JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Generating Minutes of the Meeting using NLP

[1]Tanvi Budhabaware, [2]**Siddharth Khurangale**, [3]**Atharva Wagh,** [4]**Safal Taksale,** [5]**Dr. Shilpa Khedkar,**

[(1,2,3,4)] Student, [5]Associate Professor
[(1,2,3,4,5)] Department of Computer Engineering,
[1]Modern Education Society's College of Engineering,
19, Late Prin. V.K. Joag Path, Wadia College Campus, Pune - 411001

Abstract: Online meetings and remote work have grown commonplace in the wake of the COVID-19 epidemic, underscoring the importance of effective meeting minutes creation. This study investigates the use of Natural Language Processing (NLP) methods to automate the creation of meeting minutes. The two main objectives of the study are gathering meeting transcripts and producing succinct summaries. To enable additional analysis, the recorded meeting speech is converted into text using voice-to-text technology. For the purpose of producing coherent and human-like summaries, extractive and abstractive summarization techniques—including the usage of GPT models—are examined. The system's architecture combines analogue to digital conversion with speech to text processing, text generation, and data processing to produce meeting minutes that are both accurate and efficient. The research provides insights into the developments and difficulties in this area and emphasizes the value of NLP in creating meeting minutes. In the post-COVID-19 era, the suggested automated method has the potential to increase productivity and simplify the creation of meeting minutes.

***Keywords*:** Natural Language Processing, BERT, Text Summarization, GPT, Abstractive Summarization, Extractive Summarization

I. INTRODUCTION

We have seen a significant shift in the educational and IT sectors towards online classes, online meetings, and remote employment since the COVID-19 wave hit the world. Meetings have become a regular part of our lives since COVID-19, and they frequently involve difficulties like network issues, poor connectivity, and other difficulties. The goal of meeting minutes is to summarize the conversation and act as a resource. Natural Language Processing (NLP)-based automated meeting minutes creation is crucial. Collecting meeting transcripts and creating the minutes (summary) are the two parts of the task of creating meeting minutes using NLP. Let's look more closely at these duties. The gathering of meeting minutes is the first task. The option to create meeting transcripts while the meeting is still going on is available on several online meeting platforms, including Google Meet, Microsoft Teams, and Zoom. The recorded meeting speech can be turned into text in situations where meeting transcripts cannot be obtained directly. However, the minutes creation procedure does not work well with plain text alone. To clearly summarize who said what and what was discussed throughout the meeting, it is crucial to identify the speaker.
The creation of minutes or summaries is the second task. Producing meeting minutes differs from summarizing basic material like articles, news, or stories for the reasons listed below:

1. Context and Structure: Meeting transcripts frequently feature many presenters, interruptions, and informal language in a conversational style. Meetings necessitate a grasp of their conversational nature, unlike separate bits of text with a logical framework.
2. Conversational Understanding: A better awareness of conversational dynamics is necessary to create a meeting summary, including the ability to identify speakers, discern discourse intentions, and comprehend contextual references.
3. Specific Vocabulary: Meetings frequently feature technical phrases, acronyms, and jargon that aren't always available in simple text.
4. Structured Output: Meeting summaries adhere to a structured format, unlike plain text summarization, which allows for more flexibility in terms of output structure.

Therefore, separate algorithms are needed for meeting minutes generation, distinct from the existing algorithms used for plain text summarization.

II. SPEECH TO TEXT FOR MEETING TRANSCRIPTS

We must identify each speaker in the meeting as well as the speech-to-text interaction because the recorded meeting will undoubtedly have numerous speakers. This would enable us to grasp each speaker's words clearly. We can employ agglomerative clustering to produce transcripts that include speaker names. In machine learning and data analysis, aggregative clustering is a method for combining related data points. You can use agglomerative clustering on the embeddings (representations) of speech segments in the context of speech-to-text conversion of a meeting to determine the speaker for each segment. Consider the following example:

Person A: "Did you watch the game last night?"
Person B: "Yeah! The final score was unbelievable."
Person C: "I missed it. Who won?"
Person A: "Team India won."

