Web Personalization for Enhancing e-Learning Experience

Konstantinos Markellos[1,2], Penelope Markellou[1,2], Maria Rigou[1,2], Spiros Sirmakessis[1,2,3], Athanasios Tsakalidis[1,2]
[1]Research Academic Computer Technology Institute
[2]Department of Computer Engineering and Informatics, University of Patras
[3]Technological Educational Institution of Messolongi, Department of Applied Informatics in Administration and Economics

Abstract

In this paper, we describe how web personalization techniques can be applied in a typical e-learning system. The central notion is the production of recommendations for studying new topics which come both from the automated processing of recorded online student behavior, as well as the manual specification of contextually related topics. The automated processing is based on knowledge discovery techniques and is preceded by web usage mining applied on user click-streams. The produced recommendations guide individual learners through the available material suggesting "interesting" link shortcuts, which wouldn't otherwise be discovered (at least not as many and not as fast).

Introduction

Online curriculums often demonstrate complicated structuring and students confront the problem of filtering out the subset of topics that suit their individual learning requirements. Hyperlinked course material allows users to follow any navigational path they choose and not necessarily use the structure determined by web site designers or content creators (who have a certain navigational pattern in mind). This freedom may prove a hindering factor since in many cases learners do not have the necessary maturity and skill to follow an effective path and it is often the case that they wonder around topics that are either too difficult, too easy, or just irrelevant to individual learning needs.

The answer to this problem comes from the incorporation of personalization techniques which undertake to tune the learning experience to individual (or group) requirements based mainly on tracking user browsing behavior. Such systems feature as a remedy for the problems that stem from the traditional "one-fits-to-all" approach that delivers the same static learning material to everyone, despite of individual domain expertise, information needs and preferences, which may vary dramatically (De Bra et al., 1999). Web mining (and more specifically web usage mining) plays a central role in constructing internal user models (profiles), as well as expanding them by inferring additional data (Cooley et al., 1999). This enables courses to be assembled on-the-fly addressing exactly what the learner needs to know without wasting time on topics in which the user is already proficient or not interested.

In this work we argue that the production of personalized topic recommendations may prove of great importance as refers to the efficiency of online learning. It may dramatically assist learners by guiding them through available material and avoiding disorientation in the big volume of topics usually available in learning platforms that offer numerous tutorials on many subjects. More specifically, this paper investigates the deployment of association rules mining for delivering personalization in the domain of educational applications and presents a recommendations generator that suggests learning topics based on the aforementioned technique. It concludes with discussing limitations and concerns connected with web personalization in general and its application to e-Learning.

Field Background

Learning is in general a complicated process and users need to feel that they have a unique personal relationship with the system they use. An idea to improve online learning is to apply web personalization and provide learners with added value by knowing and serving them as individuals. Personalization can be based on various techniques from the area of knowledge discovery. In our pilot recommendation generator we have incorporated association rules mining and clustering techniques.

Association Rules (ARs) connect one or more events in order to discover associations and correlations between different types of information without obvious semantic dependence (Agrawal et al., 1993), (Wang et al., 2002). The AR A→B for example can mean that when item A is observed in a database record, it is likely that item B will also appear in that same record. This kind of relationship has potential interest for producing recommendations. Although the number of the produced ARs can be significant, most of them are probably trivial. For identifying the most important rules, two interestingness measures are defined for each rule *R* of the form A→B: *support* and *confidence*. Support *S* is the number of database records which contain A∪B (often expressed as a proportion of the total number of records) and confidence *C* is the fraction:

$$\frac{\text{support for R}}{\text{support for A}} \qquad (1)$$

From (1) it is clear that confidence is computed after having the support values for the rule and its antecedent computed first. There are many algorithms for discovering ARs. The *Apriori algorithm* (Agrawal & Srikant, 1994), which is one of the earliest, performs repeated passes over the database records in order to identify interest itemsets (satisfying a predefined minimum support value). Its basic idea is that a set *A* of items is frequent only if all subsets of *A* are frequent.

Clustering (Chakrabarti, 2002), (Hand et al., 2001) is the second technique embodied in the recommendation generator. *Clustering* is a data mining technique defined as *the process of grouping data into classes (or clusters) so that objects within a cluster have high similarity in comparison to one another, but are very dissimilar to objects in other clusters.* Cluster analysis aims to discover items that have representative behavior in the collection. This technique is very suitable for the e-Learning domain, since grouping of similar students are treated differently in terms of recommended topics.

