Honours in Computer Science

Personalized Video Recommendation System

Written by Marlie Russell

Table of Contents

Abstract

This study was conducted to develop a personalized recommender system for YouTube videos. The system initially collects and ranks YouTube videos on educational classroom topics and displays them to students. Once students using the system interact with the displayed videos by liking and disliking the videos, the system combines the user history with collaborative filtering to produce a personalized list of recommended YouTube videos. The system is meant to be a supplementary tool for students to use to help them learn important classroom concepts. The system was tested out on a first year computer science course. However, the users did not interact very much with the system and few gave feedback. Despite that, the feedback was mostly positive, and we believe that this project has valuable opportunity for growth. Little work has been done for developing personalized recommender systems for supplementary education. This research project begins the groundwork of developing a tool that could have an incredible positive impact on the education of the student population.

1 Introduction

The purpose of this study was to use learning analytics software to develop and test a new collaborative filtering recommender system. Others have done similar work in this field such as Hassan Khosravi [6], the developer behind RiPPLE. RiPPLE is an online learning platform which allows students and instructors to create learning resources then recommends personalized resources to students. RiPPLE's learning resources are created by the user and can include videos, text, formulas and images. Our project narrows the focus to creating personalized YouTube video recommendations for students learning course topics since our personal experience indicates that students often spend large amounts of time using YouTube videos to supplement their in-class learning.

The method used in this system is different from the method used by YouTube to recommend videos. YouTube recommends videos based on users liking similar videos, video popularity and a user's viewing history. Our algorithm takes classroom context into consideration and weighs YouTube video statistics with classroom feedback and classmate user similarity. YouTube videos were collected and ranked based on video views, likes, dislikes and the probability that their comments had an overall positive sentiment. As users interact with the system through liking and disliking videos, the system makes personalized recommendations to the users. The idea is to offer students a video recommendation system that is personalized to both the student's viewing history and their classroom. User-personalized recommendations are important since each user is unique in their learning style. Classroom-personalized recommendations are also important since each classroom has a unique learning context which all the students in the class contribute to, so if one student in the class found a video helpful it is likely that other students will as well.

To assess the utility of our prototype, we ran a pilot study in a first year CS2 course in February 2020. Students in the class were asked to evaluate the system's effectiveness in helping them supplement classroom learning. Very few students participated in the study and even fewer gave feedback on the system. However, the overall feedback sentiment was positive, and there is promising future work for this project.

2 Literature Review

General areas of research relevant to this project are determining a resource's credibility (specifically a video resource), using collaborative filtering to generate effective recommendations, and analyzing sentiment in comments to predict a video resource's quality. All of these areas can be combined to make an effective video recommender system for students which accounts for user similarity, source credibility and user feedback.

2.1 Examining Endorsement and Viewership Effects on the Source Credibility of YouTubers

In [1], S. Fred analyzed the effects that the brand has on YouTubers and whether the amount of views a YouTube video has affects a viewer's perception of the YouTuber's trustworthiness and expertise. The study is based on the source credibility theory which uses variables of perceived expertise, trustworthiness and attractiveness to evaluate the source's credibility. In effect, the researcher's purpose is to determine how endorsement and viewership (number of views) interact to affect a YouTuber's source credibility. The study has relevance to my research since it evaluates the source credibility of YouTubers. However, it differs in that this study analyzes the credibility of beauty vloggers based on how many views they have and if they are endorsed by a brand whereas my study focuses on the credibility of educational videos based on views, ratings and comment sentiment. The author is a Masters of Arts in Mass Communication student with a concentration in Strategic Communication Management at the University of South Florida. She has three supervising Ph.D. professors.

316 females throughout the US participated in the study. They were given one of four questionnaires after watching a makeup tutorial featuring Dulce Candy. The survey and measurable scales were adapted from Ohanion's (1990) source credibility theory. Each participant watched the same video. However, one-quarter of the participants received the video with the views lowered and the brand endorsement segment of the video was excluded. One-quarter of the participants received the video with the views lowered and the brand endorsement segment was included. One-quarter of the participants received the video with the correct high view count and the brand endorsement segment excluded. One-quarter of the participants received the original video with the correct high view count and the brand endorsement segment included. So, the four video variations that were shown to study participants are as follows:

1) Non-brand endorsed-low viewership stimulus

2) Non-brand endorsed-high viewership stimulus

3) Brand endorsed-high viewership stimulus

4) Brand endorsed-low viewership stimulus

This study removed the factor of attractiveness with the justification that since the study is done on a beauty video where the vlogger is beautifying herself, so the factor of attractiveness may be excluded from testing. Instead, the researcher should have included more questions on attractiveness in part of the survey (one question was included) and analyzed the viewer's responses with respect to the variable of attractiveness since attractiveness may have swayed the results. Some viewers may have thought Dolce Cande to be attractive while other viewers may not have thought the same.

