2013 IEEE 13th International Conference on Advanced Learning Technologies

Toward a fully automatic learner modeling based on web usage mining with respect to educational preferences and learning styles

Mohamed Koutheaïr KHRIBI [1], Mohamed JEMNI [1], Olfa NASRAOUI [2], Sabine GRAF [3], and Kinshuk [3]
[1] Technologies of Information and Communication Lab,
Higher School of Sciences and Technologies of Tunis, University of Tunis, Tunisia
mk.khribi@uvt.rnu.tn, mohamed.jemni@fst.rn.tn
[2] Knowledge Discovery & Web Mining Lab,
Speed School of Engineering, University of Louisville, Louisville KY 40292, USA
olfa.nasraoui@louisville.edu
[3] School of Computing and Information Systems, Athabasca University, Canada
sabineg@athabascau.ca, kinshuk@athabascau.ca

Abstract— **In this paper, we describe a fully automatic learner modeling approach in learning management systems, taking into account the learners' educational preferences including their learning styles. We propose a composite learner model made of three components: the learner's profile, learner's knowledge, and learner's educational preferences. The learner's profile represents the learner's general information such as identification data, the learner's knowledge captures the learner's interests on visited learning objects, and the learner's educational preferences are composed of the learner's preferences (in terms of the specific attributes of the visited learning objects) and his/her learning style. In the proposed approach, all the learner model components are automatically detected, without requiring any explicit feedback. All the basic learners' information is inferred from the learners' online activities and usage data, based on web usage mining techniques and a literature-based approach for the automatic detection of learning styles in learning management systems. Once learner models are built, we apply a hierarchical multi-level model based collaborative filtering approach, in order to gather learners with similar preferences and interests in the same groups.**

Keywords: Learner Modeling; Learning Styles; Recommender Systems; Web Mining; Collaborative Filtering

I. INTRODUCTION

Learner models represent an important knowledge asset that can be built from data collected about learners. These models can then be used by the learning system to produce personalized services depending on the learner needs. There are basically two main approaches for building learner models: collaborative learner modeling and automatic learner modeling [1]. The collaborative learner modeling requires students to provide explicit information about their preferences and needs. Application forms and/or questionnaires are commonly used to collect this explicit data about the learners (preferences, interests, learning styles, etc). On the other hand, in an automatic learner modeling approach, gathering data about learners is done automatically based on implicit information such as their online behavior and activities. Collecting implicit data about learners can spare them from the onerous task of answering questionnaires and filling out forms, which are considered supplementary activities that extend beyond their normal e-learning tasks.

In this paper, we propose a fully automatic learner modeling approach in learning management systems (LMSs), which allows the automatic detection of learners' interests, preferences, and learning styles from implicit data collected about learners. So, we intend to adopt a learner model with three components: learner's profile, learner's knowledge, and learner's educational preferences. We represent the learner model LM_i related to a learner *i*, as follows : $LMi = \{LPi, LKi, LEPi\}$, where LP_i (the learner's profile) represents the learner' general information such as identification and demographic data, LK_i (the learner's knowledge) represents the learner's interests on visited learning objects, and LEP_i (the learner's educational preferences) contains the learner's preferences among visited learning objects, and his/her learning style.

Hence, compared to our work in [2], we add the learner's educational preferences component, LEP_i, including the learner's preferences and learning style, as well as considering the duration and frequency of visits as heuristics in the learner's knowledge component building. On the other hand, compared to our work in [3], we update the learners' educational preferences component by adding their learning style and preferences on visited learning objects. Furthermore, the learner's preferences and learning style are automatically inferred from tracked usage data, collected from the learning management system database, based on web usage mining techniques and a literature based-approach. Once learner models are built, we apply a hierarchical multi-level model based collaborative filtering approach, in order to gather learners with similar preferences and interests into the same groups. It is also important to note that the proposed automatic learner modeling approach falls within the scope of extending a recommendation system that we have previously developed [4]. This recommendation system relied on web usage mining techniques and scalable search engine technology to compute hybrid recommendations against a large repository of educational

978-0-7695-5009-1/13 $26.00 © 2013 IEEE
DOI 10.1109/ICALT.2013.123

IEEE computer society

resources. In [4], a range of recommendation strategies were investigated, based mainly on content based filtering and collaborative filtering approaches, each applied separately and in combination [2]. We are working on the extension of the proposed recommendation system, to be (1) included in learning management systems, and (2) to take into account the learners' educational preferences including their learning styles. The rest of the paper is organized as follows: first, we describe the main principles of the proposed learner modeling approach. Then, we describe the group modeling phase based on a proposed hierarchical multi-level model based collaborative filtering approach. Finally, we present some experiments that we carried out and finish with our conclusions.

