

Video and Text Summarisation using NLP

Rushali R. Shetty¹, Samrudhi M.², Sayed Ayman Bukhari³, Dr. Manjunath DR⁴
Undergraduate Student^{1,2,3}, Assistant Professor⁴
B.M.S. College of Engineering Bengaluru, India

Abstract:- Chunks of Information are available on the internet, it is most important to provide a solution to get information most efficiently and accurately. With the increased use of the internet and smartphones, there has been a steady increase in online learning, entertainment, and other activities. But sometimes we don't have the time to go through all the material such as videos or podcasts in case it was too long. This necessitates a mechanism that reduces the length of a document to one with limited words which essentially retains the same meaning. Automatic Summarization can come in hand in this aspect.

I. INTRODUCTION

Due to the exponential increase in the availability of data in the recent years – amounting to almost a million terabytes – the need to summarize is imperative. With the ever-increasing pace of life, the public no longer has the time to read long articles or transcripts nor can they watch long videos or audio clips; a tool to develop an effective summary is much needed.

An enormous amount of digital data is sourced from the web resources on the Internet such as websites, user reviews, news, blogs, social media networks, YouTube videos, Podcasts, so on and so forth. Furthermore, another source of digital data could be found in the archives of various news articles, novels, books, legal documents, biomedical documents, scientific papers and primarily other audio and video formats, such as any OTT platform videos like Youtube, Amazon Prime, Saavan music, Spotify, etc. There has been a noticeable growth of digital content on the internet on a daily basis. As a result, users waste a lot of time trying to find the information they desire. They can't even read and comprehend all of the textual content in search results, much less the visual and audio content. Many portions of the resulting data are repeated or unimportant. As a result, summarising and condensing text, visual, and audio resources becomes urgent and critical. Manual summarization is an expensive task which requires a significant amount of time and effort. Manually summarising this massive amount of data is extremely difficult for humans. The key solution to this quandary is Automatic Summarization (AS).

The primary goal of an automatic summarization (AS) system is to generate a summary that includes the main ideas in the input document in as minimal space as possible while minimising repetition. Users will benefit from the automatically generated summaries, which will save them a significant amount of time and effort. This allows them to easily access important data and saves time.

The proposed system in the project is to create an extension such that it summarises transcripts of YouTube videos, podcasts, etc. We are looking at two approaches, the videos for which the transcripts are available, we will be directly summarising the content, whereas for the videos for which the transcripts are not available we are looking to convert the audio file to a text file, which will finally, be summarised using the algorithm. This would turn out to be particularly useful to save time and effort of either watching the entire video or having to read through the entire transcript of the desired video. Along with which it also is useful to save space if one is looking to download and save the transcript for later.

II. PROBLEM STATEMENT

When the user uses the web application, they are given the choice to upload a video or link to the video. This video input is passed through speech-to-text API to generate the subtitle for the same. Subtitles generated are passed through the algorithm of the user's choice. When passed through the algorithm a summarised text for subtitle is generated. According to the summarised subtitle, the video is trimmed. This trimmed video is made available for the user to download.

III. LITERATURE SURVEY

Wealth of information and expertise that must be adequately summarised in order to be useful. Because of the increasing availability of documents, there is a need for extensive research in the field of natural language processing (NLP) for automatic text summarization. Automatic text summarization is the process of creating a concise and fluent summary without the assistance of a human while maintaining the sense of the original text material.

A. Key Terms

- **Automatic Summarization:** The technique, where a computer program shortens longer texts and generates summaries to pass the intended message
- **Speech-to-API:** The Speech to Text API is a basic API that, as the name implies, allows you to transform audio input into written text.
- **Image analysis:** Image analysis or imagery analysis is the extraction of meaningful information from images; mainly from digital images by means of digital image processing techniques.
- **Natural language processing:** Natural language processing is a subfield of linguistics, computer science, and artificial intelligence concerned with computer-human interaction.

B. Existing System

Approaches to Automatic Text Summarization (ATS) are either extractive, abstractive, or hybrid. The extractive method chooses the most important sentences from the input documents and concatenates them to form the summary. The abstractive method represents the input documents in an intermediate representation before generating the summary with sentences that differ from the original sentences. The extractive and abstractive approaches are combined in the hybrid approach. Despite all of the methods proposed, the generated summaries are still far from human-generated summaries. The extractive approach is the focus of the majority of studies. It is necessary to place a greater emphasis on abstractive and hybrid approaches.

Automatic video summarization is implemented as a sequential decision-making process, and a diversity-representativeness reward is used to train a Summarizer to produce diverse and representative video summaries. The diversity reward computes the distance (expressing visual resemblance) of the selected key-frames from the remaining frames of the video, whereas the representativeness reward measures the dissimilarity among the selected key-frames.

C. Shortcomings of existing system

There already exists an Automatic Text Summarization (ATS) and on another stand alone application exists an Automatic Video Summarization. But they have not been done together before and that is our project's aim.

D. NLP, LexRank and LSA in Text summarisation :

In the big data age, the amount of text data available from various sources has exploded. This lengthy text contains a LexRank is a graphical method for automatically summarising text. It's also an unsupervised method for text summarization. The eigenvector centrality concept is used, and sentence importance is calculated. In this model, the adjacency matrix of the graph representation of sentences is a connectivity matrix based on intra-sentence cosine similarity. This sentence extraction is primarily concerned with a set of sentences with the same intent, in which a centroid sentence is chosen to serve as the mean for all other sentences in the document. The sentences are then ranked based on their similarities.

