International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-6, April 2020

Video to Text conversion and Abstractive Summarization for Effective Understanding and Documentation

T Swapna, Y Sravani Devi, K Sindhura

Abstract*: Video is one in every of the sturdy sources of data and the consumption of on-line and offline videos has reached a new level within the previous few years. An elementary challenge of extracting data from videos is, a viewer should undergo the whole video to grasp the context, as against a picture wherever the viewer will extract data from one frame. Typically, protracted videos also are quite troublesome to follow because of reasons like totally different pronunciation, pace then on. Abstractive Text summarization extracts the utmost important information from a source which is a text and provides the adequate outline of an equivalent. The analysis work conferred during this paper describes a straightforward and effective methodology for video Summarization. It principally targets academic and technical videos. Speech is extracted from video. The speech is regenerate to the corresponding text using abstractive summarization technique and produces summarized text. For quicker conversion of video to text GPU can be used. This has numerous applications like lecture notes creation, summarizing catalogues for protracted documents then on.

Keywords: Video Summarization, Vision, Deep Learning, Abstractive Text summarization.

I. INTRODUCTION

Following the advances of economical knowledge storage and streaming technologies, videos became arguably the first supply of data in today's social media-heavy culture and society. Video streaming sites like YouTube are quickly replacing the standard news and media sharing strategies whom themselves are forced to adapt the trend of posting videos rather than written articles to convey stories, news and knowledge. The scholars at numerous levels also are curious about learning through the videos This abundance of videos includes new challenges concerning an efficient way to extract the subject matter of the videos in question. It might be frustrating, inefficient, unintelligent and downright not possible to observe all videos associated with one conception to know and to amass the data out of it.

Revised Manuscript Received on April 30, 2020.
***** Correspondence Author
T.Swapna*,CSE Department, GNITS, Hyderabad, India. muluguswapna@gmail.com
Y.Sravani Devi, **CSE** Department, GNITS, Hyderabad, India. y.sravanidevi@gnits.ac.in
K.Sindhura, CSE, GNITS, Hyderabad, India sindhureddy4u@gmail.com

Text report is one in all the key ideas utilized in the sector of documentation. Long documents are tough to scan and perceive because it consumes heap of your time. Text summarization solves this drawback by providing a shortened outline of it with linguistics. Within the planned work a mixture of video to text conversion and text report is enforced. This hybrid methodology can aid applications that need transient outline of long videos that is sort of helpful for documentation. The outline of videos of same kind within the text format may be done. As an example, if we have three videos on a subject then we can get outline of that topic by summarizing those three videos.

II. MOTIVATION

There are some videos that are of terribly long length. The spectator of the video may not have abundant amount of time in order to travel through the entire video. The viewer simply desires to understand the summary of that video. Again, and again it conjointly takes place that, the spectator watches the entire video on the actual concept and may conclude that, the video is not an appropriate one for which he/she is looking for. It's a great aid if user gets the outline of video in text format. It saves user's precious time. This downside may be solved by intelligent Summarization of Videos which is able to be helpful for instructional purpose wherever the time of scholars may be saved and that they will have like notes of that video.

III. RELATED WORK

Abstraction summarization of Video Sequences that used deep learning technique in order to get the linguistic interpretation associated with the video. This provides a textual-based video summary, theoretic outline, facultative users to discriminate between relevant and orthogonal data in keeping with their wants, stated by Aniqa Dilawari et.al [1] Speech to text conversion finds applications in varied situations. An efficient technique to achieve fluency in English Language that enhances the user's means of speech through correctness of pronunciation following the English phonetics was developed by Jose et al. [2]. In [3] feature extraction supported neural networks was planned that authors claim to be simpler compared to the web extractive choices. Zenkert at el. [4] introduced a cross-dimensional text summarization that uses the conception of dimensional choice and filtering.

The method was experimented exploitation the results of Multidimensional knowledge representation info.

