Analysis of Web-Mining Techniques for the Recommendation of Useful Information Using Web Personalization

***Sangeeta Bhandari**
****Dr. Sanjeev Kumar Sharma**

ABSTRACT

Millions of users are interacting daily with the websites and visiting lots of websites, leaving back a variety of information .This information must be used by the websites administrator to improve their website according to the users of the websites this is known as web usage mining.

Web personalization is a relatively new and challenging field for web content delivery. In order to meet expectations of visitors, customers and loyal users, web world is struggling to offer excellent customized services during their interaction with the system. The impact of personalization and recommendation system can be experienced by the rapid popularity that this area has gained in the last few years. Customers preferably choose to visit those websites, which understand their needs, provide them rapid value added customized services and easy access to required information in simple understandable format. Web personalization and recommendation system plays a major role in meeting this goal. In this paper, we study various methods those are used for developing recommendation systems. Also we analyses various methods that use in web usage mining and Web Personalization techniques. A short survey of various Web personalization techniques are study in this paper.

Keywords: Web Personalization, Apriori Algorithm, Fp-Growth Algorithm.

*Sangeeta Bhandari, Oriental Institute of Science and Technology, Bhopal, (M.P.), bhandari1403@gmail.com
**Dr. Sanjeev Kumar Sharma, Professor CSE, Oriental Institute of Science and Technology, Bhopal, (M.P.), sanjeevsharma@oriental.ac.in

I. INTRODUCTION:

Web mining is the application of data mining to mine knowledge from web data. Figure 1.1 shows the classification of web mining. There are following types of web mining:

a) Web Content Mining: as web Content Mining is that the method of extracting useful data from the contents of web documents. Content knowledge corresponds to the gathering of facts a web page was designed to convey to the users. it should contains text, images, audio, video, or structured records like lists and tables. Analysis activities during this field additionally involve exploitation techniques from different disciplines like information Retrieval (IR) and natural language processing (NLP).

b) Web Structure mining: The structure of a typical Web graph consists of web pages as nodes, and hyperlinks as edges connecting between two connected pages. Additionally, the content among a web page also can be organized in an exceedingly tree structured format, supported the various HTML and XML tags among the page. Thus, web Structure

Mining are often regarded as the method of discovering structure data from the web. This sort of mining are often performed either at the (intra-page) document level or at the (inter-page) hyperlink level.

c) Web Usage Mining: web Usage Mining is that the application of data mining techniques to get fascinating usage patterns from Web data, so as to know and better serves the wants of web based mostly applications. Usage information captures the identity or origin of web users alongside their browsing behavior at a web site. a number of the standard usage data collected at a web site includes IP addresses, page references and access time of the users.

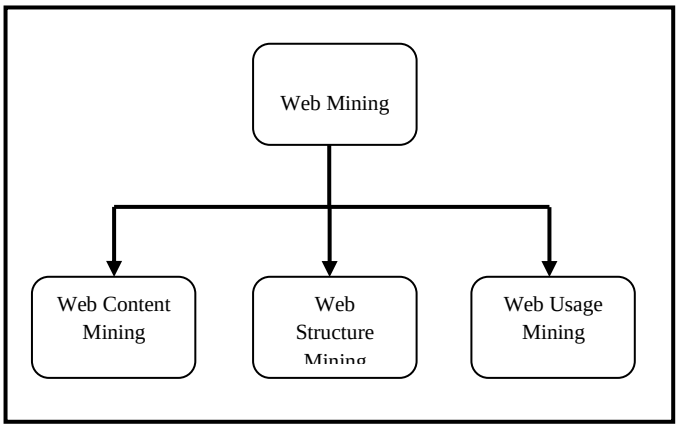

Figure 1.1: Classification of Web Mining

1.1. Brief about Web Usage Mining

When data mining techniques are applied on web usage data so as to extract helpful information relating to user behavior, it's referred to as web usage mining . It's an approach for grouping and preprocessing web usage information, so constructing models that represent the behavior and interests of users. Such models will automatically be employed by personalization system for predicting user's personal interests and so enhance his surfing experience with the web site. The web usage mining involves assortment of knowledge from numerous sources, pre-processing of collected data, discovering helpful information, and at last post process the information. The figure 1.2 provides overview of web usage mining process.

