Automatic Summarization of Instructional Videos

D. Bhavitha[1], A. Kalaivani[2]

[1]Final Year CSE Student, [2]Associate Professor,
[1,2]Department of Computer Science and Engineering,
Saveetha School of Engineering,
Saveetha Institute of Medical and Technical Sciences, Chennai
[1]duvvurubhavithareddy111@gmail.com, [2]kalaivanianbarasan@rediffmail.com

Abstract

In recent times video plays a major role in transferring the important information and a many people are showing interest for browsing, retrieval of video contents. Students now a days uses video as a tool during their revision times. Extracting audio from video is very challenging. The major goal of text summarization is to extract the important information from large amount of data and to give highly efficient content. The proposed system initially extracts the audio content from the instructional videos and then the extracted audio is converted into text transcript. The text transcripts are pre-processed and then the major features are extracted from the text and the sentence rankings have given for all the sentences which gives the summarized text. The evaluation of summarized text is very major because it shows the correctness in our summarized text. There is a high challenge to reduce the much of this converted text to smaller, high focused summarization that captures the various forms of details, such that we can easily navigate it more efficiently and effectively, as well as to check whether the data with large amount of text contains the information and data that we are looking for. It makes the selection process easier, improves the effectiveness of indexing.

Keywords: Automatic Summarization, Instructional Videos, Summary Length, Text Transcript.

1. Introduction

The main objective is to summarize the text from the instructional videos. The recorded lectures are transcripted from audio to text format. The transcripted text is then summarized using extractive summarization technique which performs operations like preprocessing (which identifies words, sentences, stop words) ,extracts the features(like word similarities, sentence length, numerical data, sentence position), sentence ranking is given , redundancy reducing is checked and then finally the summarized text is produced . The major goal is that students should find that these summary lectures are very useful for reviewing lecture material as well as for their revision. This highly improves productivity as it speeds up the surfing process instead of reading full document which may contain useless information. The importance of the project is to allow the user to read less data but still the user receive the most important information and make solid conclusions. Today's computers are far more powerful than the human mind and is most likely that computer will create a good summary before the human will have a chance to look at the article. There is a great need to reduce much of this text data to shorter, focused summaries that captures the salient details, both so that we can navigate it more effectively as well as check whether the larger documents contain the information that we are looking for. It makes the selection process easier, improves the effectiveness of indexing.

Text summarization can be categorized into two main parts— Extractive and Abstractive Summarization.

Extractive Summarization method depends on extracting various parts of information from the text given, like sentences, noun phrases, and words, from a bit of text and bind them together to create and generate as summary. Hence, identifying those correct and exact sentences for summarization is very important in an extractive summarization method.

Abstractive Summarization uses an advanced and updated NLP technique to produce a completely new text summary. Few sections of this summarization will not appear in the original text which we have in the text file. In our proposed system we are focusing on the extractive summarization technique because it extracts the text which is present in the main document only but abstractive summarization will generate entirely new summarization.

In this paper we are going to discuss about the Extractive Summarization.
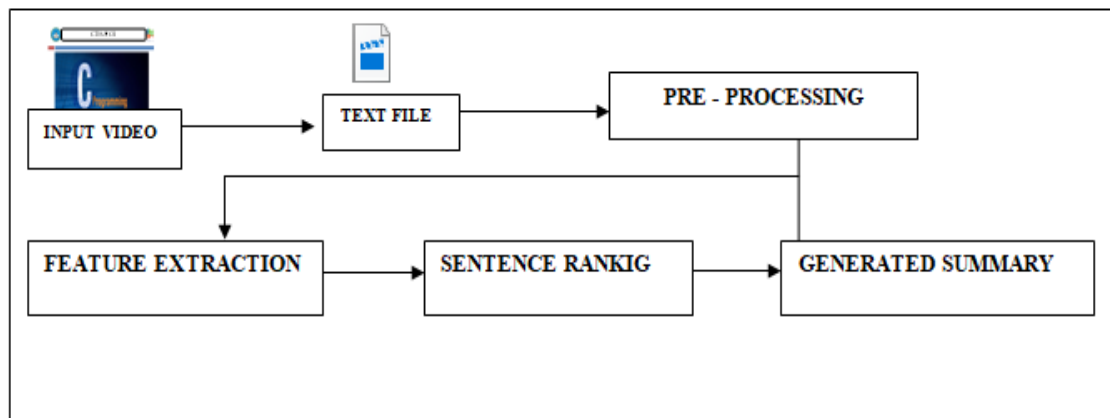

Figure 1: System Flow

The rest of the paper may be prepared as follows. In section 2 affords literature survey on video summarization. Section 3 explains the block diagram of the proposed system and section 4 describes the methodology of the proposed system. Finally section 5 highlights the outcomes at each and every section of the automatic video summarization of the instructional videos. Finally, section 6 gives the conclusions of the proposed work in the form of graphs that specifies the emotion levels of the tweets.

