Social Collaborative Filtering Approach for Recommending Courses in an E-learning Platform

Youness Madani, Mohammed Erritali, Jamaa Bengourram, Francoise Sailhan

Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 151 (2019) 1164–1169
Procedia Computer Science
www.elsevier.com/locate/procedia

International Workshop on Web Search and Data Mining
April 29 - May 2, 2019, Leuven, Belgium

Social Collaborative Filtering Approach for Recommending Courses in an E-learning Platform

Youness MADANI[a,*], Mohammed ERRITALI[a], Jamaa BENGOURRAM[a], Francoise SAILHAN[b]

[a]*Sultan Moulay Slimane University, Faculty of Sciences and Technics, Beni Mellal, Morocco*
[b]*CNAM Paris, France*

Abstract

In recent years, learning online using the e-learning platforms becomes indispensable in the teaching process. Companies and scientific researchers try to find new optimal methods and approaches that can improve education online. In this paper, we propose a new recommendation approach for recommending relevant courses to learners. Our method is based on social filtering and collaborative filtering for defining the best way in which the learner must learn, and recommend courses which better much the learner's profile and social content.

Keywords: E-learning; Recommender Systems; Collaborative Filtering; Social Filtering; Sentiment Analysis; Social Network

1. Introduction

With the exponential growth of the population that want to learn online, the e-learning platforms need to adapt and innovate in the way they suggest courses to the learners. In the literature, we find a lot of methods and approaches that try to find optimal courses to a learner like those based on algorithms such as genetic algorithm, or the use of the machine learning approaches. In recent years companies and researchers begin to use the basis of recommendation systems in e-learning.

Recommender systems (RS) help people to find products and services which much their preferences and needs, and to reduce the amount of time they spend to find the items they are looking for. They are becoming increasingly important in a range of applications, such as e-commerce, music, film and book recommendation, web search, health and in e-learning platforms(in the teaching process).

* Corresponding author. Tel.: +212649651095.
E-mail address: younesmadani9@gmail.com

10.1016/j.procs.2019.04.166

RS can be roughly divided into a number of types such as knowledge-based or content-based recommendation, Social filtering recommendation, collaborative filtering recommendation, and hybrid recommendation.

Social filtering (SF) recommendation use the social content of users based on their tracks in different social networks, to recommend relevant content that much their social profiles. On the other hand, Collaborative filtering (CF) recommend rating or product for a user based on the rating preferences of similar users. CF based on the assumption that users with similar taste have similar preference to products or items, it is divided into memory-based approach and model-based approach.

in our work, each course has a number of pedagogical objectives(PO), our goal is to predict and recommend the suitable PO for the learner. for example, the java course can have different pedagogical objectives (the installation of java- the basics of java - control structures- tables- the classes- Inheritance).

The ultimate goal of our approach is to recommend items which better match the learner's profile. items in our case are the educational objectives that represent the concepts of a course, for example, Java, Python or Php courses.

Our approach emerged from the idea that recommendation performances have a huge impact on teaching success. We use RS to estimate the potential preferences of learners and recommend relevant courses(or pedagogical objectives) for learners based on social content(using social networks) and the profile content.

The proposed RS approach is divided into two parts :

- Social filtering (SF) approach: this step consist in using the social network's profiles of each learner for calculating two factors which we have called the **productivity** and the **motivation** of learners. In this step, we use the assumption that people with similar productivity and motivation will have similar profile and skills in learning. This step will help us to find the K-nearest neighbors learners (similar learners).
- Collaborative filtering (CF) approach: In this step, we first construct a profile for each learner based on his knowledge, which means that after the construction of the profile we will find the rating of the learner for each PO. After that, we calculate the similarity between the active learner and the other learners, for recommending the PO of the most similar learner(closest learner). **CF looks into the profiles of other learners to find PO that is most suitable for the profile of the target learner.

