IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Volume-6, Issue-4 E-ISSN: 2347-2693

Comparative Analysis of Data Mining Techniques

Shaganpreet kaur[1*], Chinu[2]

[1*] Dept. of C.S.E, Baba Farid College of Engineering and Technology, Bathinda, Punjab
[2] Dept. of C.S.E, Baba Farid College of Engineering and Technology, Bathinda, Punjab

[*]*Corresponding Author: verma.chinu92@gmail.com , Tel.: 9501115498*
Available online at: www.ijcseonline.org

Received: 23/Mar/2018, Revised: 31/Mar/2018, Accepted: 21/Apr/2018, Published: 30/Apr/2018

Abstract Data mining is the area of research, which means that useful information or knowledge is extracted from previous data. Data mining defines large amounts of data as a process of finding information such as super market data for various technologies used for data mining, such as science, research, medicine, media, web, entertainment and many other areas, which is implemented with various goods, data mining model data warehouses and online analytical resources. Data mining has made a immense advancement in recent year but the problem of lost data has remained a big challenge for data mining algorithms. This paper analyzed the predictive and descriptive techniques such as classification, regression time series analysis ,predication and clustering, summarization, association rules, sequence discovery techniques on the basis of algorithms which is used to predict previously unidentified class of objects.

Keywords: Data mining, Data mining techniques: Predictive and Descriptive DM techniques.

I. INTRODUCTION

Data mining is exploring hidden information from huge data sets [1]. In this, needs to mainly focus on washing out the data so as to build it feasible for additional processing. [4].

In various domains, data mining is widely used such as insurance, banking, retail, research, astronomy, medicine, forecast of rainfall and Government security [2] Text Data Mining (Chakraborty, 2000) is a research area that includes many research areas, such as natural language processing, machine learning, information retrieval (Salton, 1989) and data mining [3]. with the large range of Data Mining technique, information or forms of data presentation it is essential to describe the restrictions of the relevance of certain methods according to the achieved objectives or provided data. It is also important to understand how the problem must be solved with the Data Mining such as clustering, classification, regression and so on[7]. The development of information communication technology has generate a huge quantity of data from dissimilar sources, that is stored in dissimilar geological locations. Every database may possibly have its individual structure to store data. multiple mining data sources dispersed at different geological location to determine helpful patterns are significant important for decision making. Data is irrelevant to each other from different Sources. Information that generate from dissimilar sources is integrated, fresh and helpful facts may appear [8].

This paper primarily focus on the predictive and descriptive techniques such as classification, regression, time series analysis ,predication and clustering, summarization, association rules, sequence discovery techniques on the basis of algorithms which is used to predict previously unidentified class of objects. Section I contains brief introduction to Data mining. Section II contains Data mining techniques and in section III Applications and comparative analysis of descriptive and predictive techniques.

II. DATA MINING TECHNIQUES

The data mining process converts information from large data sets and transforms it into some sensible form for additional uses. So it helps in achieving specific objectives. The goal of a data mining effort is usually to create a descriptive model or predictive model.

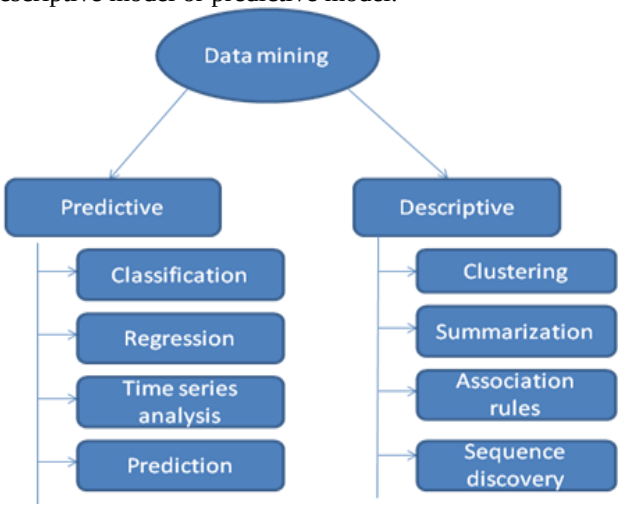

Fig 1: Data mining Techniques

Predictive and Descriptive DM techniques:

