Team Synapse @ AutoMin 2023: Leveraging BART-Based Models for Automatic Meeting Minuting

Kristýna Klesnilová and **Michelle Elizabeth**
Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics,
Charles University, Czech Republic
kristyna.klesnil@gmail.com, michelle.elizabethk@gmail.com

Abstract

This paper describes the approach we followed for our submission to the Second Run of the Automatic Minuting Shared Task. Our methodology centers around employing BART-based models fine-tuned on diverse summarization corpora. The segmented meeting transcripts are fed into the models, generating summaries that are subsequently combined and formatted into the final meeting minutes.

1 Introduction

The COVID-19 pandemic has expedited digital transformation across industries, significantly impacting the conduct of meetings. With the restriction of physical gatherings, online meetings have emerged as the primary mode of communication and collaboration. This shift towards virtual meetings has highlighted the crucial need for automatic summarization of meeting transcripts. By harnessing the power of Natural Language Processing (NLP), organizations can optimize their virtual collaboration, ensuring accurate documentation, streamlined processes, and enhanced information management.

This paper presents our endeavor to develop a robust system for generating minutes from meeting transcripts, undertaken as part of the Second Run of the Automatic Minuting Shared Task (Ghosal et al., 2022, Ghosal et al., 2023). The development of this system for automatic minuting has been influenced by previous research in the field, serving as the basis for our work. In particular, we draw inspiration from the research which pioneered the use of BART summarization models for meeting summarization tasks (Shinde et al., 2021).

We begin by looking at the related works from the previous iteration of the AutoMin Shared Task (Ghosal et al., 2021). Then we provide a concise overview of the datasets utilized in the task, followed by a comprehensive description of the system architecture we implemented. The system overview encompasses detailed explanations of the pre-processing steps, the conducted experiments, and the post-processing techniques applied to refine the generated minutes. Subsequently, we present our results and discuss potential avenues for improving the performance of our system.

2 Related Work

Automatic meeting summarization is a relatively new use case compared to the traditional task of summarizing text. The first edition of the AutoMin Shared Task (Ghosal et al., 2021) provides valuable insight into the work done in this area and explores numerous methods with which the participants approached the task.

The use of pre-trained language models, especially transformer-based architectures like BART (Shinde et al., 2021), T5-base (Mahajan et al., 2021), and GPT-2 (Garg and Singh, 2021) was a prominent approach. These models were then fine-tuned on the task-specific dataset to improve performance. One approach used for multilingual summarization involved translation from Czech to English, generating the minutes in English and then translating the results back to Czech (Yamaguchi et al., 2021).

Incorporating other techniques such as coreference resolution and dialogue partitioning during pre-processing (Žilinec and Re, 2021), syntactic phrase extraction, redundant word deletion, and vectorization with TF-IDF scores (Iakovenko et al., 2021) attempted to enhance the quality of generated summaries. Argumentation mining techniques were utilized (Yamaguchi et al., 2021) to improve coherence and internal structure, highlighting the importance of organizational and contextual coherence in meeting minutes.

Proceedings of the 16th International Natural Language Generation Conference: Generation Challenges, pages 108–113
September 11–15, 2023. ©2023 Association for Computational Linguistics

3 Dataset Description

We participated in Task A of the AutoMin 2023 Shared Task, the goal of which was to generate minutes from meeting transcripts. The task runs in two languages, English and Czech, and separate meeting corpora were available for both languages. The first edition of the AutoMin Shared Task (Ghosal et al., 2021) used the ELITR Minuting Corpus (Nedoluzhko et al., 2022). In addition to that, this year, a new meeting corpus EuroParlMin created from the European Parliamentary debates was also made available to the participants for training. Since we participated only in the minuting of English meeting transcripts, we will only describe the datasets corresponding to English.

The ELITR Minuting Corpus consists of 84, 36, and 12 transcript-minute instances for train, dev,[1] and test sets, respectively. The transcripts, which are text files, contain ASR outputs of the meetings and therefore are not very refined. Each transcript has one or more corresponding minutes generated in a specific format with details like the date, attendees, the purpose of the meeting, the summary (in bullet points), and the name of the annotator. Some transcripts have additional information on the gender of the attendees and the alignment of the transcript and minutes.

The EuroParlMin consists of 2065, 187, and 242 transcript-minute instances for train, dev, and test sets, respectively. Each dataset contains directories labeled by the date of the session. Each directory contains the transcripts and minutes of one or more chapters or sections of the meeting. Chapters are split further into parts. During the EuroParlMin transcript revision, grammar and stylistic corrections were already incorporated, resulting in reduced cleaning requirements on our part compared to the ELITR Minuting Corpus. The minutes follow a paragraph-style format and contain only a summary of the transcript. They do not report other details like date, list of attendees, etc., which were present in the minutes of the ELITR Minuting Corpus.