II. PROPOSED LEARNER MODELING APPROACH

We assume that the learner model LM_i related to a learner i, is represented as follows: $LMi = \{LPi, LKi, LEPi\}$, where

- LP_i (the learner's profile component) represents the learner's general information, which, at the bare minimum, can be reduced to their identification data that is available based on authentication techniques used in LMSs;
- LK_i (the learner's knowledge component) implies the learner's interests on visited learning objects, and represented as sets of interest measures on each visited learning object;
- LEP_i (the learner's educational preferences component) contains the learner's learning style and his/her preferences, upon visited learning objects, with respect to specific attributes such as type, format and shape.

A. The Learner's knowledge component

The learner's knowledge component represents the learner's interests on visited learning objects, from the first access to the LMS till the last one. We mean by the learner's interest on a visited learning object, his/her preference for a visited educational content on the LMS course, regardless of the type or the format of that content. We consider that if a learner visits a learning object frequently and spends much time on it, then he/she is interested by that learning object. To get such information automatically, without requiring explicit feedback from the learner, a number of necessary Web usage mining tasks must be applied on tracked usage data saved in the LMS database. Let LO be a set of n unique visited learning objects: $LO = \{ LO_1, LO_2, LO_3, \ldots, LO_n\}$, the learner knowledge model LK_i corresponding to the learner i is represented by a set of p sessions S_j^i extracted from tracked data : $LK_i = \{S_1^i, S_2^i, S_3^i, \ldots, S_p^i\}$, where each S_j^i is a subset of k weighted visited learning objects LO_l, $S_j^i : < w(LO_1^{S_j^i}), w(LO_2^{S_j^i}), w(LO_3^{S_j^i}), \ldots, w(LO_k^{S_j^i})>$, where each $LO_k^{S_j^i} = LO_l$ for some $l \in \{1, \ldots, n\}$, and $w(LO_k^{S_j^i})$ is the weight associated with the visited learning object $LO_k^{S_j^i}$ in the session S_j^i corresponding to the i^{th} learner. It should be noted that The learning object weight $w(LO_k^{S_j^i})$, also called, learning object interest measure $LOIM$, can be reduced to a binary value [5] (storing the existence or non-existence of a learning object in a session), or it can be computed as a function of a number of features (based on the frequency of occurrence of the learning object in a session and/or the time a learner spends when visiting a learning object). To compute the weights of visited learning objects within the learner's sessions, we apply two heuristics: duration and frequency of visit. Hence, the learning object weight $w(LO_k^{S_j^i})$, is obtained as follows:

$$w(LO_k^{S_j^i}) = \left(\frac{F(LO_k^{S_j^i})}{\sum_k F(LO_k^{S_j^i})} + \frac{T(LO_k^{S_j^i})}{\sum_k T(LO_k^{S_j^i})} \right) / 2 \qquad (1)$$

where $F(LO_k^{S_j^i})$ represents the frequency of visit of the learning object $LO_k^{S_j^i}$ within the session S_j^i, $\sum_k F(LO_k^{S_j^i})$ represents the sum of all visit frequency values related to the entire set of learning objects $LO_k^{S_j^i}$ within the session S_j^i, $T(LO_k^{S_j^i})$ represents the duration of a single visit to the learning object $LO_k^{S_j^i}$ in the session S_j^i, and $\sum_k T(LO_k^{S_j^i})$ represents the sum of all duration visits to the entire set of learning objects $LO_k^{S_j^i}$ in the session S_j^i. Therefore, the learner's knowledge component LK_i can be represented as a matrix $M(p, n)$, where p is the total number of learner's sessions and n the cardinality of unique visited learning objects. The element m_{xy} of this matrix is:

$m_{xy} = w(LO_y^{S_x^i}), x \in \{1, \ldots, p\}, y \in \{1, \ldots, n\}$

$$M(p, n) : \begin{pmatrix} w(LO_1^{S_1^i}) & \cdots & w(LO_n^{S_1^i}) \\ \vdots & \ddots & \vdots \\ w(LO_1^{S_p^i}) & \cdots & w(LO_n^{S_p^i}) \end{pmatrix}$$

B. The learner's educational preferences component

The learner's educational preferences component LEP_i, is composed of the learner's preferences among visited learning objects and his/her learning style. Below, we explain how the learner's preferences and learning styles are detected based on his/her behavior.

