Acquisition of User's Learning Styles Using Log Mining Analysis through Web Usage Mining Process

3 authors:

Sucheta Kolekar
Manipal Academy of Higher Education
37 PUBLICATIONS   **500** CITATIONS

SEE PROFILE

Sriram G. Sanjeevi
National Institute of Technology, Warangal
27 PUBLICATIONS   **242** CITATIONS

SEE PROFILE

D. S. Bormane
AISSMS-All India Shri Shivaji Memorial Society's College of Engineering, pune
93 PUBLICATIONS   **732** CITATIONS

SEE PROFILE

Acquisition of User's Learning Styles using Log Mining Analysis through Web Usage Mining Process

Sucheta V. Kolekar[1], Dr. Sriram G. Sanjeevi[2], Dr. D. S. Bormane[3]

[1] Research Scholar National Institute of Technology, Warangal, A.P., India
kolekar.sucheta@gmail.com
[2] National Institute of Technology, Warangal, A.P., India
sgsanjeevi@yahoo.com
[3] JSPM's Rajarshi Shahu College of Engineering, Pune, Maharashtra, India
bdattatraya@yahoo.com

Abstract: Web Usage Mining is a broad area of Web Mining which is associated with the Patterns extraction from logging information produced by web server. Web log mining is substantially the important part of Web Usage Mining (WUM) algorithm which involves transformation and interpretation of the logging information to predict the patterns as per different learning styles. Ultimately these patterns are useful to classify various defined profiles. To provide personalized learning environment to the user with respect to Adaptive User Interface, Web Usage Mining is very essential and useful step to implement. In this paper we build the module of E-learning architecture based on Web Usage Mining to assess the User's behavior through web log analysis.

Keywords: E-learning, Log Mining Analysis, Adaptive Learning styles, Web Usage Mining.

1. Introduction

Typically e-learning is Web based educational system which provides the same resources to all learners even though different learners need different information according to their level of knowledge, ways of learning style and preferences. Content sequencing of any course is a technology originated in the area of Intelligent / Adaptive Learning System with the basic aim to provide end user/student with the most suitable sequence of knowledge content to learn, and sequence of learning tasks (examples, exercise, problems, contents etc) to work with. To implement the Adaptive personalized learning system, different types of knowledge required which is related to learner's behavior, learning material and the representation of learning process [3]. Several kind of research is already addressed in the field of personalized e-learning; still there is a requirement to concentrate on the adaptation based on learning styles of the user. [7] In fact there are two basic classes of adaptation need to consider: Adaptive User Interface also called Adaptive Navigation and Adaptive content presentation.

Some of the Learning Systems focusing on static modules of contents, which can be confine the learner to gain knowledge initiatively in some degree. But because of the difference between learners in study purposes, abilities and cognizant of knowledge, it need to build the intelligent and individual learning platform for all learners to highly improve their enthusiasm for learning.

Ultimately the research emphasizes on following objectives for an Adaptive User Interface with respect to E-learning:

- Create personalized environment
- Acquisition of user preferences
- Take control of task from the user
- Adaptive display management
- Reduce information overflow
- Provide help on new and complex function

To achieve above objectives it is very essential to introduce web data mining technique. Web usage mining is dealing with the extraction of knowledge from web server log files. It mines the useful behavior to define accurate user profiles for the intelligent adaptive personalized e-learning system

The objectives of research in this paper:

1. Capture learning styles of an individual user/learner using log file method.
2. Improve the performance of web services.
3. Prepare web site structure to deal with users on an individual basis.
4. Provide accurate and complete picture of user's web activities.
5. Generate sufficient data like server side and client side logs to perform meaning full mining tasks in the phase of Pattern Analysis.

The paper is organized as follows; in section II basic architecture of Web Usage Mining is discussed with different kinds of applications. Section III talked about related work directly and indirectly existing on this issue. Section IV discussed about the proposed architecture about Web Usage Mining and detail description of steps
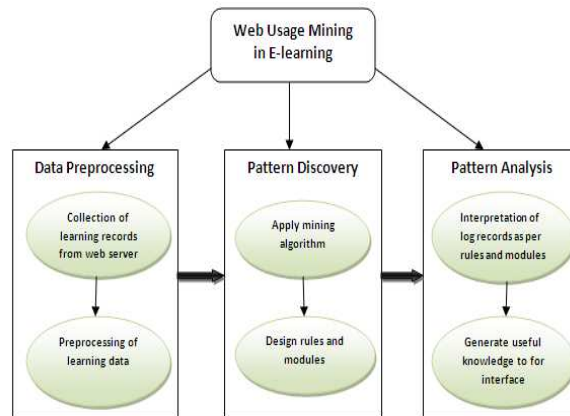
2. Web Usage Mining

Web Mining is divided into three important categories as per the part of Web based system are Web Content Mining, Web Structure Mining and Web Usage Mining. Web Content Mining deals with the discovery of useful information from the web contents. Web Structure Mining tries to discover the model of links structure from typical applications which are based on linked web pages.

