Current Data Mining Applications

Deepti Dalal
Worked as an Assistant Professor, KIIT,
Gurgaon, India

Abstract— Data mining is a process of discovering knowledge by analyzing the data store from various perspectives in huge repositories, and then the result is summarized into useful information. Data mining tools and techniques are used widely to analyze the data to give useful information which is previously unknown or unseen. The aim of this paper is to present the broad areas in the field of data mining applications which will be helpful in the further research work.

Keywords— IT, Knowledge discovery process, Data mining, Knowledge discovery, Semantic web

I.   INTRODUCTION

The advancement of Information Technology (IT) has generated huge amount of databases in various areas. The research in the field of IT and databases has led to an approach to store this huge data in data warehouses and manipulate this valuable data for further decision making. It is also known as knowledge discovery process, knowledge extraction, knowledge mining from data, or data /pattern analysis. Data mining technology is useful in automatically predicting trends, behaviours, patterns, connections, forecasting, knowledge discovery in large databases quickly, which are previously unknown or hidden that would otherwise be difficult to find. Data mining technique is extracting similar patterns, valuable information, and hidden patterns in the databases which are helpful in business decision making process. Data mining techniques are used by a number of applications to handle various challenges encountered in the business. It is used mainly in fraud detection, retail, health care, finance, telecommunication, and risk analysis. Industries and organizations adopt data mining tools and techniques to generate business intelligence so as to make improvements in the process that is required for the decision making. The different tools, techniques and methods of data mining are used to find patterns and the knowledge from these huge databases. Most of the organizations combined the data mining with other important tools such as statistics and pattern recognition. A number of organizations which are using data mining applications in their business are FBTO Dutch Insurance Company to increase efficiency in marketing and increasing cross selling to customers, ECtel Ltd., Israel to reduce fraudulent activities in more than 150 telecommunication services, . Provident Financial's Home credit Division, United Kingdom to detect and prevent fraud, Standard Life Mutual Financial Services Companies for identifying key customer types and many more.

II.   DATA MINING APPLICATIONS

A.   Healthcare And Bio-Informatics

There is a tremendous potential and use of data mining applications in healthcare which depends on the availability of clean healthcare data sharing across organizations and standardization of clinical vocabulary. It is also useful to look into how images (e.g., MRI scans) can be brought into healthcare data mining applications *(*Padhy *et al.* **2012)**. It is difficult to analyze the medical data because of its complexity and sensitivity. Data mining is proved to be very helpful in mining the medical data in health care. Data mining is used in diagnosing the diesis, health care, patient profile and history making. In the detection of breast cancer mammography is used. In mammograms, for the classification of tumor, association rule mining and neural networks with back-propagation are used. Cardiovascular disease refers to a condition that affects the heart. Kumari and Godara, 2011 compared different classification techniques of data mining to predict cardiovascular disease in patients on basis of Sensitivity, Specificity, Accuracy, Error Rate, True Positive Rate and False Positive Rate. According to their study Support Vector Machine model proved to be the best classifier technique for cardiovascular disease prediction. Patient's risks and cost of diagnosis can be reduced significantly by using the data mining algorithms. Bio-informatics deals in finding patterns in genes to cure several life threatening diseases like Cancer and AIDS. Data mining tools and techniques are used in interpreting biological structures and sequences. With DNA analysis it is now possible to detect genetic causes for diseases and disabilities in addition to diagnosis, prevention and treatment of disease.

B.   Business trends

Modern business is more dynamic it must offer high quality services and customer satisfaction. For profitable business and customer retention it must be fast in its processing. Due to increasing demand of e-commerce there is a demand to store enormous data related to different areas including sales, goods production, transportation, consumption, services offered, and customer's shopping history. Data mining techniques like classification, cluster analysis and regression are widely used in business process to support business in its decision making, enabling accurate customer's transaction,

discovering customer's shopping patterns, resulting in a strong Business Intelligence system. With Data mining technique we can find the associations between different items in Market Based Analysis. It helps the customer to match their requirements and helps to purchase the related items whereas retailers can identify customer's intension of buying different patterns. In manufacturing to control the process and to improve the quality of product first the data is analyzed and then hidden patterns can be found. Data mining can help in finding errors in the data, improving design methodology. As it is difficult to understand relationship identification in manufacturing engineering,   CRISP-DM methodology was proposed to provide the detailed steps with instructions at high level.

C. Education System

Data mining technology in education helps to advance the education system by facilitating good understanding of the students. It helps the teachers to better deal with their classes and to give hands-on feedback and guide the students. Kumar and Chadha, 2011 discussed various data mining techniques by taking real life examples which proved to be helpful to improve education system by generating strategic information. A similar study of Prabha and Shanavas, 2013 is based on the experiment done on 6th grade students which shows that data mining helps the teacher by identifying the week areas of the students. Weak areas are find when the knowledge level of students are identified and grouped accordingly and then it is analyzed that which area of a student is weak and where is to concentrate. Data mining in education helps the student, teacher and the administrators; students enhance skills by monitoring their own learning. Teachers obtain views about the performance of students which helps the teachers in their teaching accordingly. Administrators use data mining to plan policies and programs to improve teaching, learning, and retention of students. CART is the best algorithm for classification of data which will help the students and the teachers to improve the performance of the students (Yadav *et al.* 2013).

