International Journal of Innovative Research in Science and Engineering
Vol. No.3, Issue 02, February 2017
www.ijirse.com
ISSN (O) 2454-9665
ISSN (P) 2454-0663

AN OVERVIEW OF DATA MINING APPLICATION

K. Deviga, N. Krishnaveni, D.Sangeetha,

[1]Department of Computer Science, Sri Krishanasamy Arts and Science College,Mettamalai, (India)
[2]Department of Computer Science and Engineering, P.S. R Engineering College, Sivakasi, (India)
[3]Department of Electronics and Communication Engineering
B.C.M.W polytechnic college ,Ettayapuram, Tamilnadu, (India)

ABSTRACT

There is a huge amount of data available in theinformation industry.This data is no use until it convert into useful information. It is necessary to analyze huge amount of data and extract useful information from it. Data Mining converts the raw data into useful information in various research fields. It helps in finding the patterns to decide future trends in various fields.

Keyword: Data Mining, Information Prediction, Raw Data, Data Mining Techniques, Data Mining Tools

I INTRODUCTION

Data mining is defined as extracting information from huge set of data. Data ming is the procedure of mining knowledge from database, the information or knowledge is extracted. Development of information technology has created large amount of data-base and huge amount of data in various research fields. To research in knowledge mining has give rise to store data and manipulate previously stored data for further decision making process.

1.1 Association Mining

Association Rule Mining

Finding frequent patterns, associations, correlations, or causal structures between sets of items or objects in transaction databases, relational databases, and other information repositories.

Applications

Basket data analysis, cross-marketing, catalog design, loss-leader analysis, clustering, classification, etc.

- Examples.

o Rule form: ―Body ® Head [support,

ISSN (O) 2454-9665
ISSN (P) 2454-0663

confidence]‖.

- buys(x, ―diapers‖) ® buys(x, ―beers‖) [0.5%, 60%]
- major(x, ―CS‖) ^ takes(x, ―DB‖) ® grade(x, ―A‖) [1%, 75%]

1.2 Association Rule: Basic Concepts

Given: (1) database of transactions, (2) each transaction is a list of items. (purchased by a customer in a visit)

Find: all rules that correlate the presence of one set of objects with that of another set of items.

E.g., *98% of people who purchase tires and auto accessories also get automotive services done*

Applications

Maintenance Agreement (What the store should do to boost Maintenance Agreement sales)

Home Electronics (What other products should the store stocks up?)

Attached mailing in direct marketing

Detecting ping-pong‖ing of patients, faulty collisions‖

II DATA MINING PROCESS

Data mining is used to extract implicit and previously unknown information from data. Data mining is the process which provides a intention to attract attention of users due to high availability of huge amount of data and need to convert such data into useful information. So, many people use the expression ―Knowledge Discovery Device‖ or KDD for data mining. Knowledge extraction or discovery is done in seven sequential steps used in data mining.

1. Data cleaning: In this step, remove noise and inconsistent data is removed from collected raw data,
2. Data integration: In this step, multiple data sources are combined into single data store called target data. Data Selection: In this step, data relevant to analysis task are retrieved from data base as pre-processed data.
3. Data transformation: In this step, data is transformed or consolidated into standard forms appropriate for mining by performing summary and aggregation operations.
4. Data Mining: In this step, various intelligent methods are applied in order to extract data pattern or rules.
5. Pattern evaluation: At this step, data patterns are evaluated.
6. Knowledge presentation: In this step knowledge is represented using representation techniques.

The aim of knowledge discovery and data mining process is to find the patterns that are hidden among the huge set of data and interpret useful knowledge and information.

ISSN (O) 2454-9665
ISSN (P) 2454-0663

III DATA MINING APPLICATIONS

1. Market based analysis:
   Customer profiling, retention, identification of potential customer, market segmentation.
2. Fraud detection:
   Identify credit card fraud and intrusion detection.
3. Scientific data analysis:
   Identify the research decision making data.
4. Production Control:
5. Text and web mining: used to search text or information on web or given raw data. Any other applications that involve large amount of data.

