INDIAN INSTITUTE OF INFORMATION TECHNOLOGY, ALLAHABAD

PROJECT REPORT

on

SPEECH TO TEXT CONVERSION AND SUMMARIZATION

Presented by:

1. **Tejas Ramesh Pawar** (IIT2017109)
2. **Kunal Kumar Prasad** (IIT2017112)
3. **Salil Srivastava** (IIT2017113)

Under the Supervision of:

Dr. K. P. Singh

September, 2020

Contents

ABSTRACT

Speech is the most effective way of communication with which human beings communicate their thoughts and feelings. However, even while speaking in the same language, the speed and the dialect varies with each person. Speech recognition, which is an interdisciplinary computational linguistics field, helps to develop technologies that allow speech to be understood and translated into text. Text summarization extracts the most relevant information from a source, which is a text, and provides the required description of it.

Our work includes a simple and efficient method for conversion of speech to text and a description of the produced text using Natural Language Processing. It also involves the comparison of the preexisiting modern Speech-Recognition models and map the differences as to which one is the best fit.

DECLARATION BY CANDIDATES

We hereby declare that the work presented in this end semester project report of B.Tech (IT) 7th Semester entitled **"Speech to text conversion and Summarization"**, submitted by us at Indian Institute of Information Technology, Allahabad, is an authenticated record of our original work carried out in September, 2020 under the guidance of **Dr. K. P. Singh**.

Due acknowledgements have been made in the text to all other material used. The project was done in full compliance with the requirements and constraints of the prescribed curriculum.

Place: Allahabad
Date: September 16, 2020

Certified that the above statement made by the students is correct to the best of my knowledge.

Signature : ________________
Dr. K. P. Singh

1 INTRODUCTION

One of the key applications of automatic speech recognition is to transcribe speech documents such as talks, presentations, lectures and broadcast news. Although speech is the most natural and effective method of communication between human beings, it is not easy to quickly review, retrieve and reuse speech documents if they are simply recorded as audio signals. Transcribing speech is therefore expected to become a key capability for the forthcoming IT age. The original speech source is a signal, and a signal is processed to translate all the information contained in the signal into the text format. The extraction of the function is the process of taking a signal with some logic and translating it to the appropriate format. Text summarization is one of the most relevant techniques used in documentation. Long papers are hard to read and understand, since they take up a lot of time. Text synthesis addresses this problem by supplying semantics with a shorter description of it.

In our work a combination of speech to text conversion and text summarisation is to be implemented. The first and foremost step to work with NLP (Natural Language Processing) is to extract the features from the speech which have some values. If a word or a sentence is recognized as meaningless, then it becomes an obstacle to the summarization process and should be treated as obscelete. Even the punctuation plays a vital role in summarization as the semantics are important in the process.

2 MOTIVATION

The field of speech is still open for research as 100% effective system is yet to be developed. The performance of the system can be measured in terms of Accuracy, Memory and Speed. Accompanied with summarization, this work has a lot of potential as a lot of people for eg. keep up with the world affairs by listening to news bites or base investment decisions on stock market updates etc. would make making decision easier and faster.

Thus, the motivation here is to build a system which would be efficient and bridge the gap between a speech and its synopsis.

3 PROBLEM DEFINITION

In this paper, we are going to convert speech to a text document and summarize the document. The speech can either be a Voice Recording or an input from the microphone. The result should show us the speech in textual format and another text file with the summarized content.

4 LITERATURE REVIEW

1.In paper[5] published by B. H. Juang and L.R. Rabiner in 80s, it deals with use of hidden Markov models for speech recognition. HMMs predict the words according to sound pattern by taking words which have higher probability. The paper shows that this method works well in presence of large amount of training data also the recognition accuracy is high if the word under consideration is from training data set.

2. Back in 80s, a paper was published by Bruce T. Lowerre called Carnegie Mellon's "Harpy' speech system[6] which was able to understand 1000 words.

