IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Special Issue, 4, Feb 2019 E-ISSN: 2347-2693

User Web Access Record Mining for Business Intelligence

D. Kamalavadhanai[1*], N. Aarthi[2]

[1,2]Dept. of Computer Science, Idhaya College for Women, Kumbakonam, Tamilnadu, India

Corresponding Author: aarthi67@gmail.com

Available online at: www.ijcseonline.org

***Abstract*—**User's access records are captured by implementing a data mining algorithm on the website. User mostly browses those products in which he is interested. This system will capture user's browsing pattern using data mining algorithm. This system is a web application where user can view various resources on the website. User will register their profile in an exchange of a password. User will get user ID and password in order to access the system. Once the user login's to the system user will gain access to certain resources on the website. The links to the resources on the website have been modified such that a record of information about the access would be recorded in the database when clicked. This way, data mining can be performed on a relatively clean set of access records about the users. When user clicks on certain resources on the website his access records will be captured by the system this can be achieved with the help of data mining algorithm used in this system. By using this application, product based organization will get to know the demand for certain products. This system will help organization to target right consumers. This system will help product based firms to maintain good customer relationship. Hence, a good deal of business intelligence about the users' behavior's, preferences and about the popularities of the resources (products) on the website can be gained.

***Keywords*—**User Access Record, Good Customer Relationship, Web Mining, User Behavior, Web Logs.

I. INTRODUCTION

The WWW continues to grow at an amazing rate as an information gateway and as a medium for conducting business. Web mining is the extraction of interesting and useful knowledge and implicit information from facts or activity related to the WWW. Based on several research studies we can broadly classify Web mining into three do mains: content, structure and usage mining. A typical Web log format is depicted in Fig. 1. Whenever a visitor accesses the server, it leaves the IP, authenticated user ID, time/date, request mode, status, bytes, referrer, agent and so on. The available data fields are specified by the HTTP protocol. There are several commercial software that could provide Web usage statistics. These stats could be useful for Web administrators to get a sense of the actual load on the server. However the statistical data available from the normal Web log data files or even the information provided by Web trackers could only provide the information explicitly because of the nature and limitations of the methodology itself. Generally, one could say that the analysis relies on three general sets of information given a current focus of attention: (1) past usage patterns; (2) degree of shared content; and (3) inter memory associative link structures. After browsing through some of the features of the best trackers available it is easy to conclude that rather than generating statistical data and texts they really do not help to and much meaningful information.

For small web servers, the usage statistics provided by conventional Web site trackers may be adequate to analyze the usage pattern and trends. However as the size and complexity of the data increases, the statistics provided by existing Web log file analysis tools may prove inadequate and more intelligent knowledge mining techniques will be necessary.

In the case of Web mining, data could be collected at the server level, client level, proxy level or some consolidated data. These data could differ in terms of content and the way it is collected etc. The usage data collected at different sources represent the navigation patterns of different segments of the overall Web traffic, ranging from single user, and single site browsing behavior to multi-user, multi-site access patterns. Web server log does not accurately contain sufficient information for infering the behavior at the client side as they relate to the pages served by the Web server.

Web server log does not accurately contain sufficient information for infering the behavior at the client side as they relate to the pages served by the Web server. Pre-procesed

and cleaned data could be used for pattern discovery, pattern analysis, Web usage statistics and generating association sequential rules. Much work has been performed on extracting various pattern information from Web logs and the application of the discovered knowledge range from improving the design and structure of a Web site to enabling business organizations to function more efficiently. A combination of hypertext probabilistic grammar and click fact table approach is used to mine Web logs which could be also used for general sequence mining tasks. Mobasher et al.(1999) proposed the Web personalization system which consists of online tasks related to the mining of usagedata and online process of automatic Web page customization based on the knowledge discovered. LOGSOM proposed by Smith et al. (2003), utilizes a self-organizing map to organize web pages into a two-dimensional map based solely on the users' navigation behavior, rather than the content of the web pages. LumberJack proposed by Chi et al. (2002) builds up user profiles by combining both user session clustering and traditional statistical traffic analysis using K-means algorithm.

