MOOCs Recommender System using Ontology and Memory-based Collaborative Filtering

Kahina Rabahallah¹, Latifa Mahdaoui¹ and Faïcal Azouaou²

¹Department of Computer Science, USTHB University, Algiers, Algeria
²Higher National School of Computer Science, ESI, Algiers, Algeria

Keywords: MOOCs, Personalized Recommender System, Ontology, Item-based Approach, User-based Approach, Cold-Start Problem.

Abstract: With Massive Open Online Courses (MOOCs) proliferation, online learners are exposed to various challenges. Therefore, the lack of personalized recommendation of MOOCs can drive learners to choose irrelevant MOOCs and then lose their motivation and surrender the learning process. Recommender System (RS) plays an important role in assisting learners to find appropriate MOOCs to improve learners’ engagements and their satisfaction/completion rates. In this paper, we propose a MOOCs recommender system combining memory-based Collaborative Filtering (CF) techniques and ontology to recommend personalized MOOCs to online learners. In our recommendation approach, Ontology is used to provide a semantic description of learner and MOOC which will be incorporated into the recommendation process to improve the personalization of learner recommendations whereas CF computes predictions and generates recommendation. Furthermore, our hybrid approach can relieve the cold-start problem by making use of ontological knowledge before the initial data to work on are available in the recommender system.

1 INTRODUCTION

Massive Open Online Courses (MOOCs) are a current trend in the field of e-learning and it attracts millions of learners, to be engaged to enjoy massive free open education courses. Prior studies indicate that MOOCs are to date suffering from low completion rate and Dropout problem (Goldberg et al, 2015; Murphy et al, 2016; Xing et al, 2016; Dhorne et al, 2017). With proliferation of MOOC development in recent years, MOOC platforms (e.g. Coursera¹, Udacity², edX³, etc) have more than millions of learners and online courses. How to find the most suitable MOOC among all proposed within the platform? how can MOOC be efficient to address the needs of its learners? are critical challenges. In this context, we assume that more personalized MOOCs’ recommendation can answer these questions (Pang et al, 2017; Piao et al, 2016).

Previous studies have shown that E-learning Recommender System (ERS) is considered an effective key solution to overcome the information overload. The Course Recommendation System is a system that recommends the best combination of subjects wherein the learners are interested (Aher et al, 2013). In conventional recommender systems like CF, Content-Based (CB), the recommendation process are based merely on ratings. However, literature reviews have shown that these recommenders suffer from cold-start problem (Bobadilla et al, 2012; Sun et al, 2017) and do not consider the additional information about user and items in making recommendations (Adomavicius et al, 2011). The initial insufficiency of ratings or preferences leads to the occurrence of the cold start problem, hence it becomes difficult to provide reliable recommendations. Generally, the cold start problem is triggered by three factors: new community, new item and new users (Schafer et al, 2007; Sun et al, 2017). Moreover, in the context of e-learning, learners have different characteristics like knowledge level, which influence personalization of recommendations. These additional learner characteristics need to be incorporated into the recommendation process (Verbert et al, 2012; Tarus

¹https://www.coursera.org
²http://www.udacity.com/
³https://www.edx.org
et al, 2017). Our approach uses ontological knowledge to address these major problems in MOOCs recommendation process.

Traditional course recommendation systems are integrated in closed e-learning environments (Aher et al, 2013), in this paper, we propose a MOOCs recommender system combining CF and ontology to recommend personalized MOOCs to online learners which can be better applied to a course recommendation in MOOC platform. Our contributions in this paper are summarized as follow:

- Combining item-based and user-based approaches, and use ontological knowledge to address the cold-start problem;
- We propose a hybrid approach which uses the ontological knowledge to integrate the characteristics of learner and MOOC in computing similarities and generating recommendations for the learners.

The remainder of this paper is organized as follows. Section 2 gives a brief review on recommendation techniques relevant to this work, and discusses related work about MOOCs recommender systems. Section 3 presents our hybrid recommendation technique. Finally, Section 4 concludes the paper.

