

Prediction of User Browsing Behavior Using Web Log Data

Abhishek Chauhan, Sandhaya Tarar

Gautam Buddha University, Uttar Pradesh, India

ABSTRACT

Web usage mining is one of the applications of data mining techniques to discover usage patterns from web log data. It deals with log files for excerpting the useful information about user browsing behavior or we can say that it is the process of finding out what users are looking for on internet. When the user accesses to any websites, all the user's browsing behavior activities are stored in web log server. With the help of web usage mining various research had been done in this field but this paper deals with the prediction of user browsing behavior using web log data. This paper is not only provide the overview of current and past techniques used by authors to curtail the search time of user on the network but also provide the limitation on their research work.

Keywords: Web Usage Mining, Web Log, User Browsing Behavior.

I. INTRODUCTION

Nowadays, Internet is a part and parcel of our life without having it we couldn't imagine our life. When a user can access the web pages, they left some imperative information stored in their web log. This content is very useful and important in determining the web page navigational pattern of user. Web usage mining consist of three phases i.e. data pre-treatment, pattern discovery, pattern analysis. Firstly, all the web log data is being under data pre-treatment to retrieve logs with minimum redundancies and user session. Secondly, pattern discovery is used to extract user navigation patterns. At last, pattern analyzing algorithm is applied to extract data for data mining applications.

In this paper, we did literature survey on user's future request prediction - web usage mining by using various methods and algorithms which have been proposed on this work and this paper also discussed about their advantages and disadvantages. The paper has been organized as follows: Section-2 contains the related work, Section-3 contains literature review on user's browsing behavior, Section-4 contains summary of the review paper, and Section 5 contains the conclusion and future work.

II. RELATED WORK

With the growing popularity of World Wide Web, millions of user access website in all over the world. When a user browses the web pages, a large amount of data is congregated and recorded in the web log files. This series can be considered as a web access pattern which is helpful to find out web navigation behavior of user. Many techniques and algorithms has been done on determining the user future request prediction in order to get better, accurate and also efficient results. With the help of this behavior information, we can find out the accurate and efficient user next prediction which not only reduces the browsing time of web pages i.e. saves more time of user but also decreases the server load. In recent years, a huge amount of research has been done with regard to web usage mining for user web browsing behavior. The main motivation of this study is to know that what research has been done on web usage mining in prediction of user browsing behavior.

III. METHODS AND MATERIAL

A. Literature Survey

Study and Implementation Of LCS Algorithm For Web Mining: Vrishali P. Sonanane [1] proposed the prediction of user navigation patterns using LCS algorithm. They delineate the user behavior prediction process which consist of following phases i.e. data pretreatment, navigation pattern mining, navigation pattern modelling, clustering and prediction engine. Firstly, all the web log data is cleaned and filtered by going under data treatment phase. After that navigation pattern mining on the derived user access pattern and also to model navigational patterns an algorithm is used for modelling the pages accesses information as an undirected graph. In clustering phase, we try to find out groups of strongly correlated pages. At last in prediction engine, LCS algorithm is applied to classify user navigation patterns and predict user's future request. Using LCS algorithm, we can predict user's future request more accurately Evaluation of Web Usage Mining Approaches for User's Next Request Prediction: Mathias Grey, Hatem Huddad [2] suggested a framework for a recommender system that predicts the user's next request based on their behavior discovered from web log data. They had amalgamated three web usage mining techniques i.e. association rules, frequent sequence and frequent generalized sequence to predict user's next request web page. By using rule function selection, they define two prediction strategies i.e. Highest Confidence (HC) and Last Sequence (LS) to select discovered rules matching the pages requested by a user. Author performed some experiments on three collections of real usage data: one from an intranet web site and two from an internet web site. After performing the experiments all the results show that frequent sequence gives better accuracy than association and frequent generalized sequence.

An improved user browsing behavior prediction using web log analysis: Vedpriya Dongre, Jagdish Raikwal [3] proposed a system architecture for finding the hidden navigational patterns using web log data. In the architecture, more than one client is linked through the server and then server generated a log and all the data is going to the pre-processing phase where data is being cleaned and filtered. After that all the data go to the clustering phase where K-means clustering is applied to

find the similar data from huge database i.e. web pages accessed by targeted user and web page accessed by other user. For divination, all the clustered data is going for the regression analysis which is used to accurate data over the numeric values after estimating the weight for a target user.

