LIMITATIONS, PATTERNS AND INTEGRATION OF DATA WAREHOUSING, DATA MINING AND DATABASE TECHNOLOGIES-MULTIDIMENSIONAL APPROACHES

Tiruveedula Gopi Krishna Dr.Mohamed Abdeldaiem Abdelhadi M.Madusudana Subramanyam

Abstract—Data Warehousing and Data mining on large databases has been a major concern in research com-munity, due to the difficulty of analyzing huge volumes of data using only traditional OLAP tools. This sort of process implies a lot of computational power, memory and disk I/O, which can only be provided by distributed/parallel computers. We present a discussion of how database technology can be integrated to data mining techniques. This paper provides a comprehensive compilation of knowledge covering state-of-the-art developments and research, as well as current innovative activities in data warehousing and mining, focusing on the integration between the fields of data warehousing and data mining, with emphasis on the applicability to real world problems" In the last years, data In the last years, data warehousing systems have gained relevance to support decision making within organizations. All in all, we discuss the current scenario of multidimensional modeling by carrying out a survey of multidimensional design methods. We present the most relevant methods introduced in the literature and a detailed comparison showing the main features of each approach. We present a discussion of how database technology can be integrated to data mining techniques. Finally, we also point out several ad-vantages of addressing data consuming activities through a tight integration of a database server and data mining techniques.

Index Terms — Patterns,DataWarehousing, DataMining,Multidimensional frame work.,evaluations

———————————— ◆ ————————————

1 INTRODUCTION

The large amounts of data collected and stored might contain some information, which could be useful, but it is not obvious to recognize, nor trivial to obtain it. Huge amount of data collected in various processes, either manufacturing or business, (often as a side effect of computerization) should be thoroughly analyzed as they might contain some precious information for decision support. There is nothing new about analyzing data, but it is in the amount of data, where traditional methods are becoming inefficient. Modern database contain a huge amount of data. However raw data are hardly of any utility. To gain information and knowledge, data must be first analyzed and processed. The automated analysis is data in search of some information is called data mining. The growth of information resources along with the accelerating rate of technological change has produced huge amounts of information that often exceed the ability of managers and employees to assimilate and use it productively. Data must be categorized in some manner if it is to be accessed, re-used, organized, or synthesized to build a picture of the company's competitive environment solve a specific business problem. In recent, years, the need to extract knowledge automatically from very large databases has grown. In response, the closely related fields of knowledge discovery in databases (KDD) and data mining have developed processes and algorithms that attempt to intelligently extract interesting and useful information from vast amounts of raw data.

2 DATA MINING & EVOLUTION OF DATA MINING

Data mining involves the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets. These tools can include statistical models, mathematical algorithms, and machine learning methods (algorithms that improve their performance automatically through experience, such as neural network or decision trees). Consequently, data mining consist of more than collecting and managing data, it also includes analysis and prediction. Data mining can be performed on data represented in quantitative, textual, or multimedia forms. Data mining applications can use a variety of parameters to examine the data. They include associations (patterns where one event is connected to another event, such as purchasing a pen paper). Sequence or path analysis (patterns where one vent leads to another event, such as the birth of a child and purchasing diapers), classification (identification of new patterns, such as coincidences between duct tape purchase and plastic sheeting purchases) clustering (finding and visually documenting groups of previously unknown facts, such as geographic location and brand preferences), and forecasting (discovering patterns from which one can make reasonable predictions regarding future activities, such as the prediction that people who join an athletic club may take exercise classes). As an application, compared to other data analysis applications, such as structured queries (used in many commercial databases) or statistical analysis software, data mining represents a difference of Kind rather than degree. Many simpler analytical tools utilize a verification-based approach, where the use develops a hypothesis and then tests the data to prove or disprove the hypothesis. For example, a user might hypothesize that a customer, who buys a hammer, will also buy a box of nails. The effectiveness of this approach can be limited by the creativity of the user to develop various hypotheses, as well as the structure of the software being used. In contrast, data mining utilizes a discovery approach, in which algorithms can be used to examine several multidimensional data relationships simultaneously, identify those that are unique or frequently represented. For example, a hardware store may compare their customers' tool purchases with home ownership, type of automobile driven, age, occupation, income, and/or distance between residence and the store. As a result of its complex capabilities, two precursors are important for a successful data mining exercise; a clear formulation of the problem to be solved, and access to the relevant data. Reflecting this conceptualization of data mining, some observers consider data mining to be just one step in a larger process known as knowledge discovery in databases (KDD). Other steps in the KDD process, in progressive order, include data cleaning, data integration, data selection, data transformation, (data mining), pattern evaluation, and knowledge presentation. A number of advances in technology and business processes have contributed to a growing interest in data mining in both the public and private sectors. Some of these changes include the growth of computer networks, which can be used to connect databases; the development of enhanced search related techniques such as neural networks and advanced algorithms; the spread of the client. Server computing model, allowing users to access centralized data resources from the desktop; and an increased ability to combine data from disparate source into a single searchable source. "Data Mining is a non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns from data" (Srikant & Agarwal). That is a very brief definition that implies the purposes of doing mining and extracting new information from data:

