Volume 4, Issue 12, December 2014 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

A Review on User's Future Request Prediction in Web Usage Mining

Chintankumar S. Maisuriya
Research Scholar, CSE Department
Parul Institute of Engineering and Technology
Vadodara, Gujarat, India

Mr. Vaibhav Gandhi
Assistant Professor, CSE Department
Parul Institute of Engineering and Technology
Vadodara, Gujarat, India

Abstract— Today, Internet is playing such a significant role in our day-to-day life. We have witnessed the evermore-interesting and upcoming publishing medium is the World Wide Web (WWW). The rapid growth in the volume of information available over the WWW and number of its' potential users' has leads to difficulties in providing effective search service for users', resulting in decrease in the web performance. Web Usage Mining is an area, where the navigational access behaviour of users' over the web is tracked and analyzed. So that websites owner can easily identify the access patterns of its users'. By collecting and analyzing this behaviour of user activities, websites owner can enhance the quality and performance of services to catch the attention of existing as well as new customers. This research paper intends to provide an overview of past and current evaluation in users' future request prediction using Web Usage Mining.

Keywords— Future Request Predictions, Pattern Discovery, Users' Navigational Behavior, Web Mining, Web Usage Mining, World Wide Web.

I. INTRODUCTION

Today in the era of information, Data is very significant and needed to be ubiquitous. Also information dominates the world more than any time before. With the technological advancement as well as by the raised popularity of WWW, many websites typically experience thousands of visitors and its users' daily. WWW has turned to be the significant and largest source of information retrieval. It is very significant to examine the usage of the websites by its users' and web traffic due to growing rate of WWW.

Web usage mining is the process of analyzing and automatically discovering the useful knowledge from the data collected in the web log files. The collected web log file and pattern analysis click streamed knowledge helpful to web usage mining which can recommend a set of objects to the active user, possibly consisting of links, ads, text or products, tailored to user perceived preference [6].

II. MOTIVATION

Internet plays significant role in our day-to-day life and thus it becomes very difficult to survive without it. With the growing popularity of WWW, millions of users access websites in all over the world. A large amount of data such as IP address of users, URL requested, sequence of pages accessed by them, etc. information are automatically collected and maintained in the access log files by server during users' accesses websites. It is very important because many times users repeatedly access the same types of web pages. The series of accessed web pages can be considered as web browsing patterns of users' which can be helpful to find out the users' access behavior and through this we can find out the accurate user future requests predictions which helps in reducing the browsing time of web pages for users as well as decrease the server loads. In recent years, lots of research work has been done with in this field. The main motivation behind this survey is to know what research has been done web usage mining in future request prediction.

III. SURVEY OF RELATED WORK

The main focus of the literature review is the study about web usage mining which is used to find the web navigation behavior of web-users' and collecting the information about 'Uses Future Request Prediction' approach that will be used to predict users' future accesses.

In [1], Jaideep Srivastava et al., have categorized data pre-processing task into subtasks and noted that the final outcome of pre-processing should be data that allows identification of a particular user's browsing pattern in the form of page views, sessions, and click streams. Click streams are of particular interest because they allow reconstruction of user navigational patterns. Markov models have been extensively used to model web users' navigation behavior on web sites.

In [2], Alexandras Nanopoulos et al., have researched on 'Web pre-fetching' because of its significance in reducing user perceived latency present in every web based application. Due to web popularity, there is heavy traffic in the internet which results in the delay of response. The reasons of delay are the web servers are under heavy load, Network congestion, Low bandwidth, Bandwidth underutilization and propagation delay. The solution is to increase the bandwidth

but this is not proper solution because of economic cost. For that propose, this technique was proposed for reducing the delay of client future requests for web objects and getting that objects into the cache in the background before an explicit request is made for them. Technique was implemented for Web server to cooperate with a pre-fetch engine to disseminate hints every time a client request to a document of the server. Authors have presented important factors which have affect on Web pre-fetching algorithm like order to dependencies between Web document accesses and the interleaving of requests belonging to patterns with random ones within user transactions and the ordering of requests.

In [3], Yi-Hung Wu et al., have presented user behaviors by sequences of consecutive web page accesses, derived from the access log of a proxy server. Moreover, the frequent sequences are discovered and organized as an index. Based on the index, they have proposed a scheme for predicting user requests and a proxy based framework for prefetching web pages. They have performed experiments on real data. The results show that their approach makes the predictions with a high degree of accuracy with little overhead. In the experiments, the best hit ratio of the prediction achieves 75.69%, while the longest time to make a prediction only requires 1.9ms. The disadvantage is that the average service rate is very low. The other problem is the setting of the three thresholds used in the mining stage. These thresholds have great impacts on the construction of the pattern trees. The use of minimum support and minimum confidence is to prune the useless paths. Obviously, some information might be lost if the pruning effects are overestimated and the grouping confidence is only useful for the strongly related web pages due to some editorial techniques, such as the embedded images and the frames.

