Industrial Science ISSN 2347-5420
Volume-1 | Issue-10 | April-2015 Available online at www.lsrj.in

APPLICATIONS OF DATA MINING TECHNIQUES

C. Senthil Kumaran
Assistant Professor , Department of Computer Applications ,Christ College of Engineering & Technology Puducherry, India.

Short Profile

C. Senthil Kumaran is working as an Assistant Professor at Department of Computer Applications in Christ College of Engineering & Technology Puducherry, India.

Co- Author Details :

A. Jamaludeen2 and Diviyamaran3
[2]Senior Assistant Professor, Department of Computer Applications ,Christ College of Engineering & Technology , Puducherry, India.
[3]Department of Computer Applications ,Christ College of Engineering & Technology ,Puducherry, India.

ABSTRACT:
Today's technology has become an integral part of the business processes, the process of transfer of information has become more complicated. Realistic useof Database systems and Data Warehousing can contribute alot to decision support systems in all type of industries.Data Mining is the process of extracting information from large data sets through the use of algorithms and techniques drawn from the field of Statistics, artificial intelligence, information theory, Machine Learning and Data Base Management Systems.A deep understanding of the knowledge hidden information is vital to a firm's competitive position and organizational decision-making. Data mining place a vitalrole in all sectors.Finally, this paper is to discussed and concludedthat the applications of data mining techniques are adapted to improve all the sectorswith excellent results.

KEYWORDS
database,data mining,information, technology.

I.INTRODUCTION

It is well known that in Information Technology(IT) driven society, knowledge is one of the mostsignificant assets of any organization. The role of IT inall industries is well established. KnowledgeManagement in all industries offers many challenges increation, dissemination and preservation of knowledge using advanced technologies.It is actually the process of finding the hidden information/pattern of the repositories. Data mining often can improve existing models by finding additional, important variables, identifying interaction terms and detecting nonlinearrelationships.

The concepts of data mining has been evolved in modern information society as a technological solution to enhance in knowledge discovery in database.The data mining techniques developed recently, several major kinds of data mining methods, including generalization, characterization, classification, clustering, association, evolution, pattern matching, data visualization and meta-rule guided mining. Data mining is an essential step of knowledge discovery. Knowledge discovery process consists of an iterative sequence of data cleaning, data integration, data selection,data transformation, data mining pattern recognition and knowledge presentation. Once all these processes are over, we would be able to use this information in many applications such as Fraud Detection, Market Analysis, Production Control, Science Exploration, etc... Data mining tools predict behaviours and future trends, allowing businesses to make proactive, knowledge-driven decisions. Data mining derives its name from the similarities between searching for valuable information in a large database and mining a mountain for a vein of valuable ore. Both processes require either sifting through an immense amount of material, or intelligently probing it to find where the value resides.Data mining is widely used in diverse areas.A data mining process must be reliable and it must be repeatable by business people with little or no knowledge of data mining background. Data mining is a process that can be applied to any type of data ranging from weather forecasting, electric load prediction, product design, etc.

Data mining is not an easy task, as the algorithms used can get very complex and data is not always available at one place. It needs to be integrated from various heterogeneous data sources. Data mining is also used in the fields of credit card services and telecommunication to detect frauds.In fraud telephone calls, it helps to find the destination of the call, duration of the call, time of the day or week, etc. It also analyses the patterns that deviate from expected norms. Data mining is a relatively young field in computing, which broadly aims to provide tools and techniques to assist in the integration of disparate data sets and in the discovery of hidden patterns and relationships in these data sets.User interface is the module of data mining system that helps the communication between users and the data mining system. Data mining engine is very essential to the data mining system.

