A HYBRID RECOMMENDER STRATEGY ON AN EXPANDED CONTENT MANAGER IN FORMAL LEARNING

F. Madeira\textsuperscript{1}, R. Quaresma\textsuperscript{2}, S. Abreu\textsuperscript{2}

\textsuperscript{1}Institute Polytechnic of Santarém (PORTUGAL)
\textsuperscript{2}University of Évora (PORTUGAL)
filipe.madeira@esg.ipsantarem.pt, quaresma@uevora.pt, spa@di.uevora.pt

Abstract

The main topic of this paper is to find ways to improve learning in a formal Higher Education Area. In this environment, the teacher publishes or suggests contents that support learners in a given course, as supplement of classroom training. Generally, these materials are pre-stored and not changeable. These contents are typically published in learning management systems (the Moodle platform emerges as one of the main choices) or in sites created and maintained on the web by teachers themselves. These scenarios typically include a specific group of students (class) and a given period of time (semester or school year). Contents reutilization often needs replication and its update requires new edition and new submission by teachers. Normally, these systems do not allow learners to add new materials, or to edit existing ones.

The paper presents our motivations, and some related concepts and works. We describe the concepts of sequencing and navigation in adaptive learning systems, followed by a short presentation of some of these systems. We then discuss the effects of social interaction on the learners’ choices. Finally, we refer some more related recommender systems and their applicability in supporting learning.

One central idea from our proposal is that we believe that students with the same goals and with similar formal study time can benefit from contents’ assessments made by learners that already have completed the same courses and have studied the same contents. We present a model for personalized recommendation of learning activities to learners in a formal learning context that considers two systems. In the extended content management system, learners can add new materials, select materials from teachers and from other learners, evaluate and define the time spent studying them. Based on learner profiles and a hybrid recommendation strategy, combining conditional and collaborative filtering, our second system will predict learning activities scores and offers adaptive and suitable sequencing learning contents to learners. We propose that similarities between learners can be based on their evaluation interests and their recent learning history. The recommender support subsystem aims to assist learners at each step suggesting one suitable ordered list of LOs, by decreasing order of relevance.

The proposed model has been implemented in the Moodle Learning Management System (LMS), and we present the system’s architecture and design.

We will evaluate it in a real higher education formal course and we intend to present experimental results in the near future.

Keywords: Personalized Recommender Systems (PRS), collaborative filtering, collaborative learning, formal learning, sequencing, learner profile.

1 Introduction

Technological innovations, increasing research and experimentation lead learning to new scenarios where time and space are assuming different meanings. With the Internet, identifying suitable learning resources from a potentially overwhelming variety of choices became a critical service. Social networks and cloud computing are recent innovations that enable easier access to resources (in different periods and spaces and using multiple terminal devices). Distance learning (even as a supplement to classroom teaching) has assumed increasing importance.
Learner-centered instruction is another trend of the learning process. This approach defends that knowledge is created or built by self-learner and not as mainly the result of transmission by others [1]. The constructivist learning model defends that knowledge emerges as a result of social construction, in a collaborative environment with interaction between all the different learning agents. This environment requires new technologies, new teaching practices and new support tools. Different learners with different characteristics, skills, capacities and goals, seek for the most suitable learning activities and materials. Different contents and adaptive sequencing on learning activities are some of the requirements for this to happen.

The main topic of this paper is to find ways to improve learning in a formal Higher Education context. In the typical environment, the teacher publishes or suggests contents, which support learners in a given course, as supplements of classroom training. Generally, these materials are pre-stored and not changeable. Such contents are typically published in Learning Management System (LMS), like the Moodle platform (this platform emerges as one of the main choices), or in sites created and maintained on the web by the teachers themselves. Normally, these scenarios include a specific group of students (class) and a given period of time (semester or school year). Contents reutilization often needs replication and its update requires new edition and new submission by the teachers. Usually, in LMSs only teachers can publish contents (in typical formal education courses).

In the proposed solution we will consider two different systems. In the first one, the extended content management system, learners can select pre-existing materials and add new materials to a web-based platform. They are also required to evaluate and point out how much time they took to study those materials. The second system provides a hybrid strategy that combines technical recommendations with some profile-based filtering to offer adaptive and suitable sequencing learning contents to learners, in order to be able to improve personalization of learner’s learning path and also adding diversity to the learner ways of study; that is, to recommend the most interesting or relevant Learning Objects (LOs) to each learner. According to IEEE [2] a Learning Object (LO) is “any entity, digital or non-digital, that may be used for learning, education or training”.