D. Agriculture Yield Data

In agriculture data mining techniques are used to forecast weather conditions, to study soil characteristics, pollution in the atmosphere, to examine color images of fruits, to classify eggs fertility, to recognize cracks in eggs. To smell the milk, sensors are generally used whereas some taste sensors are used to get data from the process of fermentation. Ramesh and Vardhan, 2013 conducted a research on data collected from East Godavari district of Andhra Pradesh in India where crop yield is analyzed   with respect to four parameters namely Year, Rainfall, Area of Sowing and Production. To estimate crop yield analysis certain Data Mining techniques were applied on the data collected. K-Means algorithm is a good example of one such type in which clusters are made by partitioning the samples. Biclustering can provide this kind of information (Ramesh and Vardhan 2013).

E. Web Mining And Semantic Web

Text mining is searching large volume of files with keywords. Web mining integrates data and text mining in a website and hence web mining is an extension of text mining. It is the enhancement of the web site with intelligent conduct, like related links are suggested and new products are recommended to the customers according to their requirements and also some suggestions are made according to their choice for a particular search item. Search engine used in websites to search particular items according to requirements is an example of data mining. The network problems like tampering websites, webpage hang horse, phishing, Trojan horse infected host have been flourished (*Deshpande and Thakare 2010)*. Web mining is used for network security in intrusion detection system to protect the host from attack. There are three modules in the system, these are first- data acquisition module, the second is intrusion analyzing engine and the third is emergency response. Unstructured web can be organized with data mining technology turning it into a semantic web. The technology basically applied is Resource Description Framework (RDF), a technology used in describing resources. FOAF technology is used a lot for tagging in Orkut and Facebook.

F. Finance

Banks and financial institutions offer a large number of banking, credit, investment, insurance stock and services like checking, saving, transactions, mortgage, automobile loans and mutual funds. In bank, the credit department has to process huge credit data of large number of customers but analyzing this huge data is really a tough deal in terms of both manpower and economically. So due to credit industry's recent growth, the Credit scoring has become an important matter. A number of data mining techniques are used in credit risk evaluation problem like classification, support vector machine and several single and hybrid data mining methods. Data mining also help in fraud detection and thereby helps in detecting people to collect on insurance money.

G. Sports

In sports there is a need to record the data for all games and for each player, each team, for each and every match, for all seasons. If this data is recorded in statistical form then data mining can be used to discover the patterns to give useful information and may proved to a game changing decision. This information is useful for coaching, training, planning, scouting, prediction of performance, team selection for each game, and etc. The different techniques of data mining are widely used to detect the best and the optimal combination of team members of a team that may lead the team towards the victory. The 'Cy Young Award' has been presented annually to the best pitcher in the major league of baseball (Smith *et al.* 2007). The award is decided by compiling the statistics recorded over the baseball season course using Bayesian classifier.

H. Detecting Terror related Activities By intelligence Agencies

There is a huge amount of data involved in criminal and terrorist activities. The data is analyzed by the Intelligence Agencies to investigate various terrorist activities. Analyzing this huge data with data mining techniques makes this task easy like email spamming can be detected by the classification technique and deceptive information in criminal record can be found by string comparator. For detecting terror related activities this methodology is based on knowledge by using content related to web traffic as audit information. The behavior of terrorist is learned through implementing web mining algorithm to textual content of terror related websites.

I. Detecting Malicious Executables

A malicious executable is a threat to the security of a system which may damage a system or may obtain sensitive and important information without the permission of user or without his knowledge. Data mining technology is useful in detecting malicious executables before they run into the system and affect it. Some data mining techniques used to detect malicious executables are Classification algorithms, Multi-Classifier system, RIPPER and a Naive Bayes.

J. Language Research

There is a need of linguistic information to be generated about a text in language research. Data mining is helpful in automatically generating this linguistic information from text file and helps in verifying and recognizing the authorship. To select the true modern for the text, syntactic and lexical features in combination are used in a profiling system. The linguistic profiling of text effectively used to control the quality of language and for the automatic language verification (Antonie *et al.* 2001). In fact language verification is possible with this technique which verifies the text automatically whether it is original or not.

K. Telecommunication

Modern telecommunication industry is not limited to offering only local and long distance telephone services but is also involved in fax, pager, mobile phone, images, videos, e-mail, online web conferencing and data transmission. With the integration of telecommunication and internet there is a demand of data mining to identify telecommunication patterns and find fraudulent activities involved to improve the service quality.

