DATAMINING TECHNIQUE APPLICATION AND FUTURE SCOPE

Nidhi Trivedi[1]

[1] *B E Student, Computer Department, Govt. Engineering College Gandhinagar, Gujarat, India*

ABSTRACT

In the 21st century the human beings are used in the different technologies to adequate in the society . Each and every day the human beings are using the vast data and these data are in the different fields .It may be in the form of documents, may be graphical formats ,may be the video ,may be records (varying array) .As the data are available in the different formats so that the proper action to be taken. Not only to analyze these data but also take a good decision and maintain the data .As and when the customer will required the data should be retrieved from the database and make the better decision .This technique is actually we called as a data mining or Knowledge Hub or simply KDD(Knowledge Discovery Process).The important reason that attracted a great deal of attention in information technology the discovery of useful information from large collections of data industry towards field of "Data mining" is due to the perception of "we are data rich but information poor". There is huge volume of data but we hardly able to turn them in to useful information and knowledge for managerial decision making in business. To generate information it requires massive collection of data. It may be different formats like audio/video, numbers, text, figures, Hypertext formats . To take complete advantage of data; the data retrieval is simply not enough, it requires a tool for automatic summarization of data, extraction of the essence of information stored, and the discovery of patterns in raw data.

With the enormous amount of data stored in files, databases, and other repositories, it is increasingly important, to develop powerful tool for analysis and interpretation of such data and for the extraction of interesting knowledge that could help in decision-making. The only answer to all above is 'Data Mining'. Data mining is the extraction of hidden predictive information from large databases; it is a powerful technology with great potential to help organizations focus on the most important information in their data warehouses.Data mining tools predict future trends and behaviors, helps organizations to make proactive knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by prospective tools typical of decision support systems. Data mining tools can answer the questions that traditionally were too time consuming to resolve. They prepare databases for finding hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.Data mining, popularly known as Knowledge Discovery in Databases (KDD), it is the nontrivial extraction of implicit, previously unknown and potentially useful information from data in databases. It is actually the process of finding the hidden information/pattern of the repositories.

Keyword: *-Data mining, Data mining life cycle, the data mining model ,Data mining Methods,Data mining applications, Data mining future scope,Clustering methods*

1. INTRODUCTION

In the 21st century the human beings are used in the different technologies to adequate in the society . Each and every day the human beings are using the vast data and these data are in the different fields .It may be in the form of documents, may be graphical formats ,may be the video ,may be records (varying array) .As the data are available in the different formats so that the proper action to be taken. Not only to analyze these data but also take a good decision and maintain the data .As and when the customer will required the data should be retrieved from the database and make the better decision .This technique is actually we called as a data mining or Knowledge Hub or

simply KDD(Knowledge Discovery Process).The important reason that attracted a great deal of attention in information technology the discovery of useful information from large collections of data industry towards field of "Data mining" is due to the perception of "we are data rich but information poor". There is huge volume of data but we hardly able to turn them in to useful information and knowledge for managerial decision making in business. To generate information it requires massive collection of data. It may be different formats like audio/video, numbers, text, figures, Hypertext formats . To take complete advantage of data; the data retrieval is simply not enough, it requires a tool for automatic summarization of data, extraction of the essence of information stored, and the discovery of patterns in raw data.

With the enormous amount of data stored in files, databases, and other repositories, it is increasingly important, to develop powerful tool for analysis and interpretation of such data and for the extraction of interesting knowledge that could help in decision-making. The only answer to all above is 'Data Mining'. Data mining is the extraction of hidden predictive information from large databases; it is a powerful technology with great potential to help organizations focus on the most important information in their data warehouses.Data mining tools predict future trends and behaviors, helps organizations to make proactive knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by prospective tools typical of decision support systems. Data mining tools can answer the questions that traditionally were too time consuming to resolve. They prepare databases for finding hidden patterns, finding predictive information that experts may miss because it lies outside their expectations. Data mining, popularly known as Knowledge Discovery in Databases (KDD), it is the nontrivial extraction of implicit, previously unknown and potentially useful information from data in databases. It is actually the process of finding the hidden information/pattern of the repositories.

