



ISSN: 2319-5967

ISO 9001:2008 Certified

International Journal of Engineering Science and Innovative Technology (IJESIT)

Volume 2, Issue 5, September 2013

# WUM for Browsing Behavior of a User and Subsequently to Predict Desired Pages: A Survey

Hemant N.Randhir, Ravindra Gupta

*Abstract— This paper is a survey of recent work in the field of web usage mining for the benefit of research on the web log files of Web-based information services. Web usage mining is a main research area in Web mining focused on learning about Web users and their interactions with Web sites. The motive of mining is to find users access models automatically and quickly from the vast Web log data, such as frequent access paths and frequent access page. Through web usage mining, the server log, registration information and other relative information left by user access can be mined with the user access mode which will provide foundation for decision making of organizations. This article provides a survey and analysis of current Web usage mining systems and technologies. This paper also discusses an application of Web usage mining.*

*Index Terms— Web Usage Mining, Web log file, Application WUM.*

## I. INTRODUCTION

A lot of research work has been done on Web Usage Mining for web personalization and predicting their web access behavior. When the user browses the web pages, user leaves some valuable information in web log files. This web log file information is very helpful to find out web navigation behavior of user. Through his behavior we can find out what kind of information user wanted from the web sites.

Web usage mining automatic extract the knowledge from the data collected in web log file. 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 [1]. This paper did literature survey on “WUM for browsing behavior of a user and subsequently to predict desired pages”. The various techniques have been developed on the web usages mining and this article reviewing and highlights about the various techniques. This paper has been organized as follows: Section 2 presents Literature Survey, Section 3 Application of WUM and Section 4 concludes the literature survey.

## II. LITERATURE SURVEY

The aim of this literature survey is to study and analyze the available technique to predict the user behavior in interacting with the Web or with a Web site. There are many Web log analysis tools available to mine data from log record on Web page. Log record contains plenty of useful information such as URL, IP address and time and so on. Analyzing and discovering log could help organizations to find more potential customers, pages popularity number of times a page has been visited etc. that can help in reorganizing the Web site for fast and easy customer access, improving links and navigation, attracting more advertisement capital by intelligent adverts, turning viewers into customers by better site architecture.

Jaideep Srivastava et al. have categorized data preprocessing into subtasks and noted that the final outcome of preprocessing 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 behaviors on Web sites [2].

Alexandras Nanopoulos et al. have focused on ‘Web pre-fetching’ because of its importance in reducing user perceived latency present in every Web based application. From the Web popularity, there is heavy traffic in the internet and the result is that there is delay in response. The reasons of delay are the Web servers under heavy load, Network congestion, and Low bandwidth, Bandwidth under utilization 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 proposed in which 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. Author implemented a technique for Web server to cooperate with a pre-fetch engine to disseminate hints every time a client request to a document of the



ISSN: 2319-5967

ISO 9001:2008 Certified

International Journal of Engineering Science and Innovative Technology (IJESIT)

Volume 2, Issue 5, September 2013

server. Author presented important factors which affects 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 [3].

Murat et al. have implemented novel agent simulator for generating simulated user sessions. A agent simulator generates complete sessions satisfying both connectivity and timestamp rules [4].

Jose Borges et al. have proposed a dynamic clustering-based method to increase a Markov model's accuracy in representing a collection of user Web navigation sessions. This method makes use of the state cloning concept to duplicate states in a way that separates in-links whose corresponding second-order probabilities diverge. In addition, the new method incorporates a clustering technique which determines an ancient way to assign in-links with similar second-order probabilities to the same clone [5].

Jianhan Zhu et al. have proposed a clustering algorithm called citation cluster to cluster conceptually related pages. The clustering results are used to construct a conceptual hierarchy of the Web site. Markov model based link prediction is integrated with the hierarchy to assist users' navigation on the Web site. In the previous six years collection of user navigation session were presented in form of many models such as Hyper Text Probabilistic Grammar (HPG), N-Gram Model, Dynamic clustering based morkov model etc [6].

Nien-yijan et al. have proposed trend based application system to analyze user's behaviors and predict the future travelling 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 travelling 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 [7].

Prediction of user future movements and intentions based on the users' click stream data. Mehrdad Jalali et al. have developed a model for online predicting through WUM system and propose an approach for classifying user navigation patterns to predict users' future intentions. The approach is based on the using longest common subsequence algorithm to classify current user activities to predict user next movement [8].

