A DETAILED SURVEY ON TECHNIQUES AND APPLICATIONS OF DATA MINING

Mahendra Singh Panwar , Naval Kishore Pareek,

Assistant Professor, Computom Institute of Information Technology and Management,Jaipur
panwarpanwar80@gmail.com
Assistant Professor, Computom Institute of Information Technology and Management,Jaipur
email : nk18pareek@gmail.com

-- ***--

Abstract— **Data mining is the procedure of eliminating positive data, figures & styles, after a large amount of data technologies, clustering & suggestion, and regression. Real life has a variety of applications. Various devices that support different algorithms are available. The DMT (Data Mining Technique) is further developed in the research sector.The relevance of this huge collection of data is highly subjective as according to the individual business interests. Data Mining is a relatively new term in the field of informatics. Data mining is data for filtering relevant data based on a wide range of data, using a variety of techniques and algorithms, such as associations, clustering, and classics. In this paper, there techniques are explained in detailed to provide the better understanding of the data mining techniques.**

Keywords—Data Mining, Knowledge discovery data, Classification, Clustering, Association rule.

-- ***--

I.INTRODUCTION

Development of Information Technology developed a large number of databases and large data in different areas. Research done in databases and information technology The approach to keeping and valuable these valuable information takes a decision. Data mining is beneficial data& a big number of configurations. It is also known as knowledge mining from the knowledge detection process, data, knowledge detection, or data / pattern analysis.

Figure 1. Knowledge discovery Process

Data mining is a logical process for large quantities to search for useful data. The goal of this technique is to find patterns that were previously unknown. If these patterns are found, they can be more useful to make some decisions on the development of their businesses. It has three stages:

1. Exploration
2. Pattern identification
3. Deployment

Exploration: Data becomes the first form of exploration data, transformed into different formats, variables and nature of the data based on that problem are determined.

Pattern Identification: The second stage for determining the data pattern is to verify, define and define the specific variables. Identify the patterns that make the best prediction.

Alignment: Deploying patterns for the desired result.[1].

II TECHNIQUES OF DATA MINING

To analyze large amount of data, data mining came into picture and is also called as KDD process. In KDD the main and important step is data mining. KDD will turn the low level data into high level data. Data mining is predicted from the main database. It uses already built tools to get out the useful hidden patterns, trends and prediction of future can be obtained using the techniques. Data mining involves model to discover patterns which consists of various components.

Clustering, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetics Algorithm, and neighboring neighborhoods have used various algorithms and technologies..
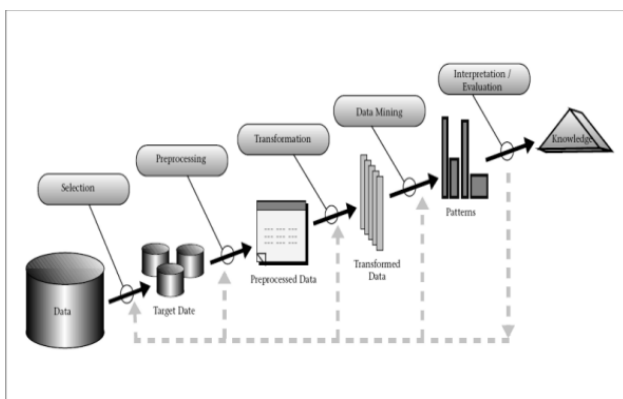
A. Classification

The most commonly used data is the Mining Technique Classification. It uses predefined examples. Fraud detection and credit apps are well suited used for this kind of analysis. It method uses classification algorithms based on TrueNeural network. The data classification process comprises study & arrangement. The arrangement algorithms examine the training data.Classification uses test data to measure the accuracy of classification rules. If accuracy is acceptable, the instructions can be useful to new data tubs. Fraud detection app includes full documentation of fraud and valid activities that determine on the basis of recording by bike. Pre-Defined Verification Examples are used by sorting-training algorithms to determine the parameters that require the correct discrimination. Algorithm These parameters are encoded as a classification called a classifier. Types of classification modes [2]:

1. Classification through decision tree induction
2. Bayesian Classification
3. Neural Networks
4. Support Vector Machines (SVM)
5. Classification Based on Associations

B. Clustering

Clustering is used to identify similar objects. Within Objective Techniques, we can identify the vast, broadest areas of the object space. All data attributes can be found in the distribution system, as well as the linkage. Clustering can be used like a pre-defined process for the attribute subset choices& classification, however it can be used according to the classification method classification groups or categories of classes. For example, to create a wavelength group of groups based on the purchase patterns into the genes with similar functions. Cluster methods:

1. Partitioning Methods
2. Hierarchical Agglomerative (divisive) methods
3. Density based methods
4. Grid-based methods
5. Model-based methods

It is also known as the Exploratory Data Analyze that there is no system of unclassified classification called clustering or a system of labeled data. The main goal of clustering technology is to extract undetected data and separate it into a complete set of nature and hidden data structures. Nonstress samples created from the same probability distribution do not have the precise nature of the system. There are two areas of clustering on broad clustering which are categorized as follows:

Hard Clustering: Hard Clustering may be a single cluster of the same object.