The study and survey may be adopted for evaluating user feedback on the credibility and trustworthiness of YouTube-sourced educational videos suggested by the recommender system I am building. Since most educational videos do not show a face, but rather use illustrations and animations to explain a topic, attractiveness will likely be excluded from the credibility mode, as with this study.

2.2 Multidimensional Credibility Model for Neighbour Selection in Collaborative Recommendation

In [2], Kwon et al developed a multidimensional model for effectively recommending neighbouring items to consumers based on a Collaborative Filtering (CF) model. The model factors in personalized consumer importance weights of credibility attributes (source expertise, trustworthiness, and similarity to the target consumer). The model uses calculations for source expertise, trustworthiness and similarity, which are then weighted by importance to select the most valuable neighbours for a user then generate a recommendation. CF methods automate “word-of-mouth” recommendations, and many students use “word-of-mouth” to recommend which videos to watch for explaining concepts in computer science (for example, how a binary tree is structured). The personalized and multidimensional CF model outlined in this study can be adapted to fit the recommender system in my software to effectively recommend educational videos based on what other students have found useful.

2.3 Extracting Attributed Verification and Debunking Reports from Social Media

In [3], S. Middleton weighs the effectiveness of filtering tweets with a semi-automated approach in regard to fake news prediction. Through the use of language pattern processing to predict fake or genuine claims in tweets and comparing tweet sources to verified lists of trusted and untrusted sources, the technology featured in the study makes leaps in the precision of classifying tweets as fake or genuine. This has applications in my research by using the language patterns to analyze comments in educational videos for positive or negative sentiment to accurately predict whether a certain video is a good educational resource.

2.4 Effective Use of Knowledge Management Systems: A Process Model of Content Ratings and Credibility Indicators.

In [4], R. Poston and C. Speier investigate how ratings affect the search and evaluation of Knowledge Management Systems (KMS) content and the nature of human judgment to develop a system to help users find high-quality, credible content without feeling overwhelmed. Content credibility was indicated by the number of raters, rater expertise, and collaborative filtering. The study remarks that ratings are subjective and provided voluntarily, so they may not provide a valid impression of the content's quality. This is important in regards to my project since I will need to investigate the validity of video ratings, perhaps through analyzing comments or the publishing channel for credibility to provide users with high-quality content based on high-validity ratings. If I employ an internal rating system for system users to evaluate recommended videos, the study suggests that ratings from users with greater expertise (such as teacher assistants and professors) do **not** have higher credibility than ratings from users with lower expertise (such as the students learning the material); therefore, the validity of internal ratings need not be weighted depending on the expertise of the user.

3 Problem Formulation and Approach

For the purpose of this study, we restricted our scope to YouTube videos. YouTube is a popular search engine used by students for educational resources (as indicated from our personal experience). Using only YouTube videos also allows us to make clear comparisons between the YouTube recommendation algorithm and our own recommendation algorithm, and to limit the scope of our research.

3.1 Overall System Architecture

Based on what we learned from the studies outlined in chapter 2, we now propose another way of building a video recommendation system for educational supplementary purposes. Figure 1 illustrates the system architecture of our approach. The system will begin by querying YouTube for videos on a particular topic then filtering out any off-topic results. The system will then analyze the remaining videos to rank each video in terms of YouTube views, likes, and dislikes. This ranking acts as a raw ranking for videos for a user before the user has personally liked or disliked the videos within the classroom context. As users interact with the system by liking or disliking videos, the system finds similar users with cosine similarity and computes a recommendation rank for a video to a user. The system combines the raw video ranking with the recommendation rank to suggest videos for a particular user. The system updates the recommended videos nightly.

