Data Mining Techniques: An Overview

Miss. Radhika K. Rathi, Dr. S. E. Yedey

***Abstract* - A potentially useful intellectual tool for researchers is the ability to make connections between seemingly unrelated facts, and as a consequence create inspired new ideas, approaches or hypotheses for their current work. This can be achieved through a process known as data mining.**

Data mining currently involves analyzing a large collection of often unrelated digital items in a systematic way and to discover previously unknown facts, which might take the form of relationships or patterns that are buried deep in an extensive collection. Data mining is a process which finds useful patterns from large amount of data. The paper discusses few of the data mining techniques, Issues in Data Mining, Tasks of Data Mining, Data mining background, Methodologies of Data Mining and Data Mining Models.

***Index Terms* - Classification, Clustering, Data Warehousing, Neural Network, Genetic Algorithm.**

I. INTRODUCTION

The past two decades has seen a dramatic increase in the amount of information or data being stored in electronic format. This accumulation of data has taken place at an explosive rate. It has been estimated that the amount of information in the world doubles every 20 months and the size and number of databases are increasing even faster. The increase in use of electronic data gathering devices such as point-of-sale or remote sensing devices has contributed to this explosion of available data. Figure1.1 from the Red Brick company illustrates the data explosion.

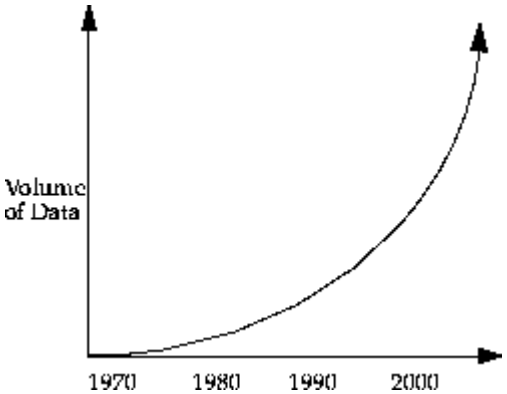

Fig1.1: Data Explosion Curve

Basically data mining is concerned with the analysis of data and the use of software techniques for finding patterns and regularities in sets of data. It is the computer which is responsible for finding the patterns by identifying the underlying rules and features in the data. The idea is that it is possible to strike gold in unexpected places as the data mining software extracts patterns not previously discernable or so obvious that no-one has noticed them before.

Data mining analysis tends to work from the data up and the best techniques are those developed with an orientation towards large volumes of data, making use of as much of the collected data as possible to arrive at reliable conclusions and decisions. The analysis process starts with a set of data, uses a methodology to develop an optimal representation of the structure of the data during which time knowledge is acquired. Once knowledge has been acquired this can be extended to larger sets of data working on the assumption that the larger data set has a structure similar to the sample data. Again this is analogous to a mining operation where large amounts of low grade materials are sifted through in order to find something of value.

Data mining is an interdisciplinary, integrated database, artificial intelligence, machine learning, statistics, etc. Many areas of theory and technology in current era are databases, artificial intelligence, data mining and statistics is a study of three strong large technology pillars. Data mining is a multi-step process, requires accessing and preparing data for a mining the data, data mining algorithm, analyzing results and taking appropriate action. The data, which is accessed can be stored in one or more operational databases. In data mining the data can be mined by passing various processes [2].