2 OVER VIEW OF DATAMINING

The development of Information Technology has generated large amount of databases and huge data in various areas. The research in databases and information technology has given rise to an approach to store and manipulate this precious data for further decision making. Data mining is a process of extraction of useful information and patterns from huge data. It is also called as knowledge discovery process, knowledge mining from data, knowledge extraction or data /pattern analysis.
Data mining is a logical process that is used to search through large amount of data in order to find useful data. The goal of this technique is to find patterns that were previously unknown. Once these patterns are found they can further be used to make certain decisions for development of their businesses. Three steps involved are Exploration Pattern identification Deployment
Exploration: In the first step of data exploration data is cleaned and transformed into another form, and important variables and then nature of data based on the problem are determined.
Pattern Identification: Once data is explored, refined and defined for the specific variables the second step is to form pattern identification. Identify and choose the patterns which make the best prediction.
Deployment: Patterns are deployed for desired outcome.

3. DATA MINIG ALGORITHM AND TECHINQUE

Various algorithms and techniques like Classification, Clustering, Regression, Artificial
Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbour method etc., are used for knowledge discovery from databases.

1) Classification:
Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. Fraud detection and creditrisk applications are particularly well suited to this type of analysis. This approach frequently employs decision tree or neural network-based classification algorithms. The data classification process involves learning and classification. In Learning the training data are analyzed by classification algorithm. In classification test data are used to estimate the accuracy of the classification rules. If the accuracy is

acceptable the rules can be applied to the new data tuples. For a fraud detection application, this would include complete records of both fraudulent and valid activities determined on a record-by-record basis. The classifier-training algorithm uses these pre-classified examples to determine the set of parameters required for proper discrimination. The algorithm then encodes these parameters into a model called a classifier.
Types of classification models:

- Classification by decision tree induction
- Bayesian Classification
- Neural Networks
- Support Vector Machines (SVM)
- Classification Based on Associations

2)Predication:
Regression technique can be adapted for predication. Regression analysis can be used to model the relationship between one or more independent variables and dependent variables. In data mining independent variables are attributes already known and response variables are what we want to predict.
Unfortunately, many real-world problems are not simply prediction. For instance, sales volumes, stock prices, and product failure rates are all very difficult to predict because they may depend on complex interactions of multiple predictor variables. Therefore, more complex techniques (e.g., logistic regression, decision trees, or neural nets) may be necessary to forecast future values. The same model types can often be used for both regression and classification. For example, the CART (Classification and Regression Trees) decision tree algorithm can be used to build both classification trees (to classify categorical response
variables) and regression trees (to forecast continuous response variables). Neural networks too can create both classification and regression models.
Types of regression methods

- Linear Regression
- Multivariate Linear Regression
- Nonlinear Regression
- Multivariate Nonlinear Regression

3) Neural networks:
Neural network is a set of connected input/output units and each connection has a weight present with it.
During the learning phase, network learns by adjusting weights so as to be able to predict the correct class labels of the input tuples. Neural networks have the remarkable ability to derive meaning from complicated or imprecise data and can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. These are well suited for continuous valued inputs and outputs. For example handwritten character reorganization, for training a computer to pronounce English text and many real world business problems and have already been successfully applied in many industries.
Neural networks are best at identifying patterns or trends in data and well suited for prediction or forecasting needs.
Types of neural networks

- Back Propagation

4. DATA MINING APPLICATION

In this section, we have focused some of the applications of data mining and its techniques are analyzed respectively Order.
1) Data Mining Applications in Healthcare:
Data mining applications in health can have tremendous potential and usefulness. However, the success of healthcare data mining hinges on the availability of clean healthcare data. In this respect, it is critical that the healthcare industry look into how data can be better captured, stored, prepared and mined. Possible directions include the standardization of clinical vocabulary and the sharing of data across organizations to
enhance the benefits of healthcare data mining applications.
1.1) Future Directions of Health care system through Data Mining Tools:
As healthcare data are not limited to just quantitative data (e.g., doctor's notes or clinical records), it International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), is necessary to also explore the

use of text mining to expand the scope and nature of what healthcare data mining can currently do. This is specially used to mixed all the data and then mining the text. It is also useful to look into how images (e.g., MRI scans) can be brought into healthcare data mining applications. It is noted that progress has been made in these areas.

