International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 06 | June-2016 www.irjet.net p-ISSN: 2395-0072

A STUDY OF ACCURATE WEB LOG MINING FOR PREDICTABLE SEARCH

T. Nandhini[1] , Dr.B.Kalpana[2]

[1]Research Scholar, [2]Professor, Department of Computer Science,

Avinashilingam Institute of Home Science and Higher Education for Women,

Coimbatore, Tamilnadu.

Abstract: *With the new upgrading technology of web mining contains large number of information to serve the people by providing the relevant knowledge. While providing the information it contains the web log data, which helps to store the user search items for providing better relevant data to the next search. Web surfer needs the predictable information while searching the content, and also they need the accurate data for further process. In order to process these data web search browser have to store the previous search result so it can help to provide the relevant search element, to processing this many ideas were developed by many researchers, this paper provide a study about various technologies and its uses for analyzing the web log details in the database.*

Keywords: *Web mining, Large Database, Web Log Mining, Predictable Information and accurate data.*

1. INTRODUCTION

In universe all the internet users completely depends on the search engine, which is not a small, contains a lot of information's, all the data's stored with semantic model. World Wide Web provides information to the people to access the content, related to their search. Nowadays it is necessary one for the people of today to done their work in easy and interactive manner. To interact with the people many providers where provide the content with the help of the web mining method, the user on web access their own information [1]. Web mining is one of the fields in data mining technique to mine the knowledge from the World Wide Web, typically referred to as Web Content Mining (WCM). Web Content Mining is the process of extort helpful information from the stuffing of Web documents. Web user want the predictable information for many reason, like timing saving, not waste in searching irrelevant information, etc., for providing these kind of information web log mining concept is arise, this helps to store the user searchable content for providing relevant information to the users. It maintain the temporary storage details for provide these information, web search will offer the information through the last search by the same computer and also in the networked computer. This mechanism refers to personalized content recommendation on portal websites involves a process of collecting and storing information about portal website users, managing the content assets, analyzing past and current user interactive actions, and, based on the analysis, delivering the right content to each user, a personalized web search engine is designed to search for information on the World Wide Web. The searches results are presented in a list of result are commonly called hits. The information may consist of web pages, information, images and other types of files. Some search engines also mine data available in databases or open directories [2]. Unlike Web directories, which are maintained by human editors, search engines operate algorithmically or are a mixture of algorithmic and human input. Most of the personalized search engines use hit based priority for recommendation. Now a day's several standard web search engines like Google provides saving choice of individual histories in order so as to recover and retrieve often viewed pages. But the limitations and convenient based issues over the search engines are the main drawback. Ineffective search history analysis and parameters which determines the interest is so old methods. The cause behind this is to find users access models robotically and rapidly from the enormous Web log data. The such as frequent access links, frequent access page groups and user clustering [3]. In particular, this states that having a web log analyzer alone is not sufficient to provide a complete picture of Search intelligence. Nevertheless, web-log analysis offers good elaborated reports on web site traffic statistics. There is a need to mine other information like competitors, market positions, and most significantly the users and their behaviors on the product and services. For this

Customized or Personalized Web search engines were user to store and retrieve the data within the web log files. This engine has the following three basic features and process. They are crawling, indexing and searching. The information available about the user's interest can be considered in some of those three phases, depending on its nature. Work on search personalization already exists. In order to solve the problems of personalizing the content in the search engine using the histories in query relation of the user search and interests, the system have developed a system for organizing those search histories in the form of clusters and generates results [4]. While generating results in the web, the result has been generated using the frequently visited websites, i.e. based on the popularity, the search engine providing the result. Many previous approach search or Search intelligence techniques are based on the concept proposed with frequent clicks of the web pages with many algorithm and technique. This paper shows the various approaches for the web log mining results with the predictable information.

2. LITERATURE REVIEW

From the author Bhagyashree D. Shendkar et al [5] evaluate broaden in reputation of web; web mining has fascinated lots of consideration. Web usage mining is an important area in web mining, the discovery of patterns in the browsing and navigation data of Web users. In organize to superior serve the needs of web based applications; web log mining is the technique to notice usage patterns from web data. In customer access log files very important information regarding log servers is present. This paper illustrate about preprocessing of Web Log Data, examination of Web Log Data to come across information on web site, users' navigational patterns production of the fastidious web site other information which will help system supervisor and Web designer to recover their system by formative the example and the procedure of web pages. The results which achieve will characterize in the graphical and tabular design. These outcomes will be used in added expansion of web site in order to augment its usefulness and they paying attention on encompass of more polished techniques for data pre-processing and acknowledgment of admittance assembly, in order to moderate general problems of Web Usage Mining. Other algorithms for model uncovering shall also be built-in in the system, so as to produce substitute system such as Fuzzy-C-Mean algorithm and different clustering technique for their additional work.

