Video Recommendation System Based on Human Interest

Shainee Jain
Computer Engineering
MCT's Rajiv Gandhi Institute of Technology
Mumbai, India
shaineejain11@gmail.com

Tejaswi Pawar
Computer Engineering
MCT's Rajiv Gandhi Institute of Technology
Mumbai, India
tejaswipawar66@gmail.com

Heth Shah
Computer Engineering
MCT's Rajiv Gandhi Institute of Technology
Mumbai, India
hethtimes@gmail.com

Omkar Morye
Computer Engineering
MCT's Rajiv Gandhi Institute of Technology
Mumbai, India
omimorye2006@gmail.com

Prof. Bhushan Patil
Computer Engineering
MCT's Rajiv Gandhi Institute of Technology
Mumbai, India
bhushan.patil@mctrgit.ac.in

***Abstract*—In today's world watching online videos have become a popular trend and a daily habit of our new generation. Videos are a reliable source for gaining knowledge and it is easier to grasp information through videos than reading. The internet is flooded with billions of videos hence it is a time consuming task for user to find a relevant video. So to save time as well as efforts there is a necessity to build a strong, efficient and accurate recommendation system which will display appropriate videos for the users. Video recommendation system saves users from browsing lots of videos to choose the appropriate ones, and on the other hand, it also brings the video websites more network traffic and user stickiness. The main task of the system is to provide personalized recommendations using Web Crawler, Rating Factor Neural Network, Slope one, and Slope one based Map Reduce of two types, one is Content-based filtering, and the other is Collaborative Filtering. Presentation of recommendations is an important part of the overall user experience. Video recommendation algorithm is the core of the system. The proposed paper is about the system which allows user to search for their favorite videos and the system recommends videos relevant to their choice.**

Keywords— webcrawler; rating; data set; recommendation;

I. INTRODUCTION (*HEADING 1*)

The explosive growth of the world-wide web and the rapid development of information technology and the Internet, the increasing performance of network facilities has resulted in continuously increasing network traffic. Video integrates such as images, audio, and animation becoming the most embracing and the most accessible form for people to gain information on the Internet. Compared with other media formation, video requires relatively higher network resource but the current Internet is no longer a bottleneck.

Typically, a wide range of media and e-commerce firms such as news websites, video providers and also social networking websites, provide data (hereafter referred as "content") on the Internet and their primary goal is to generate revenue. Not only, Content providers tend to maximize their revenue through advertisements and subscriptions but also try to reduce the cost of content distribution. Hence the providers distribute their contents across several geographical locations and also to improve and understand user experience, special analytical services would be used. Social media applications that deliver contents are completely dependent on the Users and hence make them deliver the best possible quality with minimum cost. At the same time, Content providers will now have the ability to collect, store and analyse behavioural patterns from Users.

Users are proactively engaged in integrating content information with their social information giving rise to social networking sites. Social networking sites such as Facebook, Twitter, etc., completely depend on individual users for content generation. Each of these social networking sites are Single-Platform based. With an increase in social networking sites, Single-Platform has a limitation where significant user interests are always left behind.

Similarly, studying through online courses and through videos available on Internet has become a common practise among the current youth as it is easily available and it can be used anytime and anywhere. The biggest issue with studying through videos is that there are hundreds of videos available to for each and every topic and it is difficult to select appropriate video for that topic and to know which one can best explain that topic. Also in most of the search engines there are many advertisements and non-related recommendations which might just distract the viewers.

Hence there needs to be search engine properly dedicated to videos for students studying in college so that they can waste minimum time searching for the right information and they can get it all in one place.

II. System Architecture

A. Layer Architecture

The section studies the layered architecture of the video recommendation system.

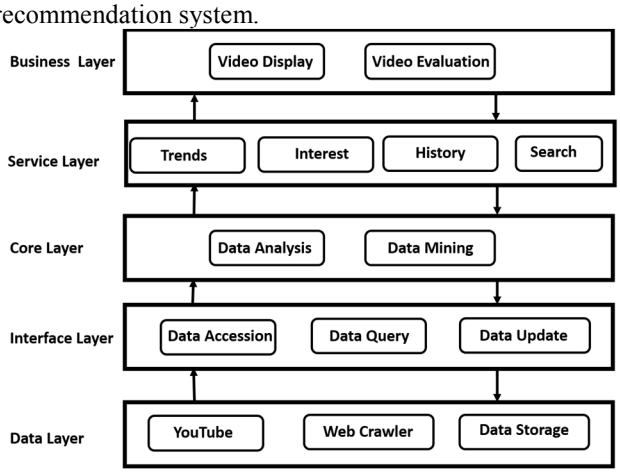

Fig. 1. System Architecture

In fig. 1. The data layer provides the basic data for system and stores the data. Data is generally collected from YouTube search engine using the web crawler which extracts all relevant information related to the video.

The interface layer provides the corresponding data accession, data processing methods according to different data sources and formats. Query processing and updates to the database are performed in this interface.

The core layer supports the core modules of the entire system, analysis of the data and data mining is performed at this stage. Data mining is the key technology for the video recommendation system.

The service layer provides the following services:

- Search: The highest rated video based on the rating factor which means the currently popular videos and depending on the user's search interest are extracted. The user can choose the most appropriate video.
- History: The service provides the number of videos viewed by the user in the past or current session.
- Interest: This service evaluates the interest in particular field as specified by the user and determines the top rated videos based on the analysis of its characteristics and the market performance of the video in the past.
- Trends: This service provides similar content videos to the user, so as to provide users with more relevant information and the trending ones.

The business layer includes delivering the video to the user and evaluation. With the help of evaluation the system decides the priorities of the user and the relevance of the video.

III. Design Structure Of the System

A. Class Diagram

Class Diagram is a type of static structure diagram in UML that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among objects. The class diagram is the main building block of object-oriented modeling.

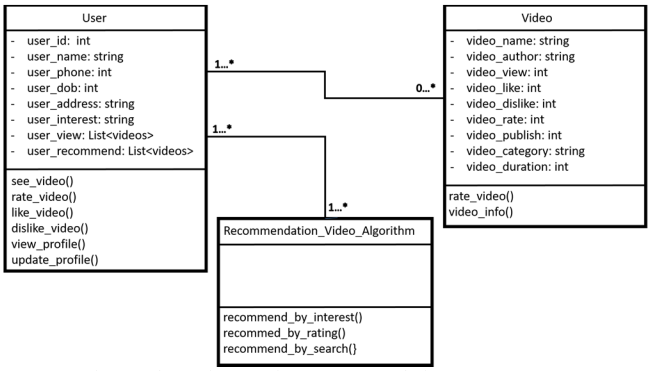

Fig. 2. Class Diagram for Recommendation System

In fig. 2. The class diagram, classes are represented with boxes that contain three sections:

- The first section contains the name of the class. It is printed in bold and centred, and the first letter is capitalized.
- The second section contains the attributes of the class. They are left-aligned and the first letter is lowercase.
- The third section contains the operations the class can execute. They are also left-aligned and the first letter is lowercase.

B. Data Flow Diagram

With the help of a data-flow diagram (DFD) representing a flow of a data of a process or a system becomes easier. The DFD also provides information about the inputs and outputs of each entity and the process itself. A data-flow diagram has no decision rules, control flow.

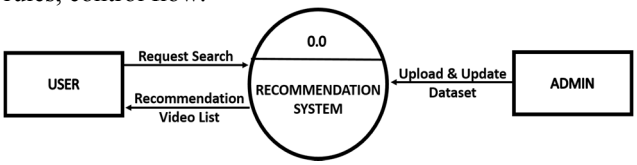

Fig. 3. Data Flow Diagram: Level 0

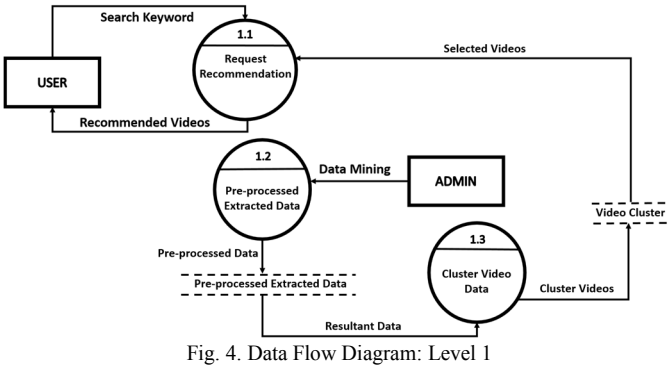

Fig. 4. Data Flow Diagram: Level 1

In fig. 3. and fig. 4. Data Flow Diagram is the overview of how the flow of each process will take place. DFD level 0 shows the general flow of the video recommendation system. Whereas DFD level 1 of different process gives the depth knowledge of method and database usage. The DFD is an easy way to visualize the step involve between the user and the database connected to various method.

IV. Recommendation Algorithm

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

A. Web Crawler

A web crawler which is also called as web spider or spider bot is an Internet bot that systematically browses the World Wide Web to read and collect information such as the URL and the metadata of the video and any other relevant information, mainly for the purpose of Web Indexing or Web Spidering.

The key component of the web search engine is the web crawler that looks into the web pages where they help in indexing the web entries and allows admin to send queries against the index to provide the web page that match the queries. Other use of web crawler is web archiving which allows large metadata of web page to be collected and archived at regular intervals.

The traditional web crawler starts with a predetermined or some initial URL seeds. The traditional web crawler can be only used to download data from static pages. However, there are many dynamic web pages on the Internet. Dynamic page is a special web page which is running on the server. Hence, the traditional web crawler could not get the right contents of the dynamic pages. Web crawlers are used in data mining wherein the pages are analyzed for different properties. In this system we take the URL from the source of videos and download the metadata from the URL. The next step is to analyze the downloaded meta data and extract what is needed for the system and store in the initial database.

B. Rating Factor

Rating factor is the core algorithm of the video recommendation system. It takes into consideration the key attributes of the videos like their views, likes, dislikes and number of comments. Depending on this attributes rating is giving to each and every video out of 10 (weight 10 being max and weight 0 being least). This rating factor is used in the recommendation system to give rank to the videos which are displayed on the user request interest where the video having highest rating factor is displayed first and so on.

$$R = (V/50000)^{1/t} + (L/1000)^{1/t} - (D/100)^{1/t} \quad (1)$$

where in Equation (1) R is the rate value, V is number of views, L is number of likes, D is number of dislikes and t is time strictly in years.

C. Hybrid Filtering

The hybrid filtering algorithm uses a combination of content based filtering and collaborative based filtering.

a) Content Based Filtering: Content-based filtering, also referred to as cognitive filtering, recommends items based on a comparison between the content of the items and a user profile. The content of the videos like the no of views, likes, dislikes and comments are taken into consideration as well as the user profile is taken depending on the history of the user as well as the reaction of the user to different videos.

$$TF_{ij} = \frac{f_{ij}}{max_z f_{zj}} \quad (2)$$

b) Collaborative Filtering: Collaborative filtering is a filtering method in which the system predicts the interest of the user by collecting preferences or information from other users on the system. The collaborative filtering works in a way that if user A has the same opinion as a user B on a video, A is more likely to have B's opinion on a different video than that of a randomly chosen user.

$$Sim(A,B) = \frac{|r_A \cap r_B|}{|r_A \cup r_B|} \quad (3)$$

Equation (3) represent Jaccard Similarity. It is used to find users from same domain

$$Sim(A,B) = \cos^{(r_A, r_B)} \quad (4)$$

Equation (4) represent Cosine Similarity. It is used to find the video preference of users who belong to same domain.

V. Implementation Steps

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.
Implementation of this system is carried out on both frontend and backend. The frontend is where the client user will use to access the system and backend is where our admin will upload data, modified data, update data for the client user in order to get best possible recommendation.

A. Back End

On backend side the following steps will be performed by backend user:

- Step 1: Upload database for the system, here the admin will upload one or many different sheets of excel file which will contain data of different information about videos like video ids, videos description, video title, and all the attribute like video likes, dislikes etc. The admin will upload this database through web portal in which he/she will login with their user id and password.

- Step2: This different data of excel sheets are combined to form single database in SQL. This Database will be filtered by the content judging function which will take video id, video likes, video dislikes and video views as a parameter. Here a rate value is assigned to each video. Higher the rate value the better is the video.
- Step3: The keyword function which will assign keywords to all videos will take video id, video title and video description as parameter. The keywords will be extracted from each video in order to increase efficiency of searching in video library.
- Step 4: For every search query from client user side the system will take query as input check for input query in keyword library. All the query matched will be extracted with rate value and sorted according to increase to decrease manner and will sent to display result.
- Step 5: For each video opened and viewed on system the system will ask for short feedback which will rating the video out of five. Also the liked and disliked on video will be update on database.
- Step 6: This feedback value will be stored on database and will be used as parameter to systems collaborative filter function. The collaborative filter function consists of Jaccard similarity and centred cosine filter will take various user feedback value to find similar users.

B. *Front End*

- Step 1: Login/Sign up. The user has to compulsory sign up or register himself/herself to the system with their credential. The old user can login or logout with their credential.
- Step 2: Assuming new user the user will put search query in search bar which will send to backend and result will be list of best possible videos displayed
- Step 3: After opening and viewing the video user will provide feedback about video from one to five
- Step 4: For experienced/existing user the list of recommendation based on collaborative filter function will also be displayed on login and opening any video.

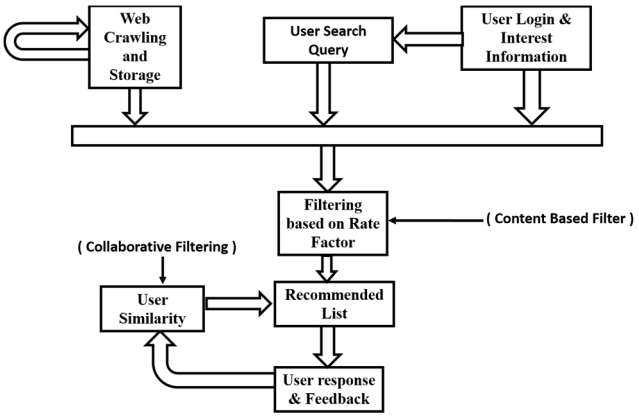

Fig. 5. Flow of the system

VI. RESULTS

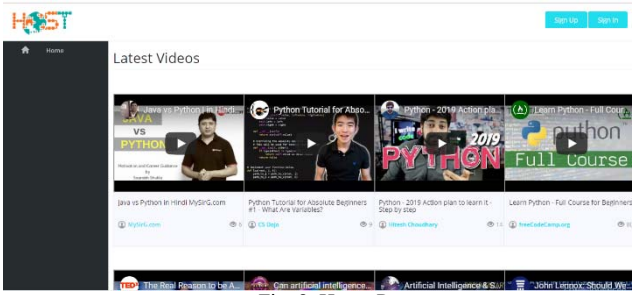

Fig. 8. Home Page

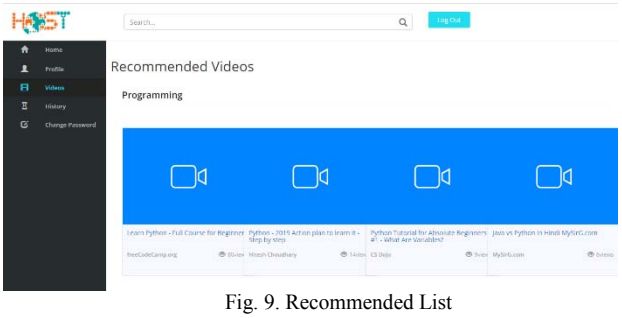

Fig. 9. Recommended List

CONCLUSION

With the development, videos are a trend now, especially young teenagers are likely to watch videos on mobile devices. The challenge that we primarily faced was with the collaborative filters which are used to find the similar interest among the users of the same domain. This paper proposes a video recommendation system that collects the reaction of the users for various videos which helps to know its relevance. Based on the viewers' watching history or browsing, the system is capable of recommending videos to the users.

REFERENCES

[1] Songtao Shang, Minyong Shi, Wenqian Shang, Zhiguo Hong, "A Micovideo Recommendation System Based on Big Data" IEEE ICIS 2016, pp 978-1-5090-0806-3,2016 *(reference)*

[2] Seong-Eun Hong, Jong Kim, "A Comparative Study of Video Recommender Systems in Big Data Era", IEEE, pp. 978-1-4673-9991-3, 2016

[3] Prof. M. P. Nerkar, Shraddha Bhurke, Nikita Kadam, Prajkta kavitake, Komal Handore, "A Micro-video recommendation system using improved slope one algorithm based on Big Data", Scientific Journal of Impact Factor, pp. 2348-6406, 2017

[4] Dongdong Jiang, Wenqian Shang, "Design and Implementation of Recommendation System of Micro Videos Topic", IEEE 2017, 978-1-50905507-4, 2017

[5] Suman Kotin, Prasadgouda Patil, "A Survey on Micro Video Recommendation System Based on Big Data" International Journal of Innovative Research in Technology, Vol 5. Issue 2, pp. 2349-6002, 2018

[6] Sao Paulo "Recommender System in Social Network", JISTEMJ Information System Technology Management, Vol 8. Issue 3, Dec 2018