International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 10 Issue 10, October-2021, Impact Factor: 7.957

Automatic Speech to Text Summarization Using Graph Based Algorithm

Dr. P.V.S.L Jagadamba[1], Ms. Garapati Sai Harshini[2], Ms. Kalla Gowri[2],
Ms. Kalla Gowri[2], Ms. Gandi Yernamma[2]

[1]Professor, Department of Computer Science and Engineering, Gayatri Vidya Parishad College of Engineering for Women, Visakhapatnam, India
[2]Pursing B.Tech 4th Year, Departmetn of Computer Science and Engineering, Gayatri Vidya Parishad College of Engineering for Women, Visakhapatnam, India

ABSTRACT

Automatic Speech to Text and Text Summarization Using Graph based Algorithms can be used in meetings, to get the brief description of the meeting for future reference. This provides signature verification using Siamese neural network to authenticate identity of the user and convert the user provided audio file which is in English into English text using the speech recognition package provided in python. Sometimes only the summarization of the meeting is required, the solution for this text summarization. So, the transcript is then summarized using the natural language processing methodologies such as unsupervised extractive text summarization algorithms.

Keywords—*Siamese Neural Network, English Speech to English Text, Natural Language Processing, Unsupervised extractive text Summarization.*

I. INTRODUCTION

Signature Verification authenticates the identity of individuals by measuring their signatures. The signature is treated as a series of movements that contain unique biometric data. The growing need for a full proof signature verification scheme ensure that the proposed scheme can provide comparable and if possible better performance than already established signature verification schemes.

Speech is the most important part of communication between human beings. Speech recognition is the process of making a machine recognize the speech of different people based on certain words or phrases it can be converted into text. Text Summarization solves this problem by providing a shortened summary of it with consumes lot of time. In the proposed work a combination of speech to text conversion and text summarization is implemented.

II. LITERATURE SURVEY

In paper performing Signature verification using Siamese Neural network [1], they presented an algorithm based on an artificial neural network, called “Siamese” time delay neural network, consists of two identical networks joined at their output. During training the network learns to measure the similarity between pairs of signatures. When used for verification, only one half of the Siamese network is evaluated. The output of this half network is the feature vector for the input signature. Verification consists of comparing this feature vector with a stored feature vector for the signer. Signature closer than a chosen threshold to this stored representation are accepted, all other signatures are rejected as forgeries.

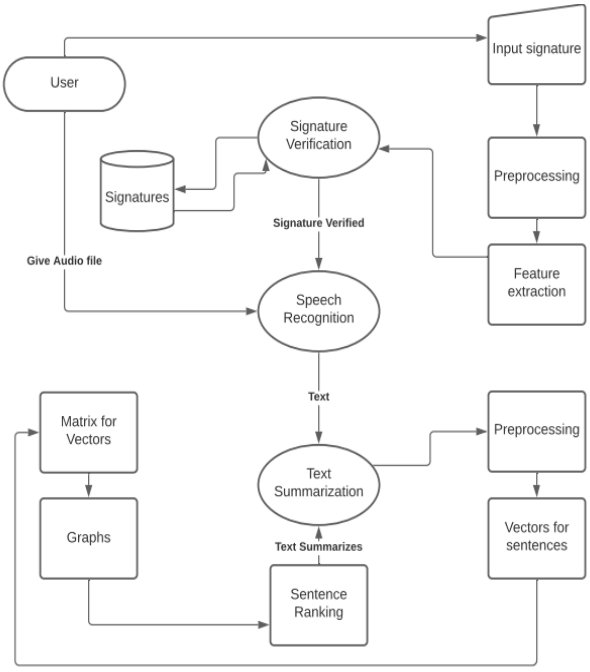
In paper Evaluating Google Speech-to-Text API’s Performance for Romanian e-Learning Resources[2] presents the history of ASR systems together with main approaches used by the algorithms behind these systems.

In a paper Automatic Text Summarization using text rank algorithm[3], had taken the data from various sources for a particular topic and summarize it for the convenience of the people, so that they don’t have to go through so multiple sites for relevant data.

