
DATA MINING IN LIBRARY SCIENCE

Dr. Nilesh V. Gore

M. Lib. Sc, SET, Ph.D.,

Librarian,

Prof. Rajabhau Deshmukh Arts College,
Nandgaon Khandeshwar, Amravati.

E-mail- ngore80@gmail.com

Abstract : *Library and information services in schools, colleges, universities, corporations and communities obtain information about their users, circulation history, resources in the collection, and search patterns. Understanding the benefits of data is essential from business point of view. Data can be used for classifying and predicting the students' behaviour, performance, dropouts as well as teachers' performance. Knowledge Discovery or Data Mining is the partially automated process of extracting patterns, usually from large data sets. The paper tries to give an overview on data sources and possible applications and techniques of data mining in the library.)*

Keywords : *Data Mining, Data Mining Techniques, Data Mining Applications.*

Introduction:

Data mining is an essential tool, which is used to predict and classify the data collected from the customers. Today, many industries use data mining to make decisions and to prepare marketing strategies towards target segmented customers for the achievement of its objectives. But many universities neglected practicing data mining techniques. Application of data mining in education sector is an emerging trend in the global competitive business. Understanding the data mining terms, tasks, techniques and application are foundation of developing data mining in education sector. Therefore, there is a need to study the role of data mining in education sector.

We live in the Age of Information. The importance of collecting data that reflect in business or scientific activities to achieve competitive advantage is widely recognized now. Powerful systems for collecting data and managing it in large databases are in place in all large and mid-range institutions and organizations.

Data Mining:

Data mining can be defined as "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 allow the prediction of future results. Typical applications include market segmentation, customer profiling, fraud detection, evaluation of retail promotions, and credit risk analysis."

Stages of the Data-Mining :

1. Data gathering, e.g., data warehousing, Web crawling.

2. Data cleansing: eliminate errors and/or bogus data,
3. Feature extraction: obtaining only the interesting attributes of the data, e.g., \data acquired” is probably not useful for clustering celestial objects, as in Skycat.
4. Pattern extraction and discovery. This is the stage that is often thought of as \data mining,” and is where we shall concentrate our effort.
5. Visualization of the data. Application of Data Mining
6. Evaluation of results; not every discovered fact is useful, or even true! Judgments are necessary before following your software’s conclusions.

Working Process of data mining :

While large-scale information technology has been evolving separate transaction and analytical systems, data mining provides the link between the two. Data mining software analyzes relationships and patterns in stored transaction data based on open-ended user queries. Several types of analytical software are available: statistical, machine learning, and neural networks. Generally, any of four types of relationships are sought:

Classes: Stored data is used to locate data in predetermined groups. For example, a restaurant chain could mine customer purchase data to determine when customers visit and what they typically order. This information could be used to increase traffic by having daily specials.

Clusters: Data items are grouped according to logical relationships or consumer preferences. For example, data can be mined to identify market segments or consumer affinities.

Associations: Data can be mined to identify associations. The beer-diaper example is an example of associative mining.

Sequential patterns: Data is mined to anticipate behavior patterns and trends. For example, an outdoor equipment retailer could predict the likelihood of a backpack being purchased based on a consumer’s purchase of sleeping bags and hiking shoes.

Data Mining Benefits:

Data mining was used in financial services, banking, retail, healthcare, manufacturing, telecommunication, food industries, engineering, motor industries, forecasting water, road traffic, pharmaceutical, drug discovery, job seekers and insurance sectors but not used in an education sector. The following may be the benefits of data mining in an education sector: identifying students’ needs and preferences towards course choices, and selection of specialization

- identifying students’ pattern trends
- predicting students’ knowledge, grades, and final results
- supporting automatic exploration of data
- Constructing students’ profiles become easy
- helping management to understand business

Data Mining Techniques:

In this section, the most common data mining techniques are discussed to understand the theory without going into details. According to Chen et al. (2005) data mining brings various techniques together to discover pattern and to construct models from database. Ngai

et al. (2009) indicates seven types of data mining model: Association, Classification, Clustering, Forecasting, Regression, Sequence Discovery and Visualizations. Huang et al. (2012) concludes that data mining technique is used to gain useful information or interesting knowledge. Perceived usefulness and perceived ease of use are the factors that affect an individual intention to use data mining tools.

