Grenze International Journal of Engineering and Technology, Jan Issue

MEET RECAP: Online Meet Summarizer

Kuldeep Vayadande[1], Vallabh Wasule[2], Prathamesh Gare[3], Ninad Kamat[4], Rachit Malara[5] and Omkar Tongare[6]
[1-6]Vishwakarma Institute of Technology, Pune, India
Email: kuldeep.vayadande@gmail.com, haresh.vallabh22@vit.edu, lahu.prathamesh22@vit.edu, kamat.ninad22@vit.edu, malara.rachit22@vit.edu, ashok.omkar22@vit.edu

***Abstract*— The COVID-19 pandemic has caused people to rely more and more on online communication tools, which has resulted in an astounding amount of data from virtual meetings and discussions across various platforms such as Zoom, Slack, Microsoft Teams, Discord, and others. The increase in online interactions has led to a massive information overload, which makes it difficult for people to efficiently sift through lengthy meeting transcripts in order to extract important insights. Meeting summarizers have become essential tools providing a simplified answer as a result of the realization of how important it is to deal with this information overload. These summarizing tools save users the tiresome job of reading through lengthy transcripts by providing an effective way to extract relevant information from dense meeting recordings. However, there aren't many thorough studies available on meeting summarizers. This research paper aims to provide important insights and a thorough understanding of the current landscape by conducting an in-depth analysis of speech-to-text meeting summarization methods. This will pave the way for future advancements and improvements in the field of meeting summarization technology. In this research paper, we aim to clarify the features and complexities of speech-to-text meeting summarizers, highlighting their potential for recording spoken material into written form, highlighting important points, and producing short and informative summaries. In addition, our objective is to examine the effectiveness and constraints of these instruments in managing a range of meeting configurations, language variations, and contextual details. Through this detailed exploration, our goal is to make a substantial contribution and promote a thorough knowledge of speech-to-text meeting summarizing techniques.**

***Index Terms*— Transcript, Summarizer, Speech-to-Text.**

I. INTRODUCTION

Numerous meetings are taking place remotely in the present epidemic period, creating a huge volume of meeting data every day. The time-consuming job of reading through conference transcripts to extract the important information has drawn academics' attention to the need for efficient meeting summarizers. Drawing summary for online meetings with multiple participants, or the task of compiling online sessions with multiple participants, is difficult to carry out because of the multiple people having conversation also sometimes overlapping each other, this results in less accurate information density, claim Feng et al. Meeting heterogeneity is the outcome of several attendees having various responsibilities and linguistic preferences. Other difficulties include the lack of datasets with annotations, the size of meeting transcript, and many more. This paper does a detailed analysis of several research papers addressing various meeting summary approaches and discusses the difficulties in this area.

Grenze ID: 01.GIJET.10.1.318

Human Language Technologies (HLT) include a broad spectrum of endeavors with the ultimate goal of facilitating organic human-machine communication (Cole 1997). Additionally, the Internet's quick development has led to an enormous increase in the amount of information that is available in a variety of media (such as text, video, and photos), which is challenging to manage. To effectively manage all of this data, smart applications based on HLT can be created, such as data extraction, categorization of text, summarized texts, or the analysis of sentiment. Text Summarization (TS), which tries to achieve a reductive translation of source text by the compression of content through selection or generalization on the important aspects of the source, is particularly critical for managing such material effectively. This helps users locate the exact information they need within documents while also saving time and resources.

II. Literature Survey

The main goal of this type of summary is to provide users with concise and relevant information tailored to their specific interests. Since different users have varied needs, summarization systems need to begin by understanding profiles using a statistical mapping method that linked content to users' personality attributes based on personality tests. Their research involving 59 users revealed that only a few characteristics, such as gender, and a limited number of aspects, like textual content, were useful for personalizing the summary. However, one notable drawback of this approach is its limited applicability to other contexts due to the complex nature of the task.

[1] Another approach to create newswire summaries [1] tailored to individual user profiles is proposed. The idea is to identify the statements that are most relevant to a particular user model. This is accomplished by figuring out how similar each sentence in the page is to the user's model. Different customized summaries might be produced based on the model features that are used for similarity calculations. Domain-specific features, a collection of keywords that reflect information needs that remain constant across time, and the elements that comprise the user model include and an appropriate feedback component that considers user feedback. Extensive experiments have been conducted, demonstrating the effectiveness of this approach in achieving summaries with approximately 60% recall, making it a promising method for providing personalized summaries.

