Personalized Recommendation Of Educational Resources In A MOOC Using A Combination Of Collaborative Filtering And Semantic Content Analysis

Youness Chaabi, Ndeye Massata Ndiaye, Khadija Lekdioui

Abstract: In this article, we present a recommending system of educational resources based on the footprint of learners on a distance-learning platform. Today, all learners follow the same learning path and therefore have access to the same training content, although their information needs are different. The proposed approach offers learners the opportunity to adapt their training path through a mechanism of recommended activities that are most relevant to them. This approach consists in automatically analyzing the interaction footprints of learners. Based on these footprints, our system automatically offers personalized recommendations (courses, course sections, activities, documents, etc.) to learners. The originality of this approach lies in the fact that the learner remains in control of his training itinerary, by increasing the information available to help him or her to build his or her training path. In this article, we present the results of the implementation of our system developed as part of an experiment with 100 students registered on our distance learning platform; as well as the research perspectives.

Index Terms: educational resources, MOOC, recommending algorithm, editorial filtering, collaborative filtering

1 INTRODUCTION

Elearning is a revolutionary way to provide quality and flexible education, compared to traditional (face-to-face) teaching. Today, learners benefit from a variety of online learning programs. However, the wide diversity of learners poses new challenges, such as a single set of learning resources that can be provided to all learners [1], [2]. In fact, learners may have different interests; they may have different levels of expertise and therefore cannot be treated the same way. It is therefore essential for the tutor to assess preferences, level of knowledge and mastery in order to adapt the content of the training. Content personalization can be achieved by using different strategies including feedback by recommending digital resources [3], [4], [5]. In this research, we offer an approach which adapt to each learner by recommending the most relevant educational resources [6]. The list of recommended resources is obtained by using a set of content-based recommendation strategies and collaborative filtering approaches. This article is organized as follows: Section 2 presents a state of the art recommendation system. Section 3 proposes a new approach to recommending educational resources. Section 4 presents a case study illustrating our approach. Finally, Section 5 presents a conclusion and some perspectives.

2 STATE OF ART

Recommendation systems (RS) are tools that provide useful resources for a particular user. Their application has become fundamental in many fields [7], [8], [9], [10]: information research, e-commerce, e-learning, web, and many others.

- *Author name is currently pursuing masters degree program in electric power engineering in University, Country, PH-01123456789. E-mail: author_name@mail.com*
- *Co-Author name is currently pursuing masters degree program in electric power engineering in University, Country, PH-01123456789. E-mail: author_name@mail.com*

In literature, we find different forms of recommendation depending on the objective and the information available. SRs are generally classified into three categories [11]: editorial filtering, collaborative filtering and content-based filtering. In this section we review the main existing approaches.

2.1 Editorial filtering

This type of recommendation is adopted when the system has no information about the learner. The objective of this approach is to attract the attention of the novice learner in order to give him/her the desire to complete part of the course. It presenting the most popular courses, documents, educational activities as well as new features. Most online retail sites use this approach.

2.2 Content-based filtering

This type of filtering [12] takes into account the content of an educational resource (document, description of a course, video, etc.) as well as the learners' needs in terms of information. This technique is based on the recommendation of items similar to those preferred by the learner in the past [13], [14]. For example: when a learner tends to consult courses in the field of artificial intelligence often, the system will propose recommendations related to artificial intelligence. Indeed, these recommended courses have common keywords such as: "Machine learning", "Deep learning" or "multi-agent system", etc. Content-based filtering techniques provide learners with educational resources through automatic data classification approaches. Classification approaches associate each learner with a classifier representing his or her profile. Classifying compares the new items with the learner's profile and recommends those that are closest [15]. Several classification techniques have been adopted by content-based recommendation systems, the most commonly used of which are: neural networks, decision trees, Bayesian networks [16], [17].

2.3 Collaborative filtering

Collaborative filtering is based on the hypothesis: word of mouth. For example, learners, who want to take a training course or read a document, ask their friends for their opinions. Thus, in collaborative filtering, the selection of courses to be offered to a learner no longer depends on content (content-based filtering), but on opinions and assessments made by learners with similar preferences. Thus, if two learners A and B have evaluated a number of courses in a similar way, chances are that A likes what B likes, and vice versa. So the courses that A liked can be recommended to B and vice versa. This approach solves an important problem faced by the content-based filtering approach, namely the processing of multimedia content [15], [18]. The general steps of a collaborative filtering system are as follows [15], [19], [20].