To perform agglomerative clustering, we first initialize clusters for each individual word spoken in the meeting.

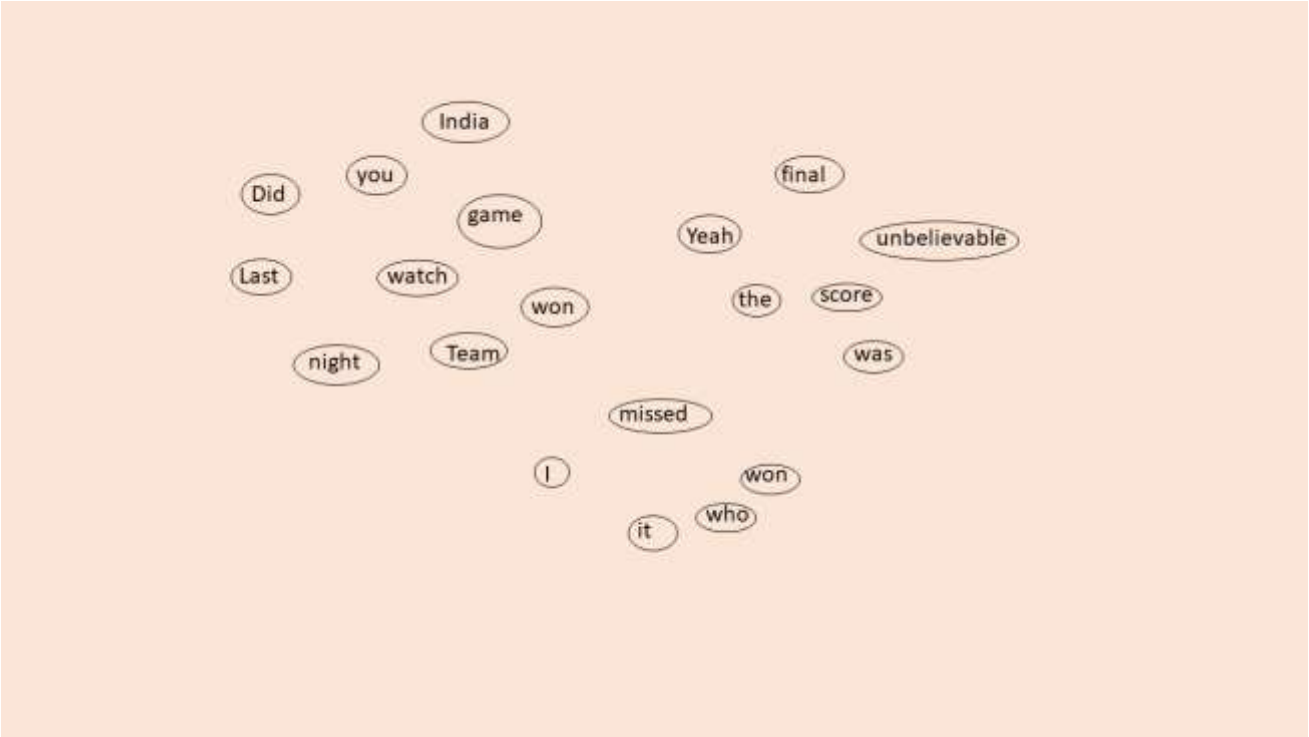

Fig. 1 Individual Clusters

Then, we calculate the similarity (distance) between the clusters using metrics like cosine similarity or Euclidean distance and merge the similar clusters.

At this point, all the sentences have been grouped into three different clusters:

Cluster 1: ["Did you watch the game last night?", "Team India won."]
Cluster 2: ["Yeah! The final score was unbelievable."]
Cluster 3: ["I missed it. Who won?"]

According to their embeddings, related sentences are grouped together through agglomerative clustering to create cohesive clusters. In this instance, it successfully distinguishes between the three distinct talks taking place within the dialogue.

OpenAI Whisper, Torch, Pyannote, and Scikit-learn are just a few of the libraries that are used to implement the transcript generation from audio (voice to text) process. These libraries aid in the audio file's conversion. Speech-to-text conversion, MP3 to WAV conversion, and speaker identification using clustering algorithms.

