iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VI **Month of publication:** June 2023

DOI: https://doi.org/10.22214/ijraset.2023.53578

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VI Jun 2023- Available at www.ijraset.com

Meeting Summarizer Using Natural Language Processing

Vishnuprasad[1], Paul Martin[2], Salman Nazeer[3], Prof. Vydehi. K[4]
Adi Shankara Institute of Engineering and Technology, Kalady, Kerala

Abstract: ***Meeting transcripts produced by tools like Microsoft Teams and Google Meet, are useful for recording discussions and decisions made during meetings. However, reading through long transcripts can be time-consuming and may not always be the most efficient way to understand the key points and conclusions of a meeting. Meeting summarization is a subfield of natural language processing that can extract important information from meeting transcripts and generate a concise summary. This summary can be used to quickly understand the key points and conclusions of the meeting, and can be especially useful for stakeholders who were not able to attend the meeting in person. Several natural language processing techniques can be used to create summaries of meeting transcripts, such as the term frequency-inverse document frequency (TF-IDF) method, PageRank algorithm, Named Entity Recognition, Topic Modeling and specific summarization algorithms. Each technique has its own advantages and limitations, and the appropriate technique can be chosen based on the specific needs and requirements of the organization, such as accuracy, efficiency, and customization.***
Keywords: TFIDF, PageRank algorithm, Glove embedding

I. INTRODUCTION

Document summarization is the process of creating a shorter version of a document that captures the most important information from the original document. This can be useful for a variety of purposes, such as quickly conveying the main points of a document to a reader who may not have time to read the entire thing, or for organizing and indexing large amounts of documents.

There are two main approaches to document summarization: extractive and abstractive. Extractive summarization techniques use mathematical and statistical methods to identify the most important words, phrases, or sentences in the original document and include them in the summary. These techniques do not generate new text, but rather extract and compile the most important information from the original document. One example of an extractive summarization technique is the singular value decomposition

(SVD) method. SVD is a mathematical technique that decomposes a matrix into its constituent parts, and it can be used to identify the most important words or phrases in a document based on their frequency and co-occurrence within the document. Other extractive techniques include keyword extraction, which involves identifying the most important words in a document based on their frequency or importance, and sentence extraction, which involves selecting the most important sentences from the original document. Abstractive summarization techniques, on the other hand, use language semantics and natural language generation (NLG) techniques to generate new text and summaries based on the content of the original document. These techniques aim to capture the main ideas and concepts of the original document, rather than just the specific words and phrases used.

One example of an abstractive summarization technique is the use of a knowledge base, which is a collection of facts and information about a specific domain. A summarization system that uses a knowledge base can generate a summary by selecting the most relevant facts from the knowledge base and combining them into a coherent summary. Another example of an abstractive technique is the use of semantic representations, which are structured representations of the meaning of words and phrases in a document. A summarization system that uses semantic representations can generate a summary by identifying the most important concepts in a document and expressing them in a way that is easy for a reader to understand.

There are a number of factors to consider when choosing an approach to document summarization. Extractive techniques are generally simpler and faster to implement, but they may not always capture the main ideas of the original document as accurately as abstractive techniques. Abstractive techniques, on the other hand, can produce more coherent and accurate summaries, but they may require more data and computation resources.

In this paper, the focus is on summarizing transcripts of meetings, such as those from Microsoft Teams. Meeting transcripts can be particularly long and complex, as they often include multiple speakers discussing a variety of topics. As a result, it can be helpful to generate a summary of a meeting transcript to quickly convey the main points and decisions made during the meeting.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VI Jun 2023- Available at www.ijraset.com

There are a number of techniques that can be used to summarize meeting transcripts. One such technique is TextRank, which is an extractive summarization technique that uses a graph-based algorithm to identify the most important words and phrases in a document. TextRank works by creating a graph of the words in a document, with edges between words that are frequently co-occurring. The most important words are then identified based on their centrality in the graph.

Another technique that can be used to summarize meeting transcripts is term frequency-inverse document frequency (TF-IDF). TF-IDF is a statistical measure that reflects the importance of a word in a document based on its frequency within the document and its rarity in a collection of documents. Another factor to consider when summarizing meeting transcripts is the level of detail that is desired in the summary. Some summaries may be intended to capture only the most important points and decisions made during the meeting, while others may require a more detailed summary that includes all of the main points and discussions.There are also a number of tools and software platforms available that can assist with the summarization of meeting transcripts. These tools often use a combination of extractive and abstractive techniques to produce summaries, and they may also allow users to customize the level of detail in the summary.