1. Propose to learner a set of educational content to evaluate them;
2. Construct the learner's profile based on his or her assessments;
3. Use this profile to help the learner in his or her future information searches.

3 PROPOSED APPROACH

Traditional learning environments offer the same content to learners, although their information needs are different. Once a learning system is developed, it will be difficult, if not impossible, to adapt it to his or her needs and preferences. The originality of this approach is that the learner remains in control of his or her training path; the recommendation mechanism increases the information at his or her disposal to help him or her build this path. This recommendation mechanism will take into account learners' preferences in terms of learning style, level of knowledge and traces of interaction [5]. To improve the quality of recommendations, taking into account the content of the items, a content-based filtering algorithm is proposed and merged with the collaborative filtering algorithm to predict missing evaluations of the -item matrix.

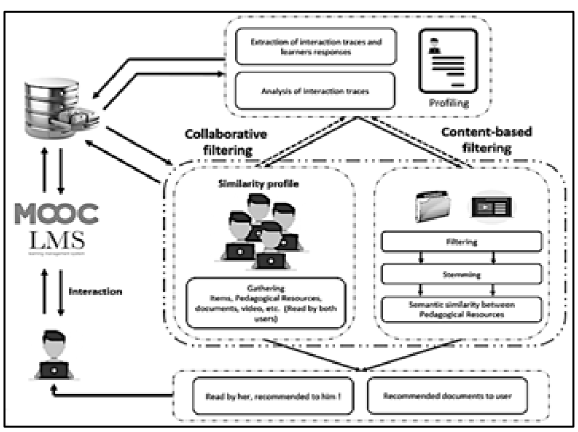

Fig. 1: *General system architecture*

The general principle of our approach is illustrated in Fig. 1: Each learner is described by information that characterizes his or her profile and is linked by friendships with other learners of the MOOC platform. This approach is based on four types of information: (1) learner profile information that describes knowledge, preferences, experiences, etc. (2) information that describes the learning styles of each learner. (3) Information describing the actions carried out by learners on educational resources (visits, evaluations, etc.). (4) Information describing the results of exercises or quizzes performed by learners.

3.1 Modeling of educational resources

Learning resource is described by a set of metadata: title, description, author, language, format, nature, creation date, modification date, knowledge domain, keywords and difficulty. The latter is used in the filtering of resources according to the learner's preferences related to that resource. In this way, resources will be filtered and ordered according to their level of difficulty.

3.2 Generation of the learner's profile

All referral systems require certain types of information about learners to generate recommendations for them. The types of information that must be present in a profile are strongly related to the scope of application of the recommendation systems. In our context, the preferences, learning style and level of knowledge of learners are of crucial importance. In this approach, the learner must answer a questionnaire to choose his or her area of interest and preferred content model. He takes two tests to identify his learning style [21], and a knowledge test consisting of pre-tests of different levels (low, medium, high) to assess his level.

3.2.1 Identification of knowledge level

In the scenario of an online course, the knowledge test is a common denominator of the course structure. Generally speaking, an online course is composed of 3 parts: (1) the entry system: it allows to manage the flow of learners at the entrance of the course. The general objectives of the course are presented in this section and an entry test is proposed. In MOOCs, a video is proposed. (2) The learning system: it includes all the contents and learning activities of the course. (3) The exit system: it consists of a summative evaluation to certify the learner's achievements. The learning system is composed of learning units created according to the objective of the course. In each learning unit is inserted a knowledge test that allows the learner to test his knowledge. The test consists of a set of multiple-choice questions (MCQ) and/or single-choice questions (QCU). A MCQ or QCU is often considered to be very guided, but it responds well to the technical constraint related to multimedia: (i) distance learning devices generally include a QCM/QCU design tool; (ii) correct answers are implemented in advance in the online training platform and (iii) feedback is automatic.

