Application of Data Mining for improving Participative Librarianship: a Brief Study

Sudeshna Karmakar
UGC-JRF Research Scholar
Department of Library and Information Science
University of Calcutta, Kolkata- 700 073, West Bengal, India

Abstract: *Data mining helps in analyzing and summarizing different elements of information. Data mining process is a form where in which all the data and information can be extracted for the purpose of future benefit. The data mining system has been designed for librarians to help them manage the library easily. It is relatively new term in the field of library science though it is being used in business organization for a long time. This paper gives an overview of data mining and its use in the field of library science; a process of data mining; advantages of data mining and also the implication of data mining in participative librarianship. Also, discuss the process of bibliomining which is very much useful for analyzing library service.*
Keywords: Data mining, Participative Librarianship, Bibliomining, Information Explosion.

1. Introduction:

Data mining and knowledge discovery in database both terms are used interchangeably. Knowledge discovery in a database is a nontrivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data. Here data means a set of facts and patterns means the expression. The KDD is a process of discovering useful knowledge and mining is a step of this process. Data mining may consist of applying data analysis and discovery algorithms that enumeration of patterns over the data. Data mining consists of a number of operations which supports various technologies like rule induction, neutral networks and conceptual clustering. Participative Librarianship and Data Mining both terms are interrelated in such a way because participative librarianship means both librarians and users are interacting so that new knowledge may be created.

2. Literature Review:

Mishra's (2007) study revealed that the data mining tools predict future trends and behavior, allowing Libraries and Information Centres managers to make proactive, knowledge-driven decisions. Mukhopadhyay and Mukhopadhyay have (2003) also stated that data mining is a form of artificial intelligence that uses automated processes to find information. Dwivedi and Bajpai (2004) stated in their study that new emerging technology in data mining which provides a road map to the next generation of library. Bansal and Sood (2011) writes that thescope and use of Data Warehousing and Data Mining in all dimensions of E-Governance like Government to Citizen (G2C), Government to Government(G2G) and Government to Business(G2B).Hanumanthappa, Prakash and Kumar (2010) observed that how data mining technique in e-government construction by taking BHOOMI project from the government of Karnataka for helps to maintain and update the land records data systematically and securely.

3. Objectives of the study:

* To find out the application of data mining in library and information science fields.
* To find out the usefulness of data mining techniques.

4. What is Knowledge Management:

KM is a process through which organizations generate value from intellectual and knowledge-based assets.

5. Knowledge Management Tools:

- Groupware systems & KM 2.0
- The internet and extranet
- Data warehousing, **data mining**, & OLAP
- Decision support systems
- Content Management System
- Artificial intelligence tools
- Simulation tools
- Semantic networks

6. Data mining:

Data mining 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. It is an interdisciplinary subfield of computer science. 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. Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD.

7. Participative Librarianship:

Simply put participative librarianship recasts library and library practice using the fundamental concept that knowledge is created through conversation. Libraries are in the knowledge business; therefore libraries are in the conversation business. Participative librarians approach their work as facilitators of conversation. Be it in practice, policies, programs and/or tools, participatory librarians seek to enrich, capture, store and disseminate the conversations of their communities.

8. Data Mining and Libraries:

Data mining provides the following facilities to libraries-

- It can provide faster and more thorough access to materials than that provided by manual cataloguing.
- It can be used by employees or users with basic computer and analytical skills so that people can easily find out their needs without the assistance of highly skilled staff.

9. Data mining techniques are useful for library environment because:

- ➢ It increases books borrowing rate.
- ➢ Attract more users to borrow books.
- ➢ Assist library professionals in making policy on the application of duplicate copies and new publications.

10. Data Mining Elements:

- ✓ Transaction data in data warehouse system.
- ✓ Store the data in multidimensional database system.
- ✓ Provide the data in IT professionals.

✓ Analyze the data by software.
✓ Present the data in useful format.

