

Data Mining and Recommender System: A Review

¹Mukul M. Bhonde, ²Chandrashekhar H. Sawarkar, ³Dr. Pramod N. Mulkalwar

¹Assistant Professor, ²Assistant Professor, ³Associate Professor

¹Department of Computer Science, Shri Shivaji Science College, Amravati, Maharashtra, India

²Department of Computer Science, Assistant Professor, Smt. Narsamma Arts, Commerce and Science College, Amravati, Maharashtra, India

³Department of Computer Science, Amolakchand Mahavidyalaya, Yavatmal, Maharashtra, India.

Abstract: Due to the enhanced capabilities to generate and collect data from varied sources, a tremendous amount of data has flooded every part of our lives. This explosion in stored data has created necessity of new techniques and tools for filtering such data into meaningful information known as data mining, also be referred as knowledge discovery from data (KDD). In terms of the scalability, Web is growing exponentially and obvious increase in redundancy of information as well. Various forms of data in unstructured, semi-structured and structured form is augmented to Web every minute. Due to this scattered and distributed nature of Web it is very challenging to surf the Web using alone search engines and plain browsers. Recommender systems (RS) are a type of information filtering system that seek to predict the 'rating' or 'preference' that user could give to an item under consideration. Recommender system is defined as a decision making strategy for users under complex information environments. Recommender systems have become prominent issue of research in recent years, and are being used for variety of web domains. All Recommender Systems (RS) apply techniques and methodologies of Data Mining (DM) for information extraction such as Similarity measures, Sampling, Dimensionality Reduction, Classification, Association-Rule- Mining (ARM) and Clustering. Recommender Systems (RS) typically apply techniques and methodologies from other neighboring areas such as Human Computer Interaction (HCI) or Information Retrieval (IR).

Keywords: Data mining(DM), Knowledge discovery from data (KDD), Information Retrieval(IR), Web mining, Recommender System(RS).

I. INTRODUCTION

Today, we are living in the data age. We are amongst the networks flooded with terabytes or petabytes of data. Vast amount of data collected daily regarding variety of aspects. To analyze such data is an important need. This necessity has led to the birth of data mining. Data mining has an important place in today's world. It becomes an important research area as there is a huge amount of data available in most of the applications. This huge amount of data must be processed in order to extract useful information and knowledge, since they are not explicit. Data Mining is the process of discovering interesting knowledge from large amount of data. Data mining can be viewed as a result of the evolution of information technology. Many people treat data mining as a term knowledge discovery from data(KDD) while others view it merely an essential step in the process of knowledge discovery. This KDD process is an interactive sequence of the steps: data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation and knowledge representation. Early four steps are different forms of data preprocessing where data are prepared for mining. So we can view it as process for discovering an interesting patterns and knowledge from large amount of data. Data mining functionalities involves characterization and discrimination, associations and correlations, classification and regression, clustering analysis. Data mining is a dynamic and fast growing field with great strengths touched each area of every field. Data mining is a process of extraction of useful information and patterns from huge data. It is also called as knowledge discovery process, knowledge mining from data, knowledge extraction or data /pattern analysis.