Web Usage Mining for Predicting Users' Browsing Behaviors by using FPCM Clustering: R.khanchana and M.Punithanali [4] come up with web usage mining technique for predicting the user's browsing behavior by FPCM clustering. They characterized the two-levels of prediction model by merging the Markov model and Bayesian theorem. But these models applied better for the general purpose. To overcome this difficulty, they uses Fuzzy Possibilistic Algorithm for clustering for heterogeneity user's behavior. Experimental results show that the hit ratio of FPCM cluster view is 63.6 % whereas the global and hierarchical view is 57.8 % and 59.9 % respectively. So they observed that the FPCM clustering has better predicting ability.

WebPUM: A Web-based recommendation system to predict user future movements: Mehrdad Jalali, Narwat Mustapha, Md. Nasir Sulaiman, Ali Mannat [5] advanced their previous work and renamed their architecture as Web PUM. They applied user navigation patterns for the navigation pattern mining phase. They used LCS for classifying all the user's activities to predict user near future movements. They oversighted two main experiments for navigation pattern mining and their system has been tested on CTI and MSNBC.

Study on Consumer Behavior Predict in E-commerce Based on Multi-Agent: Yan Rong Zhang and Zhijie Zhao [6] recommended a consumer behavior forecasting model which uses the method of user interest concept tree based on domain ontology. They uses multi-agents i.e. consumer behavior forecasting agent , user interest model management model, monitoring agent and source data monitoring agent which is used to predict the consumer behavior by combining with each other. They also suggested a consumer behavior predict expert system by using artificial intelligence technology to predict e-commerce consumer's behavior effectively.

A new classification model for online predicting users' future movements. Mehrdad Jalali, Narwat Mustapha, Md. Nasir Sulaiman, Ali Mannat [7] proposed an online and offline phase architecture. Both of this architecture

works simultaneously. In offline phase, all the web log data is being under the data pretreatment module process and reformat it to identify all web access sessions. All the similar properties or browsing behavior of the users are being clustered in navigational pattern mining module. In online phase, the underlying knowledge base is updated and the list of suggestion is appended in the prediction list. The semantic knowledge about underlying domain can be used to improve the quality of the recommendation.

Improved Web Prediction Algorithm Using Web Log Data: Megha P. Jarkad, Prof.Mansi, Bhonsle [8] scheduled a system architecture which contains five steps using classification, clustering and background algorithm. Firstly, all the web log data is being under preprocessed to abolish unwanted entries. In second phase, differentiate the potential and non-potential users using decision rule. The purpose of classification is to decrease the size of web log file. In the third phase , clustering is performed using graph partitioned algorithm is applied on smaller unit of data i.e. which divides all data subsequently at each stage and repeat the procedure until it find equal which reduces the prediction time . In last phase, prediction is given using LCS to classify user navigation patterns and predict user’s future requests. By using classification, clustering and backtracking algorithm not only enhance the

performance and reduce the time complexity of the proposed system. User Future Request Prediction Using KFCM in Web Usage Mining: Dilpreet kaur, A.P.sukhpreet kaur[9]proposed a system architecture using Fuzzy Clustering i.e. fuzzy c-means and kernelized fuzzy c-means algorithm. Firstly,web log is acquired and then it goes to the preprocessing phase. After preprocessing phase, fuzzy clustering algorithm is assigned for user’s future request prediction and then results were interpreted. Through this, authors concluded that KFCM is not only more robust than FCM but also creates better clusters for prediction.

A New Web Usage Mining Approach for Next Page Access Prediction:

Yogesh rajaram bhalerao, Prof. P.P. rokade[10] planned a system architecture to predict user navigation pattern using statistical classifier and modern techniques. In the architecture, dataset is used preprocessing from multiple users and find potential user based on the amount of time spent by them in a website from that dataset for further pairwise nearest neighbor clustering algorithm. Later, the maximum likelihood classification algorithm is applied for prediction.

