Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

Techniques for Meeting Summarization: An Analysis and Suggestions for Improvement

Viveksheel Yadav[1], Faraz Ahmad[2], Ashuvendra Singh[3]

[1, 2]Assistant Professor, Mechanical Engineering, School of Engineering, DevBhoomiUttarakhand University, Chakrata Road, Manduwala, Naugaon, Uttarakhand 248007

[3]Assistant Professor, Civil Engineering, School of Engineering, DevBhoomiUttarakhand University, Chakrata Road, Manduwala, Naugaon, Uttarakhand 248007

[1]me.viveksheel@dbuu.ac.in, [2]me.faraz@dbuu.ac.in, [3]ce.ashuvendra@dbuu.ac.in

Article Info

Page Number: 1634 - 1645

Publication Issue:

Vol 71 No. 4 (2022)

Abstract

The purpose of this research is to investigate the several approaches to summarizing meeting minutes. In addition to this, it proposes a hybridized strategy for the summarizing of meetings, which brings together abstractive and extractive methods. Audio extraction and text detection from the screen are the two methods that may be used in order to successfully extract text from meetings. After the necessary texts have been collected, a meeting recap is generated from them. Abstractive text summarizing, on the other hand, is more concerned with presenting a logical overview of the material that has been presented, as opposed to extractive text summarization, which consists of stringing together the most important lines from various paragraphs. The bulk of recent research on abstract text summarization has been built on recurrent neural networks; nevertheless, RNN-based algorithms do not perform well when dealing with lengthy sequences. [Citation needed] [Citation needed] We recommend combining these two approaches to the summarization of meetings in a sequential fashion in order to provide a more effective overall summary of the meeting. Our hybridized text summarizer model is going to be applied to the conference video and audio once it has been converted into text documents as part of the recommended research endeavor. The final product would be a condensed text document that the different parties involved, whether or not they were physically present at the conference, would be able to use as a quick reference that gets to the point.

Article History

Article Received*: 25 March 2022*

Revised*: 30 April 2022*

Accepted*: 15 June 2022*

Publication*: 19 August 2022*

Keywords:Text summarization, deep learning, extractive technique,

abstractive technique, object detection, abstractive technique, artificial intelligence, machine learning, and artificial intelligence

I. INTRODUCTION

The procedure of frame extraction is used first, and this results in the production of the text. After that, the text is recovered from the frames by using procedures that are connected to natural language processing. In addition, the audio from the meeting is taken as input, and a speech recognition model is used so that it may be converted to text. After that, the text that was produced from these sources is sent to the summarizer so that it may summarize the material.

The objective of extractive summarizing is to provide a summary of the material by arranging the phrases in a certain order. The efficacy of this strategy is very reliant on the caliber of the phrase characteristics. On the other hand, the majority of the earlier algorithms required the characteristics that were used to represent sentences to be built by hand. This was a requirement for the bulk of the older algorithms. Combining methods of deep learning that were formerly seen as distinct has become an increasingly popular practice in recent years. Techniques for embedding previously learned words have been successfully completed. Exceptional performance in a vast array of natural language processing (NLP) endeavors.

Extractive Text Summarization with BERT makes use of a pipeline that tokenizes paragraphs into sentences for the BERT (Bidirectional Encoder Representations from Transformers) model and clusters the embedding with K-means, selecting sentences that are closer to the centroid. BERT is an acronym that stands for "Bidirectional Encoder Representations from Transformers." This makes it possible to get more precise findings. The "hugging face" organization makes available the PyTorch-pre-trained-BERT library, which is used by the BERT for Text Embedding system.

We present an approach that we name TFRSP (Text Frequency Ranking Sentence Prediction). This technique combines extractive and abstractive summarization, and it makes use of both supervised and unsupervised learning algorithms. In the procedure that is known as extractive summarization, the algorithm that is known as Term Frequency-Inverse Document Frequency (TF-IDF) is paired with the method that is known as Text Rank (TR). In abstractive summarization, we make use of a model is known as the sequence-to-sequence (seq2seq) model. This is a supervised learning algorithm, which means that it uses training datasets and

testing datasets to learn from data.

