Educational Video Recommender System

Mohamed Timmi[1,*], Loubna Laaouina[2], Adil Jeghal[2], Said El Garouani[3], and Ali Yahyaouy[3]

[1]Faculty of Sciences Dhar El Mehraz, Sidi Mohamed Ben Abdellah University, Morocco
[2]LISA Laboratory, National School of Applied Science Sidi Mohamed Ben Abdellah University Fez, Morocco
[3]LISAC Laboratory, Faculty of Sciences Dhar El Mehraz Sidi Mohamed Ben Abdellah University Fez, Morocco
Email: mohamed.timmi@usmba.ac.ma (M.T.); loubna.laaouina@usmba.ac.ma (L.L.); adil.jeghal@usmba.ac.ma (A.J.); saidelgarouani@yahoo.fr (S.E.G.); ali.yahyaouy@usmba.ac.ma (A.Y.)
[*]Corresponding author
Manuscript received August 6, 2023; revised August 30, 2023; accepted September 20, 2023; published March 4, 2024

***Abstract*—In recent years, informal education has witnessed a significant upsurge, fueled by technological advancements and the ubiquitous availability of online educational content. Internet users, including students, researchers, and teachers, are increasingly seeking supplementary educational resources across diverse online repositories to augment their knowledge. Within this landscape, recommendation systems emerge as indispensable tools, aiding users in the discovery of pertinent resources aligned with their academic interests. This article proposes a novel recommendation methodology leveraging a hybrid approach, incorporating both Content-Based Filtering (CBF) and Collaborative Filtering (CF) algorithms. By harnessing information from a myriad of data repositories, this system excels in identifying and presenting the most relevant and desirable educational resources, with a particular focus on meeting the needs of students. This holistic approach embraces user profiles, contextual information, and supplementary data, underscoring its potential to revolutionize informal education in the digital age.**

***Keywords*—recommendation systems, hybrid filtering, e-learning, informal education**

I. INTRODUCTION

The increasing amount of data, and the growing use of the Internet have given rise to a new form of learning: informal education [1].

Various repositories are now utilized to search for educational content, be it for academic purposes or other. Content of the sort is thereby created, distributed, and presented in a variety of formats to cater for the needs of people with different profiles and interests, including students, teachers, and researchers.

Technologies that promote social interaction are developing rapidly, with new forms of communication, work, study, and entertainment being used by billions of people leading to the storage of an enormous amount of diverse information [2], namely audio, video, and image, as well as emails, newsgroups, blogs, wikis, and social networks [3].

With all these different tools and content spread across multiple repositories, it’s a complex task to know exactly what the users need. Recommender systems attempt to solve this problem by presenting users with resources that may be of interest to them based on knowledge extracted from historical data and recommended resources [4, 5].

With the knowledge of relevant information and the user’s interests, it has become possible to recommend items that meet the needs of a specific user or group of users. On this basis, this study advances previous work that uses the content-based filtering and collaborative filtering algorithms to recommend relevant educational resources to students in various fields.

Therefore, the purpose of this research is to suggest an alternative method to help with recommending educational resources based on the CBF and FC algorithms. Encouraging this method will lead to the development of educational resource recommendation systems that enable users to obtain content that matches their research interests.

II. RELATED WORK

A. The Main Recommendation Algorithms

Recommender systems typically fall into three categories: Content-based filtering, collaborative filtering, and hybrid filtering. The current work uses these approaches to make recommendations based on user/item properties and the information content extracted from them.

Recommender systems are a paramount area of research owing to the massive availability of practical applications that help users process the large amount of information that exists on various platforms. These systems recommend content, people, and services of interest to users on an individual or group basis.

Instances of applications may include recommendations for various products on shopping sites, such as Amazon.com or Ebay.com; videos on Youtube.com; songs on Spotify; people on Facebook and LinkedIn [6].

1) Content-based filtering

This type of filtering describes users and items based on their characteristics [7]. Defining characteristics that describe an item or user can be done by making allowance for descriptive information such as genre, media type, and duration, among others. Alternatively, semantic information is obtained through information extraction techniques to identify implicit characteristics of items and users [8].

After this characterization, descriptions are compared to check the relationships between them. An item is then inferred as related to a user if they share similar attributes.

The description of a user’s major interests can be obtained from the information provided by his or her actions in searching for resources. One way of working with this type of filtering is to ask the user to evaluate a set of items with different characteristics. After the evaluation, the system considers the relevant items to be similar to the items that the user evaluated. In this way, the system considers the items for which the user has not shown interest as irrelevant [9].

2) Collaborative filtering

This technique is based on the assumption that users who

doi: 10.18178/ijiet.2024.14.3.2058

have shown similar interests in the past will share common interests in the future. This filtering approach differs from content-based filtering in the sense that it does not require a description of the items to be recommended, based solely on the similarity between users [8, 9].

Recommendation systems that use collaborative filtering [10] have collections of user-assigned ratings for articles, and user-rated articles indicate that the article is relevant to the user's needs. This allows users to receive recommendations based on ratings provided by users with shared interests.

3) Hybrid filtering

The hybrid filtering approach seeks to combine the two recommendation techniques [11] described above to minimize the failures of each. There are different ways of combining content-based and collaborative methods in a hybrid recommender system, these ways are classified as follows [9, 12]:

1) Implementing collaborative and content-based filters separately and then combining their recommendations.
2) Integrate some content-based filtering features into the collaborative approach.
3) Integrate some collaborative filtering features into the content-based approach.
4) Build a model that unifies the features of collaborative and content-based filters.

B. Some Proposals for Recommender Systems

In this paper, some proposals for recommender systems related to the suggested work will be presented. Some relevant researches are outlined, which serve as a basis for this study, and others are considered as a trend for the system presented in this paper.

1) A health information recommendation system

Rivero Rodriguez *et al*. proposed a health recommendation system to suggest reliable videos, using information from renowned health video channels. They reliably made available information-enriched videos from YouTube video data and from a service offered by the US National Library of Medicine called "Medline Plus".

They applied four methods to generate the recommendations and evaluate their results. The method that performed best was a combination of two existing processes. The authors admit that their method requires improvements in terms of metadata enrichment to enhance the quality of the recommendations [13].

