

A SURVEY ON WEB USAGE MINING

Ankita Singh¹, Satya Prakash Singh², Meenu³

^{1,2}Department of Computer Science & Engineering,

Madan Mohan Malaviya University of Technology Gorakhpur (U.P.) (India)

³Assistant Professor, Department of Computer Science & Engineering,

Madan Mohan Malaviya University of Technology Gorakhpur (U.P.) (India)

ABSTRACT

Now a day World Wide Web becomes most popular and interactive for devolution of information. The web is giant distinct and active and thus increases the scalability, multimedia data and worldly matters. The progress of the web has issue in a giant amount of information that is now loosely offered for user access. The various types of data must be handled and managed in manner that they can be scope by various users effectively and efficiently. So, the usage of data mining manners and knowledge discovery on the web is now on the spotlight of a boosting number of researchers. Web usage mining is various of data mining manner that can be utilizable in recommending the web usage patterns with the help of users 'session and conduct. Web usage mining include three process, namely, preprocessing, pattern analysis and pattern analysis. There are different techniques already exists for web usage mining.

Keywords: Web Usage Mining, Preprocessing, Weblog, User Navigation

I.INTRODUCTION

Recently millions of electronic data are included on hundreds of millions of data that are previously on-line today. With this significant increase of existing data on the internet and because of its fast and chaos growth, the world-wide web has evolved into a network of data with no proper organization structure. In addition, survival of plentiful data in the network and the varying and heterogeneous nature of the web, web searching has become a tricky procedure for most the users. This makes the users feel confused and at time lost in overload data that persist to enlarge. Moreover, e-business and web marketing are quickly developing and significance of anticipate the requirement of their customers is evident particularly. As a result, guessing the users 'interests for rectify the usability of web or so called personalization has turn out to be very essential. Web personalization can be depicted as some action that builds the web experience of a user personalized according to the user's interest. Web usage mining is the type of data mining process for discovering the usage patterns from web information for understanding and better provide the requirements of web-based applications. Web usage mining simulations the action of human as they interact with the internet. Examination of user actions in communication with web site can offer insights causing to customization and personalization of a user's web practice. Because of this, web usage mining is of endmost attention for e-marketing and e-commerce professionals. Web usage mining included of three phases, namely, preprocessing, pattern discovery and pattern

analysis. There are different techniques available for web usage mining with its own advantages and disadvantages. The world-wide web is an immense source of data that can come either from the web content, represent by the billions of pages publicly available, or from the web usage, represented by the log information daily collected by all the servers around the world. Web mining is that area of data mining which merchandise with the extraction of interesting knowledge from the world-wide web. Fates of people begin their e-business over the web due to the ease and speed of their business affair using internet. This needs the study of people personal information, their vantage, their economy, and their buying patterns so that their marketing strategies can be improved according to the people and also can adduce the goods of their choice of interest. This leads to the web usage mining to find out the several data mining patterns like clustering, classification, association analysis which are required in the e-commerce application. Data mining used to the web is known as web mining. Basically, Web mining is divided into three parts: Web content mining, Web Structure mining and Web usage mining section2 deals with the classification of web mining.