A text instrument was developed by Devasena and Hemalatha [5] that was used to determine the structure of the text given as input. The authors claim the planned system was able to offer the results effectively that had used the automated text categorization, and text summarization. There exist totally different text summarization techniques. An in-depth summary of constant is planned in [6] by Rahimi et al. an analogous study was done by Dalal and the leader additionally [7].

IV. ARCHITECTURE

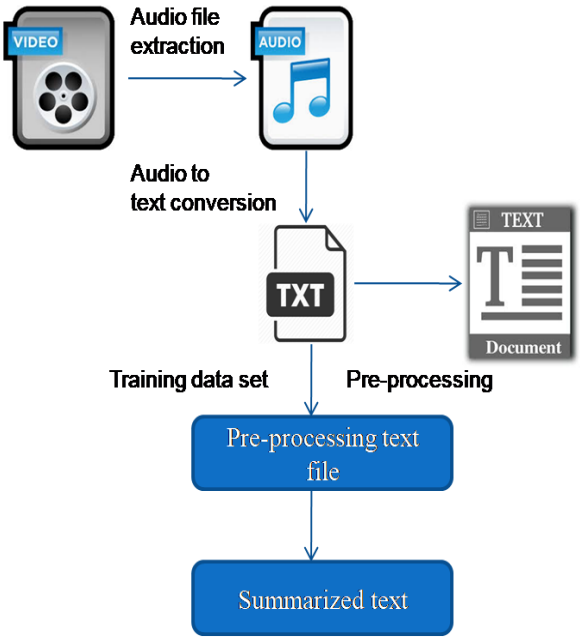

Figure1.The Proposed System

The Proposed system shown in Figure 1, mainly targets at the generation of text file from video file. This can be achieved by using various application program interfaces. The shortened form of that text file can be produced by using any summarization technique.

The process is comprised of the following steps.

- **Extraction of audio from video:** Since it is not possible to convert video into text directly, audio is extracted from video using "ffmpeg API".
- **Transformation of speech file into a text file:** The next task is to convert that audio into text using Speech recognition library. Hence the corresponding text can be obtained.
- **Preparation of Text:** The text extracted from the above step must be preprocessed. This phase is to prepare the data ready for summarization. Pre-processing involves stop words removal, stemming etc.
- **Generation of Text Summary:** The preprocessed text is given as input to the summarization module which generates synopsis of the taken video. This summarization module uses Abstractive summarization approach to get summary.

The generated summary is useful for the user to get the actual outline of the video. The user can maintain it as a lecture note. It saves user time. It can be used for documentation purpose.

The above procedure can be applied to multiple videos of the same domain. So, users instead of watching multiple videos on the same concept, just can go through the synopsis generated by this system for brief and quick understanding.

V. IMPLEMENTATION

The main task during this system is to get a good outline of the video which is influenced by the summarization technique used. Extractive summarization and Abstractive summarization are the approaches used for automatic text summarization. Extractive is retrieving and combining important sentences to generate summary whereas Abstractive Summarization involves rephrasing of the sentences most like human summarization. In general, abstraction will precise the text a lot of powerfully than extraction. There are many abstractive summarizers are available. One of such summarizers is Seq2Seq.

Sequence-to-Sequence (Seq2Seq) Modeling (Abstractive Text Summarizer)

The Seq2Seq model takes the multi-line text data as input and converts that into a precise summary. The input sentences may be lengthy, but the output is short and semantically condensed. Figure 2 is the illustration of Seq2Seq model [8]:

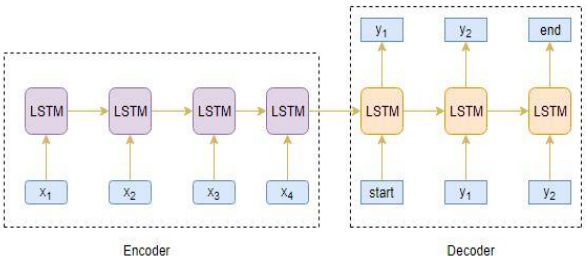

Figure2: Architecture of Seq2Seq Modeling

The important components of Seq2Seqmodel are:

1.Encoder

2.Decoder

Encoder

It uses neural network layers and transforms the input words into hidden vectors. Every vector refers the present word and the context of the word.