One example of a tool that can be used to summarize meeting transcripts is Microsoft Teams, which is a platform for online meetings and collaboration. Microsoft Teams includes a feature called "Meeting Notes" that allows users to take notes during a meeting and generate a summary of the meeting afterwards. The Meeting Notes feature uses a combination of extractive and abstractive techniques to produce the summary, and it also allows users to customize the level of detail in the summary.

In conclusion, document summarization is a useful tool for quickly conveying the main points of a document to a reader. There are two main approaches to document summarization: extractive and abstractive, with each having its own strengths and limitations. Meeting transcripts, in particular, can be long and complex, and tools such as Microsoft Teams can assist with the summarization of these transcript.

II. PROPOSED SYSTEM

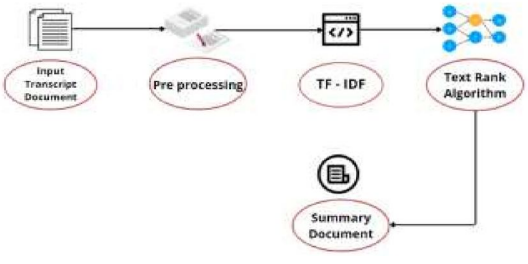
The proposed architecture for the Meeting Summarizer utilizing NLP involves taking a transcript document as input and conducting various operations to generate a summary document. Specifically designed for Microsoft Teams transcripts, the architecture begins by removing time stamps and associating each sentence with its corresponding speaker. The text is then split into individual sentences for further processing.

During pre-processing, the text undergoes standardization and stop words are eliminated. The document term feature matrix is constructed using TF-IDF (Term Frequency-Inverse Document Frequency). Term Frequency denotes the frequency of a word in a document, while Inverse Document Frequency indicates the rarity or commonality of a word across all documents. TF-IDF scores are obtained by multiplying these two factors, determining the significance of a word within a document.

Next, a document similarity matrix is created by multiplying the document term feature matrix with its transpose. This matrix captures the similarities between each pair of sentences. A document similarity graph is then generated, with sentences as vertices and the similarity scores as weight or score coefficients.

The PageRank algorithm, which assigns scores based on the importance of nodes in a network, is applied to the document similarity graph. It calculates scores for each sentence, indicating their relative significance within the overall network.

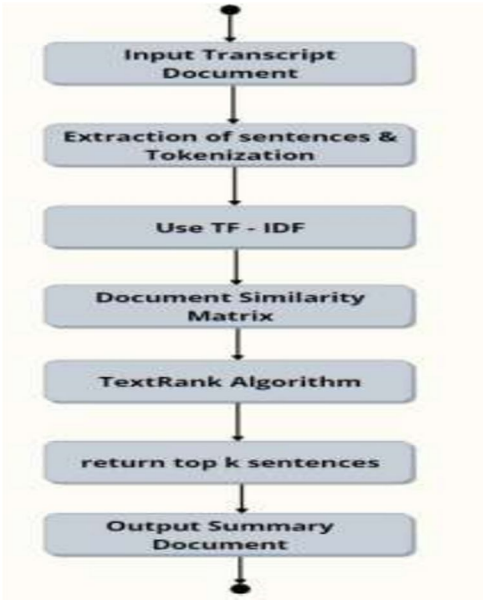
Finally, the sentences are ranked based on their scores, and the top sentences are selected to form the output summary document. These sentences represent the most relevant and informative portions of the original meeting transcript.



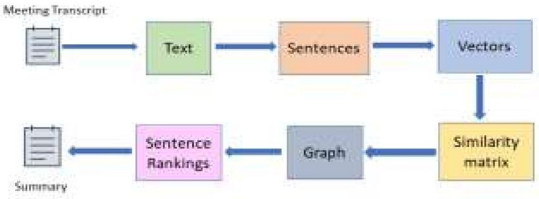
The flowchart begins by taking the transcript document as input. It then proceeds with extracting sentences and tokenizing them for further use. The TF-IDF technique is applied to generate a document similarity matrix. This matrix is then utilized by the TextRank Algorithm, which assigns rankings to each sentence. Finally, the top-ranked sentences are selected to create the output summary document.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VI Jun 2023- Available at www.ijraset.com



The transcript is tokenized using the nltk method, which splits it into individual words. Stop words are then removed, and stemming is applied to reduce words to their root form. The resulting words are transformed into vectors using numpy. These vectors represent normalized sentences. The TfidfVectorizer is imported from the sklearn module to create a document term frequency matrix. By taking the transpose of this matrix, the inverse document frequency is obtained. The number of sentences in the output summary is determined based on the number of pre-processed sentences in the input. If there are more than 30 sentences, the output will contain 20% of the total input sentences. Otherwise, 30% is considered. A similarity matrix is generated by multiplying the document term frequency matrix with its inverse. The PageRank algorithm is applied to the similarity graph, assigning rankings to each sentence based on their importance. Finally, the top-ranked sentences, determined by the rankings, are provided as the output.