Soft clustering: In this clustering same object can belong to different clusters [3].

C. Predication

Regression technique can be adapted for predication. Regression study may be utilized to model the relationship among one or more autonomous variables and reliant on variables. In data mining, free vans already know the attributes, and the response variables are predictable. Unfortunately, many real problems are not a prediction. For example, sales measurements, stock prices, and failure of the product are difficult to predict because they depend on the complex interactions in the variables of the proxies. Therefore, more complicated technologies (for example, logistic regression, decision trees or neural networks) are required to predict future values. The identical modeling approaches can frequently be utilized in regression and classification. For example, the CART (Classification and Regression Trees) decision tree algorithm can be used to build both classification trees (to classify categorical Response variables), Regression trees (to observe continuous variable variables). Glowing networks, categorization and regression models can be created. Types of regression methods:

1. Linear Regression
2. Multivariate Linear Regression
3. Nonlinear Regression
4. Multivariate Nonlinear Regression

D. Association rule

In the large data sets, occasional item detectors are formed, association and interconnected. Finding this type of catalog helps industries to make some decisions, such as catalog design, cross marketing and customer shopping behavior analysis. The Association Rule Algorithms can make laws using just one notification below. However, the number of possible association rules for a given decimon is very high, usually at a higher rate than the normal ratio.Types of association rule:

1. Multilevel association rule

2. Multidimensional association rule

3. Quantitative association rule

E. Neural networks

This is a group of input / output units connected to an incorrect network. Every connection has a load. In the study phase, the network understands the correct class labels for input taps and understands the body by adjusting the weight. Nubular networks have significant potential for finding meaning from complex or unlimited data, which can be used to extract complex trends and look for trends to look at either the game or other computer techniques. These are suitable for continuous evaluation and production. For example, manuscriptization has been successfully applied in many industries to train English text and many actual world business issues on a computer. The networks are good for identifying patterns or trends in data, for better prediction, prediction or demand. Types of neural networks:

1. Back Propagation

The main purpose of data mining is for the extraction of the useful and relevant information from the large databases or data warehouses. Applications of Data mining are mainly useful for commercial and scientific areas [1]. Isn the scientific area, it study is geerally conferred into the data mining apps. Data excavation of the scientific sector is self-extracting, and the nature of datasets is very different from the applications that lead to the data mining's traditional market. Marketing, Engineering medicine analysis, Customer Relationship Management, Skill Analysis,web mining & mobile computing and mobile [4].

IV DATA MINING APPLICATIONS

i. Marketing and retail: Data Mining, Accurate Information, Customer Buying Trends and Best Selling Products, so retail-stop managers can identify their loyal customers, configure customer discounts, and arrange shelves. Customer profiling, retention, identifying potential customer, market segmentation.

ii. Fraud detection: Infiltration and Identify credit card fraud.

iii. Scientific Data Analysis: Classify Research Decision Creating Decision.

iv. Text and web mining: Looking at the text or information on the Web or the rawinformation.

v. Business Zone: Basic use of business world data analysis, performance feedback, user feedback assessment, and stock analysis values of existing organizations and their market trends to aid in futurebusiness decisions.

vi. Bioinformatics: An assessment of the disease, the effectiveness of the treatment, genetic and proteomic data, and the medical record of the patient for diagnosis.

vii. Climate Change: Weather forecasting from time to time to predict future climate change to determine natural disasters such as hurricane andalso weather forecasting.

viii. Banking Finance: Assessment of individual banking records to generatedifferent marketing strategies for a target customer segment, loan approval, stockforecasting, checking different kinds of fraud and money laundering.

ix. Security and data integrity: Data mining can be used to monitor various systems, alarms arise whenever detect any security breaches or infiltration. It canhelp in identifying the reason for security problems in firewall.

x. E-Commerce: uses data mining technologies to e-commerce to analyze customer search patterns to promote sales and promotion. [5].

V LITERATURE SURVEY

[6] This report is mainly about data mining and its applications. Treatment includes important areas of efficiency. Health Care Management, also the detection of the fraud and also provides an overview about the customer relationship management.

