A Study of Web Usage Mining Applications and its Future Trends

[1] D. Uma Maheswari, [2] Dr. A. Marimuthu

D.Uma Maheswari, Research Scholar,Karpagam University,Coimbatore, Tamil Nadu.
Dr. A. Marimuthu, Associate Professor, Department of Computer
Science, Government Arts College, Coimbatore. Tamil Nadu

Abstract

Web usage mining is the research area of web mining is used to predict the web user behavior interacts with the web. The objective of mining is to find the user's access behavior on the web. This paper provides an area of web usage mining, with future research trends as well as commercial offerings.

Keywords - **Data Mining, Web Mining, Web usage mining, Web Log, Web Personalization, Web site structure,**

1. Introduction

The web is a huge and hard to estimate the growth of web data every day, the web provides different kind of services such as government, electronic commerce, news, etc. Mining of the web is interesting and potentially useful patterns of implicit information from World Wide Web. Web usage mining supports for the creation of web site designing; provide personalization web server and other business decision, etc. Web usage mining consists of four steps. Web usage mining consists of three stages, namely pre-processing, pattern discovery, and pattern analysis. The first step is data collection, the second step is preprocessing, the third step is pattern discovery and the final step is pattern analysis. Web usage mining is the application of data mining. : The information that formulates the hyper-link structure of a website i.e., various HTML tags used to link one page to another page and one website to another. Web usage: The information that reflects the usages of Web resources, i.e., entries in Web browser's history and Internet temporary files, proxy server and Web server logs.

The information that provides demographic information about users of the website, i.e., user's registration data and customer's profile information.

Bulent Ozel et al., [2] proposed hybrid of web link prediction. Growing web is raising the navigational problems. Using web usage can predict the user behavior and help the designer to improve the design for attracting the user's usage. In this paper the author used both association rule mining technique and Markov chain model. The processes of hybrid link cluster the similar pages for increasing the efficiency of the proposed model. Yoon Ho Cho et al., [3] presented the fast growing of e-commerce shows the customer demand on the web. In this paper e-commerce caused the overload of the customer on the web in a long time.

To overcome this overload, varies method have been used. The author used the Collaborative filtering is the method has recommended overcoming the limitations on the existing method. Rana et al., [4] focused a techniques could predict the behavior of the user's while interacting with the website. The approach is multi-disciplinary that the web is continued in growing. Thus can able to predict the user need to make them robust, scalable and efficient design. The data generated by surfer's sessions or user behaviors. They discussed the tools available in the applications of web usage mining and concluded with the challenges and future trends in the research. Zaïane *et al*. (1998) developed the knowledge discovery WeblogMiner tool from the server log files. Thus can improve the system performance, and enhance the quality website and deliver the useful data to the end users.

2.Literature Survey

A Study - Web Usage Mining Research Tools is given by Chhavi Rana et al., [21] the author presented the requirement of today's world and listed some challenges and issues of future trends of the users getting attention to the quality website. Abraham et al., [5] has combined the data mining and the World Wide Web for research. The knowledge discovery could attempt to obtain from the secondary data. The author used i-miner tool and to optimize the concurrent architecture of a fuzzy clustering algorithm for discovering information cluster to analyze the trends the fuzzy inference is used. Cooley et al., [7] described the web usage mining interesting patterns can be discovered from the uninteresting order. Several

research efforts have relied of uninteresting rules. The Web Site Information Filter (WebSIFT) system usages the web content and web structure information from the website to identify the interesting patterns for mining frequent item sets from the real world.Web usage mining is the application of data mining techniques to discover the patterns from web, in order to understand the web-based applications. Data on the web can be mined from secondary data on the web. We could classify the data that reside on the web [6]. INSITE Shahabi et al., [8] To track the user interaction with the web and produce user profile in real time by the use of Connectivity Matrix Model (CM-Model), that shows the efficiency and scalability of the user's participatory attributes on the web to visualize the user navigation path in real time.By using graph the web usage data could represent [9]. Frequently the web usage mining methods uses some background knowledge such as web content, website topology, Hierarchies, user navigation and constraints. In this paper the author proposed a new CF-based recommendation methodology that addressing the product overload problem in large E-commerce sites [10]. A Web Usage Mining Framework for Mining Evolving User Profiles in Dynamic Web Sites by Olfa Nasraoui et al., [11] the author presented a complete framework to find the user profiles for discovering patterns from the log files of the original website. The user based interest is analyzed on the website by search queries to extract the web log data. LChen et al., [12] developed WebMate; a proxy agent that helps to the user as effective browsing and searching on the web. The author experimentally described the existing and proposed systems of user profile used in various business and application point of view. A survey on web usage mining has been done by Koutri, Avouris, and Daskalaki [13].Building the user model by using the WebLogMiner techniques used for uncovering the hidden patterns within the web. Web access information is stored in the data sources. The author used these techniques for building the adaptive web site.SEWeP Eirinaki M. et al., [14] discussed the needs of the user by analysis of the navigational behavior on the web. In order to personalization the author developed as a system that makes both the web usage logs and web site contents that encapsulates knowledge discovery from the link semantics. Buchner et al., [15] introduced new algorithm called MiDAS in the wide range of web can discover the traditional sequential discovery. Thus shows the functionality as well as scalability. Ling et ai., [17] Direct marketing is the process of identify the product who purchased and who sold. Data mining tool is used for direct marketing.

