Toward a Hybrid Recommender System for E-learning Personalization Based on Data Mining Techniques

Outmane Bourkoukou*, Essaid El Bachari*

* Departement of computer science, Cadi Ayyad University, Marrakesh, Morocco
E-mail: outmane.bo@gmail.com, elbachari@uca.ma

Abstract—Personalized courseware authoring based on recommender system, which is the process of automatic learning objects selecting and sequencing, is recognized as one of the most interesting research field in intelligent web-based education. Since the learner’s profile of each learner is different from one to another, we must fit learning to the different needs of learners. In fact from the knowledge of the learner’s profile, it is easier to recommend a suitable set of learning objects to enhance the learning process. In this paper we describe a new adaptive learning system-LearnFitII, which can automatically adapt to the dynamic preferences of learners. This system recognizes different patterns of learning style and learners’ habits through testing the psychological model of learners and mining their server logs. Firstly, the device proposed a personalized learning scenario to deal with the cold start problem by using the Felder and Silverman’s model. Next, it analyzes the habits and the preferences of the learners through mining the information about learners’ actions and interactions. Finally, the learning scenario is revisited and updated using hybrid recommender system based on K-Nearest Neighbors and association rule mining algorithms. The results of the system tested in real environments show that considering the learner’s preferences increases learning quality and satisfies the learner.

Keywords—E-learning, Recommender system, Learning style, Collaborative filtering, Learning objects.

1. INTRODUCTION

Nowadays, development of searching technology provides learners a new way to break free with the more traditional educational models by exploring ways in which Web-based could adapt their behaviour to the goals, tasks, interests, and other characteristics of users [1]. In response to individual needs, personalization in education facilitates students to learn better by using different strategies to create various learning experiences [2, 3, 4]. In recent years, one of the new form of learning personalization that has been expressed as a need by several studies is to give recommendations for learners in order to support and to help them through the learning process.

Indeed, recommender systems are becoming increasingly important in various interesting application domains such as e-commerce, e-entertainment, e-health and other domains. The aim of the first Recommender Systems (RSs) is to provide useful suggestions for users (books, movies, products, etc.) among their preferences and the other similar users. In summary, recommendation strategies can be divided into three major classes: the content-based recommendation, the collaborative-based recommendation, and the hybrid-based recommendation [5,6,7]:

Content-based recommendation selects items based on the correlation between the content of the items (products, services or contents) and user profile most time by using Physiological models.

Collaborative –based recommendation also known as “people-to-people correlation.” recommends to the active user the items that other users with similar tastes liked in the past. The similarity in taste of two users is calculated based on the similarity in the rating history of users. Collaborative filtering is considered to be the most popular and widely implemented techniques in RS.

Hybrid –based recommendation combines these two techniques to improve the “quality” of recommendations and to eliminate drawbacks of each one.

On the other hand, the enormous increase of Learning Objects (LOs) and Learning Objects Repositories (LORs) on Internet in this last decade, are inspiring some researchers to design and to develop the specific recommender systems in distance learning [2, 3, 8, 9, 10].

In fact, two major challenges have emerged: 1) the existing learning environments suffer from the inability to satisfy the heterogeneous needs of learners, 2) the rapid explosive growth of repositories with digital learning resources make hard to retrieve the most appropriate ones for learning.
Until today, a lot of research works have been done about learning recommendation systems but it is still in preliminary discussions. Indeed, existing systems use either content-based recommendation often by using learning style or using collaborative-based recommendation to deliver a personalized courseware [8, 11, 12].

The objective of this work is to present a personalized framework—LearnFitII using a new hybrid-based recommendation approach to provide a relevant recommendation and also to deal when there is a lack of information about a new learner known as the Cold-Start Problem. This system is designed to:

- Achieve a personalized learning scenario according to learner’s learning style using the Felder Silverman Model [13].
- Update learner’s profile and recognize different preferences and interests of learner by mining server logs using Data mining techniques.
- Revisit the learning scenario according to the dynamic profile using our recommender system based on collaborative filtering and association rule mining. This learning provided by this system is centered on the dynamic selecting and sequencing of learning objects into a coherent, focused organization for instruction in e-learning environment.

