A Review on Data Mining and its Applications

Karishma[1,] Joni Birla[2]
[1,2]Department of Computer Science &Engineering,
Ganga Institute of Technology and Management,
Kablana, Jhajjar, Haryana, India

Abstract— **In this paper, we describe the data mining and its applications. Data mining is a relatively new technology that has not fully matured. Despite this, there are a number of industries that are already using it on a regular basis. Some of these organizations include retail stores, hospitals, banks, and insurance companies. Many of these organizations are combining data mining with such things as statistics, pattern recognition, and other important tools. Data mining can be used to find**

patterns and connections that would otherwise be difficult to find. This technology is popular with many businesses because it allows them to learn more about their customers and make smart marketing decisions. Here is overview of business problems and solutions found using data mining technology.
Keywords—**Data mining, Customer relationship management, Literature review, Classification**

I. INTRODUCTION

Data Mining is widely used in diverse areas. There are number of commercial data mining system available today yet there are many challenges in this field. In this tutorial we will applications and trend of Data Mining.

Data mining is the process of analyzing data from different perspectives and summarizing it into useful information formation that can be used to increase revenue, cuts costs, or both.

As massive amount of data is continuously being collected and stored, many industries are becoming interested in mining some patterns (association rules, correlations, clusters etc) from their database. Association rule mining is one of the important tasks that are used to find out the frequent itemset from customer transactional database. Each transaction consists of items purchased by a customer in a visit.

Data mining has its own tools and techniques to mine interesting information. When these tools and techniques are applied to the World Wide Web [as is or with some modifications and adaptations for the www environment], it can be called as Internet Mining.

So, Internet mining refers to discovery and analysis of useful information over the World Wide Web.

Internet mining can be broadly classified into three categories:
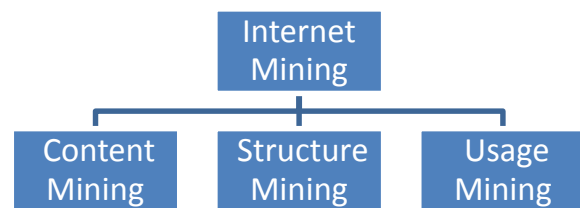
- Content Mining
- Structure Mining
- Usage Mining

Fig 1. Types of Internet Mining

A. *Content Mining:*Content Mining refers to mining of desired content over World Wide Web. Various search engines exists for the content mining, such as altavista, Lycos, WebCrawlar, MetaCrawlar etc.

B. *Structure Mining:*Structure mining tries to discover the link structure of the hyperlinks at the inter-document level to generate structural summary about the Website and Web page.

C. *Usage Mining:*Usage Mining refers to automatic knowledge mining of user access patterns from web servers. It includes,

II. DATA MINING APPLICATIONS

This tutorial discusses about the data mining applications in various areas including sales/marketing, banking, insurance, health care, transportation and medicine.
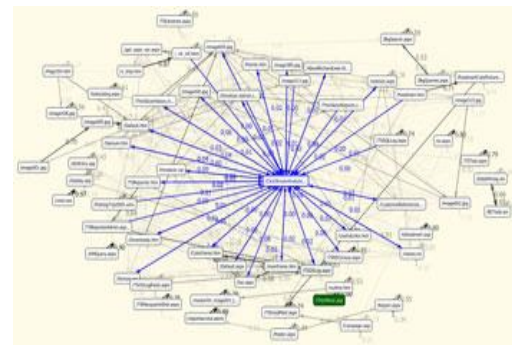

Fig 2. Data mining applications

Data mining is a process that analyzes a large amount of data to find new and hidden information that improves business efficiency. Various industries have been adopt

data mining to their mission-critical business processes to gain competitive advantages and help business grows. This tutorial illustrates some data mining applications in sale/marketing, banking/finance, health care and insurance, transportation and medicine.

III. DATA MINING APPLICATIONS IN SALES/MARKETING

Data mining enables businesses to understand the hidden patterns inside historical purchasing transaction data, thus helping in planning and launching new marketing campaigns in prompt and cost effective way. The following illustrates several data mining applications in sale and marketing.

Data mining is used for market basket analysis to provide information on what product combinations were purchased together, when they were bought and in what sequence. This information helps businesses promote their most profitable products and maximize the profit. In addition, it encourages customers to purchase related products that they may have been missed or overlooked.

Retail companies uses data mining to identify customer's behavior buying patterns.

IV. DATA MINING APPLICATIONS IN BANKING / FINANCE

Several data mining techniques e.g., distributed data mining have been researched, modeled and developed to help credit card fraud detection.

Data mining is used to identify customers loyalty by analyzing the data of customer's purchasing activities such as the data of frequency of purchase in a period of time, total monetary value of all purchases and when was the last purchase. After analyzing those dimensions, the relative measure is generated for each customer. The higher of the score, the more relative loyal the customer is.

