ISSN [Online]: 2583-2654

YouTube Transcript Summarizer

[1]Amrutha V R, [2]Dharani B,[3] Deepthika K N,[4] G Charulathika,[5]Premalatha D
Department of CSE, Dr T Thimmaiah Institute of Technology

Abstract: The YouTube Transcript Summarizer is a web-based tool designed to generate concise summaries of YouTube video transcripts using Natural Language Processing (NLP) techniques. It employs methods such as lemmatization, part-of-speech tagging, and named entity recognition to identify and extract key information. Developed using the Flask framework, the application is deployed on a web server to ensure user-friendly access. This summarizer enables users to efficiently grasp the main points of lengthy videos, saving time and enhancing content comprehension.

I. INTRODUCTION

The results but still struggled with complex language, sarcasm, and domain changes. Many systems also fail on multilingual and informal social media text, reducing their real-world effectiveness.The proposed system uses a BERT-based model fine- tuned for fake news detection, enabling better understanding of semantics and context [3]. Additional layers like Bi-LSTM/Bi-GRU or attention can enhance feature extraction and improve accuracy. A final classification layer produces a real/fake output with confidence scores, making the system more reliable and adaptable.The paper includes dataset collection, text pre- processing, BERT tokenization, model training, and evaluation [4]. The architecture processes cleaned text through transformer layers and outputs predictions. A simple user interface allows users to input news text and receive instant results, forming a complete end-to-end detection system. Many YouTube videos come with transcripts, and summarizing them becomes especially helpful when dealing with long videos or varying segment relevance. Summaries allow users to focus on the most important parts, enhancing efficiency and comprehension. The Summarizer extracts information from video subtitles and applies different summarization techniques to produce user-friendly outputs in multiple formats, ensuring accessibility and ease of use.

II. LITERATURE REVIEW

One recent study provides a detailed overview of the latest trends in video summarization, particularly focusing on deep learning methods. It begins by outlining the motivations behind the emergence of video summarization technologies, then delves into the structure of a typical deep learning-based summarization pipeline. The authors also propose a classification system for existing algorithms, illustrating how these techniques have progressed over time and offering insights for future exploration.

Another research effort highlights the growing role of video summarization and skimming in modern video content management systems. It introduces advanced methods for condensing feature-length films and surveys current research in video abstraction. The study also shares the authors' own work, which incorporates cinematic principles and analyzes audiovisual pacing to enhance movie skimming. These developments are paving the way for automated systems that allow users to efficiently search and navigate video content through genre recognition and content analysis. A third perspective emphasizes the use of machine

ISSN [Online]: 2583-2654

learning and natural language processing to generate concise summaries of YouTube subtitles, ensuring that key information is preserved. With the increasing availability of video content—especially for educational use—on platforms like YouTube, Facebook, Instagram, and Google, there is a growing need for tools that help users quickly extract meaningful insights without watching entire videos.

Unlike static images where information can be extracted from a single frame, analyzing dynamic video content poses greater challenges. To address this, one proposed solution involves retrieving transcripts from user-submitted video links, processing the text using Hugging Face Transformers, and organizing the output through a pipeline. As noted in [4], automated summarization tools offer users a fast way to locate and explore relevant content across multiple media files. With the rise of more efficient capture devices, traditional cloud-based summarization systems—often slow to respond—are becoming less favorable. Instead, a method tailored for portable devices has been suggested, capable of generating summaries in real time while a webcam records live footage. This approach considers both internal content and external metadata, such as camera settings. The system achieved F-measure scores of 0.66 and 0.84 on the SumMe and SumLive datasets, respectively, while maintaining low energy consumption at just 20 milliamps on embedded hardware.

In another study [5], researchers reviewed the field of video classification, highlighting the use of diverse features derived from textual, auditory, and visual data. The study also cataloged commonly used attributes and key research contributions, concluding with recommendations for future exploration.