2. Literature Survey

Different authors have executed their research work inside the subject of Data pre-processing on one kind of domain names and proved suitable efficiency and accuracy in getting in getting rid of noisy information with the aid of distinctive strategies. Some of the authors specially concentrated on stop phrase removal for achieving higher accuracy.

Hitoshi Nishikawa, Tsutomu Hirao, proposed a novel text summarization model, the repetition compelled rucksack model. They have added the Knapsack issue limitation to check excess in the outline. They likewise proposed a quick translating strategy dependent on the Lagrange heuristic. Investigations dependent on rouge assessments show that our recommendations outflank a best in class content rundown model, the most extreme inclusion model, in finding the ideal arrangement. They additionally show that our translating technique rapidly finds a decent inexact arrangement equivalent to the ideal arrangement of the most extreme inclusion model. It additionally improves the nature of outlines altogether contrasted with a best in class framework, the greatest inclusion model. Their model can be decoded by the Lagrange heuristic.

Laxmi B. Rananavare, P. Venkata Subba Reddy, covers about extractive, abstractive outline and assessment strategies. Outline framework should create a successful rundown in a brief timeframe with less repetition having syntactically right sentences. Both extractive and abstractive strategy yields great outcome as indicated by the setting in which they utilized. The explored writing opens up the difficult region for hybridization of these strategies to create educational, all around compacted and comprehensible outlines. Different strategies for assessing rundown frameworks are examined.

Sachan Priyamvada Rajendra, Dr. Keshaveni N , examined that in the course of recent years, there has been an enormous increment in measure of video content made. Gigantic development in video content postures issue of data over-burden and the board of substance. So as to deal with the developing recordings on the web and furthermore to separate an effective and legitimate data from the recordings, more consideration must be paid towards video and picture handling innovations. Video synopses give dense and brief portrayals of the substance of a video stream through a mix of still pictures, video sections, graphical portrayals and printed descriptors.

Over the most recent couple of decades part of work has been done around there. They have audited a portion of the ongoing work on content-based mixed media data recovery and examines their job in flow look into

bearings which incorporate perusing and search standards, client contemplates, successful figuring, learning, semantic questions, new highlights and media types, elite ordering, and assessment procedures. In light of the present best in class, they likewise talked about the significant difficulties for what's to come.

Nedunchelian.R, Saranathan.E, Muthucumarasamy.R, introduced the most significant sentences from such gigantic storehouse of information hush up troublesome and requesting task. While multi archive represents some extra overhead in sentence determination, producing synopses for multi report in a cognizant request would make more noteworthy quality. The proposed approach utilized for a few highlights to produce outlines and to go about as trainable summarizer. This methodology utilized

hereditary calculation to test the exhibition of summarizer. It is contrasted with territory of MEAD Algorithm and Naive Bayesian Classifier. The exhibition is researched at a few pressure rates from various profile of mainstream people gathered from news stories. Multi record rundown has extremely incredible effect in the web world since the time the development of online data and accessibility choosing.

The above literature surveys have given a brief explanation about both the abstractive and extractive summarization. The rouge evaluations techniques to perform a state-of-the-art text summarization model, the maximum coverage model, in finding the optimal solution is explained.
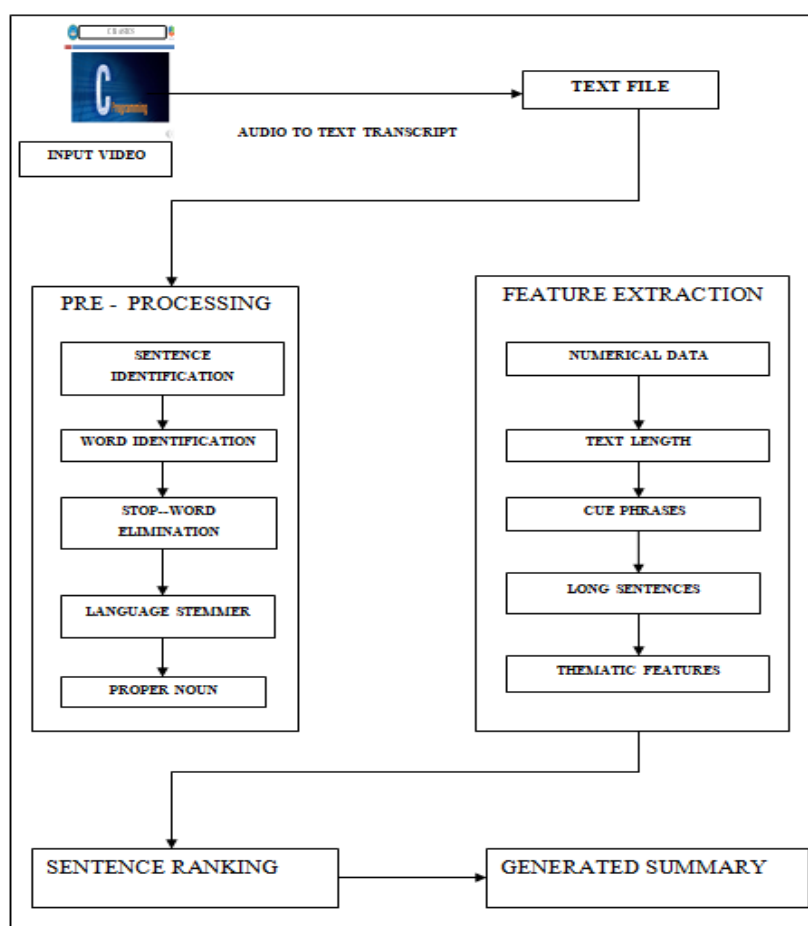
3. System Architecture & Methodology