II. LITERARY REVIEW

Within our application, we want to integrate three primary parts. The pulling out of frames, the recognizing of voice, and the summarizing of text. We are able to get frames from a video by using the movies package that Python provides. Even the number of frames per second that can be extracted is up to us to choose.

Recording speech and converting it into text may be accomplished in a number of different ways. There are many different methods that may be used, some of which include deep learning, support vector machines, minimum distance classifiers, RASTA, PCA, ICA, ZCR, and the Kalman filter. Each strategy has pros and cons. Because there is a lot of background noise in real-time audio and the speech recognition model still has to be correct despite this noise, deep learning was the best choice that was available. This article suggests leveraging Google's voice recognition for speech recording by way of a Python library called Speech Recognition and other audio libraries like PyAudio and PortAudio. With an accuracy of 79%, it is the voice recognition program with the second highest level of precision.

We have come up with a few different approaches of summarizing the material. The algorithms BERT, TSRFP, Sentence Ranking, and KNN were among those that were uncovered and evaluated by our team. The sentence ranking algorithm produced summaries that were easier for the computer to process but did not seem to be as similar to summaries that were authored by humans. Both the BERT and the KNN were computationally intensive methods, but they generated more readable and understandable summaries. We made an effort to find solutions to the issues by using the TSRFP model. Extractive summarization is then used as an input for abstractive text summarizing, which therefore results in the production of the summary. This resulted in a summary that needed fewer computer resources to produce while still delivering something that is comparable to a summary written by a person.

III. METHODS

A. Recognizing Someone's Voice

Using techniques of deep learning

The use of deep learning as a method for voice recognition is becoming more common. Voice recognition software like that developed by Google makes use of it. Artificial neural networks

are used in the prediction of speech using deep learning. [1] In order for this method to be successful, significant training data are required. At the outset, the audio data have to be segmented into two groups: the practice data (70%) and the testing data (30%). The audio input is converted into a vector by the use of the Mel-frequency cepstral coefficient, and then speech is recognized through the application of deep learning. This model [1] has an accuracy of 66.22 percent. Using a Filter Called the Bidirectional Kalman

For accurate results, speech recognition needs audio that is crystal clear. Due to the fact that the audio may include noises such as background noise, we cannot foresee being able to deliver clear audio as an input in real-time scenarios. As a consequence of these disruptions, the findings will be incorrect. The Kalman Filter is one method that may be implemented to get rid of these disruptions. The use of Kalman filtering is a fantastic strategy for lowering the amount of non-stationary background noise. In the beginning, the database is split up into a training database and a testing database. MFCC is used here for the purpose of feature extraction. The testing is carried out [2] in the presence of disturbances, and an accuracy of 90% is achieved throughout the process.

ii. Utilizing Minimum Distance Classifier and Support Vector Machine

The language English is used for the vast bulk of voice recognition work. People who are illiterate or do not understand English may be able to utilize these technologies if voice recognition can be carried out in the native languages of those countries. [3] The objective of the system that is being suggested is to create and put into operation a Speech-To-Text conversion system that is capable of handling Marathi, English, and a combined Marathi-English language. [3] The MFCC feature extraction method, the Minimum Distance Classifier, and the Support Vector Machine (SVM) approaches are used in the system that has been presented for the purpose of speech classification. In the beginning, the audio database is split up into several databases for training and testing. During the training phase, features are extracted from a training database using a training database as a reference, and a sample reference feature vector is constructed. During the testing phase, features are retrieved with the use of MFCC, a reference vector is produced, and the words that have the greatest similarity are output. It has been discovered that [3] has a precision of 93.625%.

iii. Analyzing and contrasting the various speech recognition methods

Machine learning, deep learning, and statistical approaches are all types of methods that may be used in speech recognition. Every one of these approaches comes with its own set of advantages

and disadvantages. Methods of feature extraction such as PCA (Principal Component Analysis), ICA (Independent Component Analysis), and ZCR (Zero Crossing Rate) give excellent results for voice recognition systems with a smaller vocabulary but are susceptible to noise. RASTA or RASTA- PLP may give strong performance for noisy data. [4] Both the KNN and Naive Bayes classifiers are easy to use, but their effectiveness is limited to vocabulary datasets that are quite small. SVM achieves its best results with data sets of medium size, although the training process takes much longer with big datasets [3]. A neural network [1] requires a significant amount of time for training when working with huge datasets.