2) The YouTube video recommendation system

Abbas and All used a hybrid recommendation approach to provide YouTube video recommendations. However, the paramount limitation associated with this approach is that the recommended videos are not necessarily relevant to the user's present context. It is very common for the same user to follow different interests depending on the context they are in. A recommendation system is proposed for YouTube to keep track of a user's multiple interests and recommends videos following only the current context [14].

3) A recommendation system for open educational videos based on required skills

Tavakoli *et al*. [15] proposed a new method to help learners find open educational videos related to mastering a set of chosen skills. They built a prototype that can function as follows:

1) applies text classification and text mining methods on job ads to match jobs descriptions including their required skills.
2) predicts the quality of videos.
3) creates an open educational video recommendation system to suggest personalized learning content to learners.

As a result of the performance, more than 250 videos were recommended, and 82.8% of these recommendations were considered useful by the interviewees. Furthermore, the interviews revealed that their personalized video recommendation system has the potential to improve the learning experience.

4) YouTube Recommendation Network (YRN)

Qin *et al*. [16] proposed recommender system which ranks YouTube videos based on information extracted from users' social networks, especially those with video reviews. Users write about videos that interest them. In this way, recommender system allows for a wider range of recommendations compared to YouTube [17], which are limited to topics watched and tags of user interest. We used the YouTube API to integrate features such as searching for videos and collecting data about videos and users that can be incorporated into our system.

This proposal is interesting in that it induces users' interest by recommending not only the topic of the video they are watching but also the topics that they are likely to be interested in. YouTube recommendation network does not consider quality; its purpose is simply to recommend videos that users might be interested in. However, if a user searches YouTube for themselves and utilizes a keyword such as "conjugate math", YouTube will suggest videos that are already significant based on the number of views, yet there is no way of knowing whether the video has quality in terms of didactics, audio, or image. YRN takes all of this for granted, focusing only on the most comprehensible recommendations.

5) The video content summarization for recommendation system

The paper "An Automatic Multimedia Content Summarization System for Video Recommendation", by Yang *et al*. [18], proposes to recognize the subtitle of a video from a module called Optical Character Recognition (OCR) and thus summarize its content. The proposal is to facilitate the automatic inclusion of video summaries in collections, in addition to recommending videos based on their compatibility to the user's profile. In Fig. 1, it is possible to verify the architecture of the Video Content Summarization for Recommendation (VCSR) system project.

As soon as a new video is received, the OCR engine recognizes the subtitles and generates a text document. These documents are passed through a compression engine that uses keyword extraction to generate video concept notes. Finally, the engine generates video proposition emails and infers relevance based on each student's profile. By combining these three modules, the system can automatically generate recommendations and send video proposition emails when new videos arrive. In other words, the VCSR system process is automated, and doesn't require human intervention.

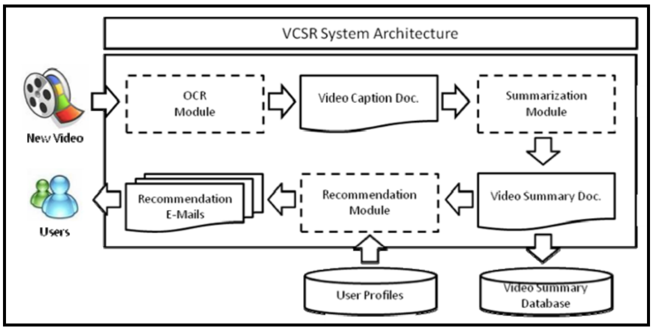

Fig. 1. VCSR system architecture [18].

The work proposed in this paper differs from the VCSR system in many regards:

- VCSR summarizes videos and sends suggestions via email, but there seems to be a substantial lack of connection among the videos.
- The objective of this paper is to use the YouTube API to obtain subtitle files.

6) Comparison between the recommendation systems

It can be noted that, although VCSR uses video recommendation, the technique used for subtitle extraction involves image text recognition; this is not necessary when using a modern video repository which facilitates access to subtitles. Even videos that aren't equipped with subtitles can be easily seen considering the speech recognition feature built into the algorithmic structure of the platform., YouTube as an example. In fact, YouTube is a very popular tool for many educational channels. Many students make use of it to study or clarify doubts if any occur. Thus, the proposed work addresses the video extraction functionality of YouTube.

All of the available works have a repository previously populated with educational content. However, the suggested work stands out in terms of using algorithms to learn how to qualify a video.

Each proposal is interesting depending on its context of application, but the present work brings a more modern vision and application to users all over the world since YouTube is now accessible everywhere. This proposal also presents another convenience that is predicated on the use of hybrid recommendation algorithm that combines the benefits of content-based recommendations and the benefits of opinion-based recommendations.

III. Methodology

The methodology of this work is organized as follows:

1) Bibliographical research to identify the literature related to the theme addressed.
2) The proposal of a recommender system method based on CBF and FC algorithms.
3) The development of a prototype based on the proposed method.
4) The evaluation of the recommendations generated by an online survey form was submitted to a target group (teachers).
5) Discussions: A bibliographic review of key issues and concepts related to the problem and its solutions is performed to provide a rationale for the study. A conceptual method is then proposed, followed by a prototype development. The prototype was divided into two main steps. In the first phase, information is extracted from selected data sources and an algorithm based on FBC and FC is conceived to create a recommended list of educational resources.

IV. Proposed System

The general structure of this method serves as a reference for implementation, following YouTube data extraction, data classification, and preprocessing, which are indispensable to generating high-quality recommendations. It is also noteworthy that this method revolves around the use of certain techniques, such as content-based filtering and user rating-based techniques. Finally, its details of implementation, advantages and disadvantages are introduced. Furthermore, usage scenarios are carefully described to illustrate the applicability of the solution to the predefined requirements.

A. Overview

Since the main purpose of the suggested system is to recommend meaningful instructional videos and ensure their educational usefulness, it is important to consider its educational and technical aspects. These latter are going to be covered in this article, but for now, it's possible to get a general idea of how the developed system operates without delving into the internal components of the recommender system.

The user can utilize the system through a web application, which will provide a search bar. Fig. 2 illustrates the general overview of the suggested system.

Step A: The user sends a query—for instance, "the affective filter hypothesis" in the field of language learning and applied linguistics. This information is going to be transferred to the Web Application Server.

Step B: The system searches YouTube for videos related to the subject matter in order to extract information and store it for later use.

