AUTOMATED GENERATION OF MINUTES OF MEETING USING MACHINE LEARNING

NITISH SINGH RAJPUROHIT, SRUJAN SP, TEJAS PANAGAR TK, Mrs. K PADMA PRIYA

Student, Department of Information Science and Engineering, AMCEC, Bangalore- 083, Karnataka
Faculty, Department of Information Science and Engineering, AMCEC, Bangalore- 083, Karnataka

ABSTRACT

The automated generation of meeting minutes using machine learning techniques has emerged as a promising solution to alleviate the time-consuming and error-prone manual process of capturing and summarizing meeting discussions. This research paper presents a novel approach for automating the generation of meeting minutes through the application of machine learning algorithms. The proposed method employs natural language processing (NLP) techniques to analyze the meeting transcripts and extract relevant information. The text is processed to identify key topics, important decisions, action items, and other significant details discussed during the meeting. Machine learning models are trained on annotated datasets to classify and extract the relevant information accurately. Furthermore, the research explores the use of advanced machine learning algorithms, such as recurrent neural networks (RNNs) and transformers, to capture the context and nuances of meeting conversations. These models enhance the accuracy and coherence of the generated meeting minutes. The evaluation of the automated meeting minutes generation system involves comparing the outputs against manually generated meeting minutes by human notetakers. Metrics such as precision, recall, and F1-score are used to assess the system's performance, ensuring the quality and accuracy of the generated minutes. The findings demonstrate that the automated generation of meeting minutes using machine learning techniques offers significant time savings, reduces human error, and improves overall efficiency in capturing and summarizing meeting discussions. The system exhibits promising potential for adoption in various industries and organizations, where efficient meeting documentation is vital. In conclusion, this research paper presents a comprehensive study on the automated generation of meeting minutes using machine learning. The proposed approach leverages NLP techniques and advanced machine learning algorithms to accurately extract and summarize meeting content. The results highlight the potential of this automated system to streamline meeting processes and enhance productivity in diverse organizational settings.

KEYWORDS: *Natural language processing, Speech recognition, Artificial intelligence machine learning, whisper AI, neural networks.*

INTRODUCTION

In today's fast-paced business environment, meetings play a crucial role in facilitating collaboration, decision-making, and information sharing among team members. However, the process of capturing and summarizing meeting discussions into comprehensive minutes can be time-consuming and prone to human error. To address this challenge, automated generation of meeting minutes using machine learning techniques has gained significant attention. This research paper focuses on exploring the potential of

machine learning, specifically with the integration of Whisper AI, to automate the generation of meeting minutes. Whisper AI is an advanced natural language processing (NLP) platform that leverages state-of-the-art machine learning algorithms to analyze and understand human language. By applying Whisper AI in the context of meeting minutes generation, we aim to develop a system that can accurately and efficiently extract key information from meeting transcripts and transform it into concise and meaningful summaries. The integration of machine learning algorithms within the Whisper AI framework enables the system to learn from annotated datasets and develop models that can recognize important meeting elements, such as agenda items, action items, decisions, and discussions. The use of machine learning algorithms, including deep learning techniques such as recurrent neural networks (RNNs) and transformers, allows the system to capture the contextual information and nuances of meeting conversations, leading to more accurate and coherent generated minutes. The research paper will delve into the technical details of the implementation, outlining the various stages involved in the automated generation process. This includes pre-processing and cleaning of meeting transcripts, feature extraction, model training, and evaluation. The evaluation of the system will be conducted by comparing the automatically generated minutes against manually created minutes by human notetakers. Performance metrics such as precision, recall, and F1-score will be employed to assess the quality and accuracy of the generated minutes. The expected outcomes of this research include a comprehensive understanding of the capabilities and limitations of automated meeting minutes generation using machine learning and Whisper AI. The findings will contribute to the growing body of knowledge in the field of natural language processing, machine learning, and automated document generation. Additionally, the research aims to demonstrate the practical application and potential benefits of this technology in various industries and organizational settings, where efficient meeting documentation is crucial for productivity and decision-making processes. Overall, this research paper aims to explore the integration of machine learning and Whisper AI for automating the generation of meeting minutes. The results of this study have the potential to revolutionize the way meetings are documented, streamlining the process and improving overall efficiency in capturing and summarizing important meeting discussions.