III. TECHNOLOGIES USED

A. TEXTRANK Algorithm

TextRank is a graph-based ranking algorithm that was inspired by Google's PageRank algorithm. It can be used to identify the most relevant sentences in a text and to extract keywords. TextRank has a number of applications in natural language processing, such as keyword extraction, automatic text summarization, and phrase ranking. To identify the most relevant sentences in a text using TextRank, the algorithm creates a graph with the vertices representing each phrase in the document and the edges linking sentences based on content overlap. This can be done by calculating the number of words shared by two sentences. The sentences are then fed into the PageRank algorithm, which determines the importance of each sentence based on the network of sentences. The most important sentences are selected and used to create a summary of the text.

TextRank can also be used to extract keywords from a text by creating a word network that identifies which words are connected to one another. If two words appear frequently next to each other in the text, a link is created between them, and the link is given more weight if the words appear even more frequently together. The PageRank algorithm is applied to the generated network to determine the significance of each word. The top third of the most significant words are selected and used to create a keywords table by grouping together relevant terms that appear in the text in close proximity. TextRank architecture, which involves creating a graph of the text, applying the PageRank algorithm to the graph, and using the resulting scores to rank the phrases or words in the text. The most important phrases or words are then selected and used for the desired task, such as summarization or keyword extraction.

©IJRASET: All Rights are Reserved

B. Term Frequency – Inverse Document Frequency (TF-IDF)

The term frequency-inverse document frequency (TF-IDF) technique is a method for evaluating the relevance of a word to a document in a collection of documents. It is widely used in information retrieval and natural language processing tasks, such as document search and classification. The TF-IDF measure is based on the idea that a word that occurs frequently in a document is likely to be important to the meaning of the document, but a word that occurs frequently in many documents is not as useful for determining the relevance of a document. To balance these two factors, the TF-IDF measure combines the term frequency, which is the raw count of the number of times a word appears in a document, and the inverse document frequency, which is a measure of how common or rare the word is in a collection of documents. To calculate the term frequency of a word, the system simply counts the number of times the word appears in the document. To calculate the inverse document frequency, the system divides the total number of documents in the collection by the number of documents that contain the word and takes the logarithm of the result. The inverse document frequency is used to downweight the importance of common words that occur in many documents. The TF-IDF weight of a word or phrase is obtained by multiplying the term frequency and inverse document frequency. The larger the weight, the more unusual the word or phrase is in the document, and the more likely it is to be relevant to the meaning of the document.

C. GLOVE Embedding

GloVe, which stands for Global Vectors for Word Representation, is an unsupervised learning algorithm used to generate word embeddings. Word embeddings are dense vector representations of words, where each word is mapped to a high-dimensional vector. These vectors capture semantic and syntactic relationships between words, allowing for better understanding of their meanings. The GloVe algorithm is based on the idea that word meaning can be inferred from the co-occurrence statistics of words in a large corpus of text. It analyzes the word co-occurrence matrix, which counts how often words appear together in a given context window. By factorizing this matrix, GloVe learns the embeddings that best capture the statistical patterns of word co-occurrence. The resulting word embeddings encode the semantic relationships between words. Similar words are represented by vectors that are close together in the embedding space, while dissimilar words are represented by vectors that are far apart. These embeddings can be used as features in various natural language processing (NLP) tasks, such as sentiment analysis, named entity recognition, machine translation, and text classification. GloVe embeddings have gained popularity due to their effectiveness in capturing semantic information and their ability to handle out-of-vocabulary words. They have been widely used in both research and industry applications to enhance the performance of NLP models.