In [4], Christos Makris et al., have proposed a technique for predicting web page usage patterns by modeling users' navigation history using string processing techniques, and validated experimentally the superiority of proposed technique. In this paper weighted suffix tree is used for modeling user navigation history. The proposed technique has the advantage that it demands a constant amount of computational effort per user action and consumes a relatively small amount of extra memory space.

In [5], Nien-Yijan and Nancy P. Lin have proposed trend based application system to analyze user behaviors and predict the future traveling path based upon the trend similarity. It is inappropriate to predict the browsing behavior of current user according to the similarity comparison with single browsing cluster of older users. Therefore a trend based prediction model is proposed to predict the future traveling path by generating ordering browsing sequence. The proposed system works in two phases. One is prediction model constructing phase and the other is predicting phase. The construction steps is proposed to help experts discover useful common browsing patterns and then used to predict the further browsing sequences. In predicting phase, the browsing behavior of the new user can be obtained to compare the similarity with the prediction model; hence the candidate's pages could be pre-fetching to improve the browsing performance. By applying on the replacement algorithm of proxy servers and the experimental result shows the performance of proposed model is useful to pre-fetch candidate's pages.

In [6], Mehrdad Jalali et al., have proposed a recommendation system called 'WebPUM', an online prediction using Web usage mining system for effectively provide online prediction and have proposed a novel approach for classifying user navigation patterns to predict users' future intentions. The approach is based on the new graph partitioning algorithm to model user navigation patterns for the navigation patterns mining phase. LCS algorithm is used for classifying current user activities to predict user next movement. The architecture of WEBPUM is divided into two parts:-

- *Offline phase:* This phase consist two main modules, which are data pre-treatment and navigation pattern mining. Data pre-treatment module is designed to extract user navigation sessions from the original Web user log files. A new clustering algorithm based on graph partitioning is introduced for navigation patterns mining.
- *Online phase:* The main objective of this phase is to classifying the user current activities based on navigation patterns in a particular Web site, creating a list of recommended Web pages as prediction of user future movement. The main online component is the prediction engine.

In [7], A. Anitha has proposed a new web usage mining approach is proposed to predict next page access. It is proposed to identify similar access patterns from web log using pair-wise nearest neighbor (PNN) based clustering and then sequential pattern mining is done on these patterns to determine next page accesses. The tightness of clusters is improved by setting similarity threshold while forming clusters. In traditional recommendation models, clustering by non-sequential data decreases recommendation accuracy. It is proposed to integrate Markov model based sequential pattern mining with clustering. A variant of Markov model called dynamic support pruned all kth order Markov model is proposed in order to reduce state space complexity. Mining the web access log of users of similar interest provides good recommendation accuracy. Thus, the proposed model provides accurate recommendations with reduced state space complexity [7].

In [8], Priyanka Makkar et al., have proposed a novel approach for increasing web performance by analyzing and predicting user access behavior both by collaborating information from web access log and website structure repository. In this first the data mining techniques are applied to extract user web access patterns from weblogs. They have done pre-fetching to improve the web performance, in the anticipation that the retrieved patterns could be served from cache in the future leads to reduce web latency. An application of petri nets (PN), a high level graphical model used in modeling system activities with concurrency, were adopted to enhance log mining. Web structure is extracted by using parsing algorithm, from which incidence matrix is built. Thus web structure information in an incidence matrix and the reachability properties obtained from the PN model helps in path completion process. Thus it pre-fetch webpages in client cache before they explicitly request for it, which decrease web latency and improves web performance.

In [9], V. Sujatha et al., have proposed the Prediction of User navigation patterns using Clustering and Classification (PUCC) from web log data. In the first stage PUCC focuses on separating the potential users in web log data, and in the second t stage clustering process is used to group the potential users with similar interest and in the third stage the results of classification and clustering is used to predict the user future requests. The first stage is the cleaning stage, where unwanted log entries were removed. In the second stage, cookies were identified and removed. The result was then segmented to identify potential users. From the potential user, a graph partitioned clustering algorithm was used to discover the navigation pattern. An LCS classification algorithm was then used to predict future requests.

In [10], Phyu Thwe has proposed a Page Rank-like algorithm is proposed for conducting web page access prediction. Once the data pre-processing is done the markov model is used to predict next page access on web session. If ambiguous results are found, page ranking algorithm is used for finding more important pages with respect to the search results. Here the use of page rank algorithm is extended for next page prediction with several navigational attributes, which are the similarity of the page, size of the page, access-time of the page, duration of the page and transition(two pages visits sequentially) and frequency of page and transition.

In [11], Dilpreet Kaur et al., have proposed a technique for predicting user future access request by using fuzzy clustering methods as fuzzy c-means (FCM) and kernelized fuzzy c-means (KFCM). Here first the collected log files are pre-processed, unwanted log entries are cleaned, users are identified by their IP address and session identification is done by taking time as threshold value. Once the pre-processing of logs is done the fuzzy clustering algorithms are applied to divide the data stored in array structure in to the clusters. In fuzzy c-means technique, each data elements are assigned by their membership level based on the distance between the data point and cluster center, which indicates the strength of association between data elements and a particular cluster. Whereas in KFCM a distance between cluster center and data point is derived by using kernelized distance matrix instead of Euclidean distance in FCM. So the webpages with highest membership value are retrieved in each clusters and weight is assigned to each webpage according to grade of membership. The page has more weight has more probability for accessing in future by user and based on that predictions are made.