PriyankaBhart and Dr.SonaMalhotra [6] author criticize the web log files in order to eliminate the functional patterns. The Web usage mining contain the data from the web server logs, poxy server logs, browser logs, user outline, register data, user sessions or dealings, cookies, user profiles, listing data and any other data as the consequences of interactions. With the unrelenting growth and large number of Web services and Web based information systems, the quantity of user data have reached through the ceiling proportions. Analyzing such data using Web Usage Mining can help to agree on the retreat happiness or needs of the web user. Lots of investigate has been done in this field but this manuscript deals with user potential request calculation using web log record or user information. This paper gives the summary of a mixture of technique of future demand prediction.

R.Shanthi and Dr.S.P.Rajagopalan [7] focuses on the competent submission of the Web Mining Algorithm for web log examination which is applied to catalog the context associated with the web design of an e commerce internet portal that demands security. As priority is given to efficiency, the proportional study made with other related algorithm like E-web Miner Algorithm and Apriori All, it has been establish that this projected Web Page Collection web mining algorithm as the best [or say the foremost suited] recitalist to manage time and space complication .thus this algorithm, improved known as competent web Miner possesses valid by computational relative routine analysis. The amount of data base inspect drastically gets compact in Web Page group algorithm. At this time it may be well-known that E -Web Miner can be functional productively in any weblog examination which includes in sequence centric system design. Web Page Collection Algorithm is introduced that uses cluster mining in classify to uncover a group of connected pages at a web site. The projected algorithm takes web server admission log as input and maps it into form clustering. After that the cluster mining is functional to the productivity data. Finally the productivity is acquire by mining the web user's logs.

From DilpreetKaur [8] analyses the web usage mining, and censure the agreement with log files for pull out the information about user browsing activities. User potential demand prediction is a move toward of web usage mining to guess the next web page for user. In their paper, KFCM method of fuzzy clustering is planned to guess the user

future requests. In this first of all log file data is unruffled and then preprocessed. After that clustering algorithm FCM and KFCM are put into operation to guess the user expectations requests. This presents the results crucial the betterment of KFCM for prediction. They execute which consist of three phases specifically Data Cleaning, User discovery and Session recognition. In Data cleanout phase surplus entries from the log file are removed and file is prearranged into prepared structure. In User discovery phase users are acknowledged based on the IP Address. In Session identification phase sessions are recognized by taking entry value of time. After preprocessing fuzzy clustering algorithms are executed for guess and results are investigate. At last user expectations request is predicted.

J. Srivastava, et al [9] analyze the great volumes of data are congregate repeatedly by Web servers and composed in right to use log files. Scrutiny of server access data can supply significant and helpful information. Web Usage Mining is the method of applying data mining method to the detection of usage model from Web data and is beleaguered towards function. It mines the secondary data consequent from the communications of the users during definite period of Web assembly. Web usage mining contains three phases, namely preprocessing, pattern discovery, and pattern analysis. Given its function probable, Web usage mining has seen a speedy augment in interest, from both in explore and perform group of people. In this paper, this applied Kohonen's SOM (Self Organizing Map) to pre-processed Web logs of our university Web server logs (http://www.um.ac.ir/) and pull out frequent patterns. Outcome of this paper would be constructive for our university Web site owner.

3. PREDICTING WEB LOGS USING VARIOUS TECHNIQUES

3.1 Association Rules

Association rule mining is used to extract the data with the help of the associated objects. In previous researchers used this technique to extract the associated information. Relatively little research has been conducted on web data. There are different types of associations: association between text content, image content, video content and non image content features. Association mining in web data can be transformed into problems of association mining in traditional transactional databases. Therefore, mining the frequently occurring patterns among different data becomes mining the frequent patterns in a set of transactions. It also extracts the concept of content-based association rules using feature localization [10, 11]. In association rule mining it extract the relevant data in the same network as well as same computer, with the help of this association data mining algorithm it tends to provide the related search contents in the web search engine.

3.2 Artificial Neural Network

ANN can be defined as a massively parallel interconnected network of simple, adaptive processing elements which is intended to interact with the objects of the real world in the same way as biological systems do [12, 13]. *A*NNs are designated by the network topology, connection strength between pairs of neurons (called weights), node characteristics, and the status updating rules. They have been applied to the tasks like IR, IE, and clustering (self organization) of web mining, and for personalization. ANNs provide a convenient method of knowledge representation for IR applications. In current trend in web log mining analyses the researchers use the ANN method to extract the relevant knowledge or information well. It provides well information but taking long time to process a single data. In this method it creates the heap inside it, using that heap what and all the user searching that kinds of information is stored in that help for providing the better and relevant user needed data in the web using Association Web log mining method.

