Advances in Computer Science and Information Technology (ACSIT)
Print ISSN: 2393-9907; Online ISSN: 2393-9915; Volume 1, Number 2; November, 2014 pp. 74-76

http://www.krishisanskriti.org/acsit.html

Overview of Data Mining Techniques

Dipeeka K. Rathod[1] and Neha Valmik[2]

[1]Everest College of Engineering Sanjay Nagar, SmashanMaruti Road, Aurangabad-43100
[2]Everest College of Engineering Jatwada Road, Aurangabad-431 119
E-mail: [1]dipeekakrathod@gmail.com, [2]nehavalmik@gmail.com

Abstract: *In terms of data processing, classical statistical models are restrictive; it requires speculation, Specialist's knowledge and experience, equations, effective knowledge of probabilities distribution and the data must have a high quality, being subject to prior processing and transformations. The concept of data mining has come forth to overcome these disadvantages, which implement knowledge extraction algorithms from the large data collections. Data mining is a process of collecting knowledge from databases or data warehouses and the information collected that had never been known before, it is valid and operational. In this paper we overviewed different tasks includes in Data mining. Data mining involves the tasks like anomaly detection, regression, classification, association rule learning, summarization and clustering.*

1. INTRODUCTION

Data mining is the extraction of hidden predictive information from a large database. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. Data mining is a strong tool because it can provide you with relevant information that you can use to your own advantage. When you have the right data, then you will need to do is apply it in the right manner, and you will be able to get beneficial result. Now a day's getting information is relatively easy. But for achieving certain desired goals it require relevant information. This is where data mining becomes a powerful tool that you will want to become familiar with. Data mining gives you the power to predict certain behaviors within a system. Data mining involves the anomaly detection, association, regression, rule learning, classification, and summarization and clustering. The goal of this technique is to find patterns that were previously unknown. Further the mined results should be valid, novel, useful, and understandable. [5]Once you have found these patterns, for solving the number of problems you can use it.

2. DATA MINING

The development of Information Technology has generated large amount of databases and huge data in different areas. The research in databases and information technology has given rise to an approach to store and manipulate this precious data for further decision making. Data mining is a process of extracting useful information and patterns from large amount of data. It is also called as knowledge discovery process, mining knowledge from data, knowledge extraction or data /pattern analysis.[4]

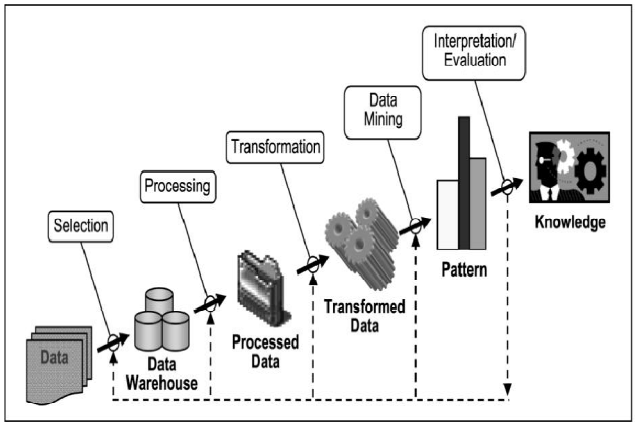

Fig. 1[2]: Knowledge discovery process

Data mining is a logical process that is used to search through large amount of data in order to find useful data. [1]The goal of this technique is to find patterns that were previously not known. After these patterns are found they can further be used to make certain decisions for development of their businesses.In data mining the data can be mined by passing various processes. Three steps in mining process involved are

- Exploration
- Pattern identification
- Deployment

Exploration: In data exploration step firstly data is cleaned and transformed into another form, and important variables and then nature of data based on the problem are determined.

Pattern Identification: Once data is examine, refined and defined for the specific variables the second step is to form pattern identification. Identify and select the patterns which

make the best prediction. For making a best prediction second step involves are identifying and choosing the patterns.

Deployment: To get the desired outcome patterns are deployed.

3. DATA MINING TASK AND TECHNIQUES

Various task and techniques like Clustering, Classification, Regression, Artificial Intelligence, Association Rules, Neural Networks, Genetic Algorithm, Decision Trees, Nearest Neighbor method etc.,[3] are used for knowledge discovery from databases.