3.2.3 Identification of learning style

Identifying the learning style of learners allows for a variety of learning strategies to meet their needs and optimize their learning. A learner's learning style is the method by which he or she learns best, but it may also present a combination of the other styles, depending on the situation. Physiological, emotional, sociological and cognitive behaviours are the indicators that identify different learning styles [22]. Learning styles are often classified as follows: adapter, diverger, assimilator and converger (Table 1).

Table 1: *Recommendations for each learning style*

Learning styles	*Description*	*Recommendations*
Adapter	*Training should include project implementation activities, collaborative peer spaces.*	*- Practical courses for the execution and realization of projects.* *- Discussion space for collaboration and exchange.*

Diverger	*A training course should include discussion and exchange periods between tutors and learners, in order to illustrate certain concepts to learners and allow them to exchange ideas.*	*- Discussion forum for and exchange of ideas between tutors and learners.*
Assimilator	*Training should include theoretical concepts such as case studies.*	*- Courses include theoretical concepts such as case studies.*
Converger	*Training should allow problems to be solved. It should include practice-oriented activities (demonstration, case studies, etc.).*	*- Training to solve problems. - Practical activities (demonstration, case studies, etc.).*

Based on the learning style model [23], we designed a questionnaire. This questionnaire is composed of 24 questions [23]. The estimated time to complete this questionnaire was approximately five minutes. Learners reported no difficulties with the questions asked in the questionnaire. The learning environment retrieves answers and test results from learners and updates their profiles. Learner interaction traces and social data are also used to update the profile.

3.3 Similarity of profiles

In the context of distance learning, learners are interested in obtaining appropriate recommendations for their educational needs. They may be interested in the most popular resources recently visited by their similar learners. A learner's neighborhood is represented by the learners who are closest to him or her, considering the information stored in each learner's profile: preferences, interests, learning styles, level of knowledge and traces of interaction. Indeed, it is enough to know the proximity of users according to their profile. We found that an effective recommendation system must have the functionality:

1. Automatic group construction and assignment of learners to different groups based on their profiles.
2. The search for relevant neighbors of the active learner.
3. Recommendation of educational resources appreciated by learners in the same groups.

Learner similarity calculation plays a very important role in neighborhood-based recommendation methods [24] because it identifies trusted neighbors. In addition, similarity has a significant impact on the accuracy and performance of recommendation systems [7]. To calculate the neighborhood V between two learners x and y, we use the Euclidean distance. Let (x1; x2; ...; xn) and (y1; y2; ...; yn) be the vector representations relating to the profiles of two learners x and y. The Euclidean distance between the two profiles is calculated as follows:

$$V(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)} = \sqrt{\sum_{i=1}^{n}(s_i)}$$

With s_i d {0, 1} such as $s_i = 0$ if $x_i = y_i$ and $s_i = 1$ if $x_i \neq y_i$

In our context, we consider only three attributes (n=3), which are the knowledge test, learning style and preferences. The referral system searches for profiles similar to the current learner, and then provides a list of useful resources liked and appreciated by the same learners in the group. The list of recommended resources r for the learner p is determined by the following formula [25]:

$$= \sum_{v \in Sim[App]}^{RV(p,r)} Ress^{-\alpha t(v,r)}$$

With:

- App: the learners of the system.
- RV (p, r) resources r which are visualized by a learner p. The RV function is equal to 1 if the learner p has consulted the resource r and 0 if not.
- Sim[App]: all learners who have a profile similar to the learner p.
- t(v,r) : the number of days since the date of the last consultation of the resource r by the learner p.
- α As a factor of decrease. The more recently a learner has consulted a resource, the higher the resource's score increases.

