Recommender System to Improve Knowledge Sharing in Massive Open Online Courses

Sarra Bouzayane[1] and Inès Saad[1,2]
[1]MIS Laboratory, University of Picardie Jules Verne, 33 rue Saint-Leu, 80039 Amiens, France
[2]Amiens Business School, 18 Place Saint-Michel, 80038 Amiens, France

Keywords: Recommender System, Knowledge Sharing, Learning Process, Leader Learner, MOOC, Periodic Incremental Prediction.

Abstract: This paper focuses on the support process, within a Massive Open Online Course (MOOC), that is currently unsatisfactory because of the very limited size of the pedagogical team compared to the massive number of the enrolled learners who need support. Indeed, many of the MOOC learners can not appropriate the information they receive and must therefore be assisted in order to not abandon the course. Thus, to help these learners take advantage of the course they follow, we propose a tool to recommend to each of them an ordered list of "Leader learners" who are able to support him throughout his navigation in the MOOC environment. The recommendation phase is based on a multicriteria decision making approach to weekly predict the set of "Leader learners". Moreover, since the MOOC learners' profiles are very heterogeneous, we recommend to each of them the leaders who are most appropriate to his profile in order to ensure a good understanding between them. The recommendation we propose is validated on real data coming from a French MOOC and has proved satisfactory results.

1 INTRODUCTION

In this paper, we discuss the learning process when the information exchange between the actors takes place exclusively online. So, we deal with the case of MOOCs (Massive Open Online Courses) which are virtual learning environments offering online courses for free and encourage the active participation of learners who must enrich the system with a digital information and then conduct a rigorous research to find the information they need. The MOOCs are accessible by a massive number of learners coming from many cultures across the globe and so characterized by very heterogeneous profiles.

A huge amount of data, in different formats (pdf, image, video), is deposited on the MOOC system either voluntarily by the learners or mandatory by the pedagogical team. The learners have to consult these data in order to interpret them in information, and then to absorb this information in order to infer their own knowledge. Finally, during their learning process, the learners can periodically answer the activities proposed by the pedagogical team, such as the automated tests and the peer assessment, using the inferred knowledge. Every week, the course materials are updated, new activities are proposed and a new forum space is created (cf. Figure 1).

Since 2008, the number of MOOCs has rapidly grown around the world (Patru and Balaji, 2016). However, despite their proliferation, the MOOCs still suffer from a high dropout rate that usually reaches 90% (Yang et al., 2014), especially because of the lack of interaction with the MOOCs instructor and the difficulty of the courses content (Hone and El Said, 2016). In fact, a MOOC is led by a small pedagogical team that is generally unable to support all of the learners. This makes it difficult to the learners to properly absorb the information they receive and they usually refer to the data exchanged via the forum whose accuracy and relevance are not always guaranteed. According to Onah et al. (2014), the excessive dropout rate of learners is one of the major recurring issues in the MOOCs.

Hence, our objective in this work is to identify, among this massive number of learners, the leader ones, so those who are able to share a correct and an immediate information with any learner in need.

To do so, we propose an approach to recommend a personalized list of "Leader learners" for each MOOC learner in need, taking into account their demographic

Bouzayane, S. and Saad, I.
Recommender System to Improve Knowledge Sharing in Massive Open Online Courses.
DOI: 10.5220/0006929900530062
In *Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2018) - Volume 2: KEOD*, pages 53-62
ISBN: 978-989-758-330-8
Copyright © 2018 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved

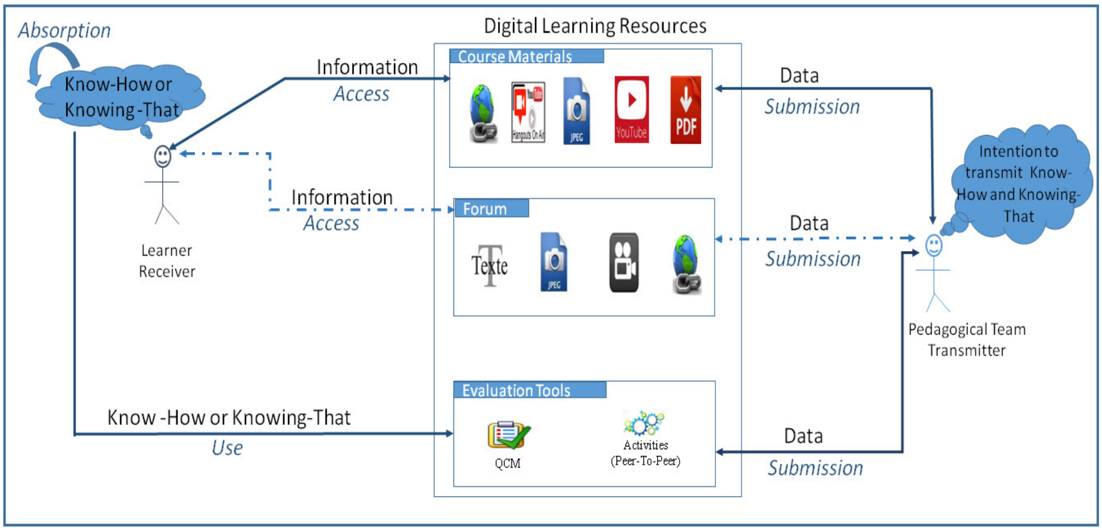

Figure 1: The Digital information system of the MOOC.