_ it is "Valid" because it looks to be well grounded in logic patterns. In order to be valid the process can be automatic or semiautomatic and there are many tools that are used to make the used algorithms and the resulting patterns as valid as possible;

_ It is a "novel" because data mining a lot of research needs to be done yet;

_ It is "potentially useful" because the results can be used in the decision making process of any organization, such as health, education, marketing etc.

_ "Understandable" patyter4ns because the results should be capable of being understood or Interpreted by users from different backgrounds and not only for researchers.

_ "Patterns" from previous data because a perceptual structure has been created as a model that can be applied to new data.

_ "Data" refers to the digitalized information in databases first and data warehouses later that can be accessed by data mining tools. Data Mining is an information extraction activity whose goal is to discover hidden facts contained in databases. Using a combination of machine learning, statistical analysis, modeling techniques and database, technology, data mining finds patterns and subtle relationships in data and infers rules that al-low the prediction of future results. The main points of this definition are the term "facts" because data mining works with "real" data and "hidden facts" because data mining shows the behaviour and performance that is not easily discovered. Commercial databases are growing at unprecedented rates. In the evolution from business data to business information, each new step has built upon the previous one. According to Kurt Thearling, these are the steps in the evolution of data mining:

3 Data Mining: Verification vs. Discovery

Decision Support System (DSS), executive information systems, and query / report writing tools are used to produce reports about data, usually aggregating it through any number of dimensions. Another use of these tools is to detect trends and patter4ns in customer data that will help answer some questions about the business. When used in this mode, a query is created to access the records relevant to the questions (s) being formulated. After the data is retrieved, it is examined to detect the existence of patterns or other useful information that can be used in answering the original question(s). We call this the verification mode. In this mode, the user of a DSS generates a hypothesis about the data, issues a query against the data and examines the results of the query looking for affirmation or negation of the hypothesis. In the first case, the process ends; in the latter case, a new query is reformulated and the process iterates until the resulting data either verifies the hypothesis is not valid for this data. Consider the following example. A sales executive has a limited to do a mailing campaign for a new product. In order to optimize the use of this money, the marketing executive wants to identify the largest set of the people that are the most likely candidates to buy the new product and which can be read within the budget limitation. To identify these customers and to verify that the customer set has been adequately narrowed to match the available promotional budgets, the executives make a hypothesis about the potential customer set. Issuing a query against the databases that contain historical data about customer purchases and demographic information respectively, the set of customers that have made significant purchases of competitive products can be obtained. Furthermore, to limit the number of customers found to a reasonable number, the executive requests to only get information about those customers that are characterized by having ages between 30 and 45 years, being heads of household with combined incomes between $25, 000 and $50, 000 and living in some specific zip code regions. If the result of this query returns a number of customers that match the available budget for mailing promotions, the process ends. However, if either significantly more (or less) customers are found than the number that can be reached with the given budget, anew query limiting (or expanding) the set of customer addresses requested must be issued. In the above example, the hypothesis used in formulating the queries where quite explicit (e.g. incomes between certain amounts). Even when the hypothesis are implicit, the process of findings useful trends or patterns using queries can be described by the above behaviour as shown in the following example involving a query drill down process. After a report about company sales shows that the last quarter sales were significantly lower than expected, the financial, officer of the company wants to discover what caused this situation. A query is first issued to return the sales figures, by region, for the last quarters. The result of this query show that all the sales are up, expect for one particular region. The financial officer begins to suspect that the problem may have occurred in some localized store. To better understand the nature of the problem, another query is issued that will return sales results for all the cities in the offending region. A result showing one city significantly lower than the rest reinforces the officers' suspicion; a result showing that sales were uniformly lower among all cities in this region requires that the initial guess about what caused the problem (i.e. the implicit hypothesis) be modified. New queries continuing to drill down looking for the results by store within an offending city follow the previous query; totally new queries need to be devised if the results of the last query contradict the implicit hypothesis. Queries such as