Data mining systems may integrate techniques from the following: Spatial Data Analysis, Information Retrieval, Pattern Recognition, Image Analysis, Signal Processing, Computer Graphics, Web Technology, Business and Bio-informatics.There is a large variety of data mining systems available.There are many data mining system products and domain specific data mining applications. The new data mining systems and applications are being added to the previous systems. Also, efforts are being made to standardize data mining languages.The actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown interesting patterns such as groups of data records, unusual records and dependencies. This usually involves using database techniques such as spatial indices. These patterns can then be seen as a kind of summary of the input data, and may be used in further analysis or, for example, in machine learning and predictive analytics.

Here the paper is to set the applications of data mining techniques. With the help of literature review, the paper is very clearly explained about the applications of data mining techniques. Then it clearly denoted about the knowledge discovery process of data mining. The paper which is identified

the various industriesusing data mining. Finally this paper concludes that retrieval of information from the database using KDD process.

2. LITERATURE REVIEW

"Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner".(Fayadd, U., Piatesky -Shapiro, G., and Smyth, P, From Data Mining To Knowledge Discovery in Databases", The MIT Press, Fayap, 1996).(Mitchell, 1997) proposed the field of data mining addresses the question of how best touse the historical data to discover general regularities and improve the process of making decisions.

(Shakil Ahmed, FransCoenen, Paul Leng, Knowledge Information System: 2006, Tree based partitioningof data for association rule mining)propose the global optimization approach to clustering and demonstrate how the supervised data classification problem can be solved via clustering. The objective function in this problem is both nonsmooth and nonconvex and has a large number of local minimizers. Due to a large number of variables and the complexity of the objective function, general purpose global optimization techniques, as a rule fail to solve such problem.(Cosper Nate, 2003) describes this trend, which is becoming increasingly accepted, has numerous implications for the drug discovery technologies companies. Most significantly, slower revenue growth will necessitate decreased expenditures throughout the organization.Data mining is an essential step of knowledge discovery. In recent years it has attracted great deal of interest inInformation industry.

(Glymour, C., D. Madigan, D. Pregidon and P.Smyth, 1996. Statistical inference and data mining. Communication of the ACM), (Stanfford, G.C., P.E. Kelley, J.E.P. Syka, W.E. Reynolds and J.F. Todd, 1984. Recent improvements in and analytical applications of advanced ion-trap technology. Intl. J. Mass Spectrometry Ion Processes).Stranieri and Zeleznikow (2000; 2004) note that the lack of Knowledge Discovery from Database applications in Law is due to the fact that most legal cases are stored in free text rather than databases. To place legal cases in databases requires massive datapreprocessing, data cleaning and data transformation efforts.

3. DATA MINING CONCEPTS

Data mining is consider to be an emerging technology that has made a revolutionary change in the information technology world. The team 'data mining' (often called as knowledge discovery) refers to the process of analysing data from different perspectives and summarizing it into useful information by means of a number of analytical tools and techniques, which in turn may be useful to increase the performance of a system. Therefore, data mining consists of major functional elements that transform data onto data warehouse, manage data in a multidimensional database, facilities data access to information professionals or analysts, analyze data using application tools and techniques, and meaningfully presents data to provide useful information's.