data. Thus, in order to categorize the learners, we propose a prediction periodic approach that is based on the DRSA (Dominance-based Rough Set Approach) (Greco et al., 2001) and aims to weekly predict the three preference ordered decision classes: Cl_1 of the "At-risk learners", Cl_2 of the "Struggling learners" and Cl_3 of the "Leader learners". This approach takes into account the preferences of the pedagogical team of the MOOC and the periodic variation of the learners' behaviour. Then, a personalized list of the predicted "Leader learners" must be recommended for each "At-risk learner" or "Struggling learner" according to her/his profile. The recommendation is based on the demographic filtering and must improve the information exchange between the "Leader learners" and the "At-risk learners" or the "Struggling learners" to help them properly understand their knowledge. According to Eastmond (1994), the communication in a context of online learning allows much more natural data exchanges without the risk that a too long waiting time between the question and the answer can lead to a disinvestment of the learner in relation to the task that is proposed to her/him. It promotes the mutual exchange of information between learners, which makes it possible to reach a mutual understanding for each information exchanged.

The remainder of this paper is organized as follows: section 2 discusses the related work. Section 3 presents the "Leader learners" recommendation process and details the periodic prediction approach based on DRSA. Section 4 is dedicated to the experiments analysis. Section 5 concludes the work and advances some prospects.

2 RELATED WORK

Thanks to their frequent use, the definitions given to a recommender system are numerous, including that of Burke (2002) defining it as a software tool that has the effect of guiding users in a personalized way to interesting or useful objects in a large space of possible options.

Today, the recommender systems are embedded in different areas that are characterized by a growing amount of data that needs to be filtered, such as the e-health field (Hoens et al., 2013; Duan et al., 2011; Sartori et al., 2018), the finance field (Kim and Ahn, 2008; Wu et al., 2015) and the e-learning (Aher and Lobo, 2013; Dascalu et al., 2015).

For the same objective of improving the knowledge sharing between the knowledge seekers and the contributors, Sartori et al. (2018) proposed a recommender system in order to deal with the complex decision-making process within the virtual community of practice. This system, called Knowledge Acquisition Framework based on Knowledge Artifact (KAFKA), is based on two types of users: the KA-Developer who is a knowledge contributor in the KAFKA and the KA-User who is a knowledge seeker. The KAFKA focuses on three types of knowledge that are the functional knowledge modeled using ontology, the procedural knowledge modeled by the Bayesian networks and the experiential knowledge that is captured by production rules. Each directed link in the Bayesian networks is associated to one or more production rules. Applied in the physical activity (PA) context, this tool permits to provide the knowledge seeker with the suggestion to increase, decrease or maintain the PA plan for the next week depends on the current week self-efficacy and MET (metabolic equivalent of task) value.

However, in the context of MOOCs there are not so many recommender systems proposed to support the learners. But, we can classify those that exist into two categories according to their purpose: the recommender systems whose objective is to help learners choose an appropriate MOOC in order to address the massive number of MOOCs (Bousbahi and Chorfi, 2015; Symeonidis and Malakoudis, 2016; Gutiérrez-Rojas et al., 2014), and the recommender systems whose objective is to help learners understand the information they receive to remedy the massification of data exchanged between the different actors of the same MOOC. In this paper, we focus on the second category.

Authors in (Yang et al., 2014) proposed a recommender model to provide each learner with a personalized list of discussions that satisfies his preferences. The recommendation requires three modelling: the modelling of the forum discussions content based on the words analysis, the modelling of the learner's preferences that are extracted from his discussion history and the modelling of the learner's social interactions with the pairs. These three models are given in input to an adaptive factorization matrix in order to predict the behaviour of the learners in the next window of time based on their behaviour during the current one. Hence, both collaborative and content-based filtering techniques are applied in order to predict the discussions that may be of interest to the learner. Experiments have shown that such a model has improved the performance of learners especially when the window of time chosen is reduced.

Authors in (Li and Mitros, 2015) proposed a recommender system to provide the learners with a rehabilitation resources list that is relevant for a given problem. This recommendation has to be more depth and less scaffolding than the forums interventions. The system is based on the crowdsourcing technique that requires a combination of what the expert learners have published to solve a problem and what the novice ones need. For each set of problems is assigned a theme and for each resource is granted a title, a link, a summary, a screenshot and a list of votes. Once recommended, the resources can be voted by the learners. Moreover, the MOOC pedagogical team can modify, delete or enrich the resources. This solution is considered more economical and more practical for creating rehabilitation tools.

Authors in (Onah and Sinclair, 2015) proposed an algorithm based on the collaborative filtering technique to recommend to a target learner the resources that are appropriate to his profile. Each learner in the system is asked to rate each used resource according to a scale from 1 to 5. Based on this rating, a prediction function is calculated to predict the degree of appreciation of the target learner to each resource.

Finally, the authors in (Labarthe et al., 2016a) proposed an integrated recommender module to provide each learner with a list of relevant learners who are available and ready to share their knowledge with the other learners. A target learner can send a private message or open a chat window with the recommended learner. She/He can also signal her/him as favoured or ignored. Unlike a favoured learner, an ignored learner will no longer be recommended to the concerned target learner. Learners are recommended based on their profiles and activities. The recommendation experiments showed a positive effect on the learners' participation level and the completeness of the MOOC (Labarthe et al., 2016b).