1) Detection of the learner's preferences

Contrary to what precedes, we are rather concerned by the learner's interests on specific attributes of visited leaning objects such as their formats. Indeed, given that educational web objects available in a LMS are characterized by many attributes (e.g. format), each of which may have several values (e.g. for format: text, image, video, etc.) that could be preferred or not by the learner, the preferences of a learner *i* upon these values can be represented, as a vector of interest measures V_{EP}^{i}: $\{LOIM\_Att_k\_LO_t^i\}_{k,t}$. The cardinality of this vector is $m = \sum_k Card(att_k)$, where $Card(att_k)$ represents the number of values related to the attribute att_k, and $LOIM\_Att_k\_LO_t^i$ represents a component of the vector V_{EP}^{i}, corresponding to the interest measure of a learner *i* on a specific value *t* of the attribute Att_k related to the learning object *LO*. The vector of interest measures V_{EP}^{i} is obtained as follows: first, we compute, for each learner's session S_j^i, the interest measures corresponding to each value of the considered attributes, based on the frequency and the duration of visits. Then, we compute the mean values of the obtained previous measures over all learners' sessions $S^i = \{S_1^i, S_2^i, \ldots, S_j^i, \ldots, S_q^i\}$. These measures are computed among all learner usage sessions since his/her first access till the last one, and updated whenever new usage sessions are registered. Therefore, usage data which has been preprocessed and prepared in the earlier phase, is used again, in order to compute the learner's interest measures on specific values of learning object attributes. In this work, we consider three attributes of educational web objects available in a LMS (Moodle as example): «learning object type», and «learning object shape», and «learning object format». The following values are associated to these attributes:

- *Type_LO (Learning object type) = {Resource, Activity}*
- *Shape_LO (Learning object shape, if Type_LO = Activity) = {Exercise, Simulation, Questionnaire, Assessment, Forum, Chat, Wiki, Assignment}*
- *Format_LO (Learning object format, if Format_LO = Resource) = {Text, HTML, Image, Sound, Video}*

Based on these considered attributes, let $LOIM\_Type\_LO_t^i$, $LOIM\_Format\_LO_t^i$, and $LOIM\_Shape\_LO_t^i$ be the interest measures composing the vector V_{EP}^{i}, which are computed over all learner's sessions S^i as follows:

$$LOIM\_Type\_LO_t^i = \frac{\sum_j LOIM\_Type\_LO_t^{S_j^i}}{Card(S^i)}, \; t \in \{1,2\} \quad (2)$$

$$LOIM\_Shape\_LO_t^i = \frac{\sum_j LOIM\_Shape\_LO_t^{S_j^i}}{Card(S^i)}, \; t \in \{1,8\} \quad (3)$$

$$LOIM\_Format\_LO_t^i = \frac{\sum_j LOIM\_Shape\_LO_t^{S_j^i}}{Card(S^i)}, \; t \in \{1,5\} \quad (4)$$

where $Card(S^i)$ represents the number of sessions S_j^i in S^i belonging to the learner *i*, $LOIM\_Type\_LO_t^{S_j^i}$, $LOIM\_Format\_LO_t^{S_j^i}$ and $LOIM\_Shape\_LO_t^{S_j^i}$ represent the interest measures computed over each learner's session S_j^i :

$$LOIM\_Type\_LO_t^{S_j^i} = \frac{\sum_k w(LO_{k,t}^{S_j^i})}{Nbre_{LO_t^{S_j^i}}}, \; t \in \{1,2\} \quad (5)$$

$$LOIM\_Shape\_LO_t^{S_j^i} = \frac{\sum_k w(LO_{k,t}^{S_j^i})}{Nbre_{LO_t^{S_j^i}}}, \; t \in \{1,8\} \quad (6)$$

$$LOIM\_Format\_LO_t^{S_j^i} = \frac{\sum_k w(LO_{k,t}^{S_j^i})}{Nbre_{LO_t^{S_j^i}}}, \; t \in \{1,5\} \quad (7)$$