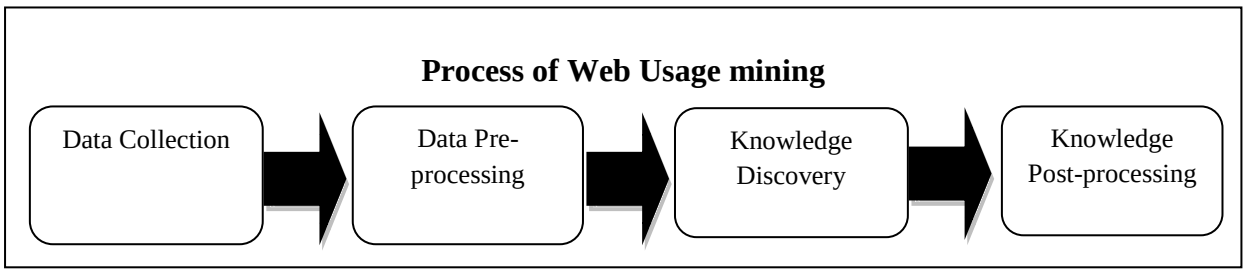

Figure 1.2: Overview of web usage mining process

Web Mining is a variation of this field that distils untapped source of abundantly available free textual information. The importance of web mining is growing along with the massive volumes of data generated in web day-to-day life. In general, web data always arrives in a multiple, continuous, rapid and time varying flow.

1.2. Web Personalization

The objective of a web personalization system is to "provide users with the information they require or needed, without expecting from them to provoke it explicitly". Personalization needs implicitly or explicitly aggregation user data. A personalization mechanism is based on explicit preference declarations by the user and on an iterative method of observance the user navigation, aggregation its requests of ontological objects and storing them in its profile so as to deliver personalized content.

- The preprocessing of web data such as content data, usage data, user profile data and structure data.
- Extracting statistical information and discovering interesting usage patterns using various data mining techniques.
- The actions to be carried out, recommended by such personalization systems.

II. RECOMMENDATION SYSTEM

Recommender system is something that every website or app that provides a reliable interface of user must have. It prevents user from wasting time in segregating what he needs and encourages efficient exploring. Various factors helps to acquire the data of user interest and utilize them in building a system which recommends them on the basis of profile of their activities and interest created. Recommender system helps to reduce the search time and prevents repetitive searching. This system helps to overcome the problem of management and manipulation of large amount of data that help the user to access the options available.

2.1. Techniques used in Recommendation Systems

Recommender systems are usually divided into 3 categories:

2.1.1. Content based recommendation

In content based method the profiling of each user is done explicitly through questionnaires, or implicitly-observed from the user's preferences. CASPER uses content based recommendation system.

2.1.2. Collaborative based recommender system

This system recommends items to the particular user on the basis of the items rated by the other users. The first recommender system which was developed was Grundy system using stereotype which build separate user models and help them to recommend relevant books to each other. Amazon uses Collaborative based recommendation system.

2.1.3. Hybrid based recommendation system

This system comprises of both content and collaborative based system for the purpose of avoiding limitations of both systems. These two systems can be combined to form hybrid by following methods:

I. Separately implementing both collaborative and content based methods and then combining their results.
II. Incorporating some content-based characteristics into a Collaborative approach
III. Incorporating some collaborative characteristics into a content-based approach
IV. Constructing a general unified model

Pandora uses Hybrid based recommendation system.