These presented works have two major limitations. On the one hand, the recommendation is specially designed for the forum participants and based on the information they share, whereas these participants represent only a limited minority between 5% and 10% of the MOOC learners (Kloft et al., 2014). On the other hand, the proposed recommendation is limited to providing the learners with the pedagogical resources that are appropriate to their profiles without ensuring that the recommended resource is correctly interpreted by the receiving learner.

3 *KTI-MOOC*: A RECOMMENDER SYSTEM FOR THE KNOWLEDGE TRANSFER IMPROVEMENT WITHIN A MOOC

In this section, we start by briefly explaining the periodic incremental prediction approach based on the DRSA and proposed to weekly categorize the MOOC learners. Then, we present the recommendation process based on the demographic filtering in order to recommend to each learner in need a list of "Leader learners" appropriate to his profile.

3.1 *MAI2P* : Multicriteria Approach for the Incremental Periodic Prediction of the "Leader learners"

This phase aims to categorize the MOOC learners during the following week of the MOOC based on their static and dynamic data of the current one. It is based on the Dominance-based Rough Set Approach (DRSA) (Greco et al., 2001) that is a supervised learning technique. DRSA relies on the preferences of the human decision makers in order to infer a set of "*If ... Then ...*" decision rules. According to our objective in this context, three categories (also called decision classes) of learners are defined:

- Cl_1. The decision class of the "At-risk learners" corresponding to learners who are likely to dropout the course in the next week of the MOOC.
- Cl_2. The decision class of the "Struggling learners" corresponding to learners who have some difficulties but still active on the MOOC environment and don't have the intention to leave it at least in the next week of the MOOC.
- Cl_3. The decision class of the "Leader learners" corresponding to learners who are able to lead a team of learners by providing them with an accurate and an immediate response.

These three decision classes are increasingly preference ordered such that learners belonging to the decision class Cl_3 are more preferred than those belonging to Cl_2 and these are more preferred than those belonging to Cl_1.

This phase is composed of three steps such that the first and the second steps are performed at the end of each week W_i of the MOOC while the third step runs at the beginning of each week W_{i+1} of the same MOOC such that $i \in \{1..t-1\}$, where t is the MOOC duration in weeks.

- ***Step 1: Construction of a Decision Table.*** This step is based on three sub-steps: The first concerns the construction of a family F of p criteria to characterize a learner's profile (for example the study level, the motivation to participate in MOOC, the score, etc.). For each criterion $g_k \in F$ a preference ordered scale is fixed according to the personal viewpoint of the decision maker who is the pedagogical team in our case (Roy and Mousseau, 1996). For example, for the criterion "Study level", the preferences applied are: 1: Scholar student; 2: High school student; 3: PhD student; 4: Doctor. This sub-step is detailed in (Bouzayane and Saad, 2017a). The second sub-step is the construction of a training sample of learners L_i, containing a set of m reference learners. This sample must be representative for each of the three predefined decision classes. Third, each learner $L_{i,j} \in L_i$, such that $j \in \{1..m\}$ and $i \in \{1..t\}$, must be evaluated on each criterion in F according to the predefined preference scale. Each evaluation vector must allow the pedagogical team to classify the learner in one of the three decision classes Cl_1, Cl_2 or Cl_3. The different sub-steps form a matrix, called decision table, whose the lines are the learners belonging to L_i, the columns are the criteria belonging to F, the content is the evaluation function $f(L_{i,j}, g_k)$ representing the assessment values of each learner $L_{i,j} \in L_i$ on each criterion $g_k \in F$ and the last column is the assignment of each learner in one of the three predefined decision classes Cl_i (see Table 1).

Table 1: Example of a decision table.

	g_1	...	g_k	...	g_p	**D**
L_1	$f(L_1, g_1)$	...	$f(L_1, g_k)$	...	$f(L_1, g_p)$	Cl_i
L_2	$f(L_2, g_1)$	...	$f(L_2, g_k)$	...	$f(L_2, g_p)$	Cl_i
...	...	...	...	...	...	...
L_m	$f(L_m, g_1)$	...	$f(L_m, g_k)$	...	$f(L_m, g_p)$	Cl_i

- ***Step 2: The Periodic Inference of a Set of Decision Rules.*** The open enrolment and the absence of a serious commitment when participating in the MOOC lead to the free entry/exit of learners during its broadcast. That is why the learning set L_i built in the first step can not be stable from one week to another. It is therefore necessary to select for each week W_i a learning set L'_i to be added to the learning set L_{i-1} such that $L_i = L_{i-1} + L'_i$. Thus, to deal with the instability of the learning set of a MOOC, we apply our incremental learning algorithm *DRSA-Incremental* (Bouzayne and Saad, 2017) that enhances the DRSA in order to periodically updates the decision rules so as to keep an up to date categorization. Each week, this algorithm takes in input the constructed decision table of the first step to infer a coherent set of decision rules using a dominance relation.
- ***Step 3: The Classification of the Potential Learners.*** This step consists in using the inferred decision rules in order to classify the potential learners at the beginning of the week W_{i+1}. We mean by "potential learners" those who are likely to be classified into one of the three decision classes.