IV. Summary of Literature Review

SNO.	AUTHORS	METHOD	AAPPLICATION	DISADVANTAGE
1	Vrishali P. Sonanane	LCS algorithm.	Prediction system architecture.	The LCS algorithm is used to edit distance between sequences, so it is much faster when the difference between two subsequence is small
2	Mathias Grey, Hatem Huddad	Association rules, Frequent Sequence and Frequent Generalized Sequence.	Evaluate AR, FS, FGS.	The main disadvantage of association mining is that it does not inherently use the notion of temporal distance but also frequent sequences cannot predict navigation patterns for data sets.
3	Vedpriya Dongre, Jagdish Raikwal.	K-means and Regression analysis.	Prediction system architecture.	The author didn’t discuss the parameters like time taken for prediction and also memory used to store the data.

4	R.khanchana and M.Punithanali.	FPCM clustering.	Predicting user's browsing behavior.	The two-levels of prediction model gives better result for general cases but it suffers from the heterogeneity user's behavior.
5	Mehradad Jalali, Narwat Mustapha, Md. Nasir Sulaiman, Ali Mannat.	Online/Offline phase of architecture, LCS Algorithm, clustering.	Online prediction future user moments.	A large websites which contain million of pages were severely affected as memory required to store web server pages is quadratic in number of pages.
6	Yan Rong Zhang and Zhijie Zhao.	Multi-agents.	Consumer behavior forecasting method.	Instead of finding weighted values for child , we can find direct frequencies.
7	Mehradad Jalali, Narwat Mustapha, Md. Nasir Sulaiman, Ali Mannat.	Online/Offline phase of architecture, LCS Algorithm, clustering.	Web PUM	The main disadvantage is that the websites made up from dynamically generated pages cannot be easily managed.
8	Megha P. Jarkad , Prof.Mansi, Bhonsle.	Classification, Clustering & Backtracking Algorithm.	Prediction System.	The author used graph partitioned clustering algorithm instead of calculating weights of the web pages direct frequencies can be used for prediction.
9	Dilpreet kaur, A.P.sukhpreet kaur	Fuzzy Clustering i.e. fuzzy c-means and kernelized fuzzy c-means algorithm.	User future request prediction.	The disadvantage in this research work is that FCM is less robust and cannot apply on the large sets.
10	Yogesh rajaram bhalerao, Prof. P.P. rokade.	Using statistical classifier and Maximum likelihood classification algorithm .	User navigation pattern.	Maximum likelihood classification algorithm estimates can be heavily biased for small samples.

V. CONCLUSION

In this survey, how the authors can use different algorithms and techniques with the help of web usage mining. Various researches had done on future request prediction approach. This survey focuses on how to advance the prediction time without compromising prediction accuracy by using various algorithms of pattern discovery techniques like graph techniques of clustering and also many types of models are developed for prediction. In future, the research can be broadened for a small number of previous log files and an in-depth

analysis to enhance the prediction accuracy level by using different techniques of data mining.

VI. REFERENCES

- [1] Vrishali P. Sonavane, "Study And Implementation Of LCS Algorithm For Web Mining", International Journal of Computer Science Issues, Vol. 9, Issue 2, No 3, March 2012.
- [2] Mathias Gery, Hatem Haddad, "Evaluation of Web Usage Mining Approaches for User's Next Request Prediction" WIDM'03 Proceedings of the 5th ACM

international workshop on web information and data management p.74-81, November 7-8,2003.

- [3] Vedpriya Dongre, Jagdish Raikwal ,”An improved user browsing behavior prediction using web log analysis”, International Journal of Advanced Research in Computer Engineering and technology (IJARCET), Vol. 4, Issue 5, , May 2015.
- [4] R.Khanchana and M. Punithavalli, “Web Usage Mining for Predicting Users’ Browsing Behaviors by using FPCM Clustering”, IACSIT International Journal of Engineering and Technology, Vol. 3, No. 5, October 2011.
- [5] M. Jalali, N. Mustapha et al ,” WebPUM: A Web-based recommendation system to predict user future movements”, in international journal Expert Systems with Applications 37 (2010) 6201–6212 .
- [6] YanRong Zhang1 and ZhiJie Zhao,” Study on Consumer Behavior Predict in E-commerce Based on Multi-Agent “,International Journal of u- and e-Service, Science and Technology Vol. 7, No. 6 (2014), pp. 403-412.