Improve Performance of Recommender System in Collaborative Learning Environment based on Learner Tracks

Qing Tang¹, Marie-Hélène Abel¹ and Elsa Negre²

¹Sorbonne Universités, UTC, CNRS UMR 7253, HEUDIASYC, 60200 Compiègne, France
²Paris-Dauphine University, PSL Research University, CNRS UMR 7243, LAMSADE, 75016 Paris, France

Keywords: Online Learning, SoIS, Collaborative Learning Environment, Recommender System, Learner Track.

Abstract: Learning with huge amount of open educational resources is challenging, especially when variety resources come from different System of Information Systems (SoIS). How to help learners obtain appropriate resources efficiently in collaborative learning environment is still a rigorous problem of research. This paper proposes a method to calculate learner’s knowledge competency by tracking and analyzing their behaviors in a collaborative learning environment based on SoIS, and combining other basic learner’s information to build a personalized recommender system to help learners select appropriate educational resources to improve their learning efficiency.

1 INTRODUCTION

With the development of Internet information technology, human society has stepped into an era of information overload. Owing to the overwhelming quantity of information, both information providers and consumers are facing challenges: information providers are willing to find the information to be transferred to the target audience while information consumers are willing to find the information most relevant to their needs (Wang, 2016).

In a collaborative learning environment of System of Information Systems (SoIS), acquiring educational resources from appropriate channels could also be much tougher (Wang, 2016). In such an environment, heterogeneous educational resources is collected from separate systems (Saleh and Abel, 2018). When faced with so many heterogeneous resources, learners have information burden problem and difficulty in resource decision-making. We want to help those learners who use SoIS-based collaborative learning environment for learning to efficiently obtain the most suitable educational resources for their current learning process.

Recommender systems in online learning is a branch of information retrieval where learning resources are filtered and presented to the learners (Chughtai et al., 2014). However, how to improve the accuracy of the recommender system is still a direction worth researching. This paper proposes a scheme for calculating learner's knowledge level from their historical learning tracks in collaborative learning environment of SoIS, and combining with basic information (e.g., degree, profession, preference, etc.) to build a personalized recommender system. This recommender system is based on learners interaction tracks and semantic description, it can help deliver appropriate educational resources to target learners, thereby reducing information burden and difficulty in decision-making and improving their learning efficiency.

The structure of this article is as follows: Section 2 reveals the background and technologies, Section 3 explains the proposed method and technical support, Section 4 discusses the core ideas and shortcomings, Section 5 makes a summary and expressed the perspective for future work.

2 BACKGROUND

This section introduces the three core theoretical parts of this paper: SoIS, Collaborative Learning Environment, and Recommender System. And gives a framework to provide theoretical support.
2.1 System of Information Systems

System of Systems (SoSs) are large-scale integrated systems which are heterogeneous and independently operable on their own but are networked together for a common goal (Carlock and Fenton, 2001) (Jamshidi, 2011). SoIS is a special type of SoS in information domain. SoIS are networks of agents interacting in a specific technology area under a particular institutional infrastructure for the purpose of creating, diffusing, and utilising technology focused on knowledge, information, and competence flow (Carlsson and Stankiewicz, 1991). SoIS are the specific clusters of the firms, technologies, and industries involved in the generation and diffusion of new technologies and in the knowledge flow that takes place among them (Breschi and Malerba, 1996).

Based on the definitions provided, SoIS has such features (Saleh and Abel, 2018):

- SoIS addresses impact of the interrelationships between different Information Systems (ISs).
- SoIS is concerned with the flow of information and knowledge among different ISs.
- SoIS is responsible for generating information from the constituent ISs.
- Information interoperability is a key issue when designing and implementing an SoIS.