2 RELATED WORK

This section gives a brief review on recommendation techniques relevant to this work, and discusses related work about MOOCs recommender systems.

2.1 Recommendation Technique

Several techniques have been proposed and used for recommendation generation.

2.1.1 Collaborative Filtering (CF)

The most successful and widely used recommendation technique is Collaborative Filtering (CF), which is founded on the basic assumption that users who have shown similar interests in the past will share common interests in the future (Goldberg et al, 1992). Collaborative filtering algorithms are located under two different categories, namely Memory-Based and Model-Based approaches. Memory-Based provides recommendations based on the similarities between users (or items) and ratings, to calculate predicted values. Two subcategories of memory-based CF are user-based and item-based CF approaches (Lü et al, 2012; Ghazarian et al, 2015) which we briefly explain below.

- **User-based**: As its name suggests, it focuses on the user. It consists in looking for users most similar to the target user, known as neighbour users, through resorting to similarity measures. After that, the unknown ratings are predicted based on the users’ similarities values and the ratings given to items by the similar users (Koohi et al, 2017; Wang et al, 2006);
- **Item-based**: Focuses on item instead of user. It uses to find k-most similar items to items that the target user has rated and then items’ similarities values and the ratings of target user in the similar items are used to predict unknown items’ rating (Sarwar et al, 2001).

Traditional similarity measures such as Pearson Correlation Coefficient (PCC), COSine Similarity (COS) and their variants, have been widely used in CFs to calculate similarity (Desrosiers et al, 2011).

2.1.2 Ontology-based (OB)

Ontology is originally defined by Gruber (1992) as an “explicit specification of a conceptualization”. Modeling the information at the semantic level is one of the main goals of using ontologies (Guarino et al, 2009). We use ontology instead database to provide a semantically rich description, which will allow automatic processing. Different ontology representation languages like Web Ontology Language (OWL) are used to create ontology. Moreover, the logical model allows the use of a reasoner which can infer facts about given situations.

In context of e-learning recommender systems, ontology is used to model knowledge about the learner (user context), knowledge about the learning resource, and the domain knowledge (Yu et al, 2007) to take them into consideration in recommendation process. Ontology-based are knowledge-based which is developed to deal with the cold start problem (Sun et al, 2017) since their recommendations make use of the ontological knowledge. We use also ontological knowledge to allow deducting additional information about the current context of learners and therefore, personalizing the recommendation of annotated MOOCs.

2.1.3 Hybrid Approaches

Combine two or more approaches, e.g. ontology-based and collaborative Filtering, to gain better
performance with less of the drawbacks of any single solution (Burke, 2007).

2.2 Recommender Systems for MOOCs

Commonly, the personalized Recommender System is usually divided into three main basic components: The first one is about the Recommendation Technique, the second one is about the recommending Item and the last one is about Personalization.

![Figure 1: Basic components of personalized RS.](image)

Recommendation system is widely becoming popular on online learning. For instance, Pang et al. (2017) propose an improved CF named as Multi-Layer Bucketing Recommendation (MLBR) to recommend courses on MOOC platform. At the same time, MLBR fixes data sparsity and cutting down the time cost on recommendation including offline similarity calculation, online similarity research, and update of similarity. Furthermore, they extend MLBR with map-reduce technique to improve the efficiency.

Bousbahi et al. (2015) propose a Recommender System (MOOC-Rec) using the Case Based Reasoning (CBR) approach and a special retrieval information technique to recommend the most appropriate MOOCs fitting her/his request based on learner profile, needs and knowledge.

Piao et al. (2016) investigated three different users modelling strategies based on the collected LinkedIn dataset, for personalized MOOC recommendations in a cold start situation. Results showed that the Skill-based user modelling strategy performs better than the Job and Edu-based ones.