1. Neural Networks: It is techniques which can be used for classification of large complex data. It can be used to study courseselection by students, student course satisfaction, and specialisation selection. Rokach (2010) state the neuralnetwork represents each cluster by a neuron or “prototype”. The input data is also represents by neurons whichare connected to the prototype neurons. Each such connection has a weight, which is learned adaptability duringlearning.

2. Decisions Tree: Decision tree is a data mining technique that can be used for classification and prediction of large data. Decision tree is used for profiling customers. Decision tree is also called rule induction technique (Luan, 2003). According to Rokach (2010) the data, in decision tree, is represented by “a hierarchical tree where each leafrefers to a concept and contains a probability description of the concept.” Furthermore, Yin et al. (2011) defines decision tree as a “systematic method that uses a tree diagram. Itconsists of nodes and branches, nodes are connected by branches, time flows from left to right, each ranchrepresents a decision or a possible event.” In addition, decision tree make classification easy and understandableand also result-oriented. Many industries use to classify and predict customers’ behaviour, acquit ion, retention,and growth. Similarly, universities can use this technique to classify students’ performance, behaviour, andexpectations.

3. Regression Analysis: Bryman and Duncan (2005) stated that regression is “a powerful tool for summering the nature of therelationship between variables and for making predictions of likely values of the dependent variables.” It is usedas a continuous variable (Nisbet, 2009). Regression analysis can be applied in data mining to predict students’GPA, expectations with regard to specialisation selection, complaints and satisfaction levels based ondemographic, geographic, psychographic and behavioural variables. Furthermore, simple linear regression andmultiple linear regressions are techniques used in regression analysis. These techniques can be done by usingSPSS software.

4. Cluster Analysis :Customer analysis is an unsupervised learning technique (Tsai et al., 2011). Customer analysis refers toidentifying groups of customers with similar characteristics (Ahn and Sohn, 2009), splitting the full data set intoa set of clusters (Baker, 2010) where categories are not known in advance. Han and Kamber (2006) indicate thatcluster analysis can be used to generate labels. The objects are clustered or grouped based on the principles ofmaximizing the intra-class similarly and minimizing the interclass similarity. Clustering is also known assegmentation (Sinha et al., 2010). Segmentation can be done based on demographic variables, (e.g., gender,income, age, qualification, religion, occupation) geographic variables, (e.g., city, country, state, zip code, region)psychographic variables (e.g., lifestyle, personality and values) and behavioural variables (e.g., benefits, andoccasion)

In brief, classification and regression models are supervised model and cluster analysis is an unsupervisedmodel (Rokach and Maimon, 2010). Both models are used to

predict and classify the large data into usable information. Consequently, universities can use cluster analysis to examine similarities and differences between colleges, students, teachers, administrative staff, courses, and examinations.

Applications DATA MINING in Library and Information Service:

Most people think of libraries as the little brick building in the heart of their community or the big brick building in the center of a campus. These notions greatly oversimplify the world of libraries, however. Most large commercial organizations have dedicated in-house library operations, as do schools, nongovernmental organizations, as well as local, state, and central governments. With the increasing use of the Internet and the World Wide Web, digital libraries have proliferated, and these serve a huge variety of different user audiences, e.g., people interested in health and medicine, science and technology, industry and world news, law, and business. With this expanded view of libraries, two key insights arise. First, libraries are nearly always embedded within larger institutions. Corporate libraries serve their corporations, academic libraries serve their universities, and public libraries serve the general public. Second, libraries play an important role within their institutions as repositories and providers of information resources. In the provider role, libraries represent in microcosm the intellectual, learning, and knowledge management activities of the people who comprise the institution.

This fact provides the basis for the strategic importance of library data mining. By ascertaining what users need to know and how well those needs are served, bibliomining can reveal insights that have meaning in the context of the library's host institution. Using bibliomining, libraries can ascertain what their constituencies want to learn, whether they find the information they seek, and whether that information satisfies their learning and knowledge needs. In corporate libraries, which serve the knowledge needs of commercial organizations, such insights can help to develop and maintain a competitive, cutting edge workforce. In special libraries, which support the research needs of government and nongovernmental organizations, these insights can influence the success of policies and programs that are informed by research. In academic organizations, accurate insights into faculty and student knowledge needs can enhance the viability of the whole institution. To understand libraries can help to achieve these insights, and thus help to enhance the effectiveness of their host organizations or communities, it is important to understand the workflow and associated dataflow that occur within a prototypical library.