Decoder

The decoder is also a network which reads the words one by one from the target sequence and estimates the same sequence offset by one-time step**.** The decoder is well trained to estimate the succeeding word in the sequence given the preceding word.

Recurrent Neural Networks (RNNs), Gated Recurrent Neural Network (GRU) or Long Short-Term Memory (LSTM), are preferred as the encoder and decoder components.

The above mentioned can capture long term dependencies by defeating the problem of vanishing gradient.
The Encoder-Decoder is established in 2 phases:

Training phase:

In this phase, firstly the encoder and decoder are set up, then a model is trained to predict the target series offset by just one-time step. Associate Encoder Long Short-Term Memory model (LSTM) can hold the entire input sentence wherein, at every time step, a single word is given as input to the encoder. It then processes the information at every time step and acquire the contextual information present in the input sequence. The decoder is also an associate LSTM network that takes one word at a time from the entire target sequence and estimates constant sequence offset by just once step. The decoder is then trained to predict ensuing word within the sequence given the preceding word.

Inference Phase

Once the trained model is ready, then it is tested on new input sentences whose target sequence is not known. So, we use the inference architecture in order to decode a test sequence. This won't work for long sentences as the decoder is looking at entire input sentence for the prediction.

Problem arises with the lengthy sentences. it's very difficult for the encoder to recall lengthy sequences into a fixed length vector. So, attention mechanism is used. It targets to predict a word by looking at a couple of clearly identified components of the sequence solely, instead of observing the whole sequence. In global attention, attention is placed on all the supply positions wherever in local attention, attention is placed on solely few positions.

VI. EXPERIMENTAL SETUP

Dataset

YouTube conceptual videos and Nptel videos are converted into text file by using ffmeg module and speech recognition module. For the construction of abstractive summarizer, CNN/Daily Mail dataset is used [9]. It has short stories related to news in English Language. The dataset was generated from CNN and Daily Mail datasets. Every story is attached with multi line summary written by human expert. The considered dataset has 287,226 training, 13,368 validation and 11,490 test tuples.

Evaluation Measure

The measure considered for the evaluation of generated text summary is ROUGE [10] metric. It estimates the quality of the system generated summary against human generated summary. It is a set of metrics includes ROUGE-1, ROUGE-2 AND ROUGE-L. These indicates the overlapping's of uni-word, Two word and lengthiest common sub sequence between system summary and manual summary.

Results and Discussion

Table1.shows the experimental results on CNN-Daily Mail Dataset

Method	Rough-1	Rough-2	Rouge-L
Seq2Seq-RNN	42.04	19.77	39.42

Table1. Experiment Results

Our model generated abstractive summary is shown below.

Original Text
Let us initial see what's R. R is open source language which is widely used as an applied mathematics software package and information analysis tool. R usually comes with the command interface. R is offered across wide used platforms, windows, linux and macOS currently, let us see, what's R Studio.

Abstractive summary
R is open source language used as applied mathematics software package and information analysis tool. R with command interface. R is offered on windows, linux and macOS. What's R-Studio.

The Proposed system has given good rouge values on machine summary over manual summary. Its performance is acceptable over small to moderate sized text documents**.** The results were shown in Table 2.