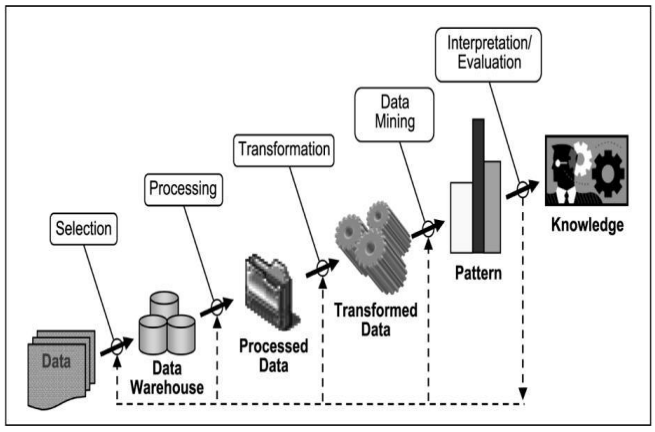

Fig.1.2: Steps in Data Mining process

Data mining consists of five major elements:

1. Extract, transform, and load transaction data onto the data warehouse system.
2. Store and manage the data in a multidimensional database system.
3. Provide data access to business analysts and information technology professionals.
4. Analyze the data by application software.
5. Present the data in a useful format, such as a graph or table [1].

***Manuscript received Dec, 2015*.**
***Radhika K. Rathi**, Department of Computer Science and Information Technology, HVPM COET, Amravati, India.*
***Dr. S. E. Yedey,** PG Department of Computer Science and Technology, HVPM COET, Amravati, India.*

II. TASKS OF DATA MINING

Data mining as a term used for the specific classes of six activities or tasks as follows:

1. Classification
2. Estimation
3. Prediction
4. Affinity grouping or association rules
5. Clustering

The first three tasks - classification, estimation and prediction rules are examples of directed data mining or supervised learning. In directed data mining, the goal is to use the available data to build a model that describes one or more particular attribute(s) of interest (target attributes or class attributes) in terms of the rest of the available attributes. The next three tasks – association rules and clustering are examples of undirected data mining i.e. no attribute is singled out as the target, the main goal is to establish some relationship among all attributes [2].

A. Classification

Classification consists of examining the features of a newly presented object and assigning to it a predefined class. The classification task is characterized by the well-defined classes, and a training set consisting of reclassified examples. The task is to build a model that can be applied to unclassified data in order to classify it. Examples of classification tasks include:

- Classification of credit applicants as low, medium or high risk
- Classification of mushrooms as edible or poisonous
- Determination of which home telephone lines are used for internet access

B. Estimation

Estimation deals with continuously valued outcomes. Given some input data, we use estimation to come up with a value for some unknown continuous variables such as income, height or credit card balance. Some examples of estimation tasks include:

- Estimating the number of children in a family from the input data of mothers education
- Estimating total household income of a family from the data of vehicles in the family
- Estimating the value of a piece of a real estate from the data on proximity of that land from a major business centre of the city.

C. Prediction

Any prediction can be thought of as classification or estimation. The difference is one of emphasis. When data mining is used to classify a phone line as primarily used for internet access or a credit card transaction as fraudulent, we do not expect to be able to go back later to see if the classification was correct. Our classification may be correct or incorrect, but the uncertainty is due to incomplete knowledge only: out in the real world, the relevant actions have already taken place. The phone is or is not used primarily to dial the local ISP. The credit card transaction is or is not fraudulent. With enough efforts, it is possible to check. Predictive tasks feel different because the records are classified according to some predicted future behavior or estimated future value. With prediction, the only way to check the accuracy of the classification is to wait and see. Examples of prediction tasks include:

- Predicting the size of the balance that will be transferred if a credit card prospect accepts a balance transfer offer.
- Predicting which customers will leave within next six months.
- Predicting which telephone subscribers will order a value added service such as three-way calling or voice mail. Any of the techniques used for classification and estimation can be adopted for use in prediction by using training examples where the value of the variable to be predicted is already known, along with historical data for those examples. The historical data is used to build a model that explains the current observed behavior. When this model is applied to current inputs, the result is a prediction of future behavior [3].

D. Association Rules

An association rule is a rule which implies certain association relationships among a set of objects (such as "occur together" or "one implies the other") in a database. Given a set of transactions, where each transaction is a set of literals (called items), an association rule is an expression of the form X Y , where X and Y are sets of items. The intuitive meaning of such a rule is that transactions of the database which contain X tend to contain Y. An example of an association rule is: "30% of farmers that grow wheat also grow pulses; 2% of all farmers grow both of these items". Here 30% is called the confidence of the rule, and 2% the support of the rule. The problem is to find all association rules that satisfy user-specified minimum support and minimum confidence constraints.

