WEB PERSONALIZATION USING WEB USAGE MINING TECHNIQUES

Anupama Prasanth
Research Scholar, Karpagam University, Coimbatore
anupama.prasanth@gmail.com

Abstract— The exponential growth in the size of data in information resources, its complicated structure and the diversity of user groups using it are increase the complexity of web usage. The heavy use of WWW as an information vault gives a lot of information Web log data. It has an inevitable importance in web based environment. If we utilize these web log data intelligently, this will become one of the essential resources to portray user access behavior. These user characteristics are in the form of hierarchical structure of related information. This information can be used for data mining tasks for user behavioral pattern analysis and therefore tailor the web page contents according to the user preferences.

Web mining is the best device which can keenly help with exchanging the gigantic gathering of information into valuable data and learning. Web usage mining is one of the real sub-regions of web mining, which is a use of data mining systems to find usage designs from web information. Some of the commonly used technologies in web mining are user access pattern analysis, clustering, classification and information filtering.

This paper presents a study based on web personalization approach using the concept of web usage mining. In the beginning it focus on web personalization strategies and then focus on one of the most promising strategy of automatic web personalization and its various advances.

Index Terms—Association Mining, Clustering, Recommendation System, Web Personalization, Web Personalization Strategies, Web Usage Mining.

I. INTRODUCTION

World Wide Web, the largest data base, is growing in unsystematic way. The web pages are linked each other, but are not logically organized. Same time probably millions of web pages are added to www and also undergo changes daily. This leads information overloading. So in this scenario, finding desired information or particular details is tiresome. Therefore we need a very efficient and effective technique to extract or access the required information. The main challenge behind is the extraction of desired information with minimum attempt and time. Another major issue is the relevancy of the extracted information.

Proper organization and management of data improve the retrieval effectiveness and efficiency. Web Mining or application of Data Mining techniques on web data is the widely accepted method for this. Web mining consists of several techniques like, Web Content Mining, Web Structure Mining, and Web Usage Mining, which helps in filtering valuable learning from web information.

To satisfy the requirements of web surfers one of the most adopted functional techniques in web mining is to analyze the user browsing patterns, web usage mining, and making recommendations, called Web Personalization[1]. Personalizing web pages is one of the efficient parts of web-mining, which enables to understand user interests to offer services and enables them to discover web pages, text documents, multimedia files, images and other types of resources from web according to

their choice [2]. Research on web mining has contributed a lot in web personalization.

In the beginning stage of Web usage mining, it was not considered for personalization. It was used only for extracting detailed knowledge to make decisions. However later research in this area leads to the innovative idea of using Web usage mining for Web personalization, ie. Change the attention from the traditional, decision-support knowledge discovery to the discovery of operational knowledge for personalization. This sort of learning can be straightforwardly conveyed back to the users so as to enhance their involvement in the site, without the intercession of any human master. Along these lines, it is presently broadly perceived that usage mining is an important source of ideas and answers for Web personalization.

Web Personalization can be used in a variety of streams, vary from enhancing the loyalty of customers on ecommerce sites [5] to facilitating searching efficiency [6]. The major goals [7] of web personalization can be:

- Converting visitors into buyers
- Improve customer loyalty and retention,
- Enhance design and usability of web sites,
- Improve search efficiency

The web page personalization based on web mining technique has two major phases, offline and online. The mission of the offline phase is to model user's navigation pattern using the history, web log data. This model is used in online phase to predict a user navigation pattern and give recommendations to them based on prediction. The aim of this is to recommend a series of links, texts, products etc to the user according to the interest and preferences.

There are a lot of web mining techniques but generating patterns from historic or related data is the most common focus of almost all techniques. This leads for a great demand on more flexible algorithms for various time critical and data intensive web applications. The main motive of this study is the need of some technique which focuses not fully on history but considering some other parameters.