those used in the previous two examples, always return records that satisfy the query predicated, thus little, new information is created in this retrieval process: either the hypothesis is verified or it is negated. The process of information finding is done by the user by successive iterations upon examining the results of query after query and linking the verified and refined hypothesis. This is the essence of a verification model. Many times, whole performing a query, a request is made to compute the functions related to the records being inspected during the query (e.g., count the number of records, find the average of a given field of the records, etc.) all these operations result in additional information being returned together with the query. For the purposes of this discussion, these derived facts are not considered. Notice that, from the user perspective, he / she is discovering facts about the data. These of queries to extract facts from databases are a common practice. There are other tools that like, query generators, are used in a mode that follows the verification model described above. Examples of these other tools are multidimensional analysis tools and visualization tools. Multidimensional tools make it easier for the user to formulate drill down queries such as those shown in the last example. Visualization tools are used, as their name implies, to present data in a visual manner and to allow the user to easily interact with the data in search of hidden patterns. The user of a visualization tool takes age of the human's visual perception capabilities to discern patterns. The three types of tools discussed above, queries multidimensional analysis, and visualization, all have in common that the user is essentially "guiding" the exploration of the data being inspected. Data mining uses a different model for the creation of information about data. We call this the discovery model. In the next section we will describe methodologies that can sift through the data in search of frequently occurring patterns, can detect trends, produce generalizations about the data etc. these tools can discover these types of information with little (or no) guidance from the user./ the discovery of these facts is not a consequence of a hap-hazard event. Quite to the contrary, a well-designed data mining tool is one that is done in such a way as to yield as a large number of useful facts about the data as possible ion the shortest amount of time. Comparing the process of finding information in a collection of data to that of mining's diamonds in a diamond mine, we can say that "verification" is like drilling individual holes in a lode with the expectation of finding diamonds. Finding all (or many) diamonds in this way can be very inefficient. "Discovery", on the other hand, is similar to scooping out all the material in the lode and dumping it on plain fields so that all the glittering stones are thrown up into the open. Diamonds are then separated from the quartz by further inspection. In data mining, large amounts of data are inspected; facts are discovered and brought to the attention of the person doing the mining. Unlike diamonds, which are easily distinguishable from quartz, business judgment must be used to separate the useful facts from those which are not. Because this last step does not involve sifting through the raw data, Data Mining is a more efficient mode of finding useful facts about data.

4 Data Mining Technology

Extracting knowledge hidden in large volumes of raw data. Researchers identify two fundamental goals of data mining: prediction and description.

Competitive advantage requires abilities. Abilities are built through knowledge. Knowledge comes from data. The process of extracting knowledge from data is called Data Mining. Typical tasks addressed by data mining include:

• Rate customers by their propensity to respond to an offer

• Identify cross-sell opportunities

• Detect fraud and abuse in insurance and finance

• Estimate probability of an illness re-occurrence or hospital re-admission

• Isolate root causes of an outcome in clinical studies

• Determine optimal sets of parameters for a production line operation

• Predict peak load of a network

Without proper analytical tools, discovering useful knowledge hidden in huge volumes of raw data represents a formidable task. The exponential growth in data, diverse nature of data and analysis objectives, the complexity of analyzing mixed structured data and text are among the factors that turn knowledge discovery into a real challenge. Data Mining provides tools for automated learning from historical data and developing models to predict outcomes of future situations. The best data mining software tools provide a variety of machine learning algorithms for modeling, such as Regression, Neural Network, Decision Tree, Bayesian Network, CHAID, Support Vector Machine, and random Forest to name a few. Yet, data mining requires far more than just machine learning. Data mining additionally involves data pre-processing and results delivery. Data pre-processing includes loading and integrating data from various data sources, normalizing and cleansing data, and carrying out exploratory data analysis. Results delivery includes model application in production environment and generating reports summarizing the results of the analysis in a simple form for business users. Megaputer's flagship data mining tool Poly Analysis supports all steps of data preprocessing and modeling and results delivering. Poly analyst enables you to solve tasks of predicting, classification, clustering, affinity grouping, link analysis, multi-dimensional analysis, and interactive graphical reporting.