11. Use or Implication of data mining in Participative Librarianship:

In the present days, library and information centres always deal with the digital data for this reason librarians want to use data mining for effective and instant retrieval of information. Exploring and analyzing data mining implies digging through heave of data to unfurl patterns and relationships contained within the organization including business activity. Data mining can be done manually by slicing the data until a right pattern becomes obvious. (Mishra,2007). To build a participative library, there are two large collections- one of the resources and another one of information about the resources. The first collection of digital content, the community repository, is built by the library and its users collaboratively. Data mining plays an important role in community repository and with the help of data mining the library and the users will be benefited for the future. The second collection, the enhanced catalog includes metadata both formal and user-created.

- Data mining is a practice of aggregating information about consumers' preferences and interests from a variety of sources including cookies, stealth data software, voluntary purchases, and mailing lists with the purpose of creating comprehensive profiles.(Mishra, 2007, p.7).

12. Bibliomining:

Use of data mining to examine library data records might be aptly termed bibliomining. With widespread adoption of computerized catalogs and search facilities over the past quarter century, library and information scientists have often used bibliometric methods (e.g. the discovery of patterns in authorship and citation within a field) to explore patterns in bibliographic information. During the same period, various researchers have developed and tested data mining techniques — advanced statistical and visualization methods to locate non-trivial patterns in large data sets. Bibliomining refers to the use of these techniques to plumb the enormous quantities of data generated by the typical automated library.

Bibliomining is used to discover patterns in what people are reading and researching and allows librarians target their community better. Bibliomining can also help library directors to focus their budgets on resources that will be utilized.

13. Factors leading towards data mining technique in libraries:

> **Growing Data Volume**

The main reason for necessity of automated computer systems for intelligent data analysis is the enormous volume of existing and newly appearing data that require processing.

> **Limitations of Human Analysis**

Two other problems that surface when human analysts process data are the inadequacy of the human brain when searching for complex multifactor dependencies in data, and the lack of objectiveness in such an analysis. A human expert is always a hostage of the previous experience of investigating other systems. Sometimes this helps, sometimes this hurts, but it is almost impossible to get rid of this fact.

> **Low Cost of Machine Learning**

One additional benefit of using automated data mining systems is that this process has a much lower cost than hiring an army of highly trained (and paid) professional statisticians. While data mining does not eliminate human participation in solving the task completely, it significantly simplifies the job and allows an analyst who is not a professional in statistics and programming to manage the process of extracting knowledge from data. (Prakash, Chand, & Gohel, 2004).

> **Information Explosion**

Information explosion is increase for the reason of invention of printing press and later usage of information technology. The rapid growth of information is a problem in the modern time. So the problem is to acquire the relevant information. For the requirement of better decisions regarding the relevant information data mining is very useful.

> **Global Business Environment:**

The application of information technology in the business organizations made the whole world as a global market. The concepts like outsourcing, e-marketing, e-business, e-commerce etc. are the results of the more competition in the business.

> **Availability of Global System Softwares:**

The availability of system software like Windows NT, Unix, Linux etc. are played a very vital role in the evolution of the data mining.

14. Objectives of Data Mining Technique in Participative Librarianship:

> **For improving the quality of data**

The main objective of data mining is use of quality data for supporting the decision-making process.

> **For providing the accurate data**

Data Mining is very useful for providing the accurate data in a limited period of time and it saves the time of the reader/user of the libraries in an effective way.

> **Providing the information in a scientific way**

Data mining provide the useful information from the data warehouse in a synthesized manner.
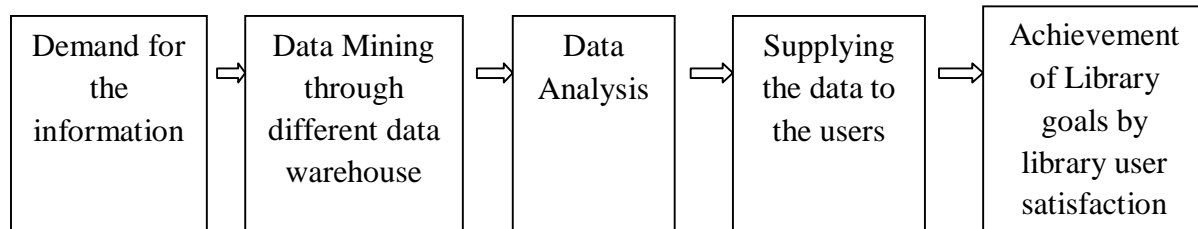
15. Operational Framework of Data Mining:

The Operational Framework highlight the concept of data mining and it provide how does the process of data mining depend on data warehousing for extracting the data for synthesization.
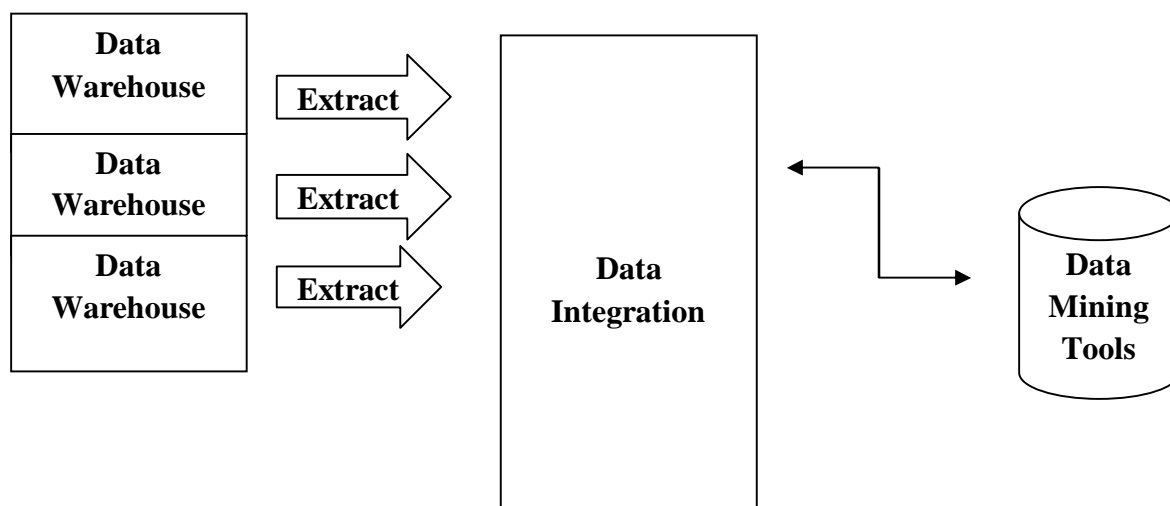
In this figure Data Warehouse is basically storehouse of knowledge where in which all the data should be loaded and this data extract and done an implementation of data integration after that the whole data should be distributed to the users by different data mining tools. In this framework Data Warehouse is playing a role of storehouse of data in the database and probably the most well-known implementation of data

integration is building an enterprise's data warehouse. The benefit of the data warehouse enables a business to perform analyses based on the data in the data warehouse.

16. Mechanism of Data Mining:

Demand for the information	⇒	Data Mining through different data warehouse	⇒	Data Analysis	⇒	Supplying the data to the users	⇒	Achievement of Library goals by library user satisfaction

To understand the concept of data mining in the participate librarianship the above figure wants to clarify that in the modern times the demand for information by library users is the first priority for the library professionals. Users demanding their approach by top-down or bottom-down way and it is very important for the library professionals to identify the selected problems of the library users. After the identification the professionals collect the selected information from the data warehouse. Then the selected information is analyzed and synthesize for making the data useful. After that the data should supply to the users. This is way in the modern scenario library professionals successfully present the information and achieve the library goals.