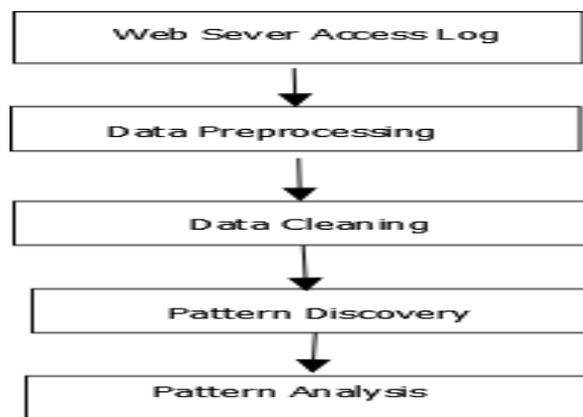


Fig. Web Usage Mining Phases

II. LITERATURE SURVEY

This section provides some discussion about various web usage mining techniques available today. The discovery of the user's navigational patterns using SOM is proposed Etmnani et al., [1]. Decuman amount of information are collected repeatedly by web servers and assembled adit log file. Analysis of server adit data can offer important and helpful data. Web usage mining is the techniques of using data mining procedure for discovering the usage patterns from web data and is targeted towards applications. It marrow the secondary information resulting from the interaction of the users through some period of web sessions. Web usage mining involves three processes, namely preprocessing, pattern analysis, pattern discovery. Provided its application possibility, Web usage mining has seen a quick raise in interest, from the research and practice area. The author used the Kohonen's SOM (Self Organizing Mape) to preprocessed web logs for extracting the common patterns. Jianxi et al., [2] presented a web usage mining 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.

Nina et al., [3] suggests a whole idea for the pattern discovery of web usage mining. 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.

Wu et al., [4] given a web usage mining technique based on the sequences of clicking patterns in a grid computing environment. Examining user's browsing pattern is a significant process of web usage mining. It can bestead the web supervisors or creators enhance the web structure or increase the performance of the web servers.

Aghabozorgi et al., [5] proposed the usage of incremental fuzzy clustering to web usage mining. Currently wide increase of information on the web has produces a decuman quantity of log records on web server databases.

[6] Personalized web page recommendation is precisian restricted by the nature of web logs, the intrinsic complexity of the problem and the higher efficiency needs.

Portly set based feature selection for web usage mining is proposed by Inbarani et al [7]. Web usage mining utilizes data mining methods to find out priceless data from navigation pattern of world wide web (WWW) users.

Jalali et al., [8] put ahead a web usage mining technique based on LCS algorithm for online predicting recommendation systems. The internet is one of the quickly developing fields of intelligence gathering.

For providing the online prediction effectively, Shinde et al., [9] provides an architecture for online recommendation for predicting in web usage mining system. The author presents the architecture of online recommendation in web usage mining (OLRWMS) for improving accuracy of classification by interaction between classification, Assessment, and present user activates and user profile in online phase of this architecture.

Zhang et al., [10] given an intelligent algorithm of data pre-processing in web usage mining. Web usage mining is the kind of data mining methods to web usage logs decuman web data stores for producing results used in some parts like web site design, web server design, user's categorization, creating adoptive web sites and web sites personalization.

Nasraoui et al., [11] provides an entire framework and finding in mining web usage navigation from web log files of a genuine web site which has every challenging characteristic of real life web usage mining, concurrently, with evolving user profiles and external data describing an antology of the web content. How their attentions modified with time, every one of which are significant questions in Customer Relationship management (CRM).

Temporal web usage mining includes application of data mining process on temporal web usage mining information for discovering temporal navigation that illustrates the temporal activities of web users. Hogo et al., [12] proposed the temporal web usage mining of web users on single educational web site with the help of the adapted kohonen SOM based on rough set property.

Along with the raped development of ecommerce, it is very necessary to recognize the user access mode. By means of web usage mining, the server log, registration data and other related data get from user access can be mined with the user access method that will afford basis for decision making of organization. DeMin et al., [13] presented a SQL server 2000 based web usage mining system.

A development of data preprocessing technique for web usage mining and information of algorithm for path closure are provided by Yan et al., [14]. Later user session detection, the missing pages in user access paths are added with the help of referrer based technique that is an effective solution to the problem because when proxy servers and local caching is used.

Sophisticated data mining techniques are expected for the knowledge to be extracted, understood and used. Baraglia et al., [15] Suggestion a web usage mining (WUM) system, called SUGGEST, which continuously creating the suggested connection to web pages of probable importance for user.

III.CONCLUSION

The paper presented an overview of the web usage mining to extract the usage patterns of the users over the web from the several log files. The also specified several data mining techniques used in the web usage mining, application of the web usage mining and various tools used in web mining. The web is a very essential means to carry out business and commerce. So, the design of web pages is highly essential for the system managers and web creators. These characteristics have decuman impact on the number of users who access the page. Therefore, the web analyzer has to examine with the data of server log file for identifying the navigation pattern.