2.1 Basic Architecture of Web Usage Mining in E-learning:

The general framework of Web Usage Mining is shown in fig. 1 for e-learning environment [12]. The first basic step of WUM is to collect and manage data related to users. It is called Data Preprocessing which includes Web Server Log files and some other important information like Student's registration details and learning information. Second step is Pattern Discovery which utilizes some mining algorithms to generate the rules and modules to extract the learning patterns of the

users based on learning styles which is recorded in log files. Pattern Analysis is the third step which is mainly converting the rules and modules into important knowledge by analyzing the user's usage which is ultimately the input to Interface component manager to change the GUI according to user's interest.

Fig. 1 Basic architecture of Web Usage Mining

2.2 Applications of Web Usage Mining:

1. Personalization Service: Personalization for a user is achieved by keeping track of previously accessed pages e.g. individualized profiling for E-Learning. Making adaptive interface on the basis of her/his profile in addition to usage behavior is very attractive and essential feature of e-learning in the field of Education. Web usage mining is an excellent approach for achieving this objective which is described in next section. It will classify the user's patterns as per the learning styles captured in log records. It can be used to find the learner's interests and preferences by mining single learner's browsing information such as[6] visiting pages, visiting frequency, content length of visit, time spent on each visit and preferences so as to provide each learner with the personalized adaptive pages which are accurate for his learning style and to forecast the learning behavior of each learned and to offer personalized education environment.

2. System Improvement: The improvement factor of the system is totally based on User's Satisfaction. The performance and quality of web site are the important measures of user's satisfaction. Web usage mining can provide useful knowledge and patterns to design of Web Server in a better way so that sever can focus on special features like [5] Page Caching, Network Transmission, Load Balancing, Data Distribution and Web Site Security.

3. Site Modification: The structure and interface of web site as per interest and contents are the key factors to attract learners to learn. Web Usage Mining can provide site improvement as per the mining knowledge and modify the structure as per the learner's navigation path and feedback. In adaptive environment of web

site, structure and interface of a web site changes automatically on the basis of usage patterns discovered from server logs.

4. Business Intelligence: Business Intelligence service is related to customer's information captured on web based system. In e-learning customers is nothing but learners whose learning behavior can be identified by web mining technique which will be ultimately used to increase the learner's satisfaction and to improve the business.

3. Related Work and Discussion

Up till now many papers have been suggested techniques related to Web Usage Mining and Log Analysis in E-learning environment. Xue Sun and Wei Zhao introduced how to use WUM in e-learning system which can be more intelligent and individual learning system and promote the interests of learners [2]. Shahnaz Nina et al. [3] propose technique of Pattern Discovery for web log records to find out the hidden information or predictive pattern by the data mining and knowledge discovery. Navin Kumar et al. surveyed about data preprocessing activities like Data Cleaning, Data Reduction. Some research is also done on personalized e-learning using different agents related to domains.

Our work is differs from above mentioned research in various aspects. As the main focus of research is to design an e-learning system with personalized adaptive interface, this paper is primarily focusing on the first step of personalization of users which is based on Web Usage Mining. The proposed framework of e-learning system is shown in fig. 2 [10] where we are implementing the Learning Style Acquisition phase using the advanced log analysis method of Web Usage Mining framework.
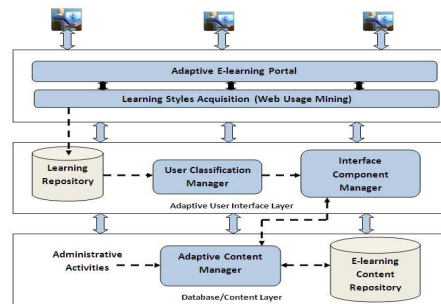
Fig. 2 Architecture diagram of E-learning System

The approach of architecture is as follows:

1. Learning Style Acquisition: In this phase Web Usage Mining technique is used to analyze the log data for identification of learning styles of different users/students.

2. User Classification Manager: The learning repository is the input for User classification manager where Back Propagation Neural Network algorithm of

Classification is performed to identify different kind of users based Learning style of Felder and Silverman.