3. In 2011, Apple launches 4s with "voice-driven assistant" feature, siri[7]. It is AI driven system which exists on remote servers by which people can access it. It is more than just NLP(Natural Language Processing), it is NLU(Natural Language Understanding) which can answers users' queries, set alarm, etc. It learns from its users and data.

4. Word error rate shows the accuracy of the system, IBM has 6.9 percent, microsoft has 5.9 percent and google has lowest 4.9 percent, recorded in Oct,2018.

5. Speech Summarization by Yousuke Shinnaka, and Chiori Hori[8] deals with summarization of text using two processes: 1.Sentence extraction by allotting scores to the sentences(Linguistic score, Significance Score, Confidence Score), 2. Sentence compaction. Low summarizing accuracy.

6. PEGASUS: A State-of-the-Art Model for Abstractive Text Summarization, June, 2020 by Peter J. Liu and Yao Zhao[9] demonstrates the working of transformer encode-decoder models[10], a recurrent neural network. It showed better performance for large dataset than its predecessors.

5 PROPOSED METHODOLOGY

Our work is mainly divided into two main parts. The first part being converting the input speech to text and the second one to summarize the text efficiently such that the actual meaning is retained as much as possible.

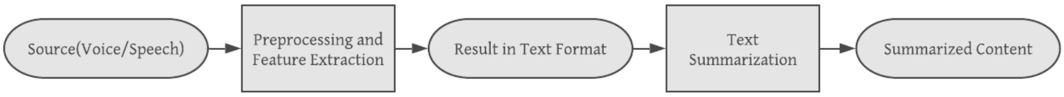

Figure 5.1: Speech recognition and text summarization process flow

There can be two types of input from the user, one can be a recorded voice clip while other can be use of the microphone i.e. a user can speak live and see his/her speech converted to text in real-time. Our work can easily shift itself according to the mode.

5.1 Speech to Text Conversion

The first part of our work is to convert the speech to text. Firstly, analog signal is converted to digital. Then these signal is distributed in phonemes. To recognize each phoneme correctly, we need to build a model using neural network, that uses large amount of data to train itself and can be used on a sample. This model is combined with language model and re-scoring algorithm to tackle linguistic problem like this sentence, “He read a book last night” but pronunciation of it would be “he red a book last night”. Here, we use the preexisting Google API. Python library SpeechRecognition uses it along with *Microsoft Bing*, *IBM Speech* to text to name a few. Out of these, Google Cloud Speech API seems more reliable when tested on multiple samples.
However, we need to take care of period(.), to add in the result. There is no denying of the fact that the speech might have a noise which is to be filtered. For this purpose, python provides inbuilt libraries which can be used to amplify the speech and succumb the noise, to further increase the efficiency of our model.

5.2 Text Summarization

There are two types of summarization when output is concerned: Extractive and Abstractive. We want our system to perform extractive text summarising and after that we also allow queries from users. For extraction, we first split the text in sentences. we allot each sentence a score. We also keep count of the words and score is update according to it. We also calculate probability of each sentence and store it in a matrix.
For query purpose, we represent words as graphs and matrix. We search through all the matrix in our search index to find the most relevant, useful results for users.
Below is the flow graph that shows the step by step working of the algorithm.