The paper a Survey on Extractive Text Summarization explains about the different types of summarization methods and describes briefly about the extractive text summarization techniques.

III. PROPOSED SYSTEM

The orientation provided by our work can be very competent in finding the authentication through digital signature verification, Audio files for speech to text conversion and then text summarization. This leads to improve authentication and summary making. Already, the users signature dataset will be stored, and user audio will also be recorded. The work is divided into three modules Signature verification, Speech to text conversion, Text summarization.



Content Diagram

IV. SIAMESE NEURAL NETWOEK

Data

The initial data collection was made by all our team members and each genuine signature is scanned and placed in an input1 folder with 250 total signatures. Later when producing genuine or forgery signatures , the signer was shown an example of the genuine signature and is placed in a folder named input2. Finally, the information about the signature whether it is a genuine or forgery is stored in a csv file with 0 or 1 representation for forgery or genuine kind signatures.

Pre-Processing

The provided 500 signatures are initially converted into black and white image. To make sure that each and every image is of the same size , resizing the signatures is required to 200 x 500 dimensions. Finally, after converting the images into same size , normalization is performed to make the values range between specific values. Now the data is ready to split into train, test, validation set. We have considered 180 images to be in the training set and 30 signatures for validation and 40 signatures for test set.

Architecture

The module needs to learn to discriminate forgeries from genuine signatures, led us to consider the use of artificial neural networks techniques. Neural networks can learn from examples and they can also be used to compress data. This work is based on a neural network architecture, called a Siamese network, which first extracts features from two signatures and then evaluates the distance between these two sets of features.

The network has two input fields to compare the two patterns and one output whose state value corresponds to a measure of the distance between the two patterns. Two separate subnetworks based on Time Delay Neural Network(TDNN) act, one on each input pattern, to extract features, then the cosine of the angle between the two feature vectors is calculated and this represents the distance value. The cosine distance measure was rather than the Euclidean distance. Using Euclidean distance and requiring two genuine signatures to have a small distance could lead to the trivial solution of zero-sized feature vectors.

The subnetworks are constrained to be identical ; they are multi-layered and feed-forward with several layers of feature extraction before distance measure. TDNN’s use local processing units having inputs that are delayed in time. In other

words, a unit receives inputs from just a portion of the signature trajectory. In a multi-layer TDNN, successive layers of units can extract features spread over a wide range of time.

To train the network , two patterns are applied, one to each input, and a desired value for the output is used to backpropagate the error. The desired output is for a small angle between the outputs of the two subnetworks (f1 and f2) when two signatures are presented, and a large angle if one of the signatures is forgery. For the cosine distance used here:

$$(f1.f2)/(|f1||f2|),$$

The desired output is in between 0 and 1 and the threshold value decides whether the respective signature is a genuine one or not.

Training the Neural Networks

During training, part of the data set is used for learning the weights of the units, another part, called the validation set, is used after a training iteration to test the test the network's performance. A measure of the performance of the network at accepting genuine pairs and rejecting forgeries is found by calculating the percentage of genuine signature pairs for which the output was greater than threshold value.

V. SPEECH TO TEXT CONVERSION

Speech recognition is the ability of a computer software to identify words and phrases in spoken language and convert them to human readable text. We do not need to build any machine learning model from scratch, this library provides with convenient wrappers for various well known public speech recognition APIs. We have used Google Speech Recognition in this application, as it's straightforward and doesn't require any API key.

Make sure the audio file in the .wav format and contains English speech. Then, load the audio file, and create a speech recognition object and then split the audio file into chunks and apply speech recognition on each of these chunks. Finally, returns the text for all chunks detected. We can also change the parameters suitable for your audio file.

VI. TEXT SUMMARIZATION

Reading the Data

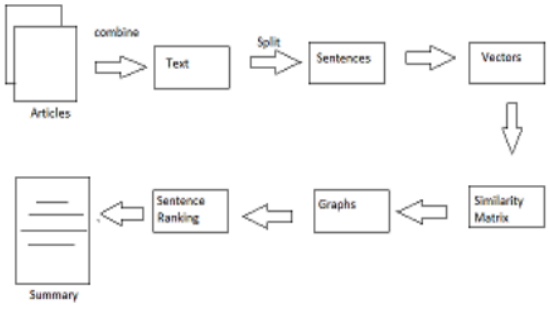
We read the data from output generated by the Google speech recognition API and now we split the text into individual sentences.

