Prediction of User Behavior in Educational Web Sites by Applying Web Mining

J.Umarani[1],G.Thangaraju[2], J.Anitha[3]
1. PhD Research Scholar, Bharathiyar University, Coimbatore.
2. PhD Research Scholar, Karpagam University, Coimbatore.
3. M.Phil Research Scholar, Thanthai Hans Roever College, Perambalur.

Abstract-A presently internet is most imperative part of human life. The internet is growing day by day, so online users are also escalating. The interesting information for knowledge of extracting from such enormous data demands for new logic and the new method. Every user spends their most of the time on the internet and their behavior is different from one and another. The application of data mining techniques is the need of web-based applications. Web usage mining is leading research area in Web Mining concerned about the web user's behavior. 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 paths and frequent access pages, links are stored in web server log files. A 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.

Keywords: web usage mining; pre-processing, Apriori algorithm, FP-Growth algorithm, log files.

I.INTRODUCTION

The World Wide Web continues to rise at an amplified swiftness in both the volume of traffic and the size and problem of web sites. It is imperative to extract knowledge from this web data to capture tastes of users. The function of a data mining techniques is a web mining used to extract functional information from web data consisting documents, hyperlinks and web site access log file etc.

Web mining is under three basic concepts,
1. Web Content Mining
2. Web Structure Mining
3. Web Usage Mining

A content mining is used to inspect the data collected by search engines. Structure mining is used to learning the data, connected to pattern of particular website and usage mining are used to inspect data related to a particular user browsers are as well as the data to gather the forms of the users may contain the submitted to through Web transactions. This information is to be gathered via Web mining. A using traditional data mining parameters are the association, clustering and classification, and examination of sequential patterns.

Web usage mining is one of the web mining techniques. Web users use the collection of web pages and web information are stored in web server. This usage of data is to provide leading paths to access the Web pages. This information is often gathered automatically access web log through the Web server.
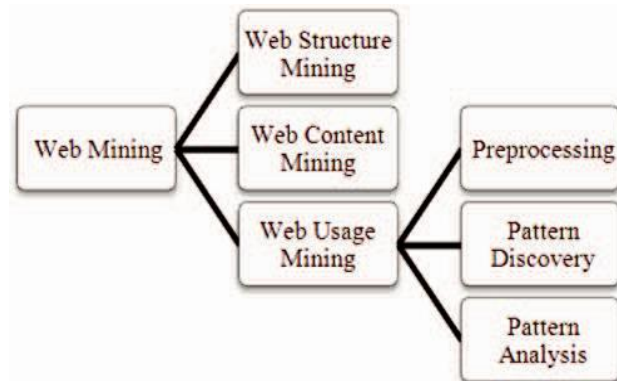
Fig. 1: Web Data Mining Structure

The Common gateway scripts of the other useful in sequence such as refer to weblogs, user subscription information and web logs. This is an most important of data mining for companies and their internet / intranet-based applications and information access. Web servers store information of each page requested by web visitors called the web access log. Web Usage Mining addresses the problem of extracting behavioral patterns from one or more web access logs. Web Usage Mining (also called as clickstream analysis) Edelstein [5] is the process of applying data mining techniques to the discovery of usage patterns from Web data, and is targeted towards applications Srivastava, et al. [8]. It tries to make sense of the data generated by the Web surfer's sessions or behaviors.

The information is to be derived from the collective information of lifetime user values. The usage of data is gathered to provide the companies with the ability to produce the results of more effective to their businesses and increasing of sales. Usage of data is applied for developing marketing skills. They will out-sell competitors and promote the company's services or product a higher level.

Therefore, it is easily determined that usage mining has valuable uses to predict web user behaviors. Each and every user thinking of thoughts and behavior is very different from one and another. This paper focuses the web usage mining process of usage mining and algorithms used for usage mining for user prediction by applying some data mining techniques.

