VIDEO SUMMARIZATION AND RE-RANKING

Aarti Raghani
Dept. of Computer Engineering
V.E.S.I.T.
Chembur, Mumbai, India

Aditi Sawant
Dept. of Computer Engineering
V.E.S.I.T.
Chembur, Mumbai, India

Kajol Chawla
Dept. of Computer Engineering
V.E.S.I.T.
Chembur, Mumbai, India

Revati Pathak
Dept. of Computer Engineering
V.E.S.I.T.
Chembur, Mumbai, India

Mrs. Lifna C.S.
Assistant Professor
Dept. of Computer Engineering
V.E.S.I.T.
Chembur, Mumbai, India

***Abstract*—With the advent of the Information Age, there has been an exponential growth in the number of videos available for any given topic. Hence, internet users are relentlessly in search for foolproof Video Classification Algorithms. The existing state-of-the-art techniques do not provide a satisfactory summary for ranking videos. The objective of the paper is to rank videos using Text Summarization techniques. This application will provide the users with the most relevant videos and their summaries, thereby saving time and increasing efficiency.**

Keywords—Text Summarization; Video Re-ranking; Natural Language Processing

I. Introduction

In today's digital era, the internet is the reliable source for obtaining videos. There has been an enormous growth in videos available for any given topic. A query for any topic yields a heap of results, and it is not easy to manually discard irrelevant results. Hence, it is essential to automatically summarize the videos to get a gist of their content.

The available video analytics systems analyze videos by using Image Processing techniques. The efficiency of these Image-based Summarization techniques heavily depends on the quality of the videos. Hence, these techniques have the following drawbacks: (1)Blurred Images; (2) Poor Illumination Effect; (3) Need for large bandwidth; (4) Storage requirements; (5) Output is also in the form of video.

These constraints degrade the quality of the summaries generated by the systems. This led us to consider Text-based Summarization techniques for summarizing the videos and evaluating the efficiency.

In short, our proposal can be summarized as follows: (1) Search for topmost videos using input keyword; (2) Extract audio from the videos; (3) Convert the audio files to text; (4) Generate document summaries; and (5) Re-rank the videos based on the summaries.

II. Motivation

For finalizing our approach towards the re-ranking of videos based on their content, we conducted extensive research. First, we looked at individual algorithms and how they performed. Paper [1] discusses the TF-ISF approach, which is a modified form of TF-IDF for single documents.

Using this approach, we can identify sentences that have a greater number of important words and lesser number of unnecessary words.

Next, we considered the sentence scoring approach discussed in paper [2]. The sentence scoring approach works to eliminate the deficiencies in individual approaches and, at the same time, gives the benefit of different approached clubbed into one. Using this as our foundation, we have then carefully chosen the factors that provide significant contribution towards creating a concise and rich summary.

Paper [3] discusses the application of machine learning to improve the summary. Improvement in summary means that we endeavour to bring it as close to human- generated summaries as possible.

III.Implemented System

The implemented system has the following modules.(1) Search Module; (2) Audio Extraction Module; (3) Speech-to-Text Conversion Module; (4) Text Summarization Module and (5) Re-ranking Module.

A. Search Module

The query input by user is searched for in the database. If found, relevant video links and generated summaries are displayed to the user. Else, the query is redirected to YouTube.

B. Audio Extraction Module

The video for which summary is to be generated is the input
and the audio is extracted from it. Since we are downloading videos from YouTube, we make use of the API provided by YouTube for downloading the top videos that appear in the search results.
The -dl command can be used for the same.

C. Speech-to-Text Conversion Module

The audio extracted in the preceding step is converted to textual format. This text document forms the input to the Summarization Module. We assessed the accuracy of several speech-to-text APIs available, including Sphinx, Google cloud speech recognition, WIT.AI, IBM Watson and Houndify.

Of these, the Google cloud speech recognition service was able to provide the most accurate and coherent results. It converts videos of greater length, too Hence, we have chosen this API for our system. However, this API does not provide punctuation, which is necessary for extractive text summarization, where two sentences need to be distinguishable to extract a sentence. To overcome this lacuna, we have used the Punctuator API. The Punctuator uses machine learning on a large corpus and provides sufficiently accurate punctuation for any non-punctuated text document.

D. Text Summarization Module

The text document obtained is summarized using extractive summarization technique. The factors that we consider for sentence scoring are: (1) Position of sentence; (2) Length of Sentence; (3) Similarity with the Title; (4) Similarity with the Query; (5) Similarity between sentences; (6) Term Frequency-Inverse Sentence Frequency; (7) Unnecessary information. These scores are combined and the best scored sentences are ranked better.

(1)]Position of sentence: The sentences in the introductory part of a text document contain more information. Hence, the starting few sentences are given more weightage than the rest of them. The scores obtained are in the range from 0 to 1.

For sentences in the first half,

$Score_i$ = 0.9 - 1.6 * i / no_of_sentences

Else,

$Score_i$ = 0.1 + 1.6 * (min ([i - no_of_sentences / 2, no_of_sentences / 2])) / no_of_sentences

(2) Length of Sentence: The lengths for different sentences are calculated and the maximum length is identified. The score awarded to each sentence is a ratio of the length of the sentence to the length of the longest sentence.

It can be given by:

$Score_i = len_i$ / max_length

(3) Similarity with the Title: We want to give a higher score to those documents that have their content strongly related to their titles. This ensures that the document delivers what it promises. Hence, sentences with title words are scored higher than other sentences. The normalised scores range from 0 to 1.

(4) Similarity with the Query: Many of the current applications do not consider similarity with the query. This can lead to higher scores for documents that are coherent and relevant as a whole, but are strongly related to what the user is looking for. Keeping this requirement in mind, we award higher scores to the documents that contain keywords, that is, words appearing in the query input by the user. The scores are normalised.

(5) Similarity between sentences: Another important factor to consider is the cohesion between the sentences within a document. For this, we use the powerful WordNet database, which provides semantic relations(like synonymy, hyponymy, etc.) between words in the document. Sentences with higher cohesion are assumed to represent the same concept and hence given higher score. This is used to weed out irrelevant sentences.

(6) Term Frequency- Inverse Sentence Frequency: This is a modified form of TF-IDF, used for a single text document. The number of occurrences of a term in a single sentence are calculated and called as term frequency. Next, the inverse sentence frequency is calculated by dividing the term frequency by the number of sentences that the term appears in. This method gives higher scores to words occurring greater number of times in a single sentence and, at the same time, diminishes the score for the words occurring in many sentences. This is effective to prevent higher scoring to common words.

(7) Unnecessary information: Conjunctions, adverbs, etc. appear many times in a sentence. However, they do not contribute to the meaningful content in a sentence. Hence, the sentences containing higher number of such words are likely to provide little to no information. Since we are aiming for a compact summary, we must ensure that such sentences receive lower scores. This is done by first using a parts-of-speech tagger to identify word type. Next, sentences having unnecessary words are scored lower to prevent them from being included in the summary.

E. Re-ranking Module

After summaries are generated for the videos, they are re-ranked on the basis of their summaries' relevance to the Query. The total score for each document is calculated as a linear combination of sentence scores. This score is then normalised to account for different number of sentences in summaries. The normalised scores are compared and the summary documents are re-ranked on this basis, with highest scores being ranked first.

The diagram shown in Fig. 1. describes the proposed system. The user inputs the search query over the internet, query is searched for in the database and if the corresponding query is accessible in the database, the results, that is, the relevant re-ranked videos, along with the videos links and the summary of those videos is displayed to user. But if query is not available, it is redirected to YouTube and the listed videos are pre-processed by initially converting them to text document, summarizing their content and finally re-ranking them based on the video content and the input query.

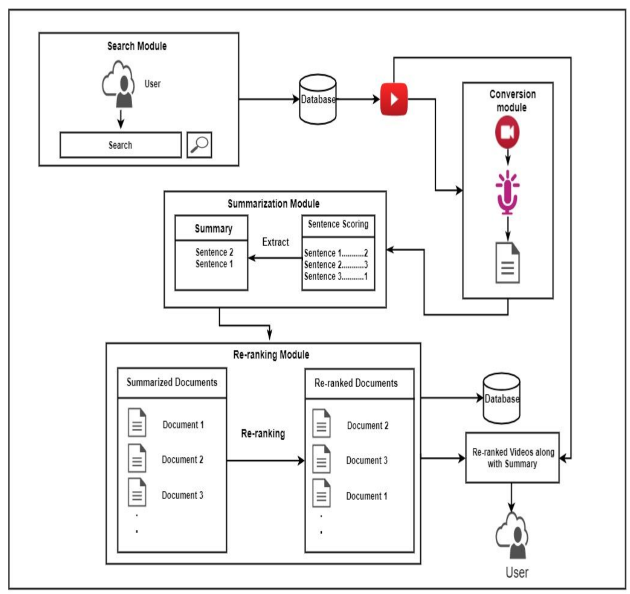

Fig. 1. Implemented System Diagram

IV. Algorithm Performance

We compared the performance of our implemented algorithm with that of other individual algorithms. The metrics used for performance comparison were Accuracy and ROUGE evaluation. Both the metrics showed significant superiority of the implemented algorithm over other individualistic approaches.Table I depicts the results obtained.

TABLE I. COMPARISON WITH OTHER ALGORITHMS

Algorithm	**TextRank**	**LexRank**	**TF-IDF**	**TF-ISF**	**Proposed System**
Accuracy	41%	44%	61%	40%	70%
Lacuna	Does not consider related words	Similarity with user query or title is not considered	Relevant for multiple documents	Does not consider other factors such as sentence scoring, position, length, etc.	Eliminates lacunae in individual algorithms by employing an integrated approach.
Evaluation	ROUGE 0.4229	ROUGE 0.4443	ROUGE 0.3101	ROUGE 0.39925	ROUGE 0.70

V. Results

We compared the summary generated by the implemented system with the summaries generated by some existing systems. We have used the Rouge metrics to compare the results.Table II describes the results obtained by performing Precision, Recall and F Score measures to the systems.

TABLE II. COMPARISON WITH OTHER SYSTEMS

	Analysis of summary obtained by different system					
Parameters	**Text Summarizer**	**Free Summarizer**	**Tools for noobs**	**Auto Summarizer**	**Text Comparator**	**Proposed System**
Precision	0.82	0.63	0.48	0.56	0.54	0.61
Recall	0.31	0.72	0.59	0.70	0.77	0.92
F Score	0.46	0.67	0.53	0.63	0.64	0.70

Summary generated by obtained Proposed System:

Input Text:

I just say Jenn the director of effortless English. I have something exciting to talk about today. I am now doing audio tweets. What is an audio tweet? What a strange word! Well, what I'm doing is I'm doing small little podcast on Twitter, and I am doing these almost everyday, because it's so easy. I use my phone my iPhone and I record a short talk about some Topic in my daily life, my normal life, and then I put it on Twitter twitter.com, and this is a very easy way for you to get more easy. English listening about daily topics about normal daily life, so I am continuing the podcast, of course, but the podcast is usually about learning ideas, teaching ideas. How can you speak English better? What is happening with effortless English, but the audio Twitter that I'm doing now? It'S about very simple day today: topics, for example. Just now I did a short audio, tweet and audio little podcast, just two and a half minutes about the weather in San Francisco. Today it is foggy in San Francisco use my phone. I recorded a short talk about the weather in San Francisco about the fog and then immediately instantly. I put it on my Twitter page. So if you want to hear new English listening topics everyday, please follow me on Twitter, so you will get longer more serious listening here at the podcast and you can get almost daily listening about easy, simple topics on my Twitter page. So how can you follow me on Twitter? It'S very very easy. Just go to twitter.com, it's Twitter. Com, A J Hoge that / a j, a h, o g e twitter.com forward, slash AJ Hoge! If you don't have a Twitter account, it's free! It'S easy! You can make a Twitter account and then follow me. There'S a button that says follow click the follow button. Then you will automatically get my short English listening topics about daily life and I will do a lot of these. It'S so easy. Since I can use my phone, my cell phone, when I'm walking around the city when I'm doing my normal life, sometimes I will stop and talk for one or two minutes about something happening in my life and put it on my Twitter page. So you can get a lot of new, easy, daily life English listening this way it's twitter.com flash AJ Hoge. Alright, I look forward to doing more of these audio Twitter's and audio tweets and also, of course, doing more podcast. I will see you on my Twitter page. I hope you enjoy these new short daily English. Listening topics see you next time, bye, bye,

Summary Generated by Implemented System:

I just say Jenn the director of effortless English.I have something exciting to talk about today. What a strange word!Well, what I'm doing is I'm doing small little podcast on Twitter, and I am doing these almost everyday, because it's so easy.I use my phone my iPhone and I record a short talk about some Topic in my daily life, my normal life, and then I put it on Twitter twitter.com, and this is a very easy way for you to get more easy.English listening about daily topics about normal daily life, so I am continuing the podcast, of course, but the podcast is usually about learning ideas, teaching ideas.How can you speak English better?Just now I did a short audio, tweet and audio little podcast, just two and a half minutes about the weather in San Francisco.So if you want to hear new English listening topics everyday, please follow me on Twitter, so you will get longer more serious listening here at the podcast and you can get almost daily listening about easy, simple topics on my Twitter page.Com, A J Hoge that / a j, a h, o g e twitter.com forward, slash AJ Hoge!There'S a button that says follow click the follow button.Since I can use my phone, my cell phone, when I'm walking around the city when I'm doing my normal life, sometimes I will stop and talk for one or two minutes about something happening in my life and put it on my Twitter page.So you can get a lot of new, easy, daily life English listening this way it's twitter.com flash AJ Hoge.

VI. Conclusion

As we can see from the results, the summary generated from the implemented methodology gives good accuracy and is closer to human-generated summaries. This is because the generated summary is a comprehensive indication of all the factors considered. The factors were chosen thoughtfully so as to eliminate the deficiencies created by using a single factor.This model can be extended to generate summaries for multilingual videos. This needs an efficient and accurate translation module.Such an application can prove to be quite useful in detecting terrorist attacks and opening up a world of information for people from different linguistic backgrounds.

VII.References

[1] Doko, Alen, Maja Stula, and Darko Stipanicev. "A recursive TF-ISF Based Sentence Retrieval Method with Local Context." *International Journal of Machine Learning and Computing* 3.2 (2013): 195.

[2] Raju, T. Sri Rama, and Bhargav Allarpu. "Text Summarization using Sentence Scoring Method." (2017).

[3] Neto, Joel Larocca, Alex A. Freitas, and Celso AA Kaestner. "Automatic text summarization using a machine learning approach." *Brazilian Symposium on Artificial Intelligence*. Springer, Berlin, Heidelberg, 2002.

[4] Allahyari, Mehdi, et al. "Text summarization techniques: A brief survey." *arXiv preprint arXiv:1707.02268* (2017).

Acknowledgement

We are sincerely thankful to Mrs. Sujata Khedkar, Associate Professor, Dept. of Computer Engineering, V.E.S.I.T. for her valuable insight and guidance in developing this approach