5 Advantages of Data Mining Marking/ Retailing

Data mining can aid direct marketers by providing them with useful and accurate trends about their customers purchasing behaviour. Based on these trends, marketers can direct their marketing attentions to their customers with more precision. For example, marketers of a software company may advertise about their new software to consumers who have a lot of software purchasing history. In addition, data mining may also help marketing in predicting which products their customers may be interested in buying. Through this prediction, marketers can surprise their customers and make the customer's shopping experience becomes a pleasant one. Retail stores can also benefit from data mining in similar ways. For example, through the trends provide by data mining, the store managers can arrange shelves, stock certain items, or provide a certain discount that will attract their customers.

7 Banking / Crediting

Data mining can assist financial institutions in areas such as credit reporting and loan information. For example, by examining previous customers with similar attributes, a bank can estimated the level of risk associated with each given loan. In addition, data mining can also assist credit card issuers in detecting potentially fraudulent credit card transaction. Although the data mining technique is not a 100% accurate in its prediction about fraudulent charges, it does help the credit card issuers reduce their losses.

8 Law Enforcement

Data mining can aid law enforces in identifying criminal suspects as well as apprehending these criminals by examining trends in location, crime type, habit and other patterns of behaviours.

9 Researchers

Data mining can assist researchers by speeding up their data analyzing process; thus, allowing those more time to work on other projects.

10 Disadvantages of Data Mining Privacy Issues

Personal privacy has always been a major concern in this country. In recent years, with the widespread use of Internet, the concerns about privacy have increased tremendously. Because of the privacy issues, some people do not shop on Internet. They are afraid that somebody and then use that information in an unethical way; thus causing they harm. Although it is against the law to sell or trade personal information between different organizations, selling personal information have occurred. For example, according to Washing Post, in 1998, CVS has sold their patient's prescription purchases to a different company. In addition, American Express also sold their customers' credit care purchases to another company. What CVS and American Express did clearly violate privacy law because they were selling personal information without the consent of their customers? The selling of personal information may also bring harm to these customers because you do not know what the other companies are planning to do with the personal information that they have purchased.

11 Security Issues

Although companies have a lot of personal information about us available online, they do not have sufficient security systems in place to protect that information. For example, recently the Ford Motor credit company had to inform 13,000 of the consumers that their personal information including Social Security number, address, account number and payment history were accessed by hackers who broke into a database belonging to the Experian credit reporting agency. This incidence illustrated that companies are willing to disclose and share your personal information, but they are not taking care of the information properly. With so much personal information available, identity theft could become a real problem.

11 Mis Use of Information/ Inaccurate Information

Tends obtain through data mining intended to be used for marketing purpose or for some other ethical purposes, may be used. Unethical businesses or people may use the information obtained through data mining to take advantage of vulnerable people or discriminated against a certain group of people. In addition, data mining technique is not a 100 percent accurate; thus mistakes do happen which can have serious consequences.

12 Data Mining Processes

Data mining can do basically six tasks. The first three are all examples of directed data mining, where the goal is to use the variable data to build a model that describes one particular variable of interests in terms of the rest of the available data. For example, analyzing bankruptcy, the target variable is a binary variable that describes if a client was declared on bankruptcy or not. In directed data mining, we try to find patterns that will make that variable have that value: 0 or 1. The next three tasks are examples of undirected data mining where no variable is singled out as a target and the goal is to establish some relationship among all the variables. These are the types of information can be obtained by data mining, summarizing from two different sources: Turban & Aronson and Berry & Linoff:

• **Classification:** consists of examining the features of a newly presented object and assigning to it a predefined class or group. The task is to build a model can be applied to unclassified data in order to classify is using the defined characteristics of a certain group (e.g.,classifying credit applicants as lo9w, medium of high risk).

• **Estimation:** Given some input data, we use estimation to come up with a value for some unknown continuous variable such as income, height or credit card balance (e.g. a bank trying to decide to whom they should offer a home equity loan based on the probability that the person will respond positively to an offer).