How to find information collections, and how to produce useful patterns of large data to find possible observation patterns [7]. This article focuses on data mining in improving the quality of the decision making procedure in the medical field.Yang Liu et al (2018) In this paper we use network footprint data from the web application app utilization and ISPs from the crawler data and the DPI data. Focusing on the most popular applications in China suggests the NFP data collection and processing framework we supply. Using the Apriory and MS-Apriory algorithm, we excavate the Association rules in NFP data.Experimental results validate our proposed method and present some interesting association rules of Apps. [8]

Ruslan Dautov et al (2018), like the recasting of output signals, creates a wrong dictionary by collecting the benefits of wavelet transformation for error-free and incorrect conditions. Classics are based on K-nearest neighbor method (k-NN), and association management mining algorithm. This error diagnostic technique has been trained and checked using the data received after the simulating behavior of the unauthorized filters-free and false behavior. As a result, there is accuracy of incorrect false provisions and wrong coverage 99, 09% and more than 99,08%. The proposed technique is completely automated and can be extended. [9]

Marwa Bouraoui et al (2017) in this paper, an efficient approach for ARM based on Map Reduce framework, adapted for processing large volumes of data. Furthermore, because real life databases lead to huge numbers of rules including many redundant rules, our algorithm propose to mine a compact set of rules with No information is lost. The results of experiments tested on large real world datasets highlight the relevance of mined data [10]

Chunkai Zhang et al (2017) this paper first compresses Constant time series. Then in order to make the mining rules reflect the characteristics of multivariate time series data, our paper designs a new algorithm called IAMTL, which can mine the rules from the fix time span. For the reason that time series data have the characteristic of continuity, so an increment version of IATML is provided. At last, we use prerequisite and the consequent windows to verify the correctness of the rules. [11]

Mengmeng Cao et al (2017) in this paper. The algorithm can streamline the size of candidate sets and minimize the number of traversing and scanning for the database. So, the efficiency of rule acquisition is significantly improved. The feasibility and validity of the improved algorithm are verified by an experiment which is conducted in a power monitoring data through association rule analysis. It provides an effective approach for excavating power monitoring data and discovering of association rules. [12]

VI CONCLUSION

Data mining is a procedure to identify beneficial design safter large data. Data mining techniques, algorithms, and some companies are finding the best results in data mining technologies to improve their industries and discuss them. Data mining is provided by data mining. Every data is in place. Data Mining is considered one of the most important components in both database and information systems Information Technology.

REFERENCES

[1] Bharati M. Ramageri, DATA MINING TECHNIQUES AND APPLICATIONS, Indian Journal of Computer Science and Engineering Vol. 1 No. 4 301-305.

[2] Baker RID, Yacef K (2009) The State of Educational Data Mining:A Review and Future Visions. JEDM - Journal of Educational Data Mining, 1: 3-16

[3] Kumar R, Kapil AK, Bhatia (2012) A Modified tree classification in data mining. Global Journals Inc. 12, 12: 58-63

[4] Sumitha Thankachan1 , Suchithra2, Data Mining & Warehousing Algorithms and its Application in Medical Science - A Survey, Data Mining & Warehousing Algorithms and its Application in Medical Science - A Survey.

[5] Aarti Sharma, Rahul Sharma, Vivek Kr. Sharma" Application of Data Mining – A Survey Paper", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2), 2014, 2023-2025.

[6] HianChyeKoh and Gerald Tan, Applications in Healthcare Information Management –Vol 19, No 2.

[7] JayanthiRanjan, ―Applicati techniques in pharmaceuticJournalof Theoretical and Applied Technology, (2007).

[8] Yang Liu, Ke Yu and Xiaofei Wu" Association Rules Mining Analysis of App Usage Based on Mobile Traffic Flow Data", 978-1-5386-4794-3/18/$31.00 ©2018 IEEE.

[9] Ruslan Dautov, Sergey Mosin" A Technique to Aggregate Classes of Analog Fault Diagnostic Data Based on Association Rule Mining", 978-1-5386-1214-9/18/$31.00 ©2018 IEEE.

[10] Marwa Bouraoui, Ines Bouzouita and Amel Grissa Touzi" Hadoop based Mining of Distributed Association Rules from Big Data", 978-1-5386-1084-8/17/$31.00 ©2017 IEEE.

[11] Chunkai Zhang, Qing Liao and Xuan Wang" MINING INTER-TRANSACTION ASSOCIATION RULES FROM MULTIPLE TIME-SERIES DATA", 978-1-5386-3016-7/17/$31.00 ©2017 IEEE .

[12] Mengmeng Cao, Chaoyou Guo" Research on the Improvement of Association Rule Algorithm for Power Monitoring Data Mining", 2473-3547/17 $31.00 © 2017 IEEE.