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

Kristýna Klesnilová, 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

Technology plays a crucial role in helping mankind navigate today's fast-paced world. With everything going online and remote work becoming the new norm, it has become increasingly important to efficiently manage the flow of information, especially during online meetings. Thus, automatic minuting of meetings is an essential tool that can enhance one's productivity and efficiency. 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 inputted into the models, generating summaries that are subsequently combined and formatted into the final meeting minutes. Through this work, we have attempted to address the needs of contemporary online communication and collaboration.

Index Terms: automatic minuting, meeting summarization, topic segmentation, multi-party dialogues

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 minuting 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. The development of this system for automatic minuting has been influenced by previous research in the field, serving as a foundation for our work. We acknowledge the invaluable insights and contributions made by researchers who have laid the groundwork for automatic summarization and minuting of meeting transcripts. In particular, we draw inspiration from the research which pioneered the use of BART summarization models for meeting summarization tasks[1].

We begin by providing 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. 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[2] used the ELITR Minuting Corpus. 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. The transcripts are not outputs of ASR and hence will not require extensive cleaning as in the case of 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.

3. 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[3] 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.

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

3.1. Pre-processing

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

For the ELITR Minuting Corpus, we apply a series of text normalization techniques, including the removal of tags and ASR stopwords and errors, deletion of punctuation at the start of sentences, removal of consecutive duplicate tokens and punctuation, and sentence normalization.

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

3.2. Segmentation

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

3.3. 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 from the Huggingface 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[4], AMI Meeting Corpus[5], SAMSUM Dataset[6], and DIALOGSUM Dataset[7]. The second model, bart-large-cnn-samsum[3] was trained on the CNN Daily Mail[8] 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.

3.4. Post-processing

After obtaining the summarization, we perform further post-processing to ensure the deidentified entities retain the correct uppercase format and the summarization sentences are formatted as minutes. We experiment with deleting some non-informative sentences from the summaries using TextRank[9]. However, ultimately, we decide to keep all the sentences to ensure coherence in the minutes.

4. Results

We evaluate the generated summaries using the ROUGE-1, ROUGE-2, and ROUGE-L metrics on the development data. Additionally, we conduct a manual evaluation to complement the automatic evaluation results. The automatic evaluation results are summarized in Table 1 and Table 2.

Table 1: *Automatic evaluation for the ELITR Minuting Corpus*

Model	Segment length 512		
	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
Model	Segment length 768		
	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
Model	Segment length 1024		
	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 2: *Automatic evaluation for the EuroParlMin*

Model	Segment length 512		
	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
Model	Segment length 768		
	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
Model	Segment length 1024		
	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

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

5. Conclusion and Future Work

In this paper, we presented our approach for automatic minuting, focusing on fine-tuning the BART summarization model using meeting summarization corpora. While our models successfully generate concise minutes as intended, we acknowledge that there are areas for improvement. Moving forward, we aim to enhance the quality of the generated minutes by implementing a sentence-ranking mechanism to eliminate irrelevant content. Additionally, improving coreference resolution capabilities will enable better tracking of speaker utterances, enhancing the coherence of the generated minutes. Exploring the application of dialogue summarization models, such as DialogLM[10], shows promise in addressing the challenge of processing lengthy meeting transcripts. These areas represent avenues for future work, where we intend to refine our system and further advance the field of automatic minuting. By addressing these issues, we strive to create a more robust and accurate automatic minuting system that caters to the nuanced requirements of meeting transcription and summarization.

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

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

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

6. References

[1] K. Shinde, N. Bhavsar, A. Bhatnagar, and T. Ghosal, "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*, 2021, pp. 26–33.

[2] T. Ghosal, O. Bojar, M. Singh, and A. Nedoluzhko, "Overview of the First Shared Task on Automatic Minuting (AutoMin) at Interspeech 2021," in *Proc. First Shared Task on Automatic Minuting at Interspeech 2021*, 2021, pp. 1