2.2 Collaborative Learning Environment

Collaborative learning is an educational approach to teaching and learning that involves groups of students working together to pursue a same learning goal, solve a learning problem, or complete a learning task (Vijayalakshmi and Kanchana, 2020). Collaborative learning environment is a community to support group members’ coordination and interaction so that they complete the task more efficiently, due to the way of group learning facilitates a comfort communication between students to share their resources, discuss their problems and receive the appropriate solutions (Riyahi and Sohrabi, 2020). A collaborative learning environment is an interconnected virtual place of people, systems, and resources to support learners by using multiple tools to access educational resources (Gütl and Chang, 2008). Collaborative learning environment consists of biotic parts which are the learning community (e.g., teachers, learners, etc.) working together with abiotic parts which are the learning utilities (e.g., technologies, Information Systems, etc.) (Álvarez-Arregui et al., 2017). The elements of collaborative learning environment are remix of different forms of technologies, devices, data repositories, information retrieval, information sharing, networks and communication (O’Connell, 2016).

Collaborative learning environment has many advantages, learners form groups, collaborate with each other and with educators, and content designed for interaction (Ouf et al., 2017). Learners do not evolve alone as single individual, but in a learning environment that includes the learners and their physical and social equipments: tools (e.g., notepad, tablet, etc.), resources (e.g., procedures, methods, instructions, course materials, notes, document, etc.), and the partners (e.g., teachers, network of experts, work colleagues, etc.) (Perkins, 1995). And this environment can be seen as a virtual learning space in which technologies that contribute to learning (e.g., hardware, software, network, etc.) are used to foster interactions between communities of actors and content. Learners who share and use resources and knowledge information about common interests, and variety of educational resources are accessible through the learner’s own memory as well as through their tools or partners (Saleh and Abel, 2018).

2.3 Recommender System

Recommender systems can be classified according to three approaches: score estimation method, the data used to estimate scores or the main objective of the system (Negre, 2015). Whatever recommendation technique is used, certain information needs to be considered in relation to users and this kind of information usually store in user’s profiles, and the common algorithms mainly are Content Based, Collaborative Filtering, Knowledge Based, and hybrid approaches (Negre, 2015). Fanaeetork and Yazdi (2013) proposed a Content Based method based on vector space model which the users’ profiles are enriched using ontologies, the ontologies are made by combining the text mining and NLP (Natural Language Processing) techniques. Yoldar and Özcan (2019) proposed a Collaborative Filtering method on an online ad dataset, which is based on bi-clustering and ordered weighted average aggregation operators, can address situations such as the lack of implicit feedback on items. Paradarami et al. (2017) present a Hybrid method with a deep learning neural network framework that utilizes reviews in addition to content-based features to generate model based on predictions for the business-user combinations.

The application of recommender systems in online learning has become a thriving research field (Pan et al., 2010). The task of a recommender system in online learning is to recommend relevant learning
materials to the students and help them in decision making (Aguilar et al., 2017). Zheng et al. (2015) proposed a recommendation approach to mitigate learning issues in online learning communities. Chen et al. (2014) proposed a recommender system to recommend learning materials in an online learning platform and their results demonstrated significant improvement in performance. Takano and Li (2010) proposed a recommender system for online learning by utilizing a feedback method that extracts student’s preference and web-browsing behavior. Experimental results of previous works show that using recommender systems on online learning community obtain significant achievements (Kardan and Ebrahimi, 2013).

### 2.4 Framework

Many learners now choose to join the online learning platform while completing their social studies. For most of the current situation, many learners join the online learning platform with great interest, however many of them are unable to persist due to the information burden of a large number of resources and the lack of precise recommendation assistant. We aim at helping such learners. From the previous sections, SoIS can provide massive and diverse learning resources for collaborative learning environment, and the recommender system can solve the information burden problem encountered by learners in collaborative learning environment. Therefore, we propose to provide learners with a complete learning framework, which contains a collaborative learning platform based on SoIS and a recommender system associated with the learner’s track. Such a learning framework can not only integrate diversified resources and enable learners to learn in a collaborative manner, but also provide accurate recommendations to maintain learner’s enthusiasm for learning.

The structure of framework: An online learning platform based on SoIS, learners are divided into different learning groups according to their learning goals, each group has a certain study subject. A learner can choose to join one or more study groups according to their learning goals. Every learner has the authority to upload and share learning resources on the platform, and these resources can come from other different resource systems. And learners can search for the resources they want in the platform; learners can discuss with other learners in the same group; learners can give scores to used resources.

