Evaluating Learning Behavior of Web-Based Training (WBT) using Web Log

Albert K W Wu and C H Leung
Faculty of Education
Chinese University of Hong Kong
HONG KONG

Abstract
With the rapid development of Internet and World Wide Web (WWW), Web-Based Training (WBT) is developed in a phenomenal pace recently. By collecting data from the relevant Web log, learners’ behavior may be evaluated and thus be used to analyze system performance and improve system design. In this paper, we report such an effort - an ongoing project towards evaluating students’ learning behaviors by tracking the Web log from a sample WBT environment.

1. Introduction

With the Web becoming the most popular media for collecting, sharing and distributing information, it is very common for educational institutions and organizations to develop Web-Based Training (WBT) systems. WBT can be explained as online environments that utilize the WWW technologies as a scaffold to facilitate teaching and learning process [3]. Instructors provide resources such as notes, multimedia, information, discussion and different learning activities in the WBT environment, while learners are expected to peruse the resource and participate in activities. Since the learner Web access data can be collected unobtrusively and automatically during the actual use of the site in an authentic context, analysis of log file data can give insight into the real-world learner’s behavior such as his/her Web-learning activities and the nature of the Web traffic [2]. By analyzing the Web log, instructors can know who is using the materials, from where and when. If the WBT site contains a high degree of interactivity, it is also important for knowing what operating systems and browsers the students use in order to optimize the commercial efficiency for most learners.

2. The Environment

In essence, the teaching environment is a teaching web-site on a commercial authoring software for the teaching of IT (Information Technology) and multimedia courses for Education major students. It consists of three major components: (1) Online-Tutorials, (2) Projects - a virtual space for collecting multimedia projects, (3) Information – consisting of a calendar schedule and message board. The learning behaviors of students in the sample site are monitored throughout the semester to collect information about students’ learning.

The Log Data
The original log file entries include the following information: the domain name (or IP address) of the request, the data and time of the request, the method of the request, the location and name of the file requested, the result of the request and the size of the data transferred. An example entry is given below:

137.189.209.6 - - [28/May/2002:23:44:46 +0800] "GET /dumper.jsp HTTP/1.1" 200 496
137.189.209.6 - - [28/May/2002:23:45:51 +0800] "POST /dumper.jsp HTTP/1.1" 200 335003
137.189.164.54 - - [29/May/2002:10:17:16 +0800] "GET /admin/login.jsp HTTP/1.1" 200 4546

3 Preliminary Result Obtained

The Challenge
There are several general challenges associated with obtaining due results from the data. Firstly, extraneous information is mixed with useful one, e.g. log entries that record the request failures, page graphics, etc. are irrelevant to our current analysis. Secondly, multiple server requests may be generated by a single user action, e.g. when a user displays all unread messages in a virtual conference, this single action generates three entries. Thirdly, multiple user actions may generate the same server request. Fourth, local activities (e.g. browser navigation using “back” and “forward” buttons) are not recorded.

Data Analysis
In general one can relatively easily get the basic descriptive statistics from a standard statistical software. Such descriptive analyses can provide some global data characteristics such as the various summaries for request,
domain, event, session, bandwidth and error. Some interesting observations:

- The Online-Tutorials session was found to be the most used components since it uses a step-by-step instruction to help students. On the other hand, the Information session was found to be the least used component with only a few students have tried to post message to the message board. This can be explained as most learners want to create or join the discussion only if marks or extra bonus would be given for their contribution [1].

- Hourly accesses on weekdays are confined largely between 10am and 10pm with still a significant number of accesses in the evening until very late at night, as well as in the very early morning! This indicates a late learning habits of students.

- In relating student access with respect to the examination and assignment submission deadline, it is found that most of the accesses (80%) concentrate on the last half-month of the semester.

- Microsoft (MS) Windows98 is the most popular operating system accounted for about 80% of all page accesses. While there are learners using Windows 95 and Windows98 (>80%), only very few are using Linux and Macintosh. And the most frequent pairing of browser and operating system is found to be Microsoft Internet Explorer (IE) 6 and Windows98. Based on this, we can then determine which browser the site should offer most support to. As almost all the learners use the MSIE, we should provide more support for the development of WBT modules on MSIE and advise users of MSIE4 to upgrade to a more recent version.

4 Conclusion

At the moment only very preliminary results and descriptive analysis are done with more in-depth analyses on learner behaviors such as behavioral patterns across users, changes over time and in various ways, etc. are to be investigated; with more specialized data mining and visualization tools for and analysis of time-series event data are also required.

References

