



RESEARCH ARTICLE

APPLICATIONS OF DATA MINING IN LIBRARY AND INFORMATION CENTRES: AN OVERVIEW

***Dr. Shantashree Sengupta**

Librarian & Head, Department of Library & Information Science, Chandmal Tarachand Bora College,
Shirur, Dist.Pune-412210, Maharashtra, India

ARTICLE INFO

Article History:

Received 03rd October, 2016
Received in revised form
18th November, 2016
Accepted 11th December, 2016
Published online 31st January, 2017

Key words:

Data Mining, Knowledge Discovery in
Databases (KDD), Data Warehouse,
Data Mining Applications in Library,
Data Mining Techniques.

Copyright©2017, Dr. Shantashree Sengupta. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Citation: Dr. Shantashree Sengupta, 2017. "Applications of data mining in library & information centres: an overview", *International Journal of Current Research*, 9, (01), 45246-45249.

ABSTRACT

The 21st Century is witnessing a massive flow of information in library. The increasing amount of information can be related to the extensive application of Information & Communication Technology. Today, one of the biggest challenges that libraries face is the enormous amount of data generation and to use this data to improve the quality of managerial decisions. Data mining techniques are analytical tools that can be used to extract meaningful knowledge from large data sets. To do this extraction, data mining combines artificial intelligence, statistical analysis and database management systems to attempt to pull knowledge from stored data. The present paper addresses the applications of data mining in library in an effective and efficient manner.

INTRODUCTION

Data mining is an interdisciplinary subfield of computer science. It is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD. In modern world a huge amount of data is available which can be used effectively to produce indispensable information. The information achieved can be used across various fields. As huge amount of data is being collected and stored in the databases, traditional statistical techniques and database management tools are no longer adequate for analyzing this huge amount of data. Data Mining also called data or knowledge discovery has become the area of growing significance because it helps in analyzing data from different aspects and condensing them into meaningful information.

Origin of the term 'Data Mining'

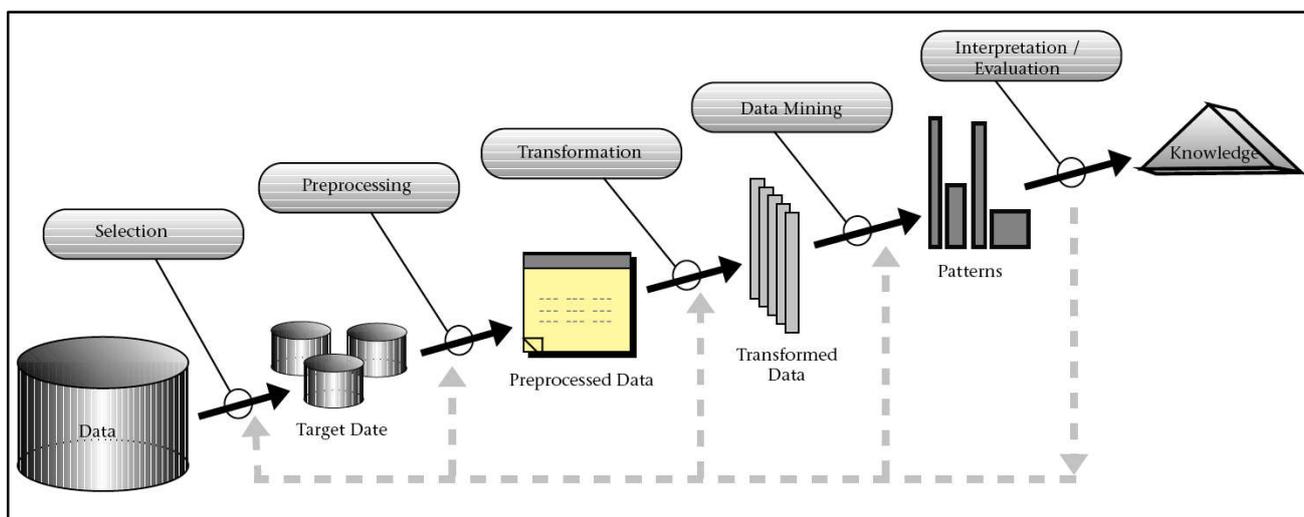
In the 1960s, statisticians used terms like "Data Fishing" or "Data Dredging" to refer to the process of analyzing data

