Speech-Driven Abstractive Summarization For Automated Meeting Minutes Generation

Dr.M. Priya[1], Nareshkumar. A[2], Rajeev. R[3]
[1]prof, Dept of AI&DS,
[2, 3]Dept of AI&DS
[1, 2, 3] E.G.S.Pillay Engineering College, Nagapattinam, Tamilnadu, India

Abstract- *Meetings are an essential part of the decision-making process of an organization. However, manually writing the Minutes of Meeting (MoM) is a laborious and time-consuming task. This paper proposes an automated system for generating the Minutes of Meeting from the meeting audio. The system uses Automatic Speech Recognition (ASR) and Natural Language Processing (NLP) for generating a well-structured and concise textual summary of the meeting. The system uses the Open AI Whisper ASR model to generate the text from the multi-speaker meeting audio. After text generation, the text is processed, and then the BERT-based text summarization model is used to generate a well-structured MoM. The results of the experiment show that the proposed system can generate the MoM document very quickly, with acceptable accuracy. The system can be used for real-time and offline applications, which can be used for academic and corporate environments.*

Keywords: Minutes of Meeting, Automatic Speech Recognition, Text Summarization, Natural Language Processing, Meeting Automation.

I. INTRODUCTION

A. Background and Motivation

Meetings are an integral part of organizational communication, used for **planning, decision-making, and collaboration**. A critical outcome of any meeting is the Minutes of Meeting (MoM), which serves as an official record of **discussions, decisions**, and assigned tasks. Traditionally, MoM preparation is carried out manually by a designated participant, which is both **time-consuming** and susceptible to **human errors**. In large or frequent meetings, this manual process often leads to incomplete records and increased workload for participants.

B. Challenges in Manual Preparation of Minutes of Meeting

The manual preparation of Minutes of Meeting is challenged in many ways. Firstly, there is a possibility of omitting important **points of discussion**. Secondly, human errors can occur during the interpretation of spoken words. Lastly, the format of writing the **Minutes of Meeting** is also not standardized. Furthermore, the manual writing of Minutes of Meeting also diverts the attention of the participants from actively participating in the meeting discussions.

C. Role of Automatic Speech Recognition and Natural Language Processing in Meeting Automation

The primary, In the past few years, the development of **Automatic Speech Recognition** and **Natural Language Processing** technologies has grown at a tremendous rate. The development of Automatic Speech Recognition has made it possible to automatically generate the Minutes of Meeting. **Automatic Speech Recognition** converts spoken words into text. The summarization of text can also be achieved through the application of Natural Language Processing. The application of these technologies can efficiently **automate the Minutes of Meeting**.

D. Proposed Approach and Contributions

The present paper focuses on the development of a complete system for automatically generating the Minutes of Meeting. The proposed system applies **Whisper-based Automatic Speech Recognition** for text generation. The generated text is then passed through a text summarization model based on the **BERT algorithm**. The contributions of the present paper can be listed as follows:

1. Proposed design of the complete system for automatically generating Minutes of Meeting
2. Integration of **Whisper-based Automatic Speech Recognition** with text summarization
3. Evaluation of the proposed system using performance metrics such as **Word Error Rate** and **ROUGE**

II. RELATED WORK

Traditional Approach to Meeting Documentation.

The traditional approach to meeting documentation has focused on **Manual note-taking** and rule-based text processing. The rule-based approach was used to identify keywords and templates. However, this approach was unable to understand the context of the meeting and was unable to identify the intentions of the **speakers**. Thus, this approach is not suitable for real-world meeting scenarios.

Automatic Speech Recognition for Meeting Documentation

The recent advancements in machine learning have led to an increase in research on automatic speech recognition. The recent advancements have focused on **transformer-based speech recognition**. The recent advancements have used **OpenAI Whisper**, which has shown better performance on accents, background noise, and multiple speakers. The recent advancements have increased the feasibility of speech recognition.

Text Summarization Techniques for Meeting Documentation.