III. LITERATURE SURVEY

This paper [1] authors had given an outline of web personalization ways. Authors had conjointly explained the potential ways that of exploitation semantics into the web personalization method. numerous theoretical approaches of personalization are referred as base models for linguistics approaches to personalization and are evaluated from prospect of usability. Several present personalization models that use semantics are represented and also the advancement they'd provided is been represented by authors considering personalization accuracy. a brand new approach for personalization within the existing system is planned i.e. "Sweb" is bestowed. As per authors the planned approach combines social networking knowledge with semantic reasoning based mostly personalization technique to extend the accuracy of recommendations system.

In line with the authors of the paper [2] web personalization recommendation systems analyze the users behavior supported the log files (usage/access history of previous users) and website Meta information. As per authors there's plenty of scope of improvement within the approach of recommender systems. during this paper, a brand new web page recommendation framework planned. First, users' navigational patterns are automatically extracted from web usage information exploitation weighted association rule mining, these navigational patterns are then accustomed generate recommendation candidates supported a user's current session. By exploitation document clusters, recommendation seed is chosen, finally, by using the HITS algorithmic program to increase the recommendation seed, recommendations have generated. Our

experimental results illustrate that exploitation this hybrid algorithmic programs during a web recommender system has the potential to improve the standard of the system and may generate higher quality recommendations than exploitation either the hybrid KNN-based recommendation approach or the Weighted Association Rule mining recommendation algorithm alone.

As per Authors of this paper [3] web personalization is that the customization of data or services provided by the web website in keeping with user wants, through the data of the navigational habits of users and individual interest, combined with the content and structure of web sites. Objective of web personalization is to produce data required by user, while not having to raise the user expressly. I n the paper authors had analyzed and enforced new approach of web personalization on web site of Japan that sales used cars. During this approach system recommend ten most relevant unseen things. These ten most relevant unseen things are derived by group action history of the signed-in user so applying K-nearest neighbour algorithmic program to it.

In this paper [4] authors had explained various data mining techniques, Web Personalization techniques. Authors had also thrown light on the problems associated with the traditional approaches of web personalization. Authors had also proposed a new approach in which they firstly they track the count of a particular link visit (foot print) then they used meta information along with foot prints. They also keep track of history and user profile based on which they provide personalized results. Along with this authors had also applied client side customization, page gather algorithm and Time spent reading as measure to implicit rating.

In this paper [5] authors explains web personalization as to make a Web site additional responsive to the exclusive and specific desires or requirements of each individual customer or set of customers. In this paper authors had provided a detailed analysis of various approaches in web personalization based on different domains with their advantages and limitations. A brief comparison has been made between the different web personalization techniques based on certain parameters. The classification made by authors is given below -

- Category based web personalization
- Multi level web personalization
- Fuzzy logic based product filtering for web personalization
- Web personalization based on user's interested domains
- Web page personalization based on weighted association rules
- Web mining based on user access patterns for web personalization
- Web personalization method based on relevance feedback on keyword space

- Application of neural network and kano's method to content recommendation in web personalization
- Web content recommender system based on consumer behavior modeling
- Techniques for adaptive website and web personalization without any user effort

According to the authors of the paper [6] the user behavior and the browsing pattern changing rapidly making it difficult & challenging task to make prediction's based on the occurrence patterns from the analysis of the device web logs, in a way ensuring that the patterns are not obsolete. In the paper authors had proposed a new approach which uses device intelligent FW, device soft sensors, WoT Sensors and browsing/web-apps usage logs. By using longitudinal context data collected from real users, authors revealed that web miners efficiently generate patterns using limited phone resources.

In this paper [7] authors have suggested a mechanism to help buyers on ecommerce website to take their purchase decision. In this approach authors had compared products on which personalized reviews summary was mentioned and the products on which no personalized reviews summary was mentioned. As per their experiment the products containing personalized reviews clears the confusion of customer more easily as compared to the products which doesn't contains the personalized reviews. Authors explain that consumers seek lesser information about the product in an e-commerce environment with personalized review summary.

