

DATA MINING TECHNIQUES AND THEIR APPLICATIONS

1. S.Saranya 2.Rajeshkumar, 3.N.Thangama

2. Assistant Professor, Department of Computer Application

3. Assistant Professor, Department of Computer Science

AJK College of Arts & Science, Coimbatore.

Abstract

Data mining is the process of analyzing data from different views and summarizing it into useful data. “Data mining, also popularly referred to as knowledge discovery from data is the automated or convenient extraction of patterns representing knowledge implicitly stored or captured in large databases, data warehouses, the Web, other massive information repositories or data streams.”. This paper provides a survey on various data mining techniques such as classification, clustering, regression, and summarization and so on. This paper also discusses some of the data mining applications

Keywords: Data Mining, Classification, Regression, Techniques, Applications

1. Introduction

Data mining, discovering of hidden predictive information from large data sets and it is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

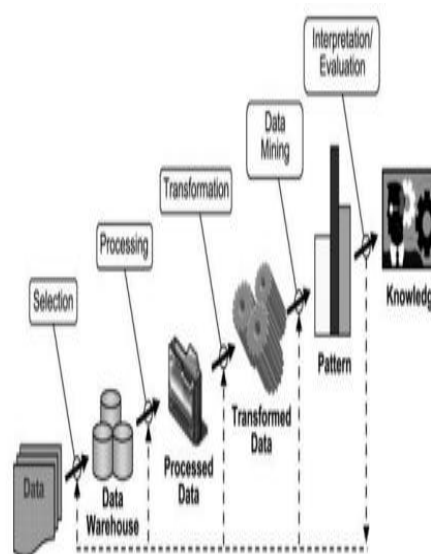


Figure 1: Data Mining Process

In Data mining, Knowledge extraction or discovery is done in seven sequential steps given below: i) Data cleaning: This is the first step to eliminate noise data and irrelevant data from collected raw data. ii) Data integration: At this step, various data



sources are combined into meaningful and useful data. iii) Data Selection: Here, data relevant to the analysis are retrieved from various resources. iv) Data transformation: In this step, data is converted or consolidated into required forms for mining by performing different operations such as smoothing, normalization or aggregation. v) Data Mining: At this step, various clever techniques and tools are applied in order to extract data pattern or rules. vi) Pattern evaluation: At this step, Attractive patterns representing knowledge are identified based on given measures. vii) Knowledge representation: This is the last stage in which, visualization and knowledge representation techniques are used to help users to understand and interpret the data mining knowledge or result.

2. Literature Survey

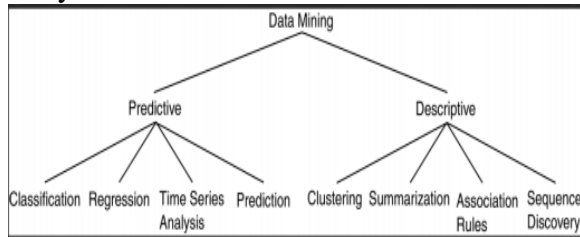
Raju and Sravanthi have studied the issues related to the social networks using web mining techniques. Their analysis focuses highly on the applications of web mining techniques and a general process for social networks. It mainly highlights the difficulties in selecting data samples, finding communities and patterns in social networks and analyzing the overlapping communities. **Rahman** deliberative a systematical data mining approach to mine intellectual knowledge from social data. The author took face book as a primary data source and proposes to use different data mining techniques to analyze the social networking site and other sites. He used K-nearest neighbor (KNN) to classify objects based on samples. **Mosley** discussed the application of correlation, clustering, and association analyses to social media. The author describes how data mining and text analytics can be applied to social media in order to

identify key themes in the data. The issues in terms of accuracy while collecting the data from social media were also highlighted. **Adedoyin-Olowe, Gaber and Stahl** have studied on the techniques that are currently used to analyze SM. It analysis of social media data has proved to be effective, this is so because of the capacity data mining possess in handling noisy, large and dynamic data. According to the authors, in future to mine the data generated on SM currently used and yet-to-be-explored data mining techniques will be used. **Hetherly et.al** explained how deduction on attacks in social networking data to foresee undisclosed information of users through various techniques and projected the efficacy of the techniques through dataset obtained from the Dallas/Fort Worth, Texas network of the Face book social networking application. **S.G.S Fernando** have studied existing web mining techniques used to mine social network data. The paper has specifically focused on the techniques used to mine social network data. Most of the algorithms are developed to mine the structure of the social network where mapping the network to a graph.

3. Data Mining Techniques

Data mining process is extraction of information from large data sets and transforms it into some understandable form for further uses. So it helps to achieve the specific objectives. The goal of a data mining effort is normally either to create a descriptive model or a predictive model. A Descriptive model presents the data in concise form which is essentially a summary of the data points, finds patterns in the data and understands the relationships between attributes represented by the data. The Descriptive model includes tasks such as

Clustering, Association Rules, Summarizations, and Sequence Discovery. The predictive model works by making a prediction about values of data, which uses known results found from different datasets. The Predictive data mining model includes classification, prediction, regression and analysis of time



series as in figure 2.

Figure 2: Data Mining Technique

3.1 Classification

Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. This approach frequently employs decision tree or neural network-based classification algorithms. The common characteristics of classification tasks are as supervised learning, categories dependent variable and assigning new data to one of a set of well-defined classes. Classification technique is used in customer segmentation, modeling businesses, credit analysis, and many other applications. E.g., classify countries based on population, or classify bikes based on mileage.