Step C: The system retrieves information about user ratings, classifies and preprocesses the data obtained from the videos.

Step D: The system uses recommendation techniques, combining the information retrieved in elements B and C to generate meaningful recommendations for educational videos based on the topic of the query sent by the user.

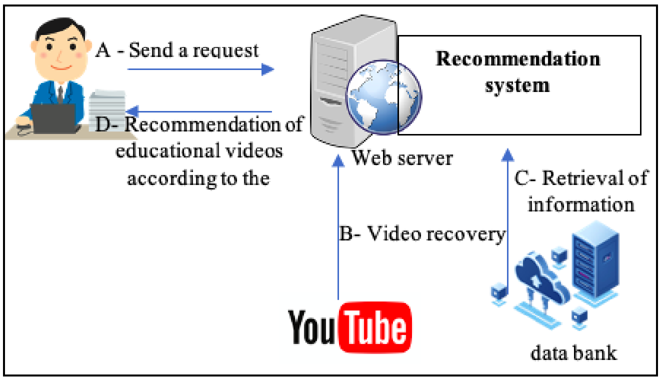

Fig. 2. Overview of the recommendation system.

The proposed solution is a web-based system consisting of an application that ranks YouTube videos on specific topics and uses that information as input to generate video suggestions related to educational resources that can be of potential utility to the user. Essentially, the generated recommendations help users tediously select quality educational videos and minimizing the time spent searching for suitable material. The benefits of this solution are not only limited to learners but also teachers may benefit from additional learning materials.

The developed solution uses the YouTube search engine to display videos as if the user were on the YouTube website itself, allowing them to perform identical actions, such as starting or pausing videos, fast-forwarding, or rewinding, and searching for videos by keywords or phrases.

The process of recommending educational videos starts when entering the keyword of a subject matter and, from there, the proposed application takes care of carrying out certain procedures through its modules, as shown in Fig. 3.

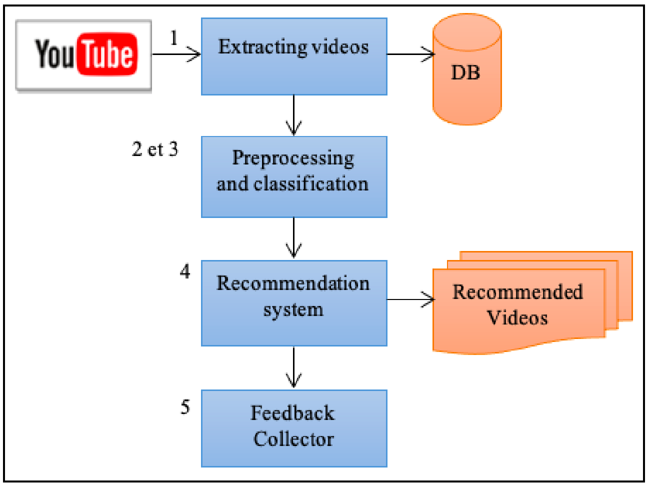

Fig. 3. Steps followed by the solution.

More details related to the modules mentioned in Fig. 3 will be further illustrated in the following sections. Below is an overview of the modules and their procedures:

1) The Video Retrieval module collects information about YouTube videos associated with terms and keywords entered by the user. This information is stored in its database for future use.
2) The pre-processing module is in charge of processing the previously extracted data. It works as a filter, selecting the relevant words and processing the data to be sent to the classification module and the recommendation engine.
3) The classification module employs the data obtained from YouTube to classify the videos, considering a set of relevant attributes put in place by the machine learning algorithms.
4) The recommendation engine is responsible for calculating the similarity between the searched topic and the classified educational videos. Moreover, it executes the recommendation algorithms in the light of the user's interests, following a wide array of characteristics and similarities, so that the process of video recommendation can begin.
5) The feedback collector is responsible for collecting user feedback regarding the generated recommendations.

The collection of user feedback is of fundamental importance to improve the machine learning algorithm used in the classification phase. In this way, the solution can evolve into a powerful educational tool, which increasingly contributes to the learning process of its users.

B. *Video Extraction*

The first step is to collect information about videos on YouTube and store them in a local database for classification and preprocessing. The video search method consists of a simple procedure where the system presents a search bar, and the user is required to employ relevant keywords to successfully execute the search.

This solution uses the YouTube Google's API, which allows this YouTube-specific functionality to be integrated into another website or application. To interact with the API, an authentication ID has to be obtained through the Google Developers Console available at https://console.developers.google.com/ [19]. There are three types of credentials available. In this work, the chosen one was "API Key".

A list of API retrievable resources is provided. However, this step required the use of the search method, which puts forward results featuring various information including channels, videos and playlists. in order to use these resources, it is necessary to complete the prerequisite methods. In the present work, the list method was used, which returns a set of search results matching the query parameters specified in the API request [19, 20]. By default, the search result set identifies videos, channels, and corresponding playlist assets. However, the user can also configure the query to retrieve only certain types of assets.

To make the HTTP request, the get method is used via the URL, https://www.googleapis.com/youtube/v3/search, and it is necessary to fill in a literal parameter called PART, which specifies a comma-separated list of one or more search resource properties that will be included in the API response. The items that can be included in the part parameter value are ID and SNIPPET [20, 21].

Each video on YouTube is uniquely identified by an eleven-character literal type, called "ID". In the feilds parameter, it is possible to specify all the information that should be returned, such as the title, description, image with default size, and the title of the channel. In addition, the max results parameter ought to be defined, which limits the number of videos returned [22].

Note that the created key must also be entered in the key field. When using data directly from YouTube, storing information in a specific database is indispensable to ensure the development of the work, as the existing dynamics of the tool could change this information. Additionally, some of the collected data must be processed and pre-processed to obtain meaningful results.

In addition to the ID, it is mandatory to capture other information that the resource list method "SEARCH" doesn't provide. Thus, it was thought necessary to make another call to the YouTube API for each video to collect more data about a given video. In order for this to happen, it was necessary to fill in the video ID as a parameter, as well as define the field part with the desired data, as shown in Fig. 4.

```
vidId = item['id']['videoId']
r = youtube.videos().list(
    part="statistics,contentDetails",
    id=vidId,
    fields="items(statistics," + \
                "contentDetails(duration))"
).execute()
try:
    duration = r['items'][0]['contentDetails']['duration']
    views = r['items'][0]['statistics']['viewCount']
    likes = r['items'][0]['statistics']['likeCount']
    favorites = r['items'][0]['statistics']['favoriteCount']
    comments = r['items'][0]['statistics']['commentCount']
    cats_info['id'].append(vidId)
    cats_info['duration'].append(duration)
    cats_info['views'].append(views)
    cats_info['likes'].append(likes)
    cats_info['favorites'].append(favorites)
    cats_info['comments'].append(comments)
except:
    pass
```

Fig. 4. Details of the videos.