Data mining is classified as primarily descriptive and predictive techniques, as shown in Figures 1, so that in order to fit a model for data, its various functions can be achieved. A predictive model predicts future values using different and historical data. The prediction models include classification, regression, time series analysis and forecast. A descriptive model recognizes hidden patterns or relationships in data. It explores the properties of the data being examined. Descriptive models include clustering, abbreviations, union rules, and series discovery [5].

Types of predictive techniques:-

- **Classification:** This is the main data mining technique used, to develop a model that maintains a set of pre-classified examples that can categorize the population of large records. It only used decision tree or neural network based classification algorithms. The general characteristics of classification work are classified as surveillance, providing new data to one of the categories dependent variables and well-defined classes. Classification techniques are used in customer segmentation, modeling business, credit analysis and many other applications. For example, classify countries classified by population, or classify bikes based on mileage.
- **Regressio**n: Regression another estimator data-mining model is also known as supervised learning technology. This technique analyzes the dependency of certain attribute values, which depends on the values of other characteristics, which are present in the same item. The purpose value of regression technique is known. For example, you can estimate child's behavior based on family history.
- **Time Series data analysis**: The time-series database uses the sequence of values or events received from the repeated capacity of the time. Prices are generally calculated at the same time interval as per hourly, daily, weekly. A series database is a chain of events in any database, sometimes the actual thoughts of time. For example, web page crossing sequences and customer purchase transactions are sequence data, but they can not be time series data.

- **Prediction:** This technique reveals the relationship between independent variable and relationships between dependent and independent variables. Predicting is to predict the future situation, rather than getting a notification for natural disasters (floods, hurricanes, ice storms, etc.), pandemics, stock crashes etc in its applications. As another example, the amount of sales of computer accessories can be estimated based on the number of computers sold in the last few months [1].

Types of Descriptive techniques:-

- **Clustering:** Clustering is a collection of identical data objects. Different object is another cluster; it is searching similarities between the data according to their features. Clustering can be measured as identifying objects of the same classes. Using grouping techniques, we can identify solid and rare areas in the object space and take into account the general distribution pattern and the correlations in the data attribute. The classification approach can also be used for efficient groups of categories of unique groups or objects, but this method is expensive so the use of clustering can be done in the form of pre-processing approach to the specialty subset selection and classification. For example, image processing, pattern recognition, city planning astronomy - the collection of stars, galaxies, or super galaxies.
- **Summarization**: Summary is referred to as the intangible or generalization of data. Summarized technique map data in subset with simple description. The summarized data set gives an overview of all the objectives of the data with the collected information. Normal methods apply to normal and general deviation to analyze statistics, automated report generation and data visualization. For pre; Length can be measured as meter, centimeter or millimeter.
- **Association**: Association technology is used to remove the relations between properties and objects. In this technique, the occurrence of another model from the occurrence of a model means i.e. the object is due to the other and related causes-and-effect. It is common for establishing a form of mathematical relationships between various mutually dependent variables of data mining; Association rules are useful for analyzing and estimating client performance. They also play an important role in the data analysis of the shopping cart, the grouping of products and the design of the catalog and the design of the store. Federation rules being created by the programming system can be used to enable the machine to learn.
- **Sequence Discover**y: Highlight the relationship between data. It is a set of everything related to its individual timeline of events, for example, analysis of scientific experimentation, natural disaster and DNA sequence [1]

III. APPLICATION OF DATA MIING[6]

Medicare and health care:- Using technical data mining, facilitating the relation between the disease, the treatment efficiency, the ordering of new drugs, the market activity in the drug distribution services etc**.**

Aid to marketing or retailing: It helps the marketers directly by providing a valuable and accurate trend on the purchasing activities of their customers and also helps them to predict what products their customers may be interested in.