Figure 2: Architecture Diagram

Video summarization process can be categorized into three broad phases: The first phase is Pre Processing phase it is the structural representation of the original text. Various features which effects the relevance of sentences and the words are been calculated. In this phase, using feature weight equation the final score of

each and every sentence is calculated and determined. In the last phase, the Top most ranked sentences have been selected for the generated final summary.

PRE-PROCESSING

Pre-preparing is organized portrayal of the first inputted content. The significance of pre-processing is utilized in pretty much every created framework related with content preparing and normal language preparing. This stage incorporates words distinguishing proof, sentences ID, stop words disposal, language stemmer for things and appropriate names, permitting contribution to legitimate arrangement and end of copy sentences or words. Pre-processing stage decrease size of content.

FEATURE EXTRACTION

One of the perspective towards text summarization, utilizes a combination of few features to achieve feature scores of each sentence.

1. Sentence Length

It is discovered that the nearness of increasingly content words in a sentence makes the sentence an enlightening one. Along these lines, we think about long sentences.

2. Sentence Position

The Sentences which are presented in the first line and last line of a paragraph are always been considered as the most informative sentences in the text.

3. Thematic Fearures:

Thematic word implies showing up more as often as possible in a book report. Nearness of this type of words in a sentence connotes the sentence as an enlightening one. In our technique, we consider top ten most incessant words from the content record and consider those sentences where these words are available.

4. Numerical data

Numerical data give more information about the numerical type of data present in a text document. A sentence which contains numerical type of data is always considered to be an important sentence.

SENTENCE RANKING

1. The initial step is connect all the content contained in the articles

2. At that point split the content into singular sentences

3. In the subsequent stage, we will discover vector portrayal (word embeddings) for every single sentence

4. Similitude's between sentence vectors are then determined and put away in a network

5. The closeness network is then changed over into a chart, with sentences as vertices and similitude scores as edges, for sentence rank count.

6. At long last, a specific number of top-positioned sentences will be taken as final generated summary.

GENERATED SUMMARY

Finally a summary has been generated which is useful for students to review the whole document in a short period of time

4. Results And Discussion

Conversion of video format is very easy. First we have to take the mp4 format video and convert it into avi format. As we are focusing on presentational videos we have to take the audio from the video. We can easily extract the audio when the video is in avi format. Then we have to extract the audio.
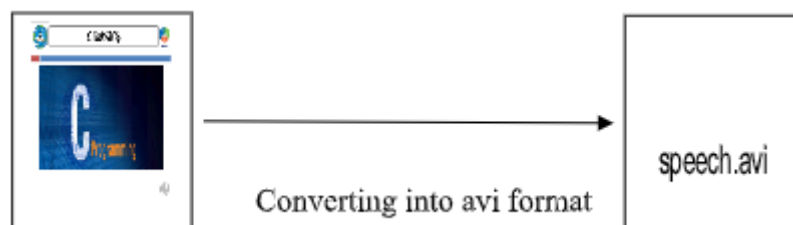

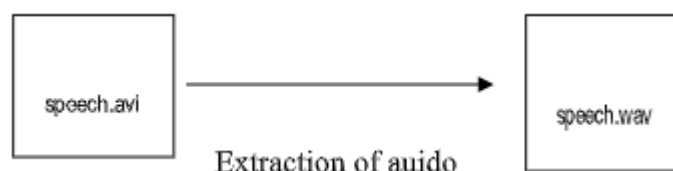Figure 3: Conversion from mp4 video format to avi format

Figure 4: Extraction of audio file from video file

Figure 5: Extraction of text transcript from the audio file

Figure 6: Pre-Processing

Figure 7: Feature Extraction

Figure 8: Generated Summary

After generating the summary we have measure whether our generated summary is correct or not. So we have use evaluation metrics for checking the correctness of summary. In this paper I am using Precision, Recall and F-Score values for checking the evaluation of summaries. Here System generated summary is the summary generated by using code and reference summary is online generated summary.

