

Article

Using a Hybrid Recommending System for Learning Videos in Flipped Classrooms and MOOCs

Jaume Jordán ^{1,*}, Soledad Valero ^{1,†}, Carlos Turró ^{2,†} and Vicent Botti ^{1,†}

¹ Valencian Research Institute for Artificial Intelligence (VRAIN), Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain; svalero@dsic.upv.es (S.V.); vbotti@dsic.upv.es (V.B.)

² Área de Sistemas de Información y Comunicaciones, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain; turro@cc.upv.es

* Correspondence: jjordan@dsic.upv.es; Tel.: +34-963-877-000

† These authors contributed equally to this work.

Abstract: New challenges in education require new ways of education. Higher education has adapted to these new challenges by means of offering new types of training like massive online open courses and by updating their teaching methodology using novel approaches as flipped classrooms. These types of training have enabled universities to better adapt to the challenges posed by the pandemic. In addition, high quality learning objects are necessary for these new forms of education to be successful, with learning videos being the most common learning objects to provide theoretical concepts. This paper describes a new approach of a previously presented hybrid learning recommender system based on content-based techniques, which was capable of recommend useful videos to learners and lecturers from a learning video repository. In this new approach, the content-based techniques are also combined with a collaborative filtering module, which increases the probability of recommending relevant videos. This hybrid technique has been successfully applied to a real scenario in the central video repository of the Universitat Politècnica de València.



Citation: Jordán, J.; Valero, S.; Turró, C.; Botti, V. Using a Hybrid Recommending System for Learning Videos in Flipped Classrooms and MOOCs. *Electronics* **2021**, *10*, 1226. <https://doi.org/10.3390/electronics10111226>

Received: 15 April 2021

Accepted: 17 May 2021

Published: 21 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: learning recommender system; learning object; learning videos; content-based; collaborative filtering

1. Introduction

New challenges in education has raised due to the students' profile changes in the last decade. They demand new ways of learning, better adapted to their way of life and moving away from classical teaching. Academic institutions must be agile in adapting their teaching methodology to the new forms required, taking into account the opportunities offered by the global world [1,2]. In this way, there has been a great increase in the supply of massive online open courses (MOOCs) by academic institutions, as well as in the number of students opting for this type of training [3]. MOOCs mainly relay on learning objects (LOs). As IEEE proposes, a LO is "any entity, digital or non-digital, which can be used, re-used or referenced during technology supported learning" [4]. Thus, MOOCs use different types of LOs, with videos being one of the most commonly used to teach theoretical concepts. In addition, new teaching approaches are emerging in higher education, such as flipped teaching [5–7], in which the theoretical content is studied at home by the students, while the face-to-face sessions are eminently practical, where the knowledge acquired is put into practice by solving problems. To this end, the lecturer instructs the students which LOs they should work on at home before the next face-to-face session. In this way, students are encouraged to acquire the theoretical concepts not only through books or specialised articles, but also through audio–visual material.

Furthermore, in this pandemic context, face-to-face classes have been replaced by online classes in many institutions, increasing the adoption of flipped learning. Lecturers need to plan subjects taking into account possible connectivity problems, as the possibility

of students being unable to attend online classes due to health problems (quarantine, hospitalisation) increases. In addition, students may have difficulties in accessing devices during working hours, because they share them with their parents, etc. All these new circumstances make it even more important to provide educators with useful tools to search for and recommend good LOs, which can be accessed by students at any time.

Universitat Politècnica de València (<http://www.upv.es>, accessed on 14 May 2021) (UPV) is a Spanish Public university which offers undergraduate degrees, dual degrees, masters and doctoral programs. UPV has more than 28,000 students. UPV has been promoting new pedagogical methodologies in their degrees in the last decade, such as flipped teaching [8]. It has also made a great effort in developing MOOCs within the edX platform (<https://www.edx.org/>, accessed on 14 May 2021), with more than 2 million enrollments and having three courses in Class Central all time top 100 MOOCs in 2019 (<https://www.classcentral.com/report/top-moocs-2019-edition/>, accessed on 14 May 2021), and two courses top 30 in 2020 (<https://www.classcentral.com/report/best-free-online-courses-2021/>, accessed on 14 May 2021). In addition, UPV participated in the movement that arose during the first months of the pandemic to offer free certificates for some of the MOOCs offered. This fact, together with the need for new training channels, has led to a significant increase in the number of students on this type of courses (<https://www.classcentral.com/report/mooc-stats-2020/>, accessed on 14 May 2021).

Students need access to a variety of resources to understand the theoretical concepts required in blended and flipped classroom environments. To facilitate this difficult work, UPV has had a long-standing digital resources project with the aim of producing video content as LOs. This video content is handled in the university central video repository, called mediaUPV (<https://media.upv.es/>, accessed on 14 May 2021). This portal is not only used in the field of MOOCs, but also in other educational projects.

mediaUPV allows UPV lecturers to upload and manage video content for students. Students access mediaUPV usually through suggestions made by their lecturers through the learning management System (LMS), but they also access the video portal and browse through the content on their own. A relevant feature of mediaUPV against other alternatives (e.g., YouTube) is that the videos have been prepared and recorded by lecturers from the institution, so the quality of the content is guaranteed.

mediaUPV portal has become an essential tool to the institution during this pandemic; however, the size of the mediaUPV content is a growing problem, and hence, it is increasingly difficult to find the most relevant content for both students to view and lecturers to suggest. Thus, UPV determined that both students and lecturers would benefit from the development of a new Learning Recommendation System (LRS), that could recommend relevant and related videos. Therefore, in our previous work [9], we presented the first recommender engine proposed to carry out that purpose, which combined two content-based techniques to recommend useful learning videos to learners and lecturers. However, with that engine it was only possible to recommend videos labelled as high quality, that is, with transcripts. Therefore, it is necessary to apply other techniques to cover the entire mediaUPV catalogue.

In this paper we present an enhancement of the previous work, which extends the proposed recommendation engine also using collaborative filtering techniques. Thus, we describe how we have designed and developed a hybrid recommender system based on both content-based techniques and collaborative filtering. Furthermore, a complete analysis of the results in production of the initial content-based LRS since its application in October 2019 until March 2021 is provided.

This article is structured as follows. In Section 2 related works are described. Following, Section 3 specifies a description about the LOs to recommend and the potential users of the system. In Section 4, the proposed recommender system is explained. Then, Section 5 shows the experimental results of the proposed recommender using the data of the mediaUPV portal. In addition, Section 6 provides an analysis of the results of the LRS used in production. Finally, Section 7 draws the conclusions and future work of this paper.

2. Related Work

Learning Recommender Systems (LRS) should assist learners in discovering relevant LO than keep them motivated and enable them to complete their learning activities [10]. Most of the LRS adopt the same techniques than regular recommender systems [10–13], such as: content-based, in which recommendations are determined considering user profiles and content analysis of the learning objects already visited by the user; collaborative filtering, in which recommendations are based on the choices of other similar user profiles; knowledge-based, in which it is inferred whether a LO satisfies a particular learning need of the user to recommend it; and hybrid, in which recommendations are computed by combining more than one of the above techniques.

In recent years, different approaches have been proposed in order to improve the efficiency and accuracy of the recommendations and retrieval of useful LOs. In this way, in [14], authors provide new metrics for applying collaborative filtering in a learning domain, so users with better academic results have greater weight in the calculation of the recommendations. However, the experiments did not carried out in a learning environment. In another proposal, Zapata et al. provide a tool for filtering the retrieved results from a user query, which uses a combination of different filtering techniques, such as content comparison, and collaborative and demographic searches [15].

