A Framework for Using Web Usage Mining to Personalise E-learning

Hafidh Ba-Omar, Ilias Petrounias and Fahad Anwar
School of Informatics, University of Manchester
{Hafidh.ba-omar;fahad.anwar}@postgrad.manchester.ac.uk
Ilias.Petrounias@manchester.ac.uk

Abstract

Web usage mining can contribute to finding significant educational knowledge. It can play a vital role in the personalisation aspects of any domain. We propose a framework for personalizing e-learning that necessitates careful attention towards individual learning styles. We focus on identifying learning patterns of learners and the sequence of choosing learning resources in relation to their learning styles. A prototype for an adaptive web based course has been developed where the learning environment is modifying its behaviour to reflect learning styles.

1. Introduction

There are opportunities in e-learning to know about learners’ preferences and rate of learning. Web usage mining in this context can be seen as extracting unknown actionable intelligence from interaction with an e-learning environment. The synergy of web usage mining and learning theories, such as the learning styles theory, is rarely considered when building learning systems. We propose a framework to enhance personalised e-learning. Considering learning styles is based upon the assumption that learners demonstrate patterns of behaviour appropriate to their learning styles. Identification of students’ learning styles is a step towards personalising web based learning that addresses different learning needs. The use of mining techniques to identify learning styles has been proposed in [1]. Here we are combining two approaches to detect and use learning styles: implicit model using two data mining techniques (association rules and sequential patterns) and explicit online learning style questionnaires utilising aspects of two learning style theories: Felder and Silverman [2] and VARK¹.

2. Experimental work (Pattern Discovery)

The scenario reported here addresses the issue of identifying learning patterns. We carried out two experiments to validate the framework. The main objective of the first one was to analyse web access logs to identify and analyse the learning patterns and purposely the learning styles patterns in offline mode. The second one was carried out in online mode to demonstrate the adaptive behaviour of an e-learning course based on changing of learning preferences. The data used in this study was collected from the web log of WebCT server in SQU University. We focused our analysis on a well-designed course as a test bed for our framework. The course was an English language (1st year students) with 47 participants. To identify the learning styles of participants, we have converted the paper based questionnaire of the VARK Model into an online questionnaire. SPSS Clementine software has been used as the data mining tool. The aim of the pre-processing phase is to convert the raw data into a suitable input for the next stage mining algorithms. There are four general tasks during this phase.

1. Data cleaning: In this stage, data has been cleaned from noise, irrelevant items were removed, such as icons or style sheets. In our case, the whole log file consists of 11,386,063 hits for a period of three months. After performing this task, only 169,669 entries were accepted to be used in the next task.

2. User identification: In the e-learning context, unlike other web based domains, user identification is a straightforward problem as in most cases the learners must login using their unique ID.

3. Server Session Identification: This task consists of grouping a learner’s page accesses in a unit named session by dividing the click stream of each learner into sessions. We have adopted a 30-minute timeout to start new session.

4. Data Enrichment: this task consists of integrating the web log with other sources about the learners, such as their learning styles, grades or demographic data (gender, age, etc). The process of mapping many URLs into meaningful and semantic structure was essential in this task. The patterns required for personalised e-learning correspond to the behaviour and interests of learners. Although, we are interested in

finding patterns that are valuable and interesting, the main focus was on supporting the potential of the proposed framework. We have focused only on sequential patterns and association rules. The main observations that were yielded are:

a) A lot of sequential patterns rules that demonstrate students learning path and the sequence of choosing learning media to achieve the learning objective (Figure 1). For illustration purposes, we focused on specific sequences which might indicate certain a learning style to be located in the existing data.

| Rule1 (visual): Page 1A ➔ Page 2A ➔ Page 3A ➔ Page 4A |
| Rule2 (Audio): Page 1B ➔ Page 2B ➔ Page 3B ➔ Page 4B. |

The first rule states if the learner chooses the learning media in this order, the interpreted behaviour is visual style. The second rule is be a sign for an audio style. The minimum support and confidence have been set to 20 and 50 respectively.

b) We used the Apriori algorithm to obtain association rules to reveal the viability of detecting students’ patterns in selecting the learning resource type (audio, textual or visual). The aim was finding association rules with a specific rule consequent (student’s learning styles) to detect the learning materials linked with a specific learning style. Educational knowledge from the pattern discovery phase is evaluated for validity and meaning. The sequential patterns were used for three purposes after applying sequential rules to existing or new data stream: Sequence detectors indicate whether a target sequence is found within the navigational profile of each learner. It was a good tool to identify students who were affected by those patterns. Sequence counters count the number of times the target sequence appears for each learner as an indicator for the intensity of pattern. Sequence predictors add predictions to the data stream based on the target sequences. The recommendations provided to learners are ranked and offered based on those predictions.

3. Adaptive web based course

We devised a prototype course to reflect the discovered patterns of learning styles (Figure 2). An experiment at the course level was carried with 10 postgraduate students. They filled in the VARK questionnaire to identify the initial learning styles. The system has the mechanism to present the learning material using the initial learning style preference. After a while the system adapts the sequence of presenting the learning resources according to the new patterns of learning styles discovered while the student is going through the material. The mining part monitors cumulative preferences of a learner during their interaction with the course. This adaptive site takes into account that, even within the same section of a course, learners may change their preferences.

4. Conclusions and further work

The focus of the framework was utilising web usage mining with learning styles for pedagogically effective and technologically possible personalised e-learning courses. Results suggest that dimensions of learning styles, i.e. preferences to learning material, can be modelled using suitable attributes and can be detected using data mining techniques. We are currently investigating using the output of the mining tool into personalised learning scenarios, in which the learners are assisted by the system based on the patterns and the preferred learning styles. We plan to compare our work with others ([3]), who have used other techniques, such as the Bayesian model or genetic algorithms.

References