Evaluation & Presentation

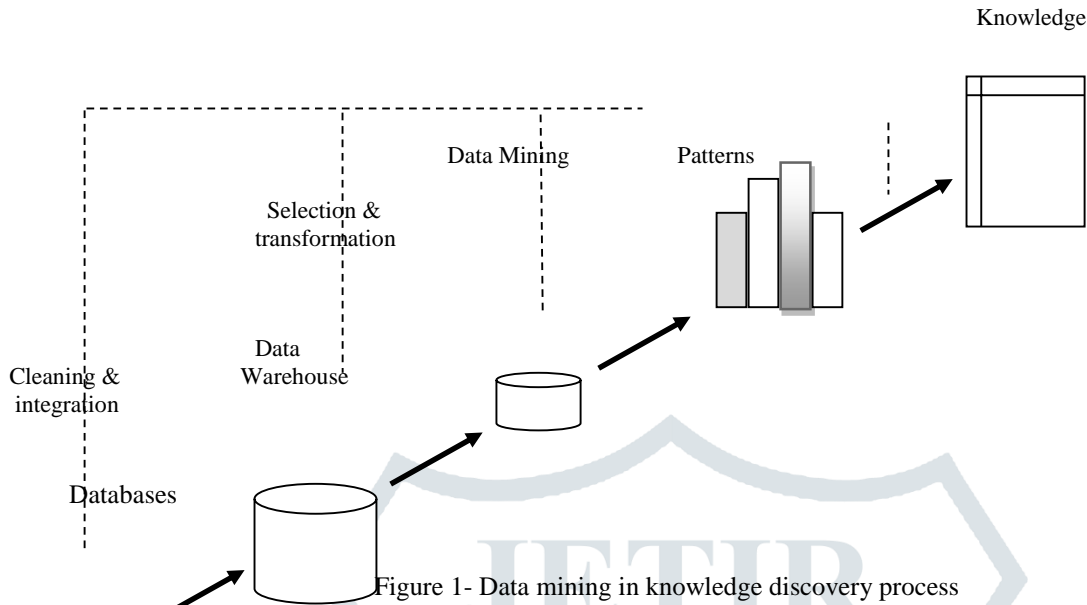


Figure 1- Data mining in knowledge discovery process

II. WEB MINING

With the explosion of the data over the Internet, WWW has become a powerful platform to store, disseminate and retrieve information to find useful knowledge by data mining. Due to the properties of the huge, diverse, dynamic and unstructured nature of Web data, Web data research has encountered a lot of challenges, such as scalability, multimedia and temporal issues etc. Web users faces the major issues with this information overload. Typical problems mentioned in Web related research and applications are Finding relevant information, finding needed information, Learning useful knowledge, Recommendation/personalization of information. Web Mining can broadly be divided into three categories according to the kinds of data to be mined:

- Web Content Mining
- Web Structure Mining
- Web Usage Mining

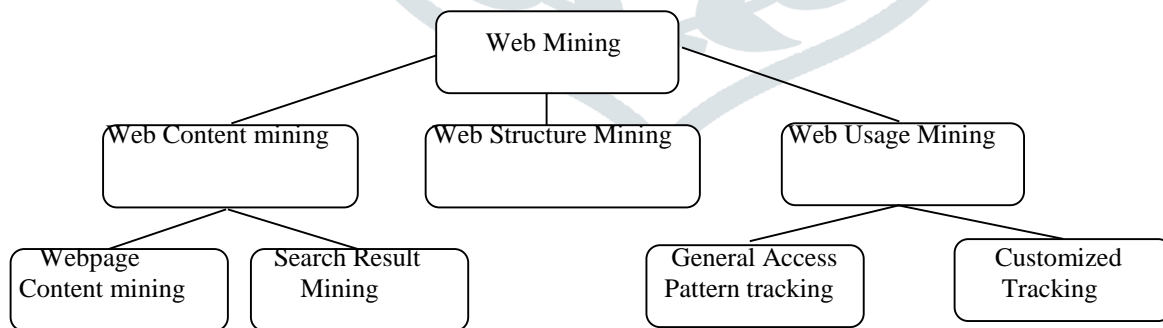


Figure 2- Taxonomy of Web Mining

- Web content mining is the task of extracting knowledge from the content of documents on World Wide Web like mining the content of html files. Web document text mining, resource discovery based on concepts indexing or agent-based technology fall in this category.
- Web structure mining is the process of extracting knowledge from the link structure of the World Wide Web.
- Web usage mining, also known as Web Log Mining, is the process of discovering interesting patterns from web access logs on servers. The Taxonomy of Web Mining is given in Figure 2.

III. INFORMATION RETRIEVAL (IR)

Over the many years people have realized the importance of archiving and finding information. Information retrieval (IR) deals with the representations, storage, organization of, and access to information items. IR consists mainly of determining which documents of a collection contain the keywords in the user query. Information retrieval plays a vital role in web search engines to access most relevant information according to the user's input query. It is a mainstream and the basics of web search engines.

IV. RECOMMENDER SYSTEM (RS)

Recommender systems are information filtering systems that deal with the problem of information overload by filtering vital information fragment out of large amount of dynamically generated information according to user's preferences, interest, or observed behavior about item. Recommender system has the ability to predict whether a particular user would prefer an item or not based on the user's profile.

Recommender systems are beneficial to both service providers and users. They reduce transaction costs of finding and selecting items in an online shopping environment. Recommendation systems have also proved to improve decision making process and quality. In e-commerce setting, recommender systems enhance revenues, for the fact that they are effective means of selling more products. In scientific libraries, recommender systems support users by allowing them to move beyond catalog searches. Therefore, the need to use efficient and accurate recommendation techniques within a system that will provide relevant and dependable recommendations for users cannot be over-emphasized.

