IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

YOUTUBE TRANSCRIPT SUMMARIZATION

Varun N, Thousif Pasha, Vishal L Prasad

Student, Student, Student

Department of Information Science and Engineering,
B.M.S. College of Engineering (Autonomous College Affiliated to VTU), Bengaluru, India

Abstract: YouTube is a massive platform that hosts a vast amount of video content. However, finding relevant information from these videos can be time-consuming and challenging, especially when one wants to understand the key points of a video quickly. This problem can be addressed by automatically summarizing the transcript of a video into a concise and informative summary. The TF-IDF (Term Frequency-Inverse Document Frequency) algorithm is used to summarize the transcript of YouTube videos. The TF-IDF algorithm is a popular information retrieval technique that measures the importance of a word in a document. The algorithm calculates the term frequency (TF) of each word in a transcript and measures the inverse document frequency (IDF) of the words across a large corpus of documents. The TF-IDF algorithm is applied to the transcript of each video to determine the most important words in the transcript. The summary is then generated by selecting a subset of the most important sentences that contain these important words. The generated summary effectively condenses the transcript into a concise and informative summary that can be quickly consumed. The results of the project indicate that the TF-IDF algorithm is an effective approach for summarizing the transcript of YouTube videos. The generated summaries accurately capture the key points of the video, making it easier for viewers to quickly understand the content of the video. This approach can be useful for a variety of applications, including content discovery, information retrieval, and education.

Keywords - Text Frequency – Inverse document, accuracy, frequency – Text ranking, - Summarization, Natural Language Processing, Extractive, YouTube.

INTRODUCTION

With over a billion active users and hundreds of hours of video content uploaded every minute, it can be overwhelming for viewers to sift through all the information and find what they are looking for. This is where transcript summarization can play a crucial role. By summarizing the transcript of a video, viewers can quickly get a sense of the content covered in the video, helping them to decide whether they want to watch the full video or not. Many videos on the platform are not properly captioned or do not have captions, making it challenging for people with hearing impairments to understand the content. A summarized transcript can provide these individuals with a concise overview of the video, making it more accessible and inclusive. Google uses transcriptions and video descriptions to rank videos in search results. By summarizing the transcript, the video description becomes more concise and relevant, which can help it to rank higher in search results and attract more viewers. It helps to save time and effort, improve accessibility, and boost the visibility of videos in search results, making it a valuable tool for both viewers and content creators. This study focuses more on extractive text summarization because the system being suggested uses it. Extractive summarization does not rewrite or alter the original material. This approach extracts a few sentences that may be used to index the material, as well as some significant key phrases.

LITERATURE SURVEY

In their work [1], the authors proposed two methods, extractive and abstractive, to generate summaries and important keywords from YouTube videos. They developed a user-friendly interface that allows users to easily obtain summaries using these methods. The interface ensures a seamless interaction, providing users with the desired information without the need to watch lengthy videos. This project effectively addresses the problem of time and effort wastage by offering only the relevant information on the topic of interest, enabling users to utilize their saved time for further knowledge acquisition.

The authors in [2] propose a system for summarizing YouTube videos using natural language processing (NLP) and machine learning techniques. With the increasing number of videos available on web platforms, including educational content on platforms like YouTube, Facebook, Google, and Instagram, it becomes challenging to extract valuable information without watching the entire video. The suggested approach addresses this issue by retrieving video transcripts from user-provided links and utilizing Hugging Face Transformers and Pipelining to summarize the text. The developed model takes video links and the desired summary duration as inputs and produces a concise transcript summary as output.

The method proposed in [3] focuses on the recent advancements in the field and offers a comprehensive overview of deep learning-based approaches for general video summarization. The authors begin by explaining the motivations behind the development of video summarization technologies and outline the key aspects of a typical deep learning-based analysis pipeline. They then present a taxonomy of existing algorithms and conduct a systematic review of relevant literature, highlighting the evolution of deep learning-based video summarization techniques. The review ultimately provides insights for future research directions in this area.

In the paper [4], the existing methods for video summarization primarily focused on ensuring diversity and representativeness in the generated summaries. The authors approached video summarization as a content-based recommender problem, aiming to extract the most valuable content from lengthy videos for users dealing with information overload. They proposed a scalable deep neural network that predicts the usefulness of video segments for users. This approach explicitly models both the individual segments and the entire video to capture their relationships. Additionally, the paper addressed scene and action recognition in untrimmed videos to uncover correlations across various aspects of video comprehension tasks. Furthermore, the study examined the impact of audio and visual features on the summarization task.