• **Prediction:** records are classified according to some predicted future behaviour or estimated future values based on patterns within large sets of data (e.g., demand forecasting or predicting which customers will leave within the next six months).

• **Association:** identifies relationships between events that occur at one time, determines which things go together (e.g., the contents of as shopping basket: fruits with snacks).

• **Clustering:** identifies groups of items that share a particular characteristic segmenting a diverse a group into a number of more similar subgroups or clusters. Clustering differs from classification in that it does not rely on predefined classes or characteristics for each group. (e.g., as a first step in a market segmentation effort, we can divide the customer base into cluster of people with similar buying habits, and ten ask what kind of promotion works best for each cluster or group).

• **Description and Visualization:** the purpose is to describe what is going on in a complicated database in a way that increases our understanding of the people, products or processes that produced the data in the first place. A good description suggests where to start looking for an explanation. (e.g., repeat visit to a super-market). Data mining is concluded against data accumulated in OLTP repositories, data warehouses, data marts and archived data. The steps for data mining follow the following pattern**:**

• Data extraction
• Data cleansing
• Modeling data
• Applying data mining algorithm
• Pattern discovery
• Data visualization

Data extraction and data cleansing can be eased with good data lifecycle management policies. Very often a data warehousing project will ensure that data extraction and meta-data standards are predefined in an organization. Data models for operational and archived data are different from data mining models. Data stored referentially in operational systems are designed for transactional sped. In data mining a unified table view is created where data of interest is stored. Most data mining vendors offer the ability to extract data from repositories and transfer to the data mining database. Not all of the data found in the data mining table view will have relevance. Other data may hold hidden patterns that can be discovered after relevancy is captured, often with external data sources.

13 Data Mining Techniques

The most commonly used techniques in data mining are (By Thearling):

_ **Artificial neural network:** Non-linear predictive models that learn through training and resemble biological neural networks in structure.

_ **Decisions Trees:** Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include classification and Regression Trees (CART) and Chi Square Automatic Interaction
Detection (CHAID).

_ **Genetic algorithms:** Optimization techniques that use processes such as genetic combination, and natural selection in a design based on the concepts of evolution.

_ **Nearest neighbour method:** A technique that classifies each record in a dataset based on a combination of the classes of the K record (s) most similar to it in a historical dataset. Sometimes called the K-nearest neighbor technique.

_ **Rule induction:** The extraction of useful if-then rules from data based on statistical significance.

14 Data mining Tools

Data mining software is based in mathematical algorithms and statistics. Developers have been working in data mining software tools to make them more users friendly and the different products available on the market have advantages and disadvantages basically related to their interface and available techniques. Today, the market offers a variety of products. The most preferred tools are:

• Clementine
• Megaputer
• SAS EM (Enterprise Miner)
• Gain Smarts and
• Easy Miner
• SGI's Mine Set
• Oracle's Darwin

15 Conducting a Data Mining Project

The following are the necessary steps for conducting a data mining project:

_ Understand the purpose of the project together with managers of the organization. Define the objectives and how the results will be applied in the organization.

_ Create the dataset that is going to be used. This may include creating a data warehouse and selecting the databases that probably need to be joined.

_ Define the tools that are going to be used: Hardware (e.g. powerful computers with a high speed

processors), Software (e.g. Database management software and data mining tool like
Clementine, SAS or SPSS) and most important, people who are going to work in the project.

_ Analyze the dataset together with the managers. It is important to define each variable. Sometimes the names of the variables do not have the real meaning or they can have different interpretations.

_ Check if data is ready to apply data mining methods. There means for example, to clean the data, and input missing data. Create dummy variables if necessary.

_ Reduce the size of the database. For example, summarize some variables that can have same meaning. Reduce the format for binary variables; they can be reduced to one bit instead of one byte. That reduces the size of the file, so it can be easier to manipulate.

_ Identify the relevance of variables. For instance, Wal-mart may require primarily transaction
variables instead frequency of transactions.

_ Take sample data and identify the best rules for decomposition of data. For example. To define what percent of available data is going to be used for learning, validation and testing data.

_ Use data mining software tools and data mining techniques to search for the models that
have better performance.

_ Test the result in a different set of data.

_ Develop an interpretation of the results, so they can be understandable not only by the analyst's and better understandable by the managers.

_ Use the results in the management decision making process. For example, choosing a better packing strategy after mining **CRM** data.