A novel online video summarization technique introduced in [6] enables rapid identification of key segments in raw, unedited footage—an otherwise time-intensive task for humans. This method uses cluster sparse coding to dynamically adapt lexical elements and build a vocabulary from the video. By combining segments that are only partially represented by this learned dictionary, a condensed version of the video is produced. Its online processing capability allows it to begin summarizing before the video ends, achieving near real-time performance with processing time closely matching the original video length.

The client attention model proposed in [7] evaluates how much attention viewers are likely to give to specific video content. This model aids in ranking and filtering videos based on cognitive engagement. The authors explore how semantic understanding and sensory cues—both visual and auditory—affect viewer focus. They also outline training strategies for attention modeling and present a summarization tool built on this framework. Notably, this method does not require deep semantic analysis or complex rule-based systems. The findings suggest that modeling user attention offers a promising alternative for understanding and summarizing video content.
Lastly, [8] points out that earlier summarization methods largely focused on optimizing the quality and variety of the generated summaries when designing their algorithms.

ISSN [Online]: 2583-2654

III. PROBLEM FORMULATION

With the surge in daily video content shared online, finding time to watch lengthy recordings has become increasingly difficult. Often, these videos exceed our expectations in duration, and without gaining meaningful insights, the time spent can feel wasted. To address this, automatically summarizing video transcripts offers an efficient way to extract key information without watching the entire content. Our system follows these steps to generate concise summaries:

1. The user initiates a request to our Flask-based backend server.
2. The server utilizes the YouTube API to fetch subtitles from the specified video.
3. Once retrieved, the backend processes and summarizes the transcript.
4. The final summary is delivered to the user in multiple languages, including Braille, Gujarati, Hindi, and English.

IV. METHODOLOGY

This section outlines the systematic approach adopted for transcript summarization, beginning with text preprocessing techniques aimed at enhancing data quality. It then introduces a punctuation restoration model, followed by three distinct extractive summarization algorithms and a hybrid method that integrates their core features. A brief overview of transformer-based models is provided, culminating in the use of three pretrained abstractive models for final summary generation.

A. Preprocessing

Preprocessing involves a sequence of operations designed to refine the input text and ensure consistency before summarization. The key steps include:

1. Text Cleaning – Unwanted elements such as symbols, special characters, and HTML tags are removed to prevent interference during summarization.

2. Tokenization – The text is segmented into individual words or sentences to facilitate further processing.

3. Stopword Removal – Common words like "and," "the," and "is" are excluded as they contribute little semantic value.

4. Contraction Expansion – Shortened forms (e.g., "isn't") are expanded to their full versions ("is not") for clarity.

5. Case Normalization – All text is converted to lowercase to maintain uniformity across the dataset.

Punctuation Restoration

Since transcripts often lack punctuation due to preprocessing, a restoration step is essential. The model used—felflare/bert-restore-punctuation from Hugging Face—is based on the bert-base-uncased architecture and fine-tuned using Yelp Reviews. It takes the merged transcript as input and outputs a version with appropriate punctuation, which is crucial for downstream extractive summarization.

B. Extractive Summarization Techniques

Three extractive methods are employed to identify and select key sentences from the transcript:

1. Luhn Algorithm

Sentences are tokenized and cleaned, then broken into words. Using NLTK, word frequencies are computed and common words are identified. Sentences are scored based on the frequency and proximity of these words, and top-ranked sentences are selected to form the summary.

ISSN [Online]: 2583-2654

2. *Keyword-Based Method*

Using SpaCy's pretrained NER models, named entities (e.g., people, organizations) are extracted. POS tagging identifies keywords from categories like nouns, verbs, adjectives, adverbs, and numerics.

Sentences are scored by keyword density, and selected sentences are arranged in their original order to produce the summary.

3. *TextRank Algorithm*

A similarity matrix is built using cosine similarity between sentences. PageRank is applied to this matrix with the following parameters:

- Damping factor: 0.85
- Convergence threshold: 0.00001
- Maximum iterations: 100

Sentences are ranked based on their scores, and the highest-scoring ones are chosen for the summary.

