Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 8, August 2014, pg.769 – 773

RESEARCH ARTICLE

Predicting Web User's Behavior Patterns

Muhammad Ramzan Talib[1], Safi Ullah Hidayat[2], Muhammad Umer Sarwar[3]

[1,2,3] Department of Computer Science, Government College University, Faisalabad, Pakistan

Abstract:

Web user prediction is a process of predicting the future set of web pages that a web user can visit based on the knowledge of formerly visited pages. Predict the behavior of a web user during internet surfing can be implemented successfully in some important applications and these applications can have the conventional tradeoffs betwixt the complexity modeling and accuracy prediction. Searching the appropriate information on the internet is mostly a strenuous and annoying task for the web users. In this context researcher are spending their efforts to introduce new and improved prediction methods to help the web users in finding relevant information. In this paper an attempt has been made to review the different patterns for predicting web user behavior that the researchers have suggested and derived after the analysis of web log data.

Keywords: Web user behavior, Web search behavior, Web log data

Introduction:

Web user prediction involves in analyzing internet server logs that have detailed knowledge of customer browsing activities. Over the past few years to inspect the web servers' log several research efforts have been focusing and concentrating on the efficient and effective ways [1].

Web mining is a significant approach to take out functional information from the web data and identify the user behavior on a web page. With the help of extracted information researchers can identify the patterns about network users' behavior, traffic and interaction between web-pages and web users. Additionally, these patterns are used to facilitate the strategies making in business management such as internet site maintaining, directional marketing, page personalization and so on. The web mining is classified into three types on the basis of data types: (1) web content mining (2) web usage mining (3) web structure mining. Web content mining (WCM) concentrates on the real information in the web pages such as graphics and text. Web usage mining (WUM) analyzes web logs counting page references, timestamps and IP addresses and so on. Web structure mining (WSM) analyzes the hyperlinks to discover web pages structure knowledge [2]. Mining of interesting information from internet data has turn out to be more famous and as a result of that internet mining is a large application of data mining you huge internet data repositories [3]. Basic principle of web mining is to find out the useful knowledge and information from those of network hyperlink structure, data used and page content. Web mining used the technology of data mining, due to the web data heterogeneity and semi-structured or structured nature [4].

Methodology:

The process that we adopted in this study is based on the research skills of the authors. For source selections certain limitations have been considered: studies incorporated in the selected sources must be relevant to our topic and these sources must be available on web. The review protocol is designed by using above mentioned keywords; web-users behavior, web search behavior and web usage mining and web log data. The sources considered to conduct this review includes: Springer, Elsevier Ltd, IEEE and JEDM. An additional step in the search process is performed by searching the related work area of the selected papers to boost the review potency by confirming that no helpful reference is fails to notice during the search process. Once the sources had been defined, it was necessary to describe the process and the criteria for study selection and evaluation. The inclusion and exclusion criterion for this study is strictly limited to studies that contain content regarding the web-users predictions behaviors, web-users interests and web-users activities.

Results and Discussion:

The most two most familiar models of web user prediction are: tree-based and sequence-based models. The HPG (Hypertext Probability Grammar) with Markov chain statement is a type of model which is sequence-based, that propose the next chain is predictable by the possibility of a given production. Hypertext Probability Grammar with concept hierarchy is a model which is tree-based proposes the mining efficacy control by the configuration of website. These studies analyze the HPG and apply matrix technique of clustering to dig out the visual representation of customer navigation behavior. The HPG shows a collection of sessions assemble by the web log analyzing to supply a model that shows a hypertext probability system traversed by the customer. It has been effectively used in different algorithms and actual applications including web usage outline mining, Markov assumptions, web navigation guide, and others [1]. An approach for web server' log analysis is also proposed in [1]. The main objective of this approach is: rearrange the web site into the hierarchical structure to increase query efficiency, incorporate the semantic constraints with the hierarchical construction to allow the customer to use normal language as an inquiry input to database and construct the customer profile matrix and compress the matrix into clusters of customer who united the common characteristics. In [2] the two-step K-means cluster algorithm proposed to web usage mining patterns with four attributes to dig out from the collected information. The four attributes are: specifically relative entropy, category number, page access times and element of categories are able to reflect consumers' activities, preference and information scope which can exemplify users' requirements in various aspects. But conventionally, web usage mining (WUM) technologies always analyze information collected from a single web site, when relate those approaches in to numerous web sites, the entire consumer's behaviors can be explained and the exposed pattern be capable of referred by on the whole web sites. Therefore, web usage mining bears the challenges of information availability, diversity of user's behaviors, mining efficiency, evolving usage patterns and features selection. In [3] web page prediction representation was designed to give the respond of web user's next contact based on model of Markov and clustering via association rule with algorithm to provide a better prediction of client's next page access as shows in fig 1.