Recognition of Handwritten Text, Option B

i. Using artificial intelligence

Handwritten identification is one of the jobs that present the greatest challenge in pattern recognition. It is used in a number of different contexts, including a bank for the processing of checks, the postal service for the identification of postal codes, and a court of law. At the outset, a test photograph is taken with the text being black and the backdrop being white. The density of each text's pixels is retrieved using the extraction procedure, and a genetic algorithm is used to analyze the properties of the handwritten text that was taken from it. It is anticipated that the accuracy of the system is somewhere around 90%.

ii. Using deep neural networks((DNN)

The identification of handwriting has seen improvements in both its speed and its accuracy as a result of the development of deep neural networks. People may now depend more on functions and algorithms that have been learnt from data as opposed to those that have been hand-crafted. [6] All of the previously proposed methods depend on a separate system in which the picture is first segmented, the angle, curve, and height of the extracted image are predetermined, and there are constraints on the orientation and size of the text. However, in fact, people's handwriting is distinct from one another, and it may sometimes even be impossible to read; as a result, these strategies are either incorrect or ineffectual. A neural network, which is taught from data and can rapidly adapt to a new dataset, performs all of these operations implicitly and without any limitations. Neural networks are taught. Figure 2 depicts the construction in its entirety. This method [6] is trained at 145 DPI, and it outperforms every other model with an error rate of 7.6% when compared to other cloud-based API, which has an error rate of 14.6%.

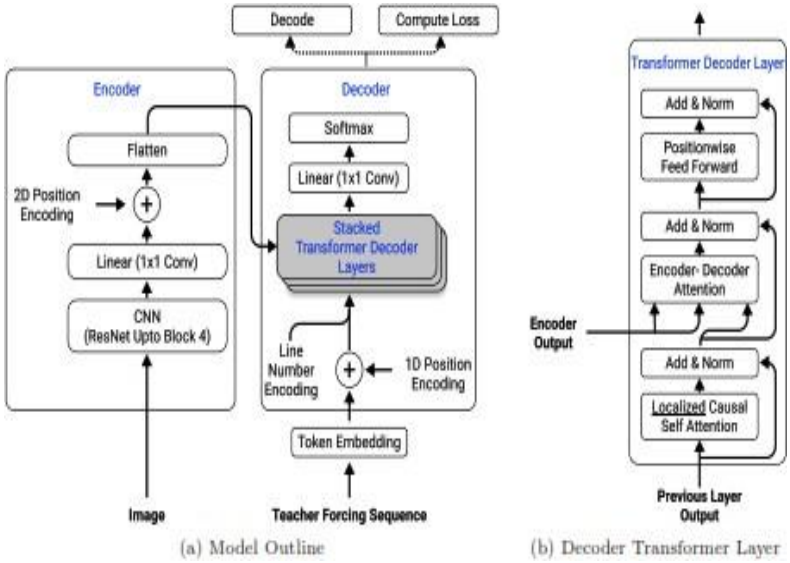

Fig 2 DNN Model Schematics

In Fig. 2, Model Schematics. Left: CNN Encoder and Transformer Decoder. On the right is a representation of the Transformer Layer, which may or may not include a Localized Self Attention. In the context of training or prediction, "Teacher Forcing Sequence" refers to the ground truth being pushed to the right.