Other proposals focus on recommending to students those LOs that can be most useful to them, providing solid arguments. This is the case of [13], which combine content-based, collaborative and knowledge-based recommenders using an argumentation-based module to recommend LOs inside a LMS. In this case, information on student profiles and learning styles is also available. An item-based collaborative filtering method is combined with a sequential pattern mining algorithm to recommend LOs to learners in [16]. In this case, LOs are ranked by the students and it is also possible to obtain the browsing sequences made by them. In a similar way, ref. [17] proposes a hybrid knowledge-based recommender system based on ontology and sequential pattern mining for recommendation of LO. Authors can adequately characterize learners and LOs using an ontology, since they have detailed information about them. Personalized learning paths (sequence of LOs) that maximizes the performance of the learner and effectiveness of learning are provided in [18], hybridizing ant colony optimization with genetic algorithm. In this case, authors use both learners and LOs attributes to determine the appropriate learning paths. In addition, in [19], Dwivedi et al. recommend learning paths using a variable length genetic algorithm. This approach considers learners' learning styles and knowledge levels extracted from the learners' registration process. Finally, in [20], authors propose a method, based in a collaborative filtering approach, for building a unified learner profile which is used to recommend LOs to a group of individuals.

Besides, other works have been done to improve the accuracy of the searches in mediaUPV. For example, in [21] a semi-supervised method is applied to cluster and classify the LOs of mediaUPV, obtaining specific keywords that represent each cluster. In [22], authors applied a custom approach for indexing and retrieving educational videos using their transcripts, which are available in mediaUPV. Videos are classified in different domains using the method described in [21]. In addition, they applied a Latent Dirichlet Allocation algorithm [23] to get a list of topics and their score. User queries are classified in one of the domains, recovering from that cluster those videos whose transcripts are the closest to the query.

As can be seen, most previous work on LRS adopts a hybrid strategy, seeking to harness the strength of each particular technique, overcoming its limitations by using them together. Furthermore, the different strategies that can be applied depend on the data available to describe LOs and users. In our case, a hybrid strategy will also be applied, combining content-based methods. On the other hand, previous experiences on the improvement of searches in the mediaUPV repository show us the usefulness of characterizing LOs using available transcripts and titles.

3. Problem Description

mediaUPV portal started in 2011 and by the end of 2019 it had 55,600 different videos mainly from STEM topics, with more than 10 million views. During 2020 until the end of March 2021, the mediaUPV catalogue increased almost a 44.8%, reaching 80,500. This significant growth may be due to the increased need for the use of online learning resources, not only for new forms of learning, but also for the institution's traditional courses that had to be converted to purely online teaching during the pandemic. From this database, only 13,232 by the end of 2019, and 20,135 by the end of March 2021 are certified as high quality LOs (an increase of 52.2%). All these high quality videos have a transcript (with more than 100 characters), a title, and an author, however the videos are not classified using any taxonomy and no useful keywords are associated to them for all cases. mediaUPV platform generates the transcripts of what is said in the videos using the poli[Trans] service (<https://politrans.upv.es>, accessed on 14 May 2021), an online platform for automated and assisted multilingual media subtitling offered by UPV. poli[Trans] service is based on transLectures-UPV Platform [24].

Hence, a 75% of the mediaUPV catalogue is composed by videos without transcripts. It was detected that a huge percentage of that videos have a medium quality and could be also interesting to lecturers and students. In fact, some of them are recordings of classes from some subjects, which can be used for reinforcing the learning of students of similar subjects in other degrees. By default, mediaUPV does not obtain the transcripts of these recordings, because is the lecturer who has the permissions to demand them.

mediaUPV portal is mainly used by students and lecturers. The students are mainly formal students of the UPV, but it also receives many visits from anonymous users, who can register in some MOOC offered by the UPV. Moreover, mediaUPV is not connected to the LMS of the UPV, so even though the user is authenticated, it is not possible to know his student profile (e.g., enrolled subjects).

Therefore, the aim of this proposal is to be able to offer recommendations not only to authenticated users but also to anonymous users of the system. In addition, the system should be able to recommend not only students, but also lecturers who want to find quality videos to suggest to their students. Therefore, the main objective is to be able to offer recommendations on learning videos (that is LOs) that meet the quality standards set by the UPV.

It is important to note that all recommendations made on the mediaUPV portal are always through a video. That is, when a user logged into the system is watching a video (which can suggest videos similar to the current video or content that the user has watched previously), or an anonymous user watching a video (which can only suggest content similar to the current video). This implies that there are no recommendations only associated with a logged-in user. Thus, a hybrid system needs to be used in which the current video, and (optionally) the logged-in user are considered.

4. Learning Recommender System Proposal

Collaborative recommendation has some well-known difficulties, such as the necessity of a huge quantity of data for making accurate recommendations. In the same way, content-based approaches also have some problems, such as the lack of serendipity. For this reason, a hybrid approach which combines the previous techniques can improve the accuracy in the provided recommendations and reduce the cold start impact.

In this way, our LRS is based on two different approaches:

- First, a content-based recommender module based on two components. One component recommends taking into account the activity of the user identified in the system, the profile-based module (PB). The other component offers recommendations based on the content of the video being watched at the time, the item-based module (IB).
- Second, a collaborative filtering module (CF), which offers suggestions based on similar other users with a viewing history similar to current user.

Thus, the computed recommendations are based on these two approaches, getting a hybrid recommendation system, so that the user receives recommendations of videos similar to the one she is watching at that moment, or also of videos that may interest her due to her viewing history, and additionally, content considered interesting due to the opinions of other users with a viewing history similar to her own. This fact increases the serendipity, discovering contents to the user that are of her interest, although they are not so similar, a priori, to learning videos already watched by the user so far.

In our proposal, LOs/videos are characterized by their title and their transcript (when available). This characterization is used for the calculation of the similarity between the LOs. As transcripts are in different languages, it is possible to recommend videos from different languages.

Because there is a large set of words in the transcript and title of the videos, it is necessary to have an algorithm that filters out the too common words, which do not serve to differentiate the content. Thus, it is possible to focus on the particular words in the entire collection, which serve to identify the content of a video. We need to use an unsupervised term weighting approach, as our video repository is not categorized. Therefore, the algorithm chosen is the well-known term frequency–inverse document frequency (TF-IDF) [25], as it is a common term weighting scheme used to represent documents. In order to improve the performance of the TF-IDF algorithm, also the stop words from the *nlTK Python* (<https://www.nltk.org/>, accessed on 14 May 2021) package are used. Furthermore, some ad hoc words have been added to this package.

The item-based module takes only the information from the content of the videos, i.e., there is no information about the users. So, we can consider this module as a recommendation by item-item similarity to be used when a (maybe anonymous) user is watching a video. To do this, we take the characteristics (transcript and title) of each of the videos to calculate the item-item matrix with the TF-IDF algorithm. Then, the cosine similarity of two instances of this item-item matrix returns the similarity among the different videos. Cosine similarity calculates the similarity between two n-dimensional vectors by the angle between them in the vector space:

$$\text{cosine_sim}(\vec{p}, \vec{q}) = \frac{\vec{p} \cdot \vec{q}}{|\vec{p}| * |\vec{q}|} \quad (1)$$

The profile-based module considers the content information (the terms extracted by the TF-IDF algorithm from the transcript and title) of the videos viewed by users. In this way, the recommendations to a user are made based on the similarity among the viewed videos of the user in mediaUPV. In this case, the similarity is calculated using the cosine profile-item matrix.

The collaborative filtering module considers the similarity between the users based on their viewed videos. Thus, two users will be more similar depending on how many videos they have both watched. In the opposite way, a user will have no similarity with another user if the sets of videos watched by each of the two users are completely disjoint. Therefore, the similarity is calculated using the cosine similarity between the users based on their watched videos. Then, the prediction values of recommendation for each user are calculated taking the dot product between the similarity of users and the viewed videos matrix, normalizing the data properly. With these process, we have the prediction to recommend each user in the system considering all users similarity (we will call this approach all users). However, we can also obtain a different prediction to recommend if we only consider the top k most similar users to the user to be recommended. To do that, we calculate the top k nearest neighbours (NN) for each user in the system, and then, the prediction values to recommend videos are also calculated (we will call this approach top k NN). We note that the computation cost of this last approach considering the top k NN is higher than the all users approach. The computation cost is also increased as k grows.