Web related information is appropriate and popular target for knowledge discovery, the knowledge discovery process concerns the web content, web structure and web usage. From the user interest the quality website can develop and discover the rules and patterns for taking some interesting measures [1]. Senkul and Salin et al., [18] generated the web usage mining techniques for web site restructuring and recommendation. The author investigated the semantic data of the web page on the patterns are generated for frequent sequences. The frequent user navigational pattern is measured by mechanism involving web page recommendation. SpeedTracer et al., [19] developed user surfing behavior by server log files. The author found the user session by reconstructing the traversal path through referrer page and generated three types of reports are prepared. Rao, Kumari, and Raju, [20] developed association rule mining with incremental method. The author used this algorithm to suit the dynamic changing log scenario with incremental techniques. This technique is more efficient to run a number of databases. Ujwala Patil et al.,[23] proposed a survey or future request prediction of the extraction from web log files .Web is the fast rising research area the log files is very useful to predict the behavior of the user in different ways. They provided the past, current evaluation and updating of web usage mining. Extraction of Business Rules from Web logs to Improve Web Usage Mining is given by Sawan Bhawaret al., [24]. The author presented the automatic data extracted by querying, organizing and analyzing. Information Extraction is the task of extracting structured information from unstructured one. Input is the log files and extracted the information from session performed by the user. This can improve the web usage. The author proposed the website with the limited web pages can easily predict the user demand from the particular web sites.

3. Applications of Web Usage Mining

Web usage mining has various application areas such as web prefetching, site reorganization, web personalization, system improvement, link prediction, Business Intelligence and Usage Characterization.

3.1. Web Prefetching

The performance of the system is very important for user satisfaction. Web usage mining is an important research area for detecting the web traffic. The communications between PC and the server constitute Web traffic. The

amount of traffic and the details of each visit are extremely valuable information to a Web-based business. The server computer records every request for a Web page by user, and determines which pages get the most attention. Web traffic analysis gives businesses concrete, reliable information on the interests of their customers. The more traffic a Web site receives, the more sessions and hits its server processes. Every time a Web server processes a file request, the computer makes an entry in a server log, a dedicated file on the server's hard. To develop the server performance new policies can be used. Downloading files can slow down the user experience. If a user scrolls through an application screen and has to wait for content to load, the application appears slow to them. To use prefetching effectively, you need to evaluate the content application uses in order to determine meaningful indexes that identify which content is appropriate for prefetching. This prefetching approach is useful for both client and server level web caching, load balancing, transmission of data distributed information are the applications of web mining.

3.2. Website Reorganization

Attractiveness of the website is an important one gives good structure of the website. The principle of website reorganization is first need to understand how users interact with web-sites, how they think and what the basic patterns of users' behavior. In website navigation, the structure of the website can rearrange. The relationships between web pages are dynamically updated. Reorganization can be performed with the extraction of frequent patterns of web usage mining. The web usage information gives the information about the user behavior's of any website. Both content and structure leads to adaptive web site.

3.3. Website Personalization

Website Personalization is gathering the personal user data, some privacy issues is a major anxiety. Personalization Consortium endorses and guide the growth of one-to-one marketing carried out. Behind of personalization technologies include collaborative filtering, in which filtering is applied with different sites for selecting the relevant information that may apply to the specific group of customers. User profiling information is collected from different web sites can result of personalized web page .By using data analysis can be predict the users future prediction and find the session with multiple websites with interesting links.

3.4. System Improvement

The major system improvement life cycle is planning, analysis, development and implementation. It should support the user demand to build a system. Developing

the system with the security can avoid the intrusion and to restrict the user's access to certain online contents. Understand the customer demand and retaining the customized products. Improve some satisfaction with the help of browsing behavior.

