A Survey : Use of Data Mining in Various Field

Ms. Mohini H. Chandwani[1] Dr. Anjali B. Raut[2]

[1]ME Student at HVPM College of Engineering & Technology, Sant Gadge Baba Amravati University, Amravati, Maharashtra, India

[2]Received PhD, Professor and Head of CSE Department HVPM College of Engineering & Technology, Sant Gadge Baba Amravati University, Amravati, Maharashtra, India

ABSTRACT

Data mining is to extract the knowledge/information from a large amount of data which stores in multiple heterogeneous databases. Knowledge /information are conveying the message through direct or indirect. This paper provides a survey of various data mining techniques. These techniques include association, correlation, clustering and neural network. This research paper also conducts a formal review of the application of data mining such as the education sector, marketing, fraud detection, manufacturing, and telecommunication. This paper discusses the topic based on past research paper and also studies the data mining techniques.

Keywords : Association, Clustering, Data mining, data mining application, knowledge discovery database.

I. INTRODUCTION

In the real world, huge amount of data are available in education, medical, industry and many other areas. Such data may provide knowledge and information for decision making. For example, you can find out drop out student in any university, sales data in shopping database. Data can be analyzed, summarized, understand and meet to challenges.[1]

Data mining is a powerful concept for data analysis and process of discovery interesting pattern from the huge amount of data, data stored in various databases such as data warehouse , world wide web , external sources .Interesting pattern that is easy to understand, unknown, valid ,potential useful. Data mining is a type of sorting technique which is actually used to extract hidden patterns from large databases. The goals of data mining are fast retrieval of data or information, knowledge Discovery from the databases, to identify hidden

patterns and those patterns which are previously not explored , to reduce the level of complexity, time saving, etc[2].

Data mining refers extracting knowledge and mining from large amount of data. Sometimes data mining treated as knowledge discovery in database (KDD)[3] . KDD is an iterative process, consist a following step shown in Figure1 [4].
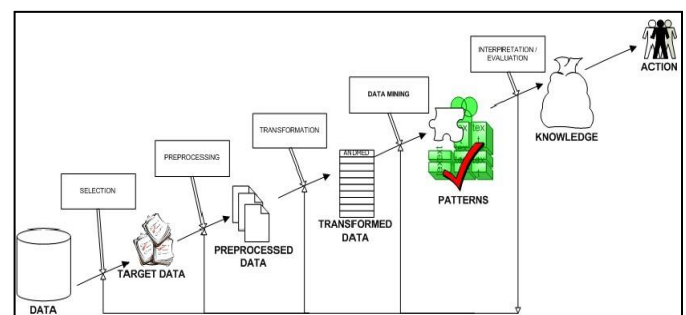

Figure 1. Knowledge data mining

- Selection: select data from various resources where operation to be performed.

- Preprocessing: also known as data cleaning in which remove the unwanted data.
- Transformation: transform /consolidate into a new format for processing.
- Data mining: identify the desire result.
- Interpretation / evaluation: interpret the result/query to give meaningful report/information

Various algorithms and techniques like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc., are meant for knowledge discovery from databases [5]. The main objective of this paper learns about the data mining. And the rest of this Section 2 discusses data mining models and techniques. Section 3 explores the application of data mining. Finally, we conclude the paper in Section 4.

II. METHODS AND MATERIAL

1. Data Mining Techniques

Data mining means collecting relevant information from unstructured data. So it is able to help achieve specific objectives. The purpose of a data mining effort is normally either to create a descriptive model or a predictive model .A descriptive model presents, in concise form, the main characteristics of the data set. The purpose of a predictive model is to allow the data miner to predict an unknown (often future) value of a specific variable; the target variable [7]. The goal of predictive and descriptive model can be achieved using a variety of data mining techniques as shown in figure 2[8].
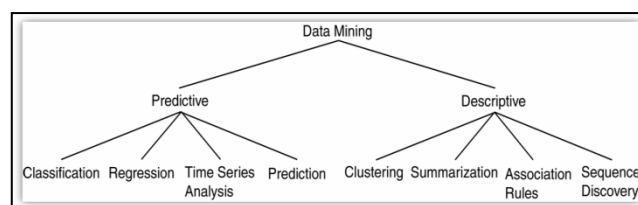

Figure 2. Data Mining Models

Classification: Classification based on categorical (i.e. discrete, unordered) . This technique based on the supervised learning (i.e. desired output for a given input is known) .It can be classifying the data based on the training set and values (class label). These goals are achieve using a decision tree, neural network and classification rule (IF- Then).for example we can apply the classification rule on the past record of the student who left for university and evaluate them. Using these techniques we can easily identify the performance of the student.

Regression: Regression is used to map a data item to a real valued prediction variable [8]. In other words, regression can be adapted for prediction. In the regression techniques target value are known. For example, you can predict the child behavior based on family history.

Time Series Analysis: Time series analysis is the process of using statistical techniques to model and explain a time-dependent series of data points. Time series forecasting is a method of using a model to generate predictions (forecasts) for future events based on known past events [9]. For example stock market.

Prediction: It is one of a data mining techniques that discover the relationship between independent variables and the relationship between dependent and independent variables [4].Prediction model based on continuous or ordered value.

Clustering: Clustering is a collection of similar data object. Dissimilar object is another cluster. It is way finding similarities between data according to their

characteristic. This technique based on the unsupervised learning (i.e. desired output for a given input is not known). For example, image processing, pattern recognition, city planning.