In this article [8] authors had described recommendation that can be used for e-commerce companies like Amazon.com. As per the authors a good recommendation system can provide customers with the most relevant products, help users discover items they might not have found by themselves and promote sales to potential customers. Based on previous work authors designed three new recommendation systems they are mentioned below-

- Item Similarity
- Bipartite Projection
- Spanning Tree

As per authors experimental analysis Item similarity algorithm takes a user's previous ratings into account, together with the similarities between the products. We have to decide the rating for user who has no other reviews solely based on the rates received by similar products. Despite it's the slowest among these algorithms, it gives a better score than the other algorithms.

As per authors experimental analysis the spanning tree algorithm predicts rating which is mainly decided by other people's review on the product and similar products. However, this algorithm is

not personalized enough. For products with ratings of 1 and 2, it often predicts higher score because most people in the dataset gives higher ratings to products.

As per authors experimental analysis the bipartite projection algorithm does not work so well in this case. Authors notice that for a portion of the users this algorithm cannot predict the rating. They believe it's because the users or products don't have enough review data.

As per Authors Item Similarity has the best result, then followed by Spinning Tree, and Bipartite Projection is the worst.

In this article [9] authors have provided details concerning numerous recommendation systems. Recommender systems are some things that each web site or app that gives a reliable interface of user should have. It prevents user from delay in segregating what he desires and encourages economical exploring.

Recommender system helps to scale back the search time and prevents repetitive looking out. This technique facilitates to beat the matter of management and manipulation of huge quantity of knowledge that help the user to access the choices accessible. As per authors recommender systems are sometimes divided into three categories:

1. Content based recommendation
2. Collaborative based recommender system
3. Hybrid based recommendation system

Authors had additionally provided the main points of phase of recommendation

1.1 Offline phase

2.1 Online phase

As per authors each the offline and on-line phases of a recommender system helps to segregate and collaborate information at a differential level that allows the user access additional correct data because the pattern tracing is practiced.

In this paper [10] authors specific their views over growing information on World Wide web. that they had additionally raised a priority relating to data overload challenges for the users. They specific their concern as retrieving the foremost relevant data from the net as per the user demand has become arduous as a result of the big assortment of heterogeneous documents. They recommended web Personalization as an answer to the matter. Consistent with the authors internet personalization is method that adjusts information/services delivered by an internet to the requirements of every user or cluster of users, taking their behavioral patterns. Frequent successive Patterns (FSPs) that are extracted from web Usage data (WUD) are vital for analyzing and understanding users' behavior to enhance the standard of services offered by the World Wide

Web (WWW). User behavioral patterns are needed to make profiles of every user, victimization that Personalization of web site is created.

In their study authors had explored completely different FSP mining algorithms, namely, WAP-tree, Prefix Span and SPADE for extraction of FSPs from WUD of the aforementioned educational site for a amount variable from weekly to quarterly. The experimental outcomes of authors study indicate that Prefix Span FSP mining algorithmic program will higher than WAP-tree and SPADE algorithms for minimum support of two.

Table 1: Recommender Systems With Different Domains [11]

Domain	Risk	Churn	Heterogeneous	Preferences	Interaction Style	Scrutiny	Examples	Recommendation Technology
E-commerce	Low	High	High	Stable	Implicit	Not required	Amazon.com, eBay	Collaborative-Filtering
Financial-services and Life-inslrrance	High	Low	Low	Stable	Explicit	Required	Koba4MS, FSAdvisor	Knowledge-Based
Job search Recruiting	High	Low	Low	Stable	Explicit	Required	CASPER	Conent-based
Movie	Low	Low	Low	Stable	Implicit	Not required	Netflix INTIMATE, Movies2Go	Collaborative and Content Filtering
Music	Low	Low	Low	Stable	Implicit	Not required	Pandora and	Content-based, Hybrid
News	Low	High	Low	Stable	Implicit	Not required	Yahoo news, ACR news and Google news, INFOrmer, NewsDude	Content-based, Collaborative-Filtering
Real Estate.	High	Low	Low	Stable	Explicit	Required	RentMe, FlatFinder	Knowledge.-based
Scientific Research papers	Low	Low	Low	Stable	Explicit/implicit	Not required	QuickStep system, Citeseer	Content-based
Software Engineering	Low	Low	Low	Stable	Explicit/Implicit	Required	[11]	Hybrid and Content- based
Tourism	High	Low	Low	Unstable	Explicit	Required	Travel Recommender	Content-based, Koowledge-based
TV Program	Low	Low	Low	Unstable	Implicit/Explicit	Not required	AVTAR	Content-based
Web Page Recommender	Low	High	High	Unsrable	Implicit	Not required	Letzia	Collaborative-Filtering, Hybrid