The quality of recommendations in the current systems to predict user future requests in a particular Web site is below satisfaction. To effectively provide online prediction, A. Anitha proposed to integrate Markov model based sequential pattern mining with clustering. With the help of proposed approach approximately 12% of prediction accuracy increases compared to traditional Markov model. The main advantage of proposed hierarchical clustering approach is that every object must be candidate of only one cluster. The traditional Markov models have serious limitation that, low order Markov models have good coverage but they lack accuracy due to poor history and high order Markov models suffers from high state space complexity, as they use long browsing history, but high order. Markov models provides good prediction accuracy for that purpose in proposed approach combined the advantages of both Markov models and in order to improve the accuracy of prediction process, sequential mining used. We have summarized various method of WUM in next session. Web in how to quickly and accurately find the user-needed information and then provides personalized service to users The increasing competitive E-commerce gradually enhances the attention paid to E-marking. Web log mining is of great importance to E-commerce, distance education platforms today [9].

The discovery of the users' navigational patterns using SOM is proposed by Etminani et al. Huge amount of information are collected repeatedly by Web servers and gathered in access log files. Analysis of server access data can offer important and helpful data. WUM is the technique of using data mining procedure for discovering the usage patterns from Web data and is targeted towards applications. It extracts the secondary information resulting from the interactions of the users through some period of Web sessions. WUM involves three processes, namely pre-processing, pattern discovery, and pattern analysis. Provided its application possibility, WUM has seen a quick rise in interest, from the research and practice area. The author used the Kohonen's SOM to pre-processed Web logs for extracting the common patterns [10].

Zhang et al. presented a WUM technique based on fuzzy clustering in Identifying Target Group. Data mining is a process of non-trivial mining of inherent, previously unidentified, and highly helpful data from very large quantity of information. Web mining can be defined basically as the usage of data mining procedures to Web data. WUM is a significant and rapidly developing field of Web mining where many research has been performed previously. The



ISSN: 2319-5967

ISO 9001:2008 Certified

International Journal of Engineering Science and Innovative Technology (IJESIT)

Volume 2, Issue 5, September 2013

author uses the fuzzy clustering technique for discovering groups that share similar interests and behaviors by examining the data gathered in Web servers [11].

Rahman et al. have suggested a complete idea for the pattern discovery of WUM. Web site creators must have clear knowledge of user's profile and site intentions and also emphasized information of the approach users will browse Web site. The creators can examine the visitor's behavior by means of Web analysis and identify patterns of the visitor's activities. This Web analysis includes the transformation and interpretation of the Web log records to identify the hidden data or predictive pattern by the data mining and knowledge discovery process. This result provides a great view coupled with the Web warehousing [12].

Hung Wu et al. have given a WUM technique based on the sequences of clicking patterns in a grid computing environment. Examining user's browsing pattern is an significant process of WUM. It can assist the Web supervisors or creators enhance the Web structure or increase the performance of the Web servers. MSCP can be regarded as a data mining task. MSCP is generally an expensive procedure because of its significant quantity of time for computation and storage for archiving a large quantity of information. Running MSCP becomes ineffective or even not practical on a computer with restricted resources. The author discovers the usage of MSCP in a distributed grid computing surroundings and expresses its effectiveness by empirical cases [13].

Aghabozorgi et al. proposed the usage of incremental fuzzy clustering to WUM. Currently wide increase of information on the Web has produces a huge quantity of log records on Web server databases. By using the WUM procedure on this huge quantity of historical information can identify potentially helpful patterns and expose user access patterns on the Web page. Cluster analysis has broadly been used to produce user behavior pattern on server Web logs. The majority of these off-line procedure have the drawback of reduction of accuracy over time resulted of new users joining or modifications of pattern for present users in model-based techniques [14].

Maratea et al. have presented a new technique to produce dynamic model from off-line model produced by fuzzy clustering. In this technique, users' transactions are used periodically for modifying the off-line model. To this intend, an enhanced technique of leader clustering along with a static technique is used to regenerate clusters in an incremental fashion. Personalized Web page recommendation is strictly restricted by the nature of Web logs, the intrinsic complexity of the problem and the higher efficiency needs. When handled by existing WUM methods, because of the existence of an large number of meaningful clusters and profiles for visitors of a usually highly rated Website, the model-based or distance-based techniques are likely to create very strong and simple assumptions or, on the other hand, to turn out to be highly complex and slow. The author designed a heuristic majority intelligence technique, which effortlessly adjusts to changing navigational patterns; with the low cost explicitly individuate them ahead of navigation. The proposed technique imitates human behavior in an unidentified environment in occurrence of several individuals working in parallel and it has the ability to predict with better accuracy and in real time the next page group visited by a user. This technique has been checked on real data from users who browse a popular Website of common content. Average accuracy on test sets is better on a 17 class problem and, most importantly, it continues to be steady as the Web navigation goes on [15,16].