Tang et al. (2017) proposed a personalized behaviour recommendation in a MOOC to predict the behaviour. They stipulate that this approach touches on factors more aligned with personalization, since the prediction of behaviour is an aggregation of the student’s cognitive abilities, affective state, and preferences. They investigated the suitability of this behavioural prediction approach by applying it to an expanded set of 13 UC Berkeley MOOCs run on the edX platform.

Zhang et al. (2017) propose the course recommendation model-oriented MOOC platform, MCRS, which greatly improves the data storage level and efficiency of calculation. The experiments are carried out on Hadoop and Spark and the results show that MCRS is more efficient than traditional Apriori algorithm and Apriori algorithm based on Hadoop.

2.3 Discussion

Previous related works on personalized MOOCs RS show that in the personalization process, the authors focus on learner behaviors, on collected details about learners by relying on their LinkedIn profiles, rather than learner’s cognitive abilities. Unlike previously mentioned approaches, our method incorporates learner’s cognitive abilities in the recommendation process to make personalization, and it attempts to solve the cold-start problem by using ontological knowledge in computing similarities. Therefore, we use these similarities values alongside ratings in computing predictions to improve the recommendation accuracy.

The novelty of our approach focuses specifically on ontology-based recommender system within MOOC platforms of which to the best of our knowledge, no study has been conducted to describe the learner and MOOC using ontology, and integrate these semantic descriptions into the recommendation process.

3 PROPOSED METHOD

By inspiring from (Tarus et al, 2017), our method’s focus is on prediction process, solving the cold start problem and to improve the personalization of learner recommendations. Indeed, we hypothesize that ontological knowledge may be the bridge that overcomes the limitations of cold-start problem and the absence of integration the learner characteristics (learner’s cognitive abilities) into the recommendation process. Our approach involves three steps, which are shown in Figure 2 and are introduced in following subsections: (1) creating ontologies to represent knowledge about the learner and MOOC, (2) computing similarities based on ontological knowledge and ratings as well as predictions for the target learner, (3) generation of top N MOOCs by the collaborative filtering recommendation engine.
3.1 Semantic Representation of MOOC and Learner

To provide personalization in MOOCs recommendations, several approaches have been proposed: Based on LinkedIn profiles (Piao et al., 2016), based on learner behaviors (Tang et al., 2017; Pang et al., 2017), etc.

In this paper, we focus on learner’s cognitive abilities, we propose a semantic representation of the learner profile inspired by (Rabahallah et al., 2016; Labib et al., 2017). It is based on two type of information which is acquired both explicitly and implicitly (see Figure 3).

- **Static Information**, including personal data (name, gender, age, username and password, type of learner which may be student; professionals; etc.), language, Format;
- **Dynamic information**, including different dimension such as “Interest”, “prerequisite” and “knowledge level” about a specific domain, “specialty”.

On the other hand, with the light of the MOOC platforms (e.g. Coursera, Udacity, edX, Udemy, FUN, etc.) we used all the information collected about MOOC for semantic description (see Figure 4). MOOC ontology contain more specific information such as MOOC’s name, specialty, domain, language, the start and the end of the MOOC’s Session, the prerequisites required to access, knowledge level which may be: beginner, intermediate and advanced, university who created it and the learning activities will propose by the MOOC.

3.2 Computing Similarities and Predictions

In order to predict unknown MOOCs’ ratings for the target learner and solving the cold start problem, our proposed method does the following steps:
• **Step 1**: It computes similarities between pairs of MOOCs \( \text{Sim}(o_{mi}, o_{mj}) \) looking for the k most similar MOOCs and between the pairs of Learners \( \text{Sim}(\hat{o}_{lv}, \hat{o}_{lu}) \) to find the k most similar learners, using both ratings and ontological knowledge;

• **Step 2**: It predicts learner’s unknown ratings on MOOCs based on similarities values and the ratings given to k most similar MOOCs by the target learner.