A. Synopsis of the Whole Text

i. Extraction of Textual Information by the Utilization of Sentence Ranking

The idea that supports extractive text summarization is founded on the premise that words that appear more often are more significant. This is done on the basis of the frequency of occurrence of each word. The process begins by constructing a frequency matrix, which is a list that specifies how often each word appears, and then it gets rid of the English phrases that are used the most frequently. When determining the significant context of a phrase, it is necessary to take into account the grade given to each individual word in the phrase. The summarizer will pull out the sentence fragments [5] that have the greatest weighted frequency in order to offer a summary of the information.

ii. Abstraction of Text by the Use of the BERT Model

Extractive text summarization utilizing BERT and K-Means makes use of a pipeline that tokenizes paragraphs into sentences for the BERT model and clusters the embedding utilizing

K-Means, selecting sentences that are closer to the centroid in order to finish the process. This is accomplished by selecting sentences that are more central to the text. This implementation of the BERT for Text Embedding technique makes use of the PyTorch-pre-trained-BERT module. K-Means and Gaussian Mixture Models are helpful tools that may be used in the process of clustering data. A very small number of phrases needed to be generated by the model in order for it to be effective in capturing the context of the whole presentation. The model [20] was including further sentences in addition to those that were requested.

iii. K The Closest of Your Neighbors (KNN)

A modification is made to the KNN model so that it takes into consideration the degree to which feature values and features are similar to one another. The K words that are most similar to one another are chosen to be the K closest neighbours, and voting on the labels of the K nearest neighbours is used to determine the label for the beginner entity. There are K words. The analysis of the text has resulted in the use of a binary classification system. The input consists of a continuous string of text that has been segmented into paragraphs [9] by using carriage returns.

iv. TFRSP

The TFRSP method utilizes both supervised and unsupervised learning algorithms in order to integrate extractive and abstractive summarization approaches. During the process of extractive summarization, the Text Rank (TR) method is used with the Term Frequency-Inverse Document Frequency (TF-IDF) algorithm. Input for the abstractive method comes from the result of the extractive text summarization. The abstractive summarization method creates the summary by using a supervised learning technique known as the sequence to sequence (seq2seq) model. This model requires both training and testing datasets. The summary results in a 38.42% rise in the ROUGE score of the procedures that are currently being used.