2) Data mining is used for market basket analysis:
Data mining technique is used in MBA(Market Basket Analysis).When the customer want to buying some products then this technique helps us finding the associations between different items that the customer put in their shopping buckets. Here the discovery of such associations that promotes the business technique .In this way the retailers uses the data mining technique so that they can identify that which customers intension (buying the different pattern).In this way this technique is used for profits of the business and also helps to purchase the related items.

3) The data mining is used an emerging trends in the education system in the whole world:
In Indian culture most of the parents are uneducated .The main aim of in Indian government is the quality education not for quantity. But the day by day the education systems are changed and in the 21st century a huge number of universalities are established by the order of UGC. As the numbers of universities are established side by side, each and every day a millennium of students are enrols across the country. With huge number of higher education aspirants, we believe that data mining technology can help bridging knowledge gap in higher educational systems. The hidden patterns, associations, and anomalies that are discovered by data mining techniques from educational data can improve decision making processes in higher educational systems. This improvement can bring advantages such as maximizing educational system efficiency, decreasing student's drop-out rate, and increasing student's promotion rate, increasing student's retention rate in, increasing student's transition rate, increasing educational improvement ratio, increasing student's success, increasing student's learning outcome, and reducing the cost of system processes. In this current era we are using the KDD and the data mining tools for extracting the knowledge this knowledge can be used for improving the quality of education .The decisions tree classification is used in this type of applications.

4) Data mining is now used in many different areas in manufacturing engineering:
When we retrieve the data from manufacturing system then the customer is to use these data for different purposes like to find the errors in the data ,to enhance the design methodology ,to make the good quality of the data ,how best the data can be supported for making the decision . But most of time the data can be first analyzed then after find the hidden patterns which will be control the manufacturing process which will further enhance the quality of the products .Since the importance of data mining in manufacturing has clearly increased over the last 20 years, it is now appropriate to critically review its history and Application.

4.1) Future Directions in the manufacturing Engineering through the Data mining Tools:
It is very tedious task to mine the manufacturing data .Generally when we mine the data in the manufacturing, we dose not give more important to the quality of the rules .After mining those knowledge which has generated is very difficult because relationship identification is too complex to understand. That's why we need the further to enhance the research methodology to know the proper knowledge. The new methodology was proposed i.e CRISP-DM which will provides the high level detail steps of instructions for using the data mining in the Data Mining Applications can be generic or domain specific.
Data mining system can be applied for generic or domain specific . Some generic data mining applications cannot take its own these decisions but guide users for selection of data, selection of data mining method and for the interpretation of the results. The multi agent based data mining application has capability of automatic selection of data mining technique to be applied. The Multi Agent System used at different levels: First, at the level of concept hierarchy definition then at the result level to present the best adapted decision to the user. This decision is stored in knowledge Base to use in a later decision-making. Multi Agent System Tool used for generic data mining system development uses different agents to perform different tasks.

5) A multi-tier data mining system is proposed to enhance the performance of the data mining process:
It has basic components like user interface, data mining services, data access services and the data. There are three different architectures presented for the data mining system namely one-tire, Two-tire and Three-tire architecture. Generic system required to integrate as many learning algorithms as possible and decides the most appropriate algorithm to use. CORBA (Common Object Request Broker Architecture) has features like: Integration of different applications coded in any programming language considerably easy. It allows reusability in a feasible way and finally it makes possible to build large and scalable system. The data mining system architecture based on CORBA is given by Object Management Group has all characteristics to accomplish a distributed and object oriented

computation. A data-centric focus and automated methodologies makes data mining accessible to no experts. The use of high-level interfaces can implement the automated methodologies that hide the data mining concepts away from the users. A data-centric design hides away all the details of mining methodology and exposes them through high-level tasks that are goal-oriented. These goal-oriented tasks are implemented using data-centric APIs. This design makes data mining task like other types of queries that users perform on the data. In data mining better results could be obtained if large data is available. It leads to the merging and linking of local databases. A new data-mining architecture based on Internet technology addressed this problem. [12] The context factor plays vital role in the success of data mining. The importance and meaning of same data in the different context is different. A data in one context is very important may not be much important in other context. A context-aware data-mining framework filters useful and interesting context factors, and can produce accurate and precise prediction using those factors.