This phase permits to categorize the MOOC learners at the beginning of each week. It is detailed in (Bouzayane and Saad, 2017b). However, the main contribution of this paper is to explain how we exploited this early and periodic categorization of learners in order to help the MOOC participants. This aid should permit to strengthen the pedagogical team by identifying learners who are able to accompany and to guide the other learners who are in need. In this case, the role of the pedagogical team may be limited to providing the course material (videos, pdf, etc.) without having to frequently intervene in the forums or to manage all the learners. Moreover, this aid aims to help the learners who are in need to obtain the appropriate and the immdiate information they seek in order to improve their learning process.

3.2 Recommender System based on the Periodic Prediction of the "Leader learners"

The objective of the recommender system KTI-MOOC is to provide each "At-risk learner" or "Struggling learner" with a personalized list of "Leader learners" who are able to support her/him during her/his participation in the MOOC. To this end, we used the demographic filtering that categorizes the users depending on their demographic data (gender, age, country, education level, etc.). This filtering assumes that two users having evolved in the same environment are more likely to have the same taste and to share the same preferences. Hence, the system must recommend to the target user ("At-risk learner" or "Struggling learner") the items ("Leader learners") appreciated by his neighbours. The recommendation process is based on three steps: first, the learner's profile modelling; second, the learner's neighbourhood identification and finally, the recommendation list prediction. It is triggered upon the connection of a learner on his personal MOOC page (cf. Figure 2).

3.2.1 Learner's Profile Modelling

The user's profile modelling is based on two key concepts that are the representation model and the information to consider. In this work, we adopt the vector representation. Moreover, the information to be included in the representation model must satisfy the purpose of the recommendation, which is the mutual understanding between the information transmitter (the "Leader learner") and the information receiver (the target learner). In other words, this information must represent the factors inhibiting the process of knowledge transfer such as the language, the field of study and the geographical distance.

- The language: the linguistic distance between the transmitter and the receiver of information has been proven by several research works as a powerful obstacle to the process of knowledge transfer. In the knowledge management field, Welch and Welch (2008) proved that the linguistic distance impacts, both the ability to transfer the knowledge by the transmitter and also the ability to absorb it by the receiver. In the context of MOOC, Barak (2015) has proved that the language barriers affect the understanding of the course content and promote the early MOOC-leaving.
- The field of study: the shared field of study allows the actors to have a similar scientific and technical language. According to Grundstein (2009), people who share the same culture may have similar patterns of interpretation that allow them to give the same meaning to a codified knowledge. Also, Gooderham (2007) highlighted the impact of the cultural distance on the quality of the knowledge transfer process. He proved that people from the same culture can understand each other better than the other people.
- The geographical distance: compared to the face-to-face interaction, the remote one puts a lot of disadvantages especially when it concerns the know-how transfer. According to Gooderham (2007), the geographic distance is an inhibitor of the knowledge transfer process. Ambos and Ambos (2009) found that the smaller the distance between the transmitter and the receiver, the higher the efficiency of knowledge transfer is.

This information is entered manually by the learner upon the registration. It is used to calculate the similarity between the "Leader learner" and the target learners. Our purpose is to minimize the linguistic, the cultural and the geographical distances between the knowledge transmitter and the knowledge receiver in order to enhance the information exchange process.

3.2.2 Learner's Neighbourhood Identification

The target learner's neighbourhood is the set of learners who are closer to her/him considering their language, their field of study, their country and their city. We are thus faced with a problem of distance minimization using the Euclidean distance.

The Euclidean distance has a lower limit of 0 indicating a perfect correspondence with no proportional upper limit. The vector representations of the profiles of two learners x and y, respectively, are $(x_1, x_2, \ldots, x_k, \ldots, x_n)$ et $(y_1, y_2, \ldots, y_k, \ldots, y_n)$. The Euclidean distance between the two profiles is calculated as shown in equation (1):

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)} = \sqrt{\sum_{i=1}^{n} z_i};\ z_i \in \{0,1\} \quad (1)$$

In our case we consider only four attributes (n = 4) which are the language, the field of study, the country and the city. For example, considering two learners X and Y characterized as follows: X=(French, Computer science, France, Paris) and Y=(French, Computer science, Belgium, Brussels). The Euclidean

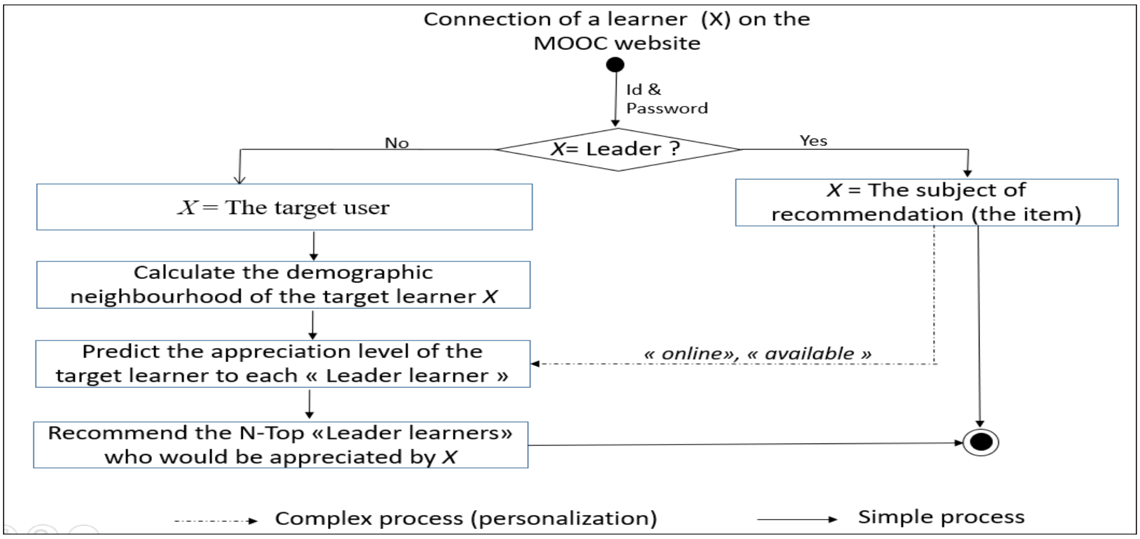

Figure 2: Recommendation process.