without a theoretical hypothesis. The term "Data Mining" appeared around 1990 in the database community. Some other synonymous terms used are Data Archaeology, Information Harvesting, Information Discovery, Knowledge Extraction, etc. Gregory Piatetsky-Shapiro coined the term "Knowledge Discovery in Databases" for the first workshop on the same topic (KDD-1989) and this term became more popular in Artificial Intelligence (AI) and Machine Learning Community. However, the term data mining became more popular in the business and press communities. Currently, Data Mining and Knowledge Discovery are used interchangeably. Fayyad et.al.(1996) however differ in their ideas and explain that KDD refers to the overall process of discovering useful knowledge from data, and data mining refers to a particular step in this process. Data mining is the application of specific algorithms for extracting patterns from data.

The distinction between the KDD process and the data-mining step (within the process) is a central point of this article. The additional steps in the KDD process, such as data preparation, data selection, data cleaning, incorporation of appropriate prior knowledge, and proper interpretation of the results of mining, are essential to ensure that useful knowledge is derived from the data. Blind application of data-mining methods (rightly criticized as data dredging in the statistical literature) can be a dangerous activity, easily leading to the discovery of meaningless and invalid patterns.

*Corresponding author: Dr. Shantashree Sengupta

Department of Library & Information Science, Chandmal Tarachand Bora College, Shirur, Pune-412210, Maharashtra, India.



[Source: Fayyad, U. et.al. (1996). From Data Mining to Knowledge Discovery in Databases. Retrieved from <http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf>]

Fig. 1. An Overview of the Steps That Compose the KDD Process

Definition

The *Online Dictionary of Library & Information Science (ODLIS)* defines data mining as “The process of using database applications to identify previously undetected patterns and relationships within an existing set of data, for example, common interests among the clientele of a business or other organization.”

Data Mining: Process

Fayyad *et al.* (1996) defined the Knowledge Discovery in Databases (KDD) process with the stages:

- Selection
- Pre-processing
- Transformation
- Data Mining
- Interpretation/Evaluation.

Data Mining Techniques

There are several major data mining techniques

1. Association

Association is one of the best-known data mining techniques. In association, a pattern is discovered based on a relationship between items in the same transaction. That 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.

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 a 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.

3. Clustering

Clustering is a data mining technique that makes a meaningful or useful cluster of objects which have similar characteristics using the 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. Book Management in the library is a major example of clustering. In a library, varieties of books are available on different subjects. The challenge is how to keep those books in a way that readers can take several books on a particular topic without hassle. By using the clustering technique, we can keep books that have some kinds of similarities in one cluster or one shelf and label it with a meaningful name. If readers want to grab books in that topic, they would only have to go to that shelf instead of looking for the entire library.

4. Prediction

The prediction is one of a data mining techniques that discover the relationship between independent variables and relationship between dependent and independent variables. For instance, the prediction analysis technique can be used in the sale to predict profit for the future if we consider the 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.

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.

6. Decision trees

A decision tree is one of the most common 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.

Two or more data mining techniques are often combined together to form an appropriate process that meets the business needs.

Advantages of Data Mining

- Helps in predicting future trends
- Helps in decision making
- Improves company revenue and lower costs
- Helps in market basket analysis
- Facilitates fraud detection

Disadvantages of Data Mining

- User Privacy/ Security Issues
- Amount of data is over increasing
- Huge cost at implementation stage
- Possible misuse of information
- Inaccurate information

Data Mining and Library & Information Centres

Data mining techniques can be applied in various fields as per their suitability. The fifth law of Library Science states that “*Library is a growing organism*”. In order to provide efficient and effective library services, the need for library automation and digital libraries have arise. Automation is not the only solution to explore the hidden information from the large amount of database. This can be done by applying data mining to library data. Following are the examples of applications of data mining in libraries:

- a) **Classification:** By using data mining we can develop a computer program that will replace the manual classification with automatic classification of library contents.
- b) **Link analysis:** Likewise the paper materials, where similar documents tend to have similar bibliographical references, and frequency of citation is often considered to reflect the quality or importance of document, link analysis assumes that higher-quality or otherwise more desirable documents will generally be linked to more frequently than other documents, and that links in a document reveal something about the content of a document. Link analysis can place frequently linked-to-documents at the top of a list or identify documents that are associated with each other.
- c) **Sequence analysis:** Sequence analysis uses statistical analysis to identify unlinked documents that users are likely to want to read together. It examines the paths that users follow when searching for information and can help identify which documents users are likely to want together.
- d) **Summarization:** Though machine generated abstracts are inferior to human-generated ones in terms of readability and content, yet they can be very useful for helping users decide what items they need. Abstract-generating software typically works by identifying significant words or phrases based on position within documents association with critical phrases.
- e) **Clustering:** Clustering is similar to classification, except that the classes are determined by finding natural groupings in the data items based on probability

analyses rather than by predetermined groupings. Clustering and classification are often used as a starting point for exploring further relationships in data. For example, various search engine break down sites by location, subject, or language before sub-arranging data.

- f) **Searching of Information (Reference Service):** Since the data of the library is continuously growing with an exponential rate and the main problem is how one can search the required information from the large amount of redundant information of the library? This can be possible by applying data mining techniques, so one can say that the data mining is the future of reference service.
- g) **Acquisition:** The third law of library science is “*Every book its reader*”. Application of data mining in the library data will help to find out the required contents that are necessary to acquire next set of documents. This will reduce the work of library staff related to the acquisition as well as will ensure efficient use of library budget.

Conclusion

Data mining has wide application almost in every industry where data is generated that's why data mining is considered one of the most important frontiers in database and information systems and one of the most promising interdisciplinary developments in the field of Information Communication & Technology. There is a need of data mining techniques that will redesign and simplify the working of library like classification, acquisition, circulation and referencing. Therefore, systematic efforts must take place to develop the application of data mining techniques and algorithms for library databases. Also, it must be remembered that tools for Data Mining are very powerful and they require very skilled specialist who can prepare the data and understand the output. Data Mining brings out the patterns and relationships, but the significance and validity of those patterns must be made by the user.

REFERENCES

- Advantages and Disadvantages of Data Mining. Retrieved from <http://www.zentut.com/data-mining/advantages-and-disadvantages-of-data-mining/>
- Clifton, Christopher, 2010. Encyclopaedia Britannica: Definition of Data Mining. Retrieved from <https://www.britannica.com/technology/data-mining>
- Data Mining, Retrieved from https://en.wikipedia.org/wiki/Data_mining
- Definition of Data Mining, Retrieved from http://www.abc-clio.com/ODLIS/odlis_d.aspx
- Dwivedi, R.K. and Bajpai, R.P.2004. Use of Data Mining in the field of Library and Information Science: An Overview. 2nd International CALIBER-2004 [pp.512-519]. New Delhi. Retrieved from http://ir.inflibnet.ac.in/bitstream/1944/375/1/04cali_67.pdf
- Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. 1996. From Data Mining to Knowledge Discovery in Databases. Retrieved from <http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf>
- Ramageri, B.M. Data Mining Techniques and Applications. Indian Journal of Computer Science and Engineering, 1(4), 301-305. Retrieved from <http://www.ijcse.com/docs/IJCSE10-01-04-51.pdf>

Rangnathan, S.R. 1993. Five laws of library science, Sarda Rangnathan Endowment for Library Science, Bangalore.

Shane, L., Bowen, S. & Jun, H.C. Data Mining. Retrieved from <http://bus237datamining.blogspot.in/2012/11/advantages-disadvantages.html>

Uppal, V. and Chindwani, G. 2013. An Empirical Study of Application of Data Mining Techniques in Library System. International Journal of Computer Applications, 74(11), 42-46. Retrieved from <http://research.ijcaonline.org/volume74/number11/pxc3890008.pdf>

What is Data Mining? Retrieved from <http://www.zentut.com/data-mining/what-is-data-mining/>
