

Impact of Data Mining and Data Warehousing in Digital Era

Mr.Shamsher Masih*, Mr.Hari Mohan**, Mr.Mohit Kumar***,Ms.Daljit Kaur****

*(Computer Applications,GGI, Amritsar
Email:shamshermasih59@gmail.com)

**(Computer Applications,GGI, Amritsar
Email: harim1882@gmail.com)

*** (Computer Applications,GGI, Amritsar
Email: mohitkumar000@gmail.com)

**** (Computer Applications,GGI, Amritsar
Email: kaurdaljit78888@gmail.com)

Abstract:

These days libraries are using concept of data mining to improve customer service, manage budgets or helps in taking strategic decision-making in their organizations. Dataware house is like a central library and data mart is departmental library. A digital library (DL) is simply an on-line service that provides access to a large variety of content as well as services. Content include electronic material such as images, video, etc., licensed databases of journals, articles and abstracts.

Keywords — Data mining, KDD, clustering

I. INTRODUCTION

Data mining is Knowledge Discovery in Databases (KDD), is the smart and mathematical process of finding patterns in large database set. The main goal of the data mining is to extract useful information from a set of data and transform it into well-organized structure for future use.

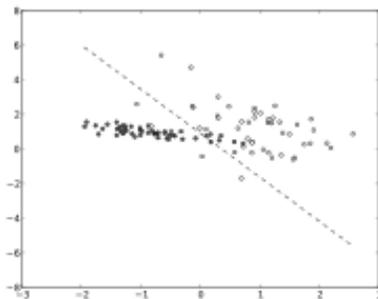


Fig.1 Data Mining

Data Mining involves six common tasks:

- 1.1 **Anomaly detection:** Finding of unusual data records, that may be interesting or data errors that require future investigation
- 1.2 **Association learning rule:** Finds relationships between

variables. Using association rule learning for determining which products are frequently bought together and use that information for other useful purpose which is also referred to as Market Basket Analysis.

- 1.3 **Clustering:** is the task of finding new groups and structures the data that are somehow related with one-another.
- 1.4 **Classification:** It is the task of generalizing known structure to apply to new dataset.
- 1.5 **Regression:** It is the attempts to find a function from large number of functions which contains the data with the least error.
- 1.6 **Summarization:** It is compact representation of the data set, including visualization and report generation. *Text mining is increasingly in many sectors such as news media, science, development, electronic texts (Web, e-mails, digital libraries, etc.), and many more.*

II KNOWLEDGE DISCOVERY IN DATABASES (KDD)

Knowledge Discovery in Databases (KDD) process contains various stages such as Selection, Pre-processing, Transformation, Data Mining and Interpretation/Evaluation.

Data is any text or numbers, or text that can be processed by a computer. As organizations are growing, large amounts

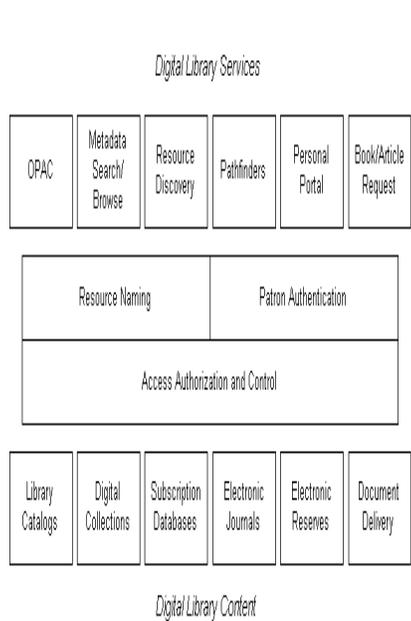
of data in different formats and different databases is also growing from time to time rapidly. [5]

Information is the patterns, associations, or relationships among all these data can provide information. Knowledge discovery allows users to integrate data from various data sources.

Data warehousing represents data is an we organized manner. It provides control repository of all data of the organization. Data warehousing is like data mining and is also defined as a process of centralized data management and retrieval.[7]

III Role of Data mining in Digital Libraries

Library is a growing organization so the volume of the library data is also growing with higher rate. Need of library automation and e-Library occur for efficiently and effectively doing the library administration and extending library services. This can be done by applying the data mining in digital library data in the following ways:



3.1 Classification: Classification mimics library cataloging procedures by grouping structured and unstructured data according to certain criteria such as source, document type, language, subject, or a number of various criteria.

3.2 Link analysis: With link analysis higher-quality or more desirable documents will generally be linked to more frequently than other documents, and that links in ac document reveal something about the content of a document. Link analysis

can place frequently linked-to-documents at top of list or to identify documents that are associated with.

3.3 Sequence analysis: It uses statistical analysis to identify unlinked documents that users are likely to want to read together. It examines paths that one follows when searching for information and can help identify which documents users want together.

3.4 Summarization: summarization is usually done to summarize the data either in tabular form or in chart form so that it can be easily understand as proper planning should be done on their basis such as abstract generating software

3.5 Clustering: Clustering is similar to classification, except that the classes are determined by finding natural groupings rather than by predetermined groupings in the data depending upon analysis. Clustering and classification are often used as a starting point for exploring further relationships in data.

IV Characteristics

4.1 Information Searching: Since the data of the library continuously growing and the main problem is how one can reference the required information form the large amount of superfluous 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.

4.2 Classification: It will replace the manual classification with the computer assisted classification, so that for the classification task less skilled person can also perform it in a fast and efficient way.

4.3 Acquisition: Library science third law “Every book its reader”. On applying the data mining in the library data it can be easily find out the required contents that are necessary to acquire next. It willremoves overburdening of library staff, as well as the efficient use of budget allocated to the library.

V Digital Library

1. Digital library is collection by acquiring, describing, storing and delivering resources.

Many digital objects can be delivered directly over the Web, while some may require special software for viewing various applications.

REFERENCES:

- [1]. <http://dataminingls-atanu.blogspot.in/>
- [2]. http://en.wikipedia.org/wiki/Data_mining
- [3]. http://ir.inflibnet.ac.in/bitstream/1944/435/1/04Planner_22.pdf
- [4]. [http://nopr.niscair.res.in/bitstream/123456789/11552/1/ALIS%2058\(1\)%207-16.pdf](http://nopr.niscair.res.in/bitstream/123456789/11552/1/ALIS%2058(1)%207-16.pdf)
- [5]. <https://net.educause.edu/ir/library/html/edu0122/edu0122.html>
- [6]. T. Li and L. Long, “Imaging examination and quantitative detection and analysis of gastrointestinal diseases based on data mining technology,” *Journal of Medical Systems*, vol. 44, no. 1, pp. 1–15, 2020.
- [7]. J. C. Zuo, “Defense of computer network viruses based on data mining technology,” *International Journal on Network Security*, vol. 20, no. 4, pp. 805–810, 2018. [8] A. Buczak and E. Guven, “A survey of data mining and machine learning methods for cyber security intrusion detection,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2017.