IV. RELATED WORKS

1) Comparative analysis of NLP models for Google Meet Transcript summarization In paper [1] “Yash Agrawal, Atul Thakre, Tejas Tapas, Ayush Kedia, YashTelkhade and Vasundhara Rathod, Comparative analysis of NLP models for Google Meet Transcript summarization, EasyChair Preprint no. 5404, 2021”. the author conveys that Manual transcription and summarization is a cumbersome process necessitating the development of an efficient automatic text summarization technique. In this study, a Chrome extension is used for making the process of transcription hasslefree. It uses the text summarization technique to generate concise and succinct matter. Also, the tool is accessorized using Google Translation, to convert the processed text into users' desired language. This paper illustrates how captions can be traced from the online meetings, corresponding to which, the meeting transcript is sent to the backend where it is summarized using an NLP model. It also walks through three different NLP models and presents a comparative study among them. The NLTK model utilizes the sentence ranking technique for extractive summarization. Word Embedding model uses pre-trained Glove Embeddings for extractive summarization. The T5 model performs abstractive summarization using transformer architecture. The working of the model is tested over meeting texts taken from various sources and results show that the NLTK model has an edge over the Word Embedding model based on ROUGE-1, ROUGE-2, and ROUGE-L scores. However, our analysis finds that T5 is generating a more concise summary.

2) Automatic Meeting transcription and summarization for in-person conversations. In paper [2] “ Yuanfeng Song, Di Jiang ,Xuefang Zhao ,Xiaoling Huang, QianXu, Raymond Chi-Wing Wong, Qiang Yang, SmartMeeting: Automatic Meeting transcription and summarization for in-person conversations. ACM International Conference on Multimedia 2021, Pages 2777-2779.” The author conveys that meetings are a necessary part of the operations of any institution, whether they are held online or in-person. However, meeting transcription and summarization are always painful requirements since they involve tedious human effort. This drives the need for automatic meeting transcription and summarization (AMTS) systems. A successful AMTS system relies on systematic integration of multiple natural language processing (NLP) techniques, such as automatic speech recognition, speaker identification, and meeting summarization, which are traditionally developed separately and

validated offline with standard datasets. In this demonstration, we provide a novel productive meeting tool named SmartMeeting, which enables users to automatically record, transcribe, summarize, and manage the information in an in-person meeting. SmartMeeting transcribes every word on the fly, enriches the transcript with speaker identification and voice separation, and extracts essential decisions and crucial insights automatically. In our demonstration, the audience can experience the great potential of the state-of-the-art NLP techniques in this real-life application.

3) Extractive summarization of call transcripts In paper [3] " Pratik K. Biswas, AleksandrIakubovich, Extractive summarization of call transcripts, arXiv Preprint arXiv: 2103.10599, 19 March, 2021. " the author conveys that text summarization is the process of extracting the most important information from the text and presenting it concisely in fewer sentences. Call transcript is a text that involves textual description of a phone conversation between a customer (caller) and agent(s) (customer representatives). This paper presents an indigenously developed method that combines topic modeling and sentence selection with punctuation restoration in condensing ill-punctuated or un-punctuated call transcripts to produce summaries that are more readable. Extensive testing, evaluation and comparisons have demonstrated the efficacy of this summarizer for call transcript summarization.

4) A survey of text summarization Techniques. In paper [4] "Nenkova A., McKeown K. (2019) ,A survey of text summarization Techniques. In: Aggarwal C., Zhai C. (eds) Mining Text Data. Springer, Boston, MA. " the author conveys that Numerous approaches for identifying important content for automatic text summarization have been developed to date. Topic representation approaches first derive an intermediate representation of the text that captures the topics discussed in the input. Based on these representations of topics, sentences in the input document are scored for importance. In contrast, in indicator representation approaches, the text is represented by a diverse set of possible indicators of importance which do not aim at discovering topicality. These indicators are combined, very often using machine learning techniques, to score the importance of each sentence. Finally, a summary is produced by selecting sentences in a greedy approach, choosing the sentences that will go in the summary one by one, or globally optimizing the selection, choosing the best set of sentences to form a summary. In this chapter we give a broad overview of existing approaches based on these distinctions, with particular attention on how representation, sentence scoring or summary selection strategies alter the overall performance of the summarizer. We also point out some of the peculiarities of the task of summarization which have posed challenges to machine learning approaches for the problem, and some of the suggested solutions.