E. Clustering

Cluster analysis can be used as a standalone data mining tool to gain insight into the data distribution, or as a pre-processing step for other data mining algorithms operating on the detected clusters. Many clustering algorithms have been developed and are categorized from several aspects such as partitioning methods, hierarchical methods, density-based methods, and grid-based methods. Further data set can be numeric or categorical. Clustering is the task of segmenting a diverse group into a number of similar subgroups or clusters. What distinguishes clustering from classification is that clustering does not rely on predefined classes. In clustering, there are no predefined classes. The records are grouped together on the basis of self similarity. Clustering is often done as a prelude to some other form of data mining or modeling.

III. DATA MINING BACKGROUND

Many researchers have contributed to the application on data mining. In this section, we will give a brief overview on a few representative works.

A. Inductive learning

Induction is the inference of information from data and inductive learning is the model building process where the environment i.e. database is analyzed with a view to finding patterns. Similar objects are grouped in classes and rules formulated whereby it is possible to predict the class of unseen objects. This process of classification identifies classes such that each class has a unique pattern of values which forms the class description. The nature of the environment is dynamic hence the model must be adaptive i.e. should be able learn. Generally it is only possible to use a small number of properties to characterize objects so we make abstractions in that objects which satisfy the same subset of properties are mapped to the same internal representation.

Inductive learning where the system infers knowledge itself from observing its environment has two main strategies:

- Supervised learning - this is learning from examples where a teacher helps the system construct a model by defining classes and supplying examples of each class. The system has to find a description of each class i.e. the common properties in the examples. Once the description has been formulated the description and the class form a classification rule which can be used to predict the class of previously unseen objects. This is similar to discriminate analysis as in statistics.
- Unsupervised learning - this is learning from observation and discovery. The data mine system is supplied with objects but no classes are defined so it has to observe the examples and recognize patterns (i.e. class description) by itself. This system results in a set of class descriptions, one for each class discovered in the environment. Again this similar to cluster analysis as in statistics.

Induction is therefore the extraction of patterns. The quality of the model produced by inductive learning methods is such that the model could be used to predict the outcome of future situations in other words not only for states encountered but rather for unseen states that could occur. The problem is that most environments have different states, i.e. changes within, and it is not always possible to verify a model by checking it for all possible situations.

Given a set of examples the system can construct multiple models some of which will be simpler than others. The simpler models are more likely to be correct if we adhere to Ockham razor, which states that if there are multiple explanations for a particular phenomena it makes sense to choose the simplest because it is more likely to capture the nature of the phenomenon.

B. Statistics

Statistics has a solid theoretical foundation but the results from statistics can be overwhelming and difficult to interpret as they require user guidance as to where and how to analyse the data. Data mining however allows the expert's knowledge of the data and the advanced analysis techniques of the computer to work together.

Statistical analysis systems such as SAS and SPSS have been used by analysts to detect unusual patterns and explain patterns using statistical models such as linear models. Statistics have a role to play and data mining will not replace such analyses but rather they can act upon more directed analyses based on the results of data mining. For example statistical induction is something like the average rate of failure of machines.

C. Machine Learning

Machine learning is the automation of a learning process and learning is tantamount to the construction of rules based on observations of environmental states and transitions. This is a broad field which includes not only learning from examples, but also reinforcement learning, learning with teacher, etc. A learning algorithm takes the data set and its accompanying information as input and returns a statement e.g. a concept representing the results of learning as output. Machine learning examines previous examples and their outcomes and learns how to reproduce these and make generalizations about new cases.