SYSTEM DESIGN

SPEECH TO TEXT

The automated generation of meeting minutes using machine learning and Whisper AI involves a comprehensive system design that encompasses data processing, machine learning models, and output generation. Whisper is an automatic speech recognition (ASR) system trained on 680,000 hours of multilingual and multitask supervised data collected from the web. Moreover, it enables transcription in multiple languages, as well as translation from those languages into English..

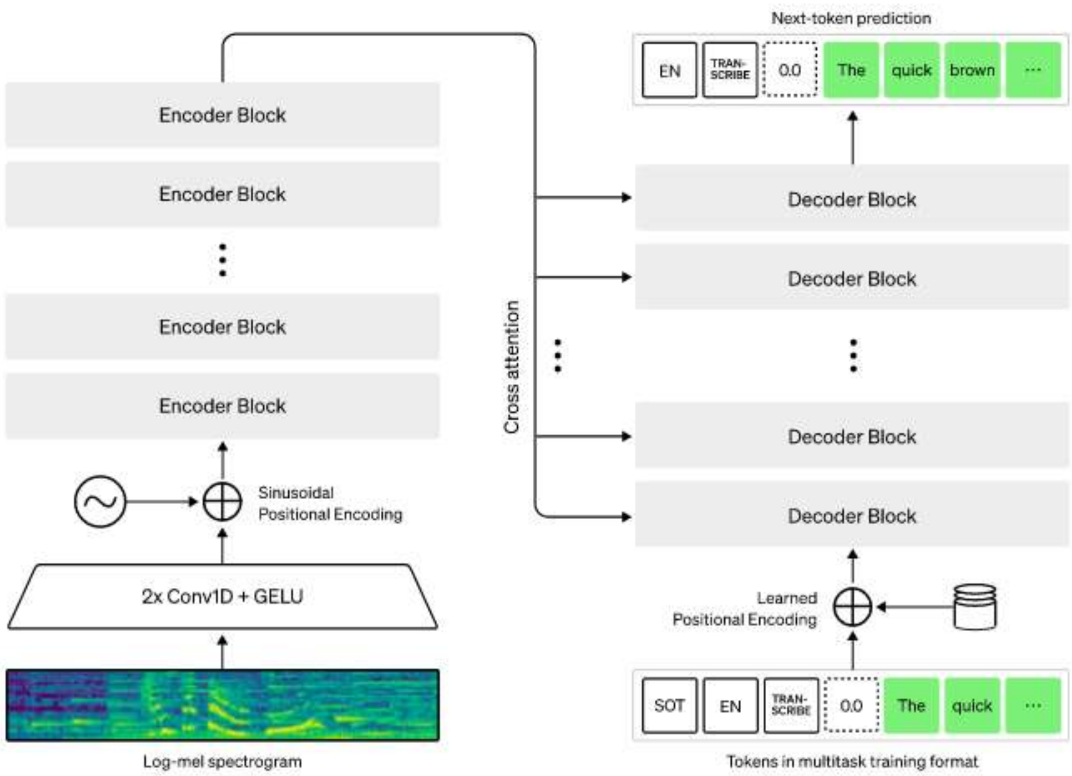


Fig. Acoustic model to recognize speech to text

The Whisper architecture is a simple end-to-end approach, implemented as an encoder-decoder Transformer. Input audio is split into 30-second chunks, converted into a log-Mel spectrogram, and then passed into an encoder. A decoder is trained to predict the corresponding text caption, intermixed with special tokens that direct the single model to perform tasks such as language identification, phrase-level timestamps, multilingual speech transcription, and to-English speech translation.