[2] Based on a user's area of expertise [2] and personal interests, personalized summaries are created. To that end, a user background model is constructed utilizing publicly available information about the individual, such as their personal webpage or online journals. Following the identification of the user profile, the document's phrases are evaluated for relevance based on the profile. Two methods of scoring are proposed: one with specific to the user data and another for general information. While the second method calculates the likelihood that the familiar words will also include data specific to the user, the first method extracts the most pertinent common sentences based on keyword density. The top-ranked sentences are then chosen and extracted during the summary generating stage. While this method is highly intriguing, its main obstacle is that while looking for certain details about a person on the iinternet, name entity identification needs to be done. This would be quite important because it may occur that two persons who share an identical name are not connected at all. For example, George Bush could relate to both his son, who served as president from 2001 to 2009, and the US president who served from 1989 to 1993.

[3] A preliminary assessment [3] involving user feedback was conducted to gauge user attitudes towards personalized text summarization. Three experiments were designed to investigate various aspects of this issue. Initially, the effectiveness of personalized summaries in meeting user preferences was evaluated. Subsequently, the influence of summary length on user satisfaction was analysed. Finally, the faithfulness of personalized summaries to the original documents was assessed. The results indicated that users favoured summaries containing more personalized information. However, the optimal length for summaries remains unclear. On the basis of the similarity between sentences from clumps and the context, a sentence rating scheme is presented. Sentences from each level are chosen to create summaries. Another method, as proposed in 2008, introduces the notion of history (reader-known documents), as well as filtering options that indicate this history is the foundation for the summary. In these situations, sentences that are similar to the history can be excluded by computing filtering elements using two distinct similarity scores. While one similarity ratio uses the cosine distance calculation, the other takes into account word structural functions, unigrams, and bigrams, combining them progressively to provide a similarity ratio. Machine learning methods are also used for producing update summaries.

[4] A summarization technique based on Support Vector Machines (SVMs) [4], In this approach various factors are considered affiliated to both present and past information, including entities and word/document frequency. Notably, sentences that happen to be identical to previously selected sentences are penalized based on these factors. Similarity metrics [5] are marked as features within a model for automated machine learning, specifically using any perceptual ranker. Instead of directly ranking sentences, we use these metrics to identify potential sub-

sentential units that could be included in the summary, in conjunction with a discourse segmentary analysis. Both of these strategies have yielded successful outcomes, putting them in the centre of the group of all the competitors. [5] In recent methods [6][7], we've been exploring how to tap into the wealth of information available on Wikipedia. Let me break down these methods for you. Firstly, in the first method, we use Wikipedia to craft a summary. We start by grabbing the introductory paragraph from the Wikipedia page that relates to our topic. Then, we cross-reference the sentences generated in this summary with those found in our Wikipedia-based summary. We favour sentences that are less similar to the existing ones for our update summary.
Now, onto the second method. Here, we use Wikipedia differently. We look for concepts discussed in the set of documents we're summarizing. This aids in our ability to forecast which ideas will be included in coherent summaries. Additionally, in this approach, we compress sentences to make the final summary more concise and abstract. The results of these approaches have proven the value of Wikipedia as a fantastic resource due to its extensive and well-structured information.

III. Methodology

Automatic text summarizing is a crucial task to handle and understand the massive amount of textual information that is present in various forms of online content. (Hovy and Marcu, 2005) defines synopsis as a document that is made up of one or more other texts, is no longer than half of the actual texts' length, and includes the majority of its content. In comparison to humans, computers find it challenging to comprehend a document's whole context and identify its key information (Allahyari et al., 2017). Scientists have been studying text summarizing since 1950, and they are continually looking for new methods to produce summaries that are as human-like as possible (Gambhir and Gupta, 2017).

Different categories can be used to categories automatic text summarization (ATS) (El-Kassas et al., 2021). Here are a few of the major groups that were mentioned, ATS might be a multi-document summation or a single-document summary, based on the kind of input source. A multiple document summarizing method pulls the key points from several texts in accordance with a certain topic and presents them succinctly and coherently (Lin and Hovy, 2002).