Generally a machine learning system does not use single observations of its environment but an entire finite set called the training set at once. This set contains examples i.e. observations coded in some machine readable form. The training set is finite hence not all concepts can be learned exactly.

IV. METHODOLOGIES OF DATA MINING

A. Neural Network

Neural Network or an artificial neural network is a biological system that detects patterns and makes predictions. The greatest breakthroughs in neural network in recent years are in their application to real world problems like customer response prediction, fraud detection etc. Data mining techniques such as neural networks are able to model the relationships that exist in data collections and can therefore be used for increasing business intelligence across a variety of business applications [5]. This powerful predictive modeling technique creates very complex models that are really difficult to understand by even experts. Neural Networks are used in a variety of applications. It is shown in fig4.1. Artificial neural network have become a powerful tool in tasks like pattern recognition, decision problem or predication applications. It is one of the newest signals processing technology. ANN is an adaptive, non linear system that learns to perform a function from data and that adaptive phase is normally training phase where system parameter is change during operations. After the training is complete the parameter are fixed. If there are lots of data and problem is poorly understandable then using ANN model is accurate, the non linear characteristics of ANN provide it lots of flexibility to achieve input output map. Artificial Neural Networks, provide user the capabilities to select the topology, performance parameter, learning rule and stopping Criteria.

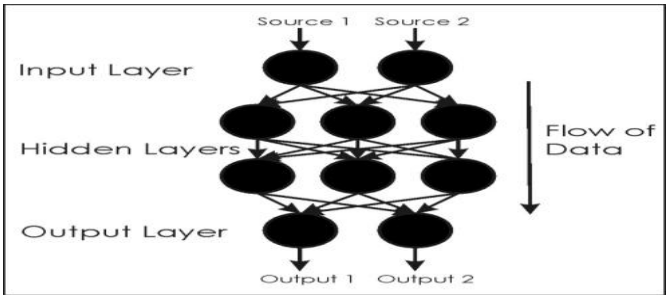

Fig4.1: Neural Network with hidden layers

B. Decision Trees

A decision tree is a flow chart like structure where each node denotes a test on an attribute value, each branch represents an outcome of the test and tree leaves represent classes or class distribution. A decision tree is a predictive model most often used for classification. Decision trees partition the input space into cells where each cell belongs to one class. The partitioning is represented as a sequence of tests. Each interior node in the decision tree tests the value of some input variable, and the branches from the node are labeled with the possible results of the test. The leaf nodes represent the cells and specify the class to return if that leaf node is reached. The classification of a specific input instance is thus performed by starting at the root node and, depending on the results of the tests, following the appropriate branches until a leaf node is reached [6]. Decision tree is represented in figure 4.2.as below:-

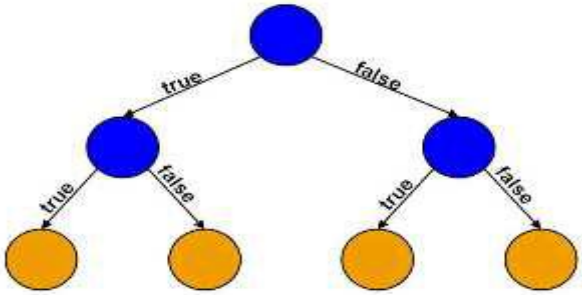

Fig 4.2: Decision tree

C. Genetic Algorithm

Genetic Algorithm attempt to incorporate ideas of natural evaluation. The general idea behind GAs is that we can build a better solution if we somehow combine the "good" parts of other solutions (schemata theory), just like nature does by combining the DNA of living beings [7]. Genetic Algorithm is basically used as a problem solving strategy in order to provide with an optimal solution. They are the best way to solve the problem for which little is known.

