YouTube Transcript Summarizer Using Machine Learning and Natural Language Processing.

Dr M A Wakchaure[1], Pratiksha A Hase[2], Shreya S Gumaste[3], Sanket D Ahire[4], Suyash A Gunjal[5]
Department Of Computer
Amrutvahini College of Engineering
Sangamner India

Abstract-This paper is about YouTube Transcript Summarizer Using Machine Learning and Natural Language Processing in which we are creating a system that is summarizing videos from the platform like YouTube. This paper uses Natural Language Processing (NLP) and Machine learning to generate a summarized transcript without missing any key points. The objectives of this research is to condense transcribed text obtained from a provided input video. The proposed methodology entails the retrieval of transcript summary through video id. The constructed model accepts video identifiers as input from users and produces a short transcript summary as output. As per the findings, the output translated text summary was generated more efficiently compared to alternative methodologies.

Keywords—Summary,YouTube, transcript.

I. Introduction

It is estimated that YouTube had more than 2.3 billion active users in 2020, and the number has shown a steady annual rise, contrary to the popular belief. Through the channel, there are 500 videos uploaded every minute. Though content is rich, identifying a particular piece of information that is scattered through the long video such as TED talks is still an issue. To cope with this, we are going to apply the Latent Semantic Analysis (LSA) Natural Language Computing (NLC) algorithm. One of the benefits is that the amount of processing power required is reduced, and it also reduces the necessitation of large training datasets. This data means therefore that YouTube is developing markedly and so the need exists thereof for better ways to get around its vast repository of content.

A. Motivation

The YouTube video transcripts summaries serve the purpose of serving as a scannable overview of the video's content enabling the users to read and get a quick grasp of the video's contents which enhances readability and the user experience. This is particularly useful when there is high pressure for instance at the end of the term when time is a constraint and time efficient revision is essential. The transcript summary is the condensed of the whole transcript that allows users to quickly grasp the major points of video.

B. Objectives

- To study and analyze machine learning and NLP algorithms.
- To study a summarization algorithm that accurately captures the main points, ideas and context of the video's content.
- To design model to get transcript text based summery of YouTube video.
- To implement proposed algorithms of transcript summarizer.
- To measure and compare proposed system result with existing system results.

II. LITERATURE SURVEY

This research paper uses extractive techniques to execute video summarization, achieved through the utilization of pipelining techniques to analyze critical segments of the video. [1] Pipelining is a methodology commonly employed in computer science and data processing, where tasks are broken down into smaller, sequential stages, with the output of one stage serving as the input for the next. This approach enables efficient and streamlined processing, allowing for the analysis of complex data in a systematic manner.[9] This employs the Algebraic Statistical method alongside the MoviePy Library to align video segments with subtitles, thereby generating the summarized text. A notable advantage of this model is its minimal processing power requirement and the absence of necessity for prior training data. In the context of our YouTube transcript summarizer, pipelining is applied to the analysis of video content.

The process begins with the extraction of the video's transcript, which is then passed through a series of stages within the pipeline. These stages encompass various techniques such as speech recognition, text preprocessing, and text summarization. Each stage focuses on a specific aspect of the analysis, progressively refining the data and extracting key information. For instance, the speech recognition stage converts the audio content of the video into text format, enabling further analysis. Subsequently, the text undergoes preprocessing, which may involve tasks such as removing noise, normalizing text, and identifying key entities or keywords. Following preprocessing, the text is subjected to summarization algorithms that identify and extract the critical points and main themes of the video.[2] Moreover, our system goes beyond mere summarization by incorporating a comparison of the similarity between the provided input video and the updated, summarized version.

This comparison provides users with insight into the extent to which the essential content of the video has been retained in the summary, thereby enhancing transparency and confidence in the summarization process. Finally, to ensure the quality and accuracy of the summarized text, our system incorporates mechanisms for evaluating the fidelity of the summary to the original video content.[3] This assessment may involve comparing the summary against the original transcript, evaluating coherence and relevance, and leveraging metrics such as precision and recall. In essence, through the strategic application of pipelining techniques, our system is able to efficiently analyze and summarize YouTube video content, while also providing insights into the similarity between the original and summarized versions, and ensuring the accuracy of the summarized text.[1] An automatic YouTube transcript summarizer is a cutting-edge tool designed to swiftly generate summaries of video content by analyzing the spoken words within the video's transcript.

This innovation is particularly valuable for users seeking to grasp the essence of a video without investing the time to watch it in its entirety. [4] Our project introduces a sophisticated system that leverages machine learning and natural language processing (NLP) methodologies to achieve this task seamlessly. Central to our system's functionality is a machine learning model that has been meticulously trained using state-of-the-art summarization algorithms.[10] This model operates in real-time, analyzing YouTube videos to accurately distill the main points and essential information from their transcripts.

