iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** VII **Month of publication:** July 2018

DOI: http://doi.org/10.22214/ijraset.2018.7124

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue VII, July 2018- Available at www.ijraset.com

Research on Web Log Mining to Predicting User Behavior through Session

Komal S. Raut[1,] Dr. S .E.Yedey[2]
[1, 2]*PG Department of Computer Science & Technology, Hanuman Vyayam Prasarak Mandal Amravati, Maharashta*

Abstract: ***Web usage mining is leading research area in Web Mining concerned about the web user's behavior. Web log mining is one of the recent areas of research in Data mining. Web Usage Mining becomes an important aspect in today's era because the quantity of data is continuously increasing. We deal with the web server logs which maintain the history of page requests. Web log files are the files which contain complete information about the users browse activities on the web server Web mining is the application of data mining techniques to discover patterns from the World Wide Web. This paper gives an attention on Web usage mining to predict the behavior of web users based on web server log files. Users using web pages, a frequent access path's and frequent access pages, links are stored in web server log files. Depending upon the frequency of users visiting each page mining is performed. By finding the session of the user we can analyze the user behavior by the time spend on a particular page. Web log along with the individuality of the user captures their browsing behavior on a website and discussing regarding the behavior from analysis of different algorithms and different methods. Web log analysis is a kind of web analytics technique that parses a server log file from a web server, and based on the values contained in the log file, derives indicators about when, how, and by whom a web server is visited. Using frequent links, we predict the user behavior and identify are all the sites mostly viewed by users. These used to predict the user behavior of and faculty in our college.***
Keywords: Web log mining, User identification, Session identification, web usage Mining.

I. INTRODUCTION

Web mining is considered as an application of data mining, it focuses on rules derived from analyzing the web. discovers a user access behavior or pattern from a web server log by data mining techniques to foresee user navigation trend. Understanding user access patterns can help a webmaster in customizing the website content in order to maximize user's browsing experience. Web navigation pattern may be characterized by analyzing web server logs that contain detailed information of user browsing activities. An instance in a web log is composed of IP address of the requesting client, date and time of the request, requesting method used, status code, the requested file, etc. It is possible to categorize the sequence of click activities into sessions, and then use them for further analysis to get the foreseen user behavior. Analyzing users' Web watching behaviors is one of the important and challenging research topics of Web usage mining. If users' interests can be automatically detected from users' Web log data, they can be used for information recommendation and marketing which are useful for both users and Web site developers. Research for analyzing Web log data has been done by many researchers in the field of Web usage mining; discovering frequent patterns of log data. The main purpose of the study of the user's visit record is to analyze the user's most concerned about the results from the mining results. By analyzing the user access to resources of the time, frequency and so on, modify the structure and design of the site to expect more customers to stay and better serve customers. User behavior analysis has become a new research hotspot. The work of this research mainly studies the web log Mining technology in user behavior analysis, and builds the user interest model based on the user interest information, and finally draws the user's interest.

II. LITERATURE REVIEW

In this research [1] propose the design and implementation of a hybrid system by combining several data mining techniques to capture user's web browsing behavior. User navigation sessions that represent the interaction with a given website are used to construct Hypertext Probability Grammar (HPG). Analyze the HPG with concept hierarchy model and apply matrix clustering technique to extract isual representation of user navigation behavior.This paper [2] describes a method for clarifying users' interests based on an analysis of the site-keyword graph. The method is for extracting sub graphs representing users' main interests from a site-keyword graph which is generated from augmented Web audience measurement data (Web log data) .Our contributions are 1) Web structure mining method (PageRank) is applicable for evaluating the qualities of selected sub graphs and 2) two kinds of data (sites and keywords) can be represented and processed in the same manner using a site-keyword graph. Our

©IJRASET: All Rights are Reserved

method is simple and it has abilities of extracting sub graphs related to users' main interests. Our work is expected to give insights for processing noisy heterogeneous graph structures.

This study of work [4] mainly introduces the method of web log mining, which can discover the mode of web pages by digging web log records. By analyzing and exploring the rules of web log records, we can identify the potential customers of the website and improve the quality of information services to users. In the stage of user behavior analysis, this study explores the differences in user browsing behavior in different types of access events, and calculates the user's interest based on the M5 model tree to analyze the analytic events.

In this study [5], we introduce "task trail" to understand user search behaviors. We define a task to be an atomic user information need, whereas a task trail represents all user activities within that particular task, such as query reformulations, URL clicks. he study analyses and comparisons to evaluate the effectiveness of task trails in several search applications: determining user satisfaction, predicting user search interests, and suggesting related queries. Experiments on large scale datasets of a commercial search engine show that: (1) Task trail performs better than session and query trails in determining user satisfaction; (2) Task trail increases web page utilities of end users comparing to session and query trails; (3) Task trails are comparable to query trails but more sensitive than session trails in measuring different ranking functions; (4) Query terms from the same task are more topically consistent to each other than query terms from different tasks; (5) Query suggestion based on task trail is a good complement of query suggestions based on session trail and click-through bipartite. The findings in this paper verify the need of extracting task trails from web search logs and enhance applications in search and recommendation systems.