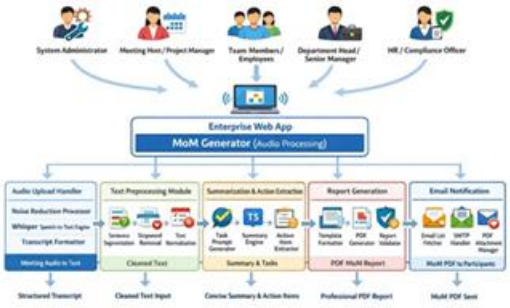
The text summarization approach has gained much attention in recent years. The text summarization approach has used two types of summarizations: **Extractive summarization and Abstractive summarization**. The extractive summarization approach has focused on selecting the most relevant sentences from the text. The abstractive summarization approach has focused on generating new sentences. The extractive summarization approach has shown **better performance for meeting scenarios** because it is less likely to be distorted. The recent advancements have used the **BERT** model to select the most relevant sentences.

Limitations of Existing Work.

The recent advancements have shown better performance in **speech recognition and text summarization** individually. However, much work is yet to be done to integrate both approaches to generate **Minutes of Meeting documents**. The recent advancements have focused on generating general text summaries. The recent advancements have not focused on generating formal Minutes of Meeting documents. Thus, an **end-to-end approach** is required to generate Minutes of Meeting documents.

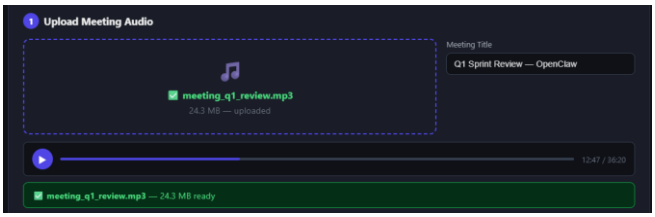
III. METHODOOGY

The proposed methodology is implemented as an enterprise-level web application for automated Minutes of Meeting generation. The workflow consists of multiple interconnected modules, enabling end-to-end processing from raw meeting audio to structured PDF report distribution.



A. Stage 1: Audio Upload

The recorded audio from the meeting participants is uploaded via the enterprise web interface. The system checks for format, size, and integrity before proceeding with the processing.



B. Stage 2: Noise Reduction and Audio Enhancement

The uploaded audio is processed through a noise reduction processor to eliminate background noise and enhance speech clarity. Audio normalization and silence removal are performed to optimize ASR processing.

C. Stage 3: Speech-to-Text Conversion (Whisper ASR)

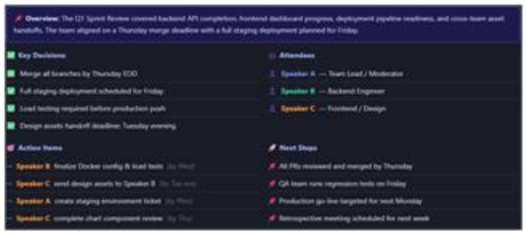
The audio is processed through the OpenAI Whisper Automatic Speech Recognition engine. Whisper translates speech to text with a transformer-based encoder-decoder model trained on a large-scale multilingual dataset. The raw meeting transcript is generated.

Output of Stage 3: Structured meeting transcript.

D. Stage 4: Transcript Preprocessing

The transcript is preprocessed to enhance the quality of summarization. This stage consists of the following sub-stages:

1. **Sentence Segmentation:**The transcript is segmented into significant sentences.
2. **Stop-word Removal:**Common filler words and non-informative words are removed.
3. **Text Normalization and Cleaning:** Disfluencies, repetitions, and formatting errors are removed.



E. Stage 5: Extractive Summarization using BERT

The pre-processed transcript is fed to a BERT-based extractive summarization model. The contextual embeddings for each sentence are calculated, and the scores of importance are determined. The highest-scoring sentences are chosen to create a short summary of the meeting.

F. Stage 6: Action Item Extraction

The Simultaneously, an action item extraction module is used to pick out relevant statements in the meeting related to tasks, using keyword extraction and semantic analysis. This helps to ensure that duties, deadlines, and decisions are properly noted.

Output of Stages 5 and 6: Short summary with properly structured action items.