Web recommender systems personalizes web search according to user information need using users' previous interaction with related and similar items. The advantage of the recommender systems is that it overcomes information overload problems on huge size web during web information retrieval. It helps users to find expected information by filtering items relevant to the web user. The filtering techniques used in recommender systems are mainly classified as: Content based, Collaborative based, hybrid and social network based approach as shown in figure 3.

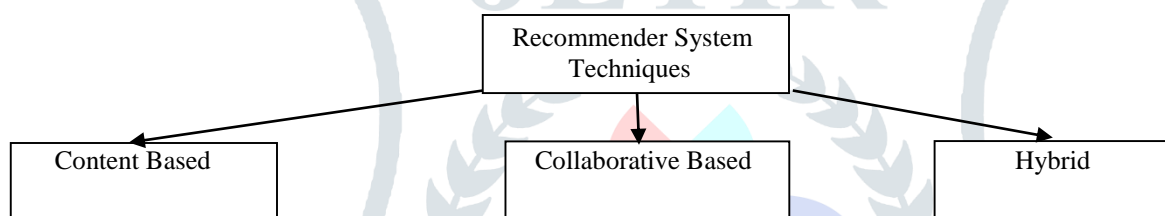


Figure 3- Anatomy of different recommendation filtering techniques.

4.1 Content based filtering

Content-based technique is a domain-dependent algorithm and it emphasizes more on the analysis of the attributes of items in order to generate predictions. When documents such as web pages, publications and news are to be recommended, content-based filtering technique is the most successful. In content-based filtering technique, recommendation is made based on the user profiles using features extracted from the content of the items the user has evaluated in the past.

4.2 Collaborative filtering

Collaborative filtering is a domain-independent prediction technique for content that cannot easily and adequately be described by metadata such as movies and music. Collaborative filtering technique works by building a database (user-item matrix) of preferences for items by users. It then matches users with relevant interest and preferences by calculating similarities between their profiles to make recommendations. Such users build a group called neighborhood. A user gets recommendations to those items that he has not rated before but that were already positively rated by users in his neighborhood. Recommendations that are produced by CF can be of either prediction or recommendation. Prediction is a numerical value, R_{ij} , expressing the predicted score of item j for the user i , while Recommendation is a list of top N items that the user will like the most as shown in [Fig. 4](#). The technique of collaborative filtering can be divided into two categories: memory-based and model-based.

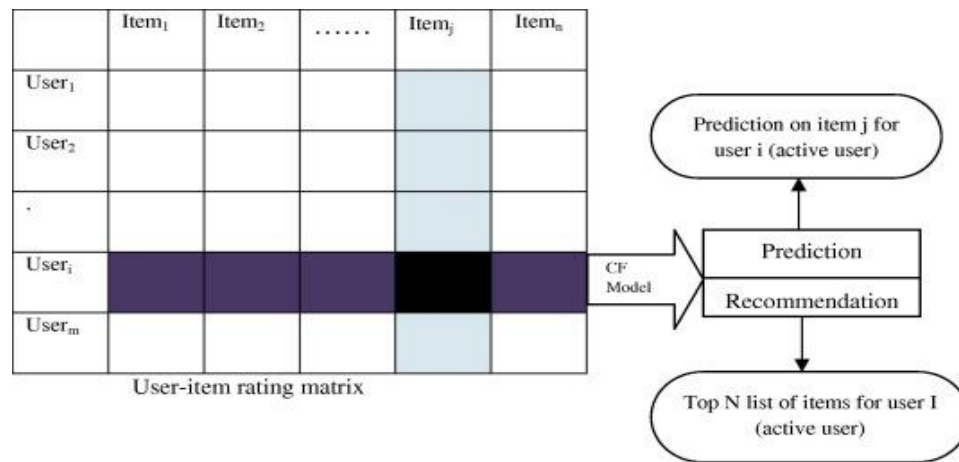


Figure 4 - Collaborative Filtering Process

4.3 Hybrid filtering

Hybrid filtering technique combines different recommendation techniques in order to gain better system optimization to avoid some limitations and problems of pure recommendation systems. The idea behind hybrid techniques is that a combination of algorithms will provide more accurate and effective recommendations than a single algorithm as the disadvantages of one algorithm can be overcome by another algorithm. Using multiple recommendation techniques can suppress the weaknesses of an individual technique in a combined model. The combination of approaches can be done in any of the following ways: separate implementation of algorithms and combining the result, utilizing some content-based filtering in collaborative approach, utilizing some collaborative filtering in content-based approach, creating a unified recommendation system that brings together both approaches.

