International Journal of Emerging Trends in Science and Technology

Open access Journal

A Survey on Web Usage Mining

Authors

S.Geetharani[1], S.Priyadharshini[2]

[1]PSG College of Arts & Science, Coimbatore
Email: *Geetharani05464@gmail.com*
[2]PSG College of Arts & Science, Coimbatore
Email: *Priyadharshini1327@gmail.com*

Abstract
Web usage mining is the third category in the web mining and it is the process of extracting useful information from the server logs. It is valuable not only to the businesses using online marketing, but also to the e-business. This types of web mining is used to gather the important information from customers visiting the site. The companies can study visitors activities through web analysis. This process is used to find out what users are looking for on the internet. In this some users might be looking for the textual data and some users might be looking for the multimedia data. Usage mining is classified into different kinds of data. Web server data, application server data and application level data. First, In web server data the user logs are collected by the web server. Second, The key feature of the application server data is the ability to track the various kinds of business events and log them in the application server logs. Finally, application level data is used to define an new kinds of events in an application.
Keywords: *server logs, web analysis, web server data, application server data, application level data*

1. INTRODUCTION

Usage mining is the main research area in web mining and it's focused on the web users and their interactions with the web sites. There are typically three main users for mining. The first, Web usage mining process is used to complete pattern discovery. This usage process contains minimal amount of information, so it is harder to track the user through a site, second is web content mining consisting of conversion of web information like text images, scripts. It is differentiated from two points of view information retrieval and database view. Web mining is an important component of content pipeline for web portals and it is used in data confirmation and validity, verification, data integrity and building taxonomies, content management, content generation and opinion mining. This helps with the clustering and categorization of web page information based on the titles specific content and images available. Finally, web structure mining analysis the structure of each page contained in the website. It uses the graph theory to analyze the node and connection structure of the website. It can be divided into two kinds: Extracting patterns from hyperlinks in the web and mining the document structure. Therefore, data usage mining has valuable uses to marketing of businesses and a direct impact to the success of their promotional strategies and internet traffic. This information is gathered on a daily basis and continues to be analyzed consistently. The gathered information will help companies to develop promotions that are more effective, internet accessibility, inter-company communication and structure and productive marketing skills through web usage mining.

2. LITERATURE SURVEY

This section provides some discussion about several web usage mining techniques available today. Suggested that HTML documents contain more number of images on the WWW. Such documents' containing meaningful images ensures a rich source of images cluster for which query can be generated. The documents which are highly needed by users can be placed near to the home page of the website. Manoj Manuja and Deepak Garg [1].Suggested that the development of web mining techniques such as web metrics and measurements, web service optimization, process mining etc… will enable the power of WWW to be realized. Jing Wang and others [2].Nina et al., suggests a complete 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. 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. It is applied for deciding business approaches via the competent use of Web Applications. It is very vital for the Customer Relationship Management (CRM) since it can guarantee customer fulfillment till the interface between the customer and the organization is concerned [3].

Data preprocessing transforms data into a format that will be more easily, and efficiently processed for the purpose of the user. The main task of data preprocessing is to select standardized data from the original log files, prepared for user navigation pattern discovery algorithm [4]. A World Wide Web usage mining and examination tool called Speed Tracer, was created by Wu et al., [5] in order to realize user browsing pattern by investigating the Web server log files with data mining procedures. As the attractiveness of the Web has exploded, there is a powerful need to recognize user browsing pattern. Conversely, it is complex to carry out user-oriented data mining and analysis straightforwardly on the server log files since they inclined to be vague and deficient. With innovative technique, Speed Tracer initially recognize user sessions by rebuilding the user traversal paths. It does not need cookies or user registration for the purpose of session identification. User privacy is protected. Onceuser sessions are recognized, data mining techniquesare then used to determine the very common traversal paths and groups of pages regularly visited simultaneously. The essential user navigation patterns are identified from the frequent traversal paths and page groups, assisting the understanding of user browsing pattern. Three kinds of reports are organized: user-based reports, path-based reports and group based reports. The author illustrates the design of Speed Tracer and shows some of its features with a little sample reports.

Nasraoui et al., [6] provides a whole framework and findings in mining Web usage navigation from Web log files of a genuine Web site which has every challenging characteristics of real-life Web usage mining, together with evolving user profiles and external data describing an ontology of the Web content. Although the Web site considered is element of a nonprofit organization that does not sell any products, it was essential to recognize who the users were, what they looked at, and how their attentions modified with time, every one of which are significant questions in Customer Relationship Management (CRM). Therefore the author provides a technique for identifying and tracing growing user profiles. The author also illustrates how the discovered user profiles can be enriched with explicit data need which is gathered from search queries extracted from Web log data. Profiles are also enhanced with other domain-specific data features that provide a panoramic view of the discovered mass usage modes. An objective validation approach is also applied to evaluate the excellence of the mined profiles, in specific their adaptability in the face of evolving user pattern. Content adaptation on the Web decreases the existing data to a subset that contests a user's anticipated requirements. Recommender techniques based on relevance scores for individual content

items; particularly, pattern-based recommendation uses co occurrences of items in user sessions to view any prediction about relevancy. To improve the discovered patterns' quality, Adda et al., [7] presents a technique with the help of metadata about the content that they imagine is stored in domain ontology. This technique includes a dedicated pattern space constructed on top of the ontology, navigation primitives, mining procedure and recommendation methods.

The discovery of the users' navigational patterns using SOM is proposed by Etminani et al., [8].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. Web usage mining 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. Web usage mining involves three processes, namely preprocessing, pattern discovery, and pattern analysis. Provided its application possibility, Web usage mining has seen a quick rise in interest, from the research and practice area. The author used the Kohonen's SOM (Self Organizing Map) to preprocessed Web logs for extracting the common patterns.

3. CONCLUSION

The internet has grown from a simple search tool to a source of supply. Companies find a new and better way to do business. Web mining is used to find the customers preference and satisfy the customers. When the companies realize the value and importance of the web mining and adopt a web mining strategy. This is used to identify the strength and weakness of the customers.

REFERENCES

1. Manoj Manuja and Deepak Garg, Semantic web mining of Un-structured Data: Challenges and Opportunities, International Journal of Engineering, 5(3) ,2011, 268-276.

2. Jing Wang, Ying Liu, Lin Zhou, Yong Shi, and Xingquan Zhu, Pushing frequency constraint to utility Mining Model, ICCS Springer-Verlag Berlin Heidlberg, LNCS 4489, 2007, 685-692

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. ShaileyMinocha, Nicola Millard, Lisa Dawson, "Integrating Customer Relationship Management Strategies in (B2C) ECommerce Environments", IFIP Conference on HumanComputer Interaction-INTERACT, 2003.

5. Chaoyang Xiang, Shenghui He and Lei Chen, "A Studying System Based On Web Mining", IEEE International Symposium On Intelligent Ubiquitous Computing and Education, pp.433-435, 2009.

6. 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.

7. Adda, M., Valtchev, P., Missaoui, R. and Djeraba, C., "Toward Recommendation Based on Ontology-Powered Web-Usage Mining", IEEE Internet Computing, Vol. 11, No. 4, Pp. 45-52, 2007.

8. 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.