L. Multimedia data Mining

Multimedia Data Mining is a new field and has prospectus in future, it includes mining of data comprising images, audio, video and animation. As a number of different multimedia items are used in the mining process so it would be a good idea to represent it in a different form than representing it in a conventional data form. One such idea is creating multimedia data cubes that might be used to change the data type of multimedia into such a form that could be easily understood able and well suited for analyzing by the techniques of data mining, but taking into account the unique characteristics of the data (Paidi, 2012). The unique characteristics may consist of the use of magnitude and measures for the shape, color, texture, and other relevant attributes.

M. Spatial And Geographic Data Mining

Spatial and geographic data have information about the satellites that orbit around the earth and also data that are transmitted by the spacecraft; it may be images or other relevant data of the earth. Astronomical data is mainly image-oriented and contain distance and topological information; if this data is analyzed properly using proper data mining techniques it can represent enormous important information. Spatial information is useful in remote sensing and navigation. Spatial data mining is the mining of relationships among special data. We can find hidden patterns in the knowledge, or some other patterns that are not clearly stored in spatial databases. According to Paidi, 2012 multidimensional structures, special methods to access the data are used to index the special data.

N. Time Series/ Sequence Data Mining

Time series and sequence data mining is the extraction of a sequence of data that may be referenced by time or data ordered in a sequence. It includes identification of movements which exists within the data also known as trend analysis. These can include trend movements, random movements, seasonal variations and cyclical variations. This type of information is useful in analyzing the customer, identifying their buying patterns, their likely follow-up purchases and much more. To extract this kind of information, we can use different techniques like sequential pattern mining, similarity search, and periodicity analysis. Periodicity analysis involves the analysis of data by identifying the various patterns which recur or repeat in a time series.

O. Constraint Based Data Mining

Currently existing data mining techniques lack the advantage of user control or any guidance. Constraint based data mining uses different constraints to guide the process which can be combined with the multidimensional mining benefits to improve the mining process. Constraints may be Knowledge-type, data constraints, Rule constraints or it may be Dimension/level constraints; a Data constraint identifies the data, Knowledge constraints specifically identify the knowledge type that is to be mined, Dimension/level constraints specify the levels or dimensions to be included in the current query, Rule constraints state the specific rules to be applied for a particular data mining query.

III. CONCLUSIONS

Data mining has its importance in pattern finding, forecasting, and discovery of the knowledge in diverse business domain. In this paper the various data mining applications and different techniques are explained that are applied to these areas of application. Data mining has broad application area, almost in every field where data is generated this is why data mining is considered most promising interdisciplinary and one of the most important frontiers in the field of IT and database systems. There are numerous research works in data mining field. This review will help the researchers to focus on the various data mining issues.

REFERENCES
[1]   Antonie M. L., Zaiane O. R., Coman, A., "Application of Data Mining Techniques for Medical Image Classification", Proceedings of the Second International Workshop on Multimedia Data Mining MDM/KDD 2001) in conjunction with ACM SIGKDD conference, San Francisco, 2001.
[2]   Chodavarapu Y., "Using data-mining for effective (optimal) sports squad selections". WebSite:http://insightory.com/view/74//using_datamining_for_effective_(optimal)_sports_squad_selections.
[3]   Deshpande D.S., "A Survey on Web Data Mining Applications". Emerging Trends in Computer Science and Information Technology - (ETCSIT2012), Proceedings published in International Journal of Computer Applications® (IJCA) pp 28-32, 2012.
[4]   Deshpande S. P., Thakare V. M., "Data mining system and applications: A review". International Journal of Distributed and Parallel systems (IJDPS), Vol. 1, No. 1, 2010.
[5]   Kumari M., Godara S., "Comparative Study of Data Mining Classification Methods in Cardiovascular Disease Prediction". International Journal of Computer Science and technology (IJCST), Vol. 2, Issue 2, pp. 304-308, 2011.
[6]   Kumar V., Chadha A., "An Empirical Study of the Applications of Data Mining Techniques in Higher Education". International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 2, No.3, pp. 80-84, 2011.
[7]   Padhy N., Mishra P., Panigrahi R., "The Survey of Data Mining Applications And Feature Scope". International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.3, pp. (43-58), 2012.
[8]   Paidi A. N., "Data Mining: Future Trends and Applications". International Journal of Modern Engineering Research (IJMER) ISSN: 2249-6645, Vol.2, Issue.6, pp. 46-57, 2012.
[9]   Prabha L. S., Shanavas A.R.M., "EDUCATIONAL DATA MINING APPLICATIONS". Operations Research and Applications: An International Journal (ORAJ), Vol. 1, No. 1,pp. 23-29, 2014.
[10]   Ramesh D., Vardhan V. B., " Data Mining Techniques and Applications to Agricultural Yield Data". International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 9, pp. 3477-3480, 2013.
[11]   Smith L., Lipscomb B., and Simkins, A., "Data Mining in Sports: Predicting Cy Young Award Winners". Journal of Computer Science, Vol. 22, pp. 115-121, 2007.
[12]   Yadav S.K., Bharadwaj B., Pal S., "Data Mining Applications: A comparative Study for Predicting Student's performance". International Journal of Innovative Technology and Creative Engineering (ISSN:2045-711),Vol.1, No.12, pp. 13-19, 2013.