RESEARCH ARTICLE   OPEN ACCESS

Different Mining Techniques for Data Mining

Fatima [1], Dr. Jawed Ikbal Khan [2]
Research Scholar [1], Magadh University Bodh Gaya,
Assistant Professor [2], Deptt. of Mathematics, Mirza Ghalib College
Gaya, Bihar, India

ABSTRACT
Data mining is a process which finds functional patterns from huge amount of data. It is powerful new machinery with great potential to help companies focus on the most important information in their data warehouses. It uses machine learning, statistical and visualization technique to discover and present knowledge in a form which is easily comprehensible to humans. This review of literature focuses on data mining techniques, issues, tools, and applications. In this paper we have focused a variety of techniques, approaches and different areas of the research which are helpful and rights as the important field of data mining technology.
Keywords:- Data Mining

I. INTRODUCTION

Data mining is the computing process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. It is an interdisciplinary subfield of computer science. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, terestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD. [1] It would be a great benefit for customers if these reviews could be processed automatically for sentiment analysis and presented in a summarized way. Text mining is a method which is used in different fields like machine learning, information retrieval, statistics and computational linguistics. Web mining is a sub discipline of text mining used to mine the semi structured web data in form of Web content mining, Web structure mining and Web usage mining. [2] The term is a misnomer, because the goal is the extraction of patterns and knowledge from large amounts of data, not the extraction (mining) of data itself. It also is a buzzword and is frequently applied to any form of large-scale data or information processing (collection, extraction, warehousing, analysis, and statistics) as well as any application of computer decision support system, including artificial intelligence, machine learning, and business intelligence. The book Data mining: Practical machine learning tools and techniques with Java (which covers mostly machine learning material) was originally to be named just Practical machine learning, and the term data mining was only added for marketing reasons. Often the more general terms (large scale) data analysis and analytics – or, when referring to actual methods, artificial intelligence and machine learning – are more appropriate. The actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown, interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection), and dependencies (association rule mining, sequential pattern mining). This usually involves using database techniques such as spatial indices. These patterns can then be seen as a kind of summary of the input data, and may be used in further analysis or, for example, in machine learning and predictive analytics. For example, the data mining step might identify multiple groups in the data, which can then be used to obtain more accurate prediction results by a decision support system

II. RELATED WORK

In [3] Velmurgun T. et al. attempted to analyze performance of K-means and Fuzzy C-means clustering techniques in the field of data mining. The performance compared on the basis of clustering result quality. In[3] Kavitha P., T. Sasipraba evaluated the performance of distributed data mining framework on Java platform. Association rule mining was used for discovering interesting patterns from a large amount of data. In [5] Yujie Zheng proposed a methodology for clustering in data mining to improve the standard of higher education used to find data segmentation and pattern information. In[6] M.Sukanya et al. used classification and clustering algorithms of data mining for the performance improvement in education sector. By using these algorithms an educational institute could predict the number of enrolled students. In [7] Manoj Bala et al. applied an application of data mining in educational institute to extract the useful information from the huge dataset and provided analytical tool to view and used this information for decision making process. They also conducted a research on student learning result based on data mining. In [8] Hamidah Jantan used a potential classification technique for academic talent forecasting in higher educational institutes. He proposed

a classification model to increase the academic talent in higher educational institutions. In[9] Tai Chang Hsia applied data mining techniques to analyze the course preferences and course completion rates of enrollees in extension education courses at a University in Taiwan. Some of the data mining algorithms like decision tree, link analysis, and decision forest were used for further analysis.

III.   DATA MINING TECHNIQUE

Data mining adopt its technique from many research areas, including statics machine learning, database systems, rough sets, visualization and neural networks.

A. Statistical Approach Statistical models are built from a set of training data. Many statistical tools have been used for data mining including, Bayesian network, correlation analysis, regression analysis and cluster analysis.

B. In the Bayesian network nodes represents states or variable while edges represents dependencies between nodes. From figure we can see that rush hour, bad weather or accident affect the traffic which in turn causes traffic jam.

C. Machine Learning Approach The most common machine learning methods used for data mining include conceptual learning, inductive concept learning and decision tree induction. By following the path from root to leaf node an objects class can be determine by decision tree.

S No	Data Mining Task	Data Mining Algorithms
1	Classification	Decision Tree, Rule Based, Neural Network, Naive Bayes, Support Vector Machine
2	Clustering	K-Means
3	Regression and Prediction	Support Vector Machine, Decision Tree
4	Association	Associations Rule Mining (ARM)
5	Summarization	Multivariate Visualization

Table 3.1 Data Mining Uses Different Algorithms

D. Decision trees are induced from the training set and decision trees give classification rules. A simple decision tree, it determines the car's mileage from its size, transmission type and weight. A simple decision tree [10] from decision tree we can conclude, for example, large size; heavy weight car will have low mileage. Nodes represent three classes of mileage

IV.   DATA MINING APPLICATION

Various fields use data mining technologies because of fast access of data and valuable information from vast amount of data. Data mining technologies have been applied successfully in many areas like marketing, telecommunication, fraud detection, and finance, medical and so on. Some of the application is listed below.