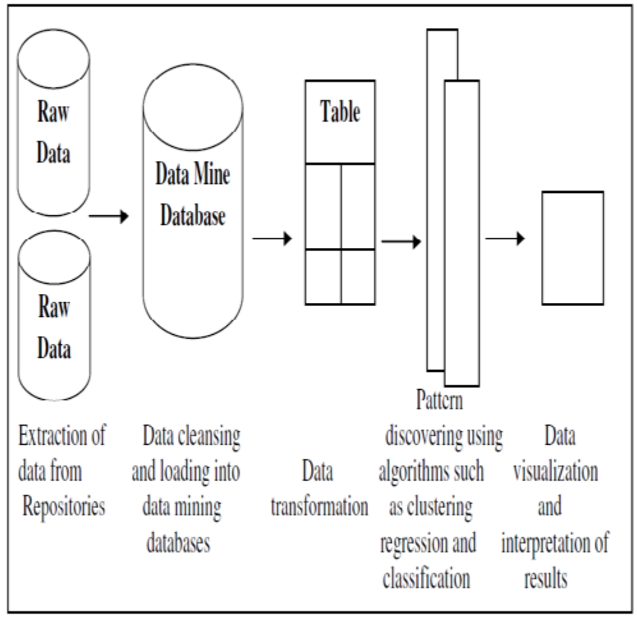

Fig. The Steps in Data Mining

16 PATTERN

Extraction of Data cleansing discovering using Data. Data from and loading into Data algorithms such visualization. Repositories data mining transformation as clustering and databases regression and interpretation of classification results.

Raw Data

Data Mine

Database Table

Data Mining Issues

17 PRIVACY

Data mining is both a powerful and profitable tool, but it poses challenges to the protection of individual privacy. Many critics wonder whether companies should be allowed to collect such detailed information about individuals. Consumer groups, privacy activists and person associated with the Federal Trade Commission are concerned about the privacy issues that arise with the vast amount of information that companies acquire when data mining. In order to protect privacy of customers, some measures were taken in the Prepared Statement of The Federal Trade Commission.

18 Database Requirements

In order to apply data mining techniques they require to be fully integrated with a data warehouse. The data warehouse grows and the organizations can continually mine the best practices and apply them to future decisions. However, if the data warehouse is not ready to mine, preparing data could take big percentage of the available resources for the project.

19 Result Interpretation

Sometimes the results that are obtained after mining data can be difficult to interpret, data mining has a variety of tools to make it easier to interpret and explain, but an intelligent human in still required to **(a)** structure be data in the first place; **(b)** interpret the data to understand the identified pattern; and **(c)** make a decision based on the knowledge.

20 Data Mining Strategic Decision Support

Data mining's main purpose is knowledge discovery leading to decision support. Data mining tools find patterns in data and may even infer rules from them. These patterns and rules can be used to guide decision-making and forecast the effect of these decisions. It is a proved that the data mining can speed analysis by focusing attention on the most important variables. Some strategic applications of data mining include (By London):

• Identifying individuals or organization most likely to respond to a direct mailing.

• Determining which products or services are commonly purchased together.

• Predicting which customers are likely to switch to competitors.

• Identifying which customers are likely to be fraudulent.

• Identifying common characteristics of customers who purchase the same product.

• Predicting what each visitor to a web site is most interesting in seeing.

Because competitor plans are never fully known ahead of time, it is essential to leverage all available information about customer's reactions to potential offers. Sometimes the information comes from the specific markets where more aggressive marketing initiatives are anticipated; sometimes the information comes from sources external to that market. In either case, effective knowledge management (KM) requires seeking diverse data about customer and competitor activities and capitalizing these data. Creating new knowledge for competitive situation requires openness to an enlarged array of data sources and the ability to capitalize on developments in data modelling and mining. Interpretation and analysis that might otherwise lead only to directional guidance can result in more specific decision parameters. Data mining can also be used to locate individual customers with specific interests or determine the interests of a specific group of customers.

