E-learning made easy with weblog mining

Raghav Jhawar
raghav.zawar007@gmail.com
Vivekanand Education Society's Institute of Technology,
Mumbai, Maharashtra

Neeraj Harjani
neeraj.harjani@ves.ac.in
Vivekanand Education Society's Institute of Technology,
Mumbai, Maharashtra

Juhi Purswani
juhi.purswani@ves.ac.in
Vivekanand Education Society's Institute of Technology,
Mumbai, Maharashtra

Sukanya Roychowdhury
sukanya.roychowdhury@vea.ac.in
Vivekanand Education Society's Institute of Technology,
Mumbai, Maharashtra

ABSTRACT

In this project, a novel concept of web log mining is performed on e-Learning Website and also the concept of intelligent navigation behavior on user browser is proposed. Web Log Mining initially performs capturing of different web log file while the user is accessing the e-Learning Website. Web log file is saved. Log files can't be directly used for pattern discovery process because it consists of irrelevant and inconsistent access information. Therefore there was a need to perform web log preprocessing which includes different techniques such as field extraction, data reduction, data cleaning, and data summarization. Field extraction and data reduction algorithm performs the process of separating fields form the single line of the log file. Data cleaning algorithm eliminates inconsistent or unnecessary items in the analyzed data. Data summarization generates a different summarized report and gives the graphical representation of the log that is being captured. Preprocessed information is given to pattern discovery and to the intelligent navigation module. Pattern discovery includes mining algorithm such as association rules, sequence patterns & clustering. Association rules are used to mine the data in order to obtain support and confidence for each rule which in turn describe the association between the subjects. Clustering approach is used to cluster web sites users into different groups based on navigation behavior of the user. Intelligent navigation module uses a concept of sequence pattern and allows the student to have most frequently used subject at the top most lists, which allows them to have easy access to the tutorial or chapter within the subject i.e. most visited subject will be on the top most preference of the user menu list. We have also performed analysis on the web usage patterns and analyzed different access navigation pattern of the student, which in turn enhances the personalization services in e-Learning Website and makes the system much effective.

Keywords: *Web log mining, Log files, Pattern discovery, Intelligent Navigation, Personalization, Cluster Analysis, Association rule.*
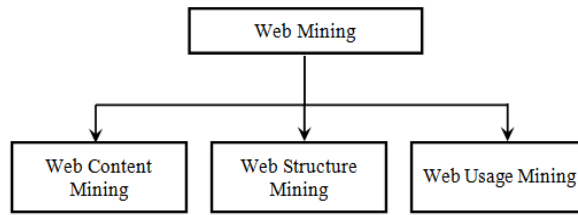
1. INTRODUCTION

Web log mining is the area of data mining which deals with the discovery and analysis of web logs, in order to improve web-based applications. Web log mining consists of three phases, preprocessing, pattern discovery, and pattern analysis. After the completion of these three phases, the user can find the required log patterns and use this information for the specific needs.

Finding the useful data from the web is a complex task, therefore the data mining algorithm is working to identify the pattern and information from the data. The log files are maintained by the web servers. By analyzing these log files gives a neat idea about the user. Our project evaluates and implements various techniques which are available for web data analysis. Therefore this project includes pattern mining algorithms for evaluation and the implementation of frequent pattern analysis from the web data using K-Means Algorithm and Web Log Mining. The challenging task for E-commerce companies is to know their customer behavior to improve the business by analyzing web log files. The analysis of web log files is used for learning the user behavior in an E-commerce system. In our case, the user behavior is the Clicks he/she does while surfing the website and Admin gets an analysis of the captured clicks in the form of pie charts.

So our aim is to build an E-learning website, which will show current trending topics to the user and pie chart analysis to admin. We aim to achieve so by capturing the web logs of authenticated users and then applying clustering algorithms (k-means clustering

algorithm) on these logs to show trend and choices to the user. Also, different analysis of the subjects and trends will be shown to admin to analyze the web usage and user behavior.

2. AN OVERVIEW OF WEB-LOG MINING

Web usage mining deals with knowledge extraction from server log files. Web log mining is a branch of web usage mining as well as an important part of web mining. Through weblog mining, we can find the models that student's access to web pages; and then we can get useful patterns and forecast the behavior of students browsing. Web log mining also refers to the automatic discovery and analysis of patterns in clickstream and associated data collected or generated as a result of user interactions with Web resources on one or more websites. E-Learning system gets the search preferences of students, habits, and mode through web log mining to adjust the design of the website. Web log mining process will generally be divided into four phases: Data collection phase, data pre-processing phase, pattern discovery phase, pattern analysis phase.

Web usage mining deals with knowledge extraction from server log files. Web log mining is a branch of web usage mining as well as an important part of web mining. Through weblog mining, we can find the models that student's access to web pages; and then we can get useful patterns and forecast the behavior of students browsing. Web log mining also refers to the automatic discovery and analysis of patterns in clickstream and associated data collected or generated as a result of user interactions with Web resources on one or more websites. E-Learning system gets the search preferences of students, habits, and mode through web log mining to adjust the design of the website. Web log mining process will generally be divided into four phases: Data collection phase, data pre-processing phase, pattern discovery phase, pattern analysis phase.