II. Web Personalization

Web personalization is a promotion tool, especially for expansive data administrations or locales that creates benefits by offering items. Web personalization provides to use the user favorites dynamically to customize web content according to particular user. So with the effect of web personalization it is implicit that the user will be provided with all information according to their expectation without anticipating from them to ask for it explicitly [1]. The personalized mechanism of explicit preference declarations is supported by an iterative procedure of checking the user access path, gathering its solicitations of ontological objects and putting away them in its profile [2].

Web personalization can be done through four stages. The first stage is for collecting explicit and implicit user information. Implicit data is the past user activity data which automatically stored in web server logs without their knowledge, at the same time the information which user register while accessing a site like username, password, and other registration questionnaires forms the explicit data. So it can be considered as the recording of visitor interest furthermore, conducts while they peruse a site. But the data which put away in the web server logs are as content file and have missing and inconsistencies, so have to clean before using for analysis purposes. This preprocessing stage will filter out all irrelevant data from analysis. The role of mining algorithm starts from this stage, appropriate mining strategies are connected to the prepared information and find user behavior pattern[12]. The usage pattern can be semantically categorized to make easier information retrieval. This stride is connected disconnected from the net for programmed client profiling without adding the weight to the web server. At last use this information for web page recommendation.

III. Web Personalization Through Various Strategies

The move from physical to virtual space leads web personalization as a necessity than an option. It retains the loyalty of customers by performing variety of functions. These functions fluctuate from straightforward routine activities

to more intricate ones which provide easy interaction with web sites. They can be separated into four techniques requested as underneath from the least difficult like client welcome to the most progressive, for example, customized content conveyance.

A. Memorization

Memorization is the simplest form of web personalization. Storing of basic information like user name, browsing time and history in the web server is the strategy used by memorization [4]. This is the oldest form of web personalization. There are two major functionalities done by memorization techniques, User Salutation and Book marks. For user salutation, the server uses this information when the user returns. They greet the user specifically by using their name along with welcome message. This strategy helps in increasing the loyalty of customers by providing them a feel of special attention rather than a regular visitor. The other usage of Bookmarking, The framework stores the Web pages that a client has gone to in the past and presents them to the client by method for a customized bookmarking pattern for that webpage [12]. This mode depends more on Web innovation than on any sort of versatile or insightful learning. It can likewise endanger client protection.

B. Customization

The content provider of a web site doesn't have a control over the view of a client, once they upload the page. This makes an opening for end-users to automate and customize the web pages of their choice. Customization [8] is an advanced web personalization strategy than memorization. In this approach, it accepts user's preferences during registration and uses their knowledge and interest for customizing the substance and structure of a page. There are different ways of customization: page layout, content, link etc. In page layout customization, the layout or color of the page changes according to the user preferences. At the same time in content customization, the page content is presented in summarized or expanded form based on user preferences.

C. Task Performance Support

Task performance support is a personalization function, inherited from personal assistant [13].

This will perform like an agent who executes a specific activity for the benefit of a client. It is one of the propelled personalization capacity, and can be implemented in client side. Personalized errand is a task support system which can do routine jobs like sending e-mails, downloading various items etc. The errands are varying from simple routine tasks to more complex ones depending upon the complexity of personalization system. Personalized query completion is another task performance support system, which work with information retrieval system and complete or even enhance the queries submitted by the user to a web index or web database framework.

D. Recommender Systems

Recommendation systems play a vital role in understanding customers' behavior, preferences, improving customer convenience, increasing service provider profits and future needs. Recommender system is much more advanced than customization and memorization. The recommender system is fully based on hyperlinks, it analyze the previous visits of a user and find out his navigational path. Next time the system automatically recommends hyperlinks to the user according to his navigational path [9]. The recommender system also has an alternative set of hyperlinks for the user according to the interest and preferences. This is supported by various systems such as the Web Personalizer [9]. This Web Personalizer base web log mining and use data mining techniques to extract knowledge for providing a personalization framework to current clients taking into account their scanning navigational history. The recommender system is used for adaptive recommendation and with the goal that it is generally executed on the Web server. The adaptive recommender system uses the browsing history and user profiles and does recommendations both implicitly and explicitly at every progression of the communication with the site. There are some other recommendation systems like WebWatcher [11] and Letizia [10] make recommendations by not only considering user profiles but check content similarity too.