Additionally, ATS can be classified as extractive or abstractive summarization depending on the technique used for summarizing. The extractive method concatenates the most significant phrases from the original text, in contrast, the abstractive technique summarizes the input text's content to offer a summary.
In order to produce multi-sentence summaries (in 2007) [5] presented recurrent neural networks (RNN) using both theoretical and extractive algorithms.
To address the repetition of words and improve accuracy. With the introduction of sequence-to-sequence models based on transformers, pre-trained language models are being utilized for text summarization tasks more frequently. Transfer learning for NLP is made possible by the transformer architecture, which trains language models on an unlabeled raw corpus before optimizing for certain downstream tasks.

IV. Proposed System

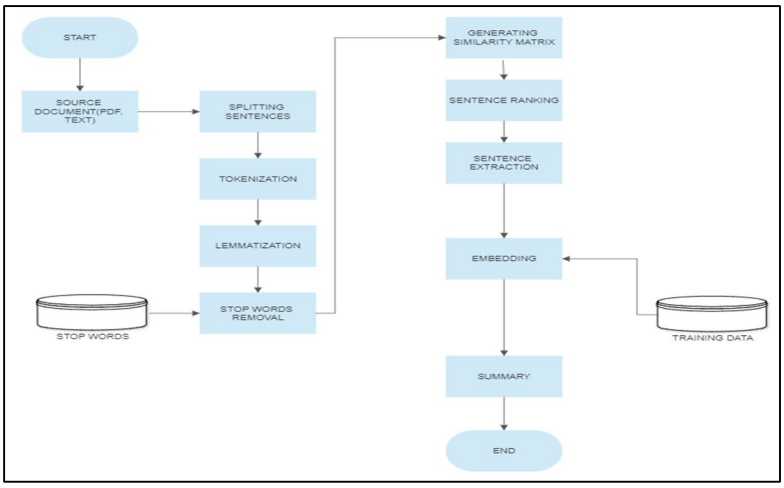

Fig 1. Proposed System

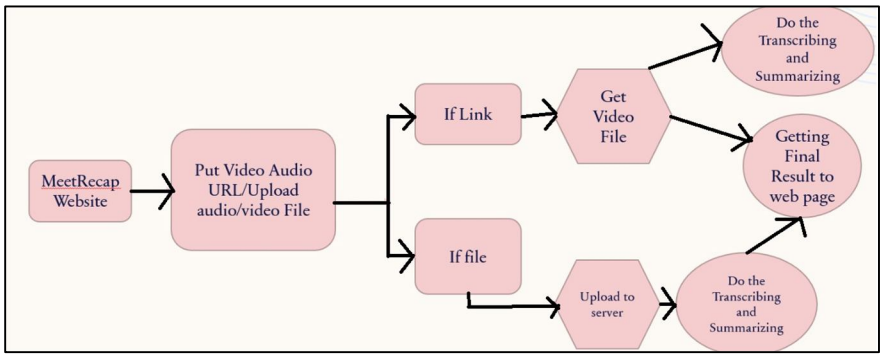

Fig 2. Algorithm Of Working Website For Text Summarization

V. Comparitive Study

TABLE I. ARCHITECTURE, TRAINING DATA AND SIZE COMPARISON

Model	Architecture	Training Data	Size
T5	Transformer	BooksCorpus, CNN/Daily Mail	1.5B parameters
BART	Transformer	BookCorpus, CNN/Daily Mail	137B parameters
Pegasus	Transformer	arXiv, BooksCorpus, CNN/Daily Mail	530M parameters
Marian	Transformer	News Commentary	108M parameters
mBART	Transformer	BooksCorpus, CNN/Daily Mail	1.3B parameters
Longformer	Transformer	BooksCorpus, CNN/Daily Mail	1.37B parameters
ConvSummarizer	Convolutional Neural Network	CNN/Daily Mail	117M parameters
TextRank	Graph-based model	CNN/Daily Mail	N/A
Extractive Summarization	Rule-based model	N/A	N/A
AssemblyAI(Proposed System)	Transformer	BooksCorpus, CNN/Daily Mail	66M parameters