Finally, our *hybrid recommendation (HR)* system consists of two components, one content-based (with two modules) and another collaborative-based (with only one module).

Thus, the recommendations are made taking into account the intersection of the videos recommended by the modules (see Equation (2)). The rest of the recommendations are obtained considering first the weight applied to each component, i.e., w_{CB} for the content-based component and w_{CF} for the collaborative filtering module. Furthermore, it is possible to balance the importance given to the two modules of the content-based component by means of other two weights, w_{IB} for the item-based module and w_{PB} for the profile-based one. In this way, in cases where there is no user (anonymous) the w_{PB} and w_{CF} are set to 0. Likewise, when an authenticated user is not yet watching a video, the w_{IB} is set to 0 (however, this case is not currently applicable to mediaUPV portal).

$$HR = (IB \cap PB \cap CF) \cup (w_{CB} \cdot ((w_{IB} \cdot IB) \cup (w_{PB} \cdot PB)) \cup (w_{CF} \cdot CF)) \quad (2)$$

In the next section, we explain the experiments carried out to determine the combination of weights which offers the best performance.

5. Experimental Results

In this section, we explain the experiments carried out with the proposed LRS. We have made tests with data from the videos watched by the users from September 2018 to July 2019, dividing this data into a training set and a test set. We used this data to be able to compare the new collaborative filtering module with the experiments already presented in our previous work [9]. In this section, we first present the general experimental setup in Section 5.1. Then, the experiments of both content-based modules are explained in Section 5.2. Section 5.3 presents the experiments with the collaborative filtering module. The experiments with the hybrid approach, which is the combination of the content-based component and the collaborative filtering module, are explained in Section 5.4. Finally, Section 5.5 contains a discussion of the results of all these experimental results.

5.1. Experimental Setup

As mentioned above, the database is comprised of all the videos available on the mediaUPV as items to be recommended, as well as the usage data (views) of the users logged into the system. Although the platform had more than 55,600 videos by the end of 2019, in the following tests we have filtered out the hidden videos (only available with a direct link), the videos that do not have a transcript, and the videos in which the transcript has less than 100 characters, leaving a total of 13,232 videos suitable for recommendation.

The data set used for these experiments is formed by of learning videos viewed by users during an academic year at the UPV, from September 2018 to July 2019. The training data set consists of the data from September 2018 to April 2019, while the testing data set consists of the data from May to July 2019, i.e., 8 months for training and 3 months for testing. The videos considered are those present on the platform until July 2019, after filtering them as described above. Thus, we try to simulate a real scenario, in which the training data represent the past activity of the users, while the test data is formed by the activity of the following 3 months (“future”). Therefore, any recommendation from the recommender that is among the videos that users have actually watched in the test set is considered a success. Additionally, in these tests we only consider five recommendations since it is the number required by the mediaUPV portal.

We use the well-known precision and recall measures to evaluate the success of the recommender. Precision can be defined as the successful recommendations made (videos that have been viewed by the user in the test set) divided by the number of recommendations made:

$$precision = \frac{success_recommendation}{recommendations_made} \quad (3)$$

Recall is defined as the successful recommendations made divided by the number of watched videos in the test set:

$$recall = \frac{success_recommendation}{watched_videos} \quad (4)$$

For our tests we considered a set of regular users (*reg_users*) that we define as those who have watched between 10 and 150 videos both in the training and test periods, having 1044 users in this set. We also considered a set of new users that have watched between 1 and 9 videos both in the training and test periods (there are 815 users in this set). If nothing is specified, the regular users set is the one used.

5.2. Content-Based Component

To test the content-based modules we use both of them together to combine their efforts. So this is like using the hybrid recommender engine but without considering the collaborative filtering module. Thus, we focus on the content-based module centred on the video being watched, that is, the item-based module (*IB*); and on the content-based module that considers the user's views, namely the profile-based module (*PB*). The weights for each module in the hybrid recommender engine are set to $w_{IB} = 50%$, $w_{PB} = 50%$ and $w_{CF} = 0%$.

5.2.1. Setting Transcript and Title Features

In this first test, we analyze the success of recommendations for the set of regular users considering different amount of features for the transcript and the title of the video to train the LRS, in order to establish the better amount of both. The graph in Figure 1 shows the precision and recall for different values of the number of features considered for the transcript, while the number of features for the title is kept at zero.

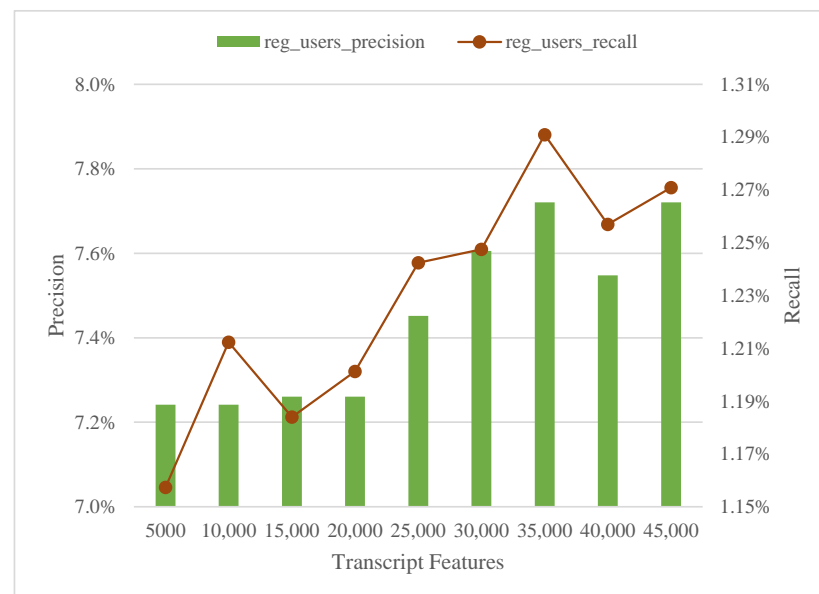


Figure 1. Precision and recall results for the regular users set with different amount of transcript features.

In general, precision and recall increase slightly as the value of the features for the transcript increases (from 7.2% to 7.7% for precision, and from 1.15% to 1.29% for recall), as would be expected when more information is available from the transcript. The best precision values are with 35,000 and 45,000 features for the transcript. However, the best recall value is in the case of 35,000 transcript features. Therefore, the best configuration for the recommender would be to use 35,000 transcript features, since this value achieves higher recall than any other and equals the precision obtained with 45,000 transcript

features. In addition, the computation of 35,000 transcript features is computationally less expensive and, in particular, implies a lower memory cost.

It should be noted that, in the experiment made with 35,000 transcript features, the number of users who have been recommended successfully is 254 of 1044, i.e., 24.33%.

Figure 2 shows the precision and recall for different values in the number of features considered for the title having the transcript features fixed to 35,000. The best values of precision and recall are obtained with the number of features of the title at 0. In addition, both values decrease slightly and in a relatively uniform way as the number of title features increases. So, apparently it is better to skip the title features. However, it is interesting to analyze this considering the nature of the LOs of mediaUPV. In this way, we have analyzed the set of words that determines the TF-IDF algorithm. As we mentioned before, the videos on the platform correspond mainly to university courses, so there is a set of terms that are certainly repetitive in the titles of the videos and do not provide any differentiating information with respect to their content. Among these terms, we find the following: {'analysis', 'calculation', 'control', 'creation', 'data', 'design', 'exercise', 'engineering', 'introduction', 'management', 'mechanism', 'model', 'module', 'practical', 'practice', 'presentation', 'simulation', 'system', 'systems', 'theme', 'unit', 'virtual'}.