### 3 THE PROPOSED METHOD

This paper focuses on improving the performance of recommender system of a collaborative learning environment in SoIS, to help learners make decisions and reduce their information burden from massive heterogeneous educational resources by recommending appropriate educational resources, and ultimately improves their learning efficiency. It proposes a method to record and analyze learner’s historical learning behaviours, calculate their knowledge level. Then, build a learner model, and use the vector space model to transform it into a multidimensional space (Sivaramakrishnan et al., 2018), and make personalized recommendations by calculating the cosine similarity of learner vectors (Geng et al., 2018).

#### 3.1 Tracking System

When learners complete learning actions in collaborative learning environment, they leave ‘digital tracks’, almost all interactions in the past represent tracks and can be regarded as learner’s study experience (Wang, 2016). The information extracted from these tracks can help in many cases (e.g., decision making, recommendation, etc.).

Reiner et al. (2001) proposed a complete set of plans to record and analyze network user behaviour, and visualize the data for easy analysis. Benevenuto et al. (2009) used web crawlers to obtain user behaviour data from websites for analysis. Alexandros and Georgios (2013) proposed a framework for recording, monitoring and analyzing learner behaviour while watching and interacting with online educational videos. Although these methods have been proven effective, they are not efficient enough and some methods are difficult to achieve. We propose to use xAPI1, which has a complete and efficient method, to help recommender system collect and organize the behaviours of learners in a collaborative learning environment.

xAPI is a standard established by the US Department of Defense and the White House Office of National Science and Technology Policy in the Advanced Decentralized Learning program, which is a new specification for learning technology that makes it possible to collect tracks data about the wide range of experiences a person has (online and offline). People learn from interactions with other people, content, and beyond. These actions can happen anywhere and signal an event where learning could

---

1https://xapi.com/
occur. All of these can be recorded with xAPI. When an activity needs to be recorded, the application sends secure statements in the form of ‘Noun, verb, object’ or ‘I did this’ to a Learning Record Store (LRS). LRS in xAPI has two types of APIs:

- Statement API for inputting and outputting the statement;
- Document APIs for access richer information (e.g., strings, word documents, pictures, videos, etc.).

Both types of APIs follow the RESTful architecture, enabling data in LRS to be processed by HTTP, including adding, deleting, querying, and modifying. LRS provides a variety of attributes\(^2\) to choose from, as shown in Table 1.

Table 1: Attributes of LRS.

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
<th>Required</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>id assigned by LRS if not set by the learning record provider.</td>
<td>Recommended</td>
</tr>
<tr>
<td>actor</td>
<td>whom the Statement is about</td>
<td>Required</td>
</tr>
<tr>
<td>verb</td>
<td>action taken by the actor.</td>
<td>Required</td>
</tr>
<tr>
<td>object</td>
<td>activity, agent, or another statement</td>
<td>Required</td>
</tr>
<tr>
<td>result</td>
<td>representing a measured outcome</td>
<td>Optional</td>
</tr>
<tr>
<td>context</td>
<td>context that gives the Statement more meaning.</td>
<td>Optional</td>
</tr>
<tr>
<td>timestamp</td>
<td>timestamp of when the events described within this statement occurred. Set by the LRS if not provided.</td>
<td>Optional</td>
</tr>
<tr>
<td>stored</td>
<td>timestamp of when this statement was recorded.</td>
<td>Set by LRS</td>
</tr>
<tr>
<td>authority</td>
<td>agent or group who is asserting this Statement.</td>
<td>Optional</td>
</tr>
<tr>
<td>version</td>
<td>the statement’s associated xAPI version.</td>
<td>Not Recommended</td>
</tr>
<tr>
<td>attachment</td>
<td>headers for attachments to the statement</td>
<td>Optional</td>
</tr>
</tbody>
</table>

In addition to the three required options (‘actor’, ‘verb’, ‘object’), here are three additional options, ‘id’, ‘timestamp’, and ‘authority’, corresponding to the elements in Table 1. After the system connected to the API of LRS and set other parameters well, LRS will record every activity generated by students in the collaborative learning environment. Simply put, no matter any learner, as he/she accesses the platform server with LRS, all his/her behaviours will be recorded in the form: Track= [id, timestamp, actor, object, authority], which can be interpreted as: [track number] [time] [who] [action] [target] [where], corresponding to the elements in Table 1.