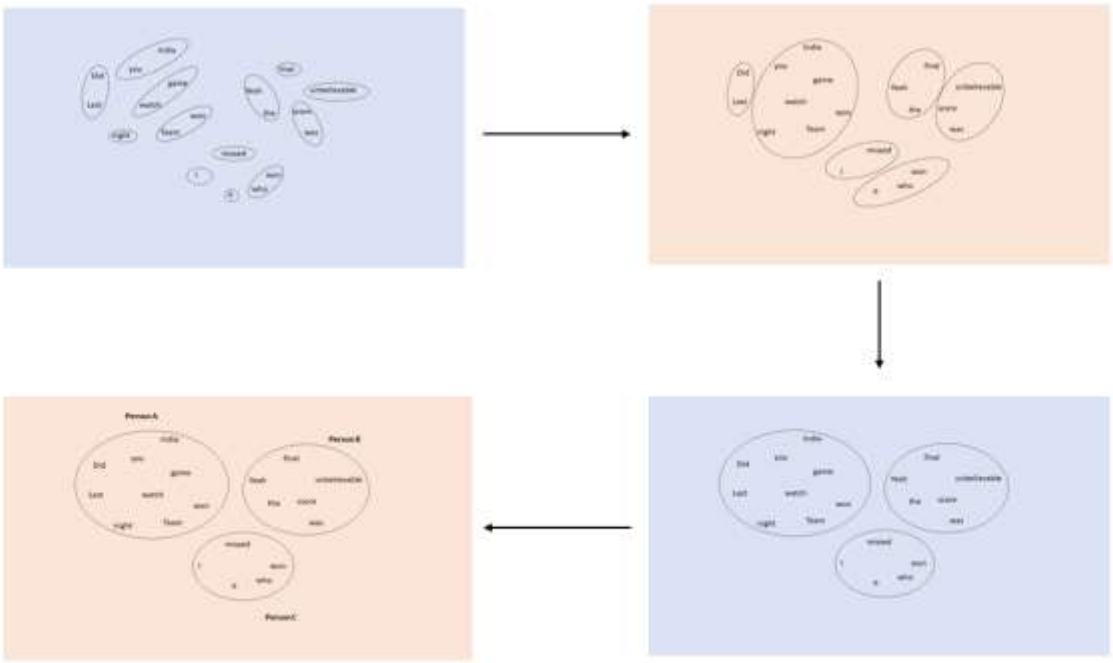

Fig. 2 Agglomerative Clustering

III. GENERATING MINUTES OF THE MEETING

There are two main methods of summarization: extractive summarization and abstractive summarization. In extractive summarization, pre-existing sentences or phrases are chosen from the source text according to criteria like prominence or relevance. Simply said, it is equivalent to underlining key passages in the text. Conversely, abstractive summarization reduces the amount of information by creating new phrases that accurately reflect the original material. It's like telling your pals the whole tale in your own words, to put it simply. Both approaches have benefits and drawbacks, so employing only one would not be suitable for all kinds of meetings. Using the model's own words and sentences while maintaining the meaning, abstractive summarization delivers a summary of the meeting sessions. It is vital to list action items in the summary for meetings where significant decisions and responsibility assignments are made. In these situations, it is recommended to employ the extractive technique, in which the model chooses portions (action items) from the meeting transcript to create the summary.

3.1 ABOUT THE LANGUAGE MODELS

The effectiveness of transformer models on various NLP tasks was recently proved by BART, T5, and GPT employing pre-trained language on large-scale datasets. These models can be used to summarize meetings. Each of these models have their own advantages and pitfalls let's look into it -

T5: T5 model is based on 'text-to-text" framework and it can be fine-tuned for various NLP tasks.

BERT: BERT is a potent model that effectively handles extended sequences and captures rich contextual representations, making it useful for summarizing tasks.

GPT Model: GPT models excel in producing coherent and human-like summaries, making them appropriate for summarization tasks because to their autoregressive nature and broad context window. They have strong language generating abilities and were trained via unsupervised learning.