where $w(LO_{k,t}^{S_j^i})$ is the interest measure (computed as above in *(1)*, based on two heuristics: duration and frequency of visit) related to the visited learning object $LO_{k,t}^{S_j^i}$, $k \in \{1 .. l\}$, belonging to a specific attribute's value *t*, *l* is the number of learning objects available in the LMS course, $Nbre_{LO_t^{S_j^i}}$ represents the number of visited learning objects belonging to a specific attribute's value *t*, within the session S_j^i. Thus, the vector V_{EP}^{i} corresponding to the i^{th} learner, based on computed interest measure values, is representing the learner's educational preferences among the considered values of learning object attributes. A Large interest measure on a specific value (e.g. Text) of an attribute (e.g. Format) implies that the learner has a strong preference for this attribute (e.g. strong preference for textual content).

2) *Detection of the learning style*

Incorporating learning styles within learning management systems, in order to perform personalization, requires selection of a learning style model for measuring the learners' preferences with respect to the dimensions of this model, and then, enabling the learning management system to recognize the learners' learning styles. The latter can be accomplished in two ways: explicitly or implicitly. The majority of existing personalized systems generally rely on an explicit approach to identify learning styles, using questionnaires. Other works have dealt with the automated detection of learning styles. In [6], Bayesian networks were used to detect and model student learning styles. The approach has been evaluated by comparing the proposed Bayesian model with results obtained using the ILS questionnaire. Cha et al., in [7], investigated the use of Decision Trees and Hidden Markov Models, while Graf et al., in [8], studied the behavior of learners during their online work to gather hints about their leaning styles. In fact, based on a rule based mechanism, learning styles have been identified using indications from learners' behaviors. The evaluation of this approach has demonstrated good results and suitability for identifying learning styles with respect to the Felder Silvermann Learning Style Model (FSLSM). FSLSM describes learning style preferences by characterizing each learner according to four dimensions: active/reflective, sensing/intuitive, visual/verbal, and sequential/global. In [9], correlation has been validated between a learner's online behavior using a LMS course and his/her learning style identified via an ILS questionnaire. Patterns of behavior for each learning style dimension are inferred using common predefined features in an LMS. These features include: content objects, outlines, examples, self-assessment, tests, exercises, and discussion forums. The navigation behavior of learners within a course are also considered [8]. To detect learning styles in our learner model, we use the approach proposed by Graf et al., [8]. Accordingly, the automated detection of learning styles is accomplished in two phases: (1) determining relevant patterns of learners' behaviors, and (2) inferring learning styles from behaviors. Thus, each learning style dimension is related to specific patterns which are represented using predefined LMS features with high or low occurrence, indicating, therefore, a specific learning style preference [7]. In our work, in order to automatically detect the learners' learning styles without requiring their explicit feedback, we use relevant patterns of learners' behaviors based on common LMS features and learners' navigational histories, and apply a literature-based approach [10]. Benefits of applying this approach are twofold: (1) learning styles are detected automatically based, on the learners' implicit feedback, which is entirely consistent with the main principle of our proposed automated learner modeling approach; and, (2) the proposed approach is suitable for LMSs (widely, commonly and successfully used in education) rather than for one specific e-learning system (generally confined to research labs).

III. Group Modeling Approach

As seen above, the learner's model LM_i related to a learner *i,* is represented as follows:

$$LM_i = \{LP_i, LK_i, LEP_i\} = \{LP_i, LK_i : \begin{pmatrix} w(LO_1^{S_1^i}) & \cdots & w(LO_n^{S_1^i}) \\ \vdots & \ddots & \vdots \\ w(LO_1^{S_p^i}) & \cdots & w\left(LO_n^{S_p^i}\right) \end{pmatrix}, LEP_i : (LS_i, V_{EP}^i)\},$$

Once learner models are built, we apply a hierarchical multi-level model based collaborative filtering approach on these models (Figure 1), in order to assign learners with common preferences and interests to the same groups, so that feedback from one learner can serve as a guideline for information delivery to the other learners within the same group. At the first level, a rule based classification is applied based on the learner's learning style. Each obtained class is corresponding to learners with the same FSLSM learning style dimension. Then, at the second level, we apply clustering techniques on the previously obtained classes, based on the learner's educational preferences represented by V_{EP}^i . Thus, the obtained groups are representing learners with similar educational preferences. At the third level, clustering techniques are applied, again, on the previously obtained groups, but this time, based on similarities and dissimilarities among the learners' knowledge components (i.e clickstreams data). Finally, association rules are extracted from latter groups. Therefore, a hierarchy of learners' groups is obtained and represented, at the leaf level, by a centroid vector, representing a group containing learners with similar learning styles, educational preferences and clickstreams data, and containing a number of extracted association rules. These mined rules are later used within a collaborative filtering recommendation strategy.