### 3.2 Learner Model

The function of the learner model is to standardize learner's information, including the learner’s basic information and the knowledge level extracted from the chaotic and disorderly behaviour tracks obtained in Section 3.1.

#### 3.2.1 Knowledge Level

Knowledge level indicates the learner's knowledge reserve and ability value in a certain field, which is an important basis for recommendation. According to the learner tracks collected in Section 3.1, the necessary information (various actions performed by learners in designated groups) can be extracted.

Each learner and each group in a collaborative learning environment has a unique ID. To calculate the knowledge level of a learner in a designated group, the necessary elements are shown in Table 2.

Table 2: Elements of knowledge level calculation.

<table>
<thead>
<tr>
<th>Learner ID</th>
<th>Group ID</th>
<th>Time span</th>
<th>Shared</th>
<th>Used</th>
<th>Reshared</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n_1)</td>
<td>(n_2)</td>
<td>(n_3)</td>
<td>(n_4)</td>
<td>(n_5)</td>
<td>(n_6)</td>
</tr>
</tbody>
</table>

Learner ID, the learner's number in the global collaborative learning environment; Group ID, the group's number in the global collaborative learning environment; Time span, the time this learner has been in designated group; Shared, the total number of resources shared by this learner in designated group; Used, the total number of resources (uploaded by this learner) used by other learners; Reshared, the total number of resources (uploaded by this learner) reshared by other learners. \(n_i\) indicate the number of times.

The calculation rule of the learner's knowledge level is shown in Equation 1:

\[
K_l^{g_j} = ts \ast w_1 + st \ast w_2 + ut \ast w_3 + rt \ast w_4
\]

\(K_l^{g_j}\) represents the knowledge level of \(l_i\) in group \(g_j\). \(g_j\) represents the j-th learning group, and \(l_i\) represents the i-th student. \(ts\) represents time span of the learner have been joined in designated group; \(st\) represents shared times completed by learner \(l_i\); \(ut\) represents used times, other learners downloaded the resources \(l_i\) uploaded, which shows that the quality of

---

\(^2\) https://github.com/adlnet/xAPI-Spec/
his/her resources has been recognized by others, and this is also a reflection of high knowledge level; $rt$ represents reshared times, other learners are willing to share again, indicating that this resource is indeed of high quality. $w_1, w_2, w_3,$ and $w_4$ are weights, which use to emphasize the importance of each element, defined according to the actual situation.

### 3.2.2 Learner Profile Standard

Ali Ben Ameur et al. (2017) proposed a learner model contains several attributes (e.g., name, age, preference, competency, etc.) of learner, however, the model contains too many dimensions for which specific precise values cannot be given. Here, redefine the learner profile standard, each dimension can reflect the important features of the learner, and the value of each dimension can be accurately defined, as shown in the following Table 3.

<table>
<thead>
<tr>
<th>ID</th>
<th>Composition</th>
<th>Dimension</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>Gender</td>
<td>G</td>
<td>1, 2</td>
</tr>
<tr>
<td>D</td>
<td>Degree</td>
<td>D</td>
<td>1, 2, 3, 4, 5</td>
</tr>
<tr>
<td>P</td>
<td>Preference</td>
<td>P</td>
<td>1, 2, 3, 4, 5</td>
</tr>
<tr>
<td>E</td>
<td>English level</td>
<td>E</td>
<td>1, 2, 3, 4, 5</td>
</tr>
<tr>
<td>K</td>
<td>Knowledge level</td>
<td>K</td>
<td>Kl</td>
</tr>
</tbody>
</table>

Define an execution standard for each dimension to prepare for the next step of learner vectorization: Gender (Male: 1, Female: 2), Degree (Elementary: 1, Bachelor: 2, Master: 3, Doctor: 4, Others: 5), Preference (Scientific papers: 1, Electronic books: 2, Video courses: 3, Report documents: 4, Others: 5), EnglishLevel (Expert: 1, Proficient: 2, Competent: 3, Advanced beginner: 4, Novice: 5), and Knowledge level is calculated from Equation 2.