V. APPLICATION OF WEB RECOMMENDER SYSTEMS

Today, the recommender systems are being used in various domains such as e-governance, e-business, e-commerce, e-library, e-learning, e-tourism, e-resource services, e-banking, etc. In e-governance for (G2C) Government to citizen model, ICT is used for strengthening the relation between governing authorities and citizens by providing cost effective e-services efficiently. E-business recommender systems generate information related to product and services for business users. In e-commerce, the online recommendations are generated to guide the users for purchase of products. Recommender systems in e-library help the users to locate intended information and knowledge sources effectively. E-learning recommender systems help users to select the specific course and learning material. E-tourism recommender systems provide the information to tourist about various possible destinations with cost effective accommodations. E-resource service recommender systems allow users to share the video, documents, audio, images on the web to share users with similar liking or interests.

VI. CONCLUSION

Recommender systems open new opportunities of retrieving personalized information on the Internet. Recommender systems are used widely to cater to users' information need in various domains and it also helps to recover the problem of information overload which is a very common phenomenon with information retrieval systems. It enables users to have access to products and services which are not readily available to users on the system. Basic recommendation techniques used on web are Content based, Collaborative, Social network and Hybrid. The recommender systems have been applied in various domains like e-learning, e-commerce, e-shopping, e-tourism, etc. The developer of a RS for a certain application domain should understand the specific facets of the domain, its requirements, application challenges and limitations. Only after analyzing these factors one could be able to select the optimal recommender algorithm and to design an effective human-computer interaction.

References

- 1) Introduction to Recommender Systems Handbook Springer by Francesco Ricci, Lior Rokach
- 2) and Bracha Shapira
- 3) J.A. Konstan, J. Riedl Recommender systems: from algorithms to user experience
User Model User-Adapt Interact, 22 (2012), pp. 101-123
- 4) C. Pan, W. Li Research paper recommendation with topic analysis
In Computer Design and Applications IEEE, 4 (2010)pp. V4-264
- 5) Pu P, Chen L, Hu R. A user-centric Evaluation framework for recommender systems. In: Proceedings of the fifth ACM conference on Recommender Systems (RecSys'11), ACM, New York, NY, USA; 2011. p. 57-164.
- 6) Hu R, Pu P. Potential acceptance issues of personality-ASED recommender systems. In: Proceedings of ACM conference on recommender systems (RecSys'09), New York City, NY, USA; October 2009. p. 22-5.

- 7) B. Pathak, R. Garfinkel, R. Gopal, R. Venkatesan, F. Yin Empirical analysis of the impact of recommender systems on sales J Manage Inform Syst, 27 (2) (2010), pp. 159-188
- 8) J. Srivastava, P. Desikan and V. Kumar (2002) "Web Mining: Accomplishments & Future Directions", National Science Foundation Workshop on Next Generation Data Mining (NGDM'02)
- 9) R. Burke Hybrid recommender systems: survey and experiments User Model User-adapted Interact, 12 (4) (2002), pp. 331-370
- 10) J. Bobadilla, F. Ortega, A. Hernando, A. Gutiérrez Recommender systems survey Knowl-Based Syst, 46 (2013), pp. 109-132
- 11) Breese J, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering. In: Proceedings of the 14th conference on uncertainty in artificial intelligence (UAI-98); 1998. p. 43–52.
- 12) G. Adomavicius, J. Zhang Impact of data characteristics on recommender systems performance ACM Trans Manage Inform Syst, 3 (1) (2012) Stern DH, Herbrich R, Graepel T. Matchbox: large scale online bayesian recommendations. In: Proceedings of the 18th international conference on World Wide Web. ACM, New York, NY, USA; 2009. p. 111–20.
- 13) J.B. Schafer, D. Frankowski, J. Herlocker, S. Sen Collaborative filtering recommender systems P. Brusilovsky, A. Kobsa, W. Nejdl (Eds.), The Adaptive Web, LNCS 4321, Springer, Berlin Heidelberg (Germany) (2007), pp. 291-324, [10.1007/978-3-540-72079-9_9](https://doi.org/10.1007/978-3-540-72079-9_9)