Our approach can deal with the traditional problems of recommender systems which are: the **cold start** problem, **sparsity**, and **scalability**, using the social content, the profiles of the learner and the similarity, for avoiding any problem that can affect the quality of recommendation.

The rest of this paper is organized as the following :

2. Related Works

In recent years a lot of approaches have been developed for improving the quality of recommendation and avoiding the problems related to cold star, sparsity, and scalability. The field of application of RS has also become widespread, it moved from the field of e-commerce (recommending products) to other fields such as recommending films, services or also in the domain of education for recommending courses(e-learning platform)[1].

Ar et al. in their article[2] proposed a new CF approach based on genetic algorithm (GA) for improving the result of prediction with the use of different similarity measures. The Proposed solution aims to refine and improve the similarity values. Authors of[3] proposed a new CF method to find users neighbors based on a user-based CF. They used the Pearson Correlation Coefficient (PCC) as a traditional similarity measure to find neighbor users, and the k-means clustering method and the Non-Negative Matrix Factorization model (NNMF) for clustering as traditional clustering methods.

Liu et al.[4] presented a new user similarity model to improve the recommendation performance when only a few ratings are available to calculate the similarities for each user. In this paper, the authors showed the drawbacks of some similarity measures such as cosine. In [5] authors also proposed a new CF similarity approach by improving the traditional Adjusted cosine similarity algorithm. The proposed CF approach is based on optimized user similarity, A balancing factor is added to the traditional cosine similarity algorithm, which is used to calculate the project rating scale differences between different users.

In [6] authors implemented two approaches which are Collaborative Filtering (CF) and Social Network Recommendations System (SNRS). They used Mean Absolute Error (MAE) and accuracy to compare the result of two mentioned approaches and found that the SNRS method as it is claimed to be improved version of CF works more efficiency.

3. Research Methodologie

Looking closely to web content, we remark that the number of learners that started learning online grows explosively. In addition to the number of courses published online. With that, it becomes necessary to use the recommender systems for finding the optimal and relevant courses for learners.

Our proposed RS is based on two new proposed approaches :

3.1. Social filtering approach (SF)

The SF is the first step in our approach. It uses the learners' social network profiles(Facebook and Twitter) to collect information such as published posts/tweets, likes, and comments. These informations will be used after to calculate/define two measures that we named **productivity** and **motivation**.

The productivity. define the period of the day in which the learner is more productive, for that we define three periods : (1) from 8 am to 12 pm, (2) from 2 pm to 6 pm and (3) from 7 pm to 11 pm. The idea is to calculate the number of posts/tweets published, liked or commented by the learner in each of these three periods. And after the productivity is for the period with maximum posts/tweets. The following formula shows how to calculate the productivity.

$$Productivity = \begin{cases} 1 & \text{if } Max[Card(P1/T1), Card(P2/T2), Card(P3/T3) = Card(P1/T1) \\ 2 & \text{if } Max[Card(P1/T1), Card(P2/T2), Card(P3/T3) = Card(P2/T2) \\ 3 & \text{if } Max[Card(P1/T1), Card(P2/T2), Card(P3/T3) = Card(P3/T3) \end{cases}$$

- $Card(P1/T1), Card(P2/T2), Card(P3/T3)$: is respectivly the number of posts/tweets published, liked or commented in the period 1, 2 and 3.
- $Max[Card(P1/T1), Card(P2/T2), Card(P3/T3)$: return the maximum number after the calculation of the cardinality in each period.

After all that, the productivity is equal to 1 if the learner is more productive in the period 1, equal to 2 if he is more productive in the period 2 and equal to 3 if he is more productive in the period 3.

The Motivation. consists in defining the motivation of each learner either he is motivated, demotivated or neutral. For that and after we define the productivity, we collect the posts(from Facebook) and tweets(from Twitter) published in the period of activity(the period in which the learner publish and tweet a lot of content).

