

HYBRID RECOMMENDATION APPROACH BASED ON A VOTING SYSTEM: EXPERIMENTATION IN AN EDUCATIONAL CONTEXT

Mohammed Baidada^{1,2}, Khalifa Mansouri³ and Franck Poirier¹

¹*Lab-STICC, Bretagne-sud University, France*

²*LIMIE, ISGA Rabat, Morocco*

³*SSDIA, ENSET, Hassan II University, Morocco*

ABSTRACT

We present in this paper the results of a second experiment that was recently conducted to evaluate a hybrid recommendation approach in an online learning environment. The approach consists of mixing the two approaches of content-based filtering and collaborative filtering to improve the relevance of the educational resources recommended to learners. A first experiment was carried out in 2019 and gave convincing results, which led us to repeat a second experimentation in order to confirm the results, on the one hand, and on the other hand, to modify the way learners evaluate the resources by transforming the "like" by a vote from one to five, in order to verify whether this will bring an improvement in the recommendations. This second experiment was also an opportunity to integrate an engine that guides learners' searches by adding criteria relating to their preferences and to check their satisfaction with the use of this engine. The results were globally positive.

KEYWORDS

E-Learning, Recommendation System, Content-Based Filtering, Collaborative Filtering, Hybrid Filtering, Experimentation

1. INTRODUCTION

Recommendation systems (RS) are increasingly used in the field of personalization in online learning environments, given their ability to determine the educational resources best suited to learners (Garrido et al., 2016; Segal et al., 2019; Souali et al., 2011). Based on filtering methods, several approaches, including hybridization, have been proposed with the aim of improving the relevance of the recommendations (Baidada et al., 2018). These approaches have also tried to exploit all kinds of interactions between learners and the learning system.

Our contribution consists of a hybrid recommendation approach that mixes the two methods of content-based filtering and collaborative filtering, to offer the educational resources best suited to a learner's preferences. We conducted a first experiment in which learners evaluated the teaching resources in a binary way with "like/dislike". The convincing results of this first experiment encouraged us to repeat it by modifying the method of evaluation by replacing the like/dislike by an evaluation on a Likert scale from 1 to 5 (vote). In this paper we present the results of the second experiment which confirmed the improvement in the relevance of the recommendations. Before that we will present a background of our work, then a state of the art of RSs in online learning environments, and a general presentation of our approach and the protocol of experimentation.

2. BACKGROUND OF OUR WORK

Our research focuses on RSs for personalization in online learning environments. We have proposed a hybrid approach mixing the two methods of content-based filtering and collaborative filtering, with the aim of

improving the relevance of the educational resources recommended to learners (Baidada et al. 2018). Our approach was tested between April and May 2019, and the results were conclusive. They shown that the hybrid recommendation approach gives better results than the content-based filtering and collaborative filtering separately, except for the private institute where we had obtained almost similar results.

In order to confirm our results, we conduct a new experiment, with two essential extensions:

1. Change the binary evaluation mode (like/dislike) for the shared resources by the learners used in the first experiment, by an evaluation method on a Likert scale with a voting system ranging from 1 to 5;
2. Propose an integrated search engine which helps the learners to find external resources corresponding to their preferences.

These two points, relating to the second experiment, are the subject of our paper and will be detailed in the section reserved for the presentation of our contribution.

3. STATE OF THE ART

3.1 Recommendation Systems

RSs are increasingly used on the web. With the aim of offering the most suitable content to users, they are found in e-commerce (Amazon, Ebay), video streaming sites (YouTube, Netflix), social networks (Facebook, Twitter, LinkedIn) (Burke et al., 2011; Klačnja-Milićević et al., 2011). They have also been used in e-learning to guarantee the personalization of the learning process by recommending the most suitable content for the learners. D. Herath and L. Jayarathne (2018) used an RS based on data mining techniques to predict learners' performance. M. Tadlaoui et al. (2015, 2018) exploited the social links connecting the learners to recommend educational resources to them; they were interested in popular, useful and recently viewed resources. A. Klačnja-Milicevic et al. (2011) proposed a system which recommends sequences of pedagogical activities to learners by considering their learning styles.

