

Applications of Data Mining in Diverse Business Domains

Dr. Surwade Dayanand
Associate Professor
dr.surwade@yahoo.com

DATTAKALA GROUP OF INSITUTUION

Abstract:-Today, Computers have become the most integral part of human life. Computers are used for various purposes in different business domains which generate huge volume of data related to different business processes. This leads to many challenging problems and issues for computer science and engineering, which includes data storage, data mining, knowledge management, security etc. Various organizations face the challenges in handling the operational data and effective use of these data to generate the knowledge. This knowledge can be used to improve the business processes as well as achieving the customer satisfaction. Data mining can be used to achieve these goals. Different data mining techniques can be used to identify the different patterns in the data. These patterns can be used to understand the outcome of existing business process. In this paper we have briefly discussed the data mining techniques and usefulness of datamining in marketing, education, pharmaceutical and health care and travels and tourism sectors.

1. Introduction

Evolution in information technology and databases generated a huge amount of data in various areas. Research in databases and information technology from 1970's has given a rise to systematic approach to store and manipulate the data in databases e.g. relational databases. Knowledge

discovery from databases (KDD)[3] is a next step in databases to extract the useful knowledge and pattern from this huge data. Using KDD we can identify the various aspects of our business. KDD includes the different steps for gathering the data, preparing the data, cleaning the data, data mining and knowledge representation.

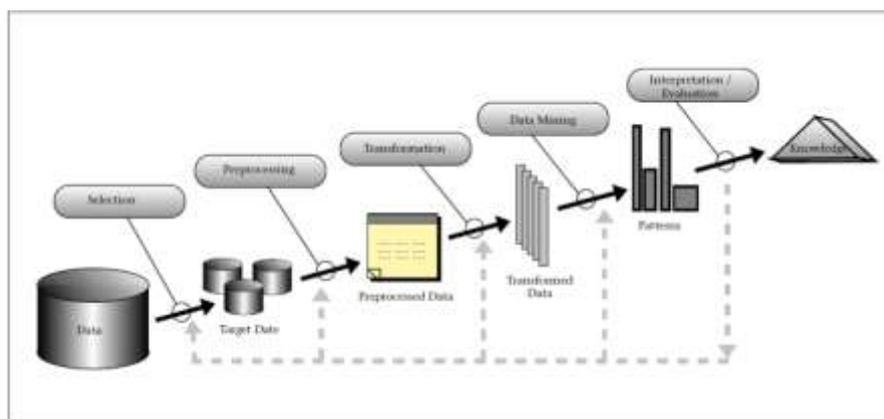


Fig.1: knowledge discovery from databases

The process of knowledge discovery inherently consists of several steps as follows:[3]

1. Understand the application domain and to formulate the problem
2. Collect and preprocess the data
3. Data mining
4. Interpret (post-process) discovered knowledge
5. Put discovered knowledge in practical use

Data Mining is the rapidly growing interdisciplinary field[1][3], which merges together database management, statistics, machine learning and related areas aiming at extracting useful knowledge from large collections of data. Data mining is a step in the knowledge discovery process consisting of particular data mining algorithms that, under some acceptable computational efficiency limitations, finds patterns or models in data.

Data Mining also can be define as an analytic process designed to explore large amounts of data in search for consistent patterns and systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data. The process thus consists of three basic stages: exploration, model building or pattern definition, and validation/verification. Ideally, if the nature of available data allows, it is typically repeated iteratively until a "robust" model is identified. However, in business practice the options to validate the model at the stage of analysis are typically limited and, thus, the initial results often have the status of heuristics that could influence the decision process.

2. Data Mining algorithms and Techniques[3]

Various algorithms and techniques like Classification, Clustering, Regression, Artificial Intelligence, Neural

Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc are used for knowledge discovery from databases.

2.1 Classification: [2]

Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large.

2.2 Clustering:[7]

Clustering approaches address segmentation problems. These approaches assign records with a large number of attributes into a relatively small set of groups or "segments." This assignment process is performed automatically by clustering algorithms that identify the distinguishing characteristics of the dataset and then partition the n-dimensional space defined by the dataset attributes along natural cleaving boundaries. There is no need to identify the groupings desired or the attributes that should be used to segment the dataset.