Market basket analysis (MBA):Actually, it helps to understand the data mining technique that according to the rules of the goods, what are the possibilities of buying together, mainly for the purpose of identifying cross-selling opportunities. Sometimes it is also known as product affinity analysis.

Intrusion detection: This is a passive approach to security because it monitors the information system and detects security violations when the alarm raises. To ascertain the signals of security procedures, this process inspects and analyzes events that occur in the computer system.

DNA Analysis: Normalized pattern patterns of each class compare the comparison between parallel search and DNA sequences (eg, sick and healthy) identify the patterns of the gene chain that play a role in various diseases Association Analysis: Cooperative DNA sequences recognise.

E-commerce: This is also the future domain for data mining. This is ideal because many materials needed for successful data mining are easily available: Data accounts are in abundance, electronic collection provides reliable data, insights can easily change into action, and return on investment can be measured**.**

Techniques	Predictive	Descriptive
1	A predictive model predicts future values using different and historical data.	A descriptive model recognizes hidden patterns or relationships in data.
2	The main motive behind the predicted data mining is to prepare the model, which is to work for example assessment and classification.	The main purpose of descriptive data mining is to get an understanding of the analysis system by highlighting relationships and patterns in large data sets.
3	Based on data and analysis, the forecast record builds the model for record, and predicts the movement and properties of unknown data.	Set the model or work important data in small, summary, educative differential forms.
4	For example, optimizing customer relationship management systems using predicted analytics.	For example, the plan checks historical power handling data to help power, and electric companies allow descriptive analytics to determine the best possible price.
5	The prediction models include classification, regression, time series analysis and forecast.	Descriptive models include clustering, abbreviations, association rules and series discovery

IV. CONCLUSION

The most challenging task in the field of data mining is extraction of useful information from large amount of data and mining of data from multiple sources. This research work mainly focus on the analysis and comparison of data mining techniques. With the continuous development of data mining technology, research in multiple data sources mining is becoming important task. It has a large range of applications in the fields such as robotics, automation and intelligent system design.

REFERENCES

[1]. Deepashri.K.S Asst. Professor, Dept. of IS&E Adichunchangiri Institute of Technology, Chikmaglur, Karnataka, India 'Survey on Techniques of Data Mining and its Applications', International Journal of Emerging Research in Management &Technology , ISSN: 2278-9359 (Volume-6, Issue-2).

[2]. Radhakrishnan Gopalapillaia* , Deepa Guptab, Sudarshan TSBa, 'Pattern Identification of Robotic Environments using Machine Learning Techniques', 7th International Conference on Advances in Computing & Communications, ICACC-2017, 22-24 August 2017, Cochin, India.

[3]. Eunseog Youn a, Myong K. Jeong b,'Class dependent feature scaling method using naive Bayes classifier for text datamining', Article history: Received 18 September 2007 Received in revised for 1August 2008 Available online 24 December 2008.

[4].Mansi Gera, Shivani Goel," Data Mining-Techniques,Methods and Algorithms:A Review on Tools and their Validity", International Journal of Computer Applications (0975 – 8887)Volume 113–No. 18, March 2015.

[5]. Ayman E. Khedra, Mona Kadryb, Ghada Walidb, 'Proposed framework for implementing data mining techniques to enhance decisions in agriculture sector Applied case on Food Security Information Center Ministry of Agriculture', Egypt International Conference on Communication, Management and Information Technology (ICCMIT 2015).

[6]. Vikas Gupta, Prof. Devanand 'A survey on Data Mining: Tools, Techniques, Applications, Trends and Issues', International

Journal of Scientific & Engineering Research Volume 4, Issue3, March-2013 1ISSN2229-5518.

[7].Kochetov Vadim National Research Nuclear University MEPhI(Moscow Engineering Physics Institute),Moscow, Russian Federation Ko4etovvadim@gmail.com,"Overview of different approaches to solving problems of Data Mining", Procedia Computer Science 123 (2018)234–239.

[8].Ruili Wang etl.,"Review on mining data from multiple data sources", January 27, 2018.