In this way, it was possible to extract other information such as date of publication, channel ID, category ID, tags, duration, subtitle indicator, number of views, number of likes, and number of dislikes [23, 24].

1) The date of publication is important, as a video published a long time ago may contain outdated information and the system may deduce this, preventing the video from being recommended.
2) Other important information such as the channel and category also make sense, as the category can only filter out educational videos and show that certain channels are reliable and can be considered safe sources.
3) Another important property is a "tag", which is widely used in other recommendation systems to infer similarity between videos.
4) The length of the video can provide implicit information about how a topic is elaborated, whether it is a summary of a given topic or a more detailed video [24].
5) The "caption" indicator informs whether the video has subtitles, a feature that will be used in future work, in which it is planned to access the text of the subtitles to transform them into keywords to increase the precision between the similarity of the videos to be recommended and the topic of interest.
6) From the number of views, we can deduce that a video has attracted the interest of users. In order for this assumption to be confirmed, the following properties, "number of likes" and "number of dislikes", will be very useful [24].

In this part, the intention is to extract information from YouTube videos, storing their status in the database. Some of the information obtained at the time of extraction will probably change over time (number of likes, views, etc.), but other steps can be taken to update the information in the local database and process it in future work.

C. Pre-Processing and Classification of Videos

As described in the previous section, the data was stored in the local database as returned by the YouTube API without any pre-processing. However, the video classification and recommendation processes require this data to be prepared in a process-oriented manner. For instance, to compare a research topic and a question video, the title should be edited to remove meaningless characters such as periods, spaces, and other meaningless special characters, as well as words with no relevant utility. The result of this rigorous process is that all meaningful words were saved in a table.

Another operation that is performed is the processing of certain information such as the "Duration" property recorded as an example in the "PT21M14S" format. P indicates the duration (years, months, weeks, or days) and the time portion starts whenever T indicates that the video is 21 minutes and 14 seconds long. The way this information is returned makes it difficult to classify the video since it constitutes two separate pieces of information within one field [25]. Pre-processing extracts durations and assigns them to classes to store times in more consistent data types. Another piece of information that needs to be managed is the tags or keywords within the video. These are returned as multiple-word literals separated by commas. Since these tags are often repeated across videos, it is necessary to treat these terms separately and explicitly to consider their usage and improve similarity calculations between terms within videos.

After pre-processing the data, it is up to the machine learning algorithm to classify, through a predictive model, whether the video is educational and whether it can be considered qualitatively sound or not. For this purpose, a hundred videos were used, containing evaluations made by specialists and students who had already seen the subject "conjugated expression in maths" in their curriculum. It is also possible to rely on the "Category" property returned by the YouTube API. The latter has defined 15 categories [26], one of which is entitled "Educational", designating that the content is thoroughly educational.

D. Recommendation System Algorithm and Results

After removing videos deemed inappropriate because of their educational inappropriateness and mismatch, or that they do not meet minimum quality standards. In this sense, the selected videos are sent to our recommendation engine. This latter is responsible for taking a list of approved videos and ranking them based on their relevance to a particular user. In this proposal, two techniques are combined to improve the recommendations and mitigate possible shortcomings of using one technique alone, CBF and FC. According to Bouazza *et al*. [11], the most appropriate classification for this proposal would be a mixed approach, in which the techniques used would be presented in a single list.

The algorithm utilized to generate the recommendations is called "hybrid", known as the only mixed-type algorithm based on two techniques CBF and FC, which are executed parallelly as shown in Fig. 5.

The video received by the recommendation engine is sent to the CBF and FC algorithms simultaneously. Each algorithm provides its recommendation list yet follows the function that recommendations appearing on both lists take precedence in the final recommendation list. Therefore, each item in both lists is given a score calculated as the sum of its position in the source list.

The final list is organized in an ascending order. The lower the rating of the video, the further to the left it is placed in the final list. Videos that do not exist t in both lists at the same time are added to the end of the list according to the following rules: One item is selected from the collaborative filter list and one item from the content-based filter list, depending on its position in the original list.

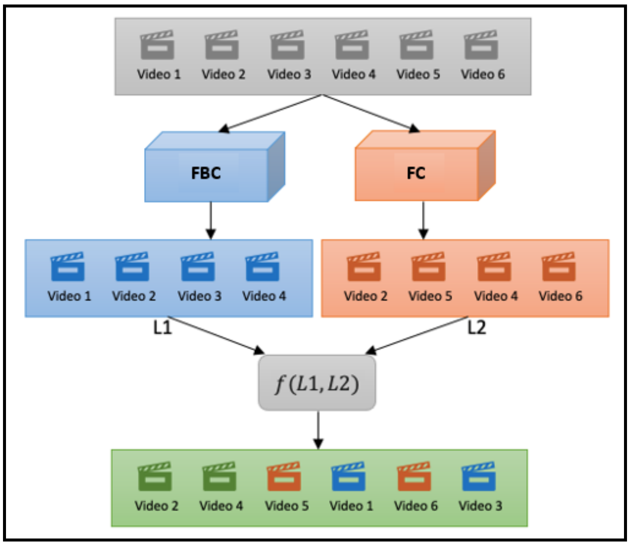

Fig. 5. Fusion algorithm applied to generate video recommendations.

In Fig. 5 shown above, we can see the following example: Appearing 2nd on the CBF list and 1st on the FC list, "Video 2" he scored 3 (3 = 2 + 1) and ranks 1st on the final list has the lowest score. "Video 4" is his second element in LF, with a score of 7 (7 = 4 + 3). These two items were the only items featured on both lists. After that, "Video 5", which was just to the left of the collaborative filtering list, was selected next. The next item will be "Video 1". It's placed further to the left of the content-based filter list, and so on.