3.2 Filtering Methods

Filtering methods are the basis of RSs. Often they are classified as follows (Isinkaye et al., 2015):

- Content-based filtering: it offers the user similar items and which have the same values for a set of attributes which describe them, it requires knowledge of a description of the items and user preferences;
- Collaborative filtering: which considers the similarities of the user to a group of users to whom he or she can be linked. There are two subtypes: user-based and item-based. The one that was used in the experiment is user-based collaborative filtering, which considers the similarity of the user to a group of users with whom he or she shares interests.
- Hybrid filtering: This is a combination of two or more filtering methods.

4. OUR CONTRIBUTION

As mentioned in the background, this new experiment brings to the first one the two extensions already mentioned, namely, modifying the resource evaluation mode by replacing the like/dislike with an evaluation on a Likert scale from 1 to 5 (vote), and also to offer learners the possibility of searching for external educational resources relating to their preferences through an integrated search engine.

The purpose of this new experiment through the integration of the discrete-scale evaluation method is to verify its impact on the recommendations of the educational resources. And then its purpose, through the proposal of the integrated search engine, is to exploit the learner's profile to filter the educational resources that he or she is likely to look for and make them correspond as closely as possible to his or her preferences.

The 2nd experiment will therefore allow us to verify the following hypotheses:

- The 2nd experiment must confirm the results of the 1st experiment by showing that hybrid filtering gives better recommendations than the other filtering approaches applied separately;

- The evaluation of educational resources with a discrete scale of 1 to 5, must have an impact on the recommendations (best values for the precision and recall indicators);
 - Learners will appreciate the external resources proposed by the integrated search engine.
- A technical description of the platform used in the first and second experiment will be given in the following section.

5. DESCRIPTION OF THE EXPERIMENT

For the first experiment we used the Moodle platform, to which we added SocialWall as a plugin which transforms the format of the courses into a social network appearance (Post, Comment, Like, etc). We have also developed recommendation modules related to the different filtering approaches used, which we have integrated into the platform. Figure 1 presents the diagram of our platform for the first experiment.

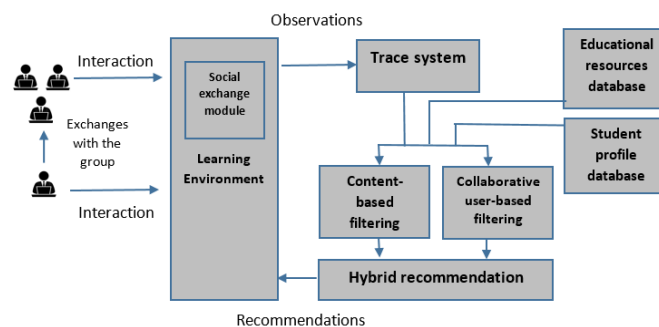


Figure 1. Diagram of the platform used for the 1st experiment in the case of the hybrid recommendation

The method of calculating recommendations according to each filtering method is described in our previous work (Baidada et al., 2018). Recall that we used a weighted hybridization approach (Burke, 2007), i.e. the results of the two filtering methods were combined numerically with a mean with equal coefficients.

Regarding the notion of the profile, we are based on the reference models IEEE PAPI (Public and Private Information for Learners) and IMS LIP (Learner Information Package) (Pavlov and Paneva, 2006; Wei and Yan, 2009), and we considered the "Preferences" component.

Subsequently, through the study of standards relating to Learning Object Metadata (LOM), namely IEEE LOM, CanCORE LOM and Dublin Core Metadata (Roy et al., 2010), we retained 3 elements relating to preferences of the learner who are General, Technical and Educational. The following table (table 1) gives a description of the final criteria retained with their possible values:

Table 1. Learner preferences metadata

Section	Element retained	Values
General	Language	English, French, Arabic
Technical	Resource format	Video, Document (pdf, slides, ...), Web article
Educational	Resource type	Course, Tests or exercises, Forum

For the second experiment, the diagram was retained in its entirety, making two essential modifications:

- Changing the like / dislike rating system by a voting system;
- Proposal of a search engine for external educational resources (Figure 2).