This paper is organized as follows: section 2 gives related works cited in literature. Section 3 presents an integrated framework for automating the e-learning personalization. Section 4 presents the methodology used for recommender system. Results and evaluation of our research are presented in section 5. Finally, Section 6 concludes the paper with a summary of the work, its limitations and potential future research directions.

II. RELATED WORKS

In last decade, a number of Learning Recommender Systems (LRSs) based on collaborative filtering have been introduced in order to support learners to achieve specific learning needs. Nevertheless, considering the various existing digital learning objects, such frameworks, could potentially play an important educational role [10, 14, 15].

One of the first attempts to develop a collaborative filtering system for digital learning objects has been the Altered Vista system [12]. This system supports discovery and automatic filtering for relevant learning resources that addresses needs of learners and educators. Another system that has been proposed for the recommendation of learning objects is the RACOFI system (Rule Applying Collaborative Filtering) [16]. The RACOFI system assists and recommends online users audio learning objects. Imran et al. [17] proposed PLORS system supports learners by providing them recommendations about which learning objects within the course are more useful for them. The recommendation mechanism uses association rule mining to find the association between LOs. The CYCLADES system has proposed by Avancini and Straccia [18] for allowing users and communities search, share and organize their information space according to their own view and evaluate learning resources available in Open Archives Initiative (OAI). The system is able to give recommendations of several types based on user and community profiles.

Dascalua et al. [19] use recommender agents for recommending online learning activities or shortcuts in a course web site based on a learner’s web logs using association rule algorithm.

However, in last few years, many researchers suggest that recommender system should combine more than technique in order to provide a better selecting, and sequencing recommendation list of learning objects to fit the specific learner’s needs and interests [7, 20]. As examples, an evolving learning management system has been developed by Tang and McCalla [21] to store, and to share digital learning resources using a hybrid recommendation process based on a clustering and collaborative filtering approach to classify students with similar interests and tastes. In his work [8] Klasnja-Milicevic et al. have developed a system called PROTUS (PRogramming TUtoring System) which can recommend relevant links and activities for learners, by considering the Felder-Silverman Learning styles Model and the learner’s level of knowledge. This system has been designed based on hybrid recommendation using the collaborative filtering and the sequential pattern mining. Li et al. [22] present a general architecture of learning recommender system for the smart learning environment. By constructing learner models and resource models, the proposed recommender system aims to recommend learning resources by using the clustering and association rule mining and to recommend peers via social interaction computing.

Bourkoukou et al. [10], propose a recommender model for e-learning environment to achieve personalized learning experiences by selecting and sequencing the most appropriate learning objects. By using a hybrid recommender system based on collaborative filtering technique and association rule mining algorithm.

Several researches on personalized learning are focused also on the learner profile based on learning style for recommendation [8, 23, 24]. Therefore, learning style is defined as a set of factors, behaviors, and attitudes that facilitate individual learning [25]. Many works on learning styles gave multiple methods and instruments to categorize students according to their difference’s, Kolb’s model [26], Felder’s model [13], and Myers-Briggs’s model [27]. Felder-Silverman model proposes the ILSQ (Index Learning Style Questionnaire) freely available, with 44 questions for assessing preferences across four dimensions [28]. These are Information Processing, Information Perception, Information Reception, and Information Understanding, each dimension are two categories shown in Table 1.

<table>
<thead>
<tr>
<th>Characteristics of Learner Based on FSLSM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Learning Style</strong></td>
</tr>
<tr>
<td>------------</td>
</tr>
<tr>
<td>D1 Active(A)</td>
</tr>
<tr>
<td>Reflective(R)</td>
</tr>
<tr>
<td>Sensing(S)</td>
</tr>
</tbody>
</table>

Table 1

"Characteristics of Learner Based on FSLSM"
Learning Style | Characteristics
---|---
Intuitive(I) | - Interested in overviews and a broad knowledge
- Interested in innovations and accept complications
- Preferring principles and theories

D3 Visual(L) | - Prefers to perceive materials as images, diagrams and films

Verbal(B) | - Prefers to perceive materials as text

D4 Global(G) | - Prefers to get the big picture first
- Assimilates and understands information in a linear and incremental step.