REFERENCES

- [1] Etminani, K., Delui, A.R., Yanehsari, N.R. and Rouhani, M., "Web Usage Mining: Discovery of the Users' Navigational Patterns Using SOM", First International Conference on Networked Digital Technologies, Pp.224-249, 2009.
- [2] Jianxi Zhang, Peiyong Zhao, Lin Shang and Lunsheng Wang, "Web Usage Mining Based on Fuzzy Clustering in Identifying Target Group", International Colloquium on Computing, Communication, Control, and Management, Vol. 4, Pp. 209-212, 2009.
- [3] Nina, S.P., Rahman, M., Bhuiyan, K.I. and Ahmed, K., "Pattern Discovery of Web Usage Mining", International Conference on Computer Technology and Development, Vol. 1, Pp.499-503, 2009.
- [4] 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", International Conference on Machine Learning and Cybernetics (ICMLC), Vol. 6, Pp. 2909-2914, 2010.
- [5] Aghabozorgi, S.R. and Wah, T.Y., "Using Incremental Fuzzy Clustering to Web Usage Mining", International Conference of Soft Computing and Pattern Recognition, Pp. 653- 658, 2009.
- [6] Maratea, A. and Petrosino, A., "An Heuristic Approach to Page Recommendation in Web Usage Mining", Ninth International Conference Conference on Computational Intelligence and Multimedia Applications, Vol. 1, Pp. 33-38, 2007
- [7] Inbarani, H.H., Thangavel, K. and Pethalakshmi, A., "Rough Set Based Feature Selection for Web Usage Mining", International Conference on Conference on Computational Intelligence and Multimedia Applications, Vol. 1, Pp. 33-38, 2007.

- [8] Jalali, M., Mustapha, N., Sulaiman, N.B. and Mamat, A., "A Web Usage Mining Approach Based on LCS Algorithm in Online Predicting Recommendation Systems", 12th International Conference Information Visualisation, Pp. 302- 307, 2008
- [9] Shinde, S.K. and Kulkarni, U.V., "A New Approach for on Line Recommender System in Web Usage Mining", International Conference on Advanced Computer Theory and Engineering, Pp. 973- 977, 2008.
- [10] Zhang Huiying and Liang Wei, "An intelligent algorithm of data pre-processing in Web usage mining", Fifth World Congress on Intelligent Control and Automation, Vol. 4, 3119- 3123, 2004.
- [11] Nasraoui, O., Soliman, M., Saka, E., Badia, A. and Germain, R., "A Web Usage Mining Framework for Mining Evolving User Profiles in Dynamic Web Sites", IEEE Transactions on Knowledge and Data Engineering, Vol. 20, No. 2, Pp. 202-215, 2008.
- [12] Hogo, M., Snorek, M. and Lingras, P., "Temporal Web usage mining", International Conference on Web Intelligence, Pp. 450-453, 2003.
- [13] DeMin Dong, "Exploration on Web Usage Mining and its Application", International Workshop on Intelligent Systems and Applications, Pp. 1-4, 2009.
- [14] Yan Li, Boqin Feng and Qinjiao Mao, "Research on Path Completion Technique in Web Usage Mining", International Symposium on Computer Science and Computational Technology, Vol. 1, Pp. 554-559, 2008.
- [15] Baraglia, R. and Palmerini, P., "SUGGEST: a Web usage mining system", International Conference on Information Technology: Coding and Computing, Pp. 282-287, 2002.
- [16] Jian Chen, Jian Yin, Tung, A.K.H. and Bin Liu, "Discovering Web usage patterns by mining cross-transaction association rules", International Conference on Machine Learning and Cybernetics, Vol. 5, Pp. 2655-2660, 2004.
- [17] Wu, K.L., Yu, P. S. and Ballman, A., "SpeedTracer: A Web usage mining and analysis tool", IBM Systems Journal, Vol. 37, No. 1, Pp. 89-105,
- [18] TheintTheint Aye, "Web Log Cleaning for Mining of Web Usage Patterns" IEEE, 2011
- [19] Omer Adel Nassar, Dr. Nedhal A. Al Saiyd" The Integrating Between Web Usage Mining and Data Mining Techniques", International Conference on Computer Science and Information Technology IEEE, 2013.