A learner may also be recommended a list of resources voted and evaluated by his or her learners in the group that are useful in relation to the learner's learning areas. The list of resources r recommended to the learner p is determined by the formula: either P is the set of all learners, R is the set of possible educational resources that can be recommended, and u is a function that measures the utility of a resource r to the learner p, i.e.

$$u: R \times R \rightarrow \mathbb{R}$$

So for each learner p ∈ P, we want to choose the resource r′ ∈ R that maximizes the usefulness to the learner:

$$\forall r \in R, i_r' = argmax_{r \in R}\, u(p,r)$$

3.4 Content-based recommendation

The content-based recommendation consists in analyzing the content of the items that are candidates for recommendation or the descriptions of these items. In this phase our system recovers the learner's level of knowledge and according to this level, the system measures the semantic similarity between the preferences, content and descriptions of the resources already preferred by this learner in the past with the resources present on the platform. Semantic similarity or semantic relationship is a concept of measuring the proximity between terms or documents in the context of their meaning. We have two different methods for calculating semantic similarity. One is to define a topological similarity, using ontology to define a distance between words. The other is based on the use of statistical means such as the vector space model to correlate words and text contexts from an appropriate corpus of text. We focus on the first approach using WordNet ontology for semantic similarity calculation [26]. The calculation of similarity in this approach is based on the fact that similarity depends on the common and distinct characteristics of the objects. WordNet is a lexical ontology. It is a semantic network developed by Princeton University that models lexical knowledge in a taxonomic hierarchy. WordNet contains three databases: one for nouns, one for verbs and one for adverbs and adjectives [27]. Semantic similarity in WordNet can be calculated by the method: the path length. This method calculates the number of nodes or relationships between nodes in the taxonomy. The advantage of this method is that it is not dependent on the static distribution of the corpus or the distribution of words. We use WordNet 2.1 which contains nine distinct name hierarchies where sometimes the path between two concepts may not exist (see Fig. 2). Therefore, we create a root node ("Entity" see Fig. 2) that includes all nine hierarchies given in WordNet.

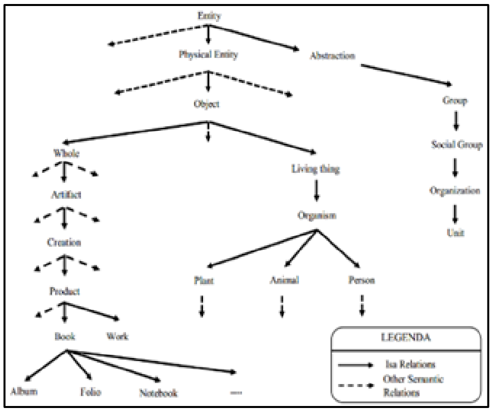

Fig. 2 : *Extract from the nominal hierarchy of WordNet*

The semantic similarity calculation process consists of three phases (see Fig. 3):

- Phase 1: Temporary construction module.
- Phase 2: Semantic calculation module.
- Phase 3: Semantic similarity measurement procedures.
-

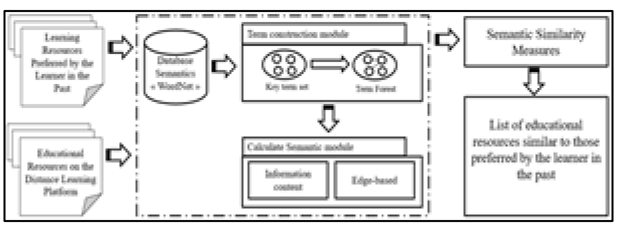

Fig. 3: *Semantic similarity calculation diagram*

Phase 1: Temporary building module
The objective of the module is to select all the words in the text that exist on WordNet and to obtain the relationship between these words. We use WordNet to generate a richer text representation. In this module, we used the hyperonyms provided by WordNet as useful features for text analysis.

Phase 2: Semantic calculation module
We use the path length algorithm that uses the semantic similarity measure to find the appropriate meanings of words according to the context at the sentence or text scale. When concepts are organized in a hierarchy, it is appropriate to measure similarity based on structural measures that find path lengths between concepts. [28], developed a measure based on the length of paths between concepts in the WordNet hierarchy. The measurement of the shortest path emphasizes the proximity of two concepts in the hierarchy.

