



# North South University

## **CSE499A: Senior Design Project I**

### **Section: 10**

### **Related Works**

#### **Group Members:**

Serial	Name	NSU ID	Section
1	Md. Saiyem Raiyan	2012468042	10
2	Samia Sultana	2014048042	29
3	Sheikh Mohammed Wali Ullah	2021186042	10
4	Zobaer Ahammod Zamil	2021796042	10

## DIALOGLM: Pre-trained Model for Long Dialogue Understanding and Summarization

(Ming Zhong, Yang Liu<sup>2</sup>, Yichong Xu<sup>2</sup>, Chenguang Zhu<sup>2</sup>, Michael Zeng<sup>2</sup>)

This work proposes a window-based denoising approach for generative pre-training for long dialogue understanding and summarization. Long dialogues are a dense medium of information, bringing challenges for users to understand the gist and extract related information quickly. DIALOGLM, a pre-trained neural encoder-decoder model for long dialogue understanding and summarization. The benefits of this model are more pronounced in datasets with longer input text. With further pre-training, the model deals with long conversation scenarios. The results demonstrate that each dialogue-inspired noise and the proposed hybrid attention methods can bring further improvements to the model.

## Abstractive Meeting Summarization with Entailment and Fusion

(Yashar Mehdad Giuseppe Carenini Frank W. Tompa Raymond T. NG)

The authors propose a framework for abstractive meeting summarization, clustering sentences into communities and building an entailment graph to identify relevant sentences. They aggregate these sentences using a word graph model and use a ranking strategy to select the best path for abstract sentences. This work presents a comprehensive pipeline for producing abstractive summaries for each meeting transcript. Word graph-based method to aggregate and generate abstractive sentence summaries, extending *Filippova's method*. Instead of utilizing the traditional NLG pipeline and a word graph technique to capture the semantics of the meeting, our objective is to construct a meeting summary using a subset of annotations, including a human synopsis and connections to source meeting sentences. The CONGA algorithm is used to identify which nodes can be clustered as a community to generate an abstract sentence, using betweenness scores to detect communities in a graph.

## NLP Based Text Summarization Using Semantic Analysis

(Hamza Shabbir Moiyadi<sup>1</sup>, Harsh Desai<sup>2</sup>, Dhairya Pawar<sup>3</sup>, Geet Agrawal<sup>4</sup>, Nilesch M. Patil<sup>5</sup>)

A technique for generating the summarization of domain-specific text by using Semantic Analysis for text summarization. The input document is first parsed or pre-processed, wherein there is a removal of unneeded words such as 'stop words' which are simply small function words, like "the", "and", "a", which do not contribute meaning to the text summary. In the summarization process, the system arranges the sentences generated from the SVD Analysis stage by semantically placing them in a way that the summary encompasses all the concepts of the original text.

MandarMitra et al., from the Department of computer science, at Cornell University proposed a similar system for text summarization, they used another method based on paragraph extraction. M. S. Patil et al., suggested a summarization system based on several extractive text summarization approaches, and on the Support-Vector- Machine (SVM). Anne HendrikBuist et al. studied the disclosure of audio-visual meeting recordings, focusing on automatic meeting summarization. Josef Steinberger et al. proposed a generic text summarization method using latent semantic analysis to identify semantically important sentences. Jen-Yuan Yeh et al. used a trainable summarizer considering features like position positive keywords, and title resemblance. Ronan Collobert et al. developed a unified architecture for Natural Language Processing. Dipanjan Das et al. explored single and multiple document summarization, emphasizing empirical methods and extractive techniques. Hovy and Lin developed SUMMARIST, a multilingual automatic summarization system, but not all modules performed optimally. Dr.A. Jaya et al. studied abstractive summarization techniques in Indian languages, highlighting limited work in this field. Michael Ji's report categorized the product evaluation process into four phases: preparation, criteria establishment, characterization, and testing. In conclusion, Research shows extractive-based summarizing implementations have more success than abstractive-based ones, but their accuracy is not as high as expected.