Collaborative filtering technology based on exchanges of opinions between users was used to create the L2 recommendation list. It implements the "word of mouth" principle that people have always used to form their opinions about videos. The steps for this technique are: First, collect user feedback on the videos they watch. The second is to integrate this information into the user's profile. Third, it makes use of that profile to help users find the following information.

One of the main drawbacks of this approach is the first reviewer problem, also known as "cold start", because new users without reviews cannot be compared to other users. As a matter of fact, this proposal employs the YouTube video repository; it provides data that can be used to generate recommendations without prior review.

Since YouTube is a platform visited by millions of users who frequently interact with videos and rate them as either good or bad, characteristics, such as the number of plays, number of likes, and dislikes are collected so that relevant information can be obtained. Essentially, this procedure serves an insightful purpose of providing Key Demographic Information. In different terms, if the users were the same or had insufficient amount of ratings, we can recommend using this data extracted from YouTube videos. The following formula was used for this:

$$\text{Weighted Average} = [(\%\ \text{Likes} * 1) + (\%\ \text{Views}) * 2]\ /3 \qquad (1)$$

YouTube videos may have an exaggerated number of views and a low number of likes. This paradox can be attributable to the fact that the audience may have disliked the content or the quality of the video. Taking this fact to heart, it is necessary to examine the total number of people who rated the video, then calculate the percentage of likes from the total number of likes and dislikes. Afterwards, the percentage of video views was calculated based on the sum of all video views. Finally, a weighted average of the values was calculated, with 1/3 weighting on the percentage of likes, and 2/3 weighting on the percentage of views, as shown in Eq. (1).

It was concluded that the number of views should be given more importance than the number of likes, considering that a newly published video may not have many likes.

As shown in Table 1, it is possible to verify that video V3 has only 7 views and that 5 people liked it. One solution to the problem of mediocre ratings is to use a way to assign a "weight" to the similarity coefficients, so that if the number of rated videos is too small, this weight will reduce the average value.

Table 1. Fusion algorithm applied to generate video recommendations

Videos	**V1**	**V2**	**V3**	**V4**
Number of Likes	2530	5142	5	17
Number of Dislikes	12	19	0	3
Number of Views	55641	96840	7	45
Likes	0.99527931	0.99631854	1	0,85
Views	0.36478008	0.63487901	0.00004589	0.00029502
Weighted Average	0.57494649	0.75535885	0.33336393	0.28353001

Content-based filtering techniques, often based on automated content analysis techniques, namely feature extraction, text analysis, and similarity comparison, were used to create L1 recommendation lists. These techniques help process your content and identify similar or related elements. For now, we used video titles and descriptions, but in the future, it is planned to use subtitle content as well to get better results for the most frequent words in a video content.

Algorithm: hybrid filtering in recommendation system

```
Begin
    # Collaborative Filtering
    user_sim = cal_user_sim ()
    # Calculate similarities between users
    user_pref = get_user_pref () # Get user preferences
    # Content-Based Filtering
    item_sim = cal_item_sim ()
    # Calculate similarities between items

    # Hybrid Filtering
    def hybrid_rec(user_id):
        # Collaborative Filtering recommendations
        cf_rec = col_fil_rec (user_id, user_sim, user_pref)
        # Content-Based Filtering recommendations
        cb_rec = con_based_fil_rec (user_id, item_sim)
        #Combine recommendations
        hybrid_rec = combine_rec (cf_rec, cb_rec)
        return hybrid_rec

    #Collaborative Filtering recommendations
    def col_fil_rec (user_id, user_sim, user_pref):
```

```
        sim_users = find_sim_users (user_id, user_sim)
        rec = gen_rec_from_sim_users (sim_users, user_pref)
        return rec

    # Content-Based Filtering recommendations
    def con_based_fil_rec (user_id, item_sim):
        user_pref = get_user_pref(user_id)
        rec = gen_rec_based_on_pref (user_pref, item_sim)
        return rec

    # Combine recommendations
    def combine_rec (cf_rec, cb_rec):
        # Apply weights or other techniques
        # to combine the recommendations
        combined_rec = cf_rec + cb_rec
        return combined_rec
    # Usage
    Print (hybrid_rec (123))
End.
```

E. Comment Collector

A comment collector is part of one of the final steps in the proposed recommender system. Once the videos show why the video was suggested, it's time to quickly rate the recommendations. Relevant users can rate the video by marking a star from 1 to 5. This rating is used as input for joint recommendations to analyze whether the recommendations are satisfactory.

F. The Interface of the Proposed System

The interface displayed in Fig. 6. seems to be the homepage of a video recommendation system tailored for educational resources, possibly named "RecSysVideos". The layout is designed to showcase various educational videos that align with the user's interests. Here's a description of what is seen on the interface:

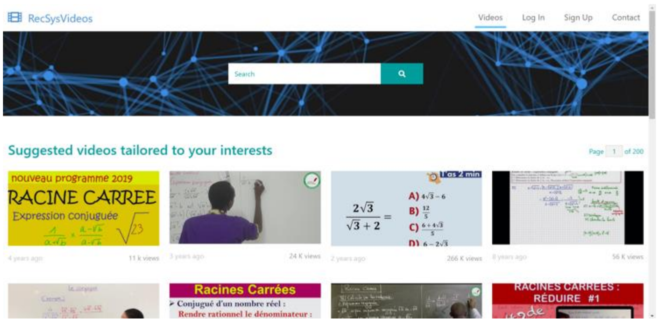

Fig. 6. The interface of the proposed system.

Navigation Bar: At the top of the page, there's a navigation bar that includes options for "Videos", "Log In", "Sign Up", and "Contact", indicating a user-friendly design that encourages interaction and navigation through the site.

Search Functionality: There's a prominent search bar in the center of the header, suggesting that users can search for videos based on specific queries.

Title of the Section: Below the search bar, the title "Suggested videos tailored to your interests" implies that the system personalizes video suggestions based on the user's browsing habits or specified preferences.

Video Thumbnails: The main area of the page displays thumbnails of suggested videos, each with distinct titles, view counts, and publication dates, allowing users to gauge the popularity and relevance of the content at a glance.

Pagination: At the bottom right, there's a pagination indicator showing "Page 1 of 200", which signifies an extensive library of videos, giving users the option to browse through a large number of pages to find the content they need.