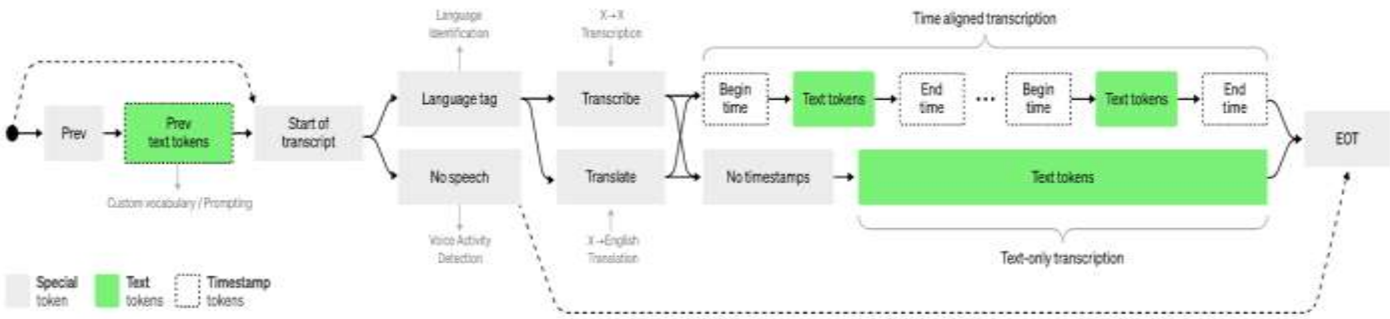


Fig. Language model to predict text

Other existing approaches frequently use smaller, more closely paired audio-text training datasets,1 2,3 or use broad but unsupervised audio pretraining.4,5,6 Because Whisper was trained on a large and diverse dataset and was not fine-tuned to any specific one, it does not beat models that specialize in LibriSpeech performance, a famously competitive benchmark in speech recognition. However, when we measure Whisper's zero-shot performance across many diverse datasets we find it is much more robust and makes 50% fewer errors than those models..

SPEAKER IDENTIFICATION

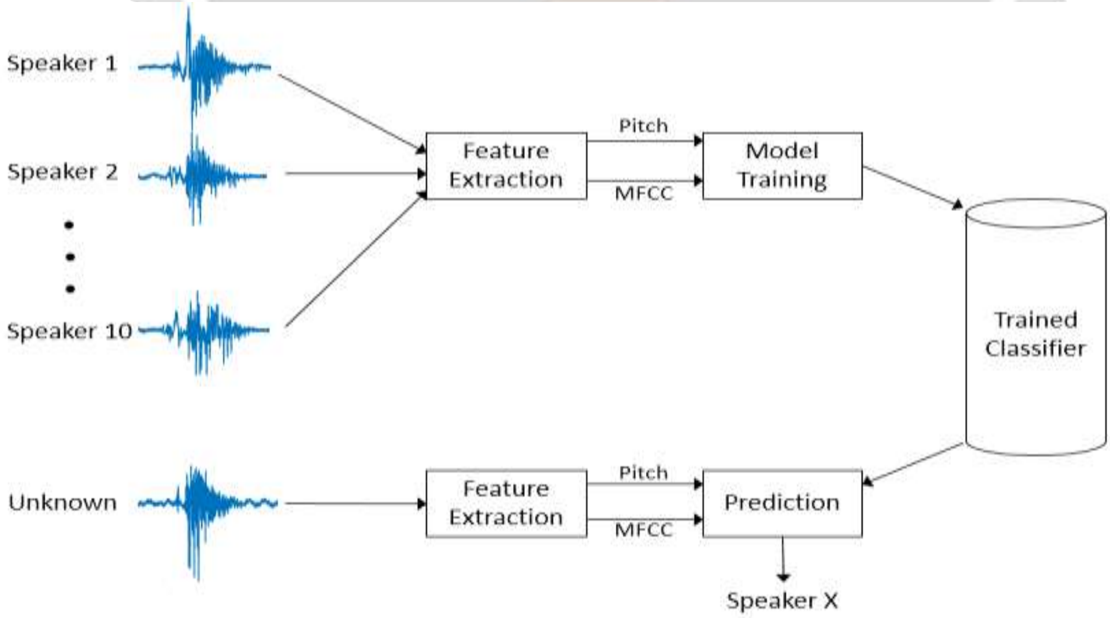


Fig. Unique audio recognition for speaker identification

MFCC, LPC, LPCC, LSF, PLP and DWT are some of the feature extraction techniques used for extracting relevant information form speech signals for the purpose speech recognition and identification. These techniques have stood the test of time and have been widely used in speech recognition systems for several purposes.