Web usage mining consists of three phases, namely preprocessing, pattern discovery, and pattern analysis (see Figure 2). The goal of web usage mining is to get into the records of the servers (log files) that store the transactions that are performed in the web in order to find patterns

revealing the usage the customers. WUM has become an active area of research in field of data mining due to its vital importance [15].
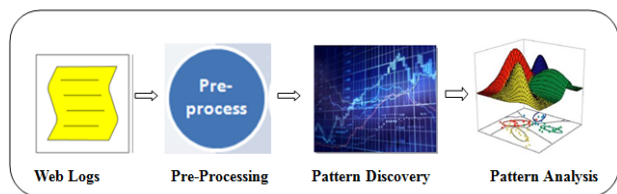

Figure 2: Web Usage Mining Process

Motivation

In the recent times more and more users are using the internet services for their necessity. This helps to do research in extracting useful information and user interest from the web using mining techniques. The web logs are one of the most utilized features to extract the user's interest measure. The web log mining is used more frequently in order to identify the user behavior based on the extent to which a user is visiting a particular web site. Since web logs are updated each time the user visits a particular web site, so it is considered as moving data. Thus algorithms which are capable of processing moving data are to be considered for the mining of the web logs.

II. RELATED WORK

The focus of related work is to study and contrast the available technique to predict the web user behavior.

Jagan and Rajagopalan [1] describe the web usage mining and algorithms used for providing personalization on the web. In this paper focused the data preprocessing and pattern analysis on the web and using the association rule mining algorithms.

Ladekar A. Pawar A. *et al*. [2] describe a web mining algorithm that aims at amending the interpretations of the draft's output of association rule mining. This algorithm is being tremendously used in web mining. The results obtained prove the robustness of the algorithm proposed in this paper.

Parvatikar S. and Joshi B. [3] this paper focused on Web Usage Mining is the user navigation patterns and their use of web resources. The different stages involved in this mining process and with the comparative analysis between the pattern discovery algorithms Apriori and FP-growth algorithm.

Deepa and Raajan [4] implemented the pre-processing techniques to convert the log file into user sessions which are suitable for mining and reduce the size of the session file by filtering the least requested pages using the preprocessing technique. Data Preprocessing is one of the important tasks before applying mining algorithms. It converts the raw log file into user session. In this work, we have briefly introduced log file preprocessing and implemented it in a CTI log file. Also, we produce the summary of the user session file. We have used filtering technique to remove least requested resources.

Anand N. [5] describe an internet usage details and provides them with the tools to understand the online behavior of their teenage children. Singh A.P. and Jain R.C. [6] Different kinds of web usage mining techniques with

their basic models and concepts are provided. In addition to that, for discovering the hidden patterns from web access log files a new model based on visual clustering is also suggested. The analysis of different methods of web usage mining.

Mishra R. and Choubey R. [7] describe the FP-growth algorithm is obtaining a most frequently access paths and pages from the web log data and providing valuable information to user behavior.

Zubi Z. S. and Riani M.S.E. [8] discusses the use of web mining techniques is used to classify the web page's type according to user visits. This classification is helps to understand the web user behavior. The classification and association rule techniques for discovering the interesting information from browsing patterns.

Avneet Saluja *et al.* [9] in their work is user future request prediction using web log records and user information. The purpose of the effort is to provide a benchmark for evaluating a various methods used in the past, a present and which can be used in a future to minimize the search time of a user on the network.

Web content mining typically involves searching through web based information repositories. Broadly speaking, there have been two approaches to searching content. The data base approach is an extension of traditional data base querying, though adapted for the nature of the Web (OKS Group, 2003; Biggs, 2003). The agent based approach is inspired by autonomous agent research and the creation of web-bots for adaptively searching or spidering through web repositories (Menczer, 2003). The driving force behind Web usage mining was business to consumer (B2C) e-commerce. By searching for usage patterns from within weblogs, one could adapt a web site to maximise the opportunity for increased sales (Doherty, 2000; Iyer et al, 2002; Manchester, 2001; Mulvenna et al, 2000; Thelwall, 2001). Both web content and web usage mining can take a retrieval approach of known or anticipated content or a discovery of new or unexpected content. There is some debate over whether content retrieval is true web mining (Hearst, 1999), however on balance, the term has been typically used to cover both aspects.

III. PROPOSED WORK

The proposed system aims to discover the behaviour of student and faculty through the web server log files. In the prediction model, the client requests the web page and links automatically get stored in a server log file. All requested web page links get stored and maintained by the client. A more noisy and dirty data's present in the log files. Preprocessing steps are necessary to remove the noise, missing and redundant data. During preprocessing, the client retrieves frequent web link used by users, the HTTP request pages, user identification in different users and sessions is identified and pattern completion. Using frequent web links, we predict the user behavior and identify what are all the sites mostly viewed by users. These techniques used to predict the user behavior of students and faculty in our college.