5) Unsupervised paradigm for information extraction from transcripts using BERT. In paper [5] "Aravind Chandramouli, SiddharthShukla, Neeti Nair, ShivenPurohit, ShubhamPandey and MuraliMohana Krishna Dandu, Unsupervised paradigm for information extraction from transcripts using BERT, arXiv Preprint arXiv:2110.00949,13 september 2021. " author conveys that audio call transcripts are one of the valuable sources of information for multiple downstream use cases such as understanding the voice of the customer and analysing agent performance. However, these transcripts are noisy in nature and in an industry setting, getting tagged ground truth data is a challenge. In this paper, we present a solution implemented in the industry using BERT Language Models as part of our pipeline to extract key topics and multiple open intents discussed in the call. Another problem statement we looked at was the automatic tagging of transcripts into predefined categories, which traditionally is solved using a supervised approach. To overcome the lack of tagged data, all our proposed approaches use unsupervised methods to solve the outlined problems. We evaluate the results by quantitatively comparing the automatically extracted topics, intents and tagged categories with human tagged ground truth and by qualitatively measuring the valuable concepts and intents that are not present in the ground truth. We achieved near human accuracy in extraction of these topics and intents using our novel approach.

6) Research on News keyword extraction Technology based on tf-idf and textrank In paper [6] "Yao, L., Pengzhou, Z., & Chi, Z. Research on News keyword extraction Technology based on tf-idf and textrank, IEEE/ACIS 18th International conference on computer and information science (ICIS), June 2019." According to the author, with the rapid development of information technology and the widespread use of the Internet, the Internet, as an information carrier, has gradually replaced the traditional media such as newspapers and television, and become the main channel for people to obtain information. This paper takes English news text as the research object of keyword extraction method. We combine TF-IDF and the TextRank algorithm to extract keywords from text by constructing word graph models, counting word frequency and inverse document frequency, and considering the weight of the positioning of headlines. A large number of experiments have been carried out with Sina News Corpus, and the performance of the algorithm is evaluated by recall rate, precision rate and macro average value. The results show that the integration of TF-IDF and the TextRank algorithm significantly outperforms the traditional algorithm in performance parameters and extraction effect.

7) Graph-based text summarization using modified textrank In paper [7] "Mallick C., Das A.K., Dutta M., Das A.K., Sarkar A. (2019) Graph-based text summarization using modified textrank. In: Nayak J., Abraham A., Krishna B., Chandra Sekhar G., Das A. (eds) soft computing in Data Analytics. Advances in intelligent systems and Computing, vol 758. Springer, Singapore" author says that, the efficient access of enormous amounts of information has become more difficult due to the rapid growth of the Internet. To manage the vast information, we need efficient and effective methods and tools. In this paper, a graph-based text summarization method has been described which captures the aboutness of a text document. The method has been developed using modified TextRank computed based on the concept of PageRank defined for each page in the Web pages.The proposed method constructs a graph with sentences as the nodes and similarity between two sentences as the weight of the edge between them. Modified inverse sentence frequency-cosine similarity is used to give different weightage to different words in the sentence, whereas traditional cosine similarity treats the words equally.The graph is made sparse and partitioned into different clusters with the assumption that the sentences within a cluster are similar to each other and sentences of different cluster represent their dissimilarity. The performance evaluation of proposed summarization technique shows the effectiveness of the method.

8) HipoRank: Incorporating hierarchical and positional Information into Graph-based. Unsupervised long document Extractive summarization In paper [8] " Yue Dong, Andrei Romascanu, Jackie C. K. Cheung, HipoRank: Incorporating hierarchical and positional Information into Graph-based Unsupervised long document Extractive summarization, arXiv, 2020,volume:abs/2005.00513. " the author proposes a novel graph-based ranking model for unsupervised extractive summarization of long documents. Graph-based ranking models typically represent documents as undirected fully-connected graphs, where a node is a sentence, an edge is weighted based on sentence-pair similarity, and sentence importance is measured via node centrality. Our method leverages positional and hierarchical information grounded in discourse structure to augment a document's graph representation with hierarchy and directionality. Experimental results on PubMed and arXiv datasets show that our approach outperforms strong unsupervised baselines by wide margins and performs comparably to some of the state-of-the-art supervised models that are trained on hundreds of thousands of examples. In addition, we find that our method provides comparable improvements with various distributional sentence representations; including BERT and RoBERTa models fine-tuned on sentence similarity.

9) Abstractive multi-modal meeting summarization. In paper [9] " Li Manling, Zhang Lingyu, Radke, Richar J, Ji Hend,Keep meeting summaries on topic: Abstractive multi-modal meeting summarization. 57th conference of the association for computational linguistics, 2021. " the author says that Transcripts of natural, multi-person meetings differ significantly from documents like news articles, which can make Natural Language Generation models generate unfocused summaries. We develop an abstractive meeting summarizer from both videos and audios of meeting recordings. Specifically, we propose a multi-modal hierarchical attention mechanism across three levels: topic segment, utterance and word. To narrow down the focus into topically-relevant segments, we jointly model topic segmentation and summarization. In addition to traditional textual features, we introduce new multi-modal features derived from visual focus of attention, based on the assumption that an utterance is more important if its speaker receives more attention. Experiments show that our model significantly outperforms the state-of-the-art with both BLEU and ROUGE measures.