Sequential(Q) | - Preferring to process information sequentially

The combination of these preferences result a total of sixteen personality types and are typically be noted by four letters to represent a person’s tendencies on the four scales [24].

Although all these works, it seems to be difficult to find a final answer how to build a list of appropriate learning objects fitting for a given learning style. In this regard, several studies have been done to show if the personalized learning according to the learner profile can improve and enhance the performance of learning process. Results seem to be mixed [8, 23, 24], but most authors agree that taking into account the learning style in e-learning system could help learners to achieve an effective and efficient learning [11, 24, 29, 30].

III. RECOMMENDER SYSTEM ARCHITECTURE FOR E-LEARNING ENVIRONMENT

In order to provide the relevant instruction to learners, the proposed framework is composed of three components which are the Learner Model, the Domain Model and finally the Recommender Model as shown in Fig.1.

Domain Model: Consist of concepts and the relations that exist between them. Typically the domain model gives a domain expert’s view of domain. Learner Model: Consists of relevant information about the user that is pertinent to the personalization of the learning style. Learner model and domain mode are described in detail in [24]. In the next this paper we describe our recommender model.

A. Recommender Model

This model represents the way used by instructor to present concepts of some domain of knowledge. In fact, a teacher can use multiple learning scenarios for a given concept. Therefore, this model constitutes the core of our framework and the decision body allowing selecting the most adequate teaching scenarios matching with learner’s preferences. A teaching scenario can be defined as the way a teacher select and sequence learning objects to achieve a learning experience. The recommender model is divided into different and independent parts, which are described in detail on the following sub-sections: Dump recommender system module and Intelligent recommender system module.

1) Dump recommender system module: The process of this adaptation is depicted in Fig 2.

At first, the system tries to find out the learner profile, if this profile has been recognized, the learning process can start by selecting the most appropriate learning strategies fitting with the learner’s preferences. Otherwise, the system invites the learner to fill the ILSQ questionnaire. Once he/she completes this task, the framework builds the learner’s profile and stored it in the database, and then the learning process can be started.

Properties learner's preferences, pertaining to education and learning, were collected from the literature [30, 31] as shown in Table 2:

<table>
<thead>
<tr>
<th>Learners’ group</th>
<th>Teaching strategy suggested</th>
<th>Learning object Media</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>Simulation, Solve Problem, Discussion group, Brainstorming, Experiment, Questions and Answers</td>
<td>Forum, Wiki, Weblog, Chat, E-mail, MCQ</td>
</tr>
<tr>
<td>Reflective</td>
<td>Presentation, Case study</td>
<td>E-book, Written text</td>
</tr>
<tr>
<td>Sensing</td>
<td>Presentation, Read Text, Solve Problem, Simulation games, Questions and Answers</td>
<td>Forum, Weblog, Wiki, Animation, Graphic, Picture</td>
</tr>
<tr>
<td>Intuitive</td>
<td>Discussion group, Simulation, Roles games, Case study, Read</td>
<td>Internet research engine, Quiz, MCQ</td>
</tr>
</tbody>
</table>

Fig. 1 System architecture process.

Fig. 2 Process of selecting the adaptive teaching strategy

TABLE II
RECOMMENDATION STRATEGY AND LEARNING STYLE IDENTIFICATION
2) Intelligent recommender system module: This module helps to determine whether the list of suggested learning objects for a specific learner profile is appropriate or not by using collaborative filtering techniques for recommendation. A learning object is a basic component or unit of a course in a tutoring system. The next section introduces the overall recommender system and describes the proposed recommendation process step by step.

B. A HYBRID RECOMMENDATION APPROACH IN E-LEARNING

Recommendation learning objects is fulfilled by following three steps, which is depicted in Figure 3.

