Data Mining and Student e-Learning Profiles

Submitted by: Pavan       Roll No: 113050010       Date: 17-08-12       Report No: 7

Philosophy: Sequential data mining algorithms can be applied to analyze computer logs to profile learners by their learning style. The data mining algorithms are employed to discover patterns which characterize the learners.

Summary: By observing how learners behave during their online self-study, we can compare, evaluate, and profile individual learners learning style and can then make suggestions to learners with similar characteristics. Here we apply data mining techniques to profile learners in contextualized learning situations.

A sequence is an ordering of events, and each event in the ordered list is called item. A sequence x is a subsequence of another sequence y if x can be formed from y by leaving out some events without disturbing the relative positions of the remaining events. For example, if x = < C, E, D, B > and y = < C, A, E, D, E, G, C, E, B, G >, then x is a subsequence of y. Given a set of sequences S, the support of a sequence x in S is the number of sequences containing x, denoted as support(x). Given a percentage min_sup as the minimum support threshold, a sequence x is a sequential pattern observed in S if support(α) ≥ min_sup*|S|, where |S| is the total number of sequences in S. By obtaining sequential patterns, we can evaluate learners activities and accordingly adapt and customize resource delivery etc.

With data mining techniques, instructors can first receive a detailed record of a learner’s behavior. Further, if they find similar groups of students, they can be classified into one group which shows the main characteristics of the group. Finally, the instructors can discover if there is any relationship between these characteristics and learners other psychological attributes. This paper uses sequential data mining methods to profile online learners by taking into account their cognitive and motivational attributes.

Many types of educational software are capable of writing event records to log files as learners interact with the software’s features. Example: In gStudy software learners can operate on multimedia documents using a variety of tools such as making notes, tagging selected content, accessing hyperlinks, constructing new terms in a glossary, drawing concept maps, searching and chatting etc. As students select and use tools in gStudy, the system collects data on the fly about their choices and constructions. Data are logged at the level of a software event and written to an XML file time-stamped to the millisecond. The XML file is a transcript containing the events that trace students interaction.

Profiling learners using sequential data mining:

- Preprocessing: The raw learning logs involve a complex series of low-level events spaced along a time dimension. Through a log parsing mechanism, a sequence of temporally ordered learner actions is generated.
- Pattern Discovery: These sequences of learner actions are then fed into the sequential mining algorithm to discover patterns across the learning logs. With these observed patterns, suggestions can be given to students.

Observations: Learning profiles became more efficient when we consider the motivation behind each action to avoid misinterpretations and also to achieve a better understanding of what we have observed on-the-fly.

References: Data Mining and Student e-Learning Profiles by Mingming Zhou at 2010 International Conference on E-Business and E-Government