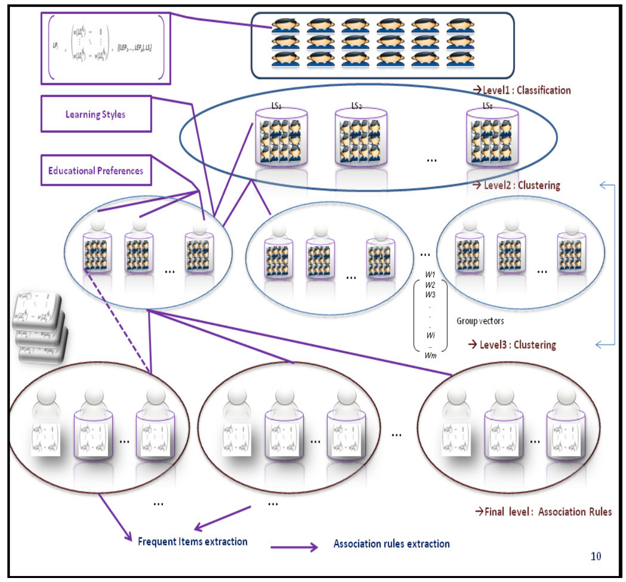

Figure 1. Group Modeling approach

Let G_t be a group, obtained after the first two levels, and composed by learners' knowledge components of a set of *h* learners, $G_t = \bigcup_i LK_i = \{S_1^i, S_2^i, S_3^i, \ldots, S_p^i, S_1^{i+1}, S_2^i, S_3^{i+1}, \ldots, S_l^{i+1}, \ldots, S_1^{i+h}, S_2^{i+h}, S_3^{i+h}, \ldots, S_s^{i+h}\}$, the cardinality of this group is obtained by summing cardinalities of LK_i : $\mathrm{Card}(G_t) = \sum_i Card(LK_i)$, S_j^i represents the j^{th} learner's session composed by *k* interest measures of visited learning objects within this session. The group G_t can be represented by a matrix $M(\mathrm{Card}(G_t), n)$, where *n* is the number of unique visited learning objects, and $\mathrm{Card}(G_t)$ represents the cardinality of the group (*i.e* the sum of learners' sessions cardinalities belonging to the same group) : $m_{xy} = w(LO_y^{S_x^i})$, $x \in \{1, \ldots, \mathrm{Card}(G_t)\}$, $y \in \{1, \ldots, n,\}$, $i \in \{1, \ldots, h\}$,

$$M(\mathrm{Card}(G_t), n) : \begin{pmatrix} w(LO_1^{S_1^1}) & \cdots & w(LO_n^{S_1^1}) \\ \vdots & \ddots & \vdots \\ w(LO_1^{S_{\mathrm{Card}(Gt)}^h}) & \cdots & w\left(LO_n^{S_{\mathrm{Card}(Gt)}^h}\right) \end{pmatrix}$$

The last level of clustering results in a set of clusters *C={C1, C2, …, Ck}*, where each cluster C_j is a subset of sessions representing a group of similar learners with similar access patterns. Each obtained cluster is represented by a group vector, built based on session vectors contained in that cluster. The group vector V_k related to a group C_k, is obtained as follows: $V_k = \frac{1}{|C_k|}\sum S^i$, where $S^i \in C_k$ represents the i^{th} session's vector belonging to the C_k group. Finally, each obtained cluster is split into co-related learning object sets using a frequent itemset mining algorithm. Then, association rules *AR* are extracted using *AR* discovery methods such as the Apriori algorithm [10], and used later in the online recommendation phase (e.g. 67.35% of learners who visited *link1* visited *link2* and *link3*. So, *link2* and *link3* can be recommended to a learner who visited *link1*).