III. PREDICTION OF USER BEHAVIOUR

Research for analyzing Web log data has been done by many researchers in the field of Web usage mining; discovering frequent patterns of log data, modeling users' navigation patterns, clustering users of specific Web site, and discovering user communities.

A. Web Log Mining

Web Usage Mining is the application of data mining techniques to discover interesting usage patterns from Web data in order to understand and better serve the needs of Web-based applications. Usage data captures the identity or origin of Web users along with their browsing behavior at a Web site.

A server log is a log file (or several files) automatically created and maintained by a server consisting of a list of activities it performed. Web server logs are used to cluster web users having similar interests. It is also defined as adapting services and information which was available on a website to the needs and the expectations of a target user, the active user; the personalization task by might benefit from the knowledge gained from an analysis of the user's navigational behavior combined with other features which are peculiar to a Web context, namely its structure and content[6].

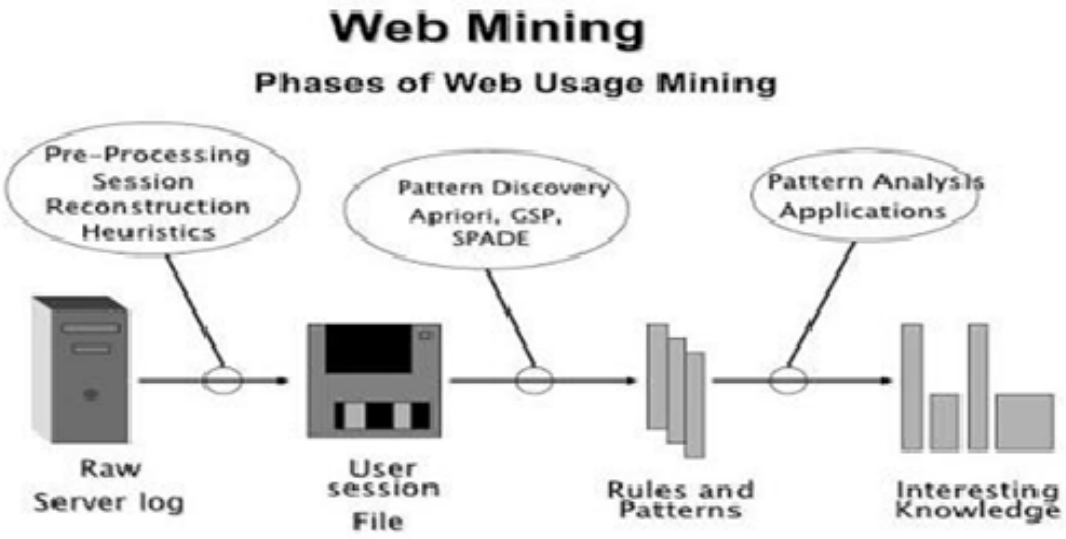


Fig:1 Phases of Web Usage Mining

©IJRASET: All Rights are Reserved

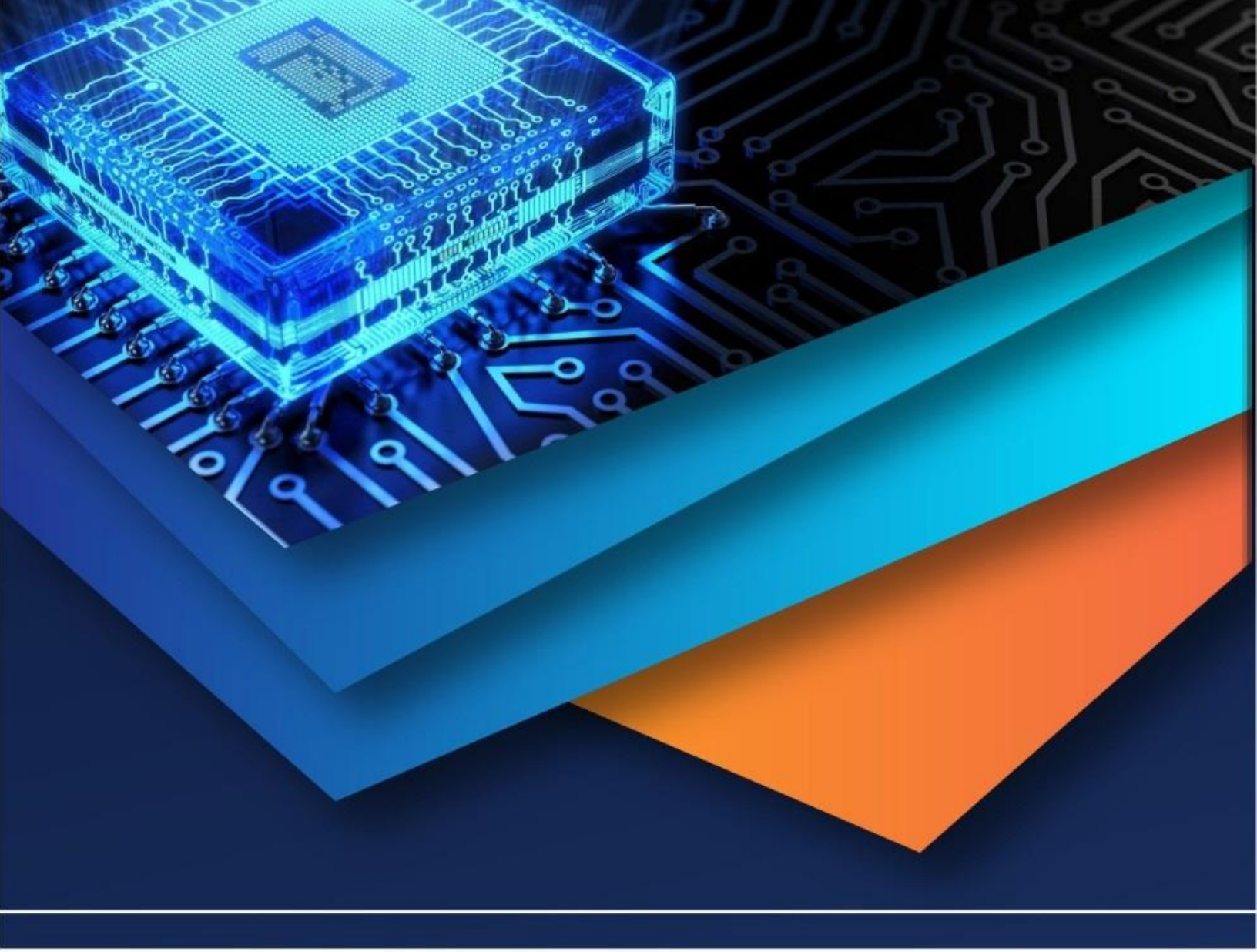
B. Session IDENTIFICATION

At present, there is no standard definition of session in the International academic community. The definition of session made by W3C Union is that it's a limit set of pages generated by users' one or several request click. In the Web log, the request records belong to the users who have different IP addresses certainly not in the same session. A user may browse the same site several times and the time interval between each browse relatively large, then that is not in the same session. Session identification's task is to divide users' browse records into appropriate session. We always first do different assumptions on user's browsing behavior on the client, and then identify the session. These sessions can be used as an input data in classification, clustering, prediction and other tasks. Based on a uniform fixed timeout a traditional session identification algorithm is used.