Summarization: Summarization is abstraction of data. It is set of relevant task and gives an overview of data. For example, long distance race can be summarized total minutes, seconds and height.

Association Rule: Association is the most popular data mining techniques and fined most frequent item set. Association strives to discover patterns in data which are based upon relationships between items in the same transaction. Because of its nature, association is sometimes referred to as "relation technique". This method of data mining is utilized within the market based analysis in order to identify a set, or sets of products that consumers often purchase at the same time [6].

Sequence Discovery: Uncovers relationships among data [8]. It is set of object each associated with its own timeline of events. For example, scientific experiment, natural disaster and analysis of DNA sequence.

III. RESULTS AND DISCUSSION

Data Mining Application

Various field adapted data mining technologies because of fast access of data and valuable information from a large amount of data. Data mining application area includes marketing, telecommunication, fraud detection, finance, and education sector, medical and so on. Some of the main applications listed below:

Data Mining in Education Sector: We are applying data mining in education sector then new emerging field called "Education Data Mining". Using these term enhances the performance of student, drop out student, student behavior, which subject selected in

the course. Data mining in higher education is a recent research field and this area of research is gaining popularity because of its potentials to educational institutes. Use student's data to analyze their learning behavior to predict the results [10].

Data Mining in Banking and Finance: Data mining has been used extensively in the banking and financial markets [11]. In the banking field, data mining is used to predict credit card fraud, to estimate risk, to analyze the trend and profitability. In the financial markets, data mining technique such as neural networks used in stock forecasting, price prediction and so on.

Data Mining in Market Basket Analysis: These methodologies based on shopping database. The ultimate goal of market basket analysis is finding the products that customers frequently purchase together. The stores can use this information by putting these products in close proximity of each other and making them more visible and accessible for customers at the time of shopping [12].

Data Mining in Earthquake Prediction: Predict the earthquake from the satellite maps. Earthquake is the sudden movement of the Earth's crust caused by the abrupt release of stress accumulated along a geologic fault in the interior. There are two basic categories of earthquake predictions: forecasts (months to years in advance) and short-term predictions (hours or days in advance) [13].

Data Mining in Bioinformatics: Bioinformatics generated a large amount of biological data. The importance of this new field of inquiry will grow as we continue to generate and integrate large quantities of genomic, proteomic, and other data [4].

Data Mining in Telecommunication: The telecommunications field implement data mining technology because of telecommunication industry have the large amounts of data and have a very large

customer, and rapidly changing and highly competitive environment. Telecommunication companies uses data mining technique to improve their marketing efforts, detection of fraud, and better management of telecommunication networks [4].

Data Mining in Agriculture: Data mining than emerging in agriculture field for crop yield analysis a 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 (SVM) [14].

Data Mining in 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 [15].Cloud computing uses the Internet services that rely on clouds of servers to handle tasks. The data mining technique in Cloud Computing to perform efficient, reliable and secure services for their users.

IV. CONCLUSION

This paper provides a general idea of data mining, data techniques and data mining in various fields.

The main objectives of data mining techniques are to discover the knowledge from active data. These applications use classification, Prediction, clustering, Association techniques and so on. Hopefully in future work we review various classifications and clustering algorithm and its significance.

V. REFERENCES

[1]. Yongjian Fu " data mining: task, techniques and application"

[2]. Er. Rimmy Chuchra "Use of Data Mining Techniques for the Evaluation of Student Performance:A Case Study" International Journal of Computer Science and Management Research Vol 1 Issue 3 October 2012

[3]. J. Han and M. Kamber. "Data Mining, Concepts and Techniques", Morgan Kaufmann, 2000.

[4]. Aakanksha Bhatnagar, Shweta P. Jadye, Madan Mohan Nagar" Data Mining Techniques & Distinct Applications: A Literature Review" International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 9, November- 2012

[5]. 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

[6]. Data mining white paper, www.ikanow.com

[7]. 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,

[8]. Dr. M.H.Dunham, "Data Mining, Introductory and Advanced Topics", Prentice Hall, 2002.

[9]. Umamaheswari. K, S. Niraimathi "A Study on Student Data Analysis Using Data Mining Techniques" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 8, August 2013

[10]. Industry Application of data mining , http://www.pearsonhighered.com/samplechapter/0130862711.pdf

[11]. David L Olson, Dursun Delen " Advance data minig techniques" springer 2008

[12]. G. V. Otari, Dr. R. V. Kulkarni, "A Review of Application of Data Mining in Earthquake Prediction" G. V. Otari et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (2) , 2012,3570-3574

[13]. D Ramesh , B Vishnu Vardhan, "Data Mining Techniques and Applications to Agricultural Yield Data" International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 9, September 2013

[14]. Ruxandra-Ştefania PETRE, "Data mining in Cloud Computing" Database Systems Journal vol. III, no. 3/2012

[15]. Bhagyashree Ambulkar and Vaishali Borkar, "Data Mining in Cloud Computing", MPGI National Multi Conference 2012 (MPGINMC-2012), 7-8 April 2012.

Author Profile

	MS. Mohini H. Chandwani. Pursuing M.E. from HVPM COET Amravati.
	DR. Anjali B. Raut. received B.E, M.E and Phd, and at present working as Professor and Head of CSE Department HVPM COET Amravati**.**