



Data Mining Techniques: A Short Review

M Shiva Prasad, Asst Professor
Department of CSE, Bharat Institute Of Engineering And Technology,
Hyderabad, INDIA
nhknhk012@gmail.com

Abstract: *The growing needs intelligently analyzing large data sets to acquire useful information gives a view for a new research field called Knowledge Discovery in Databases (KDD) or Data Mining, which attract a attention from researchers in many different fields including database design, statistics, pattern recognition, machine learning, and data visualization. Data mining is the process of discovering insightful, interesting, and novel patterns, as well as descriptive, understandable and predictive models from large-scale data. Data mining involves the tasks like anomaly detection, classification, regression, association rule learning, summarization and clustering. In this paper we overviewed different tasks includes in Data mining.*

Keywords— *Data Mining, classification, clustering, association rules*

I. INTRODUCTION

Data mining consists of extract, transform, and load transaction data onto the data warehouse system, store and manage the data in a multidimensional database system, by using application software analyse the data, provide data access to business analysts and information technology professionals, present the data in a useful format, like a graph or table. Data mining involves the anomaly detection, association, classification, regression, rule learning, summarization and clustering

Data Mining is the process of analysing data from different perspectives and summarizing the results as useful information. It has been defined as "the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data [1]. Data mining is a multi-step process, requires accessing and preparing data for a mining the data, data mining algorithm, analysing results and taking appropriate action. Data mining is the component of wider process called knowledge discovery from database. [1]. The definition of data mining is closely related to another commonly used term knowledge discovery [2].

II. CHALLENGES OF DATA MINING

The main challenges to the data mining and the corresponding considerations in designing the algorithms are as follows:

- Massive datasets and high dimensionality.
- Over fitting and assessing the statistical significance.
- Understandability of patterns.
- Non-standard incomplete data and data integration.
- Mixed changing and redundant data.

III. DATA MINING TASKS

In data mining the data is mined using two learning approaches i.e. supervised learning or unsupervised learning [5]. Data mining as a term used for the specific classes of six activities or tasks as follows:

1. Classification
2. Estimation
3. Prediction
4. Affinity grouping or association rules
5. Clustering
6. Description and visualization

The first three tasks - classification, estimation and prediction rules are examples of directed data mining or supervised learning. In directed data mining, the goal is to use the available data to build a model that describes one or more particular attribute(s) of interest (target attributes or class attributes) in terms of the rest of the available attributes. The next three tasks – association rules, clustering and description are examples of undirected data mining i.e. no attribute is singled out as the target, the main goal is to establish some relationship among all attributes [6].

Classification consists of examining the features of a newly presented object and assigning to it a predefined class. The classification task 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 unclassified data in order to classify it.

Estimation deals with continuously valued outcomes. Given some input data, we use estimation to come up with a value for some unknown continuous variables such as income, height or credit card balance.

Any prediction can be thought of as classification or estimation. The difference is one of emphasis. Predictive tasks feel different because the records are classified according to some predicted future behaviour or estimated future value. With prediction, the only way to check the accuracy of the classification is to wait and see.

An association rule is a rule which implies certain association relationships among a set of objects (such as “occur together” or “one implies the other”) in a database. Given a set of transactions, where each transaction is a set of literals (called items), an association rule is an expression of the form $X \rightarrow Y$, where X and Y are sets of items. The intuitive meaning of such a rule is that transactions of the database which contain X tend to contain Y.

Cluster analysis can be used as a standalone data mining tool to gain insight into the data distribution, or as a pre-processing step for other data mining algorithms operating on the detected clusters. Many clustering algorithms have been developed and are categorized from several aspects such as partitioning methods, hierarchical methods, density-based methods, and grid-based methods. Further data set can be numeric or categorical. Clustering is the task of segmenting a diverse group into a number of similar subgroups or clusters. What distinguishes clustering from classification is that clustering does not rely on predefined classes. In clustering, there are no predefined classes. The records are grouped together on the basis of self similarity. Clustering is often done as a prelude to some other form of data mining or modelling.

Data visualization is a powerful form of descriptive data mining. It is not always easy to come up with meaningful visualizations, but the right picture really can be worth a thousand association rules since the human beings are extremely practiced at extracting meaning from visual scenes. The important application areas of data mining are listed below

Application Areas of Data Mining

<u>Industry</u>	<u>Application</u>
Finance, Banks	Credit Card Analysis Loan approval
Insurance	Claims, Fraud Analysis
Telecommunication	Call record analysis
Transport	Logistics management
Consumer goods	promotion analysis
Retail	Customer Retention
Data Service providers	Value added data
Utilities	Power usage analysis

IV. CONCLUSIONS

Data mining involves extracting useful rules or interesting patterns from huge historical data. Many data mining tasks are available and each of them further has many techniques. In this paper some issues in Data Mining and activities used for Data mining task are discussed.

REFERENCES

- [1] Han, J., Kamber, M., *Data Mining Concepts and Techniques*, Morgan Kaufmann Publisher, 2001
- [2] Pavel Berkhin, *A Survey of Clustering Data Mining Techniques*, pp.25-71, 2002.
- [3] Oded Maimon, Lior Rokach, *Data Mining AND Knowledge Discovery Handbook*, Springer Science + Business Media, Inc, pp.321-352, 2005.
- [4] Pradeep Rai, Shubha Singh, *A Survey of Clustering Techniques*, International Journal of Computer Applications, 2010.+
- [5] K.Kameshwaran, K.Malarvizhi, *Survey on Clustering Techniques in Data Mining*, IJCSIT, Vol. 5, 2014, 2272-2276
- [6] Aastha Joshi, Rajneet Kaur, *A Review: Comparative Study of Various Clustering Techniques in Data Mining*, IJARCSSE, Vol. 3, 2013, 2277-228X