C. Hybrid Extractive Method

To enhance accuracy, the three extractive approaches are combined into a unified scoring system. Each sentence is evaluated based on:

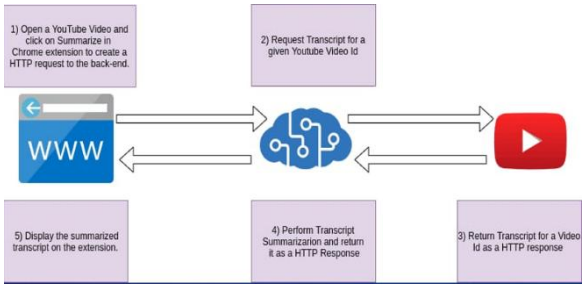
- **St(i)**: TextRank score
- **Sk(i)**: Keyword-based score
- **Sl(i)**: Luhn score

The final score is computed using the weighted formula:

$$\mathrm{Score}(i)=\frac{3\cdot St(i)+2\cdot Sk(i)+1\cdot Sl(i)}{6}$$

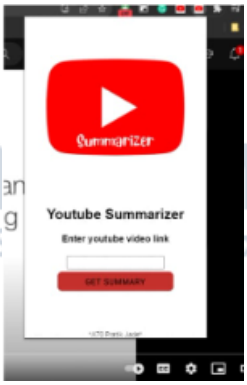
This composite score ensures a balanced consideration of frequency, keyword relevance, and sentence similarity.

When compared to simpler methods such as traditional machine-learning classifiers or single-layer neural networks, the proposed model delivered noticeably better performance. This improvement reflects the benefit of using deep contextual and sequential learning together.



V. RESULT AND ANALYSIS

Figure 2 presents the interface of the YouTube Transcript Summarizer tool. It features a navigation bar containing a field where users can paste video URLs and a "Summarize" button. Upon clicking this button, the application fetches and displays the transcript associated with the provided link.



VI. CONCLUSION

The YouTube video summarization paper has garnered significant attention from the research community, leading to the development of various algorithms and methodologies. This particular system integrates with a Chrome extension, allowing users to summarize YouTube videos directly from the Google Chrome browser. When the user clicks the "Summarize" button on the extension interface, the system retrieves the video's transcript using a Python-based API. The retrieved transcript is then processed and condensed using the Transformers library, and the resulting summary is displayed within the extension itself.

ISSN [Online]: 2583-2654

This tool offers substantial benefits by saving users time and effort, enabling them to quickly grasp the core message of a video without watching it in full. Additionally, it aids in detecting inappropriate or disturbing content, enhancing the overall viewing experience. The use of a Chrome extension ensures a seamless and user-friendly interface, eliminating the need for external applications to access the summarized content.

REFERENCES

[1] Hank Liao, Erik McDermott, Andrew Senior"Large scale deep neural network acoustic modeling with semi-supervised training data for YouTube video transcription" December 2013.

[2] Gaurav Sharma, Shaba Parveen Khan, Shivanshu Sharma, Syed Ubed Ali "Summarizer For Easy Video Assessment" Volume: 3 Issue:04-April-2021

[3] Atluri Naga, Laggisetti Valli, JahnaviDuvru "Video Transcript Summarizer" Issue: 11 March 2022

[4]AniquaDilawari, Muhammad usmanghani khan."Abstractive Summarization of Video Sequences" IEEE Access, 2019.

[5]. 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.

[6]. 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.

[7] https://journalppw.com/index.php/jpsp/article/view/9886/6457

[8] https://huggingface.co/transformers/

[9]https://atmamani.github.io/blog/building-restful-apis-with-flask-in-python/

[10] https://pypi.org/project/youtube-transcript-api/

[11] https://developer.chrome.com/docs/extensions/mv2/