3.2 Regression:

Regression is another Predictive data-mining model is also known as supervised learning technique. This technique analyzes the dependency of some attribute values, which is dependent upon the values of other

attributes mainly, present in same item. In the regression techniques target value are known. For example, you can predict the child's behavior based on family history.

3.3 Time Series data analysis:

Time-series database uses sequences of values or events obtained over repeated measurements of time. The values are typically measured at equal time interval such as hourly, daily, weekly. A sequence database is any database that consists sequence of ordered events, sometimes having concrete notions of time. For example, Web page traversal sequences and customer shopping transaction sequences are sequence data, but they may not be time-series data.

3.4 Prediction:

This technique discovers the relationship between independent variables and the relationship between dependent and independent variables. The prediction is to predict a future state, rather than a current one. Its applications include obtaining forewarning of natural disasters (flooding, hurricane, snowstorm, etc), epidemics, stock crashes, etc. As another example, the sales volume of computers accessories can be forecasted based on the number of computers sold in the past few months.

3.5 Clustering:

Clustering is a collection of similar data objects. Dissimilar object is another cluster. It is way finding similarities between data according to their characteristic Clustering can be considered as identification of similar classes of objects. By using clustering techniques we can further identify dense and sparse regions in object space and can



discover overall distribution pattern and correlations among data attributes. Classification approach can also be used for effective means of distinguishing groups or classes of object but this method is expensive so clustering can be used as preprocessing approach for attribute subset selection and classification

4. Data Mining Applications

The Data mining applications are widely used in diverse areas such as retail stores, hospitals, banks, and insurance companies. Many domains like health care, finance insurance, retail stores combines the data mining applications with statistics, pattern recognition, and other important tools to perform data analytics. Data mining is used primarily for decision making.

4.1 Medicare and health care:

Data mining in medicine enables to characterize patient activities to see incoming office visits. Data mining helps identify the patterns of successful medical therapies for different illnesses.

4.2 Education:

Educational Data Mining is a blooming field which provides knowledge from educational Environment data. The goals of EDM are identified as predicting students' learning behavior, emotions and skills. This study improves the educating methods by understanding the ward and to take accurate decisions respectively

4.3 Market Basket Analysis:

Market basket analysis is a technique that uses association rule mining to understand the purchasing behavior of the customer. It also allows the seller to understand his

business, customer's needs and to make profitable change accordingly.

4.4 Financial Banking:

Data mining can contribute to solving business problems in banking and finance by finding patterns, causalities, and correlations in business information and market prices. The managers may find this information for better segmenting, targeting, acquiring, retaining and maintaining a profitable customer.

4.5 Research Analysis:

Data mining is very useful in data pre-processing and integration of databases. Data mining allows the researchers to identify co-occurring sequences and the correlation between any activities .Data visualization and visual data mining help the researcher with a clear view of the data.

4.6 Fraud Detection:

The traditional fraud detection methods are expensive, time consuming and complex. Data mining aids in providing meaningful patterns and turning data into information. Valid and useful information is called as knowledge. The results are categorized into fraudulent or non-fraudulent.

4.7 Transportation:

Data mining helps determine the distribution schedules among warehouses and outlets and analyze loading patterns.

4.8 Agriculture:

Data mining is emerging technology in agriculture field for crop yield analysis with respect to four parameters namely year, rainfall, production and area of sowing.



Yield prediction is a very important agricultural problem that remains to be solved based on the available data. The yield prediction problem can be solved by employing Data Mining techniques such as K Means, K nearest neighbor (KNN), Artificial Neural Network and support vector machine etc. Cloud Computing: Data Mining techniques are used in cloud computing. The implementation of data mining techniques through Cloud computing will allow the users to retrieve meaningful information from virtually integrated data warehouse that reduces the costs of infrastructure and storage. Cloud computing uses the Internet services that rely on clouds of servers to handle tasks. The data mining technique in Cloud Computing helps to perform efficient, reliable and secure services for their users.

Conclusion

According to the techniques of data mining listed above, it is learned that this is a powerful and essential technique for performing manipulation of data that is data mining gives proper and targeted outcome from large and vastly growing data worldwide. This paper discusses the idea of data mining, the process of data mining, different techniques such as clustering, association, classification, prediction and so

We also discussed some insights of the data mining applications.

References:

- [1] Aarti Sharma et al, "Application of Data Mining – A Survey Paper", International Journal of Computer Science and Information technologies, Vol. 5 (2), 2014.
- [2] Smita, Priti and Sharma, "Use of Data Mining in Various Field: A Survey Paper" IOSR Journal of Computer Engineering, 8727 Volume 16, Issue 3, Ver. V (May-Jun. 2014)
- [3] Brijesh Kumar Baradwaj, Saurabh Pal" Mining Educational Data to Analyze Students Performance" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011
- [4] J. Han and M. Kamber. "Data Mining, Concepts and Techniques", Morgan Kaufmann, 2000.
- [5] Nikita Jain, Vishal Srivastava "DATA MINING TECHNIQUES: A SURVEY PAPER" IJRET: International Journal of Research in Engineering and Technology, Volume: 02 Issue: 11 | Nov-2013.
- [6] Prof. Dr. Wolfgang Karl Hardle," Time Series Data Mining Methods: A Review", Berlin, March 25, 2015.
- [7] Pradnya P. Sondwale, "Overview of Predictive and Descriptive Data Mining Techniques" IJARCSSE, Volume 5, Issue 4, April 2015
- [8] Data Mining: Concepts and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management Systems) 3rd Edition