To demonstrate the efficiency of the proposed frameworks, Web access log data at the Monash University's Web site (Monash, 2003) were used for experimentations. The University's central web server receives over 7 million hits in a week and therefore it is a real challenge to and extract hidden usage pattern information. To illustrate the University's Web usage patterns, average daily and hourly access patterns for 5 weeks (11 August '02{14 September '02) are shown in Figs. 3 and 4 respectively. The average daily and hourly patterns nevertheless tend to follow a similar trend the differences tend to increase during high traffic days (Monday{Friday) and during the peak hours (11:00{17:00 Hrs). Due to the enormous traffic volume and chaotic access behavior, the prediction of the user access patterns becomes more difficult and complex.

Self-organizing maps and fuzzy c-means algorithm could be used to segregate the user access records and computational intelligence paradigms to analyze the user access trends. Abraham (2003) and Wang et al. (2002) have clearly shown the importance of the clustering algorithm to analyze the user access trends.

The i-Miner hybrid framework optimizes a fuzzy clustering algorithm using an evolutionary algorithm and a Takagi-Sugeno fuzzy inference system using a combination of evolutionary algorithm and neural network learning. The raw data from the log files are cleaned and pre-processed and a fuzzy C means algorithm is used to identify the number of clusters.

II. METHODOLOGY

Further a novel approach called intelligent-miner" (i-Miner) is presented. i-Miner could optimize the concurrent architecture of a fuzzy clustering algorithm (to discover web data clusters) and a fuzzy inference system to analyze the Web site visitor trends. A hybrid evolutionary fuzzy clustering algorithm is proposed to optimally segregate similar user interests.

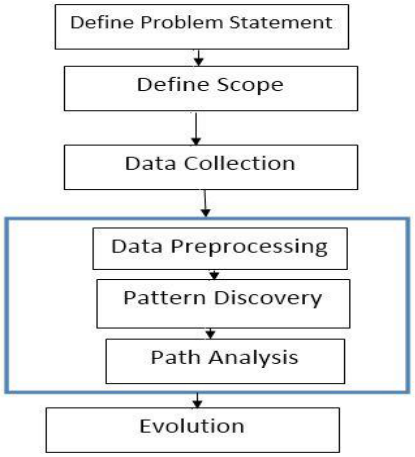

Figure 1: Research Framework

ADVANTAGES

This system will be useful for those users who often purchase products online.

User will get right product according to his preference.

This system will be useful to various product based firms who will get to know popularity of certain products on the website.

This system will help the organization to know demand for certain products.

This system will help organization to target right consumers.

This system will help product based firms to maintain good customer relationship.

III. RESULTS AND DISCUSSION

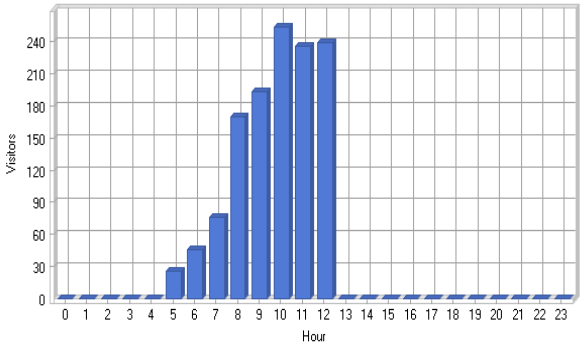

Figure 2: Hourly Website Visitor Report

The hourly basis of report of website visitor in the shown in figure3. It shows number of hits per hour, page views per hour, visitor per hour and bandwidth in KB per hour.The information in four types: first Hits, second page view, third visitors and fourth bandwidth. The first of the number of that is Hits: the number of visitor Hits, spider Hits, Average Hits per Day, Average Hits per visitor, Cached Requests, Failed Requests. Second is common statistics the shows table the number of Average page view per day and per visitor. Third is visitor and that is also provide the information of Usual visitor each Day. The last is the Bandwidth of visitor, Spider Bandwidth, Average Bandwidth per Day. This state reduces the info of total usage accessibility of website.