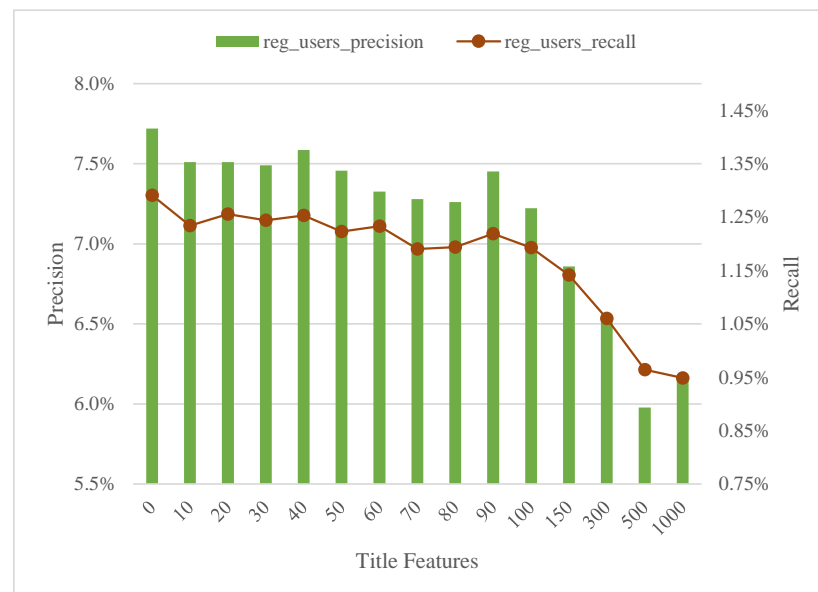


Figure 2. Precision and recall results for the regular users set with different amount of title features.

5.2.2. Filtering Title Features

In order to increase the precision of the recommendations, we decided to filter the terms of the previous list from the titles of the videos by considering them as stop words. The results of applying this correction can be seen in the Figure 3, which shows the precision and recall for 35,000 transcript features and different amount of title features. In this case, it can be seen that the best global values of precision and recall are still obtained with 0 title features. However, it should be noted that with 10 title features, for the case where the specified title terms have been filtered out, the precision and recall almost reach this base case, slightly surpassing the case without filtering (with 10 title features). In addition, for values up to 30 title features, the precision and recall are better for the case with filtered terms. However, from 40 title features on, the effect of filtering is diluted and the results are generally slightly worse, and most of the cases slightly worse than the unfiltered case.



Figure 3. Precision and recall comparison with and without filtering title terms for different amount of title features.

Consequently, we can say that filtering has significantly improved the results but it is not enough to make the title relevant. Perhaps it would be necessary a still greater filtering of terms that we have not considered ‘commo’ and that the TF-IDF algorithm has not identified as such either. However, since we are considering 35,000 transcript terms, the inclusion of 10 to 50 terms from the title can be considered irrelevant after the analysis. Furthermore, it can also be interpreted as a video being better characterized by its own transcript than by its title.

Although we have already seen that it is better to skip the title in all cases, we will analyze in detail the difference between filtering the title terms and not filtering them for 10 title terms and different amount of transcript features. Figure 4 shows this comparison (precision and recall) with and without filtering. Precision and recall are significantly better if filtering of title terms is performed in all cases, with the only exception of 45,000 transcript features where precision is slightly higher for the unfiltered case (being also the best result for the different transcript values for the unfiltered case of title terms). In this particular case, it could be that by considering only 10 features of the title, but 45,000 for the transcript, the effect of the filtering of the title is diluted. However, this is not a very significant difference. On the other hand, as previously observed, the best results, both in terms of precision and recall, are obtained with 35,000 transcript features in the case of filtering the title terms.

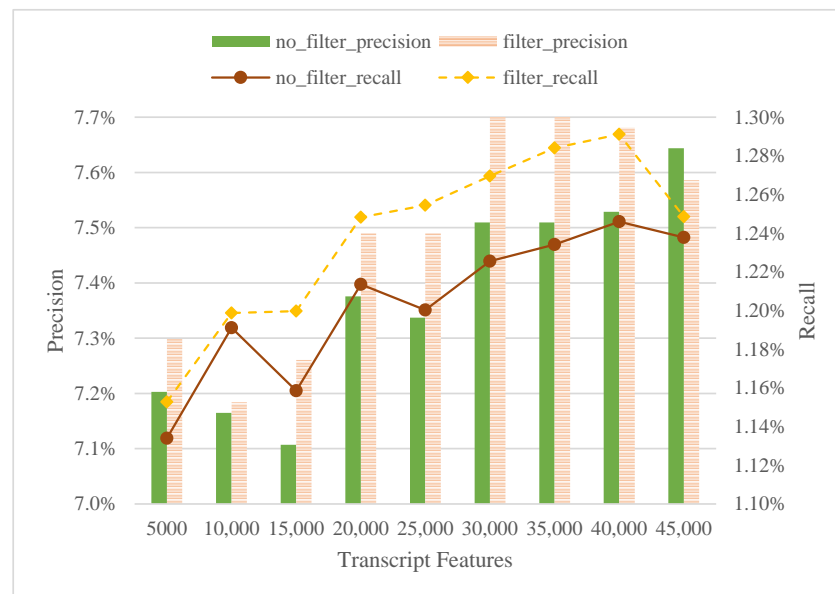


Figure 4. Precision and recall comparison with and without filtering title terms for different amount of transcript features with 10 title features.

5.3. Collaborative Filtering Component

In this subsection we conducted experiments to test the performance of the collaborative filtering module in its two variants, i.e., the one that considers the similarity with all users (which we refer to it as all users) to make recommendations (taking the videos with the highest recommendation values) versus the one that only considers the top k nearest neighbour (NN) users to make the relevant recommendations.

Figure 5 presents the results of precision and recall for the sets of regular users (watched from 10 to 150 videos) and new users (watched from 1 to 9 videos). The values shown correspond to the collaborative filtering module considering all users recommendations in the first column, and the subsequent columns consider the top k NN users for recommending. Generally, for both the regular and new users sets, the best results in precision and recall are obtained by the CF all users approach, which in fact doubles the values for almost all the top k NN approaches. In this way, it is clear that the best approach for collaborative filtering with mediaUPV data is the one that uses CF all users instead of any number of the top k NN. Additionally, this approach is less costly computationally. If we compare the different k values of the top NN, it seems that the best are 25 for the regular users set, and 50 for the new users set, slightly decreasing in both cases as k grows. All in all, the difference with the CF all users approach is significant enough to avoid any top k NN approach.

In the case of the regular users, the CF all users approach achieves almost 9.5% precision and 1.5% of recall. However, with the new users set, the precision is almost 3.7% and the recall over 10%. This difference between both users sets is mainly due to the amount of historical data of the users. In the case of the new users set, data about them is limited, as they have only watched 1 to 9 videos. This causes the precision of the recommendation to be significantly lower than the precision obtained in the regular users set, in which the amount of data is larger to build a more solid recommender. This difference is commonly known in literature as the cold start problem. However, since our recommender engine has different modules, we can leverage them to obtain satisfactory results in any case.