### III. APPLICATION OF WUM

Web usage mining is used in the following areas:

The collected public log file can be useful for following application.

- Target potential customers for electronic commerce
- Enhance the quality and delivery of Internet information services to the end user
- Improve Web server system performance
- Identify potential prime advertisement locations
- Facilitates personalization of sites
- Improve site design
- Predict user's actions (allows pre-fetching)

### IV. CONCLUSION

This paper has provided a more current evaluation and analysis of "web usage mining for browsing behavior of a user and subsequently to predict desired pages research available. Web usage mining is fast rising research area today generated log information can be useful in various ways.



ISSN: 2319-5967

ISO 9001:2008 Certified

International Journal of Engineering Science and Innovative Technology (IJESIT)

Volume 2, Issue 5, September 2013

#### REFERENCES

- [1] 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
- [2] 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.
- [3] A. Nanopoulos, D. Katsaros and Y. Manolopoulos, “Effective prediction of Web user accesses: A data mining approach”, In Proceeding of the Workshop WE-BKDD, Aristotle Univ. of Thessaloniki, Greece, 2001.
- [4] M. A. Bayir, I. H. Toroslu and A. Cosar, “A New Approach to Reactive Web Usage Data Processing”, The 2nd International Workshop on Challenges in Web Information Retrieval and Integration, ICDE 06’s Workshop, Turkey, 2006.
- [5] J. Borges and M. Levene, “A dynamic clustering-based markov model for Web usage Mining”, Journal of Computer Research Repository Report No: cs.IR/0406032, Canada, USA, 2004.
- [6] J. Zhu and J. Hong , “Using Markov Models for Web Site Link Prediction”, In Proceedings of the thirteenth ACM Conference on Hypertext and Hypermedia, College Park, Maryland, USA ACM, June 11-15, 2002.
- [7] N. yi jan and N. P. Lin, “Web User Behaviours Prediction System Using Trend Similarity”, In Proceedings of the 7th WSEAS International Conference on Simulation, Modelling and Optimization, Beijing, China, pp. 15-17, Sep. 2007.
- [8] Mehrdad Jalali and Norwati Mustapha, “ WebPUM: A Web-based recommendation system to predict user future movements”, In international journal Expert Systems with Applications, Malaysia, Vol. 37, pp. 6201-6212, 2010.
- [9] A. Anitha, “A New Web Usage Mining Approach for Next Page Access Prediction”, International Journal of Computer Applications, Tamil Nadu, India, Vol.8- No.11, October 2010.
- [10] Etminani, K., Delui, A.R., Yanehsari, N.R. and Rouhani, M., “Web Usage Mining: Discovery of the Users’ Navigational Patterns Using SOM”, In Proceedings of the First International Conference on Networked Digital Technologies, Mashhad, Iran, pp.224-249, 2009.
- [11] Jianxi Zhang, Peiyong Zhao, Lin Shang and Lunsheng Wang, “Web Usage Mining Based On Fuzzy Clustering in Identifying Target Group”, Proceedings of the International Colloquium on Computing, Communication, Control, and Management, Jinhan, China, Vol. 4, pp. 209-212, 2009.
- [12] Nina, S.P., Rahman, M., Bhuiyan, K.I. and Ahmed, K., “Pattern Discovery of Web Usage Mining” Proceedings of the International Conference on Computer Technology and Development, Sylhet, Bangladesh, Vol. 1, pp.499-503, 2009.
- [13] Chih-Hung Wu, Yen-Liang Wu, Yuan-Ming Chang and Ming-Hung Hung, “Web Usage Mining on the Sequences of Clicking Patterns in a Grid Computing Environment”, In International Conference on Machine Learning and Cybernetics (ICMLC), Kauhsiung, Taiwan, Vol. 6, pp. 2909-2914, 2010.
- [14] Aghabozorgi, S.R. and Wah, T.Y., “Using Incremental Fuzzy Clustering to Web Usage Mining”, In Proceedings of the International Conference of Soft Computing and Pattern Recognition, Kuala Lumpur, Malaysia, pp. 653-658, 2009.
- [15] Maratea, A. and Petrosino, A., “An Heuristic Approach to Page Recommendation in Web Usage Mining”, In Proceedings of the Ninth International Conference on Intelligent Systems Design and Applications, Naples, Italy, pp. 1043-1048, 2009.
- [16] Inbarani, H.H., Thangavel, K. and Pethalakshmi, A., “Rough Set Based Feature Selection for Web Usage Mining”, In Proceedings of the International Conference on Conference on Computational Intelligence and Multimedia Applications, Tamil Nadu, India, Vol.7, pp. 33-38, 2007.