## A Survey on NLP-based Text Summarization for Summarizing Product Reviews

*(Ravali Boorugu, Dr. G. Ramesh)*

This paper illustrates how to detect genuine products and choose the best options, users often read long reviews and survey various text summarization techniques, using seq2seq, LSTM, and attention mechanisms for improved accuracy. Niladri Chatterjee, Amol Mittal, and Shubham Goyal proposed an extractive text summarization technique using Genetic Algorithms. Amol Tandel et al. proposed a multi-document summarization technique that saves time and increases efficiency by condensing relevant data from multiple documents. They use LexRank to prevent score maximization of non-relevant sentences and assign lower scores to sentences with noisy data. Aditya Jain et al. proposed a Word Vector Embedding model for Extractive Text Summarization, addressing four main challenges: recognizing salient sentences, removing irrelevant information, minimizing details, and organizing information into a condensed report. Nithin Raphal, Hemanta Duwarah, and Philemon Daniel reviewed extractive methods select prominent sentences and maintain coherence, while abstractive methods create phrases or sentences that convey the complete meaning of the document. Word embedding and hot vector methods failed to detect similar words. Mikolov et al. solved this problem using the continuous skip-gram model and the continuous bag-of-words model, which project probable contextual words. Yang et al. proposed a query-based summarization model using search engines to develop background knowledge of the document's main theme.

## Summarizing Spoken and Written Conversations

*(Gabriel Murray and Giuseppe Carenini)*

This paper presents research on summarizing conversations in meetings and emails using a system that uses general conversational features across multiple domains, comparing its results with domain-dependent systems. This research takes an extractive approach to summarization, presenting a set of conversational features for locating the most salient sentences in meeting speeches and emails. Murray et al. (2005) compared textual summarization approaches with feature-based approaches, with human judges favoring feature-based approaches. Corston-Oliver et al. (2004) focused on identifying speech acts within an email, with a particular interest in task-related sentences. Carenini et al. (2007) used the Enron corpus to represent the thread as a fragment quotation graph, with a single node representing an email fragment. Rambow et al. (2004) addressed the challenge of summarizing entire threads by treating it as a binary sentence classification task. (2008) found Collective Message Summarization (CMS) to be more effective for summarizing email data, treating thread summarization as a multi-document summarization problem. Comparison: The ConverSumm system compares with state-of-the-art systems for meeting. Both systems use prosodic, lexical, structural, and speaker-related features, with two features overlapping between them. Both systems include length and position of sentences in the thread/conversation, with slight overlap in features. A general conversation summarization approach can achieve results comparable to state-of-the-art systems. It is effective in meetings and emails, and robust even in noisy ASR output. This system saves time and effort on implementing unique systems in various domains.

## Trainable, Scalable Summarization Using Robust NLP and Machine Learning

*(Chinatsu Aone, Mary Ellen Okurowski , James Gorlinsky)*

A trainable and scalable summarization system uses features from retrieval, extraction, NLP techniques, and online resources. It combines these features using a machine learning algorithm, demonstrating system scalability and usability. Trainable and scalable summarization system which utilizes features derived from information retrieval, information extraction, and NLP techniques and on-line resources. Frequency-based (Edmundson, 1969; Kupiec, Pedersen, and Chen, 1995; Brandow, Mitze, and Rau, 1995), knowledge-based (Reimer and Hahn, 1988; McKeown and Ladev, 1995), and discourse based (Johnson et al, 1993; Miike et al, 1994; Jones, 1995) approaches to automated summarization correspond to a continuum of increasing understanding of the text and increasing complexity in text processing. In both cases, they found that the effects of training increased system scores by as much as 10% F-Measure or greater. The experiments indicates that automatic summarization performance can be enhanced by discovering different combinations of features through a machine learning technique and that it can exceed lead summary performance and is affected by data source type.