REFERENCES

[1] D. M. -S. Kao | T. Ozyer | R. Alhajj,"Hybrid approach for predicting the behavior of Web users", IEEE International Conference on Information Reuse and Integration, Conf, 2005.
[2] Tsuyoshi Murata | Kota Saito,"Extracting Users' Interests from Web Log Data ", IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings)(WI'06)
[3] Mahendra Pratap Yadav | Pankaj Kumar Keserwani | Shefalika Ghosh Samaddar,"An Efficient Web Mining Algorithm for Web Log Analysis: E-Web Miner ",1st International Conference on Recent Advances in Information Technology (RAIT) 2012
[4] Xipei Luo | Jing Wang | Qiwei Shen | Jingyu Wang | Qi Qi ,"User Behavior Analysis Based on User Interest by Web Log Mining ", 27th International Telecommunication Networks and Applications Conference (ITNAC),2017
[5] Zhen Liao, Yang Song, Yalou Huang, Li-wei He, Qi He,"Task Trail: An Effective Segmentation of User Search Behavior",IEEE Transactions on Knowledge and Data Engineering,2 014.
[6] C., Ventura, S., Zafra, A., de Bra, P.: Applying Web usage mining for personalizing hyperlinks in Web-based adaptive educational systems (received January 8, 2009) (received in revised form May 4, 2009) (accepted May 4, 2009)
[7] Trapti Agrawal, Shailendra Srivastava, Abhishek Mathur," An Efficient Prediction Based on Web User Simulation Approach Using Modified Ant Optimization Model and Hierarchical Clustering," International Conference on Machine Intelligence Research and Advancement,2013
[8] G. Neelima, Dr. Sireesha Rodda, "Predicting user behavior through Sessions using theWeb log mining", International Conference on Advances in Human Machine Interaction (HMI - 2016).
[9] M.Sathya, Dr.P.Isakki@Devi," Apriori Algorithm on Web Logs for Mining Frequent Link," IEEE International Conference On Intelligent Techniques In Control, Optimization And Signal Processing,2017.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)