Study on Data Mining Approaches and Applications

Akhila S[1], Madhurima Reddy[2], Durga Bhavani[3], Boni Manasa[4]
[1]Asst.proffessor
[1, 2, 3, 4] Megha Institute of Engg & Technology for Women Edulabad,ghatkesar,rangareddy dist, telangana

Abstract- *Data mining is a technique that derives useful information or knowledge from very large unstructured and raw data. Data mining is further classified into two categories predictive data mining and descriptive data mining. Predictive data mining speculates attributes of value based on previous attributes. Descriptive data mining gives outline of hidden relationships between data. There are several approaches to achieve the goal of data mining such as clustering, classification, time series analysis, association rules, sequence discovery and etc. With its diversification of applicability data mining can be applied to solve many real world problems in various domains such as education, healthcare, weather, social media, manufacturing, smart cities, internet of things, e-governance, insurance and cyber security etc. The main objective of this paper is to present the overview of various data mining techniques and its application among different sectors. The data used in this paper is purely secondary data collected from various open access journals, online articles, and Google scholar. .*

Keywords- Data Mining, Applications of Data Mining, Data Mining Techniques, Clssification, Clustering.

I. INTRODUCTION

Data mining or knowledge discovery is about finding of hidden useful knowledge from very huge databases or data sets. This technology plays very significant role in finding of most useful information from data warehouses for most of the enterprises. Data mining examines data from several prospects and summarizes so that the resultant outcome would be useful information, which can be used by various application domains. Data mining software is an analytical tool that analyzes data from various extents or viewpoints, and identifies the hidden relationships. In Technical point of view it is technique of discovering patterns between numbers of attributes from vary huge relational data repositories. Data mining methods are operates on immerse volumes of data to find hidden patterns and relationships which are useful of decision making process. A typical data mining process is shown below (i.e. in figure1). The initial task in data mining process is data cleaning which removes the noisy and unwanted data from collected raw data. Second task in this procedure is data integration, amalgamates relevant, useful and meaningful data. Next task is data selection, as part this

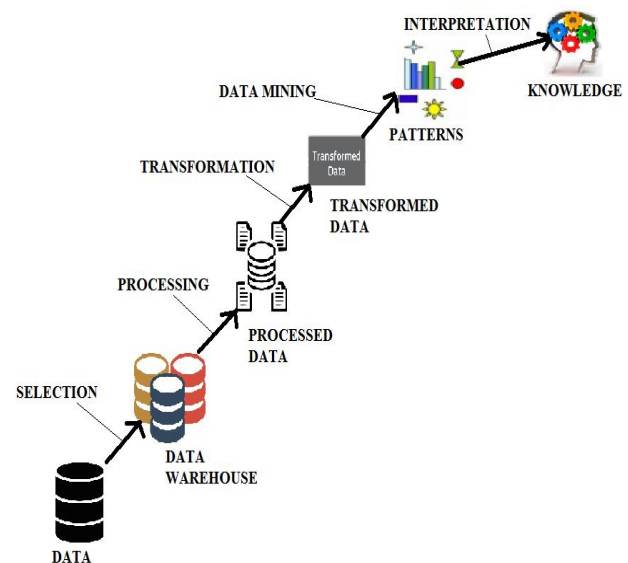data that is needful and relevant to analyze is selected from data sources.

Figure 1. Data mining Process

After data selection, a task of amalgamating and transforming of into a much needed forms used for mining process done by various methods like data normalization or aggregation is performed, this process is called data transformation. Next task is to perform data mining process on the transformed data with the help of various tools and approaches to extract desired patterns. After that pattern evaluation task is performed, identifies patterns that are needed to represent desired knowledge. After pattern evaluation task last task in this process is performed, knowledge representation i.e. information representation and visualization methods are applied on the data generated by previous task. The knowledge produced by knowledge representation task is useful for users in order to understand and elucidates the final outcome of data mining process. Data mining has its applicability various real world problems in this technological era. There are several approaches to carry out the process of data mining, which are discussed in later sections of this paper.