Following is a table demonstrating the comparison between these models –

Model	Strengths	Weaknesses	Suitable for
T5	Extractive summarization, concise and coherent summaries, fine-tuning capability	Less suited for extractive summarization, may require a smaller dataset for fine-tuning	Generating high-quality summaries for the minutes of the meeting
BERT	Abstractive summarization, good for selecting sentences from original text	May result in a summary that is a concatenation of sentences, not as suited for abstractive summarization	Generating summaries that are a concatenation of selected sentences from the original text
GPT-3	Natural language text generation, large pre-training data	May be more resource-intensive for fine-tuning and deployment	Generating summaries that are easier to understand and more readable

Fig 3. Model Comparison

Based on the comparison, we come to the conclusion that GPT should be chosen for summarizing meeting minutes since it can produce writing that is human-like and coherent, leading to summaries that are interesting and readable. Because it is autoregressive, context coherence can be ensured by generating it word by word. Additionally, the broad context window of GPT makes it possible to record long-range relationships, which makes it easier to depict meeting topics in-depth in the generated minutes.

IV. SYSTEM DESIGN

The system design for the architecture could be explained with the following diagram:

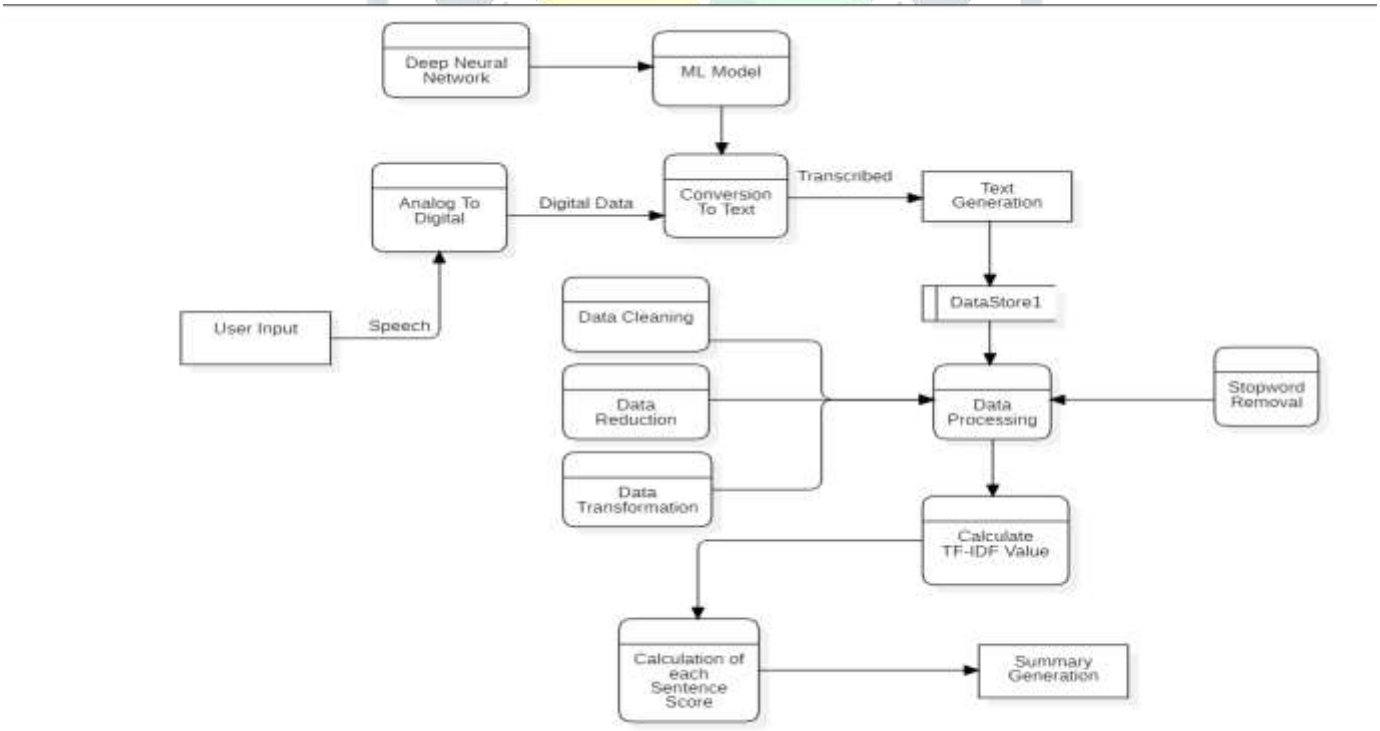

Fig. 4 System Design

A recorded meeting from the user is the first piece of input, and it is transformed from analogue to digital format. The audio data can now be further processed and analysed thanks to this conversion. Next, a deep neural network or machine learning (ML) model that has been trained particularly for speech recognition performs the speech-to-text conversion process. The meeting

transcript is created in this phase by text-to-speech converting the audio recordings. After that, a text generation module—like a summary algorithm—is applied to the resultant text. This module analyses and extracts the crucial information from the transcript using methods like natural language processing (NLP). Data processing is a step in the process of creating a summary, when pertinent information is gathered and organized. The generated summary is kept for simple access and retrieval in a DataStore1 or other defined storage system. Analogue to digital and audio to text data transformations happen throughout the entire process, allowing for effective processing and the creation of the meeting minutes.

V. CONCLUSION