4 System Overview

In this section, we provide a comprehensive overview of the system architecture implemented for the automatic minuting of meeting transcripts. We begin by presenting the pre-processing steps undertaken to prepare the input data for the summarization models. Next, we delve into the details of the experiments conducted, focusing on the fine-tuning of the BART summarization model (Lewis et al., 2019) on meeting summarization corpora. We then discuss the post-processing steps employed for the generation of concise minutes as the final output of our system. Figure 1 shows the end-to-end functioning of our system. The source code can be found at `https://github.com/klesnkri/automin-2023-team-synapse`.

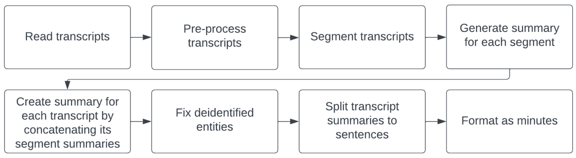

Figure 1: System diagram

4.1 Pre-processing

As a first step, we pre-process the transcript data by splitting them into speaker-utterance pairs and normalizing the utterances.

For ELITR Minuting Corpus, we apply a series of text normalization techniques, including the removal of tags (e.g., <cough/>, <laugh/>, <censored/>) and ASR stopwords and errors, deletion of punctuation at the start of sentences, removal of consecutive duplicate tokens and punctuation, and sentence normalization. Figure 2 and Figure 3 illustrate the steps involved in the pre-processing of ELITR Minuting corpus and an example of the raw text before and after pre-processing, respectively.

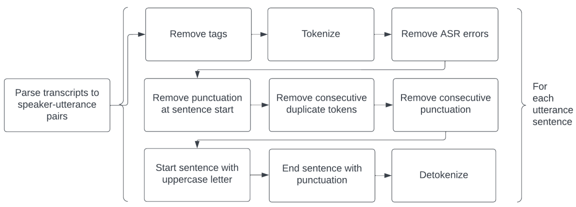

Figure 2: Pre-processing of ELITR Minuting Corpus

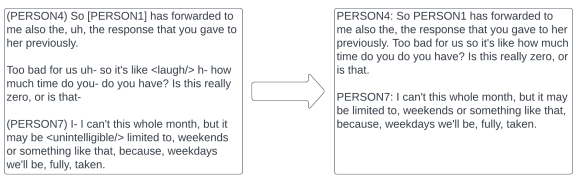

Figure 3: Example of ELITR Minuting Corpus pre-processing

[1]The dev set also includes the two test sets from the first run of AutoMin Shared Task.

Similarly, for EuroParlMin, we remove lines that were not speaker utterances, introduce the PERSON entity so the speaker-utterance pairs have the same format as in ELITR Minuting Corpus, remove punctuation, language codes, and other irrelevant information from the start of utterances, and normalize whitespaces. See Figure 4, for a detailed diagram.

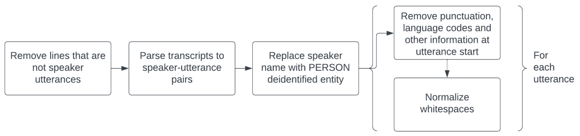

Figure 4: Pre-processing of EuroParlMin Corpus

4.2 Segmentation

To address the input length limitation of the BART architecture, we slice the speaker-utterance pairs into segments of uniform token length. We experiment with varying segment lengths of 512, 768, and 1024 tokens.

4.3 Summarization

We use three BART large summarization models trained on distinct datasets to generate summaries for the segmented data. All of the models are publicly available on the Hugging Face repository. We pass the segmented data into these models and rejoin the segment summaries to obtain the raw summary text.

The first model, MEETING_SUMMARY[2] was trained on the XSUM Dataset (Narayan et al., 2018), AMI Meeting Corpus (Mccowan et al., 2005), SAMSUM Dataset (Gliwa et al., 2019), and DIALOGSUM Dataset (Chen et al., 2021). The second model, bart-large-cnn-samsum[3] was trained on CNN Daily Mail (See et al., 2017) and SAMSUM Dataset. Finally, the third model, bart-large-xsum[4] was originally trained on the XSUM Dataset and we further fine-tuned it on the SAMSUM Dataset.

4.4 Post-processing

After obtaining the summarization, we perform further post-processing to ensure the deidentified entities retain the correct format and the summarized sentences are formatted as minutes. We experiment with deleting some non-informative sentences from the summaries using TextRank (Mihalcea and Tarau, 2004). However, ultimately, we decide to keep all the sentences to ensure coherence in the minutes.