To help bank to retain credit card customers, data mining is applied. By analyzing the past data, data mining can help banks predict customers that likely to change their credit card affiliation so they can plan and launch different special offers to retain those customers.

Credit card spending by customer groups can be identified by using data mining.

The hidden correlation's between different financial indicators can be discovered by using data mining.

From historical market data, data mining enables to identify stock trading rules.

Data Mining Applications in Health Care and Insurance

The growth of the insurance industry entirely depends on the ability of converting data into the knowledge, information or intelligence about customers, competitors and its markets. Data mining is applied in insurance industry lately but brought tremendous competitive advantages to the companies who have implemented it

successfully. The data mining applications in insurance industry are listed below:

Data mining is applied in claims analysis such as identifying which medical procedures are claimed together.

Data mining enables to forecasts which customers will potentially purchase new policies.

Data mining allows insurance companies to detect risky customers' behaviour patterns.

Data mining helps detect fraudulent behaviour.

Data Mining Applications in Transportation

Data mining helps determine the distribution schedules among warehouses and outlets and analyse loading patterns.

V. DATA MINING APPLICATIONS IN MEDICINE

Data mining enables to characterize patient activities to see incoming office visits.

Data mining helps identify the patterns of successful medical therapies for different illnesses.

Data mining applications are continuously developing in various industries to provide more hidden knowledge that increases business efficiency and grows businesses.

Data Mining Application in a Software Project Management Process

Richi Nayak
School of Information Systems, QUT, Brisbane

Tian Qiu
EDS Credit Services, Adelaide, Australia

Fig 3. Data mining applications in S/W mgt.

VI. FINANCIAL DATA ANALYSIS

The financial data in banking and financial industry is generally reliable and of high quality which facilitates the systematic data analysis and data mining. Here are the few typical cases:

Design and construction of data warehouses for multidimensional data analysis and data mining. Loan payment prediction and customer credit policy analysis.Classification and clustering of customers for targeted marketing.Detection of money laundering and other financial crimes.

VII. RETAIL INDUSTRY

Data Mining has its great application in Retail Industry because it collects large amount data from on sales, customer purchasing history, goods transportation, consumption and services. It is natural that the quantity of data collected will continue to expand rapidly because of increasing ease, availability and popularity of web.

The Data Mining in Retail Industry helps in identifying customer buying patterns and trends. That leads to improved quality of customer service and good customer retention and satisfaction. Here is the list of examples of data mining in retail industry:

Design and Construction of data warehouses based on benefits of data mining. Multidimensional analysis of sales, customers, products, time and region.Analysis of effectiveness of sales campaigns.

VIII. CUSTOMER RETENTION.

Product recommendation and cross-referencing of items. Telecommunication Industry

Today the Telecommunication industry is one of the most emerging industries providing various services such as fax, pager, cellular phone, Internet messenger, images, e-mail, web data transmission etc. Due to the development of new computer and communication technologies, the telecommunication industry is rapidly expanding. This is the reason why data mining is become very important to help and understand the business.

Data Mining in Telecommunication industry helps in identifying the telecommunication patterns, catch fraudulent activities, make better use of resource, and improve quality of service. Here is the list examples for which data mining improve telecommunication services:

IX. MULTIDIMENSIONAL ANALYSIS OF TELECOMMUNICATION DATA.

- Fraudulent pattern analysis.
- Identification of unusual patterns.
- Multidimensional association and sequential patterns analysis.
- Mobile Telecommunication services.
- Use of visualization tools in telecommunication data analysis.
-

X. BIOLOGICAL DATA ANALYSIS

Now a days we see that there is vast growth in field of biology such as genomics, proteomics, functional Genomics and biomedical research .Biological data mining is very important part of Bioinformatics. Following are the aspects in which Data mining contribute for biological data analysis:

Semantic integration of heterogeneous , distributed genomic and proteomic databases. Alignment, indexing , similarity search and comparative analysis multiple nucleotide sequences.

Discovery of structural patterns and analysis of genetic networks and protein pathways. Association and path analysis. Visualization tools in genetic data analysis.

XI. OTHER SCIENTIFIC APPLICATIONS

The applications discussed above tend to handle relatively small and homogeneous data sets for which the statistical techniques are appropriate. Huge amount of data have been collected from scientific domains such as geosciences, astronomy etc. There is large amount of data sets being generated because of the fast numerical simulations in various fields such as climate, and ecosystem modeling, chemical engineering, fluid dynamics etc. Following are the applications of data mining in field of Scientific Applications:

Data Warehouses and data preprocessing.

XII. GRAPH-BASED MINING

Visualization and domain specific knowledge.

Intrusion Detection : Intrusion refers to any kind of action that threatens integrity, confidentiality, or availability of network resources. In this world of connectivity security has become the major issue. With increased usage of internet and availability of tools and tricks for intruding and attacking network prompted intrusion detection to become a critical component of network administration. Here is the list of areas in which data mining technology may be applied for intrusion detection:

Development of data mining algorithm for intrusion detection.