6)Application of Data Mining techniques in CRM:
Data mining technique is used in CRM .Now a days it is one of the hot topic to research in the industry because CRM have attracted both the practitioners and academics. It aims to give a research summary on the application of data mining in the CRM domain and techniques which are most often used. Although this review cannot claim to be exhaustive, it does provide reasonable insights and shows the incidence of research on this subject. The results presented in this paper have several important implications: Research on the application of data mining in CRM will increase significantly in the future based on past publication rates and the increasing interest in the area. The majority of the reviewed articles relate to customer retention .

7)The Domain Specific Applications:
The domain specific applications are focused to use the domain specific data and data mining algorithm that targeted for specific objective. The applications studied in this context are aimed to generate the specific knowledge. In the different domains the data generating sources generate different type of data. Data can be from a simple text, numbers to more complex audio-video data. To mine the patterns and thus knowledge from this data, different types of data mining algorithms are used. The collection and selection of context specific data and applying the data mining algorithm to generate the context specific knowledge is thus a skilful job. In many domains specific data mining applications the domain experts plays vital role to mine useful knowledge. In the identification of foreign-accented French the audio files were used and the best 20 data mining algorithms were applied the Logistic Regression model found the most robust algorithm than other algorithm.

8) In language research and language:
Engineering much time extra linguistic information is needed about a text. A linguistic profile that contains large number of linguistic features can be generated from text file automatically using data mining. This technique found quite effective for authorship verification and recognition. A profiling system using combination of lexical and syntactic features shows 97% accuracy in selecting correct author for the text. The linguistic profiling of text effectively used to control the quality of language and for the automatic language verification. This method verifies automatically the text is of native quality. The results show that language verification is indeed possible.

9) In Medical Science:
In medical science there is large scope for application of data mining. Diagnosis of diesis, health care, patient profiling and history generation etc. are the few examples. Mammography is the method used in breast cancer detection. Radiologists face lot of difficulties in detection of tumors that's why CAM(Computer Aided Methods) could helps to the medical staff . So that they can produce the good quality of the result detection . The neural networks with back-propagation and association rule mining used for tumor classification in mammograms. The data mining effectively used in the diagnosis of lung abnormality that may be cancerous or benign . The data mining algorithms significantly reduce patient's risks and diagnosis costs. Using the prediction algorithms the observed prediction accuracy was 100% for 91.3% cases. The use of data mining in health care is the widely used application of data mining. The medical data is complex and difficult to analyze. A REMIND (Reliable Extraction and Meaningful Inference from Non-structured Data) system integrates the structured and unstructured clinical data in patient records to automatically create high quality structured clinical data. To adopt the high quality technique, we can mined the existing patient records to support guidelines and give compliance to improve patient care.

10) Data Mining methods are used in the Web Education:
Data mining methods are used in the web Education which is used to improve courseware. The relationships are discovered among the usage data picked up during students' sessions. This knowledge is very useful for the teacher or the author of the course, who could decide what modifications will be the most appropriate to improve the effectiveness of the course.In the 21st century the beginners are using the data mining techniques which is one of the best learning method in this era. This makes it possible to increase the awareness of learners. Web Education which will rapidly growth in the application of data mining methods to educational chats which is both feasible and can be improvement in learning environments in the 21st century.