3.5. Link Prediction

Link prediction is used for analyzing the nodes in a network, from the large network suggest that information can be extracted from the network topology.

3.6. Business Intelligence

Web usage mining provides data to improve the customer, sales and marketing field. It is the technology to access the data from various data sources, for business advantage, the data is gathered, stored and analyzed in organization can improve the customer needs and demands. Some decision about the business can be made to success. Thus the disciplines of Business intelligence includes decision support, data mining, online analytical processing (OLAP), querying and reporting, statistical analysis and forecasting.Some of the business intelligence tools are BizzScore Suite, IBM Cognous Series 10,WebFOCUS,QlikView,Tableau Software, Style Intelligence, Board Management Intelligence Toolkit, AS Enterprise BI Server can retrieve, analyze and generate report.

3.7.Usage Characterization

The usage of web used by the user is for various purposes. By characterizing the data usage of heavy users and normal users, and classify them and clusters according to their usage activities. The user behavior can be observed by usage regularities on the website. Characterize the users by navigational patterns and agent based approach.

4.Future Trends

The World Wide Web contains an enormous amount of data. The users of WWW need to use intelligent tools to find, sort, and filter the available data. The Web mining goals is to find and retrieve the information on the Web. By using algorithms we can find the work on web data. The user needs some tools to find the necessary data for usage patterns. This provides the quick development of master and slave intelligent system that can mine the information. Analyzing the log file is very difficult process. it is very hard to find the apt tool. Numerous tools are available but the users not liked and considered as slow process, rigid, costly and feel very hard to maintain. But some tools are using for data mining system for helping log analyze, the existing techniques have numerous i.e. moreover massive storage space, too much I/O cost, Majority of the available tools provide

clearly and statistical data. Its mission becomes more challenging one when the web is growing. Website information collection tool is huge as ease, and interactive. The data mining result would combine with dynamic website to give computerized, end-to-end task to the customer Relationship Management (CRM) and marketing systems. Mining tools in different side is important for future development. Web usage mining should handle integration of offline information with ecommerce tools, Relational Database Management, provides products catalogs and services with other applications, Few new variables are required for finding natural and meaningful and useful patterns. New tool are wanted for much process time for the period of mining practice. Standard tests always need for mining algorithms to develop the performance because of effectiveness could measure a better tool for web data mining. To develop visualization is an important for the user can understand the data. Mining of web is novel and quick developing research and application area. With research across different order such as software Engineering, marketing, artificial intelligence, database would be able to develop applications of web mining.

5. Conclusion

This paper insight the possibility of merging data mining techniques with logs for achieving a web usage mining and web application evaluations. Data mining applications are widely used in various areas such as e-business, education, telecommunications, financial, engineering, biotechnology, counter-terrorism, etc. Web mining is used for most of these applications .Web usage mining consists of three phases, namely pre-processing, pattern discovery, and pattern analysis. The study of web usage mining would help a lot of building tools that support for designing, development and maintenance. It can be applied to understand the user behavior. Through the analysis of patterns can able to develop the efficient design and powerful analysis tool.

6.References

[1] J. Borges, and M. Levene, Data Mining of User Navigation Patterns' Web Usage Analysis and User Profiling, *San Diego, CA, USA, 2000, 31-39.*
[2] Bulent Ozel and Samil Aydogan, A Hybrid Web Link Prediction System' Proceedings of 2nd International Conference on Intelligent Knowledge Systems (IKS-2005), 06-08 July 2005.

[3] Yoon Ho Cho, Jae Kyeong Kim, Application of Web usage mining and product taxonomy to collaborative recommendations in e-commerce, *Expert Systems with Applications ,Volume 26, Issue 2, February 2004, Pages 233–246.*

[4] Rana and Chhavi, A Study of Web Usage Mining Research Tools' International Journal of Advanced Networking & Applications, Vol. 3 Issue 6, May/Jun2012, p1422.

[5] Abraham, A. (2003) 'i-Miner: A Web Usage Mining Framework Using Hierarchical Intelligent Systems', *IEEE International Conference on Fuzzy Systems FUZZ-IEEE'03*, IEEE Press , pp. 1129-1134 .

[6] J. Srivastava, R. Cooley, M. Deshpande, P. Tan, Web usage mining: Discovery and applications of usage patterns from web data' *SIGKDD Explorations newsletter, 1(2), 2000, 12-23.*

[7] Cooley, R., Tan, P-N. and Srivastava, J. , 'WebSIFT: The Web Site Information Filter System',*Proceedings of Workshop on Web Usage Analysis and User Profiling WEBKDD in conjunction with ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 1999,* San Diego, California, USA.