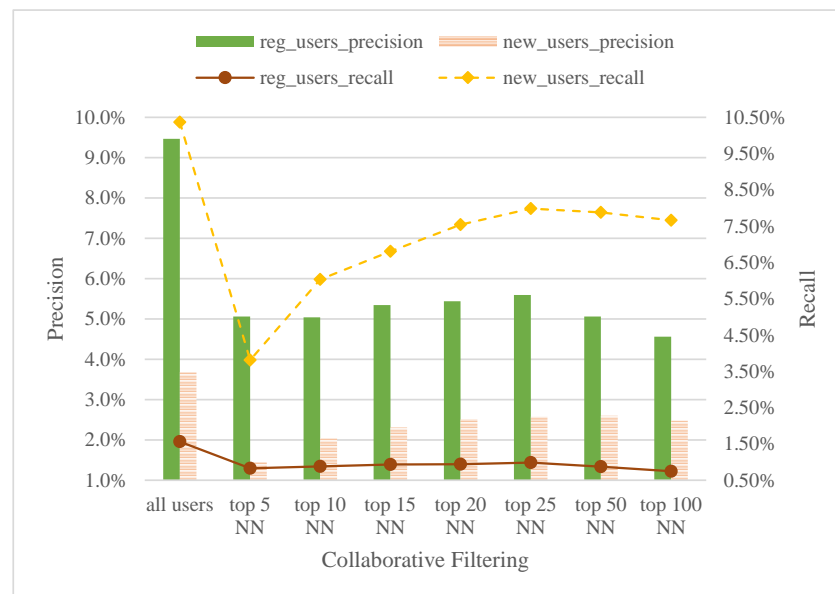


Figure 5. Precision and recall comparison of regular users and new users for collaborative filtering module with different values of top N users obtained by nearest neighbour (NN) technique.

5.4. Hybrid Weights Setting

Our LRS is a hybrid approach that considers, on the one hand, a content-based component with two modules named profile-based module and item-based module, and on the other hand, a collaborative filtering component (with a unique module) as it is specified in Equation (2). In this subsection, we first analyse the best combination of weights for the content-based modules. Once found them, we study how to tune the weights of the content-based component and the collaborative filtering component.

In Figure 6 we make a comparison of precision and recall of two different sets of users, using different weights for the content-based modules of our hybrid LRS. We first consider the set of new users of which we have little knowledge as they have only watched 1 to 9 videos in the data set (815 users). The second set are the regular users, formed by users that have already watched between 10 and 150 videos in the data set (1044 users). For regular users, the best precision and recall is obtained with balanced weights, i.e., $\{w_{PB} = 40\%; w_{IB} = 60\%\}$ and $\{w_{PB} = 60\%; w_{IB} = 40\%\}$. However, for new users, the best precision and recall values are obtained with low w_{PB} , with 15.83% of new users receiving successful recommendations.

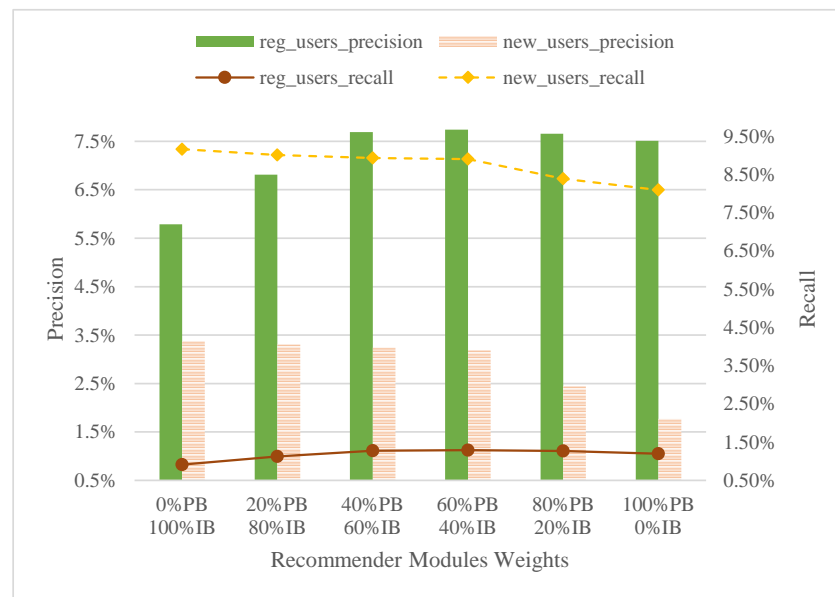


Figure 6. Precision and recall comparison of regular users and new users, using different content-based modules weights.

Having set the weights of the content-based modules to $w_{PB} = 60\%$ and $w_{IB} = 40\%$, since they obtain the best results, we can now set the weights of the combination of the content-based component (w_{CB}) with the collaborative filtering component (w_{CF}) as specified in Equation (2). For this, Figure 7 shows the results of precision and recall for both groups of regular and new users with different weights for the recommender modules, the content-based modules (CB) and the collaborative filtering module (CF). In this case, the results for the different values of weights are the same except of the extreme values of $\{w_{CB} = 0\%; w_{CF} = 100\%\}$, and $\{w_{CB} = 100\%; w_{CF} = 0\%\}$, that are slightly lower. The reason behind this is that the intersection of recommendations of both the content-based modules and the collaborative filtering module already gives the best results, and hence, any combination of the weights (except the extremes) can be considered.

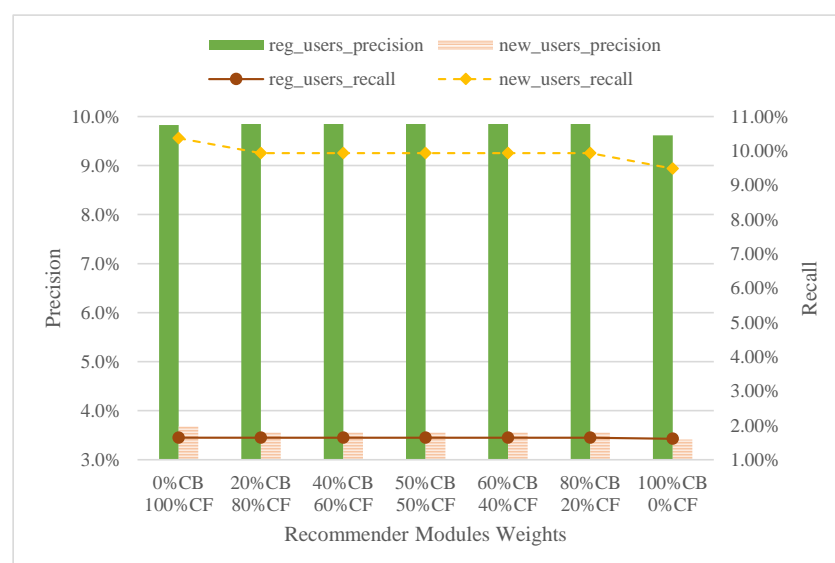


Figure 7. Precision and recall comparison of regular users and new users, using different content-based and collaborative filtering modules weights.

The incorporation of the collaborative filtering module improves the precision from 7.7% to 9.85% for regular users (for new users is increased from 3.3% to 3.56%) with respect to the version of the LRS with only the content-based modules proposed in [9]. Recall is also slightly improved from 1.3% to 1.6% for regular users (9.1% to 9.9% for new users). In this case, 28.5% of the regular users and 16.4% of the new users received a useful recommendation. We also note that the results of the collaborative filtering module alone (see Figure 5) when compared with the combined version with the content-based modules also improve from 9.5% to 9.85% of precision for the regular users. Therefore, all of these are positive results that justify the need of applying the collaborative filtering component to our LRS to improve its accuracy.

Since there is no significant difference between the weights of the content-based component and the collaborative filtering component, we propose balanced weights to apply in production $\{w_{CB} = 50\%; w_{CF} = 50\%\}$.

5.5. Discussion

For new users from which the system has few information is harder to make successful recommendations; however, this could improve in the future during the application of the LRS as the users get more engaged to mediaUPV portal. In the case of the users from which there is more historical data available, the accuracy of our LRS improves significantly, specially when using the new developed collaborative filtering module combined with the content-based modules. Furthermore, the collaborative filtering module will be able to recommend videos that do not have transcript (we remind that only 20,135 from the 80,500 videos have transcript), which will suppose a lot of more possibilities to recommend to the logged-in users when applied to production.

We emphasize that even though our hybrid LRS obtains a precision of 9.85% simulating a real environment (we improved previous results of 7.7% precision without the collaborative filtering module of [9]), 28.5% regular users and 16.4% of new users received some good recommendation. In addition, the precision and recall obtained are significantly better than a random recommendation.