They will work well in any search space because they form a very general algorithm. The only thing to be known is what the particular situation is where the solution performs very well, and a genetic algorithm will generate a high quality solution. Genetic algorithms use the principles of selection and evolution to produce several solutions to a given problem. Genetic algorithms (GAs) [8] are based on a biological applications; it depends on theory of evolution. When GAs are used for problem solving, the solution has three distinct stages:

1) The solutions of the problem are encoded into representations that support the necessary variation and selection operations; these representations, are called chromosomes, are as simple as bit strings.
2) A fitness function judges which solutions are the "best" life forms, that is, most appropriate for the solution the particular problem. These individuals are favored in survival and reproduction, thus giving rise to generation.
3) Crossover and mutation produce a new generation of individuals by recombining features of their parents. Eventually a generation of individuals will be interpreted back to the original problem domain and the fit individual represents the solution.

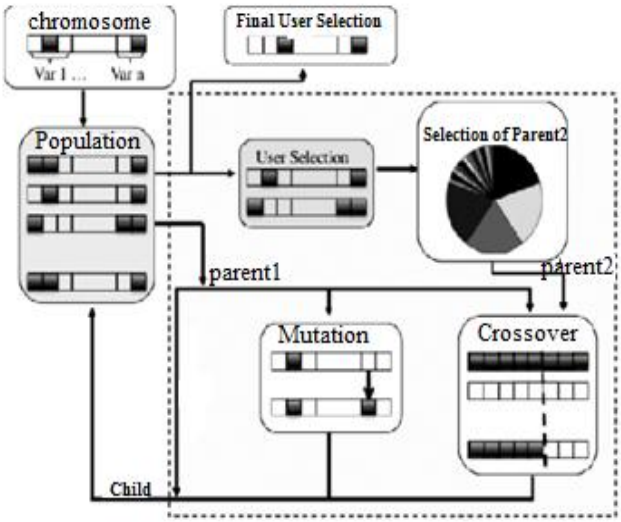

Fig 4.3: Structural view of Genetic Algorithm

D. Rule Extraction

The taxonomy of Rule extraction contains three main criteria for evaluation of algorithms: the scope of dependency on the black box and the format of the extract description. The first dimension concerns with the scope of use of an algorithm either regression or dimension focuses on the extraction algorithm on the underlying black-box: independent algorithms. The third criterion focuses on the obtained rules that might be worthwhile algorithms. Besides this taxonomy the evaluation criteria appears in almost all of these surveys rule; Scalability of the algorithm; consistency [9].

Generally a rule consists of two values. A left and a right hand consequent. An antecedent can have one or multiple conditions which must be true in order for the consequent to be true for a given accuracy whereas a consequent is just a single condition. Thus from a database antecedent, consequent, accuracy, and coverage are all targeted. Sometimes "interestingness" is also targeted used for ranking. The situation occurs when rules have high coverage and accuracy but deviate from standards. It is also essential to note that even though patterns are produced from rule induction system, they all not necessarily mean that a left hand side ("if "part) should cause the right hand side ("then") part to happen. Once rules are created and interestingness is checked they can be business where each rule performs a prediction keeping a consequent as the target and the accuracy of the rule as the accuracy of the prediction which gives an opportunity for the overall system to improve and

perform well. For data mining domain, the lack of explanation facilities seems to be a serious drawback as it produce opaque model, along with that accuracy is also required. It is becoming increasingly apparent that the absence of an explanation capability in ANN systems limits the realizations of the full potential of such systems, and it is this precise deficiency that the rule extraction experience from the field of expert systems has shown that an explanation capability is a vital function provided by symbolic.

V. DATA MINING MODELS

Data mining have identified two types of model or modes of operation which may be used to unearth information of interest to the user

A. Verification Model

The verification model takes an hypothesis from the user and tests the validity of it against the data. The emphasis is with the user who is responsible for formulating the hypothesis and issuing the query on the data to affirm or negate the hypothesis.