In a thesaurus hierarchy graph, the shorter the path between two words, the more similar these words are:

- The words are quite similar to the parents;
- Words are less similar to words far away in the network
- Pathlen (c1, c2) = number of edges of the shortest path
- Path-based similarity often involves a logarithmic transformation
- The similarity based on the length of the path is (1):

$$simpath(c1,c2) = -\log pathlen(c1,c2)$$

Phase 3: Procedures for measuring semantic similarity
Semantic vectors for T1 and T2 can be formed from T and corpus statistics. The process of deriving semantic vectors for T1 (13):

$$Word\ w, define$$
$$sim(w_1,w_2) = max_{c1,c2}[sim(c1,c2)]$$
$$sim(T1,T2) = \sum_{i=1}^{n} \frac{sim(Wi,Wi+1)}{n}$$

4 EXPERIMENTATION AND RESULTS

As announced in the introduction, the recommendation system we propose is based on the analysis of learning footprint. The characteristics of the target population are: Learners enrolled in a master's degree in computer science:

Number: 100 learners
Number of master courses: 24 courses
Course: Master 1 and Master 2

To test the recommendation system, all learning footprints concerning knowledge tests, preferences and learning styles were extracted. With this information, we created a database and developed an application (Fig. 1) to manage course recommendations. As a reminder, the recommendations are online courses from the MOOC platform. In order to evaluate our approach, we adopted as a first step, a qualitative evaluation where 100 learners were selected and more than 24 educational resources. These resources are linked to different areas of computer knowledge. We developed an evaluation questionnaire with about ten questions using Likert's score [29] on a five-point scale (A, Very satisfied; B, Somewhat satisfied; C, Neither satisfied nor dissatisfied; D, Somewhat dissatisfied; E, Very dissatisfied) [29]. The questions are classified into two categories: effectiveness, in terms of the relevance of the recommendation results; and usefulness, in terms of learning satisfaction and intention to reuse the suggested resources.

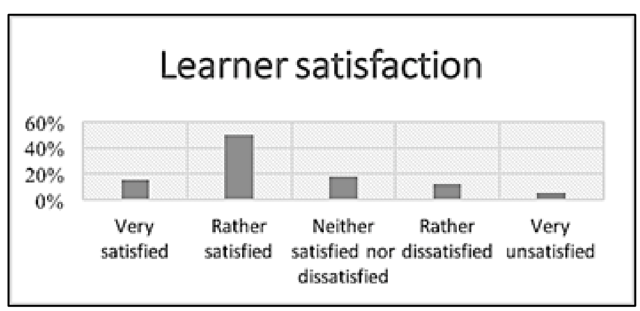

Fig. 2: *Scale Questionnaire Score*

A qualitative evaluation of our approach shows its interest for learners. However, and in order to test our approach deeper, experimentation in real training situations is necessary, by asking learners to integrate this approach into their daily practices, in order to better judge the advantages and limitations of the recommendation system.

5 CONCLUSION

In this article, we propose a personalized recommendation system combining learner profiling, knowledge assessment, learning style and trace analysis techniques to improve the learning process by recommending educational resources (courses, course sequences, videos, documents, etc.). We offer a recommendation system based on collaborative filtering and content-based filtering that automatically provides customized recommendations for each learner. The approach was then tested on a real situation, which showed its effectiveness and usefulness. In view of the development of this project, on one hand, we plan to recommend working groups. On the other, to improve our recommendation system by a hierarchical classification method to classify educational resources (documents, courses, section, video, etc.) to model learning needs.