In conclusion, after this analysis we established the parameters of the LRS to be applied in production in the mediaUPV portal as 35,000 transcript features, 0 title features, $w_{CB} = 50\% \cdot (w_{PB} = 60\%, w_{IB} = 40\%), w_{CF} = 50\%$.

6. Production Results

In this section we show the results of the application of the LRS described in [9] to the mediaUPV portal. That proposal did not contain a collaborative filtering module. It should be noted that mediaUPV did not have a LRS until the application in October 2019. The parameter configuration of the recommender system applied to production was the one that worked best in the experiments of the previous work [9], that is, 35,000 transcript features, 0 title features, $w_{PB} = 60\%$ and $w_{IB} = 40\%$ (without the collaborative filtering module which has been developed for the present work).

The graph of Figure 8 presents the global precision and recall results of the recommendations made to the users for the period from October 2019 until March 2021. We compare different time ranges, from 10 to 480 min, in which the user can watch the recommended videos. Hence, the column of 10 min represents the precision and recall of the video recommendations that are watched within the 10 min after the recommendation is made. So the precision increases as the time range grows. As it can be seen in Figure 8, the precision is significantly lower for the cases below 120 min, and then, it only increases slightly. The reason behind that might be that the users of mediaUPV portal usually watch long videos, or they even do other tasks (like homework in the case of students, or preparing other classes or new material in the case of lecturers) between watching a video and the next. This may explain why the precision increases significantly if we consider a time after recommendation of at least 120 min instead of 10 or 30 min.

Overall, a precision of around 9.5% for 120 min after the recommendation, or higher than 10% if we consider more minutes is a suitable result for our recommender system. We note that most of the users that enter in mediaUPV portal do not seek for recommendations, since they only watch the video that they need to (i.e., a student who must watch the corresponding lesson). Additionally, the production results are significantly better (around 2–3% higher precision) than the experimental results with the training and test sets of September 2018 to July 2019 originally used in our previous work [9].

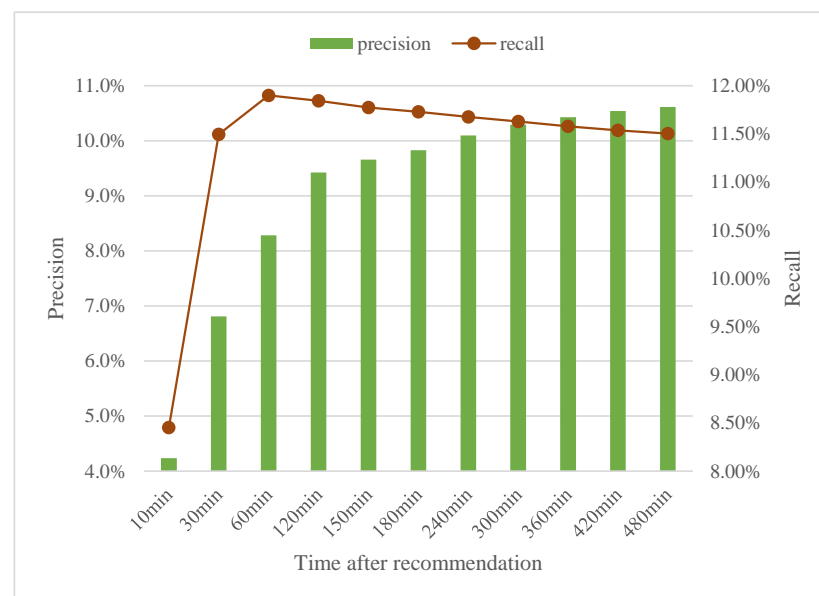


Figure 8. Production results of precision and recall considering different time range for the user to watch the recommended video after the recommendation.

Figure 9 presents the results of precision and recall for different type of users considering at most 120 min after the recommendation to watch the video. All users (all_users) is the set that includes the total amount of users that watched any video and received any recommendation in our system from October 2019 to March 2021. The set of all users is divided in other three sets, namely: new users (new_users), which are those that watched between 1 and 9 videos in the period of application of our recommender system; regular users (reg_users), the ones that watched 10 to 150 videos in the period; and top users (top_users), those who watched more than 150 videos in the period. According to this classification, we have 2297 new users, 8202 regular users, and 634 top users. The results in precision and recall for all users and regular users are reasonably similar since this last set is the larger, so it influences more the all users set, but also it is the middle point between the new users and top users set.

The main results of Figure 9 show that the precision of our recommender engine for new users is 6.3% with a recall of 12.74%. This precision is significantly lower than the precision for both the regular users and the top users, that is 9.23% and 9.8%, respectively, with a lower recall for the top users of 10.6%. From these results, we can confirm that as the known data from the user increases, the accuracy of the recommendations also increase. However, it is even hard to increase precision with top users due to the nature of the recommendations of mediaUPV portal, in which most users only come to watch the specific videos they have to. This scenario differs significantly from platforms like YouTube where most users enter looking to spend some leisure time.

Globally, we make some successful recommendation to 35% of all users. Particularly, we have successful recommendations for 24% of new users, 51% in the case of regular users, and 78,6% for the top users. This results clearly show that a successful or useful recommendation for a user is totally related to the amount of (historical) data that the system has of the user. It is important to note that almost the 80% of the top users (the

most informed set of users we defined) received any successful recommendation, which is a very positive result for the recommender system. We can assume that our recommender would be more accurate as the users are more engaged with it.

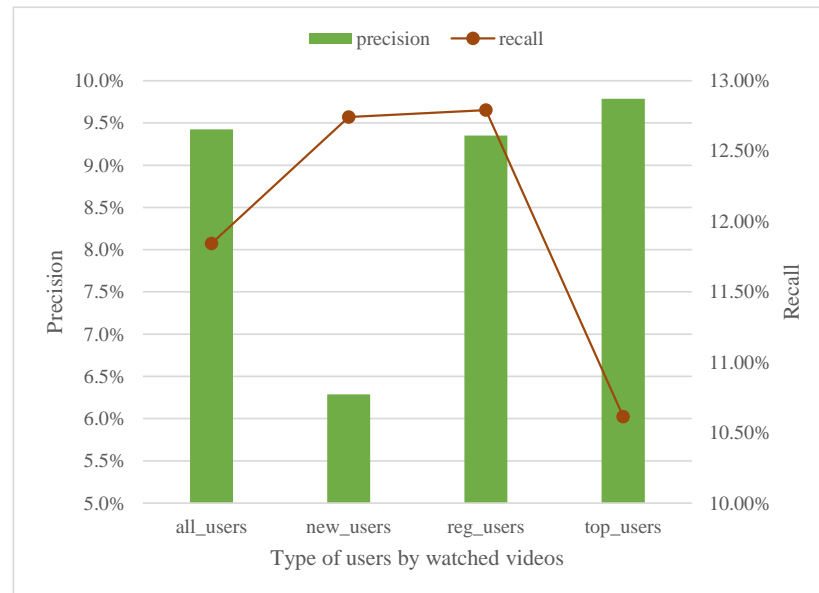


Figure 9. Production results of precision and recall by different type of users within the 120 min after the recommendation. All users include the total amount of users that received any recommendation. New users are those that watched between 1 and 9 videos, regular users watched 10 to 150 videos, and top users watched more than 150 videos.

In Figure 10, we show a violin plot (<https://towardsdatascience.com/violin-plots-explained-fb1d115e023d>, accessed on 14 May 2021) depicting the distribution of daily accesses to the videos of the mediaUPV platform during the year 2019, with no recommender (before October) and with the proposed one. As can be seen clearly, the existence of the LRS is correlated with an increase of the number of accesses to the videos by the users.