### 3.3 Learner Vectorization

This Section uses the method of Vector Space Model (VSM). VSM can help to project the learner's features that the computer cannot recognize into the multi-dimensional vector space, making the learner recognizable by the computer. C.Peterson et al. (2020) proposed to use a vector space model to predict human relational similarity. Based on VSM, W.Sholikah et al. (2020) proposed method was designed to construct a general vector space and semantic relation identification, and the results show that the use of multi-task learning with a general vector space can overcome the problem of cross-lingual semantic relation identification.

According to the implementation standard in Section 3.2.2, learner can be projected into a 5-dimensional vector space, and the dimension is expressed as Equation 2:

$$v_{li}^g = (G, D, P, E, K)$$

$v_{li}^g$ is the vector of $l_i$ in group $g_j$, $g$ represents the j-th learning group, and $l_i$ represents the i-th student. The components $(G, D, P, E, K)$ correspond to the dimension code in Table 3.

All learners in designated group can be expressed as a vector set after vectorization, shown as Equation 3:

$$V_{lj}^g = \begin{bmatrix} v_{l1}^g, v_{l2}^g, \ldots, v_{li}^g \end{bmatrix}^T$$

Since the state of the learner is changing at each time point, for example, after the learner has completed the learning behaviour, their knowledge level will improve. So, it is necessary to specify the time point $t$ when building the vector set.

### 3.4 Similarity and Recommendation

This part is mainly about similarity calculation and how to recommend to target learners based on similarity results.

#### 3.4.1 Similarity Calculation

Similarity is used to measure the common characteristics between two instances, and distance is adopted to indicate the differences between them. Many tasks, such as classification and clustering, can be accomplished perfectly when a similarity metric is well-defined (Xia et al., 2015). Many similar measures have been proposed by researchers, such as Pearson Similarity (Lü and Zhou, 2011), Jaccard Similarity (Jalili et al., 2018) and Euclidean Distance-based Similarity (Hawashin et al., 2019). Cosine similarity is a widely used metric that is both simple and effective (Xia et al., 2015). After transforming in section 3.3 to obtain the learner feature vector set, the cosine similarity algorithm can efficiently and conveniently calculate the similarity between any two learners.

Su et al. (2020) define a plan to calculate vector similarity as indicators of similarity between users,

---

revealing and measuring the difference between users’ general preferences in different scenarios. Cosine similarity can intuitively show the similarity between two vectors with same dimensions.

When recommending resource to the first learner \( l_1 \) in group \( g \), at time \( t \), recommender system needs to calculate the cosine similarity between \( l_1 \) with other members \( l_i \). Here use \((A, B)\) to represent the two learner vectors \((l_1, l_i)\), 5 dimensions in two vectors correspond to the dimension code in formula 2. The calculation is shown as Equation 4:

\[
\text{Similarity}(A,B) = \frac{\sum_{d=1}^{5} A_d \times B_d}{\sqrt{\sum_{d=1}^{5} (A_d)^2} \times \sqrt{\sum_{d=1}^{5} (B_d)^2}} \tag{4}
\]

3.4.2 Recommendation

The similarity between feature vectors is an important basis for recommender system to measure the similarity of learners, because the feature vectors are transformed from the features of the learners, and the features of the learners are an important manifestation of the current state of the learners. If two learners have a high degree of similarity, it proves that their learning situations are very similar, so the resources they currently need are very similar. Recommender system can recommend resources based on this principle.