distance between X and Y is calculated as follows: $d(X,Y) = \sqrt{0+0+1+1} = \sqrt{2}$.

3.2.3 Prediction based on Demographic Filtering

The demographic filtering is based on the ratings made by the demographic neighbourhood of the target user. The recommendation, in our case, is the "Leader learners" who have previously been rated and appreciated by the target learner's neighbourhood.

In order to recommend to a target learner the list of "Leader learners" the more appropriate to his profile, we must predict the rate of appreciation $\hat{r}_{c,l}$ of a target learner c for a "Leader learner" l, using the ratings given by his neighbourhood for this same "Leader learner".

$$\widehat{r}_{c,l} = \frac{\sum_{(v \in V_l(c))} w_{c,v} r_{v,l}}{\sum_{(v \in V_l(c))} w_{c,v}} \tag{2}$$

In equation (2), the variables c, l, v denotes respectively the target learner, the "Leader learner" and the neighbour. The set $v_l(c)$ is the neighbourhood of the target learner having already rated the "Leader learner" l. The variable $w_{c,v}$ reflects the weight of the neighbour, calculated by its distance toward the target learner. The rate $r_{v,l}$ is the evaluation given by the neighbour v to the "Leader learner" l.

The denominator of equation (2) has been added for a standardization objective to avoid the case where the sum of the weights exceeds the value 1 which can give a predicted value out of range. More details on this measure exist in (Ricci et al., 2011).

3.3 Simplifying Assumptions

In order to take into account the possible particular cases and to manage the situations of conflicts, we applied some simplifying assumptions:

- In order to cope with the high number of the "At-risk learners" and the "Struggling learners" compared to the number of the online "Leader learners", we limit the size of a recommended list to three. Also, to respect the human capacity of a "Leader learner" we propose to her/him a maximum of three "At-risk learners" or "Struggling learners" at a time. Indeed, since the discussion is supposed to be in real time, we suppose that the effectiveness of leaders can be degraded if we grant it several learners to accompany simultaneously.
- A "Leader learner" evaluated as "irrelevant" by a target learner will no longer be recommended to her/him, even if she/he has been assessed as relevant by her/his neighbourhood. Similarly, a "Leader learner" appreciated by a target learner will automatically be recommended on the header line of the list provided that she/he is online and available, thus exchanging with less than three learners.
- In case of conflict between an "At-risk learner" and a "Struggling learner", we give priority to the struggling one considering that she/he is more motivated to complete the MOOC. First, according to the classification made by our prediction model, the "At-risk learners" will no longer be connected during the next week of the MOOC. Moreover, most of the "At-risk learners" are the

lurkers who have the prior intention of abandoning the course. Thus, the majority of them does not seek to be accompanied even if they are offered support. For this reason, it is much more beneficial to accompany and to help learners who are struggling in order to prevent their transformation to "At-risk learners".

- In the case of a new "Leader learner" who is not yet evaluated or the case of lack of available "Leader learners", the system completes the list to be recommended to the target learner by the online and available "Leader learners" from his Neighbourhood. In this case, the system recommends to the target learner the "Leader learners" who are similar to her/him instead of the "Leader learners" appreciated by the learners who are similar to her/him in order to face the cold start problem.

The "Leader learners" recommendation algorithm must consider these simplifying assumptions. The 3-Top online and available "Leader learners" with the highest $\hat{r}_{c,l}$ value will be recommended and displayed on the personal page of the target learner.

4 EXPERIMENTS AND RESULTS

In this work, our application field is a French MOOC broadcasted on a French platform and proposed by a Business School in France about the "Design Thinking". This MOOC started with 2565 learners and lasted "t= 5" weeks. The pedagogical team was composed by a tutor and two assistants. The first, the second and the fourth weeks were closed with a quiz while the third and the fifth were ended with a peer-to-peer assessment. Only data about 1535 learners were used in these experiments. Learners who have been neglected are those who have not completed the registration form. The pedagogical team of this MOOC constructed a family of 11 criteria and a weekly learning set of 30 reference learners. Each week, a decision table was built and a set of decision rules was inferred and applied to categorize the MOOC leraners.

In this section we evaluate the proposed recommender system on three aspects: the prediction quality, the space coverage and the run time performance. All algorithms in this paper are coded by Java and were run on a personal computer with Windows 7, Intel (R) $Core^{TM}$ i3-3110M CPU @ 2.4 GHz and 4.0 GB memory.

4.1 Evaluation of the Prediction Quality

Figure 3 shows a comparison between the overall F-measure and the overall accuracy of the weekly MOOC prediction model. These values represent the rates of learners who are correctly classified.