The next step, is to classify the retrieved posts and tweets into three classes : **positive**, **negative** or **neutral**. that is to say, each retrieved tweet or post can either express a positive sentiment, negative sentiment or neutral sentiment. Then if the majority class (MC) is the **positive** class, we consider the learner as motivated, else if MC is the negative class, we consider the learner as demotivated, and if MC is the neutral class we consider the learner as neutral.

For making the classification, we based on our recently published article. In which we proposed a new approach based on semantic similarity for classifying a tweet into three classes. The idea of this approach and more details are presented here[7].

These two measures will be used after for clustering learners into clusters. Because the productivity and the motivation have both three possible values(1, 2 and 3 for the productivity; positive, negative and neutral for the motivation), we have 9 different clusters or classes in the form of productivity-motivation (1-positive, 1-negative, 1-neutral, 2-positive, 2-negative, 2-neutral, 3-positive, 3-negative and 3-neutral).

3.2. Collaborative filtering approach (SF)

CF is the second step. In our platform, each learner is presented with his profile of knowledge, and for constructing this profile each learner must complete a quiz which is in the form of multiple choice questions. each course in our system has its equivalent quiz. The questions of the quiz represent the different pedagogical objectives (PO) of the course, that is to say, each PO represent an item in our platform, and doing the quiz is like giving a rate to each PO.

After that the learner completes the quiz, and based on the result obtained, we construct his profile which is in the form of a vector. Each element of the vector represents a rating for a PO. The rating is equal to 1 if the learner answered incorrectly the equivalent question of the PO, and equal to 0 otherwise.

Our assumption is that if the learner did not answer a question correctly, it means he must learn the pedagogical objective related to this question, and we consider his rating as 1. Otherwise, if he answered a question correctly, he is already mastering the related educational objective, so it is not necessary to learn it again, for that we consider his rating as 0.

After we find the K-nearest neighbors(KNN) of the target learner using the social filtering approach(based on productivity and motivation), and after we construct his profile(which is in the form of vector) based on the quiz, we calculate the distance between the profile of the target learner and the profiles of the KNN learners to find the most similar learner(the closet learner). We based here on the assumption that the target learner and his closest learner have similar preferences and can learn similarly.

In our system the profiles of the learners are represented by the following notation :

$Profile = [R(PO_1), R(PO_2), R(PO_3), ..., R(PO_n)]$

- PO: is the pedagogical objective.
- R(PO): is the rating of the PO after doing the quiz (1 or 0).
- n: is the number of PO in a course.

3.3. Scalability and Sparsity

The recommender systems have some challenges related to the problems of **scalability** and **sparsity** or **cold star**, and each RS approach must take into account these two problems(dealing with them) for improving the quality of recommendation.

Our proposed approach can deal with these two problems, which helps us in avoiding any issue that can affect our system.

The first main challenge faced by learner based on our social collaborative filtering approach is the problem of **scalability** as it can be shown when searching for the closest learner if our platform contains a very large number of learners and courses. to avoid this problem we adopt two levels:

- The first level consists in clustering the learners into clusters based on the social content(productivity and motivation) for decreasing the number of learners (conserving only the most similar learners to the target learner). After that we do not need to search for the closest learner into all the learners of our platform, but only into the adequate cluster of the target learner (the cluster that contains similar learners who have similar social content with the target learner).
- The second level consists in parallelizing our work using big data technologies (Hadoop MapReduce and Hadoop Distributed File System).

The second main challenge in our system is related to the **sparsity and cold star** problems. They occur when it is not possible to make a reliable recommendation due to an initial lack of information (similar learners of the target learner). For example, for the first user of our platform, it is impossible to recommend courses based on the social content because there are no learners in the system. or if we do not find any similar learners for the target learner.