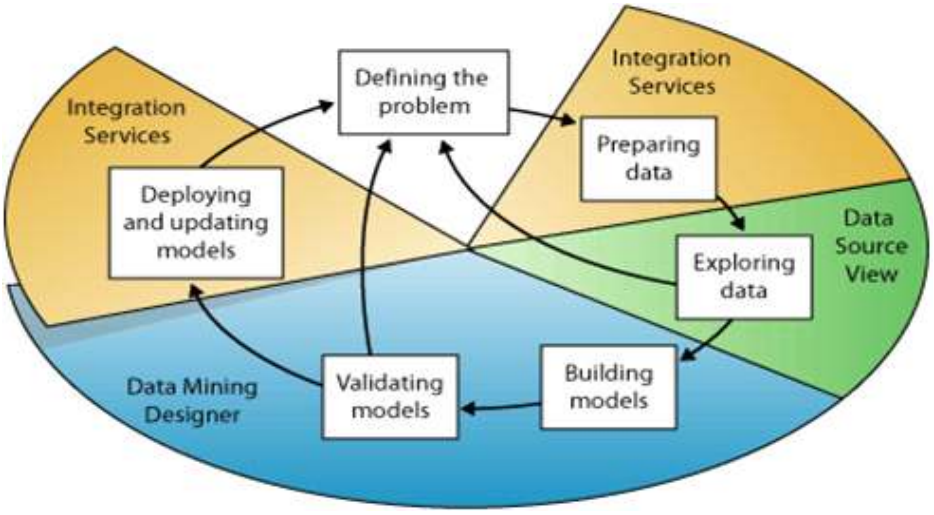

Figure 3.1: Concepts of data mining

4. KNOWLEDGE DISCOVERY

Knowledge discovery is a process that extracts implicit, potentially useful or previously unknown information from the data. The knowledge discovery process is described as follows:

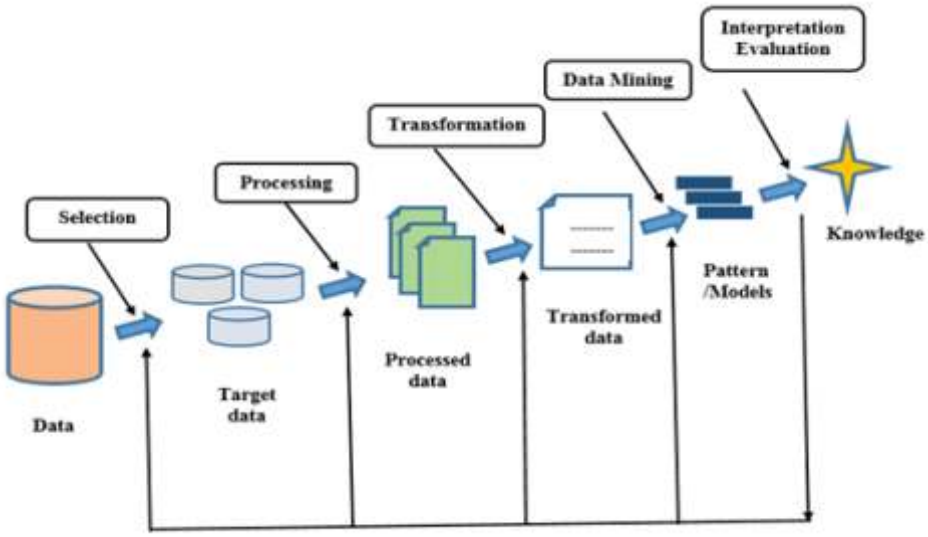

Figure 4.1: Knowledge Discovery Process

The knowledge discovery process are briefly explained below:

- Data comes from variety of sources is integrated into a single data store called target data.
- Data then is pre-processed and transformed into standard format.
- The data mining algorithms process the data to the output in form of patterns or rules.
- Then those patterns and rules are interpreted to new or useful knowledge or information.

The ultimate goal of knowledge discovery and data mining process is to find the patterns that

are hide among the huge sets of data and interpret them to useful knowledge and information. As described in process diagram above, data mining is a central part of knowledge discovery process.

Here the list of steps involved in the knowledge discovery process:

- Data Cleaning – In this step, the noise and inconsistent data is removed.
- Data Integration – In this step, multiple data sources are combined.
- Data Selection – In this step, data relevant to the analysis task are retrieved from the database.
- Data Transformation – In this step, data is transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations.
- Data Mining – In this step, intelligent methods are applied in order to extract data patterns.
- Pattern Evaluation – In this step, data patterns are evaluated.
- Knowledge Presentation – In this step, knowledge is represented.

5. DATA MINING APPLICATIONS IN DIFFERENT SECTORS

5.1 Financial Data Analysis

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

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

5.2 Retail Industry

Data mining in retail industry helps in identifying customer buying patterns and trends that lead to improved quality of customer service and good customer retention and satisfaction. Here is the list of data mining in the retail industry:

- Design and Construction of data warehouses based on the benefits of data mining.
- Multidimensional analysis of sales, customers, products, time and region.
- Analysis of effectiveness of sales campaigns.
- Customer Retention.
- Product recommendation and cross-referencing of items.