4.1.1 F-measure

It is defined as the weighted harmonic mean of the precision and recall of the test. The precision is the number of correct positive results divided by the number of all positive results returned by the classifier, and the recall is the number of correct positive results divided by the number of all relevant samples. These measures are calculated as shown in equation (3) and equation (4) respectively:

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

such that:

- True positive (TP): the number of the elements of the positive class that are correctly predicted.
- True negative (TN): the number of the elements of the negative class that are correctly predicted.
- False positive (FP): the number of the elements of the positive class that are wrongly predicted.
- False negative (FN): the number of the elements of the negative class that are wrongly predicted.

The F-measure is thus calculated as follows:

$$F-measure = \frac{2*Precision*Recall}{Precision+Recall} \tag{5}$$

4.1.2 Accuracy

It is the proportion of correct results obtained by a classifier. The accuracy is calculated as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

We notice that the accuracy provides better values than the F-measure does. Indeed, compared to an accuracy measure, the F-measure allows the distribution of errors in the predictions for a set of data. However, the accuracy only makes it possible to know whether the prediction (or the classification) is generally acceptable or not. So, it remains much more superficial than the F-measure.

4.1.3 Results Analysis

Based on the Figure 3, we can confirm that, the DRSA approach has achieved very satisfactory results. Indeed, the performance of the prediction model that we proposed gave a satisfactory F-measure that reaches 0.67 and a very satisfactory accuracy that reaches 0.89 which means tha the majority of the recommended learners were truly leaders.

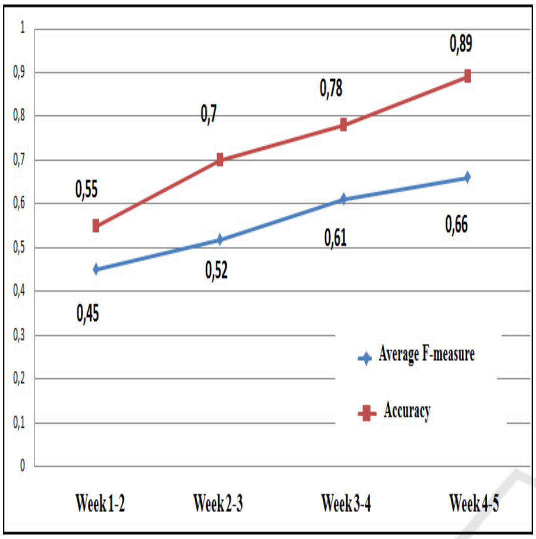

Figure 3: Comparison between the average F-measure and the overall accuracy of the decision classes during the four weeks of the MOOC.

Thus, we focus only on the F-measure values. We notice that the efficiency of the three decision classes increases from a week to another. In effect, a MOOC is known by the presence of what we call "lurkers". These are the participants who register just to discover the MOOC concept and who leave it at the first evaluation. And in spite of their activity during the first week of the MOOC, they keep having the prior intention to abandon it. This type of learners degrades the quality of the prediction model which is based on the profile and the behaviour of the learner and not on his intention. Consequently, the fewer the number of lurkers gets, the higher the prediction quality becomes.

Moreover, from a week to another, the learners enhance their participation in the forum, a thing which gives us more information concerning their profiles. In addition, the assessment activities provided by the MOOC are increasingly complex over the weeks. Obviously, compared to a Quiz , a complex assessment such as the peer-to-peer activity permits a better assessment and so, a more relevant classification. Finally, the incremental approach we developed yielded a richer training sample from one week to another. And, it is obvious that the larger the training sample, the better the model is. More details about the prediction quality are available in (Bouzayane and Saad, 2017a).

4.2 Evaluation of the Item Space Coverage

It is also important to evaluate the coverage of the item space ("Leader learners") of the recommender system. This coverage refers to the proportion of "Leader learners" recommended by the system. It represents the percentage of the recommended "Leader learners" in relation to the total number of "Leader learners". In our case the space item coverage depends on the number of "Leader learners" available and also on the number of "At-risk learners" or "Struggling learners" that we plan to help.

Figure 4 shows the results of the simulations performed on separate sets of data by modifying the size of the recommendation target set. The "Leader learners" are ordered in increasing order of frequency.

In the upper curves ((a), (b) and (c)), the recommendation concerns the "At-risk learners" classified in the decision class Cl_1 and the "Struggling learners" classified in the decision class Cl_2. However, in the lower curves ((d), (e) and (f)) the recommendation concerns only the "Struggling learners". For example in Figure 4(a) we noted that 67 of the 71 possible leader learners were selected at least once, and that the most often chosen leader learner has been recommended 310 times. We find that the coverage on the item space decreases by decreasing the target set size of the recommendation. Indeed, if we consider the ratio between the maximal number of recommended "Leader learners" and the total number of the "Leader learners" we found: $\frac{67}{71}$=0.94, $\frac{61}{89}$=0.68 and $\frac{56}{87}$= 0.64 respectively for curves (a), (b) and (c), $\frac{58}{71}$ = 0.81, $\frac{57}{89}$ = 0.64 and $\frac{46}{87}$ = 0.52 respectively for the curves (d), (e) and (f). This is due to the heterogeneity of the learners' profiles, which decreases according to their number and also to the marker we have imposed on the size of the recommendation list (3 "Leader learners" only can be recommended for each target learner) which prevented find the demographic pairs for some "Leader learners" and therefore we get smaller coverage. In this case, a "Leader learner" will be recommended several times which influences the diversity of the recommendation. However, we should note that in all of cases, more than half of the "Leader learners" were recommended.