TABLE: 1 SUMMARY OF RELATED WORK

Author	**Method**	**Application**	**Publication Year**
Alexandros Nanopoulos et al. [2]	Web page pre-fetching into cache	Prediction enabled web server	2001
Yi-Hung Wu et al. [3]	Frequent Sequence, Proxy based framework for webpage pre-fetching	Prediction system using proxy server log	2002
Christos Markis et al. [4]	Weighted suffix tree, string processing techniques	Webpage usage pattern prediction system	2007
Nien-Yijan, Nancy P. Lin [5]	Trend based prediction	Prediction system architecture	2007
Mehrdad Jalai et al. [6]	Online/Offline phase of architecture, LCS algorithm, Clustering	WebPUM	2010
A. Anitha [7]	Sequential pattern mining with k^{th} order markov model, clustering	Next page access prediction system	2010
Priyanka Makkar et al. [8]	Association rule mining, clustering, application of petri nets and parsing algorithm	User behavior prediction system architecture	2010
V. SUJATHA, PUNITHAVALLI [9]	Classification. Clustering, LCS algorithm	Users navigational pattern prediction	2012
Phyu Thwe [10]	Sequential pattern mining using markov model, popularity and similarity based page ranking	Webpage access prediction	2013
Dilpreet Kaur et al. [11]	Fuzzy Clustering	User future request prediction	2013

IV. OPEN CHALLENGES TO WEB USAGE MINING

There are some open challenges to web usage mining stated as follows:

- There is a continuously growth in the volume of data
- The pre-processing task does not gain enough analysis efforts
- The frequent pattern mining techniques often provide short and uninteresting results
- The frequent pattern mining techniques for WUM might be not appropriate for dealing with large amount of web usage data

V. APPLICATIONS OF WEB USAGE MINING

The results provided by log mining process can be used for various applications:

- Personalization of web content based on user's interest
- Target potential customers for e-commerce
- Improve web system performance
- To develop pre-fetching and caching strategies
- To predict users navigational behavior

VI. CONCLUSION

This paper has attempted to provide survey for browsing behavior of users' and subsequently to predict desired pages in the area of web usage mining. The predictions can be improved by using different techniques of pattern discovery like clustering, classification, association rules etc. so as to improve the web performance by reducing the browsing time.

REFERENCES

[1] J. Srivastava and R. Cooley, "Web Usage Mining: Discovery and Applications of Usage Patterns from Web data", SIGKDD Explor. Newsl, New York, USA, Vol.1, pp 12-23, Jan-2000.

[2] Alexandros Nanopoulos, Dimitris Katsaros and Yannis Manolopoulos "Effective prediction of web-user accesses: A data mining approach," in Proc. Of the Workshop WEBKDD, 2001.

[3] Yi-Hung Wu and Arbee L. P. Chen, "Prediction of Web Page Accesses by Proxy Server Log" World Wide Web: Internet and Web Information Systems, 5, 67–88, 2002.

[4] Christos Makris, Yannis Panagis, Evangelos Theodoridis,and Athanasios Tsakalidis "A Web-Page Usage Prediction Scheme Using Weighted Suffix Trees" © Springer-Verlag Berlin Heidelberg 2007.

[5] Nien-yi jan and Nancy P. Lin, "Web User Behaviours Prediction System Using Trend Similarity" Proceedings of the 7th WSEAS International Conference on Simulation, Modelling and Optimization, Beijing, China, September 15-17, 2007.

[6] Mehrdad Jalali, Norwati Mustapha, Md. Nasir Sulaiman, Ali Mamat, "WebPUM: A Web-based recommendation system to predict user future movements" Expert Systems with Applications 37, 2010.

[7] A. Anitha, "A New Web Usage Mining Approach for Next Page Access Prediction", International Journal of Computer Applications, Volume 8– No.11, October 2010.

[8] Priyanka Makkar, Payal Gulati, Dr. A.K. Sharma, "A Novel Approach for Predicting User Behavior for Improving Web Performance", International Journal on Computer Science and Engineering (IJCSE), Vol. 2, No. 04, 2010.

[9] V. Sujatha, Punithavalli, "Improved User Navigation Pattern Prediction Technique From Web Log Data", Procedia Engineering 30, 2012.

[10] Phyu Thwe, "Proposed Approach For Web Page Access Prediction Using Popularity And Similarity Based Page Rank Algorithm", International Journal of Scientific & Technology Research (IJSTR), Volume 2, Issue 3, March 2013.

[11] Dilpreet Kaur, A.P. Sukhpreet Kaur, "User Future Request Prediction Using KFCM in Web Usage Mining", International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), Vol. 2, Issue 8, August 2013.