5.3 Telecommunication Industry

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

- Multidimensional Analysis of Telecommunication data.
- Fraudulent pattern analysis.

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

5.4 Biological Data Analysis

In recent times, we have seen a tremendous growth in the field of biological such as genomics, proteomics, functional Genomics and biomedical research. Biological data mining is a very important part of Bioinformatics. Following aspects in which data mining contributes for biological data analysis:

- Semantic integration of heterogeneous, distributed genomic and proteomic databases.
- Alignment, indexing, similarity search and comparative analysis multiple nucleotide sequences.
- Discovery of structural patterns and analysis of genetic networks and protein pathways.
- Association and path analysis.
- Visualization tools in genetic data analysis.

5.5 Insurance

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

- Data mining is applied in claims analysis such as identifying which medical procedures are claimed together.
- Data mining enables to forecasts which customers will potentially purchase new policies.
- Data mining allows insurance companies to detect risky customer's behaviour patterns.
- Data mining helps detect fraudulent behaviour.

6. DATA MINING SYSTEM CLASSIFICATION

A data mining system can be classified according to the following criteria:

- Database Technology
- Statistics
- Information Science
- Visualization
- Other Disciplines

Figure 6.1: System Classification

7. TRENDS IN DATA MINING

Data mining are still evolving and here the latest trends in this fields are given below:

- Application Exploration.
- Scalable and interactive data mining methods.
- Integration of data mining with database systems, data warehouse systems and web database systems.
- Standardization of data mining query language.
- Visual data mining.
- New methods for mining complex types of data.
- Biological data mining.
- Data mining and software engineering.
- Web mining.
- Distributed data mining.
- Real time data mining.
- Multi database data mining.
- Privacy protection and information security in data mining.

8. DATA MINING TECHNIQUES

There are several major data mining techniques have been developing and using in data mining projects recently including association, classification, clustering, prediction, sequential patterns and decision tree. The followingdata mining techniques are shown below:

8.1 Association

Association is one of the best known data mining technique. In association, a pattern is discovered based on a relationship between items in the same transaction. That's is the reason why association technique is also known as relation technique. The association technique is used in market

basket analysis to identify a set of products that customers frequently purchase together.

8.2 Classification

Classification is a classic data mining technique based on machine learning. Basically classification is used to classify each item in a set of data into one of predefined set of classes or groups. Classification method makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics. In classification, we develop the software that can learn how to classify the data items into groups.

8.3 Clustering

Clustering is a data mining technique that makes meaningful or useful cluster of objects which have similar characteristics using automatic technique. The clustering technique defines the classes and puts objects in each class, while in the classification techniques, objects are assigned into predefined classes.

8.4 Prediction

The prediction, as it name implied, is one of a data mining techniques that discovers relationship between independent variables and relationship between dependent and independent variables. For instance, the prediction analysis technique can be used in sale to predict profit for the future if we consider sale is an independent variable, profit could be a dependent variable. Then based on the historical sale and profit data, we can draw a fitted regression curve that is used for profit prediction.

8.5 Sequential patterns

Sequential patterns analysis is one of data mining technique that seeks to discover or identify similar patterns, regular events or trends in transaction data over a business period.In sales, with historical transaction data, businesses can identify a set of items that customers buy together a different times in a year. Then businesses can use this information to recommend customers buy it with better deals based on their purchasing frequency in the past.

8.6 Decision trees

Decision tree is one of the most used data mining techniques because its model is easy to understand for users. In decision tree technique, the root of the decision tree is a simple question or condition that has multiple answers. Each answer then leads to a set of questions or conditions that help us determine the data so that we can make the final decision based on it.