![Fig. 3 Recommendation process.](image)

The data mining techniques use the collected information about learner interactions, such as navigation history and bookmarks, to build the learner profile and thereafter to build recommendations. In the following of this section, we present this approach step by step.

A. Cleaning and preprocessing

In recommender system the quality of predictions must be based on quality data. The cleaning and pre-processing is the first and the important step in data mining process, it has an high impact for detecting data anomalies, missing data (lacking attribute values or certain attributes of interest) and or missing data (containing errors, or outlier values which deviate from the expected data), rectifying them early, and reducing the data [32, 33].

B. Data normalization

After this step, the cleaned and preprocessed data should be transformed or consolidated into appropriate forms to become prepared for mining. In our case, learner’s preferences are collected from log files preprocessed and cleaned using the first step. For this purpose we defined the weight of rating for each learning activity by using the following score function $P(\theta)$:

$$P(\theta) = \frac{1}{3}(E(\theta) + I(\theta) + S(\theta))$$  

(1)

Where E is the explicit score given by the learner for each learning object $\theta$, I is the implicit score and S is the social dimension score. The implicit score is given by:

$$I(\theta) = A(\theta) + B(\theta) + C(\theta)$$  

(2)

Where A equals 1, when $\theta$ is stored in the bookmarks, 0 otherwise. The function $B(\theta) = 1 - e^{-t}$ where $t$ is the duration spending by learner during the learning object $\theta$, $C$ is the access frequency of the learning activity.

Finally the function S which is the social rating dimension is defined by:

$$S(\theta) = c \cdot e^{t}$$  

(3)

Where $t$ is the duration spending during all synchronous or asynchronous communications by using associated tools, $c$ number of contributions and interactions with these tools.

After weighting learning resources, we obtained a preference model for each learner defined as a Learner Learning Object Rating (LLOR) matrix with $n$ rows, where $n$ denotes the number of learners $L = \{l_1, l_2, ...l_n\}$, and $m$ columns, where $m$ denotes the number of learning objects $J = \{j_1, j_2, ..., j_m\}$.

The following Table 3 shows an example of Learner Learning Object Rating (LLOR) matrix.

<table>
<thead>
<tr>
<th>Learners</th>
<th>$j_1$</th>
<th>$j_2$</th>
<th>$j_3$</th>
<th>$j_4$</th>
<th>$j_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l_1$</td>
<td>0</td>
<td>5</td>
<td>3</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>$l_2$</td>
<td>6</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>$l_3$</td>
<td>8</td>
<td>3</td>
<td>8</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>$l_4$</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

This matrix use a 0-to-10 rating scale where: 10 means that the learner is strongly satisfied with the selected learning object, 5 indicates that the learner is not moderately satisfied, 1 indicates that the learner is not at all satisfied with the learner object, and finally the score 0 indicates that the learning object is not yet explicitly rated or used at all.

C. Collaborative filtering

Collaborative filtering techniques are based on the simple idea that users who share similar past choices will be interested in similar items in the future. In this paper, we use CF to predict the utility of learning objects for a particular
learner based on the learning objects previously rated by other learners.

After weighting learning objects using the first step, we apply a method based on CF in order to build virtual community of learners sharing the same interests and preferences. In fact, we have to make predictions for all learning object weighted 0 which indicates in the \textit{LLOR} matrix the unknown value. For example in Table 3, \( l_4 \) is an active learner for whom we want to make predictions on learning objects \( j_2 \) and \( j_3 \).

This step is carried out by adapting the most known classifier algorithm K-NN on E-learning scenario [7, 10]. This technique allows finding predictions by using the following steps:

3) \textit{Computing similarities between learners}: The critical step in memory based CF methods is to defined similarity and dissimilarity between users or items. The measurement for the weight for similarity between two learners \( u, v \) is the Pearson correlation coefficient calculated as follows:

\[
SC(u, v) = \frac{\sum_{j=1}^{m} (r_{u,j} - \bar{r}_u)(r_{v,j} - \bar{r}_v)}{\sqrt{\sum_{j=1}^{m} (r_{u,j} - \bar{r}_u)^2 (r_{v,j} - \bar{r}_v)^2}}
\]

In the above equation: \( \bar{r}_u \) and \( \bar{r}_v \) are the average rating of learner \( u \) and \( v \), respectively; \( r_{u,j} \) and \( r_{v,j} \) are the rating of learner \( u \) and \( v \) for learning object \( j \).