1.1 Related Work

We begin with a description of the concepts of sequencing and navigation in adaptive learning systems, followed by a short presentation of some of these systems. We then discuss the effects of social interaction on the learners’ choices. Finally, we refer some more related recommender systems and their applicability in supporting learning.

1.1.1 Adaptive Learning Sequencing and Navigation

Sequencing and navigation have been studied extensively by researchers and standardization bodies. Sequencing is the process responsible for ordering the learning objects that will be presented to students. Navigation is the process that allows the student to move from one learning object to another.

Sequence process can be based on static or adaptive models. In the first model, is the course design (or the learning event) that establishes the possible sequences that each learner can follow. The adaptive sequencing is established by a set of explicitly designed rules. Static models do not change in time and do not allow the inclusion of new content, which gives them a limited validity. As they are usually designed by a single teacher, they are also not free of errors.

1.1.2 Adaptive Learning Systems

Oppermann [3] refers that Adaptive Educational Hypermedia Systems (AEHS) should be able to “adapt their own characteristics automatically according to user needs”. Adaptability can be achieved with an adaptive selection of contents, an adaptive navigation or an adaptive contents’ presentation. The combination of these different forms is also present in several works. Adaptive systems are based mainly on the skills and competences that learners want, in their profiles and needs, in a set of rules and/or some algorithms that generate adaptability.

Some of these hypermedia adaptive systems were inspired by the AHAM architecture - Adaptive Hypermedia Application Model [4]. This architecture suggests several models. The domain model considers the learning goals and the subject domain concepts. The user model describes information and data about an individual learner, such as knowledge status, learning style preferences, etc. The
user model contains two distinct sub-models, one for representing the learner’s state of knowledge, and another one for representing learner’s cognitive characteristics and learning preferences (such as learning style, working memory capacity etc.). Media space includes the Content Management System (CMS) where learning resources are stored, and a resource description model where some pedagogical characteristics of the LOs (such as its type or its difficulty), as well as structural relationships between them (if a LO requires another to be done previously). The adaptation model defines the concept and content selection rules. The concept rules are used for selecting the appropriate concepts. The content rules are used for selecting the appropriate LOs from the CMS. Due to the problems of inconsistency and insufficiency of the defined rule sets in the adaptation model, conceptual “holes” can be generated in the produced LO sequences (or learning paths). So, one relevant problem of this rule-based design approaches, is the fact that they need a complete and correct set of adaptation rules, since existence of inconsistencies and gaps in the rules can generate sequences of incomplete concepts.

Luis de-Marcos et al. used rules combined with optimization algorithms for adaptive sequencing generation [5]. Other solutions excluded rules and are based only on algorithms like [6] and [7], or include students’ assessments to define adaptation ([8], [9]).

1.1.3 Social Sequencing

The above approaches presented have a common weakness: a mistake done by human designer affects the whole system. Aside from the LOs proposed by the teacher, students may as well discover and propose other LOs of interest to their colleagues. Also, one LOs’ sequencing design needs to evolve. Learners evolve over time, as well as their characteristics, interests and aims. Innovations, changes in formal curricula and new knowledge sources, push for new sequences of LOs.

Internet and its social applications can help with some answers to these problems. In Web2.0, contents are created, modified, shared, recorded and classified by their users. This trend has driven the emergence of learning networks where interactions between learners and between learners and teachers have dramatically increased. In learning networks connections with people and information, are developed and maintained to support one another’s learning. The behaviour and contributions of each agent in these learning networks can improve and change the learning contexts. The importance of these networks, whether formal or informal, was recognized by the EU itself, with some initiatives such as TENcompetence (http://www.tencompetence.org/).

Iglesias et al. refer that social interaction may influence the adaptive sequencing of learning activities, considering the interactions within a group [10]. Koper purposes one system that suggests the next activities based on successfully completed (by other students) activities [11]. Gutiérrez & Pardo suggest the use of annotations, indicating how many students have already done and how many have successfully completed a given activity [12].

1.1.4 Recommender Systems

One kind of social interaction can be implemented using Recommender Systems (RSs). As mentioned earlier, our proposed system makes individualized recommendations of LOs for learners. RSs are widely used in current web applications, like sales applications for books, movies, or music, among other items. The main purpose of such applications is to filter information that may interest or help each user on his choice or selection. They are based on collaborative filtering of information obtained from the behaviour of other users on the web (with messages like, "who bought this item also bought the following ... "). These applications are also considered in the group of social interaction ones. Many of these systems also include other information such as ratings and tags about the items they need to recommend.