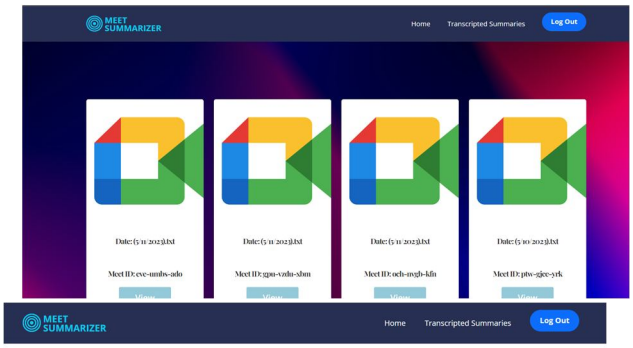
10) Leveraging BERT for Extractive text summarization on lectures In paper [10] "Derek Miller,Leveraging BERT for Extractive text summarization on lectures, arXiv preprint, arXiv:1906.04165,7 july 2019" the author says that in the last two decades, automatic extractive text summarization on lectures has demonstrated to be a useful tool for collecting key phrases and sentences that best represent the content. However, many current approaches utilize dated approaches, producing sub-par outputs or requiring several hours of manual tuning to produce meaningful results. Recently, new machine learning architectures have provided mechanisms for extractive summarization through the clustering of output embeddings from deep learning models. This paper reports on the project called Lecture Summarization Service, a python based RESTful service that utilizes the BERT model for text embeddings and KMeans clustering to identify sentences closes to the centroid for summary selection. The purpose of the service was to provide students a utility that could summarize lecture content, based on their desired number of sentences. On top of the summary work, the service also includes lecture and summary management, storing content on the cloud which can be used for collaboration. While the results of utilizing BERT for extractive summarization were promising, there were still areas where the model struggled, providing feature research opportunities for further improvement.

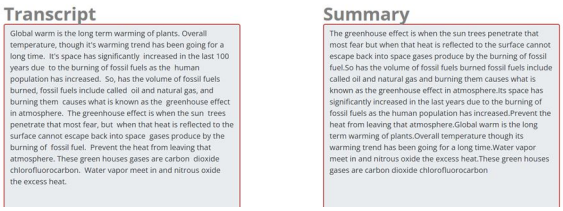
©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

11) Extractive summarization of long documents by combining global and local context In paper [11] " Wen Xiao and Giuseppe Carenini , Extractive summarization of long documents by combining global and local context, arXiv preprint, arXiv:1909.08089, 17 September 2019." the author propose a novel neural single document extractive summarization model for long documents, incorporating both the global context of the whole document and the local context within the current topic. We evaluate the model on two datasets of scientific papers , Pubmed and arXiv, where it outperforms previous work, both extractive and abstractive models, on ROUGE-1, ROUGE-2 and METEOR scores. We also show that, consistently with our goal, the benefits of our method become stronger as we apply it to longer documents. Rather surprisingly, an ablation study indicates that the benefits of our model seem to come exclusively from modeling the local context, even for the longest documents.

V. RESULT

Meeting Summarizer using Natural Language Processing (NLP), TF-IDF, and the PageRank algorithm is a concise and informative summary of the meeting. The system processes the meeting transcript, removes irrelevant information, and extracts key sentences using NLP techniques. The TF-IDF approach helps determine the importance of words in the document, and the PageRank algorithm assigns scores to sentences based on their relevance in the context of the meeting. The top-ranked sentences, identified through these techniques, are selected to form the output summary. This approach allows for efficient extraction of essential information, facilitating better understanding and decision-making based on the meeting content.





VI. CONCLUSION

Extractive summarization is a natural language processing task that involves selecting important sentences or phrases from a document and including them in a summary. It is commonly used for creating summaries of meeting transcripts, as it allows you to quickly understand the key points and conclusions of a meeting without having to read through the entire transcript. There are several algorithms that can be used for extractive summarization, such as the TextRank algorithm and the term frequency-inverse document frequency (TF-IDF) method. It is also important to perform preprocessing on the transcript to improve the quality of the text before generating the summary, which can involve tasks such as tokenization, lemmatization, and stopword removal.