IV. CONCLUSION

In this paper, we have analyzed numerous methods which are used for building recommendation systems. Also we had analyzed various methods that use in web usage mining, Web Personalization techniques A short survey of various Web personalization techniques has been given. The previous approaches increases efficiency but the overhead costs more in proportion to the efficiency improvement. The analysis suggests a need for new approach that is practically implementable and provides efficiency improvements over previous approaches.

REFERENCES

1. Dario Vuljani, Lidia Rovan, Mirta Baranovi, "Semantically Enhanced Web Personalization Approaches and Techniques", Proceedings of the ITI 2010 32nd Int. Conf. on Information Technology Interfaces, June 21-24, 2010, Cavtat, Croatia.
2. Samira Khonsha, Mohammad Hadi Sadreddini, "New Hybrid Web Personalization Framework", 978-1-61284-486-2/111$26.00 ©2011 IEEE.
3. C.S. Oemarjadi, N.U. Maulidevi, "Web Personalization in Used Cars Ecommerce Site", 2011 International Conference on Electrical Engineering and Informatics, 17-19 July 2011, Bandung, Indonesia.
4. Kanika Arora, Kamal Kant, "Techniques for Adaptive websites and Web Personalization without any user effort", 2012 IEEE Students' Conference on Electrical, Electronics and Computer Science, 978-1-4673-1515-9/12/$31.00 ©2012 IEEE.
5. Mohit Rajput, Rohit Agrawal, Dilip kumar Sharma "A STUDY AND COMPARATIVE ANALYSIS OF WEB PERSONALIZATION TECHNIQUES", Proceedings of 2013 IEEE Conference on Information and Communication Technologies (ICT 2013), 978-1-4673-5758-6/13/$31.00 © 2013 IEEE.
6. Priyesh Lakar, Suyarnbulingarn Rathinasamy Muthupandi, Siba Prasad Sarnal, Niranjan B Patil "Smart Web Miner - Extending Web Browser with Mining framework based on user behavior & Web-of-Thing patterns for web personaliztion", 978-1-4673-7910-6/15/$31.00 ©20 15 IEEE.
7. U. Mahesh Balan, S. K. Mathew, "Impact of Web Personalization of Online Word of Mouth on Buyers' Decision Process: An Experimental Study", 978-1-4673-7910-6/15/$31.00 ©20 15 IEEE.
8. Jianfeng Hu, Bo Zhang, "Product Recommendation System", CS224W Project Report, December 10, 2012.
9. Tanya Bhattacharya, Arunima Jaiswal, Vaibhav Nagpal, "Web Usage Mining and Text Mining in the Environment of Web Personalization for Ontology Development of Recommender Systems", 978-1-5090-1489-7/16/$31.00 ©2016 IEEE.
10. Doddegowda B J, G T Raju, Sunil Kumar S Manvi, "Extraction of Behavioral Patterns from Pre-processed Web Usage Data for Web Personalization", IEEE International Conference On Recent Trends In Electronics Information Communication Technology, May 20-21, 2016, India, 978-1-5090-0774-5/16/$31.00 © 2016 IEEE.
11. Sanjeev Kumar Sharma, Dr. Ugrasen Suman, "Design and Implementation of Architectural Framework of Recommender System for e-Commerce", IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS), Vol. 1, No. 2, December 2011.