Information is filtered by these systems according to next approaches: cognitive (or content-based) filtering and collaborative (or social-based) filtering. The combination of these techniques, optionally with other non-collaborative techniques originates hybrid solutions. The cognitive category of recommendations uses information about characteristics from users and from items involved in previous selections. The system recommends items with similar or related characteristics to those that exist in their profile ("show me more that I like, based on my past tastes"). The collaborative (social) approach recommends items based on what other similar users have considered (share same tastes, preferences, situations, ...) or have evaluated in the past ("tell me what is popular among my peers"). RSs also diverge in the way they get data or information about items and users. Typically, this data
can be obtained explicitly or implicitly.

There are many RSs with different types designed to support learning activities. Manouselis et al. [13] provide an introduction to RSs for Technology Enhanced Learning (TEL) settings considering the particularities of this application domain. The main purpose of such systems is to filter information which may interest or help each learner on his choice or selection. The Altered Vista system [14] recommends web addresses based on teachers and learners evaluations. Rafaeli et al. [15] propose one collaborative filtering system where users can select the users from whom they want to accept recommendations. Manouselis et al. [16] describe a case of developing a learning resources collaborative filtering service for an online community of teachers in Europe. Some proposals have been developed using multiple criteria to perform collaborative filtering, like [17] did. Some authors have been proposed the use of hybrid strategies arguing that they produce recommendations more reliable when compared with the single use of one technique ([18], [19], [20]), and some hybrid systems have also been developed ([21]; [22]; [23]). Another interesting work is developed by Drachsler et al. who compared the recommender system’s applicability in informal versus formal education [24].

Herlocker et al. review some key decisions in evaluating collaborative filtering RSs [25]. Some proposals that have been implemented have also been evaluated using different techniques, like surveys [22], metrics [26], or both surveys and metrics [16], like we did.

1.2 Organization of the Paper

The remainder of this paper is organized as follow. In section 2, we will present the overall system architecture, the details of each subsystem, and techniques we are applying to make final recommendations. In the following section, we will describe a running example. We’ll conclude this paper with some considerations, discussing limitations and improvements needed to the proposed system, as well as giving some directions for future work.

2 A HYBRID RECOMMENDER STRATEGY ON AN EXPANDED CONTENT MANAGER

Usually, in learning management systems only teachers can publish contents (in typical formal education courses). The proposed system includes a subsystem that allows learners to add new LOs. Our solution also includes a second subsystem that recommends LOs to learners (see figure 1). Note that in formal education, learners that enrol in a given course show common interests and accept the goals and skills that are implicitly established and associated to this course. Likewise, we assume that the skills that each student has already acquired were also considered in the requirements to access this course (conditions of higher education access, frequency of other courses, modules, etc.).

However, the learner’s levels of proficiency at specific competencies may vary. This, by itself, justifies that students need to work differently to achieve the same goals, at the end. In this article, content is considered as a broad concept, close to that of the learning activity, and may represent a document, a link to some resource hosted on the web, a questionnaire, an exercises file, etc.

2.1 Extended Content Manager

The extended content management system has an innovative process with several steps, which begins with the submission of one new LO. After that, next steps includes the definition of its sequence order, its prerequisites (if applied), its relationship with other LO, validation and final publication (see figure 2). All LOs from same topic must have the same order number and relationships are only established between LOs which have the same topic. For each topic, one LO must be classified as “base” (a main LO about a topic), and the others can be classified as “upgrade”, “similar” or “supplementary”. This field facilitates the identification and selection of LOs associated with a determined topic. Prerequisites are associated to topics in order to force a learner studying a necessary topic before. Each topic has only one topic as prerequisite thus all LOs of the same topic share the same prerequisite.
All of these four steps correspond to features accessible to teachers. Learners only have access to the features of step 1. They can also establish LO’s relationships, but learners aren’t able to define LOs order, nor indicate any prerequisites. The person that publishes the LO may also indicate an estimated time needed to study it. Only after approval, do the new LOs become available for publication and subsequent access by all learners.

2.2 Hybrid Recommender System for Learning Contents

In this proposal, one of our central ideas, in which we believe in, is that students with the same goals and with similar enrolments in formal courses can benefit from LOs assessments made by learners that already have completed the same courses and have studied the same LOs.

The recommender support subsystem aims to assist learners at each step suggesting one suitable ordered list of LOs, by decreasing order of relevance. To accomplish this goal, we propose a hybrid strategy with some different techniques, applicable in cascade, each one refining the received list of recommendations (see figure 3). At a final stage, we try to predict the satisfaction level of the list of contents. It can be formulated as follows:

\[(1)\]

where \(A\) is the set of students enrolled in the course, \(C\) is the set of LOs that can be recommended, \(R\) is an ordered set of recommendations and \(u_a(c)\) is the utility function that predicts the LO’s classification to the learner.