6 REFERENCES

[1] Coates, H., James, R., & Baldwin, G. (2005). A critical examination of the effects of learning management systems on university teaching and learning. Tertiary education and management, 11, 19-36.
[2] Downes, S. (2012). Learning objects: Resources for learning worldwide. In Online education using learning objects (pp. 47-57). Routledge.
[3] Bulut, B., Kaya, B., Alhajj, R., & Kaya, M. (2018, August). A Paper Recommendation System Based on User's Research Interests. In 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 911-915). IEEE.
[4] Wan, S., & Niu, Z. (2018). An e-learning recommendation approach based on the self-organization of learning resource. Knowledge-Based Systems, 160, 71-87.
[5] Ndiyae, N. M., Chaabi, Y., Lekdioui, K., & Lishou, C. (2019, March). Recommending system for digital educational resources based on learning analysis. In Proceedings of the New Challenges in Data Sciences: Acts of the Second Conference of the Moroccan Classification Society (p. 15). ACM.
[6] Drachsler, H., Verbert, K., Santos, O. C., & Manouselis, N. (2015). Panorama of recommender systems to support learning. In Recommender systems handbook (pp. 421-451). Springer, Boston, MA.
[7] Du, Y., Ranwez, S., Sutton-Charani, N., & Ranwez, V. (2019, July). Apports des ontologies aux systèmes de recommandation: état de l'art et perspectives.
[8] Baidada, M., Mansouri, K., & Poirier, F. (2019, June). Personalized E-learning Recommender System to Adjust Learners' Level. In EdMedia+ Innovate Learning (pp. 1353-1357). Association for the Advancement of Computing in Education (AACE).
[9] Ghézala, H., Kboubi, F., & Gorrab, A. (2019). Social Information Retrieval and Recommendation: state-of-the-art and future research. Revue Africaine de la Recherche en Informatique et Mathématiques Appliquées, 27.
[10] Lahlou, F. Z., Benbrahim, H., & Kassou, I. (2017). Context Aware Recommender System Algorithms: State of the Art and Focus on Factorization Based Methods. Electronic Journal of Information Technology.
[11] Negre, E. (2018). Les systèmes de recommandation: une catégorisation.
[12] Malone, T., Grant, K., Turbak, F., Brobst, S., et Cohen, M. (1987). Intelligent informationsharing systems. Communications of the ACM 30, 5, p. 390–402.
[13] Pazzani, M. et Billsus, D. (2007). The Adaptive Web, volume 4321/2007 de Lecture Notes in Computer Science, chapitre Content-Based Recommendation.
[14] Malone, T., Grant, K., Turbak, F., Brobst, S., et Cohen, M. (1987). Intelligent informationsharing systems. Communications of the ACM 30, 5, p. 390–402.
[15] Adomavicius, Gediminas, et Alexander Tuzhilin. (2005). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. Transactions on Knowledge and Data Engineering. 7(6), p. 734-749.
[16] Cohen, W.W. (1996). Learning rules that classify e_mail/ AAAI Spring Symposium on Machine Learning in information Access.
[17] Mladenic, D. (1999). Text-learning and related intelligent agents: A survey. IEEE Intelligent Systems, 14(4):44–54.
[18] Sarwar, BadruJ, George Karypis, Joseph Konstan et T. John Riedl. (2000). Analysis of recommendation algorithms for e-commerce. In Proceedings of the 2nd ACM conference on Electronic commerce: ACM.
[19] Berrut, C., et Nathalie D. (2003). Filtrage collaboratif. In Assistance intelligente à la recherche d'informations, Hermes Science Publications, p. 30.
[20] Zaier, Z. (2010). Modèle multi-agents pour le filtrage collaboratif de l'information. Thèse de

doctorat de l'université de Québec, Montréal, Canada.
[21] Vizeshfar, F., & Torabizadeh, C. (2018). The effect of teaching based on dominant learning style on nursing students' academic achievement. Nurse education in practice, 28, 103-108.
[22] O'Keefe, J. (1979). A review of the hippocampal place cells. Progress in neurobiology, 13(4), 419-439.
[23] Felder, R. M. (1996). Matters of style. ASEE prism, 6(4), 18-23.
[24] Chekkai, N., Chikhi, S., & Kheddouci, H. (2011). Nouvelle Approche à base de Graphes pour les Systèmes de Recommandation Collaboratifs. In CIIA.
[25] Guy, I., Zwerdling, N., Ronen, I., Carmel, D., & Uziel, E. (2010, July). Social media recommendation based on people and tags. In Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval (pp. 194-201). ACM.
[26] Mestrovic, A., & Calì, A. (2016, September). An ontology-based approach to information retrieval. In Semanitic Keyword-based Search on Structured Data Sources (pp. 150-156). Springer, Cham.
[27] Sudha, S. (2016). WordNet–An On-line Lexical Database.
[28] Rada, R., Mili, H., Bicknell, E., & Blettner, M. (1989). Development and application of a metric on semantic nets. IEEE transactions on systems, man, and cybernetics, 19(1), 17-30.
[29] Boone, H. N., & Boone, D. A. (2012). Analyzing likert data. Journal of extension, 50(2), 1-5.