3.3 Ant Colony Optimization

In data mining, extracting the information from hidden knowledge is not an easiest task, while extracting the data has to follow some of the technique or algorithm to find out the related data. With the help of the data web log mining, it has to extract the relevant search content which is stored in the web log data. Ant colony is one of the optimization methods to provide an optimal result to the user. It is a straightforward method, they mutually forms an ant colony which do noteworthy tasks withundeviating(shortest) path traversal to find food source and information allocation with supplementary ants by manufactureinstance. In this technique it collective astuteness of ants is distorted into useful optimization techniques that discovery uses in computer networking. This algorithm is based on the ant, which searches the food with the help of the smell, like this it search the content based on the method of ant [14, 15]. And also find the food in the shortest method to preserve the record-set and the

traffic that is incoming will be diverted to the path which has the more likelihood as in natural ant which select a path with highest pheromone. Hence using this technique it finds the related content in the web for the relevant search item.

4. CONCLUSION

Web log files have become important data source for web behaviors. Analyzing web log files is one of the significant research fields of web mining. Web log mining is used to store those searched data for providing the accurate predictable result in the future. For finding the effective data search web log provide a better predictable data to the user, for analyzing this concept of better predictable information many researchers were provide various concepts and most important research by the author in Web usage mining is how to mine predictable user behavior patterns in Web log files. This paper provide the study of various approaches to solve the predictable problem in the web search engine using the web log mining and provide the accurate result is found in the combination of hybrid techniques.

REFERENCES

1. Yi-Hung Wu and Arbee L. P. Chen, "Prediction of Web Page Accesses by Proxy Server Log" World Wide Web: Internet and Web Information Systems, 5, 67–88, 2002.
2. Huysmans, J., Baesens, B. and Vanthienen, J.), "Web Usage Mining: A Practical Study", *KatholiekeUniversitiesLeuven*,Dept. of AppliedEconomic Sciences (2003).
3. Ma Shu-yue, Liu Wen-cai Wang Shuo, "The Study on the Preprocessing in Web Log Mining" in proceedings of the Knowledge acquisition and modelling(KAM),IEEE,pp. 315-317,2011
4. BamshadMobasher, "Web Usage Mining" in Web Data Mining ExploringHyperlinks,Contents,andUsageData,Chicago,USA,Springer,ch.12,pp. 449-482,2012
5. Bhagyashree D. Shendkar et al "A Novel Approach for User Navigation Pattern Discovery and Analysis for Web Usage Mining"
6. PriyankaBhart and Dr.SonaMalhotra, " A Review Paper on Web Usage Mining and future request prediction request prediction".
7. R.Shanthi, Dr.S.P.Rajagopalan, "An Efficient Web Mining Algorithm To Mine Web Log Information".
8. DilpreetKaur and A.P. SukhpreetKaur, "User Future Request Prediction Using KFCM inWeb Usage Mining".
9. J. Srivastava, R. Cooley, M. Deshpande, and P.-N.Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data Tan, 29 September 2009.
10. S. Kim and B. T. Zhang, "Web document retrieval by genetic learning of importance factors for html tags," in *Proc. Int. Workshop Text Web Mining*, Melbourne, Australia, pp. 13–23, 2000
11. Sankar.K.Pal, VarunTalwar and PabitraMitra, "Web mining in Soft Computing Framework: Relevance, State of the Art and Future Direc-tions", IEEE Transactions on Neural Networks, September 2002
12. R. Yager, "A framework for linguistic and hierarchical queries for Document retrieval," in Soft Computing in Information Retrieval: Tech niques and Applications, F. Crestani and G. Pasi, Eds, Heidel berg: PhysicaVerlag,, vol. 50, pp. 3–20,2000.
13. Eirinaki, M. and Vazirgiannis, M. , "Web mining for web personalization", *ACM Transactions on Internet Technology*(TOIT), 2003, Vol. 3, Issue 1, Pp. 1-27.
14. Abraham, A. and V. Ramos (2003). Web Usage Mining Using Artificial Ant Colony Clustering and Linear Genetic Programming. Proc. Of the Congress on Evolutionary Computation (CEC 2003), Canberra, pp. 1384-1391. IEEE.
15. Guerbas et al., "Effective web log mining and online navigational pattern prediction", Knowledge-Based Systems,Elsevier,vol. 49,pp. 50-62, 2013