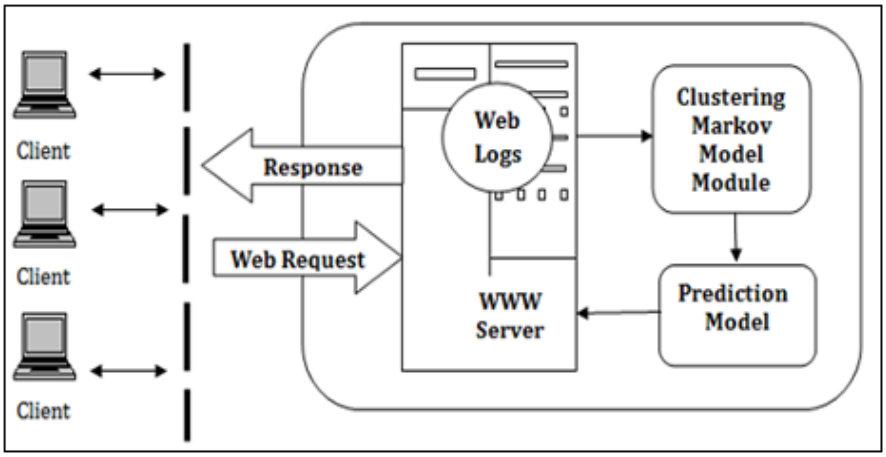

Figure 1: General architecture of web system with web access prediction

Precisely predicting web customer access behavior can reduce the user perceived latency that is crucial in the quickly growing World Wide Web. While traditional models of Markov have helped to calculate user's access behavior. These models have very serious boundaries. So the novel clustering Markov model with algorithm accessible to analyze and evaluate for the prediction of web access accurately which gives high accuracy.

In [5] analysis of users' behavior on web 2.0 sites has been done by taking data from data three social networking sites. In that analysis the creators of SNS are interviewed for the recognition of Google analytics and daily data of these sites are collected. The purpose of recognition of Google analytics was: Determine the vital measures and examine the relation between them, find out the commonalities throughout these three case companies and give administrative implications primarily based on statistical outcomes. The metrics that used in Google analytics were: Bounce rate which is the percentage of a page visits, Visitor which is a user that visits the page or site, % NV is a number of new visits by users who never visited the page or site before, ATOP is average time on a page, PV is pages number divided on the basis of visits, ATOS, which is the average time on site, Visitors is the quantity of time in which visitor has been on site.

Conclusion:

Web user prediction has become very significant in order to facilitate the web users to find relevant information. Though, identifying different user's behavior pattern by analyzing web log data is a tough and costly. But there is still a lot more work needs to be done on this area.

References:

[1]. Christoph Hölscher, Gerhard Strube, Web search behavior of Internet experts and newbies, Computer Networks, Volume 33, Issues 1–6, June 2000, Pages 337-346.

[2]. Hao Y., Bo Z., Yibo Z., Fang L. & Zhenming L. (2010) Web usage mining based on WAN users' behaviors 111 Project of Chinaunder Grant No. B08004 and the Fundamental Research Funds for the Central Universities (BUPT2009RC0123),IEEE.

[3]. Priyanka S. P. & Urmi D. A. (2013) Hybrid Technique for User's Web Page Access Prediction based on Markov Model, IEEE – 31661.

[4]. Jia, L. (2013). Advanced Information Retrieval within Blogosphere and Micro-Blogosphere (Doctoral dissertation, University of Illinois at Chicago).

[5]. Shwu-M. H. (2010) Analysis of User's Behavior on Web 2.0 Social Network Sites: An Empirical Study Seventh International Conference on Information Technology, IEEE.

