

DATA MINING CONCEPTS AND TECHNIQUES

¹Shilpa.N,²Ammulu.T,³Prameela.N

^{1,2,3} III MCA

^{1,2,3} Er.Perumal Manimekalai College Of Engineering.

Abstract:

This paper provide concepts of data mining .It is the process of patterns and large amount of data. Data mining is a multi disciplinary field machine learning, statistics, pattern recognition, information retrieval, neural networks, knowledge-based systems, artificial intelligence, high-performance computing, and data visualization. Data mining techniques allows to extract hidden knowledge. Data mining techniques can be applied to effectively detect an imposition and report in real time. The purpose of data mining is to identify valid, novel, potentially useful, and understandable correlations and patterns in existing data. The amount of data in medical database need the development of tools which are used to access the data, analyze the data, knowledge discovery, and efficient use of the stored knowledge and information. data mining technology to improve their businesses and found outstanding results. data mining techniques towards multi-class image classification and by using the following two steps procedures includes feature extraction and classifier implementation. Data mining is one of the tasks in the method of knowledge discovery from the database.

KEYWORDS

DATA WAREHOUSING, DATA MINING TECHNIQUES, DATA MINING TOOLS.

I. INTRODUCTION

Now a days, companies collect enormous volumes of data on a daily basis. Analyzing this data and discovering the meaningful information. Data mining solutions concern superior data analysis techniques used by companies for discovering unexpected patterns extracted from vast amounts of data, patterns that offer relevant knowledge for predicting future outcomes. Data mining consists of apply data analysis and discovery algorithms. Technical data mining distinguishes itself in the sense that the temperament of the datasets is often very different from traditional market driven data mining applications. Data mining is a process of removal of useful information and patterns from huge data. It is also called as knowledge discovery process, knowledge mining from data, knowledge taking out. Data mining helps marketing analysts to replicate actual behavior in different situations. Data mining is an effort to source out pattern and trends in the data and infers rules from these patterns. A technique that mines data effectively from a huge

database to get useful information is called data mining. Since data mining extracts valuable information from large data-set, it is useful for various applications, e.g., scientific areas and commercial. Data Mining is used to invent knowledge out of data and exhibiting it in a condition that is easily understandable to humans. It is a process to inspect large amounts of data collected. Information technology plays a vital role for implementing the Data mining techniques in various sectors like banking, education, etc. Data mining is attractive more and more common in both the private and public sectors.

II. DATA MINING TECHNIQUES

Data mining is the use of mechanized data analysis techniques to uncover previously unobserved relationships among data items. Data mining frequently involves the study of data stored in a data warehouse. Data mining is the middle stage of the whole process, it mainly uses the collected mining tools and techniques to deal with the data, thus the rules, patterns and trends will be found.

A. IMPORTANT OF DATA MINING TECHNIQUES

- ❖ Patterns
- ❖ Classification
- ❖ Association
- ❖ Outlier detection
- ❖ Clustering
- ❖ Regression
- ❖ Prediction

➤ *Patterns:*

Patterns is defined as different structural forms such as graphs, trees, or lattices

Which combined with item sets or subsequences.

➤ *Classification:*

Classification trees are used for the kind of Data Mining trouble which are concerned with prediction. Classification is the one of the main role in Data mining. Classification is the mission of generalizing known structure to apply to new data.

➤ *Association:*

Association rule mining represents a data mining technique and its goal is to find interesting association or association relationships along with a large set of data items. Association is correlated to track patterns, but is more exact to dependently linked variables.

➤ *Outlier detection:*

Outlier is a information spot that deviates too much from the rest of dataset. Most of real-world dataset have outlier. Outlier detection acting an essential role in data mining view Outlier Detection is functional in many fields like Network disturbance detection, Credit card format detection, stoke market analysis, detecting remote in wireless sensor network data, fault diagnosis in machines, etc.

➤ *Clustering:*

It is a supervised learning grouping of instances given un-labeled data Clustering the process of grouping physical or abstract objects into classes of similar objects.

➤ *Regression:*

Regression is used to predict a numeric or constant value while classification assigns data into distinct categories. regression is linear decay used to estimate a relationship between two variables.

➤ *Prediction:*

Predicting the personality of one thing based entirely on the explanation of another related thing. Prediction is one of the most valuable data mining techniques, since it's used to project the types of data you'll see in the future. In a lot of cases, just recognize and accepting historical trends is enough to chart a to some extent accurate prediction of what will happen in the future.