Through the utilization of advanced techniques such as speech recognition, text preprocessing, and text summarization, our system excels in delivering concise and precise summaries. [5]A pivotal component of our methodology is latent semantic analysis (LSA). LSA is a powerful technique within the field of NLP that delves into the latent semantic structure embedded in the word usage across a body of text.[7] By employing LSA, our system uncovers intricate patterns and relationships between words, thereby extracting deeper meaning from the text.

This method is particularly effective in response to user queries, commonly referred to as concept searches. Essentially,[2] LSA enables our system to grasp the contextual nuances and underlying concepts within the transcript, facilitating more accurate and insightful summarizations. Our project represents a fusion of cutting-edge technologies aimed at revolutionizing the way users interact with YouTube content. By harnessing the power of NLP, machine learning, and specifically LSA, we have developed a system capable of providing users with efficient and comprehensive summaries of YouTube videos, enhancing accessibility and usability in the digital age.

III. PROPOSED SYSTEM

In the initial phases of the process,[1] the system focuses on data classification, extracting pertinent information from the input video transcripts. Once the transcripts are obtained, the system enters the training phase, where it discerns the tone embedded within the textual content. This model serves as the foundation for generating the final summarized text, integrating insights from the transcripts while ensuring coherence and conciseness in the output.
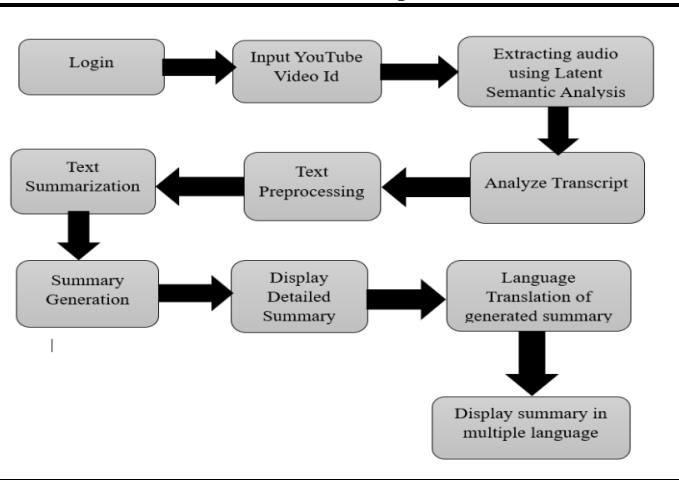

Fig1. Process Flow Diagram

By summarizing the video transcripts, each step gears toward the output of an uninterrupted brevity of operation which is concise and digestible by reading which gives the key points. First, the transcript is carefully reviewed and only relevant data is identified, which helps to omit information that is not as important and to add information that is only important in the intended summary. [8] Next the system also determines the mood of the content through the tone which permits it to discern the dominant atmospherics and implications of what is spoken in the video.

These are techniques that tend to simplify the current added-up task into small and few understandable steps as each part contributes to the overall production. The system has a continuous evolution process, including refinement and improvement [6]. Thus, the final output is well structured and is able to successfully summarize the main points of the input content. The process of interacting with AI tools is ongoing which not only helps the tool to reach an improved accuracy level but also ensures that the summaries are easily accessible and digestible by users who demand videos to be summarized quickly and easily.

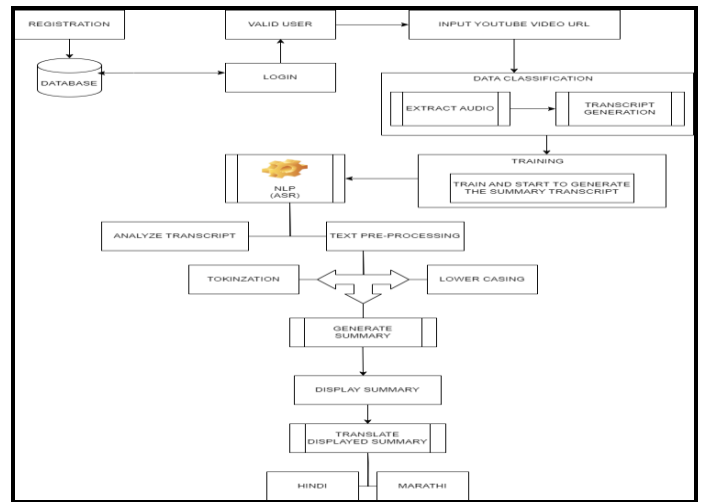The system architecture is described in below figure.