A. Financial Data Analysis The financial data in banking and financial industry is generally reliable and of high quality which facilitates the systematic data analysis and data mining. Here are the few typical cases: Design and construction of data warehouses for multidimensional data analysis and data mining. Loan payment prediction and customer credit policy analysis. Classification and clustering of customers for targeted marketing. Detection of money laundering and other financial crimes.

B. Retail Industry Data Mining has its great application in Retail Industry because it collects large amount data from on sales, customer purchasing history, goods transportation, consumption and services. It is natural that the quantity of data collected will continue to expand rapidly because of increasing ease, availability and popularity of web. The Data Mining in Retail Industry helps in identifying customer buying patterns and trends. That leads to improved quality of customer service and good customer retention and satisfaction. Here is the list of examples of data mining in retail industry:
  a) Design and Construction of data warehouses based on products, time and region.
  b) Analysis of effectiveness of sales campaigns.
  c) Customer Retention.
  d) Product recommendation and cross-referencing

C. Telecommunication Industry Today the Telecommunication industry is one of the most emerging industries providing various services such as fax, pager, cellular phone, Internet messenger, images, email, web data transmission etc. Due to the development of new computer and communication technologies, the telecommunication industry is rapidly expanding. This is the reason why data mining is become very important to help and understand the business. Data Mining in Telecommunication industry helps in identifying the telecommunication patterns, catch fraudulent activities, make better use of resource, and improve quality of service. Here is the list examples for which data mining improve telecommunication services as Multidimensional Analysis of Telecommunication
  a) Fraudulent pattern analysis.
  b) Identification of unusual patterns.
  c) Multidimensional association and sequential patterns analysis
  d) Mobile Telecommunication services.
  e) Use of visualization tools in telecommunication data

D. Biological Data Analysis Now a days we see that there is vast growth in field of biology such as genomics, proteomics, functional Genomics and biomedical research.

Biological data mining is very important part of Bioinformatics.

E. Other Scientific Applications the applications discussed above tend to handle relatively small and homogeneous data sets for which the statistical techniques are appropriate. Huge amount of data have been collected from scientific domains such as geosciences, astronomy etc. There is large amount of data sets being generated because of the fast numerical simulations in various fields such as climate, and ecosystem modelling, chemical engineering, fluid dynamics etc. Following are the applications of data mining in field of Scientific Applications.

V. CONCLUSIONS

Data miming is a computing process. Which is bring into play in different area and we are used different techniques in Data mining process. In Data mining applications Financial, Retailer, Telecommunication, Biological and other scientific applications. In our future work, we have more exhaustive various techniques in data mining such machine learning, support vector machine, naive bayes etc.

REFERENCES

[1] https://en.wikipedia.org/wiki/Data_mining
[2] Safdar, Md, and Md Jawed Ikbal Khan. "Opinion Mining For Customer Feedback: A Survey."
[3] Velmurugan T. et al., "Performance Evaluation of KMeans & fuzzy C-means Clustering Algorithm for Statistical Distribution of Input Data Points", European Journal of Scientific Research, Vol. 46, 2010.
[4] Kavitha P., T. Sasipraba, "Performance Evaluation of Algorithms using a Distributed Data Mining Framework based on Association Rule Mining", International Journal on Computer Science & Engineering (IJCSE), 2011
[5] Yujie Zheng, "Clustering Methods in Data Mining with its Application in Higher Education", International Conference on Education Technology and Computer, Vol. 43, 2012, IACSIT Press, Singapore.
[6] M.Sukan et al., "Data Mining: Performance Improvement in Education Sector using Classification and Clustering Algorithm", International Conference of Computing and Control Engineering (ICCCE) 12-13 April, 2012.
[7] Manoj Bala et al., "Study of Application of Data Mining Technique in Education", International Journal of Research in Science and Technology, Vol. No. 1, Issue No. IV, Jan-March, 2012.
[8] Hamidah Jantan, "Classification and Prediction of Academic Talent using Data Mining Technique", 14 International Conference on Knowledge based and Intelligent Information and Engineering Information pages 491-500.
[9] Tai Chang Hsia, "Course Planning of Extension Education to meet Market Demand by using Data Mining Techniques", Expert System with Applications: An International Journal, Vol.34, Issue-1, Jan 2008.
[10] M.S.Chen,J.Han,and P.S.Yu. Data Mining: An Overiew from a database Prespective. IEEE Transactions on Knowledge and Data Engineeirng, Vol.8, pp.866-883, 1996.

.