Analysis of Web Usage Behavior using Pattern Analyzing Techniques in E-Learning system

Ms.G.Dona Rashmi
Dr.V.Narayani

Abstract - The abnormal growth of web development which incorporates vast amount of user data leads to the user behavior analysis for web mining. Web mining is the recent research area which comprises data mining and World Wide Web. The problem of analyzing the user behavior for a web mining domain is an important issue in E-learning system nowadays. In this paper, we proposed a novel integrated technique for analyzing the user behavior in Digital learning in an education environment. The proposed technique is initiated with pattern analysis techniques followed by web usage analysis with the final implementation of resource analysis algorithm for an E-learning system.

In near future, we will extend our proposed novel integrated technique for analyzing web usage behavior in computer mediated Electronic communication system with heterogeneous structure. The performance analysis will also taken into account for the future enhancement with Fuzzified predictive syndromes for various architecture in an E-learning system.

Index Terms – Web mining, E-learning, Digital library, log data, path analysis.

I. WEB MINING
Web Mining is the application of data mining technique to discover and retrieve useful and interesting patterns from web data. Web data contains different kinds of information including web documents data, web structure data and web log data [1]. According to the kinds of data to be mined, Web Mining can be broadly divided into three categories: Web content mining, Web structure mining and Web usage mining.

Manuscript received March, 2013.

Ms.G.Dona Rashmi, Lecturer/MCA /Karpagam College Of Engineering, Coimbatore, India, 9715887735.

Dr.V.Narayani, ASP/MCA /Karpagam College Of Engineering, Coimbatore, India, 8903432375.

A. Web Content Mining:
Web Content Mining is the task of extracting knowledge from the content of documents on World Wide Web. The Web document usually contains several types of data, such as text, image, audio, video, metadata and hyperlinks. Some of them are semi-structured such as HTML documents or a more structured data like the data in the tables or database generated HTML pages, but most of the data is unstructured text data. The unstructured characteristic of Web data forces the Web content mining technique towards a more complicated approach. Web content mining describes the automatic search of information resources available online, and involves mining Web data content [2]. It focuses on techniques for searching the web for documents whose content meets a certain criterion. The documents found are used to build a local knowledgebase.

B. Web Structure Mining:
Web Structure Mining is the process of discovering structure information from the Web. It describes the connectivity in the Web subset based on the given collection of interconnected Web documents. The structure of a typical Web graph consists of Web pages as nodes, and hyperlinks as edges connecting related pages. A Hyperlink is a structural unit that connects a location in a Web page to different location, either within the same Web page or on a different Web page. A hyperlink that connects to a different part of the same page is called an Intra-Docment Hyperlink, and a hyperlink that connects two different pages is called an Inter-Docment Hyperlink. Mining the structure and Web page structure, can be used to guide the classification and clustering of pages to find authoritative pages to improve retrieval performance.

C. Web Usage Mining:
Web Usage Mining is the application of data mining techniques to discover interesting usage patterns from Web data, in order to understand and better serve the needs of Web based applications. Usage data captures the identity or origin of Web users along with their browsing behavior at a Web site. The strategic goals of Web usage mining are...
prediction of the user’s behavior within the site, comparison between expected and actual Web site usage, and adjustment of the Web site with respect to the users’ interests. The first type of source data for web usage mining is Web Server Data. When web users interact with a site, their behavior is recorded in web server logs on web server. These log files may contain valuable information characterizing the users’ experience in the site. Some of the typical data collected at a Web server include IP addresses, page references, and access time of the users. In addition, since in a medium size site log files amount to several megabytes a day, there is a necessity of techniques and tools to help take advantage of their content. Application Level Data is another source for web usage mining. With this type of data it is possible to record various kinds of events in an application. These data is used for generating histories about selected special events [3]. The data in this category can be divided into three categories based on the source of its collection: on the server side, the client side, and the proxy side.

II. PATTERN ANALYSIS TECHNIQUES

Pattern analysis techniques involve algorithms to discover interesting patterns from web data. Once user transactions or sessions have been identified, there are several kinds of access pattern mining that can be performed depending on the needs of the analyst. Some of these techniques are discussed below.

A. Path Analysis

Graph models are most commonly used for Path Analysis. In the graph models, a graph represents some relation defined on Web pages and each tree of the graph represents a Web site [4]. Each node in the tree represents a web page and edges between trees represent the links between web sites and the edges between nodes inside a same tree represent links between documents at a web site. When path analysis is used on the site as a whole, this information can offer valuable insights about navigational problems. Most graphs are involved in determining frequent traversal patterns and more frequently visited paths in a web site.

B. Association Rules

Predict the association and correlation among set of items where the presence of one set of items in a transaction implies with a certain degree of confidence the presence of other items. That is, it can discover the correlations between pages that are most often referenced together in a single server session/user session. It can provide the information: What are the set of pages frequently accessed together by web users? What page will be fetched next? What are paths frequently accessed by web users?. Implement association rules to on-line shopper can generally find out his/her spending habits on some related products.