Figure 2. System Architecture

A. Methodology

We were reviewing various video summarization approaches, we observed that many require substantial training and execution time. In response, we conducted an evaluation to address this issue. Our solution involves utilizing the transcripts of YouTube videos by fetching them for text summarization, rather than directly processing the video content. This approach aims to streamline the summarization process while maintaining accuracy and efficiency.

1. User Login: Implement functionality for users to register and log in securely to the system.
2. User Authentication: Verify user credentials during login to ensure authorized access.
3. Input Processing: Accept YouTube Vedio's ID as input for further processing.
4. Data Classification: Extract audio and generate a transcript from the input video.
5. Training Process: Train the system using the generated transcript to enhance performance.

6. NLP Algorithm Application: Apply Latent Semantic Analysis (LSA) to extract key sentences from documents.
7. Transcript Analysis: Analyze the generated transcript, including preprocessing steps.
8. Summary Generation: Employ tokenization and lower casing for text preprocessing to create summaries.
9. Summary Display: Showcase the generated summary to users.
10. Multilingual Translation: Enable translation of the displayed summary into multiple languages for accessibility.
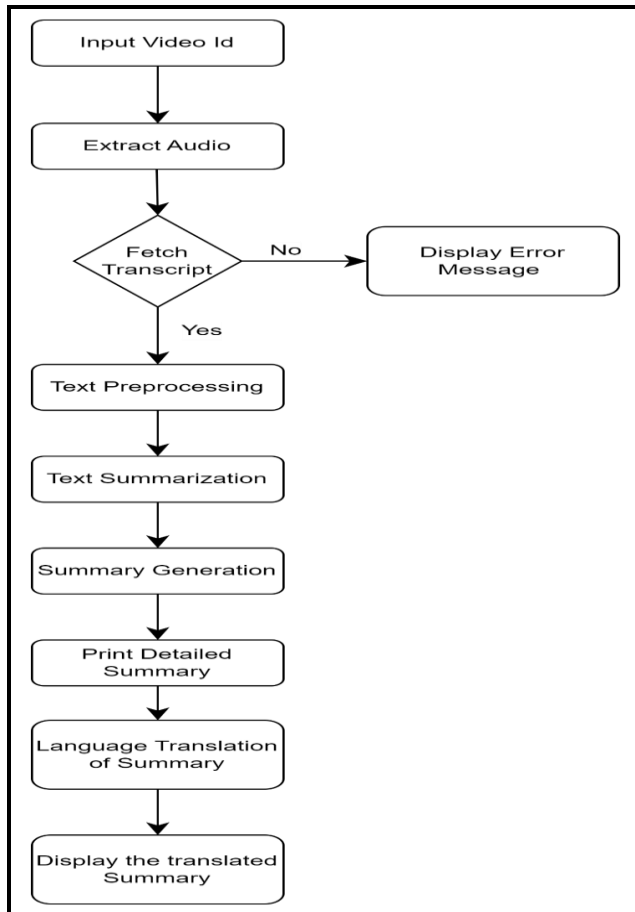

Figure 3. Process Flowchart

B. Alogrithm

Module 1: Providing the input video for extracting the necessary transcripts.

Step1: Install tkinter library which is necessary for Graphical User Interface (GUI).

Step 2: Next install youtube_transcript_api to fetch YouTube transcripts of provided video.

Step 3: Import NLTK library for natural language processing tasks.

Step 4: Give YouTube Video id of video to get the summary.

Module 2: Fetching the transcripts of provided input video into a function so that they can be processed for summarization process.

Step 1: Display the input YouTube Video Id with summarize button on GUI.

Step 2: Creating the transcript of a specified video through a function.

transcript = YouTubeTranscript_get_transcript(video_id)

Step 3: Observation of generated transcript along with its contents

full_transcript = ' '.join([entry['text'] for entry in transcript])Step 4: Perform text preprocessing.

Step 5: Perform text Summarization.

Step 6: Print the generated summary.

Module 3: Performing language translation on generated summary

Step 1: Firstly, import Translator library of googletrans

Step 2: The function used is to translate the language

translated = translator.translate(text, dest=target_language)

Step 3: Define the target languages into which the generated summary has to be translated using the following code:

text_to_translate = "Hello, how are you?"

target_language = "hi"

target_language2 = "mr"

Step 4: Print the translated summary in multiple specified target language.
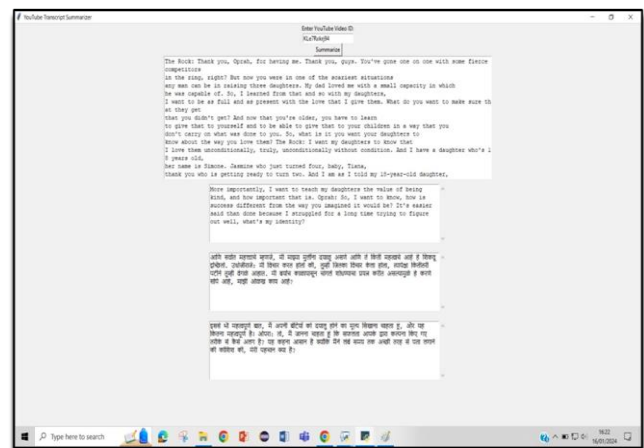