9. STRENGTH AND WEAKNESS OF DATA MINING

Data mining is an important part of knowledge discovery process that we can analyze an enormous set of data and get hidden and useful knowledge. Data mining is applied effectively not only in business environment but also in other fields such as weather forecast, medicine, transportation, healthcare, insurance, government...etc. Data mining has a lot of strength when using in a specific

industry. Besides those strength, data mining also has its own weakness e.g., privacy, security and misuse of information. The following strength and weakness of data mining in different industries are given below:

9.1 Marketing / retail

Data mining helps marketing companies build models based on historical data to predict who will respond to the new marketing campaigns such as direct mail, online marketing campaign…etc. Through the results, marketers will have appropriate approach to sell profitable products to targeted customers.Data mining brings a lot of benefits to retail companies in the same way as marketing. Through market basket analysis, a store can have an appropriate production arrangement in a way that customers can buy frequent buying products together with pleasant. In addition, it also helps the retail companies offer certain discounts for particular products that will attract more customers.

9.2 Finance / Banking

Data mining gives financial institutions information about loan information and credit reporting. By building a model from historical customer's data, the bank and financial institution can determine good and bad loans. In addition, data mining helps banks detect fraudulent credit card transactions to protect credit card's owner.

9.3 Manufacturing

By applying data mining in operational engineering data, manufacturers can detect faulty equipment's and determine optimal control parameters. For example semi-conductor manufacturers has a challenge that even the conditions of manufacturing environments at different wafer production plants are similar, the quality of wafer are lot the same and some for unknown reasons even has defects. Data mining has been applying to determine the ranges of control parameters that lead to the production of golden wafer. Then those optimal control parameters are used to manufacture wafers with desired quality.

9.4 Governments

Data mining helps government agency by digging and analyzing records of financial transaction to build patterns that can detect money laundering or criminal activities.

9.5Privacy Issues

The concerns about the personal privacy have been increasing enormously recently especially when internet is booming with social networks, e-commerce, forums, blogs…. Because of privacy issues, people are afraid of their personal information is collected and used in unethical way that potentially causing them a lot of troubles. Businesses collect information about their customers in many ways for understanding their purchasing behaviours trends. However businesses don't last forever, some days they may be acquired by other or gone. At this time the personal information they own probably is sold to other or leak.

9.6Security issues

Security is a big issue. Businesses own information about their employees and customers including social security number, birthday, payroll and etc. However how properly this information is taken care is still in questions. There have been a lot of cases that hackers accessed and stole big data of customers from big corporation such as Ford Motor Credit Company, Sony… with so much personal and financial information available, the credit card stolen and identity theft become a big problem.

9.7Misuse of information/inaccurate information

Information is collected through data mining intended for the ethical purposes can be misused. This information may be exploited by unethical people or businesses to take benefits of vulnerable people or discriminate against a group of people.In addition, data mining technique is not perfectly accurate. Therefore if inaccurate information is used for decision-making, it will cause serious consequence.

10. CONCLUSION AND FUTURE ENHANCEMENT

This paper discussed about the applications of data mining techniques. Data mining brings a lot of benefits to businesses, society, governments as well as individual. However privacy, security and misuse of information are the big problems if they are not addressed and resolved properly. Data mining can be integrated into KDD framework and enhanced the KDD process with better knowledge. It is clear that the data mining techniques will have a major impact on the practice of KDD, and will present significance challenges for future knowledge and information systems research. 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. Data mining is a way to discover new meaning in data performs data processing sophisticated data search capabilities and statistical algorithms, which can be utilized in any organization or system that needs to determine the patterns or relationships implicit ina large data warehouse for better strategies to best reach them.