The cold star and sparsity problem occur in our approach in three levels. The first is when the first user wants to use our platform to have a recommendation. The second level which we can call complete cold star (CCS) occurs when no KNN learner (of the target learner) is available in our platform (no learner has similar social content with

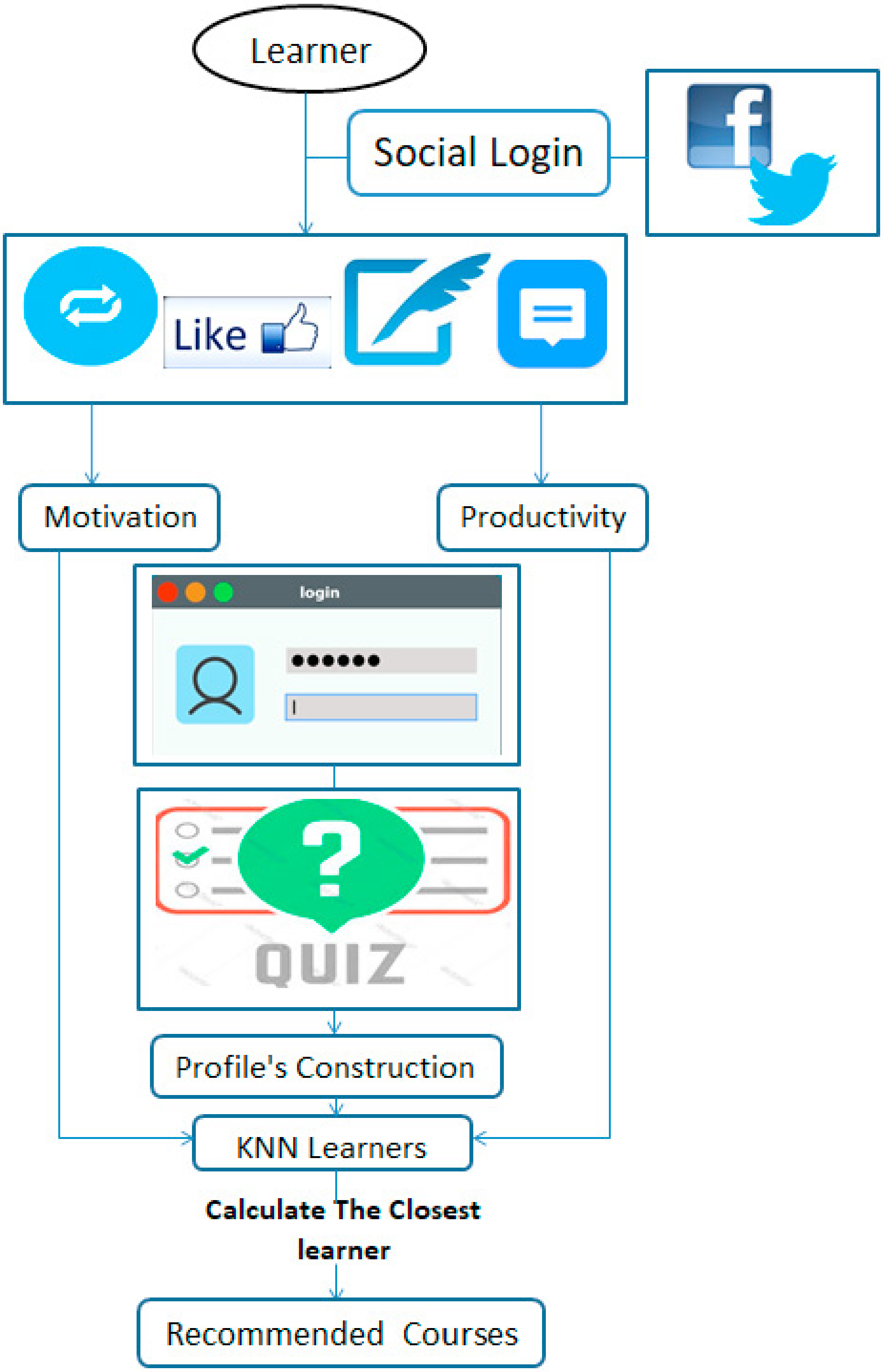

Fig. 1. **Our work Steps**

the target learner). And the third level is when the distance between the profile of the target learner and the profiles of the learners of his cluster is less than a threshold, which means that among the similar learners no one has a similar profile with the target learner.