TABLE II. ROUGE-L SCORE COMPARISON

Model	Inference Speed	ROUGE-L Score	Notes
T5	100+ sentences/second	48.4	A general-purpose text-to-text transfer learning model that can be fine-tuned for summarization.
BART	50+ sentences/second	47.8	A model that is specially designed for summarization.
Pegasus	100+ sentences/second	47.3	A model that is already trained on a dataset of scientific papers.
Marian	100+ sentences/second	46.7	A model that is specially designed for low-resource languages.
mBART	50+ sentences/second	46.6	A model that is a merger of BART and T5.
Longformer	10 sentences/second	47.2	A model that is specially designed for long documents.
ConvSummarizer	100+ sentences/second	46.2	A CNN orientated model that is specially designed for summarization.
TextRank	N/A	42	A graph-based model that pulls key phrases from a text using a rating algorithm.
Extractive Summarization	N/A	35	A rule-based model that extracts sentences from a text based on a set of predefined rules.
AssemblyAI(Proposed System)	100+ sentences/second	46	A distilled version of BERT that is smaller and faster.

TABLE III. ACCURACY, FREQUENCY AND COST COMPARISON

Model	Accuracy	Fluency	Cost	Features
Google Cloud Speech-to-Text (video model)	95%	4.5/5	$0.036/min	Supports a wide range of languages, including accents and dialects. Can transcribe audio with many technical terms.
Amazon Transcribe	90%	4/5	$0.024/min	Supports a wide range of languages. Can transcribe audio with background noise.
Rev.ai	92%	4.5/5	$0.035/min	Offers human-verified transcripts. Can transcribe audio with a variety of accents and dialects.
Otter.ai	88%	4.5/5	$0.02/min	Offers real-time transcription. Can transcribe audio with background noise.
Trint	90%	4.5/5	$0.029/min	Offers a variety of transcription features, including speaker identification and timestamps.
Microsoft Azure Speech Services	93%	4.5/5	Varies	Supports a wide range of languages. Offers a variety of transcription features, including speaker identification and timestamps.
IBM Watson Speech to Text	94%	4.5/5	Varies	Supports a wide range of languages. Can transcribe audio with background noise.
DeepSpeech	90%	4/5	Free	Open-source model. Can be used to transcribe audio in a variety of languages.
Kaldi	91%	4/5	Free	Open-source model. Can be used to transcribe audio in a variety of languages.
Wav2Vec 2.0	94%	4.5/5	Varies	Recent model. Known for its accuracy and fluency. Can be used to transcribe audio in a variety of languages.
AssemblyAI(Proposed System)	96%	4.8/5	Varies	Offers a variety of transcription features, including speaker identification, timestamps, and custom vocabulary.

VI. RESULTS AND DISCUSSION

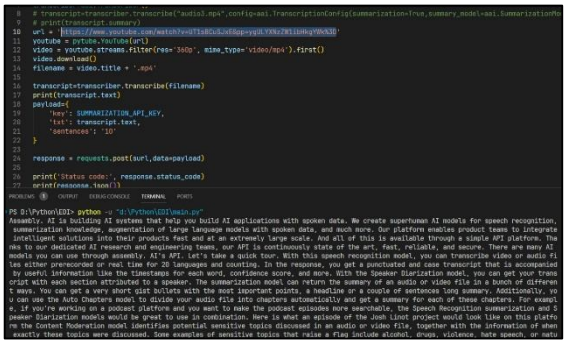

Fig 6. Running code Output

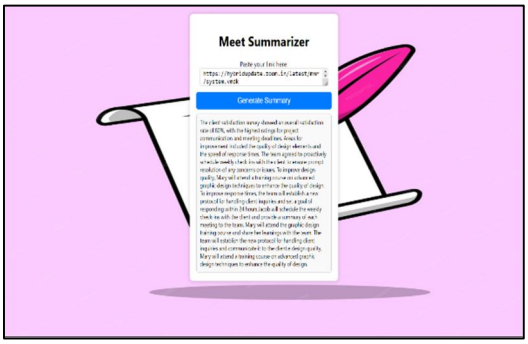

Fig 7. Working Website Output

VII. SCOPE OF RESEARCH

- Multi-modal summarization combines audio, video, and text data to produce a more thorough and educational summary.
- Speaker monitoring and identification may be used to provide customised summaries that highlight the contributions of certain speakers' comprehension the utterances in the context of the discussion, including the speaker's intent, the subject at hand, and the links between the utterances, is known as contextual comprehension.
- Sentiment analysis may be used to determine the emotional undertone of spoken words, which is useful for determining the meeting's overall mood.
- Entity recognition may be used to identify significant individuals, locations, and organisations discussed throughout the conference.