SPEECH SUMMARIZATION

BERT is an open source machine learning framework for natural language processing (NLP). BERT is designed to help computers understand the meaning of ambiguous language in text by using surrounding text to establish context. The BERT framework was pre-trained using text from Wikipedia and can be fine-tuned with question and answer datasets. BERT which stands for Bidirectional Encoder Representations from Transformers, is based on Transformers, a deep learning model in which every output element is connected to every input element, and the weightings between them are dynamically calculated based upon their connection. (In NLP, this process is called *attention*.)

```
[ ] from summarizer import Summarizer,TransformerSummarizer
```

```
Text='''
Good morning myself stages. Today we will be discussing debating about AMC engineering
college. AMC engineering college is an undergraduate college located in the 18th kilometer of Manalgata
road. We are students from the ISA department currently in the final year. Over to my point.
Good morning my name is Nitish and I wanted to share the college infrastructure and the
facilities and the education program it is very good and all the teachers and the staff are very
helpful to everyone and a good environment is formed in the college and a great learning
platform for everyone who wants to pursue engineering in any field. Over to my point.
I'm going to speak against the topic about the AMC engineering college like this is like
18 kilometers far away from the Bangalore town on Nagar where many of the people are residing
and the people who know about the college only like provide the details about for the others
who don't know about the college but like us like other famous colleges like we are VMS they don't
they don't know like from before only so this is like major cons I am thinking of thing right now
Thank you.
In this college like we didn't have any metro or any other facilities going on
they can't have the constructions and the traffic will be increased when traveling to college and
going more up and down will be increased time.
For about two semesters during the COVID we had almost no colleges
and the staff are teaching online thing but we had a college design
and we were online promoted for two sems and we didn't focus on the part which was
meant to be learned by us during that period of time.
By those consequences we are getting the trouble to implement the project right now.'''
```

```
[ ] #acceptable 0.5

bert_model = Summarizer()
bert_summary = ''.join(bert_model(Text))
print(bert_summary) #bert summarizer
```

```
bert_model = Summarizer()
bert_summary = ''.join(bert_model(Text))
print(bert_summary) #bert summarizer
```

```
Downloading (…)lve/main/config.json: 100% 571/571 [00:00<00:00, 18.2kB/s]
Downloading pytorch_model.bin: 100% 1.34G/1.34G [00:21<00:00, 61.7MB/s]
Some weights of the model checkpoint at bert-large-uncased were not used when initializing BertModel: ['cls.predictions.transform.LayerNo
- This IS expected if you are initializing BertModel from the checkpoint of a model trained on another task or with another architecture
- This IS NOT expected if you are initializing BertModel from the checkpoint of a model that you expect to be exactly identical (initial:
Downloading (…)solve/main/vocab.txt: 100% 232k/232k [00:00<00:00, 1.92MB/s]
Downloading (…)okenizer_config.json: 100% 28.0/28.0 [00:00<00:00, 1.00kB/s]
Today we will be discussing debating about AMC engineering
college. For about two semesters during the COVID we had almost no colleges
and the staff are teaching online thing but we had a college design
and we were online promoted for two sems and we didn't focus on the part which was
meant to be learned by us during that period of time.
/usr/local/lib/python3.9/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10
```

GPT-2 is an abbreviation for 'Generative Pretrained Transformer 2.' The model is open source, and is prepared on over 1.5 billion boundaries to create the following succession of message for a given sentence.

System Implementation:

Designing a speech-to-text (STT) system involves several components and considerations. Here's a high-level overview of the system design for a basic STT system:

1)Audio Input:

- Microphone or audio file input.
- Preprocessing: Handle audio normalization, noise reduction, and filtering

2)Acoustic Feature Extraction:

- Convert the audio signal into a representation suitable for speech recognition.
- Extract features such as Mel-Frequency Cepstral Coefficients (MFCCs), filter banks, pitch, and energy.
- Apply techniques like windowing, Fourier Transform, and frame-based analysis to obtain time-varying features.

3)Acoustic Modeling:

- Utilize machine learning techniques, such as deep neural networks (DNN) or convolutional neural networks (CNN), to model the relationship between acoustic features and phonetic units.
- Train the model using a large dataset of audio recordings paired with their corresponding transcriptions.

4)Language Modeling:

- Utilize language models to capture the statistical properties of the spoken language.
- Incorporate linguistic knowledge, such as grammars or n-gram models, to improve.

5)Decoding and Alignment:

- Utilize algorithms like Hidden Markov Models (HMM) or Connectionist Temporal Classification (CTC) to decode the acoustic features into a sequence of phonetic units or words.
- Perform alignment between the acoustic features and the recognized units to handle temporal synchronization.