5 Results

We evaluate the generated summaries using the ROUGE-1, ROUGE-2, and ROUGE-L metrics on the development data. The automatic evaluation results are summarized in Table 1 and Table 2. Since automatic evaluation serves only as a supplementary measure for this task, we also looked at several outputs and compared them to the minutes provided in the development datasets for both corpora. The final models were chosen based on our manual assessments of these outputs. The MEETING_SUMMARY model proved effective for the ELITR Minuting Corpus, benefiting from pre-training on similar dialogue datasets. However, it did not perform well for the EuroParlMin corpus, where speaker utterances are much longer.

According to our experiments, the MEETING_SUMMARY model with a segment length of 768 tokens is the most suitable for generating ELITR Minuting Corpus minutes, while the bart-large-cnn-samsum model with a segment length of 1024 tokens is the most appropriate for generating the EuroParlMin minutes.

6 Conclusion

In this paper, we presented our approach for automatic minuting, focusing on fine-tuning the BART summarization model using meeting summarization corpora. For ELITR corpus, we chose the MEETING_SUMMARY model with a segment length of 768 tokens, and for EuroParlMin corpus, we settled on the bart-large-cnn-samsum model with a segment length of 1024 tokens. While our current approach yields promising results, there are areas for future improvements, such as exploring dialogue summarization models like DialogLM (Zhong et al., 2022), which show potential in addressing the challenge of processing lengthy meeting transcripts. Our intention is to refine our system continuously and advance the field of automatic minuting, ultimately providing more accurate and coherent meeting minutes.

[2] https://huggingface.co/knkarthick/MEETING_SUMMARY

[3] https://huggingface.co/philschmid/bart-large-cnn-samsum

[4] https://huggingface.co/facebook/bart-large-xsum

	Segment Length 512		
Model	**ROUGE-1**	**ROUGE-2**	**ROUGE-L**
MEETING_SUMMARY	0.364	0.111	0.179
bart-large-cnn-samsum	0.331	0.121	0.170
bart-large-xsum-samsum	0.367	0.119	0.184
	Segment Length 768		
Model	**ROUGE-1**	**ROUGE-2**	**ROUGE-L**
MEETING_SUMMARY	**0.390**	**0.113**	**0.191**
bart-large-cnn-samsum	0.368	0.126	0.189
bart-large-xsum-samsum	0.388	0.113	0.194
	Segment Length 1024		
Model	**ROUGE-1**	**ROUGE-2**	**ROUGE-L**
MEETING_SUMMARY	0.379	0.102	0.190
bart-large-cnn-samsum	0.380	0.115	0.191
bart-large-xsum-samsum	0.379	0.103	0.190

Table 1: Automatic evaluation for ELITR Minuting Corpus

	Segment Length 512		
Model	**ROUGE-1**	**ROUGE-2**	**ROUGE-L**
MEETING_SUMMARY	0.225	0.072	0.145
bart-large-cnn-samsum	0.261	0.075	0.157
bart-large-xsum-samsum	0.233	0.073	0.150
	Segment Length 768		
Model	**ROUGE-1**	**ROUGE-2**	**ROUGE-L**
MEETING_SUMMARY	0.210	0.069	0.139
bart-large-cnn-samsum	0.251	0.072	0.153
bart-large-xsum-samsum	0.218	0.070	0.145
	Segment Length 1024		
Model	**ROUGE-1**	**ROUGE-2**	**ROUGE-L**
MEETING_SUMMARY	0.198	0.066	0.133
bart-large-cnn-samsum	**0.241**	**0.070**	**0.150**
bart-large-xsum-samsum	0.206	0.068	0.140

Table 2: Automatic evaluation for EuroParlMin Corpus

Limitations

While our system shows promising progress in generating meeting minutes, there are several limitations that need to be addressed to enhance its overall performance.

Our system lacks a robust sentence-ranking mechanism to filter out irrelevant content from the generated minutes. This deficiency may lead to the inclusion of extraneous information, especially when the transcripts are generated using automatic speech recognition, reducing the accuracy and conciseness of the minutes. We are not explicitly tracking speaker utterances and the references in them, and the failure to properly handle references can result in disjointed and less coherent meeting minutes.

Our current system's limited generalization to various meeting formats hampers its versatility. It may struggle to produce satisfactory minutes for informal or specialized meetings, affecting its practical applicability.

Acknowledgements

This research was partially supported by the grant 19-26934X (NEUREM3) of the Czech Science Foundation.

References

Yulong Chen, Yang Liu, Liang Chen, and Yue Zhang. 2021. DialogSum: A real-life scenario dialogue summarization dataset. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 5062–5074, Online. Association for Computational Linguistics.

Amitesh Garg and Muskaan Singh. 2021. Team Symantlytical @ AutoMin 2021: Generating Readable Minutes with GPT-2 and BERT-based Automatic Minuting Approach. In *Proc. First Shared Task on Automatic Minuting at Interspeech 2021*, pages 65–70.