Our integrated search engine is based on the CURL resource extractor module. It is a module that uses the CURL extension (Client URL) of PHP to send http requests to remote servers. These requests are initiated by the learners and enriched with the criteria related to their profiles before being launched on these servers. For example, for a learner who searches for the key words "initiation en PHP", and knowing that in his or her profile we retained his or her preferences for videos as types of resources, courses as type of activity, and language French, then the final request which will be launched will be:

"http://www.google.com/search?q=Intiation+en+PHP+cours+video&lr=lang_fr".

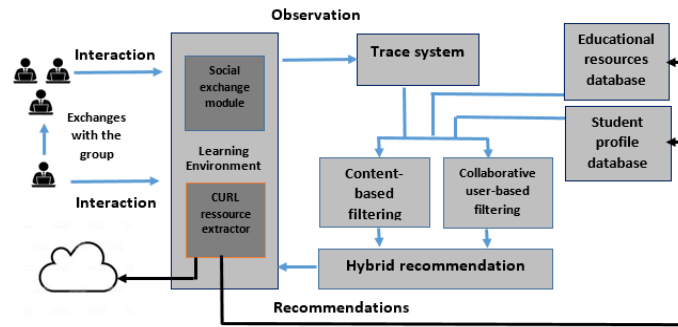


Figure 2. Diagram of the platform used for the 2nd experiment in the case of the hybrid recommendation

The experiment targeted two groups of engineering students in two different public and private institutes in Morocco. For the private institute, it is a group of 3rd year computer engineering of 33 students. For the public institute, this is a group of 1st year computer engineer cycles of 64 students. The average age of the groups is 21.1 with a standard deviation of 1.36. There are 32% female and 68% male.

Each group was divided into three subgroups:

- Subgroup 1: to whom we proposed a content-based filtering approach;
- Subgroup 2: to whom we proposed a collaborative filtering approach;
- Subgroup 3: to whom we have proposed a hybrid approach to the previous methods.

Table 2 below shows the distribution of the different groups:

Table 2. Size of groups

	Sub-group 1	Sub-group 2	Sub-group 3	Total
Public	21	21	22	64
Private	11	11	11	33
Total	32	32	33	97

Students were enrolled in a web programming course for the private institute, and an object-oriented programming course for the public institute. The experiment took place between the beginning of January and mid-February 2020.

We chose an online evaluation method with real users. This evaluation method guarantees better results than the offline evaluation method with datasets.

6. EVALUATION METHOD

To assess our recommendation system, we used the three indicators commonly used in this area (Isinkaye et al., 2015; Portugal et al., 2017), namely:

- Precision: means the percentage of results which are relevant, it measures the quality of the recommendations (the accuracy). It is given by the following report: Number of relevant recommended resources / Number of recommended resources;
- Recall: refers to the percentage of total relevant results correctly classified by the algorithm, it measures the quantity of recommendations (completeness). It is given by the report: Number of relevant recommended resources / Total number of relevant resources;
- F-measure: It is a harmonic average of the two previous indicators.

We used the same indicators for the evaluation as for the first experiment, with an essential difference in the choice of the relevance threshold. For the first experiment we considered a binary evaluation with like / dislike and we therefore considered a relevant resource any resource on which a learner has like. For this second experiment, we considered a voting system for resources with values ranging from 1 to 5. It was therefore necessary to determine the relevance threshold for resources. Instead of setting a value between 1 to 5 (Ben Ticha, 2018), we chose the average of the votes of a resource by the learners, which generated a different relevance threshold for each resource.

Regarding the analysis method, we opted for the ANOVA (ANalysis Of VAriance) method which allows us to test the significant differences between the means. It is used to measure changes in a metric variable by one or more nominal explanatory factors. In our case, we chose the ANOVA1 variance analysis to explain the variation of a single dependent metric variable by a single explanatory factor. More precisely, we seek to test the effect of the qualitative variable "group" successively on each of the quantitative variables "Precision", "Recall" and "F-measure". We applied ANOVA1 to each institute separately, and then applied it to both institutes.