## Comparative Analysis of NLP models for Google Meet Transcript Summarization

(Yash Agrawal, Atul Thakre, Tejas Tapas, Ayush Kedia, Yash Telkhade and Vasundhara Rathod)

This paper illustrates the comparison of 3 models, which is used to summarize google meetings. After removing articles, prepositions and other unnecessary parts, they used tokenization and lemmatization. Lastly, they used the NLTK (Natural Language Toolkit) model, Word Embedding Model, and T5 model. And perform ROUGE analysis. There are two models those use extractive summarization and one model for abstractive summarization. These models used frequency based, score based, matrix based, and mapping based technologies. Also used pre-trained datasets such as Glove. After training those models, using ROUGE analysis, they checked which one works more accurately without changing meaning of the context.

## A General Approach for Meeting Summarization: From Speech To Extractive Summarization

(Neslihan Akar, Metin Turan)

This research proposes a model to summarize a meeting using Natural Language Processing text summarization techniques (Extractive Summarization) supported by special dictionary usage. After extractive summarize and human summarize comparison were interpreted by an average of all similarity rates. The first only uses the TF-IDF algorithm. The second is the dictionary form that doesn't use a "word ratio". The third one used the dictionary form with 50% "word ratio". The last one used the dictionary form with 20% "word ratio". The automation of meeting summarization, and extracting pivotal information, holds immense potential to enhance operational efficiency across various business domains. The paper evaluates the summarization models by comparing the summaries generated by the model with those created by human summarizers. The research utilized the TF-IDF algorithm, which stands for Term Frequency-Inverse Document Frequency. The alpha coefficient of 10.2 is used to optimize the model. The study also compares summaries generated by the model with human summaries. This focuses on extractive summarization, a technique where significant sentences are selected directly from the original content. In conclusion, this model finds application in diverse scenarios such as condensing lengthy lectures, conferences, and meetings.

## REFERENCES

- [1] Zhong, M., Liu, Y., Xu, Y., Zhu, C., & Zeng, M. (2022). DialogLM: Pre-trained Model for Long Dialogue Understanding and Summarization. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(10), 11765-11773. <https://doi.org/10.1609/aaai.v36i10.21432>
- [2] Yashar Mehdad, Giuseppe Carenini, Frank Tompa, and Raymond T. Ng. 2013. *Abstractive Meeting Summarization with Entailment and Fusion*. In *Proceedings of the 14th European Workshop on Natural Language Generation*, pages 136–146, Sofia, Bulgaria. Association for Computational Linguistics. *Abstractive Meeting Summarization with Entailment and Fusion* (Mehdad et al., ENLG 2013)
- [3] Moiyadi, Hamza S., et al. "NLP Based Text Summarization Using Semantic Analysis." *International Journal of Advanced Engineering, Management and Science*, vol. 2, no. 10, Oct. 2016. [RIS \(Mendeley, Zotero, EndNote, RefWorks\)](#) [BibTeX \(LaTeX\)](#)
- [4] R. Boorugu and G. Ramesh, "A Survey on NLP based Text Summarization for Summarizing Product Reviews," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2020, pp. 352-356, doi: 10.1109/ICIRCA48905.2020.9183355.
- [5] Murray, Gabriel & Carenini, Giuseppe. (2008). Summarizing Spoken and Written Conversations.. 773-782. 10.3115/1613715.1613813.
- [6] Chinatsu Aone, Mary Ellen Okurowski, and James Gorlinsky. 1998. *Trainable, Scalable Summarization Using Robust NLP and Machine Learning*. In *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1*, pages 62–66, Montreal, Quebec, Canada. Association for Computational Linguistics.
- [7] Agrawal, Y., Thakre, A., Tapas, T., Kedia, A., Telkhade, Y., & Rathod, V. (2021). Comparative analysis of NLP models for Google Meet Transcript summarization. *EasyChair Preprint*, (5404).
- [8] Akar, N., & Turan, M. (2022). A General Approach for Meeting Summarization: From Speech to Extractive Summarization. *Review of Computer Engineering Research*, 9(2), 83-95.