Clustering is often one of the first steps in data mining analysis. It identifies groups of related records that can be used as a starting point for exploring further relationships [13]. This technique supports the development of population segmentation models, such as demographic-based customer segmentation. Additional analyses using standard analytical and other data mining techniques can determine the characteristics of these segments with respect to some desired outcome.

2.3 Regression: [2][13]

Regression uses existing values to forecast what other values will be. In the simplest case, regression uses standard statistical techniques such as linear regression. Unfortunately, many real-world problems are not simply linear projections of previous values. For instance, sales volumes, stock prices, and product failure rates are all very difficult to predict because they may depend on complex interactions of multiple predictor variables. Therefore, more complex techniques (e.g., logistic regression, decision trees, or neural nets) may be necessary to forecast future values. The same model types can often be used for both regression and classification.

2.4 Association Rule

The Association Rules[3] model is often associated with "market basket analysis", which is used to discover relationships or correlations among a set of items. It is widely used in data analysis for direct marketing, catalog design, and other business decision-making processes. Association rules capture the co-occurrence of items or events in large volumes of customer transaction data.

Association Rule algorithms [1] need to be able to generate rules with confidence values less than one. However the number of possible Association Rules for a given dataset is generally very large and a high proportion of the rules are usually of little (if any) value.

2.5 Neural networks

Neural networks are an approach to computing that involves developing mathematical structures with the ability to learn [1]. The methods are the result of academic investigations to model nervous system learning. Neural networks have the

remarkable ability to derive meaning from complicated or imprecise data and can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques[2]. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyze. This expert can then be used to provide projections given new situations of interest and answer "what if" questions.

Neural networks have broad applicability to real world business problems and have already been successfully applied in many industries. Since neural networks are best at identifying patterns or trends in data, they are well suited for prediction or forecasting needs.

3. Data mining applications

3.1 Marketing

Marketing is the process of communicating the value of a product or service to customers, for the purpose of selling that product or service. It's a very difficult task to be in competition in the market.

Data mining techniques help the organization to interact with existing customers and prospect customers[12]. Data mining can be used to understand the existing customer behavior. Based on this, decisions can be made to offer new products to the customers with different promotional activities like discounts and so on. This helps the organizations to control the attrition of the customers.

Secondly Customer Retention is the key concern for any industry which is directly proportional to marketing strategies. Building and developing good customer relations becomes a challenge when a company has a large number of customers. Data mining is effectively used in customer relationship management for finding out what to offer to which customer and when to offer it[12]. Hence Data mining plays important role in business development and the technique is effectively used for marketing promotions, distribution and advertising media, customer profiling, purchase frequency, recognizing success/failure of marketing programs and forecasting future sales.

3.2 Education

Another interesting application of data mining is in education sector. It helps the institutes to identify the prospective students which may be interested to enroll in the particular institute for particular course. Secondly for the enrolled student the institute's wants to know the existing skill sets of student[5]. Also the area where the students need to work on improvement. To improve the skill set of the student we need to have complete understanding of the learning process and resources required. These tasks can be carried out with the data associated with the students. These data need to be stored and effectively analyzed to gain the student enrollments, as well as to groom the existing students. It is also referred as educational data mining [6]. The application of educational mining helps to understand the patterns related with the student enhancement and discovering and improving the learning process. Data mining can play a vital role in design the curriculum for a particular course. It can be used effectively for the course and subject selection for the students. Data mining can be used to classify the students based on various factors like

demography, psychographic and skill sets. Dr. Mohd Maqsood Ali[4] has discussed the following different applications.[4]

- Students' Enrolment Prediction
- Predicting Students' Profiling
- Curriculum Development
- Students' complaints
- Course Completion
- Students' Targeting
- Library facility
- Students' course selection
- Teachers' teaching performance
- Students' Performance
- Students' Dropouts
- Students' relationship management

3.3 Healthcare and Pharmaceutical

Healthcare and Pharmaceutical industries produce huge amount of data in research, diagnosis and treatment [8]. This data cannot produce reliable reports due to lack of determined structure, as it is collected from various reports, clinics, and laboratories and hence it lacks information. Data mining helps healthcare organization to transform data into useful information. Efficient use of Data mining techniques can successfully contribute in the prediction of complex event in healthcare and pharmaceutical industries. Data mining always helps in improving decisions and reducing expenditure through disease management.