Latent semantic analysis (LSA) is a method for extracting a representation of text semantics from observed words that is unsupervised. It was first proposed in the news domain to select highly ranked sentences for single and multi-document summarization. The LSA method begins by constructing a term-sentence matrix (n by m matrix), in which each row corresponds to a word from the input (n words) and each column corresponds to a sentence (m sentences).

E. Applications of the proposed system :

- **E-learning and class assignments:** Many teachers frame their lectures with case studies and current events. Summarization can help teachers update their content more quickly by producing summarised reports on their subject of interest. This will also make it easier for the student to learn the important topics.
- **Video Scripting:** Video is quickly becoming one of the most important marketing tools. People are now sharing videos on professional networks such as LinkedIn, in addition to video-focused platforms such as YouTube or Vimeo. Scripting may be required in varying degrees depending on the type of video. When attempting to write a script that incorporates research from multiple sources, summarization can be a valuable ally.
- **Medical cases:** With the rise of tele-health, there is an increasing need to better manage medical cases, which are now entirely digital. Because telemedicine networks promise a more open and accessible healthcare system, technology must make the process scalable. When it comes to analysing medical cases and routing them to the appropriate health professional, summarization can be a critical component in the tele-health supply chain.

Media monitoring, newsletters, search marketing and SEO, internal document workflow, financial research, legal contract analysis, social media marketing, question answering, and bots are just a few examples.

IV. PROPOSED SOLUTION

A. Requirements

The proposed solution must satisfy the following functional requirements and provide the following features:

- **User Interface:** Users will link to the video to summarise or upload the file itself. Summarized video is available to download later.
- **Subtitle generation:** Subtitle is generated using speech-to-text api
- **Text summarisation:** Extracted subtitles are summarised using the extractive text summarisation approach. Algorithms such as Lex Rank and LSA are used.
- **Video trimmer:** Video is trimmed based on the summarised subtitle.

B. Key Components

- Video Input
- Subtitle generator
- Summarizer
- Video trimmer.

C. Proposed Methodology

The goal of this project is to build an accurate video summarisation webapp using its subtitles. To complete this goal we will be using the following steps -

- Accept the video input from the user in the form of actual video or URL. Accept the subtitle file if provided by the user.
- Download the video if the URL is provided.
- Generate audio files for the video if subtitle is not available. Convert the audio file to text file using speech-to-text API.
- Accept the algorithm selected by the user.
- Pass the subtitle to the selected algorithm and generate the summarised subtitle.
- According to the summarised subtitle, trim the original video.
- Trimmed video is made available for the user to download.

V. CONCLUSION

Manual summarization is a time-consuming and expensive task with numerous steps. For example, to manually summarise a single dataset, the following steps are taken: understanding what the document / video is about, extracting the "most important" parts from it, and composing a summary that meets the following requirements: summary readability and linguistic quality, summary consistency and content coverage. The absence of redundancy in the produced summary.

Because of the difficulty of manually summarising the massive amount of textual content on the Internet or in various archives, AS systems have emerged as the primary technology to address this urgent and pressing issue.

Automatic summarization is a complex task with numerous subtasks. Every subtask is capable of producing high-quality summaries. The extracted sentences are produced as a summarised text and converted into audio form. When compared to a traditional approach, the proposed model improves accuracy.

REFERENCES

- [1.] Z. Xian, F. Jiao, "Iterative video summarization, using ILS-SUMM", vol. 405, pp. 200 – 207, 2020.
- [2.] N. Gonuguntla, B. Mandal, N. Pohan et al., "Enhanced Deep Video Summarization Network," in 2019 British Machine Vision Conf. (BMVC), 2019.
- [3.] Song, Y., Vallmitjana, J., Stent, A., & Jaimes, A. (2015). Tvsum: Summarizing web videos using titles. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5179- 5187)
- [4.] Agyeman, R., Muhammad, R., & Choi, G. S. (2019, March). Soccer Video Summarization Using Deep Learning. In 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR) (pp. 270-273). IEEE.

- [5.] K. Zhou and Y. Qiao, "Deep Reinforcement Learning for Unsupervised Video Summarization with Diversity-Representativeness Reward," in Proc. of the 2018 AAAI Conf. on Artificial Intelligence, 2018.
- [6.] B. Zhao, X. Li, and X. Lu, "Property-constrained dual learning for video summarization," IEEE Trans. on Neural Networks and Learning Systems, vol. 31, no. 10, pp. 3989–4000, 2020.
- [7.] L. Lebron Casas and E. Koblenz, "Video Summarization with LSTM and Deep Attention Models," in MultiMedia Modeling, I. Kompatsiaris, B. Huet, V. Mezaris, C. Gurrin, W.-H. Cheng, and S. Vrochidis, Eds. Cham: Springer International Publishing, 2019, pp. 67–79.
- [8.] Niladri Chatterjee, Amol Mittal and Shubham Goyal's "Single Document Extractive Text Summarization Using Genetic Algorithms" (2012)
- [9.] Amol Tandel, Brijesh Modi, Priyasha Gupta, Shreya Wagle and Sujata Khedkar's "Multi-document text summarization - A survey" 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE).
- [10.] Aditya Jain, Divij Bhatia, Manish K Thakur's "Extractive Text Summarization using Word Vector Embedding"(2017).
- [11.] Shi Ziyan "The Design and Implementation of Domain-specific Text Summarization System based on Co-reference Resolution Algorithm" 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery.
- [12.] Leonhard Hennig, Winfried Umbrath, and Robert Wetzker. 2008. An ontology based approach to text summarization. In Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT'08. IEEE/WIC/ACM International Conference on, Vol. 3. IEEE, 291–294.
- [13.] Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In Text Summarization Branches Out: Proceedings of the ACL-04 Workshop. 74–81.