3. Interface Component Manager: After identifying the categories of users the Interface component manager is changing the graphical representation of user interface as per user's need.

4. Adaptive Content Manager: This phase generates the adaptive contents based on user classification with the help of administrative activities and E-learning content repository.

4. Proposed approach of Learning Style Acquisition

In the field of web based e-learning, we are mainly emphasizing on the above mentioned to application areas: (i) Personalization and (ii) Site Modification (Adaptive User Interface). When users visit the site, they are interested in some course material, so they visit different pages. The e-learning sever log the information based on their visits. Through the log analysis and mining we can get the user's interest and behavior towards the pages visited. When users log on to the portal, the system will classify the users to different classes based on the previous behavior and generates the personalized page interface by adjusting the contents continuously and timely.

The idea of the architecture implementation:

1. Activity Recorder: Authentication of the user on e-learning portal and capturing of client side information through Activity Recorder.

2. Log Information: Capturing of Server side logs and proxy side log to pass through the data pre-processing with the additional information of user.

3. Data Pre-processing: Perform data cleaning, data integration, and data reduction steps to generate useful data for mining.

4. Clustering: Apply Usage clustering method for patterns discovery. The advanced k-means clustering algorithm is used to find out appropriate clusters based on user's usage.

5. Profile Generation: Generate user's profiles and content profiles according to clusters. The user's profiles are used to generate the learning styles and content profiles used to find out the domain interest of the user.

4.1 Steps of Web Usage Mining:

(i) Data Collection: The first step in the Web usage mining process consists of gathering the relevant Web data, which will be analyzed to provide useful information about the user's behavior.

Types of log files:

1. Server Side: The Extended Log Format (W3C) [7][1], which is supported by Web servers such as Apache and Netscape, and the similar W3SVC format, supported by Microsoft Internet Information Server, include additional information such as the address of the referring URL to this page, i.e., the Web page that brought the visitor to the site, the name and version of the browser used by the visitor and the operating system of the host machine. The Server side logs should contain the information of Web server and cached pages.

2. Client Side: [5] Client side data are the local activities collected from the host that is accessing the Web site using JAVA Wrapper technique. The local activities include actions of user on desktop like save/print the page, back/forward/stop the browser, email link/page, add a bookmark etc. This information is additional and reliable to understand the accurate behavior of the learner.
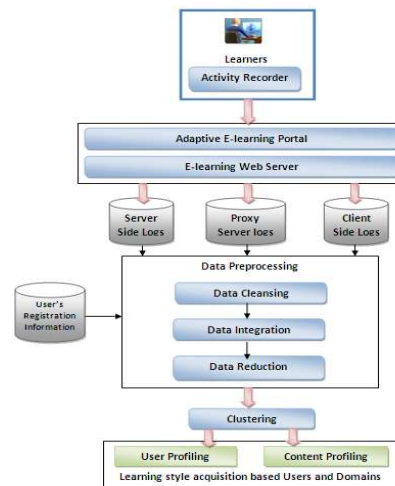
Fig. 3 Learning Style Acquisition Approach

3. Intermediary Side: Proxy Server Logs: The advantage of using these logs is that they allow the collection of information about users operating behind the proxy server, since they record requests from multiple hosts to multiple Web servers.

(ii) Data Preprocessing: The captured log files are not suitable directly for data mining techniques. Files must be gone through the three data pre-processing steps

1. Data Cleansing: Useless information removal e.g. graphical page content [6]. An algorithm for clearing the entries of log information:

(i)   Removal of picture files associated with request for particular pages:

(ii)   Remove status of error or failure on different pages.

(iii)   Automatically generated access records should be identified and removed.

(iv) Entries with unsuccessful HTTP status code should be removed. Codes in between 200 to 299 are successful entries.

2. Data Integration: Integration of cleaned data is the process of identification and reconstruction of user's sessions from log files.

This phase of prediction is divided into two basic steps:

User Identification: The identification of different users based on three ways: (i) By converting IP address to domain name exposed some knowledge. (ii) Cookies help to easily identify the individual visitors, which gives information regarding the usage of website. (iii) Records of cached pages are used to find out the profiles.

Session Identification: The need of session identification is to separate out the different sessions of same user by checking threshold value. Usually threshold of each session considered as 30 min. time interval [3].

3. Data Reduction: Need to reduce the Dimension of data to decrease the complexity. Access log files on the server side and proxy side consist of log information of user's sessions. These logs include the list of pages that a user has accessed in one single session. The log format of the file is in Extended Log File Format which includes special records. The information in this record is sufficient to obtain session information. The set of URLs of particular pages are forming a session which should satisfy the requirement that the time of elapsed between two consecutive requests is smaller than a given *t*, which is accepted as 30 minutes threshold value.