17. Data Mining Process:

- Data gathering, e.g., data warehousing, Web crawling.
- Data cleansing: eliminate errors and/or bogus data, e.g., patient fever = 125.
- Data Analysis: There are three components of data mining algorithms such as
    i)   Model representation: It is a language that describes the pattern.
    ii) Model evaluation: It is a pattern that meets the goal of KDD process.
    iii) Search: (a) Parameter search: It provides observation of data.
               (b) Model search: It is loop over the parameter search.
- Evaluation of results; not every discovered fact is useful, or even true! Judgment is necessary before following your software's conclusions.

Data Collection	→	Data Cleansing	→	Data Analysis	→	Evaluation of Result
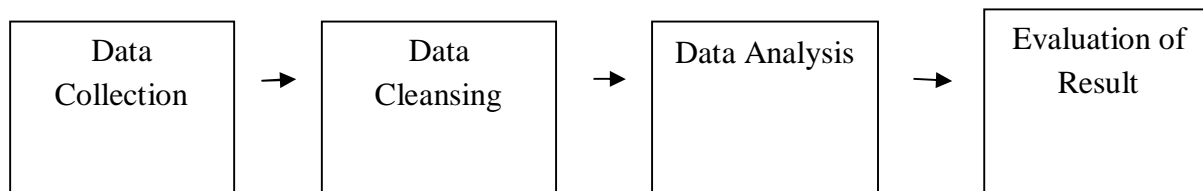
20. Future of Data Mining in Participative Library Working Environment:

In future Data Mining can provide the new road map for the next generation of library by applying it for the following activities of library.

- **Reference Service -** Since the data of the library continuously growing at an exponential rate and the main problem is how one can reference the required information form a 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. (Dwivedi &Bajpai, 2004).

- **Classification-** It will replace the manual classification of the library so that the classification task can be accomplished by less skilled person in a fast and efficient way. This will simplify the classification task of the library.
- **Acquisition-** By applying the data mining in the library data it can easily find out the required contents that are necessary to acquire next. This will reduce the work of library staff related to the acquisition as well as the efficient use of budget allocated to the library.(Dwivedi& Bajpai, 2004).
- **Clustering:** Data item is grouped according to logical relationship.
- **Summarization:** Compact description of data so that all documents are easily retrieved.

21. **Advantages of Data Mining in the Participative Library Environment because of the following reasons:**
- It is helpful to predict future trends.
- It signifies customer habits.
- Helps in decision making.
- It depends upon market based analysis.
- Quick fraud detection.

22. **Conclusion:**

It can be concluded that the data mining technique is very much useful in the library environment and it can simplify the working of library like acquisition, classification, circulation and as well as the reference service. The process of using library data more effectively begins by discovering ways to connect the disparate sources of data most libraries create. To connect the different sources in data warehouses can facilitate systematic exploration with different tools to discover behavioral patterns of the libraries primary constituencies. These patterns can help enhance the library experience for the user, can assist library management in making decisions and setting policies. So it is urgently needed that systematic efforts have been taken place to develop data mining techniques and algorithms for library database.

References:

1. *Prakash K, Chand, P, and Gohel U. (2004), Application of data mining in Library and Information Services, PLANNER.*

2. *Dwivedi, R.K., &Bajpai, R.P. (2004). Use of Data Mining in the field of Library and Information Science: An Overview. CALIBER, 512-519.*

3. *Mishra, R.N. (2007). Implication of data mining in Digital Library Environment.PLANNER.*

4. *Mukhopadhyay, Bikash.,&Mukhopadhyay, Sripate (2003). Application of Data Mining Techniques for Library Management. CALIBER.*

5.*Bansal, K.L., & Sood, S. (2011). Role of Datawarehousing and Data Mining in E-governance.International Journal of Artificial Intelligence and Knowledge Discovery, 1(1).*

6. *Prakash, B.R. (n.d.). Application of data mining in e-governance: a case study of BHOOMI project. Proceedings of the Second International Conference on Data Engineering and Management, 208-218.*

7. *(n.d.). Retrieved February 10, 2007, from http://en.m.wikipedia.org/wiki/Data_mining.*

8.*(2008,January10).Retrieved February 11, 2017, from http://yalsa.ala.org/blog/tag/participatory-librarianship/*