Table2: Rough values on sample input

	Precision	Recall	F-Score
Rouge-1	0.68	0.72	0.74
Rouge-2	0.6	0.74	0.67
Rouge-L	0.75	0.85	0.75

VII. CONCLUSION

A Video Conversion and Text Summarization are two huge areas to be explored. This analysis work aims to cut back the time, and energy of manual documentation of prolonged videos of an incident. This analysis eases the work of documentation. Video to text conversion is done using API's. An Abstractive text summarization mechanism SeqtoSeq Model with attention mechanism is employed for generating text summaries. This model can be used wherever there is a demand of summarizing prolonged lectures into precise documents, because the machine-driven system can convert the video to text, and conjointly summarize the content. It will be of nice facilitating for college students to archive lecture notes from classes, conferences or seminars. This may be enforced on NPTEL Videos that became a serious supply for college students to get aquatinted with the latest technical topics. The long run work to the current is, including pictures, written formulae, and diagrams, etc. present within the video input in the summarized output.

REFERENCES

1. Aniqa Dilawari 1 And Muhammad Usman Ghani Khan1"ASoVS: Abstractive Summarization of Video Sequences", IEEE
2. Jose D V, Alfateh Mustafa, Sharan R, "A Novel Model for Speech to Text Conversion," International Refereed Journal of Engineering and Science (IRJES), vol 3, no. 1, 2014.
3. Jain D. Bhatia and M. K. Thakur, "Extractive Text Summarization Using Word Vector Embedding," 2017 International Conference on Machine Learning and Data Science (MLDS), Noida, pp. 51-55, 2017.
4. J. Zenkert, A. Klahold and M. Fathi, "Towards Extractive Text Summarization Using Multidimensional Knowledge Representation," 2018 IEEE International Conference on Electro/Information Technology (EIT), Rochester, MI, pp. 0826-0831, 2018
5. C. Lakshmi Devasena and M. Hemalatha, "Automatic Text categorization and summarization using rule reduction," IEEE-International Conference on Advances in Engineering, Science and Management (ICAESM -2012), Nagapattinam, Tamil Nadu, pp. 594-598, 2012.

6. S. R. Rahimi, A. T. Mozhdehi and M. Abdolahi, "An overview on extractive text summarization," 2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI), Tehran, pp. 0054-0062, 2017.
7. V. Dalal and L. Malik, "A Survey of Extractive and Abstractive Text Summarization Techniques," 2013 6th International Conference on Emerging Trends in Engineering and Technology, Nagpur, pp. 109-110, 2013.
8. Yong Zhang , Dan Li , Yuheng Wang, Yang Fang and Weidong Xiao,"Abstractive text summarization with a convolutional Seq2Seq model" Appl. Sci. 2019, 9, 1665; doi:10.3390/app9081665
9. K. M. Hermann et al., ''Teaching machines to read and comprehend,'' in Proc. Adv. Neural Inf. Process. Syst., 2015, pp. 1693–1701
10. C. Y. Lin, ''Rouge: A package for automatic evaluation of summaries,'' in Proc. Text Summarization Branches Out, Workshop (ACL), vol. 8, Jul. 2004, pp. 25–26.

AUTHORS PROFILE

Mrs. T.Swapna , Assistant Professor at GNITS, Hyderabad, has 14 yrs of teaching experience. Completed M.Tech from JNTU ,Hyderabad. She has 2 papers in National and International conferences and in refereed journals. Her research interests include, Artificial Intelligence, Image Processing and Data Mining

Mrs. Y.Sravani Devi , Assistant Professor at GNITS, Hyderabad, has 8 yrs of teaching experience. Pursuing Ph.D from GITAM Deemed to be University, Hyderabad, Completed M.Tech from JNTU Kakinada. She has 7 papers in National and International conferences and in refereed journals. Her research interests include, Artificial Intelligence and Data mining and Image Processing.

Mrs.K.Sindhura, Assistant Professor at GNITS, Hyderabad, has 15 yrs of teaching experience. Pursuing Ph.D from KL Deemed to be University, Hyderabad, Completed M.Tech from JNTUA. She has 6 papers in National and International conferences and in refereed journals. Her research interests include, Artificial Intelligence and Data mining and Image Processing.