5. Data Mining Life Cycle

The life cycle of a data mining project consists of six phases[2,4]. The sequence of the phases is not rigid. Moving back and forth between different phases is always required. It depends on the outcome of each phase. The main phases are:
1) Business Understanding:
This phase focuses on understanding the project objectives and requirements from a business perspective, then converting this knowledge into a data mining problem definition and a preliminary plan designed to achieve the objectives.
2) Data Understanding:
It starts with an initial data collection, to get familiar with the data, to identify data quality problems, to discover first insights into the data or to detect interesting subsets to form hypotheses for hidden information.
3) Data Preparation:
In this stage , it collects all the different data sets and construct the varieties of the activities basing on the initial raw data.
4) Modelling:
In this phase, various modelling techniques are selected and applied and their parameters are calibrated to optimal values.
5) Evaluation:
In this stage the model is thoroughly evaluated and reviewed. The steps executed to construct the model to be certain it properly achieves the business objectives. At the end of this phase, a decision on the use of the data mining results should be reached.
6) Deployment:
The purpose of the model is to increase knowledge of the data, the knowledge gained will need to be organized and presented in a way that the customer can use it. The deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process across the enterprise.

This paper presents comparative study of different image fusion methods. In this we show the different approaches to avoid the occurrence of undesired artifacts such over-fusion, lower PSNR, higher MSE, and lack of information in the fused image (Image Entropy). And also On the basis of the Entropy, PSNR and MSE we also see that by hybrid image fusion algorithm we can get better fused image.

6. Visualizing Data Mining Model

The main objective of data visualization is the overall idea about the data mining model .In data mining most of the times we are retrieving the data from the repositories which are in the hidden form. This is the difficult task for a user. So this visualization of the data mining model helps us to provide utmost levels of understanding and trust. The data mining models are of two types:
Predictive and Descriptive.
The predictive model makes prediction about unknown data values by using the known values. Ex. Classification, Regression, Time series analysis, Prediction etc. The descriptive model identifies the patterns or relationships in data and explores the properties of the data
examined. Ex. Clustering, Summarization, Association rule, Sequence discovery etc. Many of the data mining applications are aimed to predict the future state of the data. Prediction is the process of analyzing the current and past states of the attribute and prediction of its future state. Classification is a technique of mapping the target data to

the predefined groups or classes, this is a supervise learning because the classes are predefined before the examination of the target data. The regression involves the learning of function that map data item to real valued prediction variable. In the time series analysis the value of an attribute is examined as it varies over time. In time series analysis is used for many statistical techniques which will analyze the time-series data such as auto regression methods etc.It is some times used in the two type of modelling (1) ARIMA (II)Long-memory time-series modelling .

The term clustering means analyzes the different data objects without consulting a known class levels. It is also referred to as unsupervised learning or segmentation. It is the partitioning or segmentation of the data in to groups or clusters. The clusters are defined by studying the behaviour of the data by the domain experts. The term segmentation is used in very specific context; it is a process of partitioning of database into disjoint grouping of similar tuples. Summarization is the technique of presenting the summarize information from the data. The association rule finds the association between the different attributes. Association rule mining is a two-step process: Finding all frequent item sets, Generating strong association rules from the frequent item sets. Sequence discovery is a process of finding the sequence patterns in data. This sequence can be used to understand the trend.

7. CONCLUSION

In this paper we briefly reviewed the various data mining applications. This review would be helpful to researchers to focus on the various issues of data mining. In future course, we will review the various classification algorithms and significance of evolutionary computing (genetic programming) approach in designing of efficient classification algorithms for data mining. Most of the previous studies on data mining applications in various fields use the variety of data types range from text to images and stores in variety of databases and data structures. The different methods of data mining are used to extract the patterns and thus the knowledge from this variety databases. Selection of data and methods for data mining is an important task in this process and needs the knowledge of the domain. Several attempts have been made to design and develop the generic data mining system but no system found completely generic. Thus, for every domain the domain expert's assistant is mandatory. The domain experts shall be guided by the system to effectively apply their knowledge for the use of data mining systems to generate required knowledge. The domain experts are required to determine the variety of data that should be collected in the specific problem domain, selection of specific data for data mining, cleaning and transformation of data, extracting patterns for knowledge generation and finally interpretation of the patterns and knowledge generation. Most of the domain specific data mining applications show accuracy above 90%. The generic data mining applications are having the limitations. From the study of various data mining applications it is observed that, no application called generic application is 100 % generic. The intelligent interfaces and intelligent agents up to some extent make the application generic but have limitations. The domain experts play important role in the different stages of data mining. The decisions at different stages are influenced by the factors like domain and data details, aim of the data mining, and the context parameters. The domain specific applications are aimed to extract specific knowledge. The domain experts by considering the user's requirements and other context parameters guide the system. The results yield from the domain specific applications are more accurate and useful. Therefore it is conclude that the domain specific applications are more specific for data mining. From above study it seems very difficult to design and develop a data mining system, which can work dynamically.