4.3 Run Time Performance of the Recommendation Algorithm

Finally, the figure 5 shows the results of some simulations performed according to the variation of the number of learners and also the number of ratings recor-

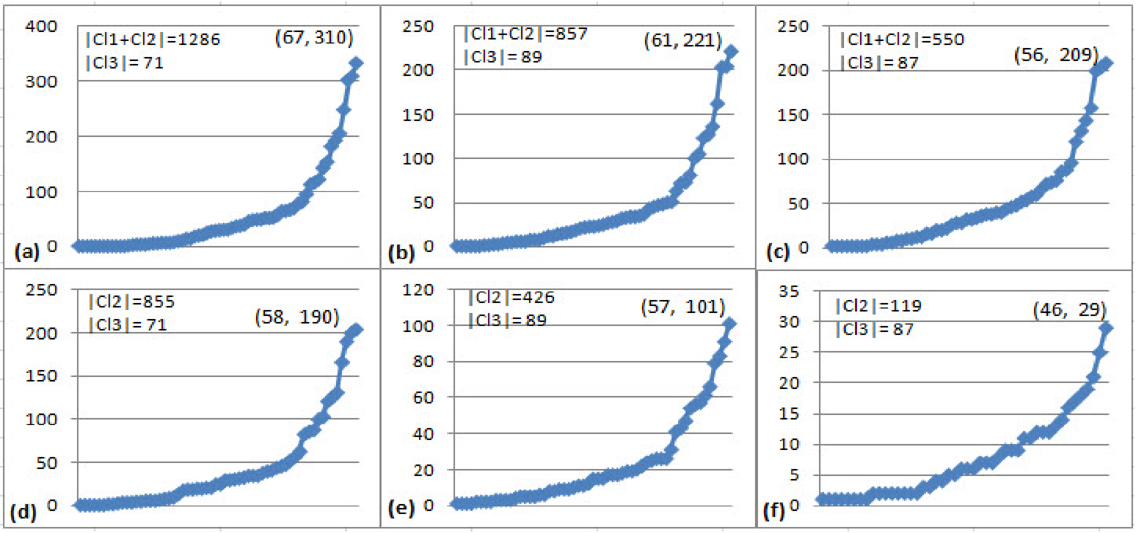

Figure 4: Coverage of the space item and diversity of proposed recommendations (X axis: Leader learner's identifier recommended; Y axis: Number of recommendations of a leader learner).

ded in the database. We find that the proposed algorithm is sensitive to the two factors. The recommender algorithm is faster when fewer learners are enrolled and fewer evaluations are given. This seems logical because demographic filtering processes the demographic data of all enrolled learners as well as the assessments they submit.

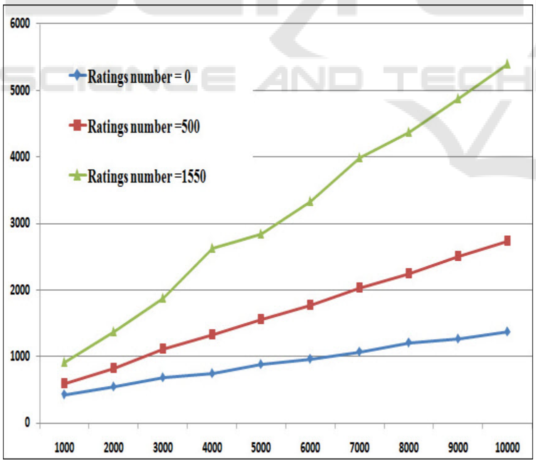

Figure 5: Execution time of the recommendation algorithm: study of the algorithm sensitivity in relation to the number of the learners and that of the stored ratings.

5 CONCLUSION

In this paper we proposed a recommender system that provides each "At-risk learner" or "Struggling learner" participating in the MOOC with a list of "Leader learners" appropriate to his profile. The "Leader learner" has the role to support the target learner throughout his learning process by providing her/him with the accurate and the immediate information she/he needs. Our objective is to help the learners absorb and understand the knowledge he receives.

Our approach is composed of two phases: (i) a periodic incremental prediction phase of the "Leader learners" based on the approach DRSA whose the objective is to weekly categorize the MOOC learners ; and (ii) a recommendation phase based on the demographic filtering and the Euclidean distance measurement that aims to recommend to each learner in need a personalized list of "Leader learners".

The proposed recommender system is a widget integrated in the personal page of an "At-risk learner" or a "Struggling learner" containing a personalized list of three "Leader learners". The quality of the recommended "Leader learners" was tested on real data from a French MOOC and proved a satisfactory F-measure that reaches 0.66. The item space coverage was also tested and yielded satisfactory rates ranging from 0.52% to 0.94%. In our future work, we intend to experiment the proposed recommender system on an online MOOC to assess its effect on the learning process of the assisted learners.

REFERENCES

Aher, S. B. and Lobo, L. (2013). Combination of machine learning algorithms for recommendation of courses in e-learning system based on historical data. *Knowledge-Based Systems*, 51:1–14.

Ambos, T. C. and Ambos, B. (2009). The impact of distance on knowledge transfer effectiveness in multinational corporations. *Journal of International Management*, 15(1):1–14.