### 3.2.1 Computing MOOCs’ Similarity

To calculate the similarity and make the predictions for the target learner, the recommendation engine will use both the ontological knowledge and the ratings given by the target learner and other learners on MOOCs. In computing the similarities, some traditional similarity measures have been widely used in CFs (Patra et al, 2015; Ahn et al, 2008). However, cosine similarity (COS) among items does not consider the difference in each user’s use of the rating scale. To address this problem, an adjusted cosine similarity (ACOS) is presented (Sarwar et al, 2001).

Figure 5 indicates Learner-MOOC rating matrix. Let \( L \) denote the set of all learners \( L = \{ l_1, l_2, l_3, ..., l_t \} \), let \( M \) be the set of all possible MOOCs \( M = \{ m_1, m_2, m_3, ..., m_l \} \), and let \( R \) be the rating function that measures the usefulness of MOOC \( m_i \) to learner \( l_j \). The possible rating values are defined on a numerical scale from 1 (very irrelevant) to 5 (very relevant).

![Learner-MOOC rating Matrix](image)

In this paper, we use an extension of Adjusted Cosine Similarity proposed in (Tarus et al, 2017) to compute the similarities of MOOCs. In the proposed extension, ontological information is utilized in computing the mean rating \( \bar{r}_{lv} \). Formally, the similarity \( \text{Sim}(o_{mi}, o_{mj}) \) between two MOOCs \( m_i \) and \( m_j \) is calculated as follows (eq. 1):

\[
\text{Sim}(o_{mi}, o_{mj}) = \frac{\sum_{l \in L} (\bar{r}_{lv} \times \bar{r}_{lm})}{\sqrt{\sum_{l \in L}(\bar{r}_{lv} - \bar{r}_{lv})^2} \sqrt{\sum_{l \in L}(\bar{r}_{lm} - \bar{r}_{lv})^2}}
\]

Where \( \bar{r}_{lv} \) denotes the rating given to MOOC \( m_i \) by learner \( l \), i and j integers \( \in \{1, 2, ..., n\} \), \( i \neq j \), \( \bar{r}_{lv} \) is the mean rating of all the ratings provided by learner \( l \) based on ontological knowledge. Unlike in pure CF, ontological information is utilized in computing the mean rating \( \bar{r}_{lv} \) (Tarus et al, 2017).

### 3.2.2 Learners’ Similarity Computation

Experimental analyses show that Pearson Correlation Coefficient (PCC) has outperformed other similarity measures in user-based CF (Aggarwal, 2016; Jannach et al, 2011). For this, we use PCC to calculate ontological similarity \( \text{Sim}(\hat{o}_{lv}, \hat{o}_{lu}) \) between two Learners \( l_i \) and \( l_j \). The PCC between users \( v \) and \( u \) can be measured through (eq. 2):

\[
\text{Sim}(\hat{o}_{lv}, \hat{o}_{lu}) = \frac{\sum_{m \in M} (r_{lv} - \bar{r}_{lv})(r_{lu} - \bar{r}_{lu})}{\sqrt{\sum_{m \in M}(r_{lv} - \bar{r}_{lv})^2} \sqrt{\sum_{m \in M}(r_{lu} - \bar{r}_{lu})^2}}
\]

Where \( r_{lv} \) and \( r_{lu} \) denote the rating given to \( m_i \) by the learner \( l_v \). \( \bar{r}_{lv} \) is the mean rating of all the ratings provided by learner \( l_v \) based on ontological knowledge, \( v \) and \( u \) integers \( \in \{1, 2, ..., n\} \), \( u \neq v \).

### 3.2.3 Making Prediction

Once we obtain the set of k most similar MOOCs (k nearest neighbors) and the k most similar learners using respectively the Adjusted Cosine Similarity and PCC similarity, the next step is to compute predictions of unknown ratings for the target learner. In the case of insufficient ratings, the principle is to predict the rating \( r_{lt,m_j} \) provided by target learner \( l_t \) for a MOOC \( m_j \) (k nearest neighbors) using the rating given to \( m_j \) by other similar learners (nearest neighbors) obtained by eq. (2).