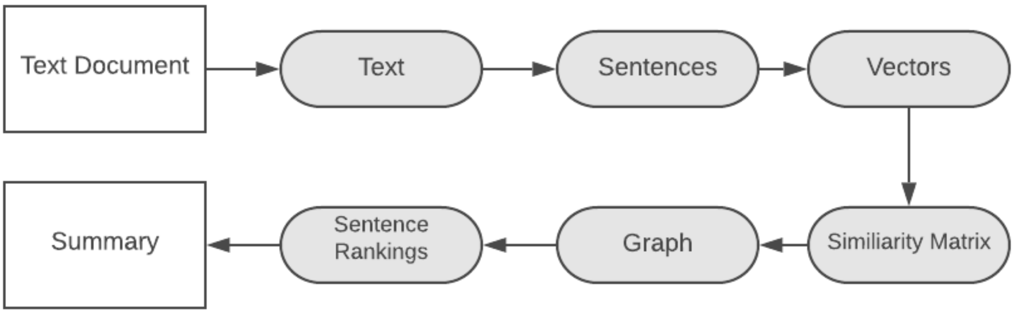

Figure 5.2: Text Summarization Algorithm process flow

6 SOFTWARE REQUIREMENTS

- **Python3**
- **pyAudio** (version 0.2.11) – pip3 install pyaudio
- **SpeechRecognition** (version 3.8.1) – pip3 install speechrecognition
- **Python NLTK libraries**
- **Google Application Programming Interface(API)**

7 CONCLUSION

Speech recognition and text summarization are two vast areas to be explored. We expect our proposed methodology would reduce the time and effort of manual documentation of lengthy speeches in an event. Even for the verification of the summarized content, with the aid of text for speech translation, the device can be automated to read out the summarized content.

One thing that can be done to improve the efficiency and accuracy of our work is by splitting words based on the pauses in the speech and speech pattern. This can be done by extensively training various speech samples on high performance machines.

This model, when complete, can be used wherever there is a requirement of summarising lengthy lectures into precise documents as the automated system will convert the speech to text and also summarise the content. It can be of great help for students to archive lecture notes from classes, conferences or seminars. Also, summarization of text can be used as tagging for recommendation, for fighting abuse, autocomplete and even extracting detail from the text.

8 REFERENCES

[1] **Sadaoki Furui, Tomonori Kikuch**, *"Speech-to-Speech and Speech-to-Text Summarization "* : Feb, 2003 [Accessed: Sep, 2020]

[2] **Vinnarasu A., Deepa V. Jose**, *"Speech to text conversion and summarization for effective understanding and documentation"*: Oct, 2019 [Accessed: Sep, 2020]

[3] **Dhilip Subramanian**, [Online] Available: https://towardsdatascience.com/easy-speech-to-text-with-python-3df0d973b426 [Accessed: Sep 2020]

[4] **Michal Rott**, *"Speech-to-Text Summarization Using Automatic Phrase Extraction from Recognized Text"* : Sep, 2016 [Accessed: Sep, 2020]

[5] **B. H. Juang and L. R. Rabiner**, *"Hidden Markov Models for Speech Recognition"* : Mar, 2012 [Accessed: Sep, 2020]

[6] **Bruce T. Lowerre**, *"Harpy Speech recognition system, 1976 "* [Online] Available : https://stacks.stanford.edu/file/druid:rq916rn6924/rq916rn6924.pdf [Accessed: Sep, 2020]

[7] **The Guardian**, *"Voice recognition: has it come of age?, 2011"*

[Online] Available : https://www.theguardian.com/technology/2011/nov/20/voice-recognition-apple-siri [Accessed: Sep, 2020]

[8] **Yousuke Shinnaka, and Chiori Hori, Member, IEEE**, *"Speech Summarization of Spontaneous Speech "*
[Online] Available :https://www.csie.ntu.edu.tw/ b97053/paper/SpeechToText.pdf[Accessed: Sep, 2020]

[9] **Peter J. Liu and Yao Zhao, Software Engineers, Google Research**, *"PEGASUS: A State-of-the-Art Model for Abstractive Text Summarization Tuesday, June 9, 2020"*
[Online] Available : https://ai.googleblog.com/2020/06/pegasus-state-of-art-model-for.html [Accessed: Sep, 2020]

[10] **Jakob Uszkoreit**, *"Transformer: A Novel Neural Network Architecture for Language Understanding, 2017"*
[Online] Available :https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html [Accessed: Sep, 2020]

[10] **Alex Graves, Abdel-rahman Mohamed and Geoffrey Hinton**, *"Speech recognition with deep recurrent neural networks"*
[Online] Available :https://ieeexplore.ieee.org [Accessed: Sep, 2020]