V. Discussions and Evaluation

A. Evaluation

Before the development stage of the proposed system, an online survey form was sent to a group of 75 people, including trainees from the Regional Center for Education and Training in Meknes and teachers from educational institutions in the Taunate Provincial Directorate, Morocco, where experiments were conducted. It is designed to help you understand how the YouTube Platform has been made use of, including its usage characteristics. As shown in Fig. 7, we found that a high percentage of respondents use YouTube daily.

Q1: How often do you use YouTube?

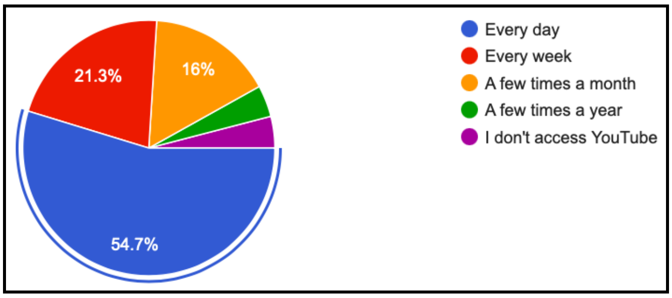

Fig. 7. Answers on the frequency of YouTube utilization.

When asked whether the research audience used YouTube for educational purposes, it was noted that almost half of the respondents used it at least weekly for this purpose, as shown in Fig. 8.

Q2: How often do you use YouTube for mere educational purposes?

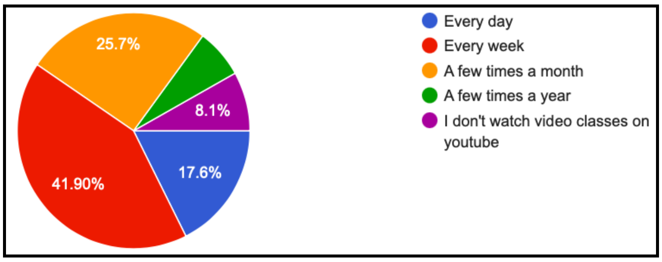

Fig. 8. Answers on the use of YouTube for educational purposes.

The remaining information retrieved relates to the main characteristics users count on when choosing a video. The most important features are the picture and sound quality, followed by the video title and cover image. The initial step in evaluating the suggested system was to search for videos directly on the YouTube platform, known for its diverse collection of videos and its search engine. It is important to emphasize that user authentication is not performed during the initial search so as not to affect user preferences. For example, searching for a term like "conjugation", which is typically related to the field of mathematics, suggested several videos, including a video on the conjugation system in chemistry, a video lesson on conjugating verbs in French, and a video on conjugating expressions in mathematics at the bottom of the list. This shows how hard it is to find educational videos with a sufficient amount of information.

Any information that is of no relevant utility to the user is called noise.

It is possible to verify that the same search was performed, using the same term, but this time with user authentication, significant results on the theory of the expression “conjugate in math” were displayed. It is noticeable that the video that was previously ranked at the bottom of the list has been ranked among the first.

This leads to the conclusion that YouTube continues to search for data such as actions that indicate that a user has watched a video. This data is used to improve our recommendations, but noise can still occur, and users should seek out videos that solve their problems. One way is to lucidly select the search terms until you find the ideal video. Still, it may take some effort. Another possible alternative is to use a recommendation system that filters educational videos to remove content related to music, humor, or irrelevant educational information. Therefore, only the most meaningful videos are recommended based on topic and user ratings.

When using the YouTube API to search for videos, even when enriching searches with words sharing the same signification, the list of videos returned by the API is not satisfactory. There are many videos in other languages, videos related to other topics, which have nothing to do with the expression “conjugated”; needless to mention low-quality videos as well. This supports the idea that enriching the search with synonymous terms can help with facilitating the process. Query enrichment is an important step in the process, but it is not enough to find good result options, it is important to eliminate bad options in order to minimize the noise that eventually gets to the user. employing Thanks to the implementation of query enrichment, it was possible to notice that the YouTube API returned, in its 3rd position, a video on “conjugated”, which corresponds to the conjugated system of chemistry, and in 4th position a video on the conjugation of the French language, which, according to our assessment, is viewed as a noise and irrelevant. Considering the latter observation, we concluded that YouTube results can and should be improved in the educational context.

Based on these indications, the proposed system has been developed using artificial intelligence techniques to recommend meaningful educational videos related to the searched topic. As previously mentioned, the recommendation engine of this proposal uses the fusion algorithm to combine the “content-based filtering” and “ratings-based filtering” techniques; however, before presenting the results of this combined technique, it is possible to verify the recommendations generated by the CBF technique which were applied separately.

The system recommended videos with the highest possible degree of accuracy as far as the terms given are concerned, ignoring all sorts of YouTube user ratings. It was observed that the first recommended video is relatively recent, with only 1 like and 23 views. Despite the title informing that it is a “math conjugate” concept, the video’s audio content has noise, and it was added to YouTube on 20 November 2021. Thus, we can deduce that this would not be a good recommendation, given that with over a year of availability on the platform, the video has only received one like. We also notice that the first 9 recommendations have the term “conjugate”, thanks to the algorithm which compares the search terms to those found in the “title” and in the “tags”. Recommendations 3 and 4 have a good image and sound quality and are in the same category.

It can be seen that the video, which was in 4th place in the content-based filtering, now takes first place in the merge. This happened because when calculating the score of the technique, the CBF score was added (which was 4) to the FC score (which was 22), so that the score of the video was 26, considering that it was the lowest in the calculation made by the Fusion algorithm. It was also noticed that more videos from the same channel appeared in the next seven positions. These videos appear to have been professionally edited, with better sound and image quality, which may have led to the positive reviews they received. Although the videos that were in the 3rd and 4th positions did not have the phrase “conjugate math” in the titles or tags, the recommendations were relevant since the phrase “conjugate square root” is a complementary topic to “conjugate math”. Furthermore, they were the only suggestions in the top 10 that did not include the phrase “conjugate math “. The number of these recommendations would have been higher if a cut-off point for the CBF technical score had not been included. Scores below the cut-off, 0.30, were rejected and remained useless in the fusion technique. This resulted in more relevant and higher-scoring recommendations in the top positions.