II. DATA MINING APPROACHES

A. *Clustering*

Clustering is a popular data mining approach in which knowledge is extracted by process of collecting similar data objects. It an unsupervised learning approach that groups the data objects into clusters that share similar characteristics. Additionally, with help of clustering we can discover dense and sparse regions of object space and overall distributed patterns and correlations between data values. Most popular clustering algorithm is k-means clustering, hierarchical clustering, grid based clustering and model based clustering algorithms. Clustering can be used in many real world applications such as financial related, energy consumption at home or organizations, spatial data analysis etc.

B. Classification

Classification is another popular approach of data mining. Classification is process that categorizes immerse volume data and sorts them into pre-classified set of classes. Data classification has four sub processes. They are data preprocessing, data modeling, feature selection and data validation. Classification is a predictive and supervised learning approach. Using classification we can derive object's class by its attributes/values. Attribute vectors and their classes represent training sets. Building classification model is done with help if analysis of relationships among classes and attributes of objects in training set. Those models are useful for classifying of future objects and evolving of well understanding of objects and classes in the database. Few of popular classification approaches are neural networks, decision trees, Bayesian classification, association rules. Classification approach has wide variety of applications such as banks, fraud detection etc.

C. Time Series data analysis

In time series analysis approach, extraction of sequence patterns or events over repeated time intervals are used by time series databases on hourly, daily and/or weekly basis stores them into sequence database. Sequence database contains sequence ordered events and strong time notions.

D. Regression

Regression is a predictive and supervised learning approach. In Regression, relationships are constructed among one or more dependant and independent attributes from a data set in a way that independent variable are used to predict the dependant variable by using support vector machines. Linear regression, non-linear regression, logistic regression, multivariate regression and multivariate non linear regression are some examples of regression. By using regression we can predict the behavior of children with analyzing family history,

and we also predict stock market and future resource utilization etc.

E. Summarization

In summarization approach, data is mapped into subsets with simple details. This basic idea of summarization approach is to derive accumulated information.

F. Association

The Association approach finds associations, basic structures, and relationships between attributes and objects. In this approach existence of on model is entailed by another model. Association rules are used to analyze and predict customer behavior and shopping data analysis, advertising, and product management etc. It is most popular approach of data mining. In this approach strong rules are discovered in very large databases using interestingness.

III. DATA III. MINING APPLICATIONS

A. Healthcare

Data mining approaches can be applied in healthcare sector to identify patient's activities, predicting cancers, heart diseases and many other diseases. It also useful for keeping track of patients visits to the hospitals and finding the patterns of successful medication for various diseases. We also analyze patient's biological data for analyzing patient's health status. Many of researchers are trying to improve data mining algorithms so that they can be applied in predicting and estimating patients and diseases for better healthcare prospects.

B. Financial

Data mining in financial sector makes banking and financial industries more reliable and offers high quality services to customers. Its applications in financial sector includes analyzing market trends, customer behavior, stock market prediction, credit related applications etc.

C. Educational

Data mining can be applied in educational domain to give knowledge about educational environment data. With help of data mining we can predict student learning outcomes, behavior, and asses their skills frequently for improvements.

D. Cyber Security

Data mining is useful for analyze and predict future cyber attacks. By applying data mining algorithms we can provides better intrusion detection system for cyber security.

E. Agriculture

Data mining can be used to analyze the crop yield using factors like year, percentage of rainfall, production quantity and quality and using geographical area. Prediction of yield is significant issue in agriculture domain. We can overcome this problem by applying various clustering and classification algorithms like K-means clustering, KNN (K-nearest neighbor) algorithm, neural networks, artificial intelligence, and support vector machines.

Above mentioned are some of example domains of data mining applications but it has enormousness applicability. Like analysis of satellite data and astronomical data for predicting environment disasters and behavior. Data mining is very useful technology for cloud computing environment. By applying data mining algorithms over cloud users can extract needful information for virtual databases which further reduces maintenance cost of storage and infrastructure. Using data mining we can offer reliable, efficient, and secure cloud services to its users. In retail sector data mining used analyze large amount of customer data to predict customer buying behavior, sales prediction, and track product transportation etc. Data mining is applied to various Internet of Things (IoT) devices to keep track of events and predict the future events by analyzing the very large data collected from various things such as fitness trackers, energy consumption trackers, traffic analysis and prediction in smart cities using sensors etc.