While extractive summarization can be a useful tool for quickly understanding the content of a meeting transcript, it does not always generate the most concise summary, as it may include irrelevant or redundant information from the original document. Abstractive summarization, on the other hand, involves generating new sentences that capture the main points of the original document and can be more effective at generating concise summaries. However, abstractive summarization is generally more challenging to implement than extractive summarization, as it requires the system to understand the meaning of the text and generate new sentences.

VII. FUTURE SCOPE

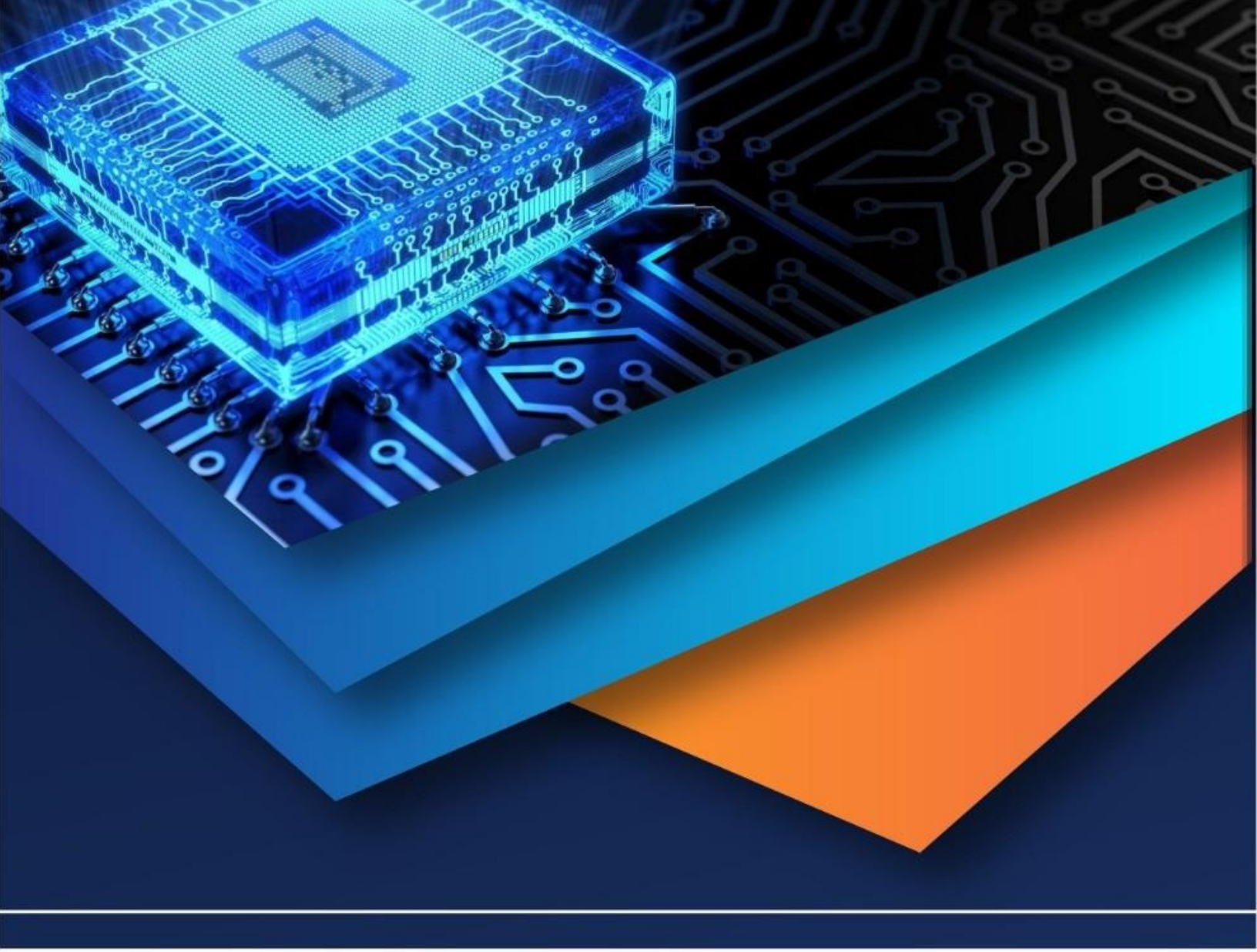
The future scope of Meeting Summarizer using Natural Language Processing (NLP) is promising. Advancements may include multi-modal summarization by integrating audio, video, and text, enhancing speaker identification and tracking for personalized summaries. Contextual understanding, such as sentiment analysis and entity recognition, can lead to more accurate summaries. Real-time summarization for live meetings, user customization, and integration with collaboration tools offer improved productivity. Robust evaluation metrics are needed to assess summary quality. Multilingual support can broaden its applicability. These advancements aim to optimize information extraction, knowledge management, and decision-making in organizational settings, leading to more efficient and effective meeting summaries.