Pre Processing

One of the major forms of preprocessing is to filter out useless data. In natural language processing, useless words, are referred to as stop words. We apply preprocessing on the text which enables us to remove all the extra data from it which are of no use example: "the", "a" ,"an", "in" etc.

Word Embeddings, Similarity Matrix and Graphs

Word embeddings are vector representation of words. These word embeddings will be used to create vectors for our sentences. The Words from vocabulary is mapped to vectors or numbers. Word embeddings are in fact a class of techniques where individual words are represented as real-valued vectors in a predefined vector space. This is achieved by mapping words into a meaningful space where the distance between words is related to semantic similarity. Based on vector of sentences we prepare a similarity matrix, i.e., find similarities between the sentences, and we will use the cosine similarity approach . Let's first define a zero matrix and populate it with cosine similarities of the sentences . Cosine similarity to compute the similarity between a pair of sentences and initialize the matrix with cosine similarity scores. Now convert the matrix into graph. The nodes of this graph will represent the sentences and the edges will represent the similarity scores between the sentences.



Text Summarization process

Page Rank Algorithm

Page rank is used primarily for web pages, we would have to compute a score called page rank score. The undirected graph is then passed it's time to extract the top N sentences based on their rankings for summary generation. For our, based on similarity we take out top 2 or 3 sentences and summarize them. The basic idea of Text Rank is to provide a score for each sentence in a text, then you can take the top-n sentences and sort them as they appear in the text to build an automatic summary.

CONCLUSION

The web application is designed with sign in and signs up, Initially the user authentication with the signature verification using Siamese neural network model is the one module of the project. Once the user is verified means, the audio file is chosen from the files and then, on clicking submit button , audio file goes to speech recognition process, where python code is used and then the audio file is converted to text file. The text file goes through page rank algorithm and summarized text is produced. Finally, The converted text and summarized text are produced as the output of the project.

The real idea of the project is to prepare summary of a meetings, using meeting recorded audio files. This will save time and also authentication using signature verification helps in not accessing or entering of any other persons into meeting. Text summarization refers to the technique of shortening long pieces of text.

The proposed method is useful to preparing notices and summary and also the safety of user's data using authentication. In the future, The accuracy of the models used in this project can be increased, For speech recognition new functions and new models can be developed, For authentication biometric models can be used, Can add recording option which will automatically recorded and covert to text. This project can be used in many areas such as Meetings, Summarizing large data, Making notes, Conferences, Hospitals. This project intention is to create a coherent and fluent summary having only the main points outlined in the document.

REFERENCES

[1]. Jane Bromley, James W. Bentz, Leon Bottou, Isabelle Guyon, Yann Lecun, Cliff Moore, Eduard Sackinger and Roopak Shah "Signature Verification Using A "Siamese" Time Delay Neural Network" By Carnegie Mellon University on 10/14/14, AT&T Bell Laboratories, Crawfords Corner Road Holmdel, NJ 07733, USA and NCR Corporation, E&M Atlanta 2651 Satellite Boulevard, Duluth, GA 30135, USA.

[2]. Iancu, Bogdan. (2019). Evaluating Google Speech-to-Text API's Performance for Romanian e-Learning Resources. Informatica Economica. 23. 17-25. 10.12948/issn14531305/23.1.2019.02.

[3]. Tanwi, Satanik Ghosh, Viplav Kumar, Yashika S Jain, Mr. Avinash B "Automatic Text Summarization using Text Rank" International Research Journal of Engineering and Technology (IRJET) Volume: 06 Issue: 04 | Apr 2019

[4]. N.Moratanch ,S.Chitrakala "A Survey on Extractive Text Summarization" IEEE International Conference on Computer, Communication, and Signal Processing (ICCCSP-2017).