Finally, we tested the effect of the qualitative variable "Institute" on each of the indicators. For this we used the Student T-test, which compares the means of two sample groups.

In order to test the equality of the means between the three groups, we used the null hypothesis:

- H_0 : there is no difference between the three groups
- H_1 : at least one group is different from the others

We used the Tukey test which proposes a significance level of 0.05, called sig (Tukey significant difference). The hypothesis 0 will be rejected if the threshold obtained is lower than 0.05, and therefore we will deduce that at least one group is different from the others.

Note that we have adopted the same methods and analytical approaches as for the first experiment, in order to confirm the results.

7. RESULTS

7.1 ANOVA1 for the Two Combined Institutes

We used SPSS software to get the results. As a first step, we considered the two institutes together and an inter-group analysis was performed. This analysis gave a significance level (sig) of 0.00. Hypothesis 0 was rejected, and we deduced that at least one group is different from the others. This required a multiple comparison of the means to see which group is different. Table 3 summarizes the results for the three indicators precision, recall and F-measure:

Table 3. Multiple comparison of the three indicators - the two institutes grouped

	Sig(G1 vs G2)	Sig(G1 vs G3)	Sig(G2 vs G3)
Precision	0.29	0.00	0.00
Recall	0.25	0.00	0.00
F-measure	0.26	0.00	0.00

We find that group 3 compared to the other two groups 1 and 2 have a significance level (sig) equal to 0, confirming that it is different from the other two groups. We then proceeded to a classification in homogeneous subsets, to find the groups which present similarities and the groups which differ, this by considering the harmonic averages. Table 4 presents the results:

Table 4. Classification into homogeneous subsets for the three indicators - the two institutes grouped

Groups	Size of group	Precision		Recall		F-measure	
		Subset		Subset		Subset	
		1	2	1	2	1	2
Group 1	32	0.38		0.24		0.29	
Group 2	32	0.42		0.27		0.33	
Group 3	33		0.59		0.43		0.49

For the three indicators, group 3 has a harmonic mean much higher than the other groups, and it is classified each time in a single subset, while the other two groups are classified in the same subset.

7.2 ANOVA1 for the Private and Public Institute Separately

For the two institutes considered separately, the inter-group analysis gave a significance level sig less than 0.05. It has once again been deduced that at least one group is distinguished from the others.

Following the same approach, the multiple comparison of the three indicators respectively for the two institutes gave the results presented in Tables 5 and 6:

Table 5. Multiple comparison of the three indicators - Private Institute

	Sig(G1 vs G2)	Sig(G1 vs G3)	Sig(G2 vs G3)
Precision	0.621	0.001	0.014
Recall	0.955	0.000	0.001
F-measure	0.849	0.001	0.003

Table 6. Multiple comparison of the three indicators - Public Institute

	Sig(G1 vs G2)	Sig(G1 vs G3)	Sig(G2 vs G3)
Precision	0.378	0.000	0.000
Recall	0.147	0.000	0.000
F-measure	0.216	0.000	0.000

For the two institutes group 3 presents a difference compared to the others with significance thresholds lower than 0.05. We must proceed to the classification into homogeneous subsets given by Tables 7 and 8:

Table 7. Classification into homogeneous subsets for the three indicators - private institute

		Precision		Recall		F-measure	
		Subset		Subset		Subset	
Groups	Size of group	1	2	1	2	1	2
Group 1	11	0.33		0.23		0.27	
Group 2	11	0.37		0.24		0.29	
Group 3	11		0.49		0.35		0.41

Table 8. Classification into homogeneous subsets for the three indicators - public institute

		Precision		Recall		F-measure	
		Subset		Subset		Subset	
Groups	Size of group	1	2	1	2	1	2
Group 1	21	0.40		0.24		0.30	
Group 2	21	0.45		0.29		0.35	
Group 3	22		0.63		0.45		0.53

For the two institutes separately, each time group 3 has a better harmonic mean and is classified in a subset while the two groups 1 and 2 are classified in another.