REFERENCES

1.Adriaans, P., and Zantige, D. (1996) Data Mining. Harlow, UK: Addison-Wesley.
2.Bayardo, R.J., and Agrawal, R. (1999) Mining the most interesting rules. Proc. 5th ACMSIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-99). New York: ACM Press, pp. 145–154.
3.Berson, A., Smith, S. and Thearling, K. Building Data Mining Applications for CRM, McGraw-Hill Professional, 1999.
4.Cosper Nate, Insightful Strategies for Increasing Revenues in thePharmaceuticals Industry: Data Mining for Successful Drugs,2003.
5.Fayadd, U., Piatesky -Shapiro, G., and Smyth, P, From Data Mining to Knowledge Discovery in Databases", The MIT Press, ISBN 0–26256097–6, Fayap, 1996.
6.Feelders, A., Daniels, H. and Holsheimer, M. (2000) 'Methodological and Practical Aspects of Data Mining', Information andManagement, pp.271-281.
7.Frawley and Piatetsky-Shapiro, 1996. Knowledge Discovery in Databases: An Overview. The AAAI/MIT Press, Menlo Park, C.A.

8.Friedman, J. (1997) On bias, variance, 0/1 loss, and the curse of dimensionality. Data Mining and Knowledge Discovery, pp. 55–77.
9.Glymour, C., D. Madigan, D. Pregidon and P.Smyth, 1996. Statistical inference and data mining. Communication of the ACM, pp: 35-41.
10.Han, J., and Kamber, M. (2000) Data Mining: Concepts and Techniques, San Francisco, CA: Morgan Kaufmann.
11.Hand David, MannilaHeikki and Smyth Padhaic. (2001) Principles of Data Mining, The MIT Press, pp.9-10.
12.Holsheimer, M., Kersten, M., Mannila, H., and Toivonen, H. (1995) A perspective on databases and data mining. Proceedings of the First International Conference on knowledge discovery and data mining, Fayyad, U.M., and Uthu-rusamy, R. (eds.), Menlo Park, CA: AAAI Press, pp. 150–155.
13.Imielinski, T., Virmani, A., and Abdulghani, A. (1999) DMajor application programming interface for database mining. Data Mining and Knowledge Discovery, 3(4), pp. 347–372.
14.Mannila, H. (1997) Inductive databases and condensed representations: Concepts for data mining. International Logic Programming Symposium 1997, Cambridge, MA: MIT Press, pp. 21–30.
15.Mitchell, T. Machine Learning. McGraw-Hill, New York, 1997.
16.Shakil Ahmed, FransCoenen, Paul Leng, Knowledge Information System: 2006, Tree based partitioning of data for association rule mining.
17.Piatetsky-Shapiro, G. (1999) The data-mining industry coming of age. IEEE Expert, 14(6), pp. 32–34.
18.Sarawagi, S., Thomas, S., and Agrawal, R. (2000) Integrating association rule mining with relational database systems: alternatives and implications. Data Mining and Knowledge Discovery, 4, pp. 89–125.
19.Stanfford, G.C., P.E. Kelley, J.E.P. Syka, W.E. Reynolds and J.F. Todd, 1984. Recent improvements in and analytical applications of advanced ion-trap technology. Intl. J. Mass Spectrometry Ion Processes, 60: 85-98.
20.Stranieri, A., and Zeleznikow, J., Knowledge discovery for decision support in law. ICIS 2000: 635-639, 2000.
21.Ullman, J.D. (1988) Principles of Database and Knowledge-Base Systems, vol. 1. Rockville, MD: Computer Science Press.
22.Van Rijsbergen, C.J. (1979) Information Retrieval. London: Butterworth Press.
23.Weiss, S.M., and Indurkhya, N. (1998) Predictive Data Mining: A Practical Guide. SanFrancisco, CA: Morgan Kaufmann.
24.Witten, I.H., and Franke, E. (2000) Data Mining: Practical Machine Learning Tools andTechniques with Java Implementations. San Francisco, CA: Morgan Kaufmann.
25.Zhang, T., Ramakrishnan, R., and Livny, M. (1997) BIRCH: an efficient data clustering method for very large databases. Data Mining and Knowledge Discovery, 1(2), pp. 141–182.