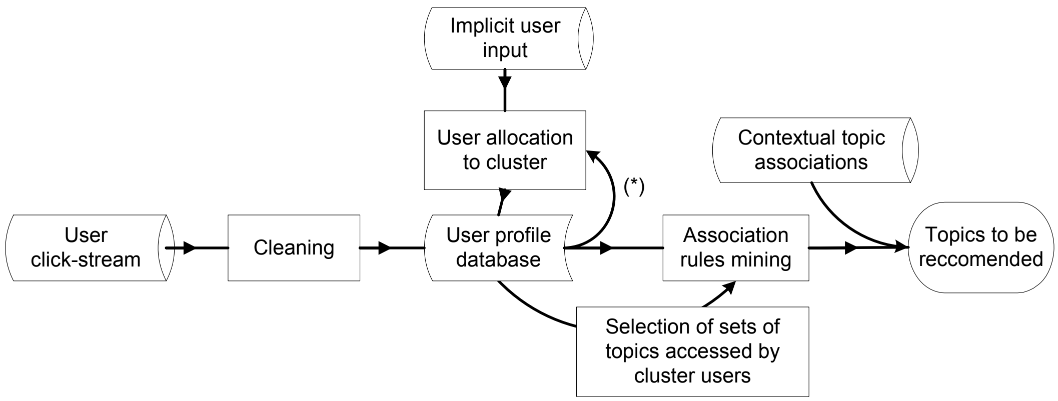
Theory to Action

The recommendations generator integrates various sources of data, both *explicit* (demographics, preferences, and domain background information collected through the mandatory student registration questionnaires), and *implicit* (student actions recorded using server logs and cookies). The latter type of data poses a number of research challenges since apart form recording user click-streams, a phase of information extraction from web logs and mapping with specific user actions (starting a new lesson, doing a test or sending an e-mail to the tutor, etc.) should precede before making any sense of the large volume of collected raw data. The extracted information along with all data supplied explicitly form a user, are stored in the corresponding database record known as the *user profile*. The generator accesses user profiles to trigger the recommendation process.

The generator delivers two types of suggested sets of topics, one that is computed dynamically based on the recorded student click-streams and another one based on contextual topic correlations defined by tutors (referring both to prerequisite topics, as well as suggestive topics to proceed with, according to a pre-decided teaching scenario).

- **Implicit user input:** this information is used to classify users into one of a set of predefined groups (clusters). The factors that determine a specific cluster can be tuned and typically include domain knowledge, as well as learning needs and preferences while in some cases they may also comprise age, sex, nationality, preferred learning style, etc. depending on the variety of adaptations the system can support. In the pilot we distinguish 3 levels of background skill and 5 categories of learning preferences. A user cluster for instance includes all students that have average MS Office skills (with the “average” meaning that the computation of a weighed function of Word, Excel, and Access skills lies within a predefined threshold). As depicted in figure 1 cluster membership may be revised based on collected profile data. More specifically, when the generator records the complete set of topics in a lesson as accessed, then the domain background for the specific user must change to reflect it (see arrow marked with (*)). Student assignment to clusters might also be used for provoking interactions among students and enhancing this way collaboration and communication, as well as for allowing tutors to perceive a useful insight of the "classroom".

Figure 1: Recommendations Generator Operation



- **User click-stream:** the generator monitors all student actions as successive page requests recorded in server logs. Log files are then cleaned form all redundant information (such as secondary, automatically generated requests for page images, etc). Combining the remaining requests with information about the way learning content is structured, the generator distills user accesses to topic pages. The set of topics that have been accessed by a certain user during all past visits to the system are stored in the user profile, and this is where the generator seeks for discovering association rules. An association rule example follows:

$$\{\mathrm{topic_i},\ \mathrm{topic_j}\} \rightarrow \{topic_x\}\text{, support=0.02, confidence=0.68} \qquad (2)$$

The rule in (2) conveys the relationship that users who accessed $\mathrm{topic_i}$ and $\mathrm{topic_j}$ also tend (with a confidence of 68%) to be interested in $\mathrm{topic_x}$. Support represents the fact that the set $\{\mathrm{topic_i}, \mathrm{topic_j}, \mathrm{topic_x}\}$ is observed in 2% the sets of topics accessed.
The discovered association rules may use as input, either the sets of topics accessed by all users, or just the ones accessed by users that belong to the same cluster as the current one. Another option is to use both approaches and suggest the union of discovered topics. This scenario is very useful when association rules mining inside clusters

fails to produce reliable recommendations due to lack of adequate input.