Conclusion:

Owing to automation of libraries, they have gathered data about their collections and users for years, but have rarely used those data for better decision-making. By taking a more active approach based on applications of data mining, data visualization, and statistics, information organizations can get a clearer picture of their information delivery and management needs. At the same time, libraries must continue to protect their users and employees from misuse of personally identifiable data records. Libraries must compete against online booksellers, downloadable audio books, and the vast supply of "free" information of varying quality from the Internet, librarians must begin to take the initiative in

using their systems and data for competitive advantage and to justify continued support and funding of libraries.

References :

- Aggarwal, C. Charu and Yu, S. Philip. "Data Mining Techniques for Associations, Clustering and Classification." in Zhong, Ning and Zhou, Lizhu (Eds.) methodologies for knowledge discovery and data mining, third Pacific Asia Conference, PAKDD, Beijing, China, April 26-28, 1999 proceedings, Springer, New York
- Baritchi, Andi. Data Mining and Knowledge Discovery, in (Eds.) Business Intelligence in the Digital Economy, Raisinghani, S. Mahesh, 2004, Idea Group Inc, 2004,
- Delen, Dursun. "Analysis of Cancer Data: A Data Mining Approach." Expert Systems, vol.26 No.1(2009)
- Gopalan, N.P and Sivaselvan, B. Data Mining: Techniques and Trends, PHI, New Delhi, 2009
- Guo, W. William, "Incorporating statistical and neural network approaches for student course satisfaction analysis and prediction." Expert System with Applications, 37 (2010) 3358-3365
- Han, Jiawei and Kamber, Micheline. Data Mining: Concepts and Techniques, Second Edition, MorganKauffmann Publishers.
- Kardan, A. Ahmad., Sadegh, Hamid, Ghidary, Shiry. Saeed., and Sani, Reza Fani., Mohammad. "Prediction of student course selection in online higher education institutions" 65 (2013): 1-11
- Kumar, A.N.V. and Uma, G.V. "Improving academic performance of students by applying data mining techniques." European Journal of Scientific Research, No. 4 (2009): 526-534
- Nisbet, Robert and Elder, John and Miner, Gary., Handbooks of Statistical analysis and Data Mining Applications, Academic Press Publications, 2009
- Maqsood, A. Mohammed, "Customer Relationship Management in B-Schools: An Overview." International Journal of Computer Sciences and Management Research, Vol. 2 issue 4 (April 2013):
- Mark, Last. Data Mining, in Cyber Warfare and Cyber Terrorism, Janczewski, Lech and Colarik, Andrew., Information Science Reforms, IGI Global, (2007): 358-365
- Olson L. David and Delen, Dursun. Advanced Data Mining Technique, 2008, Springer
- Pujari, K. Arun. Data Mining Techniques, Universities Press India Pvt. Ltd., Hyderabad, 2001
- Rajamani, Karthick., Sung, Sam., and Lox, Alan. "Extending the applicability of association rules." The proceedings of Third Pacific Asia Conference, PAKDD 99, Beijing, China, April 26-28, 1999
- Ramageri, M. Bharti "Data Mining Techniques and Applications." Indian Journal of Computer sciences and Engineering, Vol.1 No.4, (Dec-Jan 2010-2011): 301-305



- Sinha, Nidhi., Vandana, Ahuja., and Medury, Y. “Effective Brand Management through customerprofiling using clustering, IUP Journal of Brand Management, 2010, 21-33
- Tiwari, Mahendra.,Randhir Singh and NeerajVimal., “An Empirical Study of Application of DataMining Techniques for Predicting Student Performance in Higher Education.” International Journal ofComputer Sciences and mobile Computing, Vol.2 Issue 2 (Feb 2013): 53-57
- Tsai, C.F., Tsai, C.T., Hung, C.S., Hwang, P.S., “Data Mining Techniques for identifying students at riskof failing a computer proficiency test requires for graduation.” Australian Journal of EducationalTechnology, 2011, 27(3): 481-498