IV. EXPERIMENTAION

To implement the proposed learner modeling approach, we chose Moodle as the learning management system, since it is one of the most used platforms in universities. Moodle has a modular architecture and is open source, so it can be easily extended to add new components. We used an online hybrid course *(C2i)* with 704 enrolled learners. Tracked data was extracted from various Moodle tables (primarily from mdl_log table). More than 54364 requests were initially selected, and 37903 requests remained after preprocessing tasks. The sessionizing task resulted in 3172 different sessions, and after removing non trivial sessions, only 1516 sessions remained. Then we applied the proposed hierarchical multi-level model based collaborative filtering approach, using *k-way* Clustering via Repeated Bisections as the clustering algorithm [11] with the cosine similarity and a number of clusters *k=5*. We obtained 5 groups of similar sessions containing respectively 375, 215, 133, 494, and 299 sessions. Finally, we applied the Apriori algorithm on each of these obtained clusters, using a minimum confidence value of 0.4 and a minimum support value 0.04 (after trial and error). The proposed approach of learner and group modeling is currently being experimented as part of a recommendation approach, where recommendations are computed with respect to the active (i.e current and online) learner's short term usage history, his/her learning style and educational preferences, as well as exploiting similarities and dissimilarities among the learner models and educational content.

V. CONCLUSION

In this paper, we have outlined the general principles of an entirely automated approach for modeling learners in learning management systems, taking into account their educational preferences and learning styles. We considered a learner model with three components: the learner's profile that represents the learner' general information such as identification data, the learner's knowledge that represents the learner's interests on visited learning objects, and the learner's educational preferences that contains the learner's preferences among visited learning objects, and his/her learning style. These learner model components are inferred automatically from usage data, based on web usage mining techniques and a literature-based approach for the automatic detection of learning styles in LMS. Then, a hierarchical multi-level model based collaborative filtering approach is applied for modeling learners into groups. Currently, we are applying the proposed approaches within the Moodle LMS, in the context of a recommendation system aiming to automatically recommend links to learners taking into account their educational preferences, including their learning styles.

REFERENCES

[1] P. Brusilovsky, "Methods and techniques of adaptive hypermedia," *User Modeling and User Adapted Interaction*, vol. 6 (2–3), pp. 87-129, 1996.

[2] M.K. Khribi, M. Jemni, and O. Nasraoui, "Automatic Recommendations for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval," *Educational Technology & Society*, vol. 12(4), pp. 30-42, 2009.

[3] M.K. Khribi, M. Jemni, and O. Nasraoui, "Toward Integrating the Pedagogical Dimension in Automatic Learner Modeling within E-learning Systems," in *The 9th IEEE International Conference on Advanced Learning Technologies*, 2009, pp. 642-644.

[4] M.K. Khribi, M. Jemni, and O. Nasraoui, "Toward a Hybrid Recommender System for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval", *In Proceedings of World Conference on E-Learning in Corporate*, Government, Healthcare, and Higher Education 2007, Volume 7, Issue 1,pp. 6136-6145.

[5] C. Shahabi, A. . Zarkesh, J. Adibi, and V. Shah, "Knowledge Discovery from User's Web-page Navigation," in *7th IEEE Intl. Conf. On Research Issues in Data Engineering*, 1997, pp. 20-29.

[6] P. García, A. Amandi, S. Schiaffino, and M. Campo, "Using Bayesian Networks to Detect Students' Learning Styles in a Web-Based Education System," in *the Argentine Symposium on Artificial Intelligence*, 2005, pp. 115 - 126.

[7] H. J. Cha, Y. S. Kim, S. H. Park, T. B. Yoon, Y. M. Jung, and J.-H. Lee, "Learning Style Diagnosis Based on User Interface Behavior for the Customization of Learning Interfaces in an Intelligent Tutoring System," in *the 8th International Conference on Intelligent Tutoring Systems, Lecture Notes in Computer Science,*2006, pp. 513-524.

[8] S. Graf, Kinshuk, and T.-C. Liu, "Identifying Learning Styles in Learning Management Systems by Using Indications from Students' Behaviour," in *the IEEE International Conference on Advanced Learning Technologies (ICALT 2008)*, 2008, pp. 482-486.

[9] S. Graf, "Adaptivity in Learning Management Systems focussing on Learning Styles," Vienna University of Technology, 2007.

[10] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," in *the 20th Int'l Conference on Very Large Databases*, 1994, pp. 487-499.

[11] Y. Zhao and G. Karypis, "Soft clustering criterion functions for partitional document clustering: a summary of results," in *CIKM'04*, 2004, pp. 246-247.