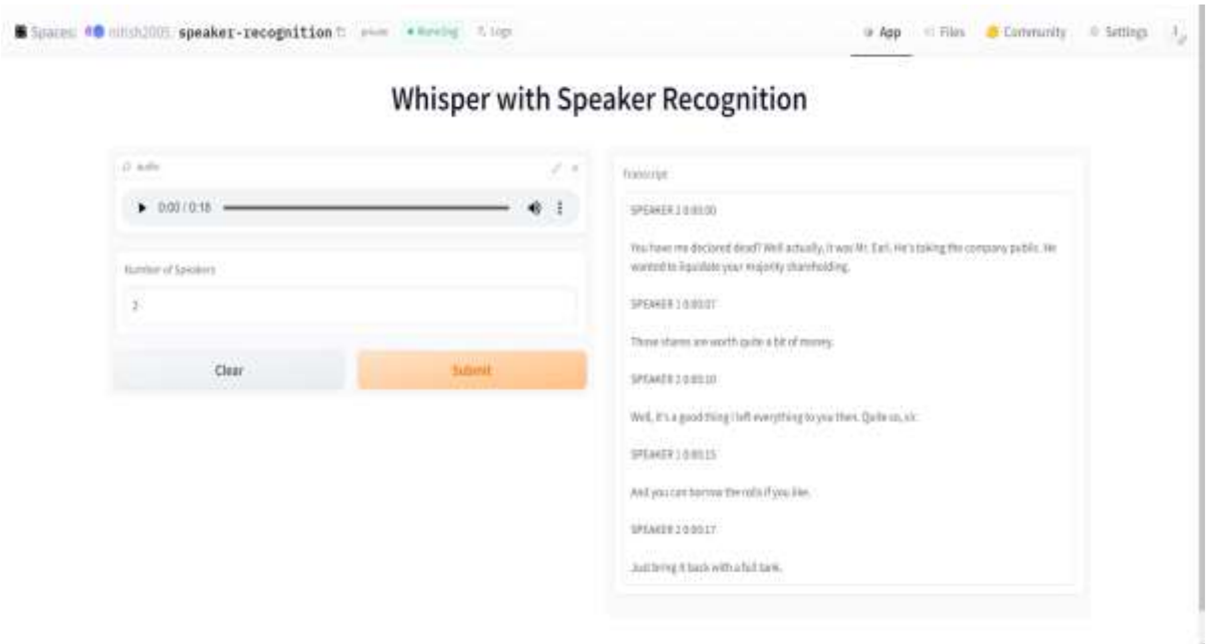
6)Post-processing:

- Apply post-processing techniques to improve the accuracy and readability of the recognized text.
- Handle tasks such as capitalization, punctuation restoration, homophone disambiguation, and error correction.

RESULTS:

The research paper presents detailed results and analysis of the performance of different speech recognition and summarization models. It highlights the strengths and weaknesses of each model, identifies the factors influencing their performance, and provides insights into the trade-offs between accuracy, speed, and summary quality.

The research paper presents detailed results and analysis of the performance of different speech recognition and summarization models. It highlights the strengths and weaknesses of each model, identifies the factors influencing their performance, and provides insights into the trade-offs between accuracy, speed, and summary quality.



CONCLUSION:

The research paper concludes by summarizing the findings of the comparative study and discussing the implications for future research and applications in speech recognition and summarization. It highlights the importance of selecting appropriate models and techniques based on specific use cases and provides recommendations for improving the accuracy and efficiency of speech recognition and summarization systems. Overall, this research paper contributes to the existing body of knowledge by providing a comprehensive analysis of speech recognition and summarization models and techniques. The findings and insights can guide researchers and practitioners in selecting and developing effective solutions for speech recognition and summarization tasks.An automatic notes generator is a useful tool for creating concise, organized summaries of audio or text content. It can save time and effort by automatically identifying key points and main ideas, and organizing them into coherent notes. Automatic notes generators can be particularly useful for students and professionals who need to take notes from lectures, meetings, or other spoken or written materials. They can also be useful for language learners, as they can help to extract the most important information from listening or reading materials.

There are a number of different techniques and approaches that can be used to build an automatic notes generator, including speech recognition, natural language processing, and text summarization. The specific method or combination of methods used will depend on the specific requirements and goals of the application. Notes is an important part of daily meeting, classes and seminars, it is a must to write notes during attending the meeting and keeping track of all points discussed and express new ideas and suggestion in meeting. But sometimes it is difficult to write notes simultaneously listening to points and the main concept is to keep the points, agenda of meeting to be kept safely. There are indeed several challenges and limitations to speech-to-text conversion technology, which is a form of natural language processing that involves transcribing spoken language into written text.