IV. METHODOLOGY

A. Association rule Mining

A various prediction methods are available to predict the web user behavior. Mining of association rules is imperative research in web usage mining. There are many algorithms are applied in web usage mining. Association rule mining is an innovation and correlation a set of frequent item sets or pages. In most frequent algorithms are Apriori, Apriori TID, STEM, DIC, Partition Algorithm, Eclat and FP-Growth, etc. Since the algorithms are mining frequent item sets.

a. Apriori algorithm

Apriori algorithm is a step-wise search, the n number of item sets is used to discover a n+1 item sets. A set of usual sets, scanning the database to gather the count for each items, and collecting those item sets is satisfying the minimum support. The resulting set is denoted as one common item set. Next, the common time set is used to find next interesting item sets, this process is continue until get the most frequent item sets. A final iteration, you will end up with many n-item sets, this is basically called association rules. To pick interesting rules from the set of all prospective rules and various constraint measures such support and confidence is applied.

b. FP-Growth

This algorithm calculates frequent items sets from large data sets. The advantage of Frequent Pattern Growth is compared to Apriori is that it uses only two times data scans is therefore often valid even on large data sets.

c. Frequent pattern discovery

Pattern Discovery is one of the Web Usage Mining process. It is one of the techniques in association rule mining, According to the most recent of web log analysis methodology; data mining methods are more reliable and efficient for the hidden pattern discovery. But most of the authors are working on supervised methods. Unsupervised methods are not finding the pattern extraction from web logs.
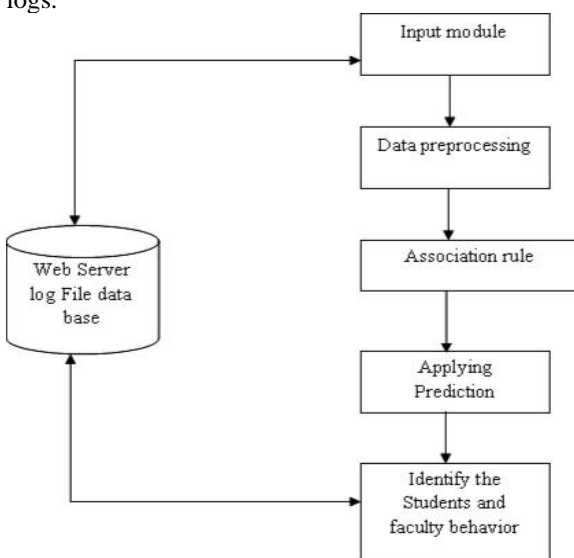

Figure 3: Methodology

d. Pattern Analysis

Pattern Analysis is the goal of the process. In this process is used to extract the appealing patterns from the log files. It is one of the pattern discovery process. Reject an un-relevant data from web log files. Pattern discovery and pattern analysis are the preprocessing stages in web usage mining.

They are similar to those developed for non-Web domains such as statistical analysis, clustering, and classification, but those methods must take into consideration the different kinds of data abstractions and prior knowledge available for Web Mining. In Web Usage Mining, a server session is an ordered sequence of pages requested by a user. Pattern analysis is to filter out the uninteresting rules or patterns from the dataset found in the pattern discovery phase. The exact methodology used for analysis is usually governed by the application for which Web mining is to be done. The most common form of pattern analysis consists of a knowledge query mechanism such as SQL. Another method is to load usage data into a data cube to perform OLAP operations. Visualization techniques, such as graphing patterns or assigning colors to different values, can highlight patterns. The content and structure information can be used to filter out patterns which contain pages of a certain use type or content, or pages that match a certain hyperlink structure.

V. CONCLUSION

Outcomes of the Web Usage Mining process always depend upon the problem of interesting application. Data Mining is the process to mine the interesting knowledge from the enormous amount of data. In previous work, they predict the future request given by the user based on the current request, analyze the pattern and they also predict the kids' behavior. In this paper, we discuss various methods of data mining techniques and describes the way how to apply data mining techniques. The proposed work is to develop a prediction model to predict user behavior based on web server log files. The client analyzes the links used by users and predicts the behavior of the student and faculty based on the websites they are used.

REFERENCES

[1] S. Jagan, and S.P. Rajagopalan, "A survey on web personalization of web usage mining", IRJET-International Research Journal of Engineering and Technology, 2015.
[2] A. Ladekar, P. Pawar, D. Raikar and J. Chaudhari, "Web Log Based Analysis of User's Browsing Behavior", IJCSIT - International Journal of Computer Science and Information Technologies, Vol. 6 (2), 2015.
[3] S. Parvatikar and B. Joshi, "Analysis of User Behavior through Web Usage Mining", ICAST - International Conference on Advances in Science and Technology, 2014.
[4] A. Deepa, and P. Raajan, "An efficient preprocessing methodology of log file for Web usage mining", NCRIIAMI - National Conference on Research Issues in Image Analysis and Mining Intelligence, 2015.
[5] N. Anand, "Effective prediction of kid's behavior based on internet use", International Journal of Information and Computation Technology, 2014.
[6] A.P. Singh, R. C. Jain, "A Survey on Different Phases of Web Usage Mining for Anomaly User Behavior Investigation", IJETTCS - International Journal of Emerging Trends & Technology in Computer Science, Vol 3, 2014.
[7] R. Mishra, A. Choubey, "Discovery of Frequent Patterns from Web Log Data by using FP-Growth algorithm for Web Usage

Mining", International Journal of Advanced Research in Computer Science and Software Engineering, Vol 2, 2012.

[8] Z.S. Zubi, M.S. Riani, "Applying web mining application for user behavior understanding", Recent Advances in Image, Audio and Signal Processing.

[9] A. Saluja, B. Gour, and L. Singh., "Web Usage Mining Approaches for User's Request Prediction: A Survey", IJCSIT-International Journal of Computer Science and Information Technologies, Vol. 6 (3), 2015.

[10] L. Tamrakar, M. Ghosh., "Identification of Frequen Navigation Pattern Using Web Usage Mining", IJARCST-International Journal of Advanced Research in Computer Science & Technolog, Vol. 2, 2014.

[11] S.P. Ajeetkumar, P.K Anagha, "Review on Exploring User's Surfing Behavior for Recommended Based System", IJETTCS - International Journal of Emerging Trends & Technology in Computer Science, Vol 3, 2014.

[12] A. D. Kasliwal, and G. S. Katkar, "Web Usage mining for Predicting User Access Behavior", IJCSIT-International Journal of Computer Science and Information Technologies, Vol. 6 (1), 2015.

[13] V. Neha, Patil *et al*, "Prediction of Web Users Browsing Behavior: A Review", International Journal of Computer Science and Mobile Computing, Vol.4,209-212, 2015.

[14] S. Khan, Y. Singh and A. Kumar. Sachan, "Web Mining Approach in Analysing User Behavior and Interest for Website Modification", International Journal of Advanced Research in Computer Science and Software Engineering, Vol 5, 2015.

[15] A. Vishwakarma, and K.N. Singh, "A survey on web log mining pattern discovery", IJCSIT - International Journal of Computer Science and Information Technologies, pp: 7022-7031, 2014.

[16] S. Parvatikar, B. Joshi, "Analysis of user behavior through web usage mining", ICAST - International Conference on Advances in Science and Technology,pp: 27-31, 2014.

[17] Y.C. Liu, Y.C. Hsu, "Predicting Adolescent Deviant Behaviors through Data Mining Technologies", Educational Technology & Society, 2013.

[18] J. Vellingiri, S. C. Pandian., "A Survey on Web Usage Mining", Global Journal of Computer Science and Technology, Volume 11 Issue 4, 2011.

[19] S.Vijiyarani, E. Suganya, "Research Issues in Web Mining", IJCAx-International Journal of Computer-Aided Technologies, Vol.2, 2015.

[20] U. Patil1, S. Pardeshi., "A Survey on User Future Request Prediction: Web usage Mining", International Journal of Emerging Technology and Advanced Engineering, Vol 2, 2012.

[21] S. Khan, and A. K. Sachan., "Web Mining Approach in Analysing User Behaviour and Interest for Website Modification", International Journal of Advanced Research in Computer Science and Software Engineering, Vol 5, 2015.

[22] H.N. Randhir, R. Gupta, "WUM for Browsing Behavior of a User and Subsequently to Predict Desired Pages: A Survey", IJESIT - International Journal of Engineering Science and Innovative Technology, Vol 2, 2013.