In fact, there was also another recommendation on the “Conjugate system of chemistry” in the 10th position. This latter had absolutely nothing to do with the search terms, except for the word “conjugate”. This recommendation emerged because of the good ratings the video received and the fact that it had a good score in the CBF technique: 0.72. However, as it is a topic with a homonymous term, with identical pronunciation and spelling, but alien to the field, it is considered an irrelevant recommendation. Although it is presented in a remote placement as opposed to the former, it is still possible to use techniques to discard the homonymous terms found. One of these is obtained from the evaluations of the recommendations themselves. It is noteworthy that ontologies can also be used to solve the problem.

The evaluations presented below in Table 2 show that the proposed system has succeeded in obtaining a satisfactory acceptance rate by its users. In this sense, pre-processed lists were submitted and put online to a group of 21 users at the Regional Center for Education and Training Professions FES-MEKNES, Morocco, mostly trainee teachers in computer science, 17 in total, and 4 computer science trainers, and they were asked to analyze the quality of the video recommendations, taking into consideration the subject “conjugated math” and the noise (non-pedagogical videos, videos in other languages, videos on other subjects, videos of poor quality) presented. Four lists were generated, each with different characteristics:

1) List A: videos returned by YouTube when searching for the term “conjugate” without authentication.
2) List B: Videos returned by YouTube during the process of searching for the chosen term with authentication provided that the user has seen videos related to the topic.
3) List C: videos returned by the YouTube API, with the search terms enhanced by query enrichment.

4) List D: videos selected by the proposed system, using, in addition to query enrichment, the hybrid recommendation technique and YouTube ranking.

Table 2. The four lists generated, with different characteristics

List	Rejection rate	Users' opinions
A	80%	More effort to find an interesting video on this list.
B	75%	More effort to find an interesting video.
C	60%	Put effort into finding an interesting video.
D	15%	Users reported achieving their goal by choosing one of the first 3 recommended items.

In addition to the query enrichment applied in the last list, several techniques were used, including Artificial Intelligence, such as the fusion algorithm, which has been widely commented on in the hybrid filtering approach to recommender systems [12, 27]. These factors have certainly contributed to the positive reviews received. Among others, the positive reviews stand out in terms of classifying videos as "educational" or "non-educational", which succeeded in eliminating most videos. In addition, the use of other information about the video, such as language, duration, and YouTube user ratings, served to classify the videos and generate another pile of rejections, as was the case with videos in a language other than the user's language.

The remaining videos underwent two recommendation methods. The initial one is content-based filtering, which computes the similarity between the searched topic terms and the video's title and description, and then sorts the videos by relevance based on the processed content. The second method, collaborative filtering, executes calculations based on the ratings of YouTube users and returns a list of videos sorted by relevance based on their ratings. The final list prioritizes the videos that appear in both sets, while the remaining ones are included in an alternating manner according to their position in the generated sets.

B. *Discussion*

The comparison among recommendation systems sheds light on critical factors affecting their efficacy and suitability in the realm of educational video suggestions. In this context:

The choice between image text recognition, as used in VCSR, and the utilization of readily available subtitles in contemporary video platforms like YouTube is pivotal. Image text recognition may introduce complexity and potential inaccuracies, while platforms like YouTube offer user-friendly, precise subtitle access, significantly enhancing content accessibility for users with varying language preferences or hearing abilities.

The selection of a video repository holds great significance. YouTube's global presence, vast educational content library, user-friendly interface, speech recognition capabilities, and diverse content range render it an ideal choice for educational purposes, as emphasized in this work.

The incorporation of learning algorithms for video evaluation represents a forward-looking strategy. It empowers the system to adapt and enhance recommendations continually. In contrast, traditional recommendation systems reliant solely on predefined criteria may struggle to keep pace with evolving user preferences and content quality.

The adoption of a hybrid recommendation algorithm, amalgamating content-based and opinion-based approaches, strategically enhances recommendation accuracy. Content-based analysis ensures alignment with subject matter, while opinion-based recommendations factor in user preferences, increases user satisfaction and accommodating diverse interests and needs. In conclusion, this discussion underscores the importance of embracing modern platforms like YouTube and advanced technologies, which streamline processes such as subtitle extraction and offer a more engaging user experience. Furthermore, integrating learning algorithms and hybrid recommendation strategies demonstrates a proactive response to the dynamic nature of educational content and user preferences, affirming the commitment to improving the accessibility and quality of educational video recommendations for a global audience.

VI. Conclusion

In conclusion, this article focuses on the development and implementation of a recommendation system for educational videos on the YouTube platform. The system utilizes content-based and collaborative filtering algorithms to generate personalized recommendations based on user-selected themes and video evaluations. The findings demonstrate the effectiveness of incorporating YouTube user ratings and the native language of videos in improving recommendation accuracy. The contributions of this work include the implementation of a versatile recommendation system that can be applied to various application domains. The system's architecture and functionality serve as a reference for other applications, and its rest services enable easy integration and consumption. The user evaluations indicate a satisfactory acceptance rate, validating the system's usefulness in assisting users in finding quality educational videos.

While the developed system shows promising results, there are opportunities for future enhancements and research. These include sentiment analysis of user comments, identification of reliable video channels, addressing the challenge of unrelated videos, and utilizing video subtitles to improve content-based filtering. Further exploration of video metadata and the use of machine learning algorithms for quality analysis are potential avenues for future investigation.

This article presents a successful implementation of a recommendation system for educational videos on YouTube. The system's capabilities provide valuable assistance to users in discovering relevant and high-quality educational content. The findings contribute to the field of recommendation systems and offer insights for improving the recommendation process in the context of informal education. Future work can build upon these findings to refine and expand the system, benefiting both learners and educators in their pursuit of educational resources online.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

The authors collectively contributed to all aspects of this

research endeavor. The conceptualization of the study was a collaborative effort led by M.T., L.L., A.J., S.E.G., and A.Y. Methodology, software development, and validation were meticulously executed by M.T., L.L., A.J., S.E.G., and A.Y. Formal analysis and investigation were equally shared among the authors, including M.T., L.L., A.J., S.E.G., and A.Y. The allocation of resources and data curation tasks was a joint undertaking, with contributions from M.T., L.L., A.J., S.E.G., and A.Y. The initial draft of the manuscript was prepared collaboratively by M.T., L.L., A.J., S.E.G., and A.Y., and subsequent review and editing were conducted by the entire team. Visualization efforts were shared among M.T., L.L., A.J., S.E.G., and A.Y. Supervision and project administration responsibilities were collectively managed by M.T., L.L., A.J., S.E.G., and A.Y. Additionally, M.T., L.L., A.J., S.E.G., and A.Y. played pivotal roles in securing funding for the research project. All authors had approved the final version.