According to the cosine similarity calculated in the previous section, recommender system can recommend resources to learners who study in the same group. The processes are:

- If Learner A and Learner B have a highest similarity, from the resources Learner B has shared recently, select the latest resource and recommend that resource to Learner A.
- If none of the resources Learner B has shared recently, then, from the other Learner C in the group with the second highest similarity with Learner A, choose the resources that Learner C has shared recently, select the latest resource and recommend to Learner A.
- If there is still no result, continue from the second step, until find a qualified resource.

Simply put, the basis of this recommendation method is that learners with similar features will select similar resources.

4 DISCUSSION

In this article, our theoretical basis is building a precise recommender system in collaborative online-learning environment of SoIS based on learner’s tracks, to reduce the burden of learners to find suitable resources and improve their learning efficiency. Unlike many existing online learning environments, such as Coursera\(^4\), DZone\(^5\), the environment we designed has a personalized recommender system is based on learners internal (knowledge level) and external (gender, degree, preference, and language) information, which means that the recommender system is personalized. The system uses the xAPI standard protocol to collect and record learner’s historical learning behaviours in collaborative online-learning environment. At this step, for the recommender system, knowing the knowledge level of the learner is very important for recommending suitable resources. The system will extract information from these historical learning behaviours to calculate learner’s knowledge level and form learner models. Then the learner model will be projected to the multi-dimensional vector space through the vector space model, so that the learner becomes comparable. Finally, the similarity between learners is judged by calculating the similarity between the vectors, so as to make resource recommendations to each other learners in the same group.

From the perspective of calculation method, we chose the method of calculating the similarity of learners, which considers the features of learners and the connections between learners, but it didn’t take the resource features and the result feedbacks (attitudes of learners after using resources) in to consideration. In many cases, the features of resources are also important factors that affect the choice of learners, and the interactive results feedbacks between learners and resources are important indicators that reflect whether the resources meet the learners’ requirements. If we collect the result feedbacks generated between different learners and different resources, separate the positive result feedbacks from the negative result feedbacks, and form a data set, the direction can be transformed into supervised machine learning. Thus, it will be possible to study learners with ‘some features’ who used resources with ‘some features’ and come out the conclusion of ‘positive’ or ‘negative’. These data could be recorded and used as training data set for

\(^4\) https://www.coursera.org/

\(^5\) https://dzone.com/
supervised learning in machine learning to train a machine learning model, this model is familiar with the possible results of different learners and different resources. For the target learner, the model can judge which resources combined with him/her will produce positive results, and then recommend these resources to the him/her.

5 CONCLUSIONS AND FUTURE WORK

This paper proposed a method to improve the performance of recommender system in collaborative learning environment based on learner tracks. Although similar research has already existed, for example, Kanoje et al. (2016) used Data Mining and Information Retrieval technologies to obtain user information and use this to build a recommender system; Pan et al. (2020) proposed a dynamic user profile called User profile refactoring (builds a dynamic user profile and determines the weights of the extended tags in the profile.), combined with Tag cloud generation (discovers potentially relevant tags in an application domain) and Tag expansion (finds a sufficient set of tags upon original tags) to build a recommender system. However, none of these methods mentioned the concept of extracting information from the learner's track, because a lot of valuable information is often reflected in the behaviour. Therefore, this article proposes to obtain internal information (knowledge level) from track and combine the basic information (e.g., gender, degree, language skills, etc.) to make the recommendations has good interpretability and personalization.

To test our work, we chose the MEMORAe® collaboration platform as our experiment environment. This web platform already has the expected functions: It has the theoretical structure of the collaborative learning environment we mentioned before; the semantic collaboration model makes it possible to track learner activities (e.g., sharing, voting, etc.). It was tested as collaborative learning platform (Abel, 2015). Install and set xAPI on the web server, it will provide the learner's track information to the recommender system. Then the recommender system calculates the resources that each student may need at each moment and feeds it back to the learning platform. And as we have mentioned in the discussion section, in the later work, this article can be considered as a priori step of supervised machine learning, and it is very promising to train the machine learning model as the decision-making core of the recommender system.

REFERENCES

O’Connell, J., 2016. Networked Participatory Online Learning Design and Challenges for Academic Integrity

http://memorae.hds.utc.fr/demo/lab