VIII. ACKNOWLEDGEMENT

First, the authors wish to express our sincere gratitude to our project guide, Prof. Vydehi K, for her enthusiasm, patience, insightful comments, practical advice and unceasing ideas that have helped us tremendously at all times during our research. Her immense knowledge, profound experience, and professional expertise in NLP has enabled us to complete this research successfully which would not have been possible without her support and guidance. We also wish to express Our sincere thanks to Adi Shankara Institute of Engineering and Technology, for their consistent support.

REFERENCES

[1] Yash Agrawal, Atul Thakre, Tejas Tapas, Ayush Kedia,YashTelkhade and Vasundhara Rathod, Comparative analysis of NLP models for Google Meet Transcript summarization, EasyChair Preprint no. 5404, 2021.
[2] Yuanfeng Song, Di Jiang ,Xuefang Zhao ,Xiaoling Huang, QianXu, Raymond Chi-WingWong, Qiang Yang, SmartMeeting: Automatic Meeting transcription and summarization for in-person conversations. ACM International Conference on Multimedia 2021, Pages 2777-2779.
[3] Pratik K. Biswas, AleksandrIakubovich, Extractive summarization of call transcripts, arXiv Preprint arXiv: 2103.10599, 19 March, 2021.
[4] Rani, Ujjwal and Karambir Bidhan, Comparative assessment of Extractive summarization: textrank, tf-idf and lda, Journal of Scientific Research 65.1 (2021): 304-311.
[5] Nenkova A., McKeown K. (2012) ,A survey of text summarization Techniques. In: Aggarwal C., Zhai C. (eds) Mining Text Data. Springer, Boston, MA.
[6] AravindChandramouli, SiddharthShukla, Neeti Nair, ShivenPurohit, ShubhamPandey and MuraliMohana Krishna Dandu, Unsupervised paradigm for information extraction from transcripts using BERT, arXiv Preprint arXiv:2110.00949,13 september 2021.
[7] Yao, L., Pengzhou, Z., & Chi, Z. Research on News keyword extraction Technology based on tf-idf and textrank, IEEE/ACIS 18th International conference on computer and information science (ICIS), June 2019.
[8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pretraining of deep bidirectional Transformers for language understanding, arXiv Preprint arXiv:1810.04805.
[9] Xingxing Zhang, MirellaLapata, Furu Wei, and Ming Zhou. Neural latent Extractive document summarization, conference on empirical methods in Natural Language
Processing, 2018.
[10] Mallick C., Das A.K., Dutta M., Das A.K., Sarkar A. (2019) Graph-based text summarization using modified textrank. In: Nayak J., Abraham A., Krishna B., Chandra Sekhar G., Das A. (eds) soft computing in Data Analytics. Advances in intelligent systems and Computing, vol 758. Springer, Singapore.
[11] Yue Dong, Andrei Romascanu, Jackie C. K. Cheung, HipoRank: Incorporating hierarchical and positional Information into Graph-based Unsupervised long document Extractive summarization, arXiv, 2020, volume:abs/2005.00513.
[12] Derek Miller,Leveraging BERT for Extractive text summarization on lectures, arXiv preprint, arXiv:1906.04165,7 july 2019.
[13] Wen Xiao and Giuseppe Carenini , Extractive summarization of long documents by combining global and local context, arXiv preprint, arXiv:1909.08089, 17 September 2019.
[14] Li Manling, Zhang Lingyu, Radke, Richar J, Ji Hend,Keep meeting summaries on topic: Abstractive multi-modal meeting summarization. 57th conference of the association for computational linguistics, 2021.
[15] Shashi Narayan, Shay B. Cohen, Mirella Lapata, Ranking sentences for Extractive summarization with reinforcement learning. arXiv preprint, arXiv:1802.08636,23 february 2018.Proceedings of the Sixth International Conference on Trends in Electronics and Informatics (ICOEI 2022)IEEE Xplore Part Number: CFP22J32-ART; ISBN: 978-1-6654-8328-5978-1-6654-8328-5/22/



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)