In a marketing division for example with a limited budget for a mailing campaign to launch a new product it is important to identify the section of the population most likely to buy the new product. The user formulates an hypothesis to identify potential customers and the characteristics they share. Historical data about customer purchase and demographic information can then be queried to reveal comparable purchases and the characteristics shared by those purchasers which in turn can be used to target a mailing campaign. The whole operation can be refined by `drilling down' so that the hypothesis reduces the `set' returned each time until the required limit is reached.

The problem with this model is the fact that no new information is created in the retrieval process but rather the queries will always return records to verify or negate the hypothesis. The search process here is iterative in that the output is reviewed, a new set of questions or hypothesis formulated to refine the search and the whole process repeated. The user is discovering the facts about the data using a variety of techniques such as queries, multidimensional analysis and visualization to guide the exploration of the data being inspected.

B. Discovery Model

The discovery model differs in its emphasis in that it is the system automatically discovering important information hidden in the data. The data is sifted in search of frequently occurring patterns, trends and generalizations about the data without intervention or guidance from the user. The discovery or data mining tools aim to reveal a large number of facts about the data in as short a time as possible.

An example of such a model is a bank database which is mined to discover the many groups of customers to target for a mailing campaign. The data is searched with no hypothesis in mind other than for the system to group the customers according to the common characteristics found.

VI. DATA WAREHOUSING

Data mining potential can be enhanced if the appropriate data has been collected and stored in a data warehouse. A data warehouse is a relational database management system (RDMS) designed specifically to meet the needs of transaction processing systems. It can be loosely defined as any centralized data repository which can be queried for business benefit but this will be more clearly defined later. Data warehousing is a new powerful technique making it possible to extract archived operational data and overcome inconsistencies between different legacy data formats. As well as integrating data throughout an enterprise, regardless of location, format, or communication requirements it is possible to incorporate additional or expert information. It is the logical link between what the managers see in their decision support EIS applications and the company's operational activities. In other words the data warehouse provides data that is already transformed and summarized, therefore making it an appropriate environment for more efficient DSS and EIS applications [9].

- *Characteristics of a data warehouse*

According to Bill Inman, author of Building the Data Warehouse and the guru who is widely considered to be the originator of the data warehousing concept, there are generally four characteristics that describe a data warehouse:

1. *Subject-Oriented:* data are organized according to subject instead of application e.g. an insurance company using a data warehouse would organize their data by customer, premium, and claim, instead of by different products (auto, life, etc.). The data organized by subject contain only the information necessary for decision support processing.
2. *Integrated:* When data resides in many separate applications in the operational environment, encoding of data is often inconsistent. For instance, in one application, gender might be coded as "m" and "f" in another by 0 and 1. When data are moved from the operational environment into the data warehouse, they assume a consistent coding convention e.g. gender data is transformed to "m" and "f".
3. *Time-Variant:* The data warehouse contains a place for storing data that are five to 10 years old, or older, to be used for comparisons, trends, and forecasting. These data are not updated.
4. *Non-Volatile:* Data are not updated or changed in any way once they enter the data warehouse, but are only loaded and accessed.

VII. DATA MINING PROBLEMS/ISSUES

Data mining systems rely on databases to supply the raw data for input and this raises problems in that databases tend be dynamic, incomplete, noisy, and large. Other problems arise as a result of the adequacy and relevance of the information stored.

A. Limited Information

A database is often designed for purposes different from data mining and sometimes the properties or attributes that would simplify the learning task are not present nor can they be requested from the real world. Inconclusive data causes problems because if some attributes essential to knowledge about the application domain are not present in the data it

may be impossible to discover significant knowledge about a given domain. For example cannot diagnose malaria from a patient database if that database does not contain the patients red blood cell count.

B. Noise and missing values

Databases are usually contaminated by errors so it cannot be assumed that the data they contain is entirely correct. Attributes which rely on subjective or measurement judgments can give rise to errors such that some examples may even be missing classified. Errors in either the values of attributes or class information are known as noise. Obviously where possible it is desirable to eliminate noise from the classification information as this affects the overall accuracy of the generated rules.