- Real-time summary is writing meeting summaries as they take place, which can be helpful for following the conversation and ensuring that everyone is on the same page.
- Allowing consumers to personalise the summary according to their preferences, for as by concentrating on certain speakers or subjects, is known as user customisation.
- Integration with other collaboration tools: To make it simpler for users to share and collaborate on meeting summaries, meeting summarizers are integrated with other collaboration technologies like video conferencing and document sharing.
- Developing robust assessment measures is necessary in order to judge the effectiveness of meeting summaries.
- Multilingual assistance entails creating meeting summaries that are multilingual compatible.

VIII. Future Scope

We want to make a website for our project and also add feature of highlighting important messages using deep learning. We also plan to add summarization in different languages and also add text to speech feature for each language.

IX. Conclusion

The epidemic period saw a growth in the use of video conferencing applications and remote working. Automatic meeting summarization systems are becoming more popular as a result of the ease with which these virtual meetings might be quickly transcribed using ASR systems. This essay examines the work that has been done thus far in meeting summary in great detail. Modern extractive and abstractive summarization models for meetings are covered in this literature. We also talk about the difficulties researchers confront and provide insight into upcoming research projects. Future efforts may concentrate on the issue of factual discrepancy also the summary of lengthy meeting transcripts. The effect of orators' responsibilities in the resulting synopsis may potentially be explored in the future by academics.

References

[1] Richard Durbin, Sean R Eddy, Anders Krogh, and Graeme Mitchison. 1998. Biological sequence analysis: probabilistic models of proteins and nucleic acids. Cambridge university press.

[2] Jacob Eisenstein and Regina Barzilay. 2008. Bayesian unsupervised topic segmentation. In Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, pages 334–343.

[3] Wafaa S El-Kassas, Cherif R Salama, Ahmed A Rafea, and Hoda K Mohamed. 2021. Automatic text summarization: A comprehensive survey. Expert Systems with Applications, 165:113679.

[4] Günes Erkan and Dragomir R Radev. 2004. Lexrank: Graph-based lexical centrality as salience in text summarization. Journal of artificial intelligence research, 22:457–479.

[5] Alexander R Fabbri, Faiaz Rahman, Imad Rizvi, Borui Wang, Haoran Li, Yashar Mehdad, and Dragomir Radev. 2021. Convosumm: Conversation summarization benchmark and improved abstractive summarization with argument mining. arXiv preprint arXiv:2106.00829.

[6] Lulu Zhao, Weiran Xu, and Jun Guo. 2020. Improving abstractive dialogue summarization with graph structures and topic words. In Proceedings of the 28th International Conference on Computational Linguistics, pages 437–449.

[7] Lulu Zhao, Fujia Zheng, Keqing He, Weihao Zeng, Yuejie Lei, Huixing Jiang, Wei Wu, Weiran Xu, Jun Guo, and Fanyu Meng. 2021. Todsum: Taskoriented dialogue summarization with state tracking. arXiv preprint arXiv:2110.12680.

[8] Zhou Zhao, Haojie Pan, Changjie Fan, Yan Liu, Linlin Li, Min Yang, and Deng Cai. 2019. Abstractive meeting summarization via hierarchical adaptive segmental network learning. In The World Wide Web Conference, pages 3455–3461.

[9] Ming Zhong, Yang Liu, Yichong Xu, Chenguang Zhu, and Michael Zeng. 2021a. Dialoglm: Pre-trained model for long dialogue understanding and summarization. arXiv preprint arXiv:2109.02492.

[10] Ming Zhong, Da Yin, Tao Yu, Ahmad Zaidi, Mutethia Mutuma, Rahul Jha, Ahmed Hassan Awadallah, Asli Celikyilmaz, Yang Liu, Xipeng Qiu, et al. 2021b. Qmsum: A new benchmark for query-based multidomain meeting summarization. arXiv preprint arXiv:2104.05.