If we encounter one of these two problems, we will be based only on the profile of learners and the pedagogical objectives for recommending the most suitable PO. The idea is that after the construction of the profile by completing the quiz we construct the profile's vector with a rating for each concept. We represent also each PO in the form of a vector by giving 1 as a rating for the concept that this PO treats and 0 if a concept is not treated by the PO.

After that, we calculate the distance between the learner's profile and the different PO for defining the closest PO for the learner(the PO that must learn).

3.4. Our work steps and Algorithm

As presented earlier, our approach is based on a number of important steps. The first is by using the social network profiles of learners(Social login [8]) for extracting some social factors like tweets, Facebook posts, likes, comments. These factors will be used in the calculation of the productivity and the motivation, which in turn will help us define the K-nearest neighbors. The second step is the construction of learners profiles based on a quiz. And finally, the calculation of the distance between the profile of the target learner and the profiles of the KNN learners for defining the Recommended courses(or Recommended pedagogical objectives).

Figure 1 (the figure above) shows the different steps of our work.

4. Conclusion

In this paper, we have presented a new approach for recommending courses to learners based on their social contents and their profile (knowledge). Our method is a social collaborative filtering approach that uses the social content of learners such as the tweets, facebook posts, likes and comments for grouping them into clusters, and also it is based on the knowledge of learners for finding the K-nearest neighbors learners of the target learner.

Our approach deals with the problems of scalability, sparsity and cold star for improving the quality of recommendation.

Our next work will consist in developing our approach and apply it in different domains like in an e-commerce website or in recommending films. and Also in proposing a new approach for calculating the similarity between the learners' profiles.

References

[1] Y. Madani, J. Bengourram, M. Erritali, B. Hssina, M. Birjali, Adaptive e-learning using Genetic Algorithm and Sentiments Analysis in a Big Data System, INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS 8 (8) (2017) 394–403.

[2] Y. Ar, E. Bostanci, A genetic algorithm solution to the collaborative filtering problem, Expert Systems with Applications 61 (2016) 122–128. doi:10.1016/j.eswa.2016.05.021.

[3] H. Koohi, K. Kiani, A new method to find neighbor users that improves the performance of Collaborative Filtering, Expert Systems with Applications 83 (2017) 30–39. doi:10.1016/j.eswa.2017.04.027.

[4] H. Liu, Z. Hu, A. Mian, H. Tian, X. Zhu, A new user similarity model to improve the accuracy of collaborative filtering, Knowledge-Based Systems 56 (2014) 156–166. doi:10.1016/j.knosys.2013.11.006.
URL http://www.sciencedirect.com/science/article/pii/S0950705113003560

[5] H. Chen, Z. Li, W. Hu, An improved collaborative recommendation algorithm based on optimized user similarity, The Journal of Supercomputing 72 (7) (2016) 2565–2578. doi:10.1007/s11227-015-1518-5.
URL https://link.springer.com/article/10.1007/s11227-015-1518-5

[6] M. Nayebzadeh, A. Moazzam, A. M. Saba, H. Abdolrahimpour, E. Shahab, An Investigation on Social Network Recommender Systems and Collaborative Filtering Techniques 6.

[7] Y. Madani, M. Erritali, J. Bengourram, Sentiment analysis using semantic similarity and Hadoop MapReduce, Knowledge and Information Systems (2018) 1–24doi:10.1007/s10115-018-1212-z.
URL https://link.springer.com/article/10.1007/s10115-018-1212-z

[8] Y. Madani, J. Bengourram, M. Erritali, Social Login and Data Storage in the Big Data File System HDFS, in: Proceedings of the International Conference on Compute and Data Analysis, ICCDA '17, ACM, New York, NY, USA, 2017, pp. 91–97. doi:10.1145/3093241.3093265.
URL http://doi.acm.org/10.1145/3093241.3093265