Barak, M. (2015). The same mooc delivered in two languages: Examining knowledge construction and motivation to learn. *Proceedings of the EMOOCS*, pages 217–223.

Bousbahi, F. and Chorfi, H. (2015). Mooc-rec: a case based recommender system for moocs. *Procedia-Social and Behavioral Sciences*, 195:1813–1822.

Bouzayane, S. and Saad, I. (2017a). Prediction method based drsa to improve the individual knowledge appropriation in a collaborative learning environment: Case of moocs. In *Proceedings of the 50th Hawaii International Conference on System Sciences*, pages 124–133.

Bouzayane, S. and Saad, I. (2017b). A preference ordered classification to leader learners identification in a mooc. *Journal of Decision Systems*, 26(2):189–202.

Bouzayne, S. and Saad, I. (2017). Incremental updating algorithm of the approximations in drsa to deal with the dynamic information systems of moocs. *international conference on Knowledge Management, Information and Knowledge Systems (KMIKS)*, pages 55–66.

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12(4):331–370.

Dascalu, M.-I., Bodea, C.-N., Moldoveanu, A., Mohora, A., Lytras, M., and de Pablos, P. O. (2015). A recommender agent based on learning styles for better virtual collaborative learning experiences. *Computers in Human Behavior*, 45:243–253.

Duan, L., Street, W. N., and Xu, E. (2011). Healthcare information systems: data mining methods in the creation of a clinical recommender system. *Enterprise Information Systems*, 5(2):169–181.

Eastmond, D. V. (1994). Adult distance study through computer conferencing. *Distance Education*, 15(1):128–152.

Gooderham, P. N. (2007). Enhancing knowledge transfer in multinational corporations: a dynamic capabilities driven model. *Knowledge Management Research & Practice*, 5(1):34–43.

Greco, S., Matarazzo, B., and Slowinski, R. (2001). Rough sets theory for multicriteria decision analysis. *European Journal of Operational Research*, 129(1):1–45.

Grundstein, M. (2009). Gameth®: a constructivist and learning approach to identify and locate crucial knowledge. *International Journal of Knowledge and Learning*, 5(3-4):289–305.

Gutiérrez-Rojas, I., Leony, D., Alario-Hoyos, C., Pérez-Sanagustín, M., and Delgado-Kloos, C. (2014). Towards an outcome-based discovery and filtering of moocs using moocrank. *Proceedings of the Second MOOC European Stakeholders Summit*, pages 50–57.

Hoens, T. R., Blanton, M., Steele, A., and Chawla, N. V. (2013). Reliable medical recommendation systems with patient privacy. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 4(4):67.

Hone, K. and El Said, G. (2016). Exploring the factors affecting mooc retention: A survey study. *Computers & Education*, 98:157–168.

Kim, K.-j. and Ahn, H. (2008). A recommender system using ga k-means clustering in an online shopping market. *Expert systems with applications*, 34(2):1200–1209.

Kloft, M., Stiehler, F., Zheng, Z., and Pinkwart, N. (2014). Predicting mooc dropout over weeks using machine learning methods. In *Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs*, pages 60–65.

Labarthe, H., Bachelet, R., Bouchet, F., and Yacef, K. (2016a). Increasing mooc completion rates through social interactions: a recommendation system. *EMOOCs, Research Track*, pages 471–480.

Labarthe, H., Bouchet, F., Bachelet, R., and Yacef, K. (2016b). Does a peer recommender foster students' engagement in moocs? In *9th International Conference on Educational Data Mining*, pages 418–423.

Li, S.-W. D. and Mitros, P. (2015). Learnersourced recommendations for remediation. In *Advanced Learning Technologies (ICALT), 2015 IEEE 15th International Conference on*, pages 411–412. IEEE.

Onah, D. F. and Sinclair, J. (2015). Collaborative filtering recommendation system: a framework in massive open online courses. *INTED2015 Proceedings*, pages 1249–1257.

Onah, D. F. O., Sinclair, J., and Boyatt, R. (2014). Dropout rates of massive open online courses: behavioral patterns. *6th international conference on education and new learning technologies, published in: Edulearn14 proceedings*, pages 5825–5834.

Patru, M. and Balaji, V. (2016). Making sense of moocs: a guide for policy makers in developing countries. Technical report, UNESCO.

Ricci, F., Rokach, L., and Shapira, B. (2011). *Introduction to recommender systems handbook*. Springer.

Roy, B. and Mousseau, V. (1996). A theoretical framework for analysing the notion of relative importance of criteria. *Journal of Multi-Criteria Decision Analysis*, 5:145–159.

Sartori, F., Melen, R., and Pinardi, S. (2018). Cultivating virtual communities of practice in kafka. *Data Technologies and Applications*, 52(1):34–57.

Symeonidis, P. and Malakoudis, D. (2016). Moocrec. com: Massive open online courses recommender system.

Welch, D. E. and Welch, L. S. (2008). The importance of language in international knowledge transfer. *Management International Review*, 48(3):339–360.

Wu, D., Zhang, G., and Lu, J. (2015). A fuzzy preference tree-based recommender system for personalized business-to-business e-services. *IEEE Transactions on Fuzzy Systems*, 23(1):29–43.

Yang, D., Piergallini, M., Howley, I., and Rose, C. (2014). Forum thread recommendation for massive open online courses. In *Educational Data Mining 2014*, pages 257–260.