Figure 4. Screenshot of Resulted Output

C. Requirements

The system undergoes testing with a specific set of inputs to assess its accuracy and functionality. The operational requirements for implementation include transcription, summary transcription, and text analysis. The software prerequisites entail:
1. Programming Language: Python
2. Operating System: Windows 7 or newer
3. Development Environment: Google Colab

■ Advantages

i. Time-Saving: The software enables viewers to quickly assess the relevance of video content by providing concise summaries, saving time and effort in content consumption.
ii. Content Optimization: Content creators benefit from improved discoverability and a deeper understanding of viewer engagement through keyword extraction and categorization.
iii. Accessibility: The summarization feature enhances accessibility for individuals with hearing impairments, providing closed captions and text-based con tent summaries.
iv. Customization: Users can tailor their summarization preferences and categories, allowing for a personalized and enhanced user experience.
v. Data Insights: The software offers data insights and user feedback through comment analysis, assisting content creators in refining their content and strategy.

■ Disadvantages

i. Accuracy Constraints: Summarization accuracy depends on the quality of NLP algorithms and speech recognition technology, which may result in occasional inaccuracies.
ii. Language and Dialect Variability: Summarization may be less effective for content in languages or dialects with limited NLP and speech recognition support.
iii. Privacy Concerns: The handling of user-generated content and comments raises privacy and ethical considerations that must be addressed.

■ Conclusions and future work

In conclusion, the "YouTube Transcript Summarizer" project presents a significant advancement in online video content interaction, promising enhanced accessibility, efficiency, and informativeness. Through the utilization of Latent Semantic Analysis (LSA), key sentences are extracted from video transcripts, allowing for concise and relevant summaries. This implementation ensures that users can quickly grasp the essence of video content without having to watch the entire video, thereby saving time and improving the overall viewing experience.

Furthermore, the project's future scope includes potential enhancements such as sentiment analysis, topic modeling, and advanced natural language processing techniques to provide even deeper insights into video content. Additionally, integrating multilingual support and real-time summarization features could further broaden its impact globally. By addressing current digital media needs and continually evolving to meet future demands, this project has the potential to significantly shape the landscape of video content consumption and creation, fostering a more dynamic and engaging online environment.

● References

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

2. Foong, O.-M., Yong, S.-P., & Jaid, F.-A. (2015). Text Summarization Using Latent Semantic Analysis Model in Mobile Android Platform. 2015 9th Asia Modelling Symposium (AMS). doi:10.1109/ams.2015.15

3. Sulochana Devi, Rahul Nadar, Tejas Nichat, and Alfredprem Lucas. Abstrac tive summarizer for youtube videos. In International Conference on Appli cations of Machine Intelligence and Data Analytics (ICAMIDA 2022), pages 431–438. Atlantis Press, 2023.

4. Eesha Inamdar, Varada Kalaskar, and Vaidehi Zade. Survey paper on youtube transcript summarizer.

5. Dan Jurafsky. Speech & language processing. Pearson Education India, 2000. [6] Yue Kang, Zhao Cai, Chee-Wee Tan, Qian Huang, and Hefu Liu. Natural language processing (nlp)

6. James Rumbaugh, Michael Blaha, William Premerlani, Frederick Eddy, William E. Lorensen, et al. Object-oriented modeling and design, volume 199. Prentice-hall Englewood Cliffs, NJ, 1991.

7. M. A. Wakchaure and S. S. Sane, "An Algorithm for Discrimination Prevention in Data Mining: Implementation Statistics and Analysis," 2018 International Conference On Advances in Communication and Computing Technology (ICACCT), Sangamner, India, 2018, pp. 403-409, doi: 10.1109/ICACCT.2018.8529489.

8. Shitole, Manodnya, and M. A. Wakchaure. "Survey: Techniques Of Data Mining For Clinical Decision Support System." International Journal of Advance Research and Innovative Ideas in Education.

9. T. Patten and P. Jacobs, "Natural-language processing," in IEEE Expert, vol. 9, no. 1, pp. 35-, Feb. 1994, doi: 10.1109/64.295134.

10. M. Banane and A. Erraissi, "A comprehensive study of Natural Language processing techniques Based on Big Data," 2022 International Conference on Decision Aid Sciences and Applications (DASA), Chiangrai, Thailand, 2022.