ORIENTAL JOURNAL OF COMPUTER SCIENCE & TECHNOLOGY
An International Open Free Access, Peer Reviewed Research Journal
Published By: **Oriental Scientific Publishing Co., India.**
www.computerscijournal.org

ISSN: 0974-6471
June 2013,
Vol. 6, No. (2):
Pgs.45-48

Data Mining: Exploring New Horizons from Huge Data Store

TRUSHAR.B PATEL[1] and PREMAL SONI[2]

[1] Assist. Prof, Shree P.M Patel Institute Of Business Administration, Anand (India).
[2] Assist. Prof, Shree P.M Patel College of Computer Science and Tech., Anand (India).

(Received: June 02, 2013; Accepted: June 12, 2013)

ABSTRACT

DM i.e. Data Mining is the technique to find out the hidden facts from the large amounts of data. From the historic data, we derive knowledge. Data Mining is important aid for decision making in organizations. Data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown interesting patterns from the data store.

Key words : Data mining, Selection, Transforming, DM, Clustring.

INTRODUCTION

Data Mining or KDD (Knowledge Discovery in Database) is related with finding hidden facts and knowledge from the large databases or data warehouses. Data mining implements the extensive algorithms to find out the facts those cannot be extracted by simple analysis on the data. The data mining applies the algorithms and methods to find out hidden patterns from the data. This technique helps in making efficient decisions. Data mining can be applied for a variety of fields, retail sales, bioinformatics, pattern reorganization, machine learning, scientific research, weather forecast and so on. Generally data mining uses the artificial intelligence techniques and algorithms to analyze the data. Data mining extracts so called *hidden knowledge* from the data. This kind of facts cannot be derived by simple data analysis.

Data mining techniques include association rules, classifications and clustering. The extracted knowledge can be used for efficient decision making and providing the management with the options those were not opened before.

Data Mining Techniques

There are several techniques used for finding put the hidden patterns and hidden knowledge from the data some of them are as given below:

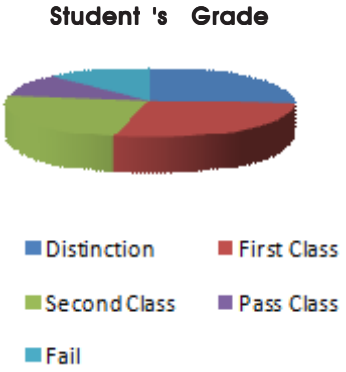
- Classification
- Clustering
- Neural Networks
- Decision Trees
- Association Rules

To better understand these techniques we provide a basic description about them.

Classification

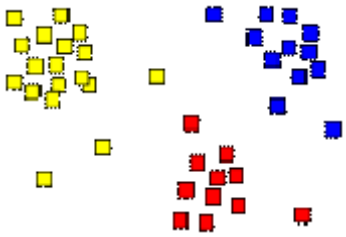
This technique is the commonly applied data mining technique. Classification uses a set of pre-classified examples to develop a model

that can classify the range of records at large. It uses the classification algorithms based on neural networks and decision trees. In classification data are used to estimate the accuracy of the classification rules. The classifier-training algorithm uses these pre-classified examples to determine the set of parameters required for proper discrimination. The algorithm then encodes these parameters into a model called a classifier. There are several algorithms used for classification of data: *ID3 algorithm, C4.5 algorithm, Nearest neighbor, Naïve- Bayes, Decision Table, etc.*



Clustering

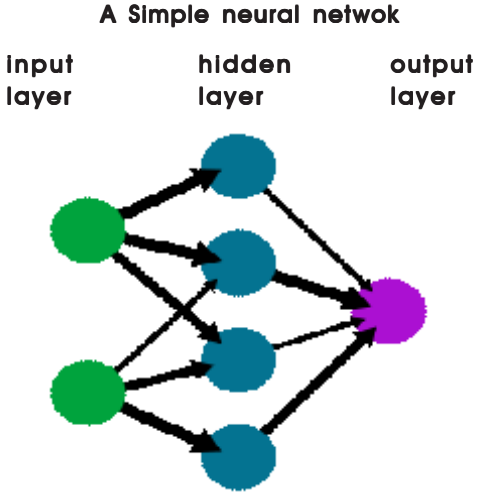
Clustering can be said as the identification of same types of objects. We can find out distribution pattern in data and relations among data attributes. This technique can also be used to differentiate classes of data that can be further used for classification technique.