C. Sequential Patterns

Sequential patterns discovery is to find the inter-transaction patterns such that the presence of a set of items is followed by another item in the time-stamp
ordered transaction set. Web log files can record a set of transactions in time sequence. Using sequential pattern discovery, useful user trends can be discovered, predictions concerning visit pattern can be made, website navigation can be improved and adopt web site contents to individual client requirements or to provide clients with automatic recommendations that best suit customer profiles.

D. Decision Trees

A decision tree is essentially a flow chart of questions or data points that ultimately leads to a decision. They are great for situations in which a visitor comes to a Web site with a particular need [5]. But once the decision has been made, the answers to the questions contribute little to targeting or personalization of that visitor in the future.

E. Clustering

Clustering identifies visitors who share common characteristics. After you get the customers’/visitors’ profiles, you can specify how many clusters to identify within a group of profiles, and then try to find the set of clusters that best represents the most profiles [6].

III. METHODOLOGY

A. Association Rule

- More number of III years and less number of II Year students test their skill in Placement Training after their reference of technical Level 1, Level 2, and Level 3 Pages.
- Students who spend more time on Research subject refer less in Placement Training Pages.
- Students who refer all the web pages sequentially/hierarchically surprisingly hold good CGPA results.

B. Sequential Pattern

- Students refer Level 1 in General Knowledge refers Level 2 often rather than with Level3.
- Student refer Advanced Technical Subjects also refers Research Webpage.

C. Decision Trees

Placement Motive Students => Test Skills in Placement Level1, Level2 and Level3.
Research Motive Students => Research subject Scope Webpage References.

IV. RESULT AND DISCUSSION

In a Abcd Engineering college, we perform our research analysis over a sample space of 282 students for accessing the E-Learning environment. We obtain the following results.
The students accessing feature sustains in the following tree structure.

```
General
  Level 1 41 53 39
  Level 2 09 21 13
  Level 3 03 12 02

Subject
  Academic 23 56 26
  Technical 20 40 20
  Research 13 06 24
```

Fig. 3: Tree Structure of student's access

<table>
<thead>
<tr>
<th>Category/Year</th>
<th>I Year</th>
<th>II Year</th>
<th>III Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total students</td>
<td>70</td>
<td>114</td>
<td>98</td>
</tr>
<tr>
<td>General</td>
<td>41</td>
<td>53</td>
<td>39</td>
</tr>
<tr>
<td>Subject</td>
<td>56</td>
<td>102</td>
<td>70</td>
</tr>
</tbody>
</table>

The Final year students give more importance to research rather than with the first and second year students; the subject refreshment for E-Learning is not an important criteria for the Final year students.

### Table 2: Students Browsing Type

<table>
<thead>
<tr>
<th>Year</th>
<th>Type of Browsing Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>I year Students</td>
<td>Subject Pages</td>
</tr>
<tr>
<td>II Year Students</td>
<td>Subject &amp; Technical Pages</td>
</tr>
<tr>
<td>III Year Students</td>
<td>Technical &amp; Research Pages</td>
</tr>
</tbody>
</table>

### Table 3: Analysis of students browsing

<table>
<thead>
<tr>
<th>Predicts Opportunities Placement</th>
<th>GAT</th>
<th>20/70</th>
<th>40/114</th>
<th>20/98</th>
</tr>
</thead>
<tbody>
<tr>
<td>More Opportunities for Research</td>
<td>GAR</td>
<td>13/70</td>
<td>6/114</td>
<td>24/98</td>
</tr>
<tr>
<td>More Research / Less Placement</td>
<td>GTR</td>
<td>13/70</td>
<td>6/114</td>
<td>20/98</td>
</tr>
<tr>
<td>More Placement / Less Placement</td>
<td>ATR</td>
<td>13/70</td>
<td>6/114</td>
<td>20/98</td>
</tr>
<tr>
<td>GATR</td>
<td>13/70</td>
<td>6/114</td>
<td>20/98</td>
<td></td>
</tr>
</tbody>
</table>

The E-Learning environment in Abcd Engineering College provides an optimal source for knowledge update and commercial value incorporating rather than with the ideal research improvements.

V CONCLUSION

The Web is an excellent tool to deliver on-line courses in the context of E-learning System. However, counting only on web traffic statistical analysis does not take advantage in the potential of hidden patterns inside the web logs. Web usage mining is a non-trivial process of extracting useful implicit and previously unknown patterns from the usage of the Web. Significant research is invested to discover these useful patterns to increase profitability of e-Learning sites. While some tools using data
mining techniques to help educators and learners are being developed, the research is still in its infancy. In addition, with the awareness of the potential advantages of integrated web usage mining and the insufficient data recorded by web servers, there is a need for more specialized logs from the application side to enrich the information already logged by the web server. This added value by specific event recording on the e-learning side will give click streams and the patterns discovered a better meaning and interpretation.

REFERENCES