21 Limitations of Data Mining

While data mining products can be very powerful tools, they are not self sufficient applications. To be successful, data mining requires skilled technical and analytical specialists who can structure the analysis and interpret the output that is created. Consequently, the limitations of data mining are primarily data or personnel related, rather than technology related. Although data mining can help reveal patterns an relationships, it does not tell the user the value or significance of these patterns. These types of determinations must be made by the user. Similarly, the validity of the patterns discovered is dependent on how they compare to "***real world***" circumstances. For example to assess the validity of a data mining application designed to identify potential terrorist suspects in a large pool of individuals, the user may test the model using data that includes information about known terrorists. However, while possibly reaffirming a particular profile, it does not necessarily mean that the application will identify a suspect whose behaviour significantly deviates from the original model. Another limitation of data mining is that while it can identify connections between behaviours and/or variables, it does not necessarily identify a causal relationship. For example, an application many identify that a pattern of behaviour, such as the propensity to purchase airline tickets just shortly before the flight is scheduled to depart is related to characteristics such as income, level of education, and Internet use. However, that does not necessarily indicate that the ticket purchasing behaviour is caused by one or more of these variables. In fact, the individual's behaviour could be affected by some additional variable (s) such as occupation (the need to make trips on short or a hobby) taking advantage of last minute discounts to visit new destinations).

Acknowledgment

I would like to thank to all who helped me out to publish my paper.

Conclusions

Data Warehousing, Data mining and its application on large databases have been extensively studied due to the increasing difficulty of analyzing large volumes of data using only OLAP tools. This difficulty pointed out the need of an automated process to discover interesting and hidden patterns in real-world data sets. The ability to handle large amounts of information has been a major concern in many recent data mining applications. Since we still lack a standard multidimensional terminology and terms used among methods to describe the multidimensional concepts may vary. For each method we captured its main features that were mapped onto different criteria. If a method introduced a new criterion, the rest of works were analyzed to know their assumptions with regard to this criterion. All in all, we have provided a comprehensive framework to better understand the current state of the area as well as its evolution.

References

[1] Agrawal, R., Ghosh, S., Imelinski, T., Iyer, B., & Swami, A., An Interval classifier for database mining applications, Proc. of VLDB Conference, Vancouver, Canada, pp. 560-573, 1992.

[2] Agrawal, R., Imielinski, T., & Swami, A., Mining association rules between sets of items in large databases, Proc. of Int. Conf. ACM SIGMOD, Washington D. C. pp. 207-216, 1993.

[3] Agrawal, R., Gehrke, J., Gunopulos, & D., Raghavan, P., Auto-matic Subspace Clustering of High Dimensional Data for Data Mining Applications, Proc. of the ACM SIGMOD Int. Conf. on Management of Data, Seattle, Washington, 1998.

[4] Agrawal, R., Metha, M., Shafer, J., & Srikant, R., The Quest Data Mining System, Proc. of the 2nd Int. Conf. on Knowledge Discovery in Databases and Data Mining, Portland, Oregon, 1996.

[5] Agrawal, R., & Shim, K., Developing Tightly-Coupled Data Mining Applications on a Relational Database System, Proc. of 2nd Int. Conf. on Knowledge Discovery in Databases and Data Mining, Portland, Oregon, 1996.

[6] Agrawal, R., & Srikant, R., Fast Algorithms for mining association rules, Proc. of the 20th VLDB Int. Conf., Santiago, Chile,1994.

[7] Bigus, J. P., Data Mining with NeuralNetworks, McGraw-Hill, 1996.

[8] Breiman, L., Friedman, J., Olshen, R., & Stone, C., Classification and Regression Trees, Wadsworth International Group, 1984.

[9] Chen, M. S., Han, J., & Yu, P. S., Data Mining: An Overview from Database Perspective, IEEE Trans. on Knowledge and Data Engineering, Vol. 8, No. 6, pp. 866-883, 1996.

Author*[1]*: Tiruveedula Gopi Krishna, Lecturer, Department of Computer Science, Faculty of Arts and Science, Hoon, AlJufra, Sirt University, Libya. Currently pursuing Ph.D in DataMining. Research Scholar.
Email:gktiruveedula@gmail.com
Author*[2]*: Dr.Mohammed Abdeldaiem Abdelhadi, Hod, Department of Computer Science,Faculty of Arts & Science, Hoon, Al-Jufra,Sirt University,Libya.
Email:mabdeldaiem2009@gmail.com
Author*[3]*: M.Madusudana Subramanyam,Ex.Lecturer, Dhammam Universiy,SaudiArabia,Research Scholor.
Email: mmsnaidu@yahoo.com.