Hits	
Total Hits	60,233
Visitor Hits	59,899
Spider Hits	334
Average Hits per Day	60,233
Average Hits per Visitor	48.31
Cached Requests	9,500
Failed Requests	1,404
Page Views	
Total Page Views	2,797
Average Page Views per Day	2,797
Average Page Views per Visitor	2.26
Visitors	
Total Visitors	1,240
Average Visitors per Day	1,240
Total Unique IPs	1,195
Bandwidth	
Total Bandwidth	5.09 GB
Visitor Bandwidth	5.04 GB
Spider Bandwidth	56.67 MB
Average Bandwidth per Day	5.09 GB
Average Bandwidth per Hit	88.62 KB
Average Bandwidth per Visitor	4.16 MB

Table 1: General Activity Statics of Website

IV. CONCLUSION AND FUTURE SCOPE

Recently Web usage mining has been gaining a lot of attention because of its potential commercial benefits. The proposed i-Miner framework seems to work very well for the problem considered. The empirical results also reveal the importance of using soft computing paradigms for mining useful information. Several useful information's could be discovered from the clustered data. FCM clustering resulted in more clusters compared to SOM approach. Perhaps more clusters were required to improve the accuracy of the trend analysis. The main advantage of SOMs comes from the easy visualization and interpretation of clusters formed by the map. The knowledge discovered from the developed FCM clusters and SOM could be a good comparison study and is left as a future research topic. In future research will be oriented in this direction by incorporating more data mining paradigms to improve knowledge discovery and association rules from the clustered data.

REFERENCES

[1]. Abraham, A (2001). Neuro-fuzzy systems: State-of-the- art modeling techniques, connectionist models of neurons, learning processes, and arti_cial intelligence. In Lecture Notes in Computer Science 2084, J Mira and APrieto (eds.), Germany, Spain: Springer-Verlag, pp. 269-276.

[2]. Cord_on, O, F Herrera, F Ho_mann and L Magdalena (2001). Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases. Singapore: World Scienti_c Publishing Company.

[3]. Hall, LO, IB Ozyurt and JC Bezdek (1999). Clustering with a genetically optimized approach. IEEE Transactions on Evolutionary Computation, 3(2), 103-112.

[4]. Pal, SK, V Talwar and P Mitra (2002). Web mining in soft computing framework: Relevance, state of the art and future directions. IEEE Transactions on Neural Networks, Vol. 13, No. 5, pp. 1163-1177.

[5]. Paliouras, G, C Papatheodorou, V Karkaletsisi and CD Spyropoulous (2000). Clustering the users of large websites into communities. In Proc. of the 17th International Conference on Machine Learning (ICML'00), pp. 719-726. USA: Morgan Kaufmann.

[6]. Pazzani, M and D Billsus (1997). Learning and revising user pro_les: The identi_cation of interesting web sites. Machine Learning, 27, 313-331.

[7]. Spiliopoulou, M and LC Faulstich (1999). WUM: A web utilization miner. In Proc. of EDBT Workshop on the Web and Data Bases (WebDB'98), pp. 109-115. Springer Verlag.

[8]. Srivastava, J, R Cooley,MDeshpande and PN Tan (2000). Web usage mining: Discovery and applications of usage patterns from web data SIGKDD Explorations, 1(2), 12-23.

[9]. Wang, X, A Abraham and KA Smith (2002). Soft computing paradigms for web access pattern analysis. In Proc.of the 1st International Conference on Fuzzy Systems and Knowledge Discovery, pp. 631-635.

[10]. Yang, Q and HH Zhang (2003). Web-log mining for predictive web caching. IEEE Transactions on Knowledge and Data Engineering, Vol. 15, No. 4, pp. 1050-1053.