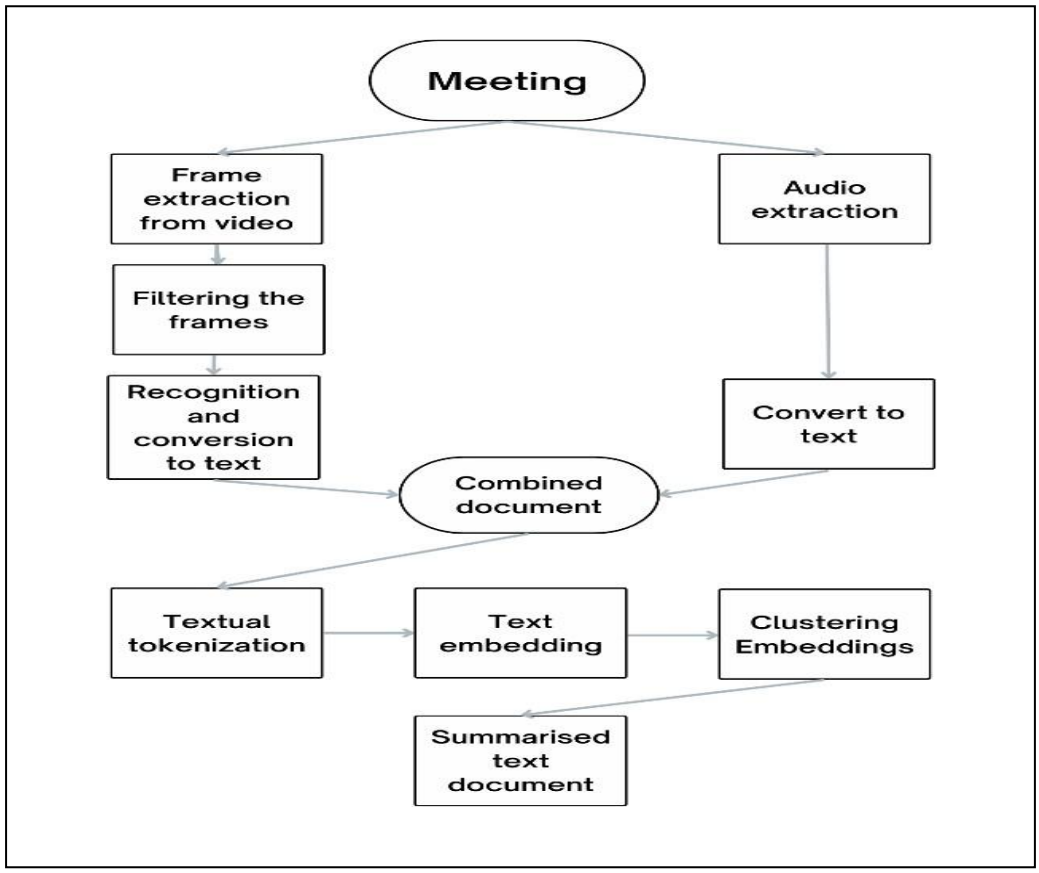

Fig. 3. Architectural diagram

V. PROPOSED SYSTEM

We propose to summarize a meeting in the format depicted in Figure 3, by first collecting the data from the presenter's voice and the presentation, and then translating it into text with the use of speech recognition and handwriting text recognition, respectively. After the data has been gathered, it is sent into an extractive text summarizer, which then tokenizes the data and filters out the phrases that are most relevant to the topic. After that, this input is sent to the abstractive text summarization model so that it may provide a summary that is far briefer and more accurate.

There is more than one model of implementation available for each component of the proposed task. It will not be easy to choose the perfect model and then make more improvements to it. Every person's interpretation of the meeting's takeaways will be quite different. The models that were implemented are separate from one another. itself but have not yet been integrated into a single, comprehensive application.

TABLE 1: Summary of the models

Models	Advantages	Disadvantages
Sentence Ranking	Extremely fast and doesn't demand many computational resources.	The summary that is generated doesn't sound human.
Abstractive text summarization using the BERT model	Generates a very humanly sounding summary around the main keywords of thedocument	Takes a lot of time and is computationally taxing.
KNN Model	As it uses,the relation between the keywords to generate a summary. The summary is accurate.	Since there is a lot of processing it is computationally moretaxing.
TFRSP	This model uses both extractive and abstractive techniques and can hence reduce both the computational power and accuracy of themodel.	This takes different models and combines them to generate a summary so this might more time when compared to othermodels.

In addition, to the best of our knowledge, no other programs have succeeded in producing a real-time summary of a conference. As a result of this, we want to create a web application that can be used for lecture summaries and can be installed as a browser extension. By carrying out this study, we expect to be able to reduce a meeting that lasts an hour into a single document that can be easily referred to and gives a summary that is clear and succinct. This application has potential applications in both the business world and the academic world.

Our research of a wide variety of libraries, including the hugging face transformers, MoviePy, Speech Recognition, PyAudio, and PortAudio, gave us an in-depth knowledge of the libraries

that are currently on the market.

It is conceivable to conduct research in the proposed path since there is not a single integrated application currently available on the market for effectively summarizing the meeting. The models that would be utilized in the proposed study are summarized in Table 1, which may be found here.

CONCLUSION

This article offers a full and in-depth description of the numerous different approaches that may be used to summarize meetings. A system that would combine extractive and abstractive text summarization techniques is proposed. This system would produce a better overall summary that would benefit different stakeholders who might not be interested in listening to the entire meeting recording but instead plan to focus on key points, or the essence of the meeting. In addition to examining the literature, this system would produce a better overall summary that would benefit different stakeholders.

REFERENCES

[1] Kongthon, Alisa; Sangkeettrakarn,Chatchawal; Kongyoung, Sarawoot; Haruechaiyasak,Choochart (October 27–30, 2009). Implementing an online help desk system based on conversational agents. MEDES '09: The International Conference on Management of Emergent Digital EcoSystems. France: ACM.doi:10.1145/1643823.1643908.