Association and correlation analysis, aggregation to help select and build discriminating attributes.

- Analysis of Stream data.
- Distributed data mining.
- Visualization and query tools.

XIII. DATA MINING SYSTEM PRODUCTS

There are many data mining system products and domain specific data mining applications are available. The new data mining systems and applications are being added to the previous systems. Also the efforts are being made towards standardization of data mining languages.

Choosing Data Mining System

Which data mining system to choose will depend on following features of Data Mining System:

Data Types - The data mining system may handle formatted text, record-based data and relational data. The data could also be in ASCII text, relational database data or data warehouse data. Therefore we should check what exact format, the data mining system can handle.

XIV. SYSTEM ISSUES

We must consider the compatibility of Data Mining system with different operating systems. One data mining system may run on only on one operating system or on

Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCETEMS-2015 Conference Proceedings

several. There are also data mining systems that provide web-based user interfaces and allow XML data as input.

Data Sources - Data Sources refers to the data formats in which data mining system will operate. Some data mining system may work only on ASCII text files while other on multiple relational sources. Data mining system should also support ODBC connections or OLE DB for ODBC connections.

Data Mining functions and methodologies - There are some data mining systems that provide only one data mining function such as classification while some provides multiple data mining functions such as concept description, discovery-driven OLAP analysis, association mining, linkage analysis, statistical analysis, classification, prediction, clustering, outlier analysis, similarity search etc.

Coupling data mining with databases or data warehouse systems - Data mining system need to be coupled with database or the data warehouse systems. The coupled components are integrated into a uniform information processing environment. Here are the types of coupling listed below:

- No coupling
- Loose Coupling
- Semi tight Coupling
- Tight Coupling

XV.
XVI.   SCALABILITY - THERE ARE TWO SCALABILITY ISSUES IN DATA MINING AS FOLLOWS:

Row (Database size) Scalability - Data mining System is considered as row scalable when the number or rows are enlarged 10 times, It takes no more than the 10 times to execute the query.

Column (Dimension) Salability - Data mining system is considered as column scalable if the mining query execution time increases linearly with number of columns.

Visualization Tools - Visualization in Data mining can be categorized as follows:

- Data Visualization
- Mining Results Visualization
- Mining process visualization
- Visual data mining

Data Mining query language and graphical user interface - The graphical user interface which is easy to use and is required to promote user guided, interactive data mining. Unlike relational database systems data mining systems do not share underlying data mining query language.

Trends in Data Mining

Here is the list of trends in data mining that reflects pursuit of the challenges such as construction of integrated and interactive data mining environments, design of data mining languages:

- Application Exploration
- Scalable and Interactive data mining methods

- Integration of data mining with database systems, data warehouse systems and web database systems.
- Standardization of data mining query language
- Visual Data Mining
- New methods for mining complex types of data
- Biological data mining
- Data mining and software engineering
- Web mining
- Distributed Data mining
- Real time data mining
- Multi Database data mining
- privacy protection and Information Security in data mining.

XVII.   CONCLUSION

In this paper we have study the different types of data mining and there application. There are few components that are : Structure mining, User Mining, Data mining applications- In Banking, In Marketing, In Retail Industry, Data Mining in Medicines, In Financial, Customer Retention, Graphic base mining, Other scientific application and etc..

REFERENCE

[1] Kantardzic, Mehmed (2003). Data Mining: Concepts, Models, Methods, and Algorithms. John Wiley & Sons. ISBN 0-471-22852-4. OCLC 50055336.
[2] "Microsoft Academic Search: Top conferences in data mining". Microsoft Academic Search.
[3] "Google Scholar: Top publications - Data Mining & Analysis". Google Scholar.
[4] Proceedings, International Conferences on Knowledge Discovery and Data Mining, ACM, New York.
[5] SIGKDD Explorations, ACM, New York.
[6] Gregory Piatetsky-Shapiro (2002) KDnuggets Methodology Poll
[7] Gregory Piatetsky-Shapiro (2004) KDnuggets Methodology Poll
[8] Gregory Piatetsky-Shapiro (2007) KDnuggets Methodology Poll
[9] Óscar Marbán, Gonzalo Mariscal and Javier Segovia (2009); A Data Mining & Knowledge Discovery Process Model. In Data Mining and Knowledge Discovery in Real Life Applications, Book edited by: Julio Ponce and Adem Karahoca, ISBN 978-3-902613-53-0, pp. 438–453, February 2009, I-Tech, Vienna, Austria.
[10] Lukasz Kurgan and Petr Musilek (2006); A survey of Knowledge Discovery and Data Mining process models. The Knowledge Engineering Review. Volume 21 Issue 1, March 2006, pp 1–24, Cambridge University Press, New York, NY, USAdoi:10.1017/S0269888906000737
[11] Azevedo, A. and Santos, M. F. KDD, SEMMA and CRISP-DM: a parallel overview. In Proceedings of the IADIS European Conference on Data Mining 2008, pp 182–185.