The predicted ratings are computed from the k most similar MOOCs (k nearest neighbors) obtained by eq. (1) and the ratings given on it by the target learner. To compute the predictions of ratings, we use the following prediction algorithm (eq. 3).

\[
\text{P}_{lt,m_j} = \frac{\sum_{m_i \in E} [\text{Sim}(m_i,m_j) \times r_{lt,m_i}]}{\sum_{m_i \in E} [\text{Sim}(m_i,m_j)]}
\]

Where \( \text{P}_{lt,m_j} \) is the predicted rating for unrated MOOC \( m_j \) by the target learner \( l_t \). K denotes the MOOC \( m_i \)’s similar MOOC set (K nearest neighbors), \( \text{Sim}(m_i,m_j) \) is the similarity between
two MOOCs $m_i$ and $m_j$, and $r_{l, m_j}$ is the rating of MOOC $m_j$ (the neighbors) by the target learner $l_t$.

### 3.3 Generating Individual Recommendation List

After predicting all unknown MOOCs rating in a Learner-MOOC matrix, it is necessary to generate recommendations list (top N) for the target learner. For this step, the CF recommendation engine based on the predicted ratings for the target learner and ontological knowledge to generating recommendation. After that, the filtering process consists to eliminate the MOOCs that do not adapt for the target learner’s profile (i.e. remove the MOOC that don’t correspond to the preference and need of learner, like language, knowledge level). The list (top N) generated is ranked according to their similarities with MOOC $m_i$. Algorithm 1 shows how the top N MOOCs recommendation list is generated.

**Algorithm 1: Generate Recommendation List**

- **Input:**
  
  - Set of MOOCs: $M = \{m_1, m_2, \ldots, m_n\}$
  - Set of learners: $L = \{l_1, l_2, \ldots, l_n\}$
  - Ontological Knowledge: $O = \{\text{Learner, MOOC}\}$
  - Learner ratings on MOOCs: $R = \{r_{l_1, m_1}, r_{l_1, m_2}, \ldots, r_{l_n, m_n}\}$ where $R \in \{1, 2, 3, 4, 5\}$

- **Output:** Top N recommendations list

- **Method**

  1. for each $m_i \in M$, o $\in O$, do
     
     2. Compute ontological similarity $\text{Sim}(o_{m_i}, o_{m_j})$ using eq. (1)
     
    end for

  3. for each $l_t \in L$, o $\in O$, do
     
     4. Compute ontological similarity $\text{Sim}(o_{l_t}, o_{l_s})$ using eq. (2)
     
    end for

  5. for each unknown ratings of the target learner
     
     6. Compute predicted ratings $P_{l_t, m_i}$ using eq. (3)
     
    end for

  7. Filter the MOOCs according to the learner profile

  8. return Top N recommendations list for the target learner $l_t$.

### 4 CONCLUSIONS

In this paper, we propose a hybrid recommendation approach based on ontology and collaborative filtering to recommend MOOCs to learners within MOOC platforms. The proposed hybrid recommendation algorithm incorporates ontological knowledge about the learner and MOOC into recommendation process, while Collaborate filtering predicts ratings and generates recommendations.

The novelty of our approach focuses specifically on ontology-based recommender system within MOOC platforms of which to the best of our knowledge, no study has been conducted to describe the learner and MOOC using ontology.

Directions for future research include the experimental of the proposed approach in real situation to show that the proposed hybrid algorithm can obtain better performance and accuracy than other related algorithms. We plan also to integrate other techniques such as machine learning like support vector machine (SVM).

### REFERENCES


Goldberg, L.R., Bell, E., King, C., O’Mara, C., McInerney, F., Robinson, A., Vickers, J., 2015. Relationship between participants’ level of education
and engagement in their completion of the Understanding Dementia Massive Open Online Course. BMC medical education, 15(1), 60.

In The adaptive web (pp. 291-324). Springer Berlin Heidelberg.