The above image shows the results in three different colored clusters.

Neural Networks

Neural networks is related to the artificial intelligence in which an interconnected group of nodes based on mathematical or computational model for information processing on the bases of connectionist model. This is the best technique to find out hidden knowledge and patterns from an imprecise data collection. Also they are best at forecasting and predictions. An example of a neural network is shown in the Fig below



Decision Trees

Decision trees are decision support technique which uses a tree-like model of decisions. Decision trees are commonly used for decision analysis. They can be used to generate the rules for classification of datasets.

Association Rules

The association rules are used to find out frequent item sets from a large dataset. Many algorithms can be applied for this. E.g *Apriori Algorithm.*

Data Mining Process

The process of Data Mining consisting of the following steps.

- Selection
- Pre-processing
- Transformation
- *Data Mining*
- Interpretation/Evaluation

Selection

A large data is generally needed for data mining. Generally the data from the large database or the Data Warehouse is used for data mining. A portion or the whole data can be used for knowledge discovery from data warehouse.

Preprocessing

After the selection of the data, now we need to process the data so that we can efficiently use them for our purpose. Generally the real world data is incomplete, inefficient and noisy so we need to remove these faults from the data. Data preprocessing refers to the fact that before using the data for mining, the data must be, complete and efficient. We can provide *data cleaning, data reduction, normalization etc.* for preprocessing the data.

Data cleaning fills the missing values, clears the noise from the data. Data cleaning removes incomplete and inaccurate data from the dataset.

Data reduction is the process of converting the numerical and alpha-numerical data in to correct order. Large amounts of the data are converted down to the meaningful parts.

Normalization is the process in which the data tables in the database are normalized means the redundancy of the data is removed and data tables are converted into more efficient and consistent form. Normalization provides integrated data in the database. According to the purpose we can provide the normalization in several forms. e.g 1st Normal Form, 2nd Normal Form, 3rd Normal Form etc.

Transformation

Before we can go for mining we must make sure that the source data is converted to the format that we want for data mining. Transformation is the process that converts the source data into the appropriate format for data mining.

Data mapping can be done to map the source system data into the destination system data format. This is a very important step in data mining that makes the source data into proper format.

Data Mining

After the completion of these steps listed above it comes data mining. The process of data mining is performed based on several algorithms that hugely analyze the large amounts of the data and finds out the hidden patterns and knowledge from the data.

Interpretation/Evaluation

Data Mining provides the organization/ company management with the *knowledge* that is derived from the *information*. On the basis of this knowledge, the management can take efficient decisions and can also evacuate their previous decisions and can lead their organization to a better future.

Advanced Applications of DM

Data Mining is used for some advanced fields like *Spatial Data Mining:* mining the spatial (geographic) data to find out the patterns to help in the field of GIS (Geographic Information System), Visual Data Mining, *Sensor Data Mining*: using the data from wireless data networks, Pattern Mining etc.

CONCLUSION

For this research paper we can get understand the concept of data mining that is pretty useful for analyzing large data stores. From the historic data we can gain the knowledge about the future trends and also find out the solution of problems that rose in the past. We can say that data mining is the technique to analyze the data and gain the knowledge from that data.

REFERENCES

1. Kantardzic, Mehmed . *Data Mining: Concepts, Models, Methods, and Algorithms*. John Wiley & Sons (2003).
2. Rhind, David W., *Geographic Information Systems: Principles and Applications*,
3. Miller, Harvey J.; and Han, Jiawei; *Geographic Data Mining and Knowledge Discovery*
4. Witten, Ian H.; Frank, Eibe; Hall, Mark A. *Data Mining: Practical Machine Learning Tools and Techniques*
5. Pieter Adriaans.; Dolf Zenting. Data Mining