7.3 T-Student Test between the Two Institutes

We used T-Student test to check whether the variable "Institute" has an influence on each of the indicators. Referring to the results table (Table 9), we note that the significance level sig, for the three indicators, is less than 0.05, which leads us to confirm a difference between the two institutes.

Table 9. T-Student test

		sig
Precision	equal variances hypothesis	0.002
	unequal variances hypothesis	0.001
Recall	equal variances hypothesis	0.016
	unequal variances hypothesis	0.009
F-measure	equal variances hypothesis	0.007
	unequal variances hypothesis	0.004

According to the statistics table provided by SPSS (Table 10), we get higher values for the public institute compared to the private institute. It's especially the precision that makes a big difference.

Table 10. Group statistics

	Institute	Size	Average	Standard deviation	Mean standard error
Precision	Private	33	0.40	0.12	0.02
	Public	64	0.50	0.15	0.18
Recall	Private	33	0.27	0.09	0.15
	Public	64	0.33	0.11	0.14
F-measure	Private	33	0.32	0.10	0.17
	Public	64	0.40	0.13	0.16

7.4 Questionnaire Results

To support the results obtained, a questionnaire was distributed to the learners who participated in the experiment. 66 responded, here are some of the statistics that emerged:

- For the satisfaction with the experiment platform : 59% respond "very good"
- For the satisfaction with the exchange module : 50% respond "very good"
- For the relevance of shared resources in the group : 54% respond "good"
- For the relevance of recommended resources : 64% respond "very good"
- For matching recommended resources to their preferences : 54% respond "a lot"

With regard to the use of the search engine, out of 48 who evaluated it, 62% confirm that the resources proposed correspond "a lot" to their preferences (Figure 3).

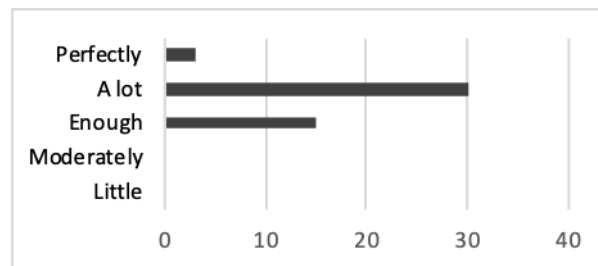


Figure 3. Matching of researched resources to preferences

8. DISCUSSION

Through the results obtained, this second experiment confirms the results of the first, by showing that group 3 to which we applied the hybrid recommendation approach presented better results compared to the other two groups, for the case of the two combined institutes, and for the public institute and this time also for the private institute. On the other hand, the public institute always presents better results, the thing that we had explained for the first experiment also, by the overall level of students in the public sector which is generally better, the thing which influences the quality of shared resources and their relevance when evaluating these resources.

One of the objectives of this second experiment was to check whether the evaluation with a voting system would improve the quality of recommendations compared to a binary evaluation with like / dislike. The results obtained rather show a slight improvement for the recall indicator, which leads us to confirm that we have gained in terms of the completeness of the recommended resources rather than in their quality.

The questionnaire made it possible to confirm the general appreciation of the learners, both with regard to the use of the platform, the social exchange module, and the relevance of the resources recommended by the system, and that these corresponded very much to their preferences. Learners also appreciated the integrated search engine, confirming that the results it generated correspond to their preferences.

9. CONCLUSION

We have verified through the first experiment that the hybridization of the two methods of content-based filtering and collaborative filtering improves the relevance of the recommendations in an online learning context. The second experiment was an opportunity to confirm this result, but also to verify the impact of the evaluation method on the relevance of the recommendations. It was also an opportunity to extend the platform to a proposal of external resources that best matched the learners' preferences, through the integrated search engine, and to check their satisfaction with the results returned.