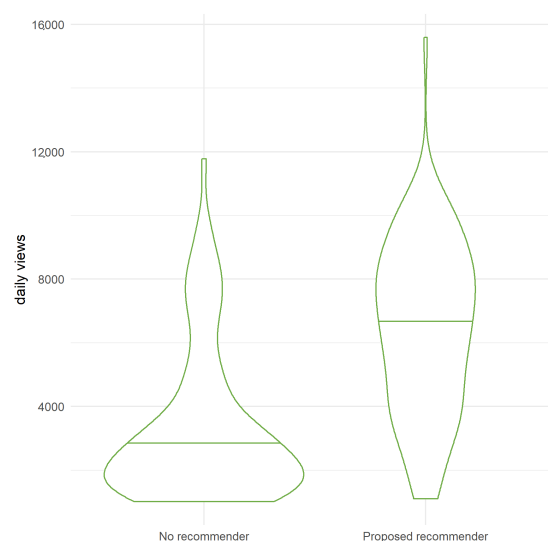


Figure 10. Video access with and without a recommender in production.

A common way in industry to measure relative quality of a recommender system is the Click-Through Rate (CTR) (<https://en.wikipedia.org/wiki/Click-throughrate>, accessed on 14 May 2021), that measures the percentage of clicks in the recommender per number

of views. As the CTR is used by the ads industry, there is an ongoing interest in CTR prediction techniques [26]. In the case of a generic recommender system, anything above 0.35% means you are doing a good job (www.acquisio.com/blog/agency/what-is-a-good-click-through-rate-ctr/, accessed on 14 May 2021). As can be seen in Figure 11, the CTR is 1.28% on average, with notable peaks over 4%. These results can be considered quite satisfactory, since they imply that the recommendations made to users generate interest in them.

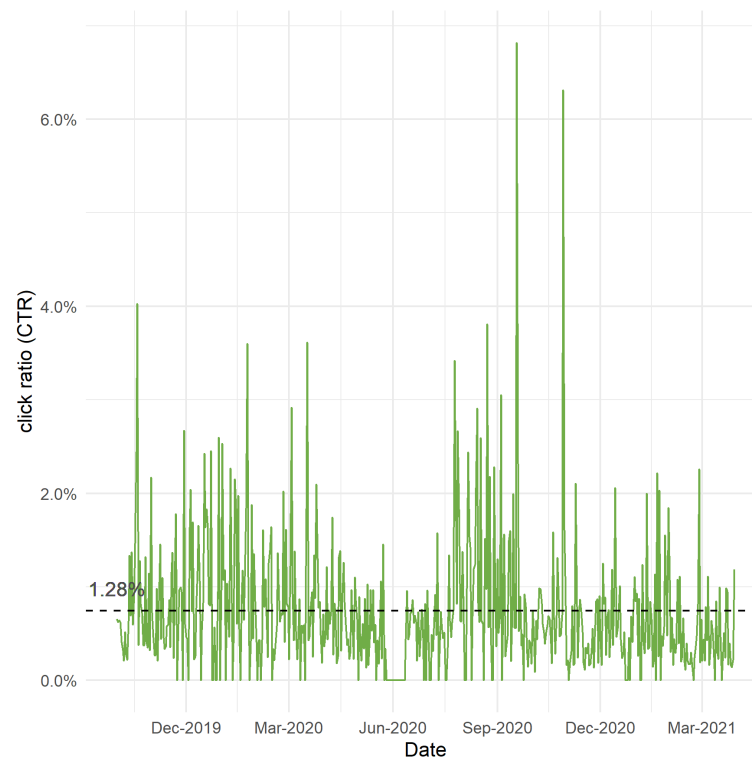


Figure 11. Clicks in recommender (CTR) in production.

Finally, we point out the percentage of clicks on each of the recommended videos according to their order in the list, demonstrating the relevance of this order to users (see Figure 12). About 28% of users click on the first video, and more than half of the users click on the first two recommendations with a distribution that seems to be heavy tailed. So the most relevant video by far for users is the first one and then the rest of the videos follow a decreasing order of importance, which also points to a reasonable work of the presented LRS.

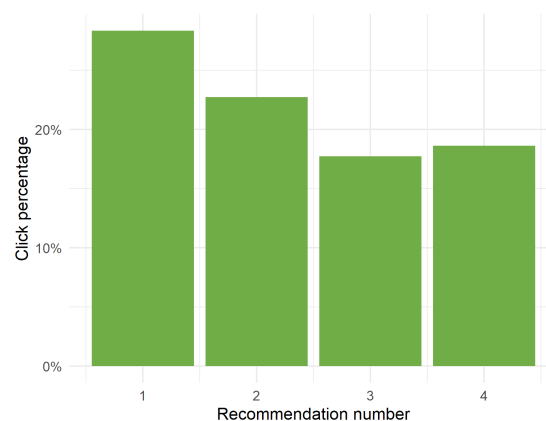


Figure 12. Percentage of clicks in recommendations per list position.

7. Conclusions

This work proposes a new hybrid LRS based on collaborative filtering and content-based components capable of recommending learning videos based on viewing history and current video content. Thus, the LRS proposed is able to recommend not only to authenticated users but also to anonymous users from the mediaUPV portal, independently if they are lecturers or students. In fact, mediaUPV portal has not information about learners' profiles or needs, as it is not connected with any LMS.

The hybrid LRS has been applied to a simulated environment, using a data set of learning videos and user profiles from the 2018-2019 academic year at UPV. The best hybrid LRS configuration obtained 9.85% of precision and 1.6% of recall, where 28.5% of regular users received some useful recommendation.

Furthermore, the content-based component of the approach has been applied in a real scenario, the mediaUPV portal of the UPV from October 2019 to March 2021. This portal is mainly used by learners and trainers to access to useful LOs for their MOOCs and flipped classrooms. We can state that the application of this LRS to the mediaUPV portal was positive as it improved the precision of the original experimental results of [9] (from 7.7% to 9.85%), it brought an increase in visits to the videos, and it had a significant CTR of 1.28% on average, with notable peaks of over 4%.

The results of our LRS must be seen in the context of its application. In this respect, it should be remembered that we are dealing with videos of university lectures or subject-specific lessons. This means that users of the system do not usually enter for leisure purposes as on YouTube or Netflix, or to make purchases as on platforms such as Amazon. Thus, users of this system usually enter to watch a specific lesson of the subjects they are studying, or to search for a specific video about a particular topic. Therefore, it is difficult for a recommender to obtain better results than the ones we show. In addition, it is important to highlight what has been observed with the results in production with respect to the time that elapses between the recommendation and the moment in which the users watch one of the recommended videos. In this sense, it has been shown (Figure 8) that the precision of the recommender almost doubles if instead of considering the 30 min after the recommendation we consider 120 min or more.

As future work, we want to analyse the results in production of both components working together, the collaborative filtering and content-based one. Thus, the collaborative filtering component presented in this work is able to recommend videos that have no transcript, which opens up more possibilities to increase the serendipity of the recommendations. In addition, it would be interesting to evaluate whether to add the classification of the mediaUPV videos obtained by [21] to the current characterization of the videos used by our proposal, which is currently based on the video transcript and collaborative filtering.

We also want to test if we achieve better results by changing the term frequency-inverse document frequency (TF-IDF) algorithm in the content-based module using other techniques such as delta TF-IDF [27] or TF.IDF.ICF [28], which try to avoid the TF-IDF problem of not considering intraclass or interclass distributions. In addition, we need to further study if it will be possible to apply other variant weighting approaches of TF-IDF, such as the presented in [29], in which the number of occurrences of a term, the number of documents that include the term, and the number of classes in which the term appears, are used to obtain a more accurate set of characteristic features.

Finally, we would like to conduct random user surveys to analyse the precision of the recommender in a less automatic way and to get direct feedback from users. In this way, we could know better the precision of the recommendations as the users could answer if these recommendations are useful instead of basing them on whether they have seen the video or not. This would avoid the uncertainty of the current analysis in which we do not know for sure if users do not find the recommendations useful or if they only enter the platform to watch the content they need without paying attention to any recommendation, whether it is useful or not.