IV DATA MINING SYSTEMS

There is a large variety of data mining system available. Data mining systems may integrate techniques from the following:-

- Spatial data analysis
- Information retrieval
- Pattern recognition
- Image analysis Signal processing
- Computer graphics
- Web technology
- Business
- Bio information System

V DATA MINING TECHNIQUES

1. Association: Association is one of the best known data mining technique. In association, a pattern is discovered based on a relationship between items in the same transaction. That's why association technique is also known as relation technique. The association technique is used in market basket analysis to identify a set of products that customers frequently purchase together. Association rules mining has many applications other than market basket analysis, including applications in marketing, customer segmentation, medicine, electronic commerce, bioinformatics and finance. The patterns discovered with this data mining technique can be represented in the form of association rules.

2. Classification: Classification is a classic data mining technique based on machine learning. Essentialy classification is used to classify each item in a set of data into one of predefined set of classes or groups. Classification method frame the use of mathematical techniques such as decision trees, linear programming, neural network and statistics. In classification, we develop the software that can learn how to classify the data objects into groups. For example, we can apply classification in application that given all datas of employees who left the company, predict who will probably leave the company in a future period. In this case, we divide

ISSN (O) 2454-9665
ISSN (P) 2454-0663

the records of employees into two groups that named leave and stay. And then we can ask our data mining software to classify the employees into separate groups.

3**.** Prediction**:** The prediction, as it name implied, is one of a data mining techniques that discovers relationship between dependent and independent variables. For instance, the prediction analysis technique can be used in sale to predict profit for the future if we consider sale is an independent variable, profit could be a dependent variable. Then based on the ancient sale and profit data, we can draw a fitted regression curve that is used for profit prediction.

VI MINING ASSOCIATION RULES IN LARGE DATABASES:

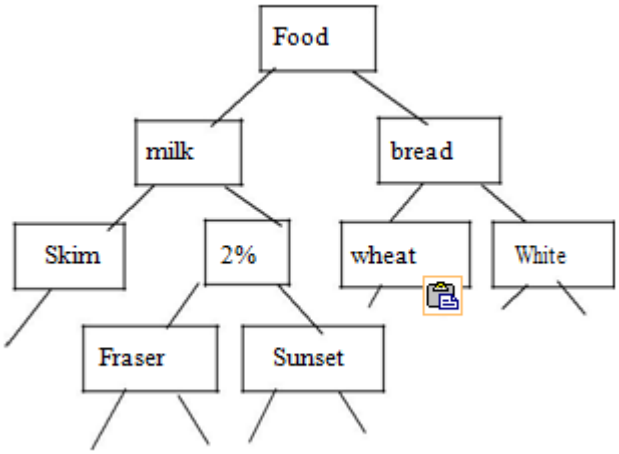
- Association rule mining.
- Mining single-dimensional Boolean association rules from transactional databases.
- Mining multilevel association rules from transactional databases
- Mining multidimensional association laws from transactional databases and data warehouse.
- From association mining to correlation analysis
- Constraint-based association mining.

TID	Items
T1	{111, 121, 211, 221}
T2	{111, 211, 222, 323}
T3	{112, 122, 221, 411}
T4	{111, 121}
T5	{111, 122, 211, 221, 413}

VII MULTIPLE-LEVEL ASSOCIATION RULES

Items often form hierarchy. Items at the lower level are expected to have lower support. Rules regarding item sets at appropriate levels could be quite useful. Transaction database can be encoded based on dimensions and levels. We can explore shared multi-level mining.

ISSN (O) 2454-9665
ISSN (P) 2454-0663



VIII MULTI-LEVEL ASSOCIATION

8.1 UNIFORM SUPPORT VS. REDUCED SUPPORT

Uniform Support: the same minimum support for all levels

o + One minimum support threshold. No need to examine itemsets containing any item whose ancestors do not have minimum support.