III. Data warehousing

A data warehouse is a subject-oriented, included, time-alternative and non-unstable collection of data in support of management's decision making process. Data warehousing is the development of constructing and via the data warehouse. A data warehouse is constructed by integrating the data from multiple mixed sources. It supports systematic coverage, structured and/or ad hoc queries, and decision making.

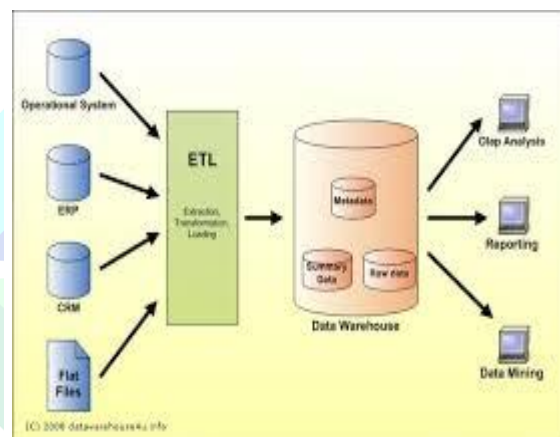


FIG 1:Data Warehousing

A. Types of data warehouse

- Enterprise Data Warehouse
- Virtual Warehouse
- Data mart

i. **Enterprise Warehouse:**

covers all areas of interest for an organization

ii. **Data Mart:** covers a subset of corporate-wide data that is of interest for a specific user group (e.g., marketing).

iii. **Virtual Warehouse:** offers a set of views constructed on

demand on operational databases. Some of the views could

be materialized (precomputed).



FIG:2Types Of Data Warehouse

and data sources around the screen and
FIG3: Data Mining Tools

A. Rapid miner

Rapid Miner is one of the best predictive analysis system developed by the company with the same name as the Rapid Miner. Rapid Miner comes with model based frameworks that allow speedy delivery with reduced number of errors.

B. Mahout

Mahout is an open source machine learning files from Apache. The algorithms it apparatus fall under the broad umbrella of machine learning or collective intelligence. This can denote many things, but at the moment for Mahout it means mainly recommender engines clustering, and classification.

C. Orange

Orange can read files in native tab-delimited format, or can fill data from any of the major normal spreadsheet file types, like CSV and Excel. Native design starts with a header row with characteristic (column) names. The second title row gives the attribute type, which can be continuous, discrete, time, or string.

IV. Data mining tools

The Most Popular Data mining tools and Applications

- Rapid miner
- Mahout
- Orange
- Weka
- Data melt

join them together into the configuration you want.

D. Weka

WEKA provides implementations of culture algorithms that you can simply relate to your dataset. WEKA contains some incremental algorithms that can be used to process very large datasets. The Knowledge Flow interface lets you drag boxes representing learning algorithm

E. Datamelt

DataMelt is an crack to create a data-study situation using open-source association with a coherent user interface and tools competitive to commercial programs.

DataMelt has its roots in element physics wherever data mining is a prime task.

V. Conclusion

This paper gives a general introduction of data mining, the process of discover interesting knowledge from great amount of data stored in information repositories. It also discusses conditions on data mining and methods to combine uncertainty in data mining. Data mining has large purpose grassland almost in every industry where the data is generated that's why data mining is exact one of the most important frontiers in database and

information systems and one of the most promising interdisciplinary developments in Information Technology.

IJCSMC, Vol. 6, Issue. 3, March 2017.

[3] Annaluri Sreenivasa Rao, Attili Venkata Ramana, S.Ramakrishna, Implementing the Data Mining Approaches to Classify the Images with Visual Words, ISSN: 2277-3878, Volume-7 Issue.

[4] Tariq O. Fadl Elsid, Mirghani. A. Eltahir, Data Mining: Classification Techniques of Students' Database A Case Study of the Nile Valley University, North Sudan, (*IJCTT*) – volume 16 number 5 – Oct 2014.

Reference

[1] Neelamadhab Padhy, Dr. Pragnyaban Mishra , and Rasmita Panigrahi, The Survey of Data Mining Applications And Feature Scope, *IJCSC* Vol.2, No.3, June 2012.

[2] Sumitha Thankachan, Suchithra, Data Mining & Warehousing Algorithms and its Application in Medical Science - A Survey,

[5] Hemlata Sahu, Shalini Shurma, Seema Gondhalakar, Data Mining: Classification Techniques of Students' Database A Case Study of the Nile Valley University, North Sudan, (*IJCTT*) – volume 16 number 5 – Oct 2014