According to the findings presented in reference [5], video summarization and skimming have become essential tools in practical video content management systems. The mentioned paper offers a tutorial on the existing techniques for abstracting generic videos and introduces cutting-edge approaches for skimming feature films. It also discusses the authors' recent advancements in movie skimming, which involve audiovisual tempo analysis and adherence to specific cinematic rules. Considering the advancements in movie genre classification, content comprehension, and video abstraction techniques, an automated system for analyzing movie content to facilitate navigation, browsing, and searching for desired movie content could become a reality in the near future.

According to reference [6], automatic summarization methods provide users with a convenient means of quickly accessing important content from a collection of media and later exploring media of their choice. Due to advancements in capturing devices, cloud-based summarization solutions with longer processing times are becoming less popular among end users. In this research paper, the author introduces a real-time video summarization technique specifically designed for mobile platforms. This approach involves analyzing the video in real-time during camera recording and generating instantaneous summaries. The technique utilizes intrinsic video data, such as the video stream content, as well as extrinsic metadata like external camera information. The proposed technique achieves an f-measure of 0.66 and 0.84 on the SumMe and SumLive datasets, respectively. Furthermore, it successfully limits the overall power consumption to 20 milliamps on an embedded system.

The authors of reference [7] present a method called online video highlighting, which aims to create concise summaries of unedited and unstructured videos in a systematic manner. Manual processing of such videos is time-consuming and expensive. The proposed approach involves learning a dictionary from the given video using group sparse coding. The dictionary is updated dynamically during the process. The generated summary video is a combination of segments that cannot be sparsely reconstructed using the learned dictionary. One key advantage of their online method is its ability to process videos of arbitrary length and start generating summaries before reaching the end of the video. Additionally, the proposed method achieves a processing time that is similar to the original video length, resulting in near real-time summarization speed.

According to a previous study [8], a user attention model was introduced as a general framework for estimating the attention viewers might allocate to video content. This model takes into account human attention, which is a powerful mechanism for prioritizing and filtering information. By considering visual and auditory stimuli, as well as some level of semantic understanding, the user attention model defines viewer attention. The study also proposed various modeling methods for visual and auditory attentions. One important application of this user attention model is video summarization, where a practical solution is implemented to demonstrate the model's effectiveness, robustness, and applicability. Results from a user study on video summarization showed promising outcomes, indicating that the user attention model can serve as an alternative approach to understanding video content.

According to the authors in [9], they suggested that video summarization could be improved by integrating more external information, specifically user-based information that is collected non-invasively. This integration aims to address long-standing challenges like the semantic gap and to generate video summaries that are more personalized and relevant to individual users.

The authors of [10] conducted a survey on the literature related to video classification. They observed that features used for video classification are derived from three modalities, namely text, audio, and visual. They noted that a wide range of combinations of these features and classification methods have been explored in the research. The authors provided an overview of the commonly chosen features and summarized the existing research in this field. The paper concludes by suggesting potential areas for future research.

PROPOSED SYSTEM

This document presents a proposed system for YouTube transcript summarization utilizing the TF-IDF (Term Frequency-Inverse Document Frequency) algorithm. The objective is to automatically generate concise summaries of YouTube video transcripts, enabling users to quickly grasp the key information contained within the videos. The system leverages the TF-IDF technique to identify important words and sentences within the transcript and constructs a summary based on their significance. This proposed system aims to enhance accessibility and improve the user experience by providing efficient and informative summaries of YouTube videos.