Author Contributions: Conceptualization, J.J., S.V. and V.B.; Data curation, J.J. and C.T.; Formal analysis, J.J. and S.V.; Funding acquisition, V.B.; Investigation, J.J., S.V. and C.T.; Methodology, V.B.; Project administration, V.B.; Resources, C.T.; Software, J.J.; Supervision, C.T. and V.B.; Validation, J.J., S.V. and C.T.; Writing—original draft, J.J. and S.V.; Writing—review & editing, J.J., S.V., C.T. and V.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially supported by MINECO/FEDER RTI2018-095390-B-C31 and TIN2017-89156-R projects of the Spanish government, and PROMETEO/2018/002 project of Generalitat Valenciana.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CB	Content-Based Recommender Modules
CF	Collaborative Filtering Module
CTR	Click-Through Rate
HR	Hybrid Recommender
IB	Item-Based Recommender Module
LMS	Learning Management System
LRS	Learning Recommender System
LO	Learning Object
MOOC	Massive Online Open Course
PB	Profile-Based Recommender Module
TF-IDF	term frequency—inverse document frequency
UPV	Universitat Politècnica de València

References

- Maassen, P.; Nerland, M.; Yates, L. (Eds.) *Reconfiguring Knowledge in Higher Education*; Higher Education Dynamics, 50; Springer International Publishing: Cham, Switzerland, 2018.
- Zajda, J.; Rust, V. *Globalisation and Higher Education Reforms*; Springer: Cham, Switzerland, 2016; Volume 15.
- van Dijck, J.; Poell, T. Higher Education in a Networked World: European Responses to U.S. MOOCs. *Int. J. Commun. IJoC* **2015**, *9*, 2674–2692.
- Institute and Committee of Electrical and Electronics Engineers; Learning Technology Standards. *IEEE Standard for Learning Object Metadata*; IEEE Standard 1484.12.1; IEEE: Piscataway, NJ, USA, 2002.
- O’Flaherty, J.; Phillips, C. The use of flipped classrooms in higher education: A scoping review. *Internet High. Educ.* **2015**, *25*, 85–95. [[CrossRef](#)]
- Roehl, A.; Reddy, S.L.; Shannon, G.J. The flipped classroom: An opportunity to engage millennial students through active learning strategies. *J. Fam. Consum. Sci.* **2013**, *105*, 44–49. [[CrossRef](#)]
- Tucker, B. The Flipped Classroom. Online instruction at home frees class time for learning. *Educ. Next* **2012**, *2012*, 82–83.
- Turró, C.; Morales, J.C.; Busquets-Mataix, J. A study on assessment results in a large scale Flipped Teaching Experience. In Proceedings of the 4th International Conference on Higher Education Advances (HEAD’18), Valencia, Spain, 20–22 June 2018; pp. 1039–1048.
- Jordán, J.; Valero, S.; Turró, C.; Botti, V. Recommending Learning Videos for MOOCs and Flipped Classrooms. In *Advances in Practical Applications of Agents, Multi-Agent Systems, and Trustworthiness*; Demazeau, Y., Holvoet, T., Corchado, J.M., Costantini, S., Eds.; The PAAMS Collection; Springer International Publishing: Cham, Switzerland, 2020; pp. 146–157.
- Klašnja-Miličević, A.; Ivanović, M.; Nanopoulos, A. Recommender systems in e-learning environments: A survey of the state-of-the-art and possible extensions. *Artif. Intell. Rev.* **2015**, *44*, 571–604. [[CrossRef](#)]
- Burke, R. Hybrid Recommender Systems: Survey and Experiments. *User Model. User-Adapt. Interact.* **2002**, *12*, 331–370. [[CrossRef](#)]
- Herlocker, J.; Konstan, J.; Terveen, L.; Riedl, J. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.* **2004**, *22*, 5–53. [[CrossRef](#)]
- Rodríguez, P.; Heras, S.; Palanca, J.; Duque, N.; Julián, V. Argumentation-Based Hybrid Recommender System for Recommending Learning Objects. In *Multi-Agent Systems and Agreement Technologies*; Springer International Publishing: Cham, Switzerland, 2016; pp. 234–248.
- Bobadilla, J.; Serradilla, F.; Hernando, A. Collaborative filtering adapted to recommender systems of e-learning. *Knowl.-Based Syst.* **2009**, *22*, 261–265. [[CrossRef](#)]
- Zapata, A.; Menéndez Domínguez, V.; Prieto, M.; Romero, C. A Hybrid Recommender Method for Learning Objects. *Int. J. Comput. Appl. (DEDCE)* **2011**, *1*, 1–7.

16. Chen, W.; Niu, Z.; Zhao, X.; Li, Y. A Hybrid Recommendation Algorithm Adapted in E-Learning Environments. *World Wide Web* **2014**, *17*, 271–284. [[CrossRef](#)]
17. Tarus, J.K.; Niu, Z.; Yousif, A. A hybrid knowledge-based recommender system for e-learning based on ontology and sequential pattern mining. *Future Gener. Comput. Syst.* **2017**, *72*, 37–48. [[CrossRef](#)]
18. Vanitha, V.; Krishnan, P.; Elakkiya, R. Collaborative optimization algorithm for learning path construction in E-learning. *Comput. Electr. Eng.* **2019**, *77*, 325–338. [[CrossRef](#)]
19. Dwivedi, P.; Kant, V.; Bharadwaj, K.K. Learning Path Recommendation Based on Modified Variable Length Genetic Algorithm. *Educ. Inf. Technol.* **2018**, *23*, 819–836. [[CrossRef](#)]
20. Dwivedi, P.; Bharadwaj, K.K. e-Learning recommender system for a group of learners based on the unified learner profile approach. *Expert Syst.* **2015**, *32*, 264–276. [[CrossRef](#)]
21. Stoica, A.S.; Heras, S.; Palanca, J.; Julian, V.; Mihaescu, M.C. A Semi-supervised Method to Classify Educational Videos. In *Hybrid Artificial Intelligent Systems*; Springer International Publishing: Cham, Switzerland, 2019; pp. 218–228.
22. Turcu, G.; Heras, S.; Palanca, J.; Julian, V.; Mihaescu, M.C. Towards a Custom Designed Mechanism for Indexing and Retrieving Video Transcripts. In *Hybrid Artificial Intelligent Systems*; Springer International Publishing: Cham, Switzerland, 2019; pp. 299–309.
23. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
24. MLLP Research Group (Universitat Politècnica de València). TLP: The transLectures-UPV Platform. Available online: <https://www.mllp.upv.es/tlp/> (accessed on 14 May 2021).
25. Salton, G.; Buckley, C. Term-weighting approaches in automatic text retrieval. *Inf. Process. Manag.* **1988**, *24*, 513–523. [[CrossRef](#)]
26. Richardson, M.; Dominowska, E.; Ragno, R. Predicting clicks: Estimating the click-through rate for new ads. In Proceedings of the 16th International Conference on World Wide Web, Banff, AB, Canada, 8–12 May 2007; pp. 521–530.
27. Martineau, J.; Finin, T. Delta tfidf: An improved feature space for sentiment analysis. In Proceedings of the International AAAI Conference on Web and Social Media, San Jose, CA, USA, 17–20 May 2009; Volume 3.
28. Ren, F.; Sohrab, M.G. Class-indexing-based term weighting for automatic text classification. *Inf. Sci.* **2013**, *236*, 109–125. [[CrossRef](#)]
29. Kansheng, S.; Jie, H.; Liu, H.T.; Zhang, N.T.; Song, W.T. Efficient text classification method based on improved term reduction and term weighting. *J. China Univ. Posts Telecommun.* **2011**, *18*, 131–135.