Missing data can be treated by discovery systems in a number of ways such as;

a) Simply disregard missing values
b) Omit the corresponding records
c) Infer missing values from known values
d) Treat missing data as a special value to be included additionally in the attribute domain or average over the missing values using Bayesian techniques.
e) Noisy data in the sense of being imprecise is characteristic of all data collection and typically fit a regular statistical distribution such as Gaussian while wrong values are data entry errors. Statistical methods can treat problems of noisy data, and separate different types of noise.

C. Uncertainty

Uncertainty refers to the severity of the error and the degree of noise in the data. Data precision is an important consideration in a discovery system.

D. Size, updates, and irrelevant fields

Databases tend to be large and dynamic in that their contents are ever-changing as information is added, modified or removed. The problem with this from the data mining perspective is how to ensure that the rules are up-to-date and consistent with the most current information. Also the learning system has to be time-sensitive as some data values vary over time and the discovery system is affected by the `timeliness' of the data.

Another issue is the relevance or irrelevance of the fields in the database to the current focus of discovery for example post codes are fundamental to any studies trying to establish a geographical connection to an item of interest such as the sales of a product [12].

VIII. POTENTIAL APPLICATIONS

Data mining has many and varied fields of application some of which are listed below.

A. Retail/Marketing

Identify buying patterns from customers; find associations among customer demographic characteristics predict response to mailing campaigns, Market basket analysis.

B. Banking

Detect patterns of fraudulent credit card use Identify `loyal' customers, Predict customers likely to change their credit card affiliation ,Determine credit card spending by customer groups, Find hidden correlations between different financial indicators, Identify stock trading rules from historical market data.

C. Insurance and Health Care

Claims analysis - i.e. which medical procedures are claimed together Predict which customers will buy new policies, Identify behavior patterns of risky customers, Identify fraudulent behavior .

D. Transportation

Determine the distribution schedules among outlets & Analyze loading patterns

E. Medicine

Characterize patient behavior to predict office visits. Identify successful medical therapies for different illnesses.

IX. DATA MINING – THE NEXT WAVES

Data mining is a promising area of engineering and it does have wide applicability. It can be applied in various domains. Data mining, as the confluence of multiple intertwined disciplines, including statistics, Machine learning, pattern recognition, database systems, information retrieval, World Wide Web, visualization, and many application domains, has made great progress in the past decade. Following are the working area and the constraint in data mining:-

A. Data Mining in Security and Privacy Preserving

Security and privacy are not very new concepts in data mining, but there is too much that can be done in this area with data mining.

B. Challenges in Mining Financial Data

There are many motivating factors for the study of this area.

Biggest is profit everyone wants profit may it be investor, speculator or operator in trading. He presents models of assets prices, and presents the modeling of relative changes of stock prices

C. Detecting Eco-System Disturbances

This is another promising area. It comprises of many areas such as remote sensing, earth-science, biosphere, oceans and predicts the ecosystem. There are also issues in mining the earth science like high dimensionality because long time series data are common in data mining.

Study of this area is important due to radical changes in ecosystem has led to floods, drought, ice-storms, hurricanes, tsunami and other disasters

D. Distributed Data Mining

Conventional data mining is thought to be as containing a large repository, and then mine knowledge. But there is an eminent need for mining knowledge from distributed resources. Typical algorithms which are available to us are based on assumption that the data is memory resident, which makes them unable to cope with the increasing complexity of distributed algorithms .Similar issues also rise while mining data in sensor network, and grid data mining. We need distribution classification algorithms. A technique called partition tree construction approach can be used for parallel

decision tree construction. We also need distributed algorithms for association analysis [10].