[2] Mitchell, Tom (1997). Machine Learning. New York: McGraw Hill. ISBN 0-07-042807-7. OCLC36417892.

[3] A.P.Singh,R.NathandS.Kumar,"ASurvey: Speech Recognition Approaches and Techniques," 2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), 2018, pp. 1-4, doi:10.1109/UPCON.2018.8596954.

[4] N. Chumuang and M. Ketcham, "Model for Handwritten Recognition Based on Artificial Intelligence," 2018 International Joint Symposium on Artificial Intelligence andNatural Language Processing (iSAI-NLP), 2018, pp. 1-5, doi:10.1109/iSAI-NLP.2018.8692958.

[5] J. N. Madhuri and R. Ganesh Kumar, "Extractive Text Summarization Using Sentence

Ranking," 2019 International Conference onData Science and Communication (IconDSC), 2019, pp. 1-3, doi:10.1109/IconDSC.2019.8817040.

[6] S. S. Desai, D. Rajput and K. Patil, "An approach for Text Recognition from Document Images,"2020IEEEBangaloreHumanitarianTechnology Conference (B-HTC), 2020, pp. 1-5, doi: 10.1109/B-HTC50970.2020.9297939.

[7] Du, Y. et al. (2020) "PP-OCR: A practical ultra lightweight OCR system," arXiv [cs.CV]. doi:10.48550/ARXIV.2009.09941.

[8] Ghadage, Y. H. and Shelke, S. D. (2016) "Speech to text conversion for multilingual languages," in 2016 International Conference on Communication and Signal Processing (ICCSP). IEEE, pp.0236–0240.

[9] Jo, T. (2017) "K nearest neighbor for text summarization using feature similarity," in 2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE). IEEE, pp.1–5.

[10] Jolad, B. and Khanai, R. (2019) "An Art of Speech Recognition: A Review," in 2019 2nd International Conference on Signal Processing and Communication (ICSPC). IEEE, pp.31–35.

[11] Lakkhanawannakun, P. and Noyunsan, C. (2019) "Speech Recognition using Deep Learning," in 2019 34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC). IEEE, pp.1–4.

[12] Ozsoy, M. G., Alpaslan, F. N. and Cicekli,I. (2011) "Text summarization using Latent Semantic Analysis," Journal of information science, 37(4), pp. 405–417. doi: 10.1177/0165551511408848.

[13] Raundale, P. and Shekhar, H. (2021) "Analytical study of Text Summarization Techniques," in 2021 Asian Conference on Innovation in Technology (ASIANCON). IEEE, pp.1–4.

[14] Sharma, N. and Sardana, S. (2016) "A real time speech to text conversion system using bidirectional Kalmanfilter in Matlab," in 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE, pp.2353–2357.

[15] Singh, S. S. and Karayev, S. (2021) "Full page handwriting recognition via image to sequence extraction," in Document Analysis and Recognition – ICDAR 2021. Cham: Springer International Publishing, pp.55–69.

[16] Sinha, A., Yadav, A. and Gahlot, A. (2018) "Extractive Text Summarization usingNeuralNetworks," arXiv [cs.CL]. doi: 10.48550/ARXIV.1802.10137.Zhang,

Y., Meng, J. E. and Pratama, M. (2016) “Extractive document summarization based on convolutional neural networks,” in IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society. IEEE, pp. 918–922.

[17] (No date) Researchgate.net. Available at:https://www.researchgate.net/publication/348531563_Text_Summarization_Using_Text_Frequency_Ranking_Sentence_Prediction (Accessed: June 22,2022).

[18] Dalal, V. and Malik, L. (2013) “A Survey of Extractive and Abstractive Text Summarization Techniques,” in 2013 6th International Conference on Emerging Trends in Engineering and Technology. IEEE, pp.109–110.

[19] Miller, D. (2019) “Leveraging BERT for Extractive Text Summarization on Lectures,” arXiv [cs.CL]. doi: 10.48550/ARXIV.1906.04165