IV. PROGRAMMED PERSONALIZATION BASED ON WEB USAGE MINING: A STUDY

The personalized website content leads to popularity in between customers and increase

return on marketing investment [20]. it's been a foregone conclusion and because of that most of the companies are putting resources into innovations, capacities, and assets to bring the idea of web substance personalization to life. In the context of web personalization, the web usage data has a very important role. Usage data represent a web site's visitor IP, visiting time and date, files or directories they accessed, duration etc. [3]. These data gives the identity of the user along with browsing behavior. Researches in the field of recommender system create more and more advancement in personalization and so as leads to automatic personalization. It is proved that web usage mining can be used as an aid to overcome the problems with traditional web personalization techniques. An investigation [14] on usage mining notify that usage based personalization will become a challenging task if there is little usage data or regular changes in site content. This problem can be overcome by a framework which integrates both content and usage attributes [14].

The kNN mechanism of standard collaborative filtering for personalization involves in real time evaluation of user records both current and historical. But the major drawback of this mechanism is as the number of users increase this will become impractical, ie. kNN cannot deal with large volume of data. To maintain scalability a new framework has designed [15] in that the personalization has done with associate rule mining on click stream data. It has an efficient data structure to store frequent items and a recommendation algorithm. Combined with the data structure the recommendation algorithm generates required association rules from items set instead of all.

Web usage mining techniques has the power to effectively capture the relevant user navigational patterns for web personalization. The approach proposed in [16] do an automatic discovery of user interest domain. They adopted the method of generating concept hierarchy based clusters and its effectiveness has been tested with fuzzy proximity relations. So in effect web site permits clients to settle on their decisions in common way. The specialists suggested this model is equipped for delivering proposal sets with high scope and high F1 measure with moderate accuracy.

Most of the popular search engines adopted link based ranking approach, but in some cases they are one-sided towards to most well known importance and tend to produce results based on that. As of late, customized internet searchers have been presented with the point of focusing so as to enhance indexed lists on the client conduct, as opposed to on their submitted questions. One of such approach is explained in [17], Hierarchical Clustering Engine SNAKET, can be plug on top of any impersonalized internet searcher with a specific end goal to acquire a type of personalization. The user query is accepted by SNAKET and it choose related labeled folder hierarchy and helps to select appropriate folder label that best fit to the user query. SANKET personalizes the ranked list by filtering out the unwanted folders. This approach does not require the precise login information of the user or client profile, or the example of client past pursuit conduct.

Association mining had proved that it can easily predict user browsing behavior for web personalized recommendation. Even though it predicts the pattern but because of the low coordinating rate of the subsequent tenets and different in scanning conduct of the user the application of it for predicting future pattern is low. But the research presented in [18] proposes another suggestion framework by coordinating grouping with affiliation mining procedure. They utilized another bunching technique, called HBM (Hierarchical Bisecting Medoids Algorithm) to group clients in light of the time-encircled route sessions. After that association mining is used to analyze those sessions to build up a suggestion model for similarity.

The opportunity of mass customization process in web-based companies is supported by its capability to gather point by point use information at the level of individual mouse clicks. They heavily rely on human participation for collecting their profiles. Since it is collecting at the time of usage, the data may be subjective, or out of date as when user preference changes over time. So an approach [19] presented a different way of programmed Web personalization taking into account Web use information. They dispatched a successful bunching procedure utilizing affiliation principle mining to learn covering client profiles, and talked about how the removed information can be utilized as a part of constant to give navigational pointers to clients.