Data mining has importance regarding finding the patterns, forecasting, discovery of knowledge etc., in different business domains. Data mining techniques and algorithms such as classification, clustering etc., helps in finding the patterns to decide upon the future trends in businesses to grow. Data mining has wide application domain almost in every industry where the data is generated that's why data mining is considered one of the most important frontiers in database and information systems and one of the most promising interdisciplinary developments in Information Technology. In this paper we briefly reviewed the various data mining applications. This review would be helpful to researchers to focus on the various issues of data mining. In future course, we will review the various classification algorithms and significance of evolutionary computing (genetic programming) approach in designing of efficient classification algorithms for data mining. Most of the previous studies on data mining applications in various fields use the variety of data types range from text to images and stores in variety of databases and data structures. The different methods of data mining are used to extract the patterns and thus the knowledge from this variety databases. Selection of data and methods for data mining is an important task in this process and needs the knowledge of the domain. Several attempts have been made to design and develop the generic data mining system but no system found completely generic. Thus, for every domain the domain expert's assistant is mandatory. The domain experts shall be guided by the system to effectively apply their knowledge for the use of data mining systems to generate

required knowledge. The domain experts are required to determine the variety of data that should be collected in the specific problem domain, selection of specific data for data mining, cleaning and transformation of data, extracting patterns for knowledge generation and finally interpretation of the patterns and knowledge generation. Most of the domain specific data mining applications show accuracy above 90%. The generic data mining applications are having the limitations. From the study of various data mining applications it is observed that, no application called generic application is 100 % generic. The intelligent interfaces and intelligent agents up to some extent make the application generic but have limitations. The domain experts play important role in the different stages of data mining. The decisions at different stages are influenced by the factors like domain and data details, aim of the data mining, and the context parameters. The domain specific applications are aimed to extract specific knowledge. The domain experts by considering the user's requirements and other context parameters guide the system. The results yield from the domain specific applications are more accurate and useful. Therefore it is conclude that the domain specific applications are more specific for data mining. From above study it seems very difficult to design and develop a data mining system, which can work dynamically for any domain.

8. REFERENCES

[1] Introduction to Data Mining and Knowledge Discovery, Third Edition ISBN: 1-892095-02-5, Two Crows Corporation, 10500 Falls Road, Potomac, MD 20854 (U.S.A.), 1999.

[2] Larose, D. T., "Discovering Knowledge in Data: An Introduction to Data Mining", ISBN 0-471-66657-2, ohn Wiley & Sons, Inc, 2005.

[3] Dunham, M. H., Sridhar S., "Data Mining: Introductory and Advanced Topics", Pearson Education, New Delhi, ISBN: 81-7758-785-4, 1st Edition, 2006

[4] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. and Wirth, R... "CRISP-DM 1.0 : Step-by-step data mining guide, NCR Systems Engineering Copenhagen (USA and Denmark), DaimlerChrysler AG (Germany), SPSS Inc. (USA) and OHRA Verzekeringenen Bank Group B.V (The Netherlands), 2000".

[5] Fayyad, U., Piatetsky-Shapiro, G., and Smyth P., "From Data Mining to Knowledge Discovery in Databases," AI Magazine, American Association for Artificial Intelligence, 1996.