IV. CONCLUSION

Data mining is utmost and essential technique that has high influence on data. Since data is generation is very high rate and most of data is dynamic data in this technological era importance of data mining approaches are expanded to analyze those large amounts of data and give useful knowledge. This paper successfully presented overview of various data mining approaches and their applications. Data mining approaches like clustering, classification, times series analysis, prediction can be applied on data of any domain to get desired knowledge about that domain.

REFERENCES

[1] Fabio T. Matsunaga, Jacques D. Brancher, Rosângela M. Busto, "Data mining applications and techniques: a systematic review", Londrina State University (UEL).

[2] Mehdi Allahyari, Seyedamin Pouriyeh, Mehdi Assefi, "A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques", KDD Bigdas, August 2017.

[3] S.Brindha, Dr.K.Prabha, Dr.S.Sukumaran, "A SURVEY ON CLASSIFICATION TECHNIQUES FOR TEXT MINING", ICACCS -2016.

[4] Janak Thakkar, Dr. Mehul Parikh, "A Survey on Efficient Frequent Pattern Mining Techniques", IJSRSET, Volume 4, Issue 2, 2018.

[5] Prof.S M.Inzalkar1, Jai Sharma, "A Survey on Text Mining- techniques and application", IJRISE, Special Issue, Techno-Xtreme 2016.

[6] Dr. Suresh Jain, Rajeev Raghuvanshi, Ilyas, "A Survey Paper on Overview of Basic Data Mining Tasks", IJIACS, Volume 6, Issue 9, September 2017.

[7] Fathimath Zuha Maksood, Geetha Achuthan, Analysis of Data Mining Techniques and its Applications", IJCA, Volume 140, Issue No. 3, April 2016.

[8] Shweta Bhatia, Sweety Patel, "Analysis on different Data mining Techniques and algorithms used in IOT", IJERA, Volume 5, Issue 11, November 2015.

[9] Prempratap Singh, Gouri Gosawi, Sulakshana Dubey, "Application of Data Mining", Binary Journal of Data Mining & Networking 4, 2014.

[10] S.D.Gheware, A.S.Kejkar, S.M.Tondare, "Data Mining: Task, Tools, Techniques and Applications", IJARCCE, Volume 3, Issue 10, October 2014.

[11] Priya Bhardwaj, Priyanka Khosla, "Review of Text Mining Techniques", National Conference on Emerging Trends in Information Technology- Advances in High Performance Computing, Data Sciences & Cyber Security, Volume 8, Issue 1. January-June 2017.

[12] Mani, Bharti Suri and Manoj Kumar, "Performance Evaluation of Data Mining Techniques",

[13] Deepashri.K.S, Ashwini Kamath, "Survey on Techniques of Data Mining and its Applications", ICETE, Volume 6, Issue 2, February 2017.

[14] Liao, S.-H., Chu, P.-H., and Hsiao, P.-Y. (2012). Data mining techniques and applications – A decade review from 2000 to 2011. Expert Systems with Applications.

[15] Prof. Dr. Wolfgang Karl Hardle,"Time Series Data Mining Methods: A Review", Berlin, March 25, 2015.

[16] Pradnya P. Sondwale, "Overview of Predictive and Descriptive Data Mining Techniques", IJARCSSE, Volume 5, Issue 4, April 2015

[17] Data Mining: Concepts and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management System`s) 3rd Edition.

[18] Jasvinder Kumar, "A Comprehensive Study of Educational Data Mining", ijeecse, Special Issue - TeLMISR 2015.

[19] Mr. Abhishek B. Mankar, Mr. Mayur S. Burange, "Data

Mining - An Evolutionary View of Agriculture", ijaiem, Volume 3, Issue 3, March 2014. Volume 3,