V. CONCLUSION

This paper introduced the various advances of web usage mining techniques in the area of web personalization. Web personalization is turned into an unavoidable piece of e-business applications. On the off chance that the site proprietor can incorporate the personalization apparatus with the site then they can ready to better examine and judge the offerings. Electronic business is one of the real strengths that permit the Web to prosper, however the achievement of electronic trade relies on upon how well the webpage proprietors comprehend clients' conduct and needs. Web utilization mining can be utilized to find intriguing client route designs, which can then be connected to genuine issues, for example, Web website/page change, extra item/subject proposals, client/client conduct concentrates on, and so forth.

REFERENCES

[1]. M. Eirinaki, M. Vazirgiannis, (2005), "Web Mining for Web Personalization", *ACM Transactions on Internet Technology.*

[2]. A. Antoniou, M. Paschou, E. Sourla and A. Tsakalidis, (2010),"A Semantic Web Personalization Technique: The case of bursts in web visits", *IEEE Fourth International Conference on Semantic Computing.*

[3]. Chen L, Sycara K , (1998), "A personal agent for browsing and searching". *In proceedings of the 2nd International Conference on Autonomous Agents, Minneapolis/St. Paul.*

[4]. A. J. Ratnakumar,(2005), "An implementation of Web Personalization using Web Mining Techniques", *Journal of Theoretical and Applied Information Technology.*

[5]. Schafer J, Konstan J and Reidel J , (1999), " Recommender systems in E-Commerce", *In Proceedings of ACM Conference E-Commerce.*

[6] Joachim T, (2002), "Optimizing search engines using click through data", *In proceedings of the 8th ACM SIGKDD Conference.*

[7] Olfa Nasraoui,(2005), "World Wide Web Personalization", *Invited chapter in Encyclopedia of data mining and Data ware housing*.

[8]. Michael Bolin, Matthew Webber, Philip Rha, Tom Wilson, and Robert C. Miller, (2005), "Automationand customization of rendered web pages", *Proceedings of UIST, ACM Press.*

[9]. Mobasher, B., Cooley, R., and Srivastava, J. (2000). Automatic personalization based on web usage mining, *Communications of the. ACM*

[10]. H. Lieberman, Letizia,(1995), An agent that assists web browsing, *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence.*

[11]. T. Joachims, D. Freitag, and T. Mitchell. (1997), Webwatcher:A tour guide for the world wide web. *15th International Conference on Artificial Intelligence.*

[12]. Suguna, R. and Sharmila, D.(2012), "Association Rule Mining for Web Recommendation", *International Journal on Computer Science & Engineering.*

[13]. Mitchell,T.,Caruana, R.,Freitag,D.,McDermott,J. and Zabowski,D.: 1994, *Experience with a learning personal assistant, Communications of the ACM*

[16] Bhawesh Kumar Thakur , Syed Qamar Abbas and Mohd. Rizwan Beg (2014), Web Personalization using Clustering of Web Usage Data, *International Journal in Foundations of Computer Science & Technology (IJFCST), Vol.4, No.5.*

[17]Paolo Ferragino, Antonio Gulli; (2005), A personalized search engine based on web-snipped hierarchical clustering, Proceeding of www, *Special interest tracks and posters of the 14th international conference on world wide web; ACM Communications.*

[18] Feng-Hsu Wanga, Hsiu-Mei Shao, (2004), Effective personalized recommendation based on time-framed navigation clustering and association mining; Expert Systems with Applications ; *Elseiver.*

[19] Feng, Xiuzhen, Yang Peng, Haoran Xie, and Ziyu Yan. (2011), "Role-Based Learning Path Discovery for Collaborative Business Environment", *International Conference on Control Automation and Systems Engineering.*

[20]. Magdalini Eirinaki,(2003), "Web mining for web personalization", *ACM Transactions on Internet Technology.*