Recommendation techniques are based on the interests of learners, defined in their profile. In this learner profile we have considered the minimum desired satisfaction level (scale 1 to 5), the maximum duration of the study for each LO, the sequence length (i.e. what are the previously selected contents and in what order) and the desired minimum assessment. These values should be explicitly added by each learner to their profile. Note that values on the profile of each learner correspond to their interests, and not necessarily reflect their behaviour. For example, one learner may have an average duration of study of 30 minutes for each LO, but if he has availability to take up more time with each one, he can set in his profile a higher duration than this average.

Before asking for a new recommendation, it is checked whether there is no any LO to be finalized by the active learner. If this condition is verified, the first technique (relational filtering) is applied to determine which LOs the learner may choose. This process excludes LOs that have already been selected by the active learner and those LOs that have prerequisites not yet attended (i.e. they need other LOs to be done firstly).

In phase 2, if the learner has defined, on his profile, the maximum duration of study time he is prepared to spend in a single LO, the list obtained from phase 1 is revised taking in account the study time indicated by the other students for each LO in the list (social filtering). This average time calculation includes the suggested duration indicated by the LO’s publisher (teacher or learner). This solution solves the cold-start problem when one LO has no selections. The result's list includes only the LOs with an average time shorter than or equal to the time defined in the profile of the active learner. If the learner does not define any minimum time in his profile, it is maintained the same list that has resulted from stage one. For each LO \(c\), this average time is obtained by the following formulation:

\[(2)\]

where \(A\) is the set of all the learners that have selected LO \(c\) and have defined its value for the duration field, plus its author (learner or teacher). \(t_a(c)\) is the duration of LO \(c\) defined by learner \(a\) or by its author.

The next step (3) begins with the calculation of the similarity between learners (defined between a minimum of 0 and a maximum of 1). This similarity is calculated from two metrics. The first metric considers the average grade achieved by each learner in the active course and the grade value defined in the active learner’s profile. For colleagues with a grade greater than or equal to that value,
the maximum value of similarity (1) is assigned. The complete formulation is:

$$\text{similarity} = \text{metric of similarity}$$

where \( \text{similarity} \) is the metric that evaluates the grade-similarity between learners \( \text{learner}_1 \) and \( \text{learner}_2 \). \( \text{grade}_1 \) is the grade defined in the active learner profile and \( \text{grade}_2 \) is the grade achieved by the learner.

The second measure considers the size of the sequence defined in the active learner’s profile. We propose that the selection of the same object and at the same order, in the past, by other learners means that they have more similarities among themselves, so future choices should consider this proximity. The value defined in the profile of each student sets the sequence length for the latest LOs studied. Learners who selected the same LOs in exactly same order (even if in different positions), will have the maximum similarity. For new students, who have not yet selected any LO, this metric is not calculated. This measure is calculated using expression 4:

$$\text{sequencing-similarity} = \text{metric of sequencing-similarity}$$

where \( \text{sequencing-similarity} \) is the metric that evaluates the sequencing-similarity between the learners \( \text{learner}_1 \) and \( \text{learner}_2 \).

The final learner similarity between two learners is obtained by the arithmetic average of the two metrics. If one metric value is 0, then the value of the other metric will be the only one to be considered.

Finally, to complete this stage of collaborative filtering, for all LOs that belong to the previous output list we will predict the expected satisfaction value of the active learner.

To calculate these prediction values we consider usual formulations from memory-based collaborative filtering algorithms [27]:

$$\text{prediction} = \text{metric of global similarity} \times \text{metric of satisfaction level}$$

where \( \text{prediction} \) is the prediction of the satisfaction level for LO \( \text{LO}_1 \) to the active learner \( \text{learner}_2 \). \( \text{metric of global similarity} \) is the metric that evaluates the global similarity between learners \( \text{learner}_1 \) and \( \text{learner}_2 \). \( \text{metric of satisfaction level} \) is the satisfaction level explicitly defined by learner \( \text{learner}_2 \) for LO \( \text{LO}_1 \). \( \text{average satisfaction level} \) considering all selected LOs by learner \( \text{learner}_1 \) is the same concept but for learner \( \text{learner}_2 \).

The last formulation (7) provides some solutions for the cold-start problem. When new students enrol a course there are neither selections nor LOs’ evaluations. The same situation occurs when a new LO is imported. In these cases or if the prediction value is zero, the final prediction value for the satisfaction level of a given LO, is the default value assigned by the teacher.