Recall tells that how much of the reference summary is the system summary recovering or capturing. If we are just considering the individual words, it can be computed as:

Recall = Number of overlapping words between original text and reference summary / Total number of words in reference summary

In terms of precision, what you are essentially measuring is, how much of the system summary *was* in fact relevant or needed? Precision is measured as:

Precision = Number of overlapping words between original text and system generated summary/ System Generated Summary

The F score is used to provide and give the more realistic measures of a test evaluated performance by using both the precision and recall methods.

F-Score = (2 * Precision * Recall) / (Precision + Recall)

The below table gives the values for recall, precision and f-score of different instructional videos.

Table 1: Comparison Table

	RECALL	PRECISION	F-SCORE
Video 1	0.6785714285714286	0.6082474226804123	0.6414878397711016
Video 2	0.6	0.6206896551724138	0.6101694915254238
Video 3	0.575	0.5466666666666666	0.5604754829123328
Video 4	0.6213592233009708	0.5	0.5541125541125541
Video 5	0.5773195876288659	0.5827338129496403	0.5800140655968288

From the table we can see difference between the values of precision and recall and it clearly states the system generated summarization is having a slight higher efficiency when compared to online generated summary.

5. Conclusion

In this project, we given a new method to summarize the visual and audio content in the instructional videos. We first presented a probabilistic model which extracts the audio content and converts that into text format, then we applied extractive techniques to summarize the transcripted text.

The Feedback has taken from the students and as a result it indicates that they have felt the summarized text we have produced were very useful to study. It also helps the students to understans the topic easily even though if they have missed the session during lecture time. Providing text along with the video will be useful for some students during their revision time.

In future we are going to improve that the best quality of the recording which should be as good as possible in terms of quality, using better equipment than we used before, with good lighting environment and sound recording. The technology should be robust so that breaks in transmission do not occur, and streaming is recommended for summary videos.

References

[1] Gang Pan, Yaoxian Zheng, Rufei Zhang, Zhenjun Han, Di Sun, Xingming Qu, "A bottom-up summarization algorithm for videos in the wild", EURASIP Journal on Advances in Signal Processing , December 2019.

[2] Shruti Palaskar, Jindrich Libovický, Spandana Gella, Florian Metze, "Multimodal Abstractive Summarization for How2 Videos", Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 6587–6596 Florence, Italy, July 28 - August 2, 2019.

[3] Martin Malchow, Matthias Bauer, Christoph Meinel, "Enhance Learning in a Video Lecture Archive with Annotations" , IEEE Global Engineering Education Conference (EDUCON), 17-20 April, 2018, Santa Cruz de Tenerife, Canary Islands, Spain. 2018.

[4] Shashi Narayan , Shay B. Cohen, Mirella Lapata "Ranking Sentences for Extractive Summarization with Reinforcement Learning ", Proceedings of NAACL-HLT 2018, pages 1747–1759 New Orleans, Louisiana, June 1 - 6, 2018.

[5] Laxmi B. Rananavare, P. Venkata Subba Reddy, "An Overview of Text Summarization ", International Journal of Computer Applications (0975 - 8887) Volume 171 - No.10, September 2017.

[6] Hitoshi Nishikawa , Tsutomu Hirao, "Text Summarization Model based on Redundancy-Constrained Knapsack Problem " , Proceedings of COLING 2012: Posters, pages 893–902, COLING 2012, Mumbai, December 2012.

[7] Chekuri Choudary, " Summarization of Visual Content in Instructional Videos ", IEEE TRANSACTIONS ON MULTIMEDIA, 2008

[8] C. Choudary and T. Liu, "Extracting content

from instructional videos by statistical modeling and classification," Int. J. Pattern Anal. Applicat., vol. 10, no. 2, pp. 69–81, 2007

[9]   Janice Whatley and Amrey Ahmad, "Using Video to Record Summary Lectures to Aid Students' Revision" Interdisciplinary Journal of Knowledge and Learning Objects Volume 3, 2007.

[10]   Lu, S., Lyu, M. R., King, I., & Sar, H. K, "Semantic Video Summarization Using Mutual Reinforcement Principle and Shot Arrangement Patterns.", 2005.