[8] Shahabi, C. Faisal, A. Kashani, F. B. and Faruque, J,'INSITE: A Tool for Real-Time Knowledge Discovery from Users Web Navigation', *Proceedings of the 26th International Conference on Very Large Databases (VLDB),* Cairo, Egypt, pp. 635-638,2000.

[9] A. G. Buchner, M. Baumgarten, , S. S Anand, M. D. Mulvenna, and J. G. Hughes, Navigation Pattern Discovery from Internet Data, *Proceedings of Web Usage Analysis and User Profiling at the International WEBKDD99 Workshop*, 2000, 74- 91.

[10] Seok Kee Lee, Young-Gab Kim, and Do-Gil Lee, A Collaborative Recommender System: Lexicographic. Consensus and Web Usage Mining Approach' *ASTL Volume 2, 2012*.

[11] Olfa Nasraoui, Maha Soliman, Esin Saka, Antonio Badia and Richard Germain ,A web usage mining framework for mining evolving user profiles in dynamic web sites'*Journal IEEE Transactions on Knowledge and Data Engineering ,Volume 20 Issue 2, February 2008,Pages 202-215.*

[12] L. Chen, and K. Sycara, WebMate: A Personal Agent for Browsing and Searching, *Proceedings of the 2nd International Conference on Autonomous Agents*, Minneapolis MN, USA, 1999, 132-139.

[13] M. Koutri, N. Avouris, and S. Daskalaki, A survey on web usage mining techniques for web-based adaptive hypermedia systems, *Adaptable and Adaptive Hypermedia Systems Idea*, 2004, 1-23.

[14] Eirinaki, M., Vazirgiannis, M. and Varlamis, I. (2003) 'SEWeP: using site semantics and a taxonomy to enhance the Web personalization process', *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining,* pp. 99-108.

[15] A. G. Buchner, M. Baumgarten, , S. S Anand, M. D. Mulvenna, and J. G. Hughes, Navigation Pattern Discovery from Internet Data, *Proceedings of Web Usage Analysis and User*

www.ijert.org

Profiling at the International WEBKDD99 Workshop, 2000, 74-91.

[16] Zaïane, O.R, Xin, M. & Han, J. (1998) Discovering Web Access Patterns and Trends by Applying OLAP and Data Mining Technology on Web Logs, *Proc. Advances in Digital Libraries Conf.*, pp. 19-29.

[17] Ling, C.X. & Li, C. (1998) Data Mining for Direct Marketing: Problems and Solutions, *Proc. KDD'99*, pp. 73-79. Mulvenna, M.D., Norwood, M.T. & Büchner, A.G. (1998) Data-driven Marketing, *The Int'l Journal of El ectronic Commerce and Business Media*, **8**(3):32-35.

[18] P. Senkul, and S. Salin, Improving Pattern Quality in Web Usage Mining by Using Semantic Information, Knowledge and Information Systems, 1(6), 2011, 400- 404.

[19] Wu, K. L., Yu, P. S. and Ballman, A. (1998) 'SpeedTracer: A Web usage mining and analysis tool',
IBM Systems Journal on Internet Computing, Vol. 37, pp. 89-105.

[20] M. Rao, M. Kumari, and K. Raju, Understanding User Behavior using Web Usage Mining', international *Journal of Computer Applications,* 1(7), 2010, 55–61.

[21] Chhavi Rana,A Study of Web Usage Mining Research Tools, *Int. J. Advanced Networking and Applications Volume:03 Issue:06 Pages:1422-1429 (2012) ISSN : 0975-0290.*

[22] Myra Spiliopoulou,Carsten Pohle and Lukas C.Faulstich ,Improving the Effectiveness of a web site with Web Usage Mining'

[23] Ujwala Patil and Sachin Pardeshi,   A Survey on User Future Request Prediction: Web Usage Mining' International Journal of Emerging Technology and Advanced Engineering ,ISSN 2250-2459, Volume 2, Issue 3, March 2012.

[24] Sawan Bhawsar, Kshitij Pathak, Sourabh Mariya and Sunil Parihar, Extraction of Business Rules from Web logs to Improve Web Usage Mining' *International Journal of Emerging Technology and Advanced Engineering ,ISSN 2250-2459, Volume 2, Issue 8, August 2012.*

http://www.passionned.com/business-intelligence/business-intelligence-tools/list-of-business-intelligence-bi-tools/