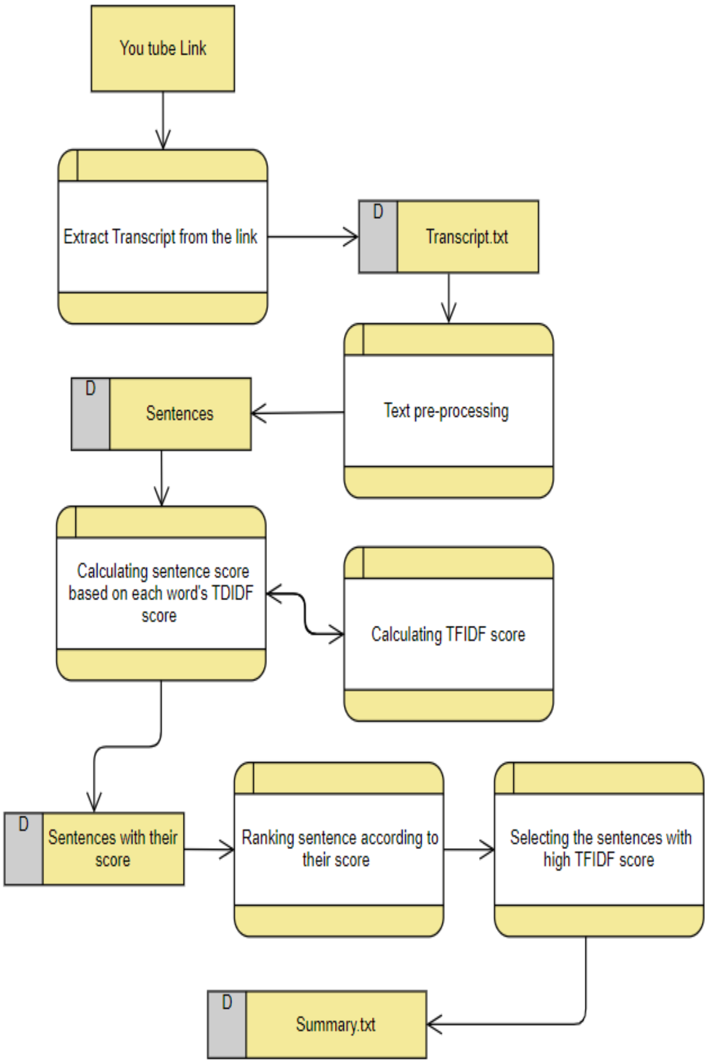

Fig. 1 High Level Design

1. Youtube link

The youtube_transcript_api is a Python module that provides a convenient way to fetch transcripts from YouTube videos. It allows you to retrieve the text-based transcription of a video's spoken content. The "list_transcripts()" function in Python is used to retrieve a list of available transcripts. This function retrieves all available transcripts for the video in different languages. The function returns a Transcript List object, which can be iterated over and offers methods for filtering transcripts based on language and type.

2. Tokenization

Tokenization involves breaking down large blocks of text into smaller units known as tokens. Tokens can be words in a sentence or sentences in a paragraph. The initial step is to tokenize the content by dividing it into sentences, and then assigning weights to these sentences. The weight assigned to each token can be determined based on factors like frequency and uniqueness. The NLTK library's tokenize module submodule is commonly used for tokenization. Ultimately, this process results in a matrix that contains all the identified tokens along with their respective weights.

3. Term Frequency (TF)

Once the weight matrix is obtained, the system proceeds to calculate the Term Frequency (TF) for each word in a paragraph. The TF is defined as the number of times a specific term appears in a document divided by the total number of terms in that document. The resulting TF values for each word are then stored in a matrix known as the TF matrix.

Definition of TF

TF(term) = (Frequency of term appearing in a document) / (Total number of terms in the document)

.

4. **Inverse Document Frequency (IDF**):

The Inverse Document Frequency (IDF) for each word in a paragraph, considering the paragraph as a document and each word within the paragraph as a term.

Definition of IDF,

IDF IDF(term) = log(total number of documents / number of documents containing the term)

In this formula, the IDF of a term is determined by taking the logarithm of the ratio between the total number of documents in the corpus and the number of documents that contain the term. The purpose of IDF is to assign higher weights to terms that are relatively rare across the corpus, as they are considered more informative and discriminative.

5. **TF-IDF Score**

   Using the matrices of individual TF and IDF scores, the system calculates the combined TF-IDF score for each word. This is achieved by multiplying the TF score of the word with its corresponding IDF score.

   TF-IDF(term) = TF(term) * IDF(term)

6. **Scoring the sentences**

   In various algorithms, scoring a sentence involves different methods. However, in this particular approach, the TF-IDF scores of the words in a sentence are utilized to calculate the sentence score. The score of a sentence is determined by calculating the average TF-IDF value of all the words within that sentence.

   Sentence score = (Sum of TF-IDF values of all the words in a sentence)/(Total number of words in the sentence).

7. **Finding the threshold**

   The process of scoring sentences differs across various algorithms once the TF-IDF scores of each word have been calculated. In this approach, the sentence score is determined by utilizing the TF-IDF scores of the words within the sentence. The score of a sentence is computed by taking the average of the TF-IDF values of all the words present in that sentence.