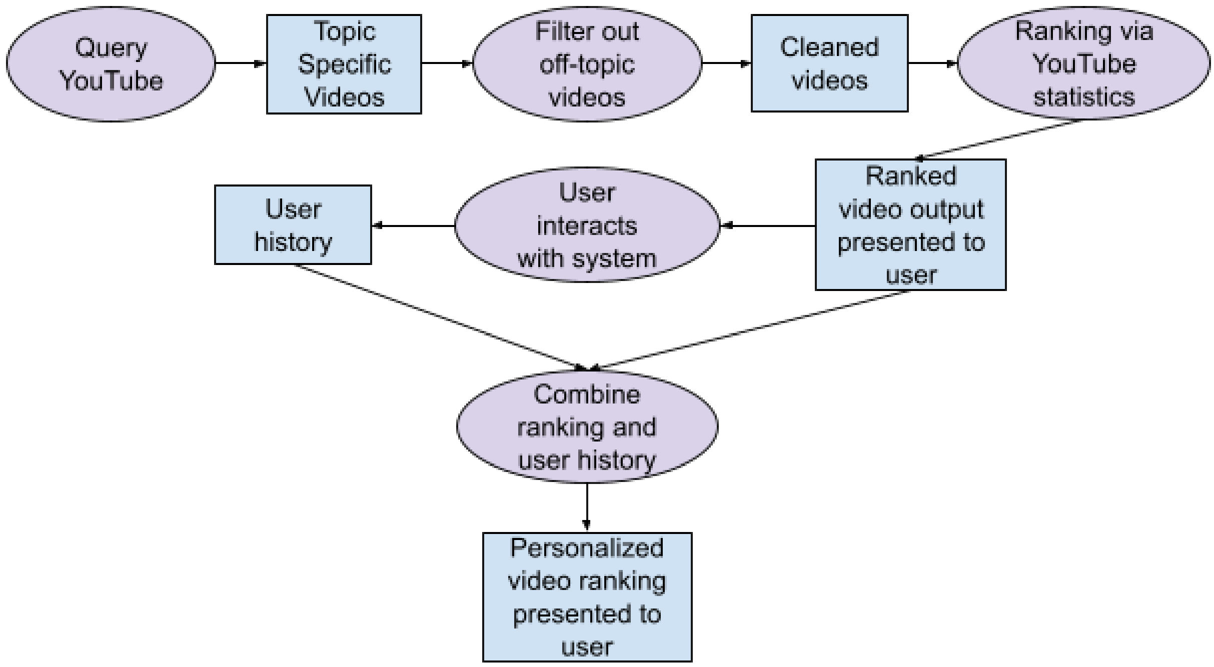

Figure 1. System architecture

As shown in Figure 1, the system first queries YouTube for topic-specific videos then filters out any off-topic videos. These cleaned videos are ranked using YouTube statistics and are presented to the user. The user interacts with the videos and the user's likes and dislikes are recorded in user history. The user history and video ranking are combined to produce a personalized ranking of videos for the user. The next few sections will explain some of the non-trivial processes in more detail.

3.2 Filtering Process

In the filtering process, videos collected from the YouTube query are cleaned to filter out any irrelevant videos. When we first queried YouTube for videos on a specific topic, we received many off-topic videos and videos in languages other than English (which is not ideal since the Computer Science courses taught at UBC Okanagan are all in English). We filter out off-topic videos by checking the title of each video with the specific query topic. If the query topic is not found within the video title, the video is filtered out. We filter out different language videos by specifying English as our relevance language and only retrieving results playable in Canada during the YouTube query. Running the YouTube query on the topic "Binary Tree", I retrieved 124 videos. We ran these videos' titles through a language detector API [7] which detected two video titles to be in a non-English language. One of the videos was correctly categorized in Chinese. The other one was English but incorrectly categorized as Scots, which is not accurate. We ran the Language Detection API again, but this time with both the video title and cleaned (punctuation removed, lemmatized and stemmed) video description as the text for language detection, and zero videos were detected to be non-English, which is not accurate either since one of the videos is Chinese. Due to this incorrect categorization, we chose not to use the Language Detection API to filter out non-English videos. Instead, the system should recommend those videos less and less as users interact with the system if it proves to be a video they do not like.

3.3 Ranking via YouTube Statistics

When new users sign on, they have no likes or dislikes of videos, so the new user will not receive any personalized recommendations. The problem was how to rank videos before a user rates anything so that they receive a list of quality videos. One option was to present videos in order of decreasing YouTube views so that the highest viewed video on a topic would be presented at the top of the list. Instead, I set the rating to be a combination of likes, dislikes, and views as presented in the equation below:

$$rank = \left(\frac{likes - dislikes}{views}\right) * log\left(\frac{views}{average\ views}\right)$$

The first part of the equation is the Feedback Ratio. Video feedback is represented by likes and dislikes. Dislikes are subtracted by likes so that if a video has more likes than dislikes, the outcome would be positive whereas if the video has more dislikes than likes, the outcome would be positive. A large disparity between the likes and dislikes would result in a more strongly positive or negative part of the equation. Dividing the feedback by a video's views gives a ratio of how many people interacted with the video to how many people viewed the video. For instance, if many people viewed the video, but few gave feedback with likes or dislikes, the ratio between feedback and views would be small whereas if many people gave feedback for a video after viewing it, the ratio would be large. Whether positive or negative, a large ratio gives a lot of valuable information about the quality of a video and if it should be included as a recommendation. The value of that first portion of the equation is always between negative one (-1) and one (1) since a video's feedback cannot exceed its views.

The next part of the equation is the View Weighting. Videos with more views than average receive a multiplier that will increase the strength of the Feedback Ratio while videos with fewer views than average receive a multiplier that decreases the strength of the Feedback Ratio. Average views represent the average amount of views for videos on a specific topic. Since some videos have upwards of a million views while other videos have only a handful of views, the View Weighting ranges from values in the hundredths to values in the hundreds. Applying a logarithm to the ratio of views to average views lessens the dramatic variety of view ratios and evens out the ranking.

If a video has zero views, the system simply sets the rank to zero since the rank equation cannot divide by zero, and a log can never equal zero.

3.4 User History and Collaborative Filtering

Once users begin interacting with the system and giving feedback on videos by liking or disliking them through the website, the system creates a rating matrix of users and videos. If a user likes a video, the rating matrix marks the corresponding cell with one (1). If a user dislikes a video, the rating matrix marks the corresponding cell with a negative one (-1). From this rating matrix, the system uses cosine similarity to determine user similarity based on the ratings that users give videos. To estimate missing ratings for a video from a user, the user's top ten most similar users are collected, and their similarities are weighted to add to one (1). If none of the similar users have rated the video, zero is returned as the user's predicted rating for a video. Otherwise, the system applies the following equation:

$$r_{u,v} = r_u + \Sigma_{k\ users} w_{u,k}(r_{k,v} - \overline{r_k})$$

Where $r_{u,i}$ is the predicted rating for a user, u, on a video, v; *k users* is u's k similar users where k equals ten (10); $w_{u,k}$ is the similarity weighting between u and k; $r_{k,v}$ is the rating given by k for v; and $\overline{r_k}$ is k's average ratings for videos. For each similar user, the system takes the similar user's rating for the video and subtracts the similar user's average rating then multiplies this value by the similar user's similarity weighting before summing this value with all the other similar users. The sum is added to the user's average rating which is returned as the predicted rating for a user on a video. Each predicted rating is between negative one (-1) and one (1).

3.4 Combine Ranking and User History

How to combine video ranking and collaborative filtering to recommend videos for users? For each user-video combination, the system collects the video rank and the collaborative filtering predicted rating. We shall call the video rank *r* and the collaborative filtering predicted rating *p*. The final combination which will be used to decide whether or not to recommend a video to a user will be called the recommendation weight *w*. The value of *w* is computed differently based on the returned values of *r* and *p* for a user-video combination. It is summarized in the table below.

	$p > 0$	$p < 0$	$p = 0$
$r > 0$	$w = p * r$	$w = p * r$	$w = r$
$r < 0$	$w = p + r$	$w = -(p * r)$	$w = r$
$r = 0$	$w = p$	$w = p$	$w = r$

Table 1. Determining Recommendation Weighting

Recall that *r* is simply an initial ranking of the video from YouTube data whereas *p* is a personalized predicted rating determined by user feedback for a video where the users are all within a classroom. For this reason, *p* is a more important measure of video quality than *r*. When both *r* and *p* are positive values, *w* is their product which is also positive. When *r* is positive and *p* is negative, *w* their product since the classroom feedback has determined that the video is not very good quality. This may happen when the video is popular on YouTube but it is off-topic for the specific classroom content so the personalized rating is negative while the video rank is positive. When *r* is negative and *p* is positive, *w* is the sum of *r* and *p* since the classroom feedback is positive about the video, but the quality of the video may not be very good. In this case, the video rank, r, is still considered for the recommendation weighting, w, since poor quality videos should not be as highly recommended even though the classroom approves of the video. When r and p are both negative, w is their negative product. The negative sign is added to the product since a negative multiplied by a negative is positive, but if both the video rank and predicted rating are negative, the recommendation weight should also be negative. If ever p is zero due to having no similar users, w is set to be r, the video's rank. If the opposite happens and r is zero due to having no views, w is set to be p, the predicted rating. If both r and p are zero, w is zero.

If the final recommendation weighting, w, for a user-video combination is greater than or equal to zero, the video is added to the user's list of personalized recommended videos for learning. When a user logs into the system, each topic is displayed in its own row with the videos sorted by decreasing recommendation weight. Users may also view all videos for a topic besides their recommendations.

This program is run nightly to provide users with new and better-personalized recommendations each day.

4 Nature of the Study

The study was conducted on students in a first year computer science class at UBC Okanagan, COSC 121. There were 227 students registered in the class. The class started January 6, 2020 and ended April 8, 2020. We created a website for users to interact with the system where they could receive ranked and recommended videos on course topics in their class. We presented our website to the class in the ninth week of lectures, giving them five weeks to use the system. The website only included new topics which the students had not learned yet but would be learning in lecture in the five weeks that they had the system.

4.1 System Implementation

The system was hosted on Heroku. The main page shows rows of videos, each row being a specific topic containing recommended videos for that topic as shown in Figure 2. The rows are scrollable, and the videos in each row are sorted by the personalized video ranking in descending order. If the user has not liked or disliked any videos, the videos in each row are sorted by the ranking via YouTube statistics in descending order. The “Videos By Topic” tab in the navigation bar lists all the topics, and the user can click on a topic to be directed to the section of the page that the topic’s row is.

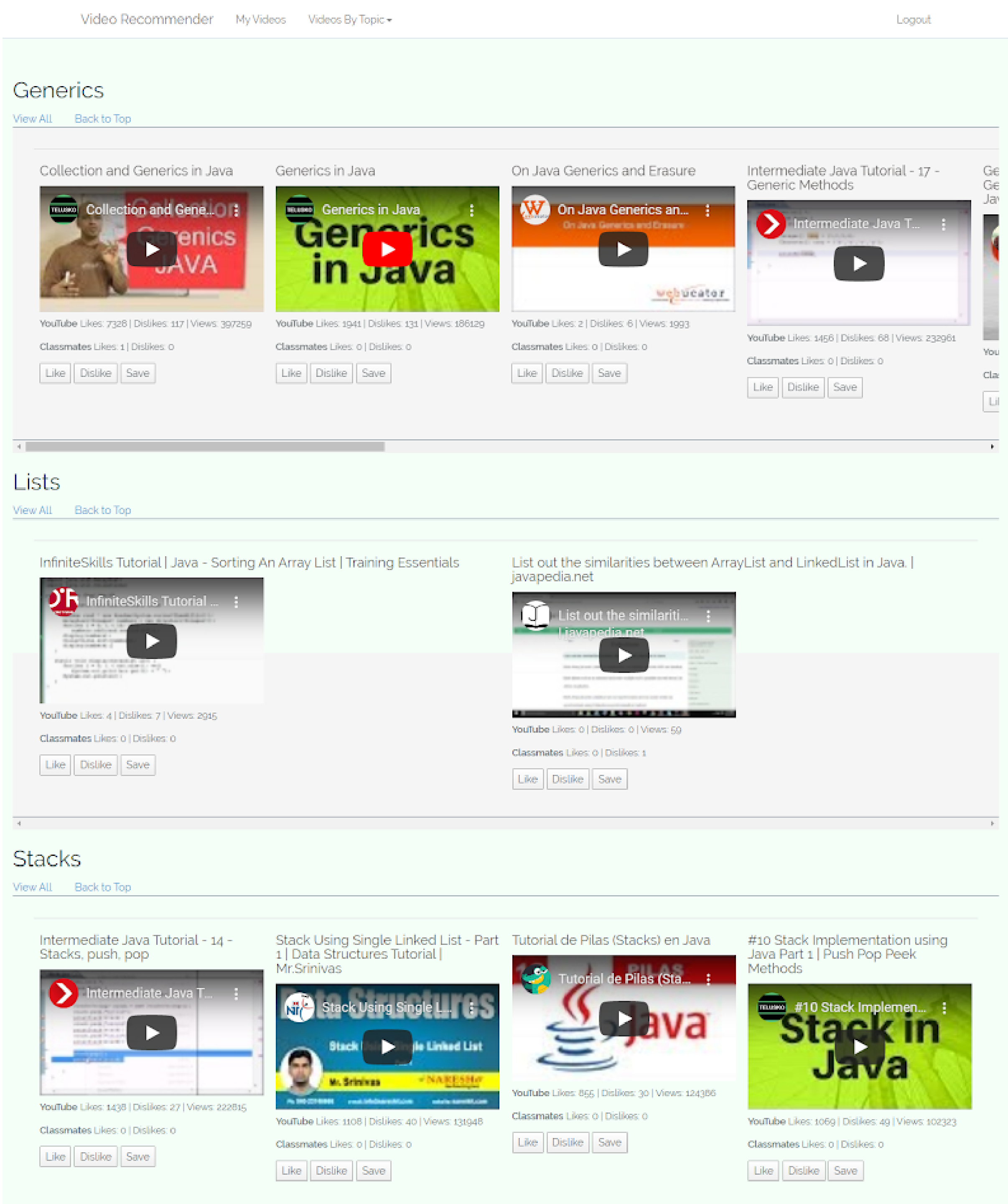

Figure 2. Website main page with topics in rows

The user may like, dislike or save any video. Clicking on "View All" for any video topic will direct the user to a page with all the videos on that topic as shown in Figure 3. These videos can be sorted by YouTube views descending or Recommended descending (which is the personalized rank).

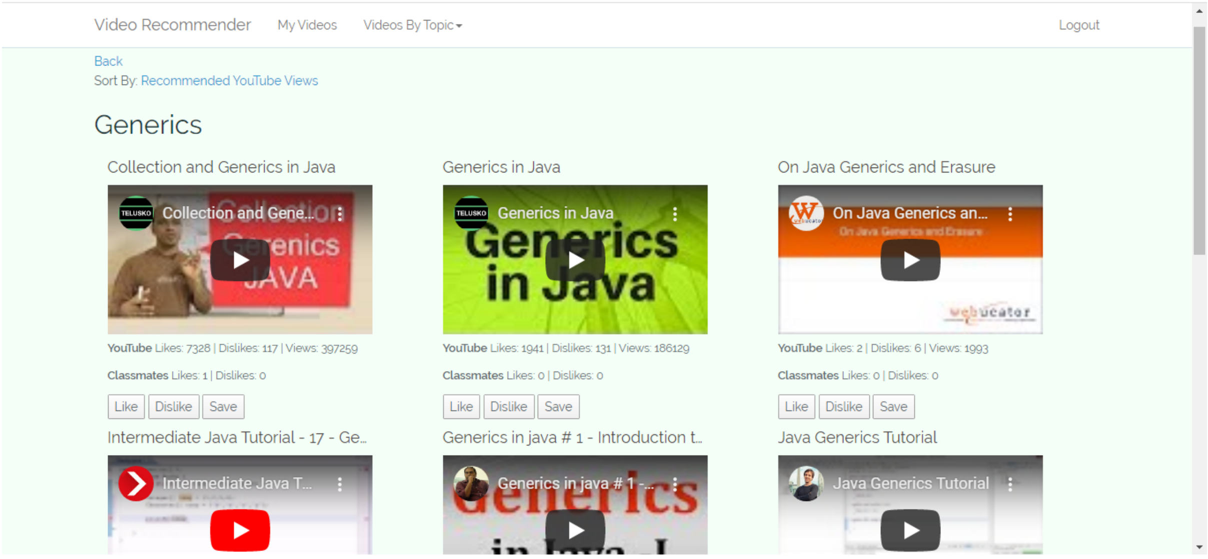

Figure 3. View all page for a topic

Saved video can be accessed under the tab "My Videos" as shown in Figure 3. From "My Videos", the user can like, dislike or un-save any saved videos.

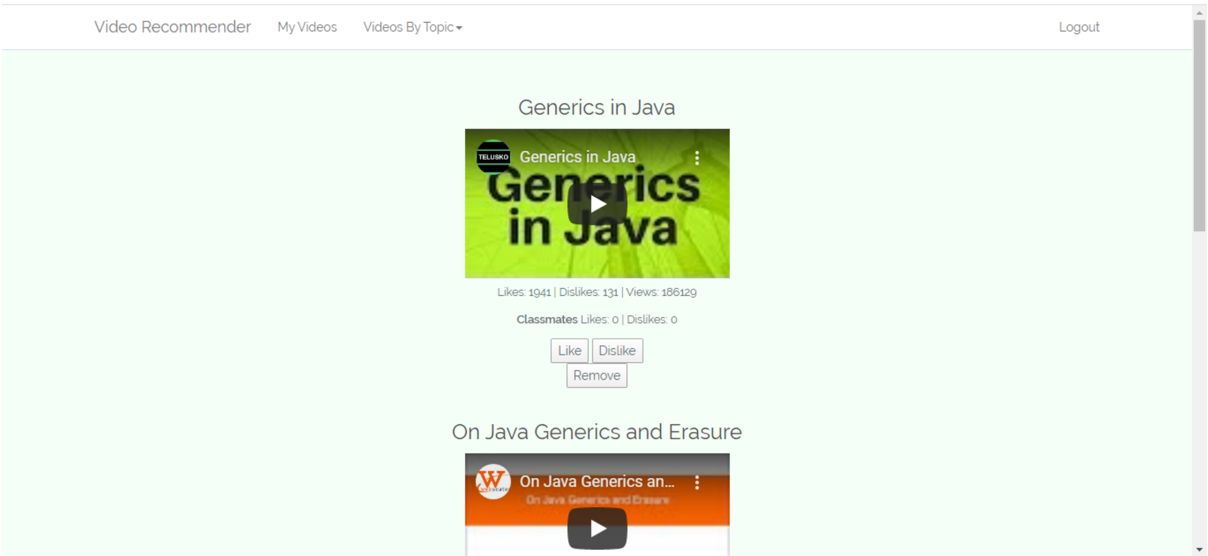

Figure 4. My Videos page

4.2 Post-Questionnaire

Students were asked to complete a short questionnaire at the end of the five week period to give feedback on the system. The questionnaire asked the students basic demographic questions as well as questions about the system and whether the system's recommendations improved over time. The questions are all listed in Table 2 in section 5.3.

5 Preliminary Results

5.1 Data Logs

Ideally, we hoped to collect user views, likes and dislikes for videos, user login history and how video recommendations for a user changed over time as they interacted with the system. We were able to collect user likes and dislikes for videos, but due to a shortage of time, we were not able to implement tracking user views and user login history. We were able to log how recommendations changed over time, but unfortunately, the recommendations never changed from the first day since no users made any new likes or dislikes throughout the trial period.

We had thought that there would be an increase in usage from the users when COVID-19 caused classes to go online. However, this prediction was not true in terms of new likes and dislikes for videos. One user suggested this could be due to assignments and labs being cancelled, so students had no need to learn new material.

5.2 General Usage

Thirty-three (33) users registered for the Video Recommender website. Of those, only three interacted with the system in the five week period, and all in the first day of using the system. There was not enough data to identify any trends or patterns with usage or recommendations. The personalized recommendation formula only generates new recommendations if there are new likes or dislikes, so no new video recommendations were made after the first day.

5.3 Post-Questionnaire Feedback

Of the thirty-three (33) students who registered for the website, only four (4) gave feedback via a UBC Qualtrics survey. The students ranged in age from 18 to 20 years old and were split between year 1 and year 2 students. Three students were male and one was female. The feedback from the students is listed in Table 2 below:

Question	Student 1	Student 2	Student 3	Student 4	Tally
1.) Please give your year of study	Year 1	Year 1	Year 2	Year 2	**Year 1:** 50% **Year 2:** 50%
2.) Please enter your age	18	19	20	19	**18:** 25% **19:** 50% **20:** 25%
3.) Please give your gender	Male	Female	Male	Male	**Male:** 75% **Female:** 25%
4.) A. Do you normally use videos as a supplementary source to lectures?	No	Yes	No	No	**Yes:** 25% **No:** 25%

4.) B. If so, which websites did you use?		Youtube			**YouTube:** 25%
5.) The video recommendation website was easy to use.	Strongly agree	Somewhat agree	Somewhat agree	Somewhat agree	**Strongly Agree:** 50% **Somewhat agree:** 50%
6.) In the first two weeks of using the video recommendation website, I found the suggested videos useful.	Neither agree nor disagree	Somewhat agree	Strongly agree	Somewhat agree	**Neither agree nor disagree:** 25% **Somewhat agree:** 50% **Strongly agree:** 25%
7.) The quality of the video recommendations improved over the five weeks.	Neither agree nor disagree	Neither agree nor disagree	Strongly agree	Neither agree nor disagree	**Neither agree nor disagree:** 75% **Strongly agree:** 25%
8.) I would use a video recommendation website for other courses.	Strongly agree	Strongly agree	Somewhat agree	Somewhat agree	**Strongly agree:** 50% **Somewhat agree:** 50%
9.) I would recommend other students taking this course to use the video recommendation website.	Somewhat agree	Strongly agree	Strongly agree	Somewhat agree	**Strongly agree:** 50% **Somewhat agree:** 50%
Other comments, questions, or suggestions for improvements:	Pretty good!				

Table 2. User Feedback Post-Questionnaire and results tally

As the results tally shows, the video recommendation website was fairly easy to use, according to the user feedback. User feedback also indicated that the ranking via YouTube statistics produced fairly useful suggestions, since the users gave positive feedback on the usefulness of the recommendations in the first two weeks of using the system. Three out of the four user feedbacks show that the recommendations did not become more or less useful over the course of the five week trial period. This can be accounted for by the lack of user interaction with the system over the trial period. The user feedback indicates an overall positive interaction with the system since they suggested they would use the system in other classes and recommend the system to other students taking COSC 121.

It should be noted that user feedback of only four students out of thirty-three participants does not provide indicative or conclusive results about the success or failure of the system. Further work should be done in the user-testing stage.

6 Conclusion

Through this research project, I learned the importance and application of personalized recommender systems and how they can be applied to education. Recommender systems are common in E-Commerce and media services, but there is little research done for using recommender systems in education. Although this project did not have the user interaction and user feedback that we desired, it was still a valuable learning experience in researching and developing a personalized YouTube recommender system.

Furthermore, this was a valuable opportunity to spearhead a research project for UBC focused on educational recommender systems for supplementary learning. The tool we developed can be adapted and used for future courses for online and remote support to students' learning. This is especially true with respect to the migration of education to online platforms in light of the COVID-19 global pandemic. The video recommender system can be an invaluable tool for UBC.

7 Future Work

Based on my experience with this project, I had hoped to implement tracking user logins and views on videos. However, due to time constraints, I was not able to do so. Future work should be done with more detailed tracking of user activity.

Future work should also be done in expanding the website. Currently, the website is specific to COSC 121 at UBC Okanagan and only has topics from the last half of the semester. The website could be scaled to be a multi-page site with users split into instructors and students. Instructors would be able to sign up and create a recommendation page for their class on topics of their choice. Students would use their class's page to access video resources. An instructor should be able to create multiple course pages. Administrators would also be required for managing instructors (deleting if necessary) and managing course pages.

Bibliography

[1] Fred, S. (n.d.). Examining Endorsement and Viewership Effects on the Source Credibility of YouTubers. 56.

[2] Kwon, K., Cho, J., & Park, Y. (2009). Multidimensional credibility model for neighbor selection in collaborative recommendation. Expert Systems with Applications, 36, 7114–7122. https://doi.org/10.1016/j.eswa.2008.08.071

[3] Middleton, S. (2015, September). Extracting attributed verification and debunking reports from social media: MediaEval-2015 trust and credibility analysis of image and video. 3. Retrieved from https://eprints.soton.ac.uk/382360/

[4] Poston, R. S., & Speier, C. (2005). Effective Use of Knowledge Management Systems: A Process Model of Content Ratings and Credibility Indicators. MIS Quarterly, 29(2), 221–244. https://doi.org/10.2307/25148678

[5] YouTube Data API. https://developers.google.com/youtube/v3/getting-started

[6] Khosravi, Hassan & Kitto, Kirsty & Williams, Joseph. (2019). RiPPLE: A Crowdsourced Adaptive Platform for Recommendation of Learning Activities.

[7] Language Detection API. https://detectlanguage.com/documentation.