3.1 Classification

Classification consists of examining the features of a newly presented object and assigning to it a predefined class. The classification method is characterized by the well-defined classes, and a training set consisting of reclassified examples. The task is to build a model that can be applied to data which not classified in order to classify it. The derived model is based on analysis of set of data objects whose class label is known i.e. training data[3]Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. Particularly this type of analysis includes Fraud detection and credit risk.[2]

3.2Clustering

Clustering can be said as identification of similar classes of objects. With the help of clustering techniques we can further identify dense and sparse regions in object space and can discover overall distribution pattern and correlations among data attributes. Classification technique can also be used for effective means of distinguishing groups or classes of object but it becomes costly so clustering can be used as preprocessing approach for attribute subset selection and classification. For example, based on purchasing patterns forming group of customers, to categories genes with similar functionality.[3]

3.3 Predication

Prediction is very similar to classification. The difference in classification and prediction is the class is not a qualitative discrete attribute but a continuous one. The goal of prediction is to find the numerical value of the target attribute for unseen objects; this problem type is also known as regression, and if the prediction manages with time series data, then it is often called forecasting. Regression analysis, neural nets, and decision trees generally apply.[5]

3.4 Association rule

Association and correlation is usually to find frequent item set findings among large data sets. This item sets helps businesses to make certain decisions, such as, cross marketing, catalogue design and customer shopping behavior analysis. In Association Rule technique algorithms should be able to generate rules with confidence values less than one. However the number of possible Association Rules for a given dataset is generally very large and a high proportion of the rules are usually of little (if any) value.[3]

4. LITERATURE SURVEY

Data mining techniques provide a popular and powerful tool set to generate various data driven classification system. In data mining the data is mined using two learning approaches i.e. supervised learning and unsupervised learning.[2]

4.1. Supervised Learning

In supervised learning (often also called directed data mining) the variables under investigation can be split into two groups: explanatory variables and one (or more) dependent variables. The goal of the analysis is to determine a relationship between the dependent variable and explanatory variables the as it is done in regression analysis. To proceed with directed data mining techniques the values of the dependent variable must be known for a sufficiently large part of the data set. Classification, prediction rules, and estimation are example of direct data mining or supervised learning.

4.2. Unsupervised Learning:

In unsupervised learning, all the variables are treated in same way; there is no difference between dependent and informative variables. However, direct contrast to the name undirected data mining, still there is some target to achieve.[2] This target might be as data reduction as general or more specific like clustering. The dividing line between unsupervised learning and supervised learning is the same that distinguishes discriminate analysis from cluster analysis. Supervised learning requires, target variable should be well defined and that a sufficient number of its values are given. Unsupervised learning typically either the target variable has only been recorded for too small a number of cases or the target variable is unknown.

Clustering, association rules and description are example of example of unsupervised learning.

5. CONCLUSION

Data mining has importance regarding finding the patterns, discovery of knowledge, forecasting etc., in different business domains. Data mining algorithms and techniques such as clustering, classification, etc., helps in finding the patterns to decide upon the future trends in businesses to grow. Data mining has broad application domain almost in every industry where the data is generated that's why data mining is

considered one of the most important frontiers in database and information systems and one of the most promising interdisciplinary developments in Information Technology. This paper has presented the conceptual overview of various techniques in data mining.

6. ACKNOWLEDGMENT

I would like to thanks to our principal Dr. VenkateshGaddime sir, Head of the Department Prof. R. A. Auti, my guide Prof. N. A. Valmik, and Faculty of Computer Science And Engineering, Everest Engineering College For Providing the necessary facilities for the preparation of the paper.

REFERENCES

[1] DivyaChaudhary **"** Data Mining: Techniques and Algorithms**"** Department of Computer Science & Applications M.D. University, India, Volume 3, Issue 8, August 2013

[2] Anand V. Saurkar, VaibhavBhujade, PritiBhagat , Amit Khaparde "A Review Paper on Various Data Mining Techniques" Department of Computer Science&Engg , DMIETR, Sawangi(M), Wardh, Maharashtra, India., Volume 4, Issue 4, April 2014.

[3] Mrs. Bharati M. Ramageri "Data mining techniques and applications" Lecturer, Modern Institute of Information Technology and Research,Department of Computer Application, Yamunanagar, Nigdi, Pune, Maharashtra, India - 411044. Vol. 1 No. 4 301-305

[4] Jiawei Han and MichelineKamber, "Data Mining" Morgan Kaufmann Publishers, Second edition, 2006, pp 1-45

[5] Joyce Jackson, " DATA MINING: A CONCEPTUAL OVERVIEW , Management Science Department University of South Carolina, joyce.jackson@sc.edu, Communications of the Association for Information Systems (Volume 8, 2002) 267-296

About Author:-

Author 1-

Dipeeka K. Rathod -I am pursuing degree in computer Science & engineering from Dr. SeemaQuadri Institute of Tech, Aurangabad, My area of interest is DBMS, Data structure, Software testing.

Author 2:-

Neha Khatri Valmikis currently working as Assistant Professor in the Department of Computer, Dr. SeemaQuadri Institute of Tech, Aurangabad, India. She have 5 years and 4 months of teaching experience. Her research area include Security, image processing, Data Structure, Android.