Tirthankar Ghosal, Ondřej Bojar, Marie Hledíková, Tom Kocmi, and Anja Nedoluzhko. 2023. Overview of the second shared task on automatic minuting (automin) at inlg 2023. In *Proceedings of the 16th International Conference on Natural Language Generation: Generation Challenges*. Association for Computational Linguistics.

Tirthankar Ghosal, Ondřej Bojar, Muskaan Singh, and Anja Nedoluzhko. 2021. Overview of the First Shared Task on Automatic Minuting (AutoMin) at Interspeech 2021. In *Proc. First Shared Task on Automatic Minuting at Interspeech 2021*, pages 1–25.

Tirthankar Ghosal, Marie Hledíková, Muskaan Singh, Anna Nedoluzhko, and Ondřej Bojar. 2022. The second automatic minuting (AutoMin) challenge: Generating and evaluating minutes from multi-party meetings. In *Proceedings of the 15th International Conference on Natural Language Generation: Generation Challenges*, pages 1–11, Waterville, Maine, USA and virtual meeting. Association for Computational Linguistics.

Bogdan Gliwa, Iwona Mochol, Maciej Biesek, and Aleksander Wawer. 2019. SAMSum corpus: A human-annotated dialogue dataset for abstractive summarization. In *Proceedings of the 2nd Workshop on New Frontiers in Summarization*, pages 70–79, Hong Kong, China. Association for Computational Linguistics.

Olga Iakovenko, Anna Andreeva, Anna Lapidus, and Liana Mikaelyan. 2021. Team MTS @ AutoMin 2021: An Overview of Existing Summarization Approaches and Comparison to Unsupervised Summarization Techniques. In *Proc. First Shared Task on Automatic Minuting at Interspeech 2021*, pages 59–64.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2019. BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *CoRR*, abs/1910.13461.

Parth Mahajan, Muskaan Singh, and Harpreet Singh. 2021. Team AutoMinuters @ AutoMin 2021: Leveraging state-of-the-art Text Summarization model to Generate Minutes using Transfer Learning. In *Proc. First Shared Task on Automatic Minuting at Interspeech 2021*, pages 34–40.

Iain Mccowan, J Carletta, Wessel Kraaij, Simone Ashby, S Bourban, M Flynn, M Guillemot, Thomas Hain, J Kadlec, V Karaiskos, M Kronenthal, Guillaume Lathoud, Mike Lincoln, Agnes Lisowska Masson, Wilfried Post, Dennis Reidsma, and P Wellner. 2005. The ami meeting corpus. *Int'l. Conf. on Methods and Techniques in Behavioral Research*.

Rada Mihalcea and Paul Tarau. 2004. TextRank: Bringing order into text. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 404–411, Barcelona, Spain. Association for Computational Linguistics.

Shashi Narayan, Shay B. Cohen, and Mirella Lapata. 2018. Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. *ArXiv*, abs/1808.08745.

Anna Nedoluzhko, Muskaan Singh, Marie Hledíková, Tirthankar Ghosal, and Ondřej Bojar. 2022. ELITR minuting corpus: A novel dataset for automatic minuting from multi-party meetings in English and Czech. In *Proceedings of the Thirteenth Language*

Resources and Evaluation Conference, pages 3174–3182, Marseille, France. European Language Resources Association.

Abigail See, Peter J. Liu, and Christopher D. Manning. 2017. Get to the point: Summarization with pointer-generator networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1073–1083, Vancouver, Canada. Association for Computational Linguistics.

Kartik Shinde, Nidhir Bhavsar, Aakash Bhatnagar, and Tirthankar Ghosal. 2021. Team ABC @ AutoMin 2021: Generating Readable Minutes with a BART-based Automatic Minuting Approach. In *Proc. First Shared Task on Automatic Minuting at Interspeech 2021*, pages 26–33.

Atsuki Yamaguchi, Gaku Morio, Hiroaki Ozaki, Ken ichi Yokote, and Kenji Nagamatsu. 2021. Team Hitachi @ AutoMin 2021: Reference-free Automatic Minuting Pipeline with Argument Structure Construction over Topic-based Summarization. In *Proc. First Shared Task on Automatic Minuting at Interspeech 2021*, pages 41–48.

Ming Zhong, Yang Liu, Yichong Xu, Chenguang Zhu, and Michael Zeng. 2022. Dialoglm: Pre-trained model for long dialogue understanding and summarization.

Matúš Žilinec and Francesco Ignazio Re. 2021. Team Matus and Francesco @ AutoMin 2021: Towards Neural Summarization of Meetings. In *Proc. First Shared Task on Automatic Minuting at Interspeech 2021*, pages 53–58.