The results were positive, the hybrid recommendation always gives better results, the evaluation method with a scale from 1 to 5 impacts the Recall indicator, and finally the learners were globally satisfied with the use of the integrated search engine and the results it offers.

This will encourage us to develop our platform by taking full advantage of learners' interactions with the system to better understand their needs and interests. This can be achieved by exploiting the queries made on the search engine, by analyzing those that most interest the learners.

REFERENCES

- Baidada, M., Mansouri, K., Poirier, F. 2018. Hybrid Recommendation Approach in Online Learning Environments, In: Rocha Á., Serrhini M. (eds) Information Systems and Technologies to Support Learning. EMENA-ISTL 2018. Smart Innovation, Systems and Technologies, vol 111. Springer, Cham. pp 39-43. https://doi.org/10.1007/978-3-030-03577-8_5
- Ben Ticha, S. 2018. Recommendation personnalisée hybride, Phd Thesis, University of Lorraine, France in joint supervision with the University of Tunis El Manar, Tunisia
- Burke, R. 2007. Hybrid web recommender systems. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4321 LNCS, 377–408.
- Burke, R. et al., 2011. Recommender systems: An overview. *AI Magazine*, 32(3), 13–18. <https://doi.org/10.1609/aimag.v32i3.2361>
- Garrido, A. et al., 2016. On the use of case-based planning for e-learning personalization. *Expert Systems with Applications*, 60, 1–15. <https://doi.org/10.1016/j.eswa.2016.04.030>
- Herath, D. and Jayarathne, L. 2018. Intelligent Recommendations for e-Learning Personalization Based on Learner's Learning Activities and Performances, *International Journal of Computer Science and Software Engineering*, Vol.7, Issue 6, pp.130-137. <https://doi.org/10.13140/RG.2.2.27785.34406>
- Isinkaye, F. O. et al., 2015. Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, 16(3), 261–273. <https://doi.org/10.1016/j.eij.2015.06.005>
- Klašnja-Milićević, A. et al., 2011. E-Learning personalization based on hybrid recommendation strategy and learning style identification. *Computers and Education*, 56(3), 885–899. <https://doi.org/10.1016/j.compedu.2010.11.001>
- Pavlov, R. and Paneva, D. 2006. Personalized and adaptive learning—approaches and solutions. In *The Proceedings of the Third CHIRON Open Workshop "Visions of Ubiquitous Learning"*, 20 June, 2006, Stockholm, Sweden, pp. 6-19
- Portugal, I. et al., 2017. The use of machine learning algorithms in recommender systems: A systematic review. *Expert Systems with Applications*, 97, 205–227. <https://doi.org/10.1016/j.eswa.2017.12.020>
- Roy, D. et al., 2010. A Comparative Study of Learning Object Metadata, Learning Material Repositories, Metadata Annotation & an Automatic Metadata Annotation Tool. *Boley & Akerkar*, 2, 103–126.
- Segal, A. et al., 2019. A difficulty ranking approach to personalization in E-learning. *International Journal of Human Computer Studies*, 130(October 2017), 261–272. <https://doi.org/10.1016/j.ijhcs.2019.07.002>
- Souali, K. et al., 2011. A new recommender system for e-learning environments. *International Conference on Multimedia Computing and Systems -Proceedings*, 1–4. <https://doi.org/10.1109/ICMCS.2011.5945630>
- Tadlaoui, M. et al., 2015. Approche pour recommandation de ressources pédagogiques basée sur les liens sociaux, EIAH'2015, June 2015, 7ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain, Agadir, Morocco, pp.192-203
- Tadlaoui, M. 2018. Système de recommandation de ressources pédagogiques fondé sur les liens sociaux formalisation et évaluation, PhD Thesis, INSA Lyon, France, in joint supervision with Tlemcen University, Tlemcen, Algeria
- Wei, X. and Yan, J. 2009. Learner profile design for personalized E-learning systems. *Proceedings - 2009 International Conference on Computational Intelligence and Software Engineering, CiSE 2009*. <https://doi.org/10.1109/CISE.2009.5363560>