X. CONCLUSIONS AND FUTURE WORK

In this paper we put forward a Data mining involves extracting useful rules or interesting patterns from huge historical data. Many data mining tasks are available and each of them further has many techniques. Data mining is an interdisciplinary, artificial intelligence, integrated database, machine learning, statistics, etc. Data mining are a large number of incomplete, noisy, fuzzy, random applications of the data found in hidden, regularity which are not known by people in advance, but is potentially useful and ultimately understandable information and knowledge of non-trivial process In this paper we discusses some issues in Data Mining and activities used for Data mining task. At present data mining is a new and important area of research and ANN itself is a very suitable for solving the problems of data mining because its characteristics of good robustness, self-organizing adaptive, parallel processing, distributed storage and high degree of fault tolerance. The commercial, educational and scientific applications are increasingly dependent on these methodologies.

ACKNOWLEDGMENT

I express my sincere gratitude to Resp. Prof. P. L. Ramteke, Head of the Department of Computer Science & Information Technology & Resp. Dr. S. E. Yedey for providing their valuable guidance and necessary facilities needed for the successful completion of this seminar throughout.

Lastly, but not least, I thank all my friends and well wishers who were a constant source of inspiration.

REFERENCES

[1] Nikita Jain, Vishal Srivastava, Data Mining Techniques: A Survey Paper, International Journal of Research in Engineering and Technology , pissn: 2321-7308Volume: 02 Issue: 11 Nov-2013.

[2] Anand V. Saurkar, Vaibhav Bhujade, Priti Bhagat Amit Khaparde, A Review Paper on Various Data Mining Techniques, International Journal of Advanced Research in Computer Science and Software Engineering.

[3] Manish Verma, Mauly Srivastava, Neha Chack, Atul Kumar Diswar, Nidhi Gupta, A Comparative Study of Various Clustering Algorithms in Data Mining,, International Journal of Engineering Reserch and Applications (IJERA), Vol. 2, Issue 3, pp.1379-1384, 2012.

[4] Aastha Joshi, Rajneet Kaur, A Review: Comparative Study of Various Clustering Techniques in Data Mining, IJARCSSE, Vol. 3, 2013, 2277 128X .6

[5] R. Andrews, J. Diederich, A. B. Tickle," A survey and critique of techniques for extracting rules from trained artificial neural networks", Knowledge-Based Systems,vol.- 8,no.-6, pp.-378-389,1995.

[6] Lior Rokach and Oded Maimon,"Data Mining with Decision Trees: Theory and Applications(Series in Machine Perception and Artificial Intelligence)", ISBN: 981-2771-719, World Scientific Publishing Company, 2008.

[7] AnkitaAgarwal,"Secret Key Encryption algorithm using genetic algorithm", vol.-2, no.-4, ISSN: 2277 128X, IJARCSSE, pp. 57-61, April 2012.

[8] Li Lin, Longbing Cao, Jiaqi Wang, Chengqi Zhang, "The Applications of Genetic Algorithms in Stock Market Data Mining Optimisation", Proceedings of Fifth International Conference on Data Mining, Text Mining and their Business Applications,pp- 593-604,sept 2005.

[9] Han, J., Kamber, M., Data Mining Concepts and Techniques, Morgan Kaufmann Publisher, 2001

[10] Dharminder Kumar, Deepak Bhardwaj, Rise of Data Mining: Current and Future Application Areas, International Journal of Computer Science Issues, Vol. 8, Issue 5, No 1, September 2011

[11] Xingquan Zhu, Ian Davidson, "Knowledge Discovery and Data Mining: Challenges and Realities", ISBN 978- 1-59904-252, Hershey, New York, 2007.

[12] Oded Maimon, Lior Rokach, Data Mining AND Knowlwdge Discovery Handbook, Springer Science + Business Media.Inc, pp.321-352, 2005.

AUTHOR DETAIL:

Miss. Radhika K Rathi is currently pursuing Master in Computer Science and Information Technology at HVPM COET Amravati, Maharashtra (India).

Dr. S. E. Yedey is currently working as associate professor in PG Department of Computer Science & Technology, HVPM COET Amravati, Maharashtra (India).

