chances that there are missed orwrong information. This wrong information will reduce the accuracy of the results.

Data mining tools and techniques

Various kinds of data mining tools are available for any particular hardware and software. Depending upon the requirement, a particular tool can be selected and put for use. Each of the available tools has its own strength's and weaknesses. It is the duty of the user to select a particular tool. Normally the data mining tools are classified into Traditional data mining tools, Dashboards, and Text-mining tools (Institute of Internal Auditors, 2013). The traditional data mining tools have a set of standard algorithms and mainly finds the pattern and the accuracy of the required parameters (IBM, 2013). These tools do not require any complex hardware or software and can be implemented on a normal personal computer. Dashboards are another type of data mining tool which can directly handle the large data sets and updates the result continuously with the addition of records. It is automated and displays the required parameter and pattern automatically at regular intervals as set by the user (Sean, 2012). This tool is largely used by the corporation to regularly monitor the company's performance. Text mining tools are simple and can take the data in simple formats. It can take Microsoft Word and pdf, e-mails, internet pages, audio data and video data formats. This tool can take data from a large number of formats and convert it into the format as required by the tool. The input data can be typed in a word document in a required format and given as input to this tool. It provides capability to the user to access the data in an easy way. It reduces the user's work to a large extent and provides the user with many features.

There are various techniques to perform the data mining. The normally used techniques are artificial neural

networks, decision trees and, the nearest neighbor method (Institute of Internal Auditors, 2013). The artificial neural technique is applicable for the non-linear data. Some data does not vary in a linear manner with time and fluctuates. The artificial neural network technique is suitable for such data. This technique is normally used to identify fraud activities. The decision tree technique has certain rules, based on which the data is classified. The rules are determined from the tree shaped structures. The relationship among the classified data is then determined and information is extracted. The nearest neighbor technique works on the principle of finding similarity among the neighbor dataset. This technique normally finds a criterion from a record or set of record and searches for the same in all the other available records. All the matching records are clustered together and the data mining is carried out. The nearest neighbor technique is also called as *k*-nearest neighbor technique. Apart from the discussed techniques of data mining there are also other data mining techniques such as using of genetic algorithms, rule induction and data visualization techniques (Sean, 2012). Each of these tools has certain requirement on the hardware and software required by the personal computers. A tool handling huge data needs to have a high configuration computer. Normally, the computers linked to data mining are powerful than the normal computers. All these techniques are less used compared to the discussed techniques.

Advantages and Disadvantages of Data mining

Unlike any other process the data mining has also got its own advantages and disadvantages. The advantages are however more than the disadvantages.

The major advantage of data mining is its ability to forecast on a particular problem. Using the available data the user can find out the change pattern and the repetition in the pattern can be used to forecast about the future (Shane Li/Bowen Shi/Jun Heng Cai, 2012). Data mining increases the decision making capability, it provides a short and informative point about the large data. Data mining is cost effective and the things used to carry out data mining are only a tool and a technique implemented on a personal computer. The user must have the knowledge to operate the tool. Data mining provides short and clear information on a large data. The user needs not to scan the whole data and perform the function manually. Data mining reduces the confusion created by the visual inspection of a large data (Zentut, 2011). Data mining helps in decision making on a number of occasions. In short, it can be said that data mining helps in forecasting, prediction, decision making, arriving at research results and fraud detection (Microsoft, 2012).

The disadvantages associated with data mining are risk on private data security (Shane Li/Bowen Shi/Jun Heng Cai, 2012). The data used for mining will actually contain the company past performance a user personal details. The usage of such data may introduce issues related to privacy (Zentut, 2011). The accuracy of the result is influenced by the correctness of the data and any wrong information may bring down the accuracy. There are chances for misuse of the information. The data mining of a company data by a third party may result in a leaking of company information to external source.

Who use Data mining?

Data mining is used by many people for performing their activities. It helps the marketer and the retailers to maintain the stocks. The retailers can forecast the time in the year for which maximum sales will happen for a product. Depending on the forecasting the retailers can keep stock of items. Data mining can also help to predict the characteristic of a product, for example, a financial institution such as bank can decide upon providing the loan based on the past history of a human community. Based on their loan repaying capacity, the bank can decide whether to give loan or not. Data mining can assist the researchers and statistician to take decisions on a task. Data mining can help them to arrive at a result depending on the available input. Data mining can help the business man and corporate to improve the company revenue by helping them in their decision making. It provides a quick overview of the data and helps the corporate to take decisions on a matter. Data mining helps in fraud detection. Out of huge available data, the unusual information can be detected easily. It helps the government agencies for planning a particular activity, for example, in order to bring about city planning, the agencies can mine the history information about human spread in a particular area and more resources can be provided to the area with dense human population. Data mining also help the government agencies to take decisions in the area of finance and welfare schemes.

Data Mining Results

The data mining results are normally in the form of histogram, bar charts, scatter diagram, normal XY plots, ratio and decision trees (Murat Kantarcıo, Jiashun Jin, Chris Clifton, 2002). While using the data mining software the user inputs the data file in a particular format. The tool takes the data and displays all the possible results. The user has to select the data as per his/her needs.

Conclusion

While doing this work, various aspects of data mining was studies. I have improved my knowledge in the area of data mining. Different tools and techniques associated with data mining were studied. Data mining is useful in many ways and has many benefits. It helps in forecasting, prediction, decision making, arriving at research results and fraud detection. The data mining tools are used by many individuals and agencies such as companies, banks, statistician, auditors and government establishments. In short, it can be said that data mining in cost effective and it simplifies human task to a great extent.

References

Anderson (2012). *Data Mining: What is Data Mining?*. http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm.

Doug Alexander (2011). *Data Mining* http://www.laits.utexas.edu/~anorman/BUS.FOR/course.mat/Alex

Institute of Internal Auditors (2013). *Data Mining Tools & Techniques*Available at: http://www.theiia.org/intAuditor/itaudit/archives/2006/august/data-mining-101-tools-and-techniques

Shane Li/Bowen Shi/Jun Heng Cai (2012). *Advantages and Disadvantages of Data Mining*. [ONLINE] Available at: http://bus237datamining.blogspot.in/2012/11/advantages-disadvantages.html.

Murat Kantarcıo, Jiashun Jin, Chris Clifton, (2002). When do Data Mining Results Violate Privacy?. *UT Dallas Publications*. (), pp.6

Microsoft (2012). *Data Mining Helps You Make Better Decisions*. http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm.

Oracle (2012). *What Is Data Mining?*. http://docs.oracle.com/cd/B28359_01/datamine.111/b28129/process.htm#DMCON002.

Sean (2012). *10 Ways Data Mining Can Help You Get a Competitive Edge*Available at: http://blog.kissmetrics.com/data-mining/.

Zentut (2011). *Advantages and Disadvantages of Data Mining* http://www.zentut.com/data-mining/advantages-and-disadvantages-of-data-mining

IBM (2013). *Data mining techniques*. [ONLINE] Available at: http://www.ibm.com/developerworks/opensource/library/ba-data-mining-techniques/index.html.