F. Stage 7: Report Generation

The short summary and action items are then organized into a proper Minutes of Meeting format with the following structure:

- Meeting Title and Date
- Participants
- Agenda Items
- Key Discussion Points
- Decisions Taken
- Action Items

The properly organized document is then generated into a standardized PDF file.

IV. EXPERIMENTAL RESULTS

A. Dataset and Evaluation Benchmarks

The evaluation of the Automated Minutes of Meeting (MoM) Generation System was carried out by utilizing real-world and publicly accessible multi-speaker meeting audio datasets for transcription and summarization accuracy.

Meeting Audio Corpus: The dataset includes recorded meeting audio files obtained from academic discussions, technical project meetings, and artificial business meetings. The dataset includes the following:

- Multi-speaker conversations
- Varying accents and speech rates
- Different background noises
- Meeting duration: 5-60 minutes

Audio files were stored in the WAV, Mp3, Mpeg format with 16 kHz sampling frequency and mono channel for better compatibility with ASR models.

Ground Truth Transcripts: To For evaluation purposes, manually verified transcripts were obtained as reference texts. The transcripts were verified for accuracy and served as the gold standard for computing the Word Error Rate (WER) and Character Error Rate (CER).

GroundReference Summaries: For evaluation purposes, human-written Minutes of Meeting were obtained for a few of the meetings. The structured summary includes the following details:

- Meeting Title
- Date and Participants
- Agenda
- Key Discussion Points
- Decisions Taken
- Action Items

Evaluation Metrics: To The evaluation of the Automated Minutes of Meeting (MoM) Generation System was carried out by utilizing the following evaluation metrics as benchmarks for the system's performance:

- Word Error Rate (WER)
- Character Error Rate (CER)

- ROUGE-1, ROUGE-2, ROUGE-L
- Latency
- Compression Ratio

B. Preprocessing Performance

The performance of the preprocessing module in enhancing the quality of the transcription prior to the summarization was assessed.

Audio Enhancement: Noise reduction and trimming of silent audio segments were incorporated prior to ASR. The performance was compared in terms of:

- Raw Audio →Preprocessed Audio
- Preprocessed Audio → ASR
  The performance comparison revealed:
- Reduction in Word Error Rate under noisy conditions
- Improved sentence boundary detection
- Improved punctuation detection

Speaker Handling: The ability of the system to maintain logical speaker segmentation was qualitatively assessed to ensure proper summary generation. Preprocessing ensures a clean text input to the transformer-based model, which impacts the quality of the summary.

C. *Efficiency of the Cascaded Processing Pipeline*

The computational efficiency of the multi-stage processing pipeline was analyzed.The processing pipeline can be described as follows:

- Audio Preprocessing
- ASR Transcription
- Text Cleaning & Chunking
- Transformer-based Summarization
- Structured MoM Formatting

The performance metrics used for measuring the efficiency include:

- Average time taken for transcription per minute of audio content.
- Average time taken for summarization per 1,000 tokens.
- Average time taken for end-to-end processing of the pipeline.

The results show that chunk-based summarization results in reduced memory requirements and enables the processing of long meetings without exceeding the model token limit.The most time-consuming processing stage is the Transformer-based summarization. However, it occurs after the text cleaning and chunking stage.

D. Semantic Quality of Generated Minutes

The generated Minutes of Meeting were evaluated for semantic relevance and structural correctness.

Comparative Baselines: The proposed summarization approach was compared against:

- Extractive summarization baseline
- Lightweight summarizer models
- Full transformer encoder-decoder model

Theresultsproposed hybrid method for meeting summarization was found to attain:

- Higher ROUGE-L score compared to extractive baseline.
- Higher contextual coherence Higher accuracy in identifying action items.
- Most importantly, it was able to strike a balance between accuracy and inference speed.