After preprocessing of log files the following fields are used for research:

1. Users: In e-learning system it refers as a learner who visits the e-learning portal with different learning styles.

2. Page view: A page view can get users by clicking on the page once which can be used to represent as one learning behavior.

3. No. of click streams per session: Click streams are nothing but the user's page requests which can be considered as a learning sequence.

4. User Session: All sequence of clicks from that user visits from whole website i.e. aggregation behavior of the user.

Evaluation of Parameters for Method:

1. **Topics (T)** are related to the contents of the web site and are defined by the owner of the portal.

2. **Weight (W)** defines the importance of the topic based on the actions.

3. **Actions (A)** is nothing but the clicks of the student on particular content type like text links, video lectures, downloadable link etc. Each action can be defined by the weight as per the importance. E.g. A1=PageRequest and weight WA1=1 or A2=VideoLecture and weight WA2= 4 or A3= Download PDF and weight AW3= 8.

   Action A1 is default action for any other action.

4. **Duration (D)** is the time; a student spends on a page which will give us the interest area based upon actions on pages.

   To calculate the exact duration, there is one problem related to the time a student spends. You cannot predict whether the student is really reading the page or might be sleeping. To solve this problem we are considering the time duration up to timeout of the login.

$$D(i) = \frac{\sum_j (D(P_j) \cdot W(P_j))}{\sum_k D(P_k)}$$
(i)

(P_j): each visited page that contains Topic TP_j

$W(P_j)$: weight of the page associated according to contents

Different pages visited as per Learning Styles:

The Web Usage Mining architecture we propose aims to find a mapping of student's actions on the browsers to learning styles they fit. Based on the formula derived in equation (i) we can find the duration spent on number of pages of E-learning portal.

The observed actions are as follows:

1. Access of contents and Reading Material:

$$D(i) = Content_{Visit}$$
$$D(i) = Content_{reading}$$
$$D(i) = ContentList_{Visit}$$

2. Access of Examples:

$$D(i) = Example_{reading}$$

3. Exercises or Quiz

$$D(i) = Quiz_{Visit}$$
$$D(i) = Quiz_{solving}$$
$$D(i) = Downloading\ Exercise$$

4. Chat Usage/Forum Usage/Email Usage: Student may use the chat or forum or email service for social communication based on contents.

(iii) Clustering: There is a need of clustering the user's profiles and the contents profiles based on the log information. Clustering is an unsupervised classification method which groups the objects together based on the similarity feature into the same cluster [2]. Clustering can be possible by two ways which are partitioned based and hierarchical based. Partitioned clustering is to separate the records of n objects into *k* clusters which include most similar objects into *n* groups; the separation depends on distance measure. In this approach we are using popular *k-means* clustering method with some advance features to find out most frequent pages access by the user on specific contents of domain.

```
192.168.100.3-[30/Dec/2010:18:36:55-0800] "GET /templates
/dbms/introvideo1/video1.avi: HTTP/1.1" 200 188 "http://www.e-
learning.recoe.in/" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 7 )"
192.168.100.3  -  [30/Dec/2010:18:56:56  -  0800]  ""  http://www.e-
learning.recoe.in/" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 7);
```

Fig.4. Web Log Record

Page Contents	No. of users accessing contents or frequency
Text Data/Read:View	312
Video Lectures	64
Graphical Data	47
PPTs/Demos	23
Downloadable pdfs	77
Exercise	55
Reference Links/Online Resources	22

Fig. 5. Web page contents and frequency
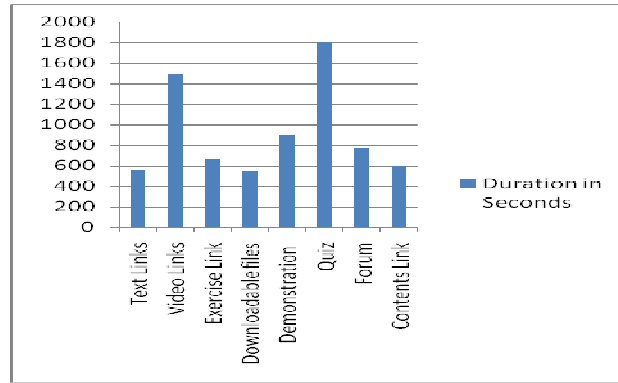
Fig. 6. Histogram of calculated durations of different pages and contents.