o -Lower level items do not occur as frequently. If support threshold

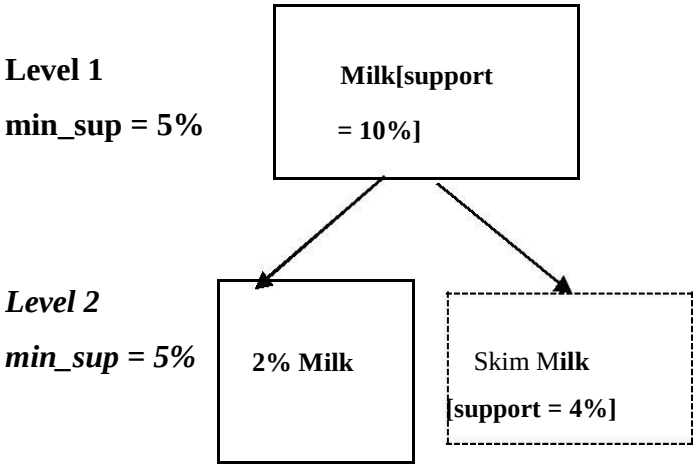
- too high miss low level associations
- too low generate too many high level associations

Reduced Support: reduced minimum support at lower levels

There are 4 search strategies:

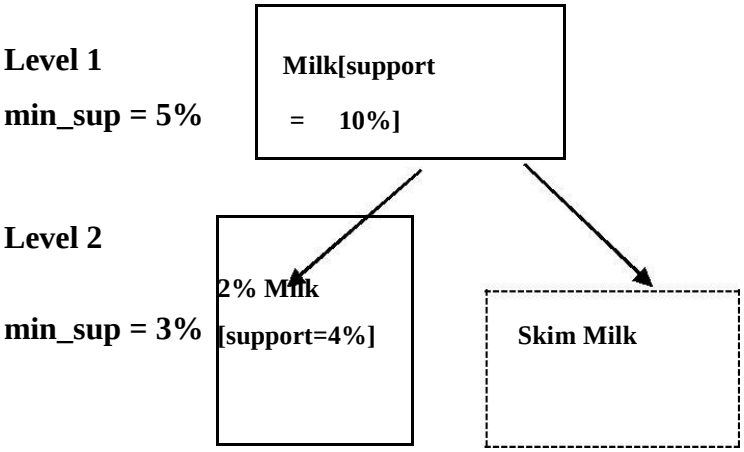
- Level-by-level independent
- Level-cross filtering by k-itemset
- Level-cross filtering by single item
- Controlled level-cross filtering by single item.

8.2 MULTI-LEVEL MINING WITH UNIFORM SUPPORT



International Journal of Innovative Research in Science and Engineering
Vol. No.3, Issue 02, February 2017
www.ijirse.com
ISSN (O) 2454-9665
ISSN (P) 2454-0663

8.3 MULTI-LEVEL MINING WITH REDUCED SUPPORT.



IX MULTI-LEVEL ASSOCIATION

9.1 Redundancy Filtering

- Some rules may be redundant due to ―ancestor‖ relationships between items.
- Example
  - milk wheat bread [support = 8%, confidence = 70%]
  - 2% milk wheat bread [support = 2%, confidence = 72%]
- We say the first rule is an ancestor of the second rule.

X DATA MINING SYSTEM CLASSIFICATION

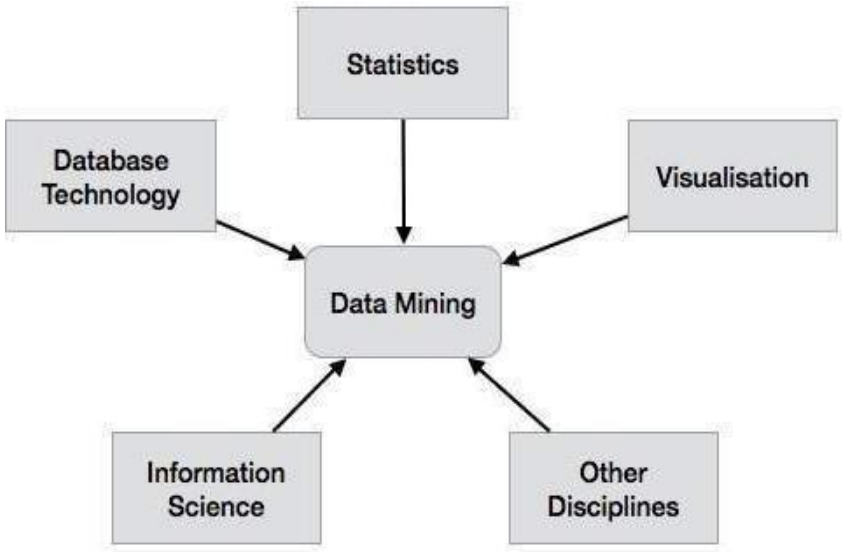


Fig 10.1. Data Mining System Classification

ISSN (O) 2454-9665
ISSN (P) 2454-0663

XI DATA MINING TOOLS

Data mining tools are categorized as stand-alone and client/server solutions. Client/server solutions are dominating.They are specially designed for business users. They are available for different platforms, including Windows, MAC OS, Linux, or special mainframe supercomputers. Many numbers of JAVA-based systems are being developed that are platform-independent for researchers and applied researchers.

11.1 WEKA

The original version of WEKA was non-JAVA and was developed to analyze data from the agricultural domain. the JAVA version of WEKA, l is very sophisticated and used in various applications to visualize , analyze and predict. Its a open ware under the GNU General Public License, Users can customize the tool.

11.2 Rapid Miner

Rapid Miner is Java based tool that offers advanced analytics through template-based frameworks. This Tool has been offered as a service, rather than a local software. Rapid Miner also provides functionalities like data preprocessing and visualization, predictive analytics and statistical modeling, evaluation, and deployment

11.3 R – Programming

R – Programming is developed from C and Fortran. It's a freeware that provide software programming language and software environment for statistical computing and graphics. Data miners to develop statistical software and data analysis with the help of R- programming.

11.4 Orange

Orange, a Python-based, powerful and openware. It has components for machine learning, bioinformatics and text mining. It's wrapped with characteristics for data analytics.

11.5 KNIME

KNIME is a Java based. KNIME does all the three process of extraction, transformation and loading of data .It provides a GUI that allows to assemble the nodes for data processing It is an open source that is able to do data analytics, reporting and integration platform. KNIME also integrates various components for machine learning and data mining through its modular data pipelining concept. It is also able to perform business intelligence and financial data analysis. KNIME is easy to extend and to add plugins.

11.6 NLTK

In this paper the process of KDD and relevance of data mining in various sectors is discussed. The data mining functionalities –predictive mining (classification, regression,

prediction and decision trees) and descriptive minining (clustering, association, summarization) is also summarized. The essential of data mining in commercial, educational, medical, scientific fields are highlighted. NLTK is phython based can be customized. NLTK provides a pool of language processing tools including data mining, machine learning, data scraping, sentiment analysis and other various language processing tasks.

XII CONCLUSION

In this paper the process of Data Mining and its relevance of data mining in various sectors is discussed. The data mining functionalities –predictive mining (classification, regression, prediction and decision trees) and

ISSN (O) 2454-9665
ISSN (P) 2454-0663

descriptive maiming (clustering, association, and summarization) is also summarized. The essential of data mining in market based analysis field is highlighted.

REFERENCES

[1]. Kalyani et al., International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X ,Volume 2, Issue 10, October 2012 .

[2]. Shalini Sharma, Vishal Shrivastava, International Journal on Recent and Innovation Trends in Computing and Communication , ISSN 2321 –8169 Volume: 1 Issue: 4, March 2013.

[3].Megha Gupta, Vishal Shrivastava, International Journal on Recent and Innovation Trends in Computing and Communication, ISSN 2321 – 8169Volume: 1 Issue: 8,August 2013.

[4]. S.Vijiyarani S.Sudha, Disease Prediction in Data Mining Technique – A Survey, International Journal of Computer Applications & Information Technology, ISSN: 2278-7720 Vol. II, Issue I, January 2013 .

[5]. Vili Podgorelec, Peter Kokol, Bruno Stiglic, Ivan Rozman, Decision trees: an overview and their use in medicine, Journal of Medical Systems, Kluwer Academic/Plenum Press,Vol. 26, Num. 5, pp. 445-463, October 2002.

(Books):

[6]. Jiawai Han and Kamber, ―DataMining and Concepts Techniques‖, 2nd March 2006,Chapter.