

AN EMPIRICAL STUDY ON DATA MINING TECHNIQUES AND APPLICATIONS

Neelam Rani¹, Er. Harpreet Kaur²

¹ Research Scholar, Department of CSE, UIET SBBSUPadhiana, Jalandhar,Punjab, India

² Assistant Professor & Dean Academics, Department of CSE, UIET SBBSU Padhiana, Jalandhar,Punjab,India

Received: May 16, 2018

Accepted: June 28, 2018

ABSTRACT

Data mining is the way or process of extraction of useful information from large datasets is an innovative technology. The main aim of the data mining is to extract knowledge from a large data set and transform it into an understandable or suitable structure for further use. In this paper, we discuss data mining, its techniques. Also, discuss the various application of data mining and its tools.

KEYWORDS:Data mining, Clustering, Classification, KDD, Application

I. INTRODUCTION

Data mining is the technique or process of extraction of useful knowledge from the large volume of business data or data sets, is a great innovative technology, it is useful for data analysis and decision making. Knowledge is discovered from the dataset and presenting in the suitable form that is easily understood by humans. Data mining using various type of technique to analyze the very large dataset. Data mining is a rapidly expanding field. It is also called as knowledge mining from data, knowledge discovery process, knowledge extraction .Data mining is logical process means that is used to find useful data after analysis a large amount of data. The main of data mining to find the usage patterns and relationships within our data. Ashish Kumar Dogra and Tanuj Wala (2015) describe data mining "Data mining is the process of extraction hidden knowledge from large volumes of raw data. Data mining has been defined as the nontrivial extraction of previously unknown, and potentially useful information from data. "[9]

The following two main reasons to use data mining as a rapidly growing in demands of data. These are:

1. A Large amount of data and too little information.
2. There is a need to extract useful information (knowledge) from the data and to interpret the data.

A) Knowledge Discovery in Database in Data mining consists of the following steps:-

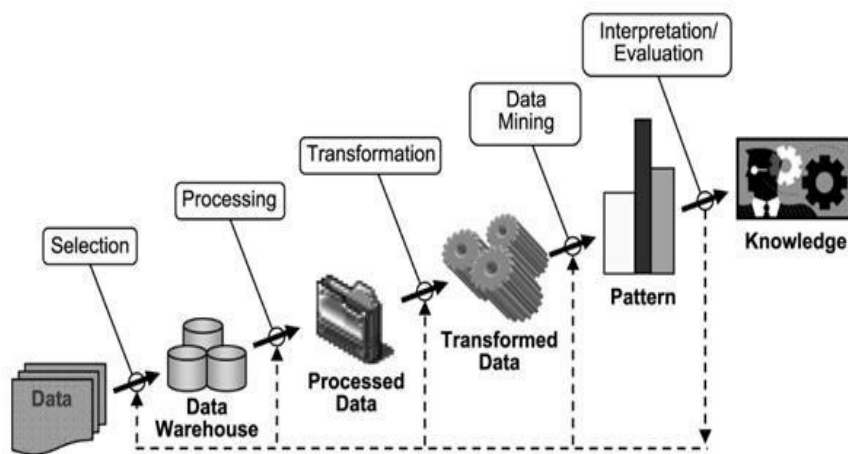


Fig.1 Knowledge Discovery process in Data Mining [1]

1. Data cleaning:-It is also called data cleansing, at this phase in which noise data and inconsistent data removed from the collected data.
2. Data integration:-In this phase, multiple data sources, maybe integrate with the common source.

3. Data selection:- In that step the data relevant to analysis is selected on and retrieved from the data collection.
4. Data transformation:-It is also called data consolidation, at this phase in which the selected data is transformed into particular forms (convert data into two-dimension form).
5. Data mining:-In that step in which clever techniques are applied to discover patterns or knowledge.
6. Pattern evaluation:-In this phase, interesting patterns representing knowledge are identified based on given measures.
7. Knowledge representation:-It is the final step in which the visually represented the discovered knowledge to the user.

II. DATA MINING TECHNIQUES

Following are the several major data mining techniques:-

A) Classification Approach

Data classification is a process through the given data is classified into different classes or groups according to a classification model.

Classification is one of the useful data mining technique which is used for predicting group membership for data instances. Credit-risk applications and fraud detection are particularly well suited to this type of analysis. This approach employs neural network-based or decision tree classification algorithms. The classification process involves two step that is learning and classification. In Learning the training data are analyzed by the particular classification algorithm. In classification step, test data are used to estimate the accuracy and efficiency of the classification rules.

Classification is a supervised machine learning method. Data can be trained by providing the training to data and we can capable for predict the future of data. In that case, the prediction is in the form of predicting the class to which data can belong. Providing training to data is based on the training sample provided.

Basically, there are two types of attributes available that are dependent on attribute or output and the independent attribute or input. In the supervised learning classification, mapping the input data set to the finite set of discrete class labels.

Dataset as an input $x \in R^i$, where the input space dimensionally is i and the discrete class label $y \in 1.....t$, where t is the total number of class types. And this is modeled in the term of equation $y=y(x, w)$, w is the vector of adjustable parameters.

Types of classification models:

- Neural Networks
- Classification by decision tree induction
- Bayesian Classification

B) Clustering Approach

Clustering is an unsupervised classification or it is also known as exploratory data analysis in which no provision of labeled data. The main task of clustering is to maximize the similarity between the objects and data set of classes and to reduce the similarity between the classes.