4) \textit{Selecting K learners neighbors} : After the similarity between two learners is calculated, an \( N \times N \) similarity matrix is generated, where \( N \) is the number of learners. Then, to predict the unrated learning object \( j \) in the rating matrix by the active learner \( u \), the \( K \) most similar learners will be selected and used as input to compute prediction for \( u \) on \( j \).

5) \textit{Computing predictions}: To make a prediction and generate recommendations for an active learner \( u \) on certain learning objects \( j \), we can take a weighted average of all the ratings on those learning objects according to the following formula:

\[
P_{u,j} = \bar{r}_u + \frac{\sum_{v=1}^{k} SC(u, v)(r_{v,j} - \bar{r}_v)}{\sum_{v=1}^{k} |SC(u, v)|}
\]

In equation (5), \( r_{v,j} \) denotes the rating value given by the user \( v \) for the selected learning object \( j \).

D. Association rules mining

Association Rule Mining (ARM) techniques are one of themost popular ways of representing discovered knowledge and describe a close correlation between frequent items in a database. Association rule mining has been applied to E-learning systems aims to intelligently recommend learning object to learners based on the actions of previous learners to improve course content navigation as well as to assist the learning process [7].

1) \textit{Data transformation}

After, generating a Learner Learning Object Improved (LLOI) matrix to predict unknown ratings using collaborative filtering step, we select for each learner all learning objects with ratings score higher than or equal 3 in order to prepare an input data for the next step of our recommendation model. Fig. 3, present an example of different matrix transformations of learners’ preferences models.

2) \textit{Generate recommendation}

After generating a Learner Learning Object Improved (LLOI) matrix and selecting the most appropriate learning objects using collaborative filtering step detailed in the previous section, we use an association rule mining algorithm to retrieve the most frequent sequence of learning objects in this matrix.

Consequently, using the ARM analysis those sequences of learning objects could selected as the most appropriate learning scenario to achieve an optimal learning experience. Indeed, a learning scenario is defined as the manner an instructor or tutor could present and sequence a list of learning objects to conduct instructional activities. This scenario is designed in a way that the learner is encouraged to observe, to analyze and to learn efficiently. For example a learning scenario can be achieved by the sequence of learning objects composed with a lecture, a video presentation, read text, questions and answers and assessment.
in mind their own preferences or behaviors. However, a personalized E-learning experience could be designed and presented in a nonlinear manner in order to build for each learner the «optimal sequence» of learning objects. We defined an optimal sequence the best learning scenario can be recommended for a given learner. In this personalized scenario some learning objects like LO_10 can be ignored or isolated by the system since they are not fitting with the learner profile.

In our recommender process, we used the Generalized Sequential Pattern (GSP) algorithm to generate recommendation list, the main procedures can be described as follows [34]:

- First pass: determines the support for each item (learning object) and find the frequent 1-sequence that have the minimum support.
- Candidate generation: generates new candidate sequence (next level) from the previous frequent set of all candidates.
- Prune candidates: deletes candidate sequences that have support is less than the minimal support threshold.

IV. RESULTS AND DISCUSSIONS

We have set up experimentation to compare our approach with a classical one that does not use an adaptive teaching strategy, by measuring the student understanding after learning process. We have conducted a research on LearnFit’s effectiveness in learning “Java programming”. Our main research question was: “Does adaptive teaching strategies based learner’s profile affect the learning outcome?”

Participants for this experimentation were drawn from a pool (n = 163) of Computer Information Systems Bachelor’s degree students at ENS, Cadi Ayyad University Marrakesh Morocco in four months of 2016. Indeed, students had to study the four learning chapters in LearnFit environment. Chapter 1 “Java introduction”, Chapter 2 “Java language fundamentals”, Chapter 3 “Java Classes and methods” Chapter 4 “Framework Collection”. Four versions of subject material were implemented in LearnFit to provide personalized learning environments for students with different learning styles.