Natural language processing (NLP)-based automated meeting minutes creation is crucial, especially given the growing prevalence of online meetings and distant work. Gathering meeting transcripts and creating the minutes (summary) are the procedure' two key duties. Deep neural networks or machine learning models can be used to convert speech to text to create meeting transcripts. Extractive or abstractive summarizing techniques must be used to create minutes, with GPT models standing out for their capacity to provide cohesive and human-like summaries. To enable the quick and precise creation of meeting minutes, the system architecture combines analog-to-digital conversion, speech-to-text processing, text production, and data processing. In the post-COVID-19 era, the automated generation of meeting minutes can substantially speed the process and increase productivity thanks to developments in NLP and the availability of potent language models.

VI. REFERENCES

[1] Nenkova, A.(2011).“Automatic summarization, Foundations and Trends in Information Retrieval”,5(2),103-233

[2] J.N.Madhuri . "Extractive Text Summarization Using Sentence Ranking"

[3] Goldstein.J,carbonell.J,Kantrowitzt.M(1998).“Multiple document summarization by sentence Extraction”40-48

[4] Shivakumar K M , Varsha V Jain and Krishna Priya P. "A study on impact of Language Model in improving the accuracy of Speech to Text Conversion System "

[5] Mayank Ramina , Nihar Darnay, Chirag Ludbe, Ajay Dhruv. " Topic level summary generation using BERT induced Abstractive Summarization Model "

[6] N. Moratanch and S. Chitrakala, A survey on abstractive text summarization,” 2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT), Nagercoil, 2016, pp. 1-7.J. Breckling, Ed., The Analysis of Directional Time Series: Applications to Wind Speed and Direction, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.

[7] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate”, 2014 arXiv:1409.0473.

[8] Ilya Sutskever, Oriol Vinyals, Quoc V. Le, “Sequence-to-Sequence Learning with Neural Networks”, 2014, arXiv: 1409.3215.

[9] Y.Bengio, P. Simard, and P. Frasconi. Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks, 5(2):157–166, 1994.

[10] K. Cho, B. Merrienboer, C. Gulcehre, F. Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Arxiv preprint arXiv:1406.1078, 2014.

[11] Ramesh Nallapati, Bowen Zhou, Cicero dos Santos, Caglar Gulcehre, Bing Xiang, “Abstractive text summarization using sequence to-sequence RNNs and beyond”, 2016, arXiv:1602.06023.

[12] G. E. Dahl, D. Yu, L. Deng, and A. Acero. Context-dependent pre-trained deep neural networks for large vocabulary speech recognition. IEEE Transactions on Audio, Speech, and Language Processing - Special Issue on Deep Learning for Speech and Language Processing, 2012. “PDCA12-70 data sheet,” Opto Speed SA, Mezzovico, Switzerland.