The final list of LOs’ recommendations is ordered by the prediction values obtained, in a decreasing order, and filtered based on the minimum desired level of satisfaction indicated by the active learner. Only LOs with a value superior or equal to the one indicated in the learner profile will be presented to him. From this final list, each learner can select one or more LOs, not necessarily the highest ones of the list.

3 **RUNNING EXAMPLE**

In this section, we introduce a simple running example to better illustrate a typical flow of the proposed system. Suppose there are four learners (…) and six LOs (…). Some LOs have been selected by some learners (sequence column). Learners have defined some of the fields in their profiles, and some have an average grade (grade column) calculated from topics that they had completed before. Each learner registered the necessary study time (\( d \)) and satisfaction level (\( s_l \)) for each LO that they
have already done. This example is shown in Table 1.

After step 1, where relational filtering occurs, the resulting list of LOs for learner  is formed by the following set of LOs (not yet been selected by him): .

In the next step, called social filtering, the study time of other students is used to calculate the average time for each LO. This value is compared with time defined in the learners’ profile to decide if it will be maintained or removed from list of recommendations. Thus, for LO , which has an estimated duration assigned by the teacher of 50 minutes, the average duration is 52.5 (less than profile value of 55 for learner , so this LO will remain in the list). Because LO has not yet been selected, its default value (60 minutes) will be considered. Hence, this LO will be removed from the list of recommendations for the same learner . At the end of this step, the list is formed by LOs .

The calculation of the grade-similarity is the first step of collaborative filtering phase. Considering the data of Table 1, =0.76 ([1-(17-13)/17]); =1 (the grade achieved by learner is equal to that defined in the profile of learner , i.e. 14 values); and =0.88 and  can’t be calculated because there is no grade defined for learner .

After that, it follows the calculation of sequence-similarity. Its value is 1 for , because learner has done the same sequence of LOs (length equal to 3) that learner did ; = 0.4 and  = 0.6.

As already mentioned, global similarity metric is obtained by the arithmetic average of the two metrics, which have been calculated previously. So,  = 0.88,  = 0.64 and  = 0.6.

After that, it follows the calculation of the prediction for the satisfaction value for each LO, to the active learner. From our example, the list of recommendations in this step contains LOs and . =4.75 and =4.08. Although it is not necessary to calculate the value for , its value would be =4 (satisfaction level assigned by the teacher).

Finally, the resulting list of recommendations for learner shows LOs and , by this order, with the predict value for the satisfaction level of 4.75 and 4.08, respectively.

4 CONCLUSION

In this paper, we describe one LMS with some extra functionality not usually available in traditional ones like Moodle. We enable learners to add additional LOs. We also propose relationships between contents with different semantic meanings. The access to these new contents, submitted by learners, needs teacher approval, since we are in a formal course learning setting. We defend that these contents might be available for several years and be used for future learners of the same courses, which is very different from traditional approaches where course contents are published to be accessible during just one semester or one academic year. This will enable the system to have more contents for learners to choose from, and analyses of all previous learners’ interactions with the system. This analysis is the basis, for our hybrid recommender strategy, which will permit recommending suitable LOs to learners. We also believe that this extended system, will increase learning motivation without mischaracterizing the current formal learning model.

In a system with a wide range of contents, it is important to have some support to select the most suitable contents for each learner. We want to enhance the suitability of their choices to their interests. So we propose a model based on a hybrid recommendation strategy that considers profile information (study time, sequencing and learner grade and satisfaction level). These four fields do not form a completed or unique solution, but rather a custom solution. There will be, of course, other criteria that may be included in the recommendation strategy, such as contents’ difficulty level or its validity. We want to test if this solution can contribute to create different sequencing learning activities and unique experiences of learning. One relevant formulation for the recommendation strategy is the one presented for the calculation of the similarity between learners. It reflects the interests of the student (profile) in the calculation of grade-similarity metric (greater for learners with grades greater than or equal to that defined in profile) and in the calculation of sequencing-similarity (where the presence of the same contents and their order of selection will be valorised for the latest contents that were chosen by active learner).
We have developed a prototype of this system as a module in Moodle (version 2.2.1). It is published on Apache web server, and it uses MySql for database management. We used PHP as the programming language for developing the referred module.

As future work, we will analyse the level of student involvement in the publication of new content, the diversity of learning paths taken by several learners and the usefulness of the recommendation techniques proposed. We are interested in testing the behaviour of the metrics in use, analyse, measure and evaluate on their suitability and the quality of the predictions (using metrics such as Mean Absolute Error). We will try to optimize this model or at least identify its limitations.

The next step of this work is the experimental evaluation using the prototype in a real Higher Education course.

REFERENCES