Data mining applications help to evaluate effectiveness of medical treatments by comparing causes, symptoms and course of action [9]. It can help in identifying successful standard treatments for specific disease. Data mining primarily helps in finding medical insurance fraud and abuses. Also pharmaceutical industry is widely benefited by data mining. It helps in tracking which physician prescribes which drug for which disease so that pharma companies can decide whom to target for their specific drug composition. Pharma companies are now effectively using data mining techniques in drug development which has basic steps: finding new drugs, development tests and prediction of drug behavior, clinical trials in human and its commercialization [8]. It also helps in Clinical data analysis to see trends, irregularities and calculation of risk factors, marketing and sales analysis, targeting physicians who have high prescription rate of a certain drug and finally managing all inventory [9].

3.4 Travel & Tourism industry [10][11]

Travel & Tourism is the most revenue generating business domain for many countries. This industry generates a significant amount of data in terms of travel packages, travelers' destination preferences, traveling media and hospitality[10]. Managing this data for future marketing strategies is always a challenge for tourism companies. Data mining provides a key to this challenge in terms of understanding the traveler profile and traveling patterns, so that tourism organizations can market their destination and packages in better way. Also data mining techniques can help finding customer's choices for places to travel, how

and when they spend on traveling and their hospitality preferences. Good customer relationship is the base for tourism industry[11]. Data mining technology allows hotel companies to predict consumer behavior trends, which are useful for marketing strategies.

4. Conclusion

In this paper we have briefly discussed the data mining techniques and its use in various business domains. Different data mining techniques like classification, clustering, association rules, regression, neural networks etc can be effectively used to improve the business outcomes. Data mining can be used in marketing to provide better services to the customer. In education sector data mining will be useful in various functional areas related with the student, teacher and curriculum development. In pharmaceutical and health care sector, data mining can be used to reduce the time required in new drug development, promotion of new drugs and diagnosis. For the travel and tourism sector, data mining can be helpful to understand the travel pattern of the customer. This leads to a large research opportunity in terms of data mining and its application in various business sectors.

5. References

- [1]. Daniel T. Larose, "Data mining methods and models", John Wiley & Sons, INC Publication, Hoboken, New Jersey (2006)
- [2]. Mehmed Kantardzic, Data Mining: Concepts, Models, Methods, and Algorithms, John Wiley & Sons, (2003)
- [3]. Fayyad et al., machine learning and data mining, American Association for Artificial Intelligence. PP 37-54, (1996)
- [4]. Dr. Mohd Maqsood Ali, Role Of Data Mining In Education Sector, International Journal Of Computer Science And Mobile Computing, Vol. 2, Issue. 4, April 2013, Pg.374 – 383
- [5]. Jing Luan, Data Mining Applications in Higher Education, SPSS executive report. 2004
- [6]. Baker, Data Mining for Education, International Encyclopedia of Education (3rd edition), Oxford, UK: Elsevier, 2010, Pages 112-118
- [7]. S. Anitha Elavarasi, A Survey on partition clustering algorithm, International Journal of Enterprise Computing and Business Systems, Vol. 1 Issue 1(2011)
- [8]. Jayanthi Ranjan, Applications of data mining techniques in pharmaceutical industry, Journal of Theoretical and Applied information technology, (2005-2007)
- [9]. Sandhya Joshi, Hanumanthachar Joshi, Application of data mining in health and pharmaceutical industry, International Journal of Scientific & Engineering Research, Volume 4, issue 4, April 2013.
- [10]. Pairaya Juwattanasamran et.al., Applying data mining to analyze travel pattern in searching travel destination choices, The International Journal Of Engineering And Sciences(IJES), Volume 2, Issue 4, 2013.
- [11]. Danubianu Mirela et al, Towards a Distributed data mining system for tourism industry, University of Suceava faculty of Electrical Engineering and Computer Science, Romania.
- [12]. Radhakrishnan B, Shineraj G et. al, Application of Data Mining In Marketing, International Journal of Computer Science and Network, Volume 2, Issue 5, 2013
- [13]. Pavel Berkhin, Survey of Clustering Data Mining Techniques, Grouping Multidimensional Data In Grouping Multidimensional Data (2006)