The sessions were arranged at the beginning of the course and during eight weeks of experimentation, the students studied the learning material using one of those approaches in the same conditions. The student achievement was measured at the end of each chapter and at the end of each course using MCQ. MCQ consists of questions divided into three levels of difficulty (easy, medium, and difficult). Scores for this experience were calculated on the scale of 0 to 20.

Fig. 5 shows a typical adapted course instances for learner with profile RSLQ. In this case, the concept is about the ‘while loop’ statement. The system presents the concept’s theory as simulation and associates an electronic media as image.

Fig. 6 shows a typical adapted course instances for learner with profile RSLG. In this case, the concept is about the ‘for loop’ statement. The system presents the concept’s theory as simulation and associates an electronic media as video.

When the learner completes the sequence of learning materials, the LearnFitII system evaluates the learner’s acquired knowledge for each chapter and each course. In our experimentation, the final exam for each course is considered to be acquired by the learner, if he obtains a score of control higher than 10/20. The learners’ scores can be interpreted according to the percentage of correct answers, as follows: (excellent) (18–20), (very good), (16–18) (good), (14–16), (average) (10–14) and (marginal) (<10).

Fig. 7 Comparison between traditional and personalized learning strategy distance

V. EVALUATION AND DISCUSSION

Several studies suggest that students’ satisfaction is an important factor in order to measure the effectiveness success of the E-learning process. Therefore, the satisfaction statistics are necessary to understand the perception of learners in relation to the learning process, including contents, methodology and adaptation.

To get subjective evaluation of our system, at the end of the course we organized a mandatory questionnaire that collected learners’ opinions about the main features of the system, in order to answer the following questions: Did the presentation of the entire content into different media presentations help you in your understanding? Did you find the adaptive framework easy to use? Did you enjoy learning through this adaptive system? Out of 163 learners, 145 filled in the questionnaire.

In Fig. 8, the student’s degree of satisfaction with regard to different adaptive educational systems is reported. These results are obtained by explicitly asking them their opinion. The rating score related to our adaptive system is from 1 very low to 5 very high.
Results show that most learners think that adaptive educational system is good for learning and that their requirements are satisfied. Most of them found that the system is user-friendly. They were satisfied to learn with the preferred learner’s profile and willing to use the system in the future. The results seem to support earlier studies which concluded that using learner profile matching with the learners’ psychology is helpful to students in enhancing both learning efficacy and efficiency [8, 11, 29, 30].

VI. CONCLUSION

Nowadays, recommender systems are used to support individual learning in E-learning context. Indeed, personalized learning occurs when E-learning environments make deliberate efforts to design, to elaborate and to accomplish educational experiences that fit the needs, goals, talents and interests of their learners. Furthermore, the issues concerning personalization in learning process have been widely discussed in the past decades and remain the focus of attention of many researchers to day. However, there are several limitations when applying the existing recommendations algorithms. To address these limitations, we propose in this paper an automatic courseware authoring based on learning identification and collaborative filtering techniques. The main idea is to deliver personalized recommendations for each learner by selecting and sequencing the most appropriate learning objects into a coherent, focused organization in online distance education. To deal with the absence of data about learner and his/her preferences during the first connection, the framework offers a content-based recommendation based only on the learning style. This teaching strategy will be adjusted by the decision body of the system using filtering collaborative methods in order to achieve the desired fit.

Our experimental results show that a combination of the learners’ learning styles and hybridge recommendation techniques has the potential to improve the quality of an intelligent e-learning system, as well as keep the recommendation up-to-date. In the future, we plan to refine the recommender model to deal with several inherent issues such as data sparsity and data correlation. Since CF methods are known to be vulnerable to these problems in recommendation. In addition, we will consider more complex recommendation approaches, by including other factors such as learner motivation, knowledge level, and apply other intelligent artificial techniques.

REFERENCES