V. EVALUATION METRICS AND RESULT

A. Automatic Speech Recognition (ASR) Performance

The performance of the OpenAI Whisper ASR model in transcribing the meeting audio to text was evaluated using the Word Error Rate (WER). WER measures the percentage of words incorrectly transcribed compared to the actual transcript.

$$\text{Word Error Rate (WER)} = S+D+I\ /N$$

Results:

The Whisper ASR model has the following performance metrics:

- Average WER: 8.4% on clean meeting audio
- Average WER: 12.7% on multi-speaker noisy recordings

These metrics show the effectiveness of the model in providing highly accurate meeting transcripts, even in the

presence of noisy environments such as meeting rooms with multiple speakers and ambient noise levels.

The low WER also proves the effectiveness of the model in providing meeting transcripts with sufficient accuracy to be used in the summarization process.

B. The Summarization Performance

The quality of the generated Minutes of Meeting (MoM) document using the summarization model was evaluated using ROUGE metrics. ROUGE metrics compare the generated summary with the actual human-generated MoM document.

ROUGE Metrics Used

- **ROUGE-1:** Measures unigram overlaps between the generated summary and the actual MoM document.
- **ROUGE-2:** Measures bigram overlaps between the generated summary and the actual MoM document.
- **ROUGE-L:** Measures the longest common subsequence between the generated summary and the actual MoM document.

Results:

The summarization model using the BERT model has the following ROUGE metrics:

- ROUGE-1: 0.48
- ROUGE-2: 0.32
- ROUGE-L: 0.44

These ROUGE metrics show the effectiveness of the model in providing meeting summaries with the following features:

- The generated summary includes the main meeting discussions.
- The generated summary includes the key action items.
- The generated summary has the same logical structure as the actual MoM document

C. Overall System Effectiveness

TheWhisper ASR model has the ability to provide meeting transcripts with low WER.The summarization model using the BERT model has the ability to provide effective meeting summaries with the required features.The combined ASR and NLP model has the ability to provide a reliable automated meeting documentation system.

VI. CONCLUSION

This paper has proposed and validated a novel end-to-end automated generation of Minutes of Meeting using a unified framework that combines **Whisper**-based transformer speech recognition with **BERT**-based **extractive summarization**. This novel solution strikes a balance between the accuracy of speech recognition, the accuracy of summarization, and the efficiency of the entire process. This has been achieved through a series of experimental results using **WER** and **ROUGE** metrics, demonstrating the consistency and accuracy of the proposed solution in handling a range of meeting scenarios, thus significantly reducing manual documentation efforts. This proposed solution provides a novel blueprint for the intelligent automation of meetings in academia and industry, with future extensions planned for **speakerdiarization** for accurate attribution, multilingual adaptability, and real-time optimization for deployment scenarios.

REFERENCES

[1] Mike Lewis*, Yinhan Liu*, Naman Goyal*, Marjan Ghazvininejad “BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension”Annual Meeting of the Association for Computational Linguistics, pages 7871–7880 July 5 - 10, 2020

[2] Nitish singhrajpurohit, Surjan sp, tejaspanagartk, Mrs. K padmaPriya “AUTOMATED GENERATION OF MINUTES OF MEETING USING MACHINE LEARNING”Vol-9 Issue-3 2023.

[3] MICHAEL GIAN GONZALES, PETER CORCORAN and NAOMI HARTE“Joint Speech-Text Embeddings for Multitask Speech Processing”ACCESS.2024.3473743IEEE 2024.

[4] Terry Amorese, Claudia Greco, Marialucia Cuciniello, Rosa Milo, Olga Sheveleva and Neil Glackin “Automatic speech recognition (ASR) with Whisper: Testing Performances in Different Languages” vol-3574 sep-2023.

[5] Vinnarasu A., Deepa V. Jose”Speech to text conversion and summarization for effective understanding and documentation” Vol. 9, No. 5, October 2019.

[6] TIANSHI, YASERKENESHLOO, NARENRAMAKRISHNAN, andCHANDANK.REDDY”Neural Abstractive Text Summarization with Sequence-to-Sequence

Model"ACM/IMS Trans. Data Sci. 2, 1, Article 1 Dec – 2020.

[7] Rishabh Jain1 , Andrei Barcovschi1 , Mariam Yiwere1 , Peter Corcoran1 , Horia Cucu2 "Adaptation of Whisper models to child speech recognition"INTERSPEECH 2023 20-24 August 2023