REFERENCES

[1] Z. Shen, M. Pritchard, S. Tan, and C. Noteboom, "Educative sensemaking on social media: An empirical investigation of informal learning on YouTube," in *Proc. the 55th Hawaii International Conference on System Sciences*, 2022.

[2] M. Timmi, L. Laaouina, A. Jeghal, S. Garouani, and A. Yahyaouy, "Recommendation system based on user trust and ratings," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 7, 2022.

[3] Y. Deldjoo, M. Schedl, P. Cremonesi, and G. Pasi, "Recommender systems leveraging multimedia content," *ACM Computing Surveys (CSUR)*, vol. 53, no. 5, pp. 1–38, 2020.

[4] B. Li *et al.*, "A personalized recommendation framework based on MOOC system integrating deep learning and big data," *Computers and Electrical Engineering*, vol. 106, 108571, 2023.

[5] B. Hui, L. Zhang, X. Zhou, X. Wen, and Y. Nian, "Personalized recommendation system based on knowledge embedding and historical behavior," *Applied Intelligence*, pp. 1–13, 2022.

[6] S. G. K. Patro, B. K. Mishra, S. K. Panda, and R. Kumar, "Fuzzy logics based recommendation systems in e-commerce: A review," in *Proc. 2nd International Conference on Smart Computing and Cyber Security: Strategic Foresight, Security Challenges and Innovation (SMARTCYBER 2021)*, 2022: Springer, pp. 107–120.

[7] M. R. Julianti, Y. Heryadi, B. Yulianto, and W. Budiharto, "Recommendation system model for personalized learning in higher education using content-based filtering method," in *Proc. 2022 International Conference on Information Management and Technology (ICIMTech)*, 2022, IEEE, pp. 1–6.

[8] S. A. Thorat, G. Ashwini, and M. Seema, "Survey on collaborative and content-based recommendation systems," in *Proc. 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2023, IEEE, pp. 1541–1548.

[9] L. Zhou, F. Zhang, S. Zhang, and M. Xu, "Study on the personalized learning model of learner-learning resource matching," *International Journal of Information and Education Technology*, vol. 11, no. 3, pp. 143–147, 2021.

[10] R. Kuo and S.-S. Li, "Applying particle swarm optimization algorithm-based collaborative filtering recommender system considering rating and review," *Applied Soft Computing*, 110038, 2023.

[11] H. Bouazza, B. Said, and F. Z. Laallam, "A hybrid IoT services recommender system using social IoT," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 8, pp. 5633–5645, 2022.

[12] B. Walek and P. Fajmon, "A hybrid recommender system for an online store using a fuzzy expert system," *Expert Systems with Applications*, vol. 212, 118565, 2023.

[13] R. Mervin, D. Mavalaru, and T. Thomas, "Knowledge-driven and intelligent computing in healthcare," *Handbook of Intelligent Healthcare Analytics: Knowledge Engineering with Big Data Analytics*, pp. 167–188, 2022.

[14] M. Abbas, M. U. Riaz, A. Rauf, M. T. Khan, and S. Khalid, "Context-aware Youtube recommender system," in *Proc. 2017 international conference on information and communication technologies (ICICT)*, 2017, IEEE, pp. 161–164.

[15] M. Tavakoli, S. Hakimov, R. Ewerth, and G. Kismihok, "A recommender system for open educational videos based on skill requirements," in *Proc. 2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT)*, 2020, IEEE, pp. 1–5.

[16] S. Qin, R. Menezes, and M. Silaghi, "A recommender system for youtube based on its network of reviewers," in *Proc. 2010 IEEE Second International Conference on Social Computing*, 2010, IEEE, pp. 323–328.

[17] J. Sultana, "A survey on YouTube recommendation system," 2020.

[18] J. C. Yang, Y. T. Huang, C. C. Tsai, C. I. Chung, and Y. C. Wu, "An automatic multimedia content summarization system for video recommendation," *Journal of Educational Technology & Society*, vol. 12, no. 1, pp. 49–61, 2009.

[19] A. A. Vera and C. M. Gonz ález, "Educational Resource Recommender Systems Using Python and Moodle," in *International Conference on Computational Science and Its Applications*, 2022, Springer, pp. 15–30.

[20] R. Ilieva, Y. Yonkov, and A. Tonchev, "Building independent models of YouTube content networks as a tool for monitoring health misinformation," in *Proc. 2022 XXXI International Scientific Conference Electronics (ET)*, 2022, IEEE, pp. 1–6.

[21] T. Alrashed, J. Almahmoud, A. X. Zhang, and D. R. Karger, "Scrapir: making web data apis accessible to end users," in *Proc. the 2020 CHI conference on human factors in computing systems*, 2020, pp. 1–12.

[22] S. Garg, Saurabh, and M. Breja, "Social network analysis of YouTube: a case study on content diversity and genre recommendation," in *Computational Methods and Data Engineering: Proceedings of ICMDE 2020, Volume 1*, 2021, Springer, pp. 25–37.

[23] M. Kurdi, N. Albadi, and S. Mishra, ""Think before you upload": An in-depth analysis of unavailable videos on YouTube," *Social Network Analysis and Mining*, vol. 11, no. 1, p. 48, 2021.

[24] A. C. Munaro, R. Barcelos, E. C. Maffezzolli, J. P. Rodrigues, and E. Paraiso, "To engage or not engage? The features of video content on YouTube affecting digital consumer engagement," *Journal of consumer behaviour*, vol. 20, no. 5, pp. 1336–1352, 2021.

[25] R. F. Alhujaili and W. M. Yafooz, "Sentiment analysis for youtube videos with user comments," in *Proc. 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, 2021, IEEE, pp. 814–820.

[26] A. R. Shaikh, H. Alhoori, and M. Sun, "YouTube and science: models for research impact," *Scientometrics*, vol. 128, no. 2, pp. 933–955, 2023.

[27] S. Sharma, V. Rana, and M. Malhotra, "Automatic recommendation system based on hybrid filtering algorithm," *Education and Information Technologies*, pp. 1–16, 2022.

Copyright © 2024 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).