- **Contextual topic correlations:** they are defined by tutors as a way to capture domain related dependencies among topics that are not indicated by their allocation in lessons in the curriculum structure (for example the topic “How the Save As dialog works” in the Excel lesson can be associated with “How to create a new folder” in the Windows lesson). Such correlations are internally represented in the form of rules just like the dynamic associations discovered by web usage mining. There is also the possibility to suggest a complete sequence of links implementing a specific teaching scenario. It is important to stress that during the early phases of generator operation all topics recommended are determined by this type of correlations, since there in no (or not enough) recorded usage information.

Limitations and Concerns

Determining and delivering personalized learning is a data intensive task that requires the execution of numerous processing steps. This usually causes intolerably *long response* times, which in turn may lead to abandonment. To avoid this obstacle, software designers should consider either having part of the process execute offline or deploying special algorithms, structures and configurations to assure fast online operation.

Apart from requiring fast delivery of adaptations, it is equally crucial to assure accuracy, in the sense that recommendations that are not successful slow down the learning process by confusing and disorienting users. It is much better not to deliver any recommendations than deliver a set of useless and harmful ones.

Another concern results from the fact that in learning environments that allow for *dynamic content updating*, software designers have to resolve issues concerning the newly added topics and more importantly the updated ones. For instance, should the system recommend to a student that has already studied the previous version of a topic, its new one? Or, on what kind of input (required form the tutors’ part) should this decision be based upon? Should there be a student assessment process to indicate the required topics that will provide the student with the skills not acquired yet? All the questions raised so far and a large number of others that come to mind relate directly to the chosen teaching strategy. And this in turn is determined and implemented by *adequately skilled teachers* who are also assigned the task of defining content attributes to lessons, prerequisite skills, topic correlations, topic updates and a number of other contextual dependencies.

Another area of concerns relates to human aspects. Adaptive systems are striving to identify the user, record the user’s online behavior in as much detail as possible and extract needs and preferences in a way the user cannot notice, understand or control. This situation brings up the invaded *privacy* hazard that remains an unresolved problem for the web personalization domain. Last but not least, the produced adaptations should be delivered in the appropriate way (avoiding learner intrusion and loss of concentration) and should not deprive users *control* over the learning process.

Conclusion

In this work we have investigated the delivery of personalization for learning environments on the web through recommendations. A recommendations generator based both on association rules mining applied on user click-streams, and topic correlations defined by tutors, combines the aforementioned sources of information with student clusters and delivers topic recommendations that help users navigate faster the curriculum space and identify the information of interest.

References

Agrawal, R., & Srikant, R. (1994). Fast Algorithms for Mining Association Rules. Proceedings of the 20th VLDB Conference, Santiago, Chile, 487-499.

Agrawal, R., Imielinski, & T., Swami, A. (1993). Mining Association Rules between Sets of Items in Large Databases. Proceedings of the ACM SIGMOD International Conference on Management of Data, 207-216.

Chakrabarti, S. (2002). Mining the Web: Discovering Knowledge from Hypertext Data. Morgan Kaufmann.

Cooley, R., Mobasher, B., & Srivastava, J. (1999). Data Preparation for Mining World Wide Web Browsing Patterns. Journal of Knowledge and Information Systems 1, 1.

De Bra, P., Brusilovsky, P., & Houben, G.J. (1999). Adaptive Hypermedia: From Systems to Framework. ACM Computing Surveys, 31, 4.

Hand, D., Mannila, H., & Smyth, P. (2001). Principles of Data Mining. MIT Press.

Wang, D., Bao, Y., Yu, G., & Wang, G. (2002). Using Page Classification and Association Rule Mining for Personalized Recommendation in Distance Learning. Proceedings of the 1st International Conference Advances in Web-Based Learning, Hong Kong, China, 363-374.

[1]Research Academic Computer Technology Institute, 61 Riga Feraiou str., 26221, Patras, Greece
[2]Department of Computer Engineering and Informatics, University of Patras, 26504, Patras, Greece
[3]Technological Educational Institution of Messolongi, Department of Applied Informatics in Administration and Economics, 30200, Messolongi, Greece
Tel: +30-2610-960335, Fax: +30-2610-960374
e-Mails: kmarkel@cti.gr, markel@cti.gr, rigou@cti.gr, syrma@cti.gr, tsak@cti.gr