Clustering is an effective data mining technique that makes the valuable cluster of objects and data sets. The clustering technique describes the classes of similar object and puts objects in each class, in the classification techniques, objects are given into predefined classes. For example in a library, the large numbers of books in various titles are available. Then the challenge is how to keep those books in such a way that readers can take books in a particular topic without occurring any difficulty. By using clustering technique, keep the books that have some kinds of similarities in one cluster and label it with a meaningful and effective name.

Types of clustering methods

- Partitioning Methods
- Density-based methods
- Grid-based methods
- Model-based methods
- Hierarchical Agglomerative (divisive) methods

C) Association Rule

Association uses to find frequent item set findings among the large data sets. This type of technique helps businesses to make certain decisions, such as particular design, customer shopping, and cross-marketing behavior analysis. Association algorithms need to be able to generate rules with confidence and efficient values less than one. The number of possible Association Rules for a given dataset is very large and a high proportion of the rules are usually of little value.

Types of association rule

- Multilevel association rule
- Quantitative association rule
- Multidimensional association rule

D) Prediction

The prediction through determine the relationship between independent and dependent variables. In the prediction technique analysis can be used for sale to predict profit, in that sale is an independent variable and the profit could be a dependent variable. Then based on the past profit and sale data, a regression curve that is used for profit prediction.

Neelamadhab Padhy et al. (2012) describes the difficulty in predict a data is a complex. Actually, no approaches or tools can guarantee to generate the accurate prediction in the organization. In this paper, they have analyzed the different algorithm and prediction technique. In spite the fact that least squares regression median known to produce better effective results than the classifier linear regression techniques from the available set of the attributes. As the comparison, they found the Linear Regression technique which takes the less time as compared to Least Square Regression Median.

E) Neural Networks

The neural network is a set of input/output connected units and each connection has a weight. In the learning phase, network learns by adjusting weights so that is able to predict the correct class labels of the input tuples.

In the neural network have the ability to derive the meaning from very complicated dataset or data and can be used for extract patterns and detect trends from data that are very complex to be noticed by either humans or any other different computer techniques. These are very suitable for continuous-valued inputs and outputs. For example, the handwritten character reorganization, provide training to the computer to pronounce.

Types of neural networks

- Back Propagation

Following are the tools for data mining:

- Tool 1-Orange
- Tool 2- WEKA
- Tool 4- Apache Mahout
- Tool 5- R Software Environment
- Tool 8-NLTK (Natural Language Tool Kit)

III.DATA MINING APPLICATIONS:

Data mining is an analytical approach that has been quickly adapted and used in domains that were already using statistics. The applications areas are :

A) Financial Industry, E-commerce, Banks and Businesses

- Investment analysis and stock.
- Risky customer and identify loyal customers.
- Predict customer spending.

B) Decision support and Database analysis

- Management and Market analysis.
- Market basket analysis, target marketing, customer relation management.
- Management and risk analysis.
- Competitive analysis, forecasting and quality control.
- Management and Fraud Detection.

C) Pharmacy/ Medical

- Prediction/ Characterization of patient's response to product dosage.
- Identification of successful prescription patterns (successful medical therapies).
- Expert systems learning (Computer Assisted Diagnosis).

IV.CONCLUSIONS AND FUTURE WORK

This study gives an idea about the data mining, its techniques, and application. Therefore, Interesting knowledge extracting from large amounts of data stored either in data warehouses, databases or other information repositories that is the data mining process. The different techniques used for the mining of data. The future scope provides efficiency or enhancement of data in the system. They could lead to faster, better and qualitative exaction of data with better tools and techniques.

REFERENCES

1. Mrs. Bharati, M. Ramageri : DATA MINING TECHNIQUES AND APPLICATIONS.
2. Mansi Gera,Shivani Goel: Data Mining - Techniques, Methods and Algorithms: A Review on Tools and their Validity, March 2015
3. K.Suguna , Dr.K.Nandhini: LITERATURE REVIEW ON DATA MINING TECHNIQUES.
4. Meenakshi Sharma: Data Mining: A Literature Survey, February 2014
5. Ms. Aruna J. Chamatkar¹, Dr. P.K. Butey: Importance of Data Mining with Different Types of Data Applications and Challenging Areas, May 2014
6. Nirmal Kaur, Gurpinder Singh:A Review Paper On Data Mining And Big Data,Volume 8, No. 4, May 2017 (Special Issue).
7. Deepashri.K.S, Ashwini Kamath : Survey on Techniques of Data Mining and its Applications, February 2014
8. Anu Sharma, Dr. M.K Sharma & Dr. R.K Dwivedi : Literature Review and Challenges of Data Mining Techniques for Social Network Analysis 2017
9. Ashish Kumar Dogra¹, TanujWala²: A Review Paper on Data Mining Techniques and Algorithms,May 2015
10. Mohammed J. Zaki,Department of Computer Science, Rensselaer Polytechnic Institute: DATA MINING TECHNIQUES.
11. Chinky Gera Kirti Joshi:A Survey on Data Mining Techniques in the Medicative Field, March 2015
12. ¹Manoj and ²Jatinder Singh :Study and Analysis of Data Mining Techniques,2011