3. PROPOSED MODEL FOR ASSOCIATION RULE MINING

In the web log extractor, we will take care of session management of web links that are visited by the user. And in data pre-processing module we will use some freely available data mining software systems that can be used for discovering association rules in the web log usage data. For this we will use C++ language or WUM prep scripts for web log pre-processing. The process to implement a new (and simple) web association rule mining system Figure 1.
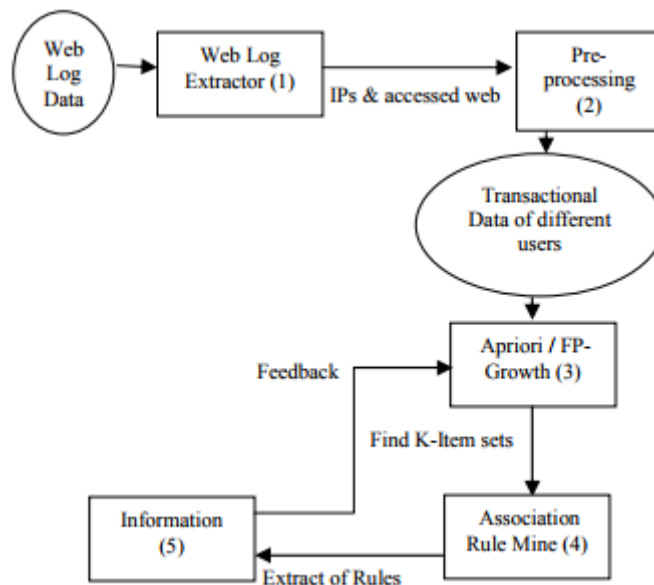
Figure 1 Proposed Model

We designed an efficient web association rule mining system, the best way to find areas where exactly improvement is needed. After finding, make a conscious how it can be an improvement. Following are the phase, which needs to be, take care during system development.

- Web log Extractor Module: We have developed this web extractor to extract the IPs and Web links from a web log file. It gives a File 1, which is used as an input for data preprocessing module.
- Data Preprocessing Module: This is a program which is developed by us to make input file containing navigational profile entries for APRIORI. We used link list data structure to solve our purpose. This Data preprocessing module produces an input file for APRIORI which contains an entry of different navigational profile. Here each IP works as user id and web line necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

4. SYSTEM ARCHITECTURE

In the traditional model of the online education system, student communicates with the web server through the internet and whenever a student is accessing a subject, there is no dynamism in the navigation of the pages i.e. the link available to each user remains fixed on the left navigation menu. Intelligent navigation provides a procedure to have dynamism in the navigation menu while the user is accessing the e-education system as shown.

Recommended modules includes are as follows:

- Log Capture module
- Intelligent Navigation module
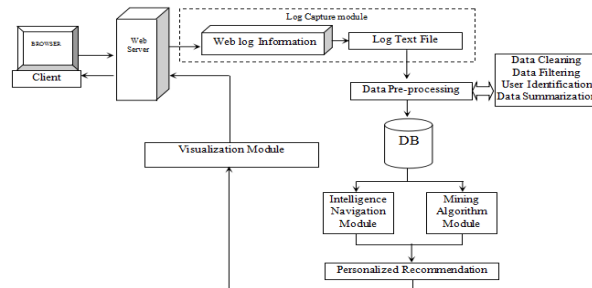- Mining Algorithm module
- Visualization module

Fig. Architecture of the system

Description:

Log capture module:

We have considered client log files which mostly consist of client IP address & authentication information. The client log files help to uniquely identify the users and the subject which the user will be accessing. So after capturing information in the log file (.txt format), we then converted it into the table with the attributes such as client-IP, user-id, subject-id, subject-count, position, visibility.

Intelligence navigation module:

User (student) is supposed to register for intelligence navigation system website and then allowed to log in to e-education system sites. The system records for the clickstream of the user on each subject and updates the subject-count on each access of the user to a specific subject.

Mining algorithm module:

Mining module incorporates association rules algorithm and clustering algorithm. Association log and clustering log table are been generated from the summarized log table. Association rules algorithm considers the counting occurrence of each rule followed by the support which describes, how often the rules occur in the database and confidence which measures the strength of the rules. Clustering algorithm includes clustering based on dissimilarity measure and real-time fuzzy clustering which works on the real-time log captured by the user and recommending the user to select the cluster in which user want to belong.

Visualization module:

Visualization module consists of a menu representation for the user. Menu consist list of subject-name and the assigned chapter to each subject. We have used relational database technology (SQL) to carry out a selection of the subject-name which the user request and display it in the navigation section of the user browser.

5. REFERENCES

[1] Dilip Singh Sisodia, Shrish Verma, Web Usage Pattern Analysis through Web Logs: A Review, 2012 IEEE 978-4673-1921-8/12.
[2] ThienThientAn, Web log cleaning for Mining of Web Usage Patterns, IEEE 978-1-61284-840-2/11.
[3] Renata Ivancsy, Ferenc Kovacs, Clustering Techniques Utilized in Web Usage Mining: Proceedings of the 5th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases, Madrid, Spain, February 15-17, 2006 (pp 237-242)
[4] A Survey on Methods used in Web Usage Mining, www.irjet.net, Rajinder Singh Rao, Jyoti Arora
[5] ISSN: 2454-132X (Volume2, Issue5) Available online at: www.Ijariit.com Web Usage Mining Tools & Techniques: A Survey
[6] Automatic Recommendations for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval.