Distance Education: a Web Usage Mining Case Study for the Evaluation of Learning Sites

Leticia dos Santos Machado Karin Becker

Faculdade de Informática – Pontifícia Universidade Católica do Rio Grande do Sul - Brazil
E-mail: lmachado@inf.pucrs.br, kbecker@inf.pucrs.br

Abstract
Web Usage Mining (WUM) focus on the interaction behavior between web users and requested Web pages in order to identify navigation patterns. This work describes a case study aimed at investigating the potential of WUM as a framework for supporting the validation of learning site designs. The goal was to model the domain in terms of a WUM application, and to explore abstractions and types of patterns that can help site usage evaluation.

1 Introduction
Web-based learning environments are designed as a set of web pages that constitute the educational site. The distributed and virtual interaction model of web-based learning environments makes difficult the observation and evaluation of the actual usage of the learning resources available at the site. To monitor students’ behavior in terms of site usage, web course management infrastructures (e.g. WebCT1) typically offer functionality such as access statistics, recently accessed pages, or participation in communication tools. These options present limited capacity in the analysis of students’ usage tendency and perception of a learning site. As a consequence, the evaluation of the adequacy and effectiveness of an educational web site as a learning environment is hard and subjective.

Web Usage Mining (WUM) addresses the application of data mining techniques over web data in order to identify and characterize users’ navigation behavior patterns and tendencies [1]. The extraction of information about users’ accesses is mainly obtained from Web server logs. The WUM process is divided into the following phases [1]: data pre-processing (used to select, clean and prepare the log raw data), pattern discovery (application of data mining algorithms, such as association, sequence, etc), and pattern analysis (evaluation of yielded patterns to seek for unknown and useful information). WUM has been extensively applied in the electronic commerce area (e.g. [2], [3], [1]), and its benefits are being extended to other areas such as distance education [4]. This work describes a case study aimed at investigating the potential of WUM as a framework for supporting the validation of learning site designs. The case study was developed at the distance education department of a large Brazilian university.

The goal of the case study was to model the domain in terms of a WUM application, and to explore types of patterns that could help in site usage evaluation. The paper describes the process framework applied, which is learning activity-oriented and focuses on students’ actions over technological resources available at the learning site.

2 A Case Study
This case study aimed at investigating the potential of WUM as a framework for supporting the validation of learning site designs, given that functionality offered by WebCT for interpreting users’ behavior was considered too simple. The analyzed course lasted 11 days, resulting in a log of 15,953 accesses. In order to understand how students interact with the learning web site, specific interpretations were made over the server log data, which took into account the characteristics of the domain. Emphasis was settled into site topology and navigability. The goal was to model the domain in terms of a WUM application, and to explore abstractions and types of patterns that could help in site usage evaluation.

2.1 Pre-Processing
The analysis performed revealed that the main focus was in the relationship between the learning activities/tasks and the technological resources needed for their execution. Therefore, each activity planned for the course was used as a unit for guiding and orienting a WUM dedicated process. The decision to develop the WUM process centered on the activities has defined a paradigm for the search of navigation patterns, which is focused on the actions taken by students over the technological resources in order to perform each activity planned for the course. The analysis of students’ navigation behavior based on course activities enables to evaluate the structure of the site model, that is, how the pages of the environment are perceived and used, as well as to evaluate students’ learning models with regard to site structure.
The activity paradigm resulted in a dedicated WUM process based on the following criteria:

- Activity-oriented process
- Taxonomy of technological resources
- Learning activity session
- Search for patterns focused on use of resources.

The pages of a virtual environment learning can be perceived according various abstraction levels, constituting a multilevel resource taxonomy (e.g. a communication resource encompasses specific ones such as email, chat, etc). The resulting taxonomy represents the mapping of the technological resources needed for the execution of learning activities. The activity paradigm is also used to limit the duration of an activity session and to relate students’ accesses to the execution of an activity under study. An activity session includes all accesses for the execution of a learning activity, which can range from minutes to days, possibly implying that students logged in and out the learning site several times.

2.2 Pattern Discovery

Pattern discovery focuses on the use of learning resources for the execution of an activity. This case study explored two mining techniques: association and sequence [1]. Figure 1 exemplifies a sequential pattern extracted for a learning activity analyzed. Sequential patterns describe related page accesses in a specific order (e.g. page “a” followed by page “b”). In the figure, numbers represent auxiliary pages (e.g. “course environment”), whereas labels represent an abstraction of a specific page in term of a resource class, according to the built taxonomy (e.g. CT – Chat, SU – Assignment Submission).

![Figure 1. Navigation Sequence for Example 1](image)

This pattern captures users that initially log in the system (class AT - Authentication), access auxiliary pages 15 and 26, and then access a page of the class Submission (SU). Then, they return to page 15, access auxiliary page 206, and finally reach a Chat page (CT).

This pattern might reflect a problem related to the submission of assignments specific to the WebCT Infrastructure. Given that students have difficulty on the technical operation of the WebCT submission functionality, very frequently they use the chat for clarifying doubts. This interpretation was advanced by the domain expert, and the WUM validation phase must confirm or reject this interpretation hypothesis.

2.3 Pattern Evaluation

In this phase, patterns obtained are evaluated in terms of their validity and meaning in the domain, which is by no means an easy task. The evaluation phase was not completed in this case study, given that our purpose was merely to demonstrate possible patterns that could be extracted, and to understand their power as an instrument for site evaluation and restructing, considering the criteria and abstractions applied to obtain them.

3 Conclusions and Future Work

This case study approached the evaluation of educational site design as a WUM application. Focus was set on site structure, as well as usage and combination of technological resources for accomplishing a learning activity. The WUM process adopted considers characteristics that are specific to the learning context. Even considering that the pattern evaluation phase was not completed, the results obtained demonstrated that the abstractions adopted and the yielded pattern types are suitable and useful for site usage evaluation. The domain expert has shown particular interest on sequential patterns, for they show most frequent students’ trails on the site. These were considered a powerful tool for identifying possible site restructuring points. It is hoped that site restructuring becomes a less hazardous activity if a comprehensive WUM infrastructure becomes available. The findings of this case study, of course, lack generalization, and the application of this process in other courses is planned. Other research topics currently under investigation include: definition of an infrastructure targeted at web learning WUM, development of tools for supporting the pattern evaluation phase, variations on the concept of learning transactions, etc.

References