5. Experimental Details and Results

Although user registered his or her favorite domains at the time of registration, user's identity or work will influence the style of content reference of each visit, for example user submits the domain of interest called "database management systems" but to understand the concepts of the database user always prefer the contents based on his/her individual learning style. The styles of user we can get by analyzing the log records. It's sure that the definition of interests is not easy. In order to get satisfied criteria for interests, we have done a great deal of experimentation. We employ the extended data which is a kind of client level Web log data and server level log data that are recorded, since the data contains behaviors of a large numbers of users among the investigation. Based on the mentioned steps of web usage mining we built a system on extracting users' Interests. The experiment on used web log data, collected from www.e-learning.rscoe.in web server (see in Figure 4). This record is used get web pages access the by user. The accessed web pages contain different type of contents for the particular search topic. According to user's interest they can access different links. We have recorded 600 users log records and the frequency of the contents (see in figure 5). After Clustering we can define the 7 types of clusters as per web page contents to decide user profiling and content profiling. The graph (see in figure 6) shows the number of users accessing and spending time on the different pages of portal which are useful input for Neural Network Algorithm to classify different Learning Styles.

6. Conclusion

In this paper we proposed approach of Web Usage Mining by surveying data preprocessing activities and different kinds of log records. Web Usage Mining for e-learning environment mines the log records to find the user's usage patterns to provide users with personalized and adaptive session. The next phase of research is to use the effective User's profiles as input parameters to Neural Network based algorithm to classify the users as per the Felder & Silverman Learning style model. According to classified users the interface components can be changed adaptively on website by using adaptive contents and administrative activities.

References

1.   "Extended Log File Format": http://www.microsoft.com/technet/prodtechnol/WindowsServer2003/Library/IIS/676400bc-8969-4aa7-851a-9319490a9bbb.mspx?mfr=true

2.   Xue Sun, Wei Zhao (2009), *"Design and Implementation of an E-learning Model based on WUM Techniques"*, IEEE International Conference on E-learning, E-business, Enterprises Information Systems and E-government.

3.   Shahnaz Nina, Md. Mahamudur Rahaman, Md. Khairul Islam, Khandakar Entenam Unayes Ahmed (2009), *"Pattern Discovery of Web Usage Mining"*, International conference on Computer Technology and Development, Published by IEEE Computer Society.

4.   Rozita Jamili Oskouei (2010), *"Identifying Student's Behaviors Related to Internet Usage Patterns"*, T4E 2010, 978-1-4244-7361-8/ 2010 IEEE.

5.   Xinjin Li, Sujing Zhang (2010), *"Application of Web Usage Mining in e-learning Platform"*, International conference on E-business and E-government, published by IEEE Computer Society.

6.   Navin Kumar Tyagi, A.K. Solanki, Sanjay Tyagi (2010), *"An algorithmic approach to data preprocessing in Web Usage Mining"*, International Journal of Information Technology and Knowledge Management July-Dec 2010 Vo. 2 No. 2 pp 269-283.

7.   Mohammed Koutheair Khiribi, Mohamed Jemni, Olfa Nasraoui (2009), *"Automatic Recommendations for E-learning Personalization based on Web Usage Mining Techniques and Information Retrieval"*, Eight IEEE International Conference on Advanced Learning Technologies, published by IEEE Computer Society.

8.   Farah Chanchary, Indrani Haque, Md. Saifuddin Khalid (2008), *"Web Usage Mining to evaluate the transfer of learning in a Web-based Learning Environment"*, Workshop on Knowledge Discovery and Data Mining, published by IEEE Computer Society.

9.   Ling Guo, Xin Xiang, YuanChun Shi (2004), *"Use Web Usage Mining to Assist Online E-learning Assessment"*, Proceeding of the IEEE International Conference on Advanced Learning Technologies.

10.   Sucheta Kolekar, Dr. Sriram G. Sanjeevi, D. S. Bormane (2010), *"The Framework of an Adaptive User Interface for E-learning Environment using Artificial Neural Network"*, 2010 WORLDCOMP/ International Conference on e-Learning, e-Business, Enterprise Information Systems, and e-Government (EEE'10: July 12-15, 2010, USA).

11.   Penelope Markellou, Loanna Mousourouli, Sirmakessis Spiros, Athanasasios Tsakalidis (2009), *"Using Semantic Web Mining Technologies for Personalized E-learning Experiences"*, International Journal in E-learning.

12.   Resul Das, Ibrahim Turkoglu (2009), *"Creating mining data from web logs for improving the impressiveness of a website by using path analysis method"*, Journal of Elsevier, Expert Systems with Applications 36 6635-6644 Science Direct.