8. **Generating summary**

   The summary is created based on a specific threshold score. Sentences that have a TF-IDF score higher than the threshold are included in the summary, while those with lower scores are disregarded. This process results in the final summary of the collected data..

RESULTS

The following is obtained as a result of the text summarization process:
Step 1:
User interface:
The web page which acts as an interface for the user to enter or paste link.

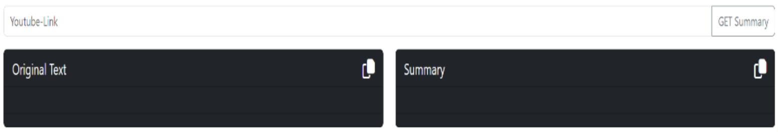

Fig. 2 A sample picture showing the interface before uploading the you tube transcript

Step 2:
Final summary:
Transcript of the particular link in fetched using the API. Left part represents the original transcript pf the video and right the summary of the original transcript. Key word with highest TF-IDF score is display below the summary.

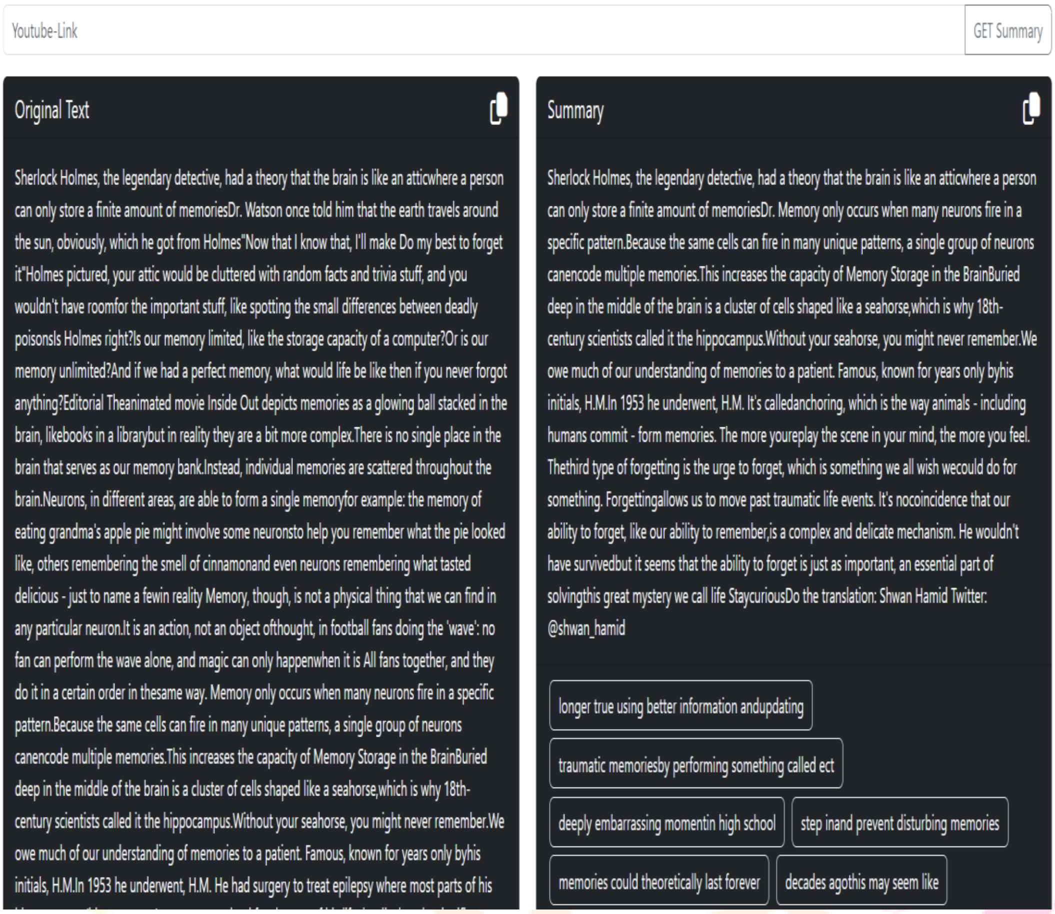

Fig. 3 A sample picture showing the interface after uploading the you tube transcript with the summary

Conclusion

YouTube transcript summarizer utilizing the TF-IDF algorithm presents an effective approach for generating summaries from YouTube video transcripts. By employing TF-IDF scores, sentences are evaluated based on their importance and relevance to the overall content. The algorithm selects sentences with higher TF-IDF scores, indicating their significance in capturing key information. By setting a threshold score, the summarizer determines which sentences should be included in the final summary. This method provides a concise and informative summary, leveraging the textual content extracted from YouTube transcripts. However, it is important to note that the TF-IDF algorithm is just one approach in the field of transcript summarization, and other techniques may offer further improvements in terms of summarization quality and accuracy.

References

[1] Shraddha Yadav, Arun Kumar Behra , Chandra Shekhar Sahu, Nilmani Chandrakar, " SUMMARY AND KEYWORD EXTRACTION FROM YOUTUBE VIDEO TRANSCRIPT", International Research Journal of Modernization in Engineering Technology and Science Volume:03/Issue:06/June-2021 Impact Factor- 5.354 .

[2] A. N. S. S. Vybhavi, L. V. Saroja, J. Duvvuru and J. Bayana, "Video Transcript Summarizer," 2022 International Mobile and Embedded Technology Conference (MECON), 2022, pp. 461-465, doi: 10.1109/MECON53876.2022.9751991.

[3] E. Apostolidis, E. Adamantidou, A. I. Metsai, V. Mezaris and I. Patras, "Video Summarization Using Deep Neural Networks: A Survey," in Proceedings of the IEEE, vol. 109, no. 11, pp. 1838-1863, Nov. 2021, doi:10.1109/JPROC.2021.3117472.

[4] Yudong Jiang, Kaixu Cui, Bo Peng, Changliang Xu; "Comprehensive Video Understanding: Video Summarization with Content-Based Video Recommender Design"; Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 0-0.

[5] Ying Li, Shih-Hung Lee, Chia-Hung Yeh and C. . -C. J. Kuo, "Techniques for movie content analysis and skimming: tutorial and overview on video abstraction techniques," in IEEE Signal Processing Magazine, vol. 23, no. 2, pp. 79-89, March 2006, doi: 10.1109/MSP.2006.1621451.

[6] P. Choudhary, S. P. Munukutla, K. S. Rajesh and A. S. Shukla, "Real time video summarization on mobile platform," 2017 IEEE International Conference on Multimedia and Expo (ICME), 2017, pp. 1045-1050, doi: 10.1109/ICME.2017.8019530.

[7] Bin Zhao, Eric P. Xing; Quasi Real-Time Summarization for Consumer Videos; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 2513-2520

[8] Yu-Fei Ma, Xian-Sheng Hua, Lie Lu and Hong-Jiang Zhang, "A generic framework of user attention model and its application in video summarization," in IEEE Transactions on Multimedia, vol. 7, no. 5, pp. 907-919, Oct. 2005, doi: 10.1109/TMM.2005.854410.

[9] Video summarization: A conceptual framework and survey of the state of the art, Journal of Visual Communication and Image Representation, Volume 19, Issue 2,2008, Pages 121Arthur G. Money, Harry Agios, - 143, ISSN 1047-3203.

[10] D. Brezeale and D. J. Cook, "Automatic Video Classification: A Survey of the Literature," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 38, no. 3, pp. 416-430, May 2008, doi: 10.1109/TSMCC.2008.91

[11] S. JUGRAN, A. KUMAR, B. S. TYAGI and V. ANAND, "Extractive Automatic Text Summarization using SpaCy in Python & NLP," 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2021, pp. 582-585, doi: 10.1109/ICACITE51222.2021.9404712.

[12] J. Chen and F. You, "Text Summarization Generation Based on Semantic Similarity," 2020 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), 2020, pp. 946-949, doi: 10.1109/ICITBS49701.2020.00210.

[13] M. Afsharizadeh, H. Ebrahimpour-Komleh and A. Bagheri, "Query-oriented text summarization using sentence extraction technique," 2018 4th International Conference on Web Research (ICWR), 2018, pp. 128-132, doi: 10.1109/ICWR.2018.8387248.

[14] R. Bora-Kathariya and Y. Haribhakta, "Natural Language Inference as an Evaluation Measure for Abstractive Summarization," 2018 4th International Conference for Convergence in Technology (I2CT), 2018, pp. 1-4, doi: 10.1109/I2CT42659.2018.9057819.

[15] K. U. Manjari, S. Rousha, D. Sumanth and J. Sirisha Devi, "Extractive Text Summarization from Web pages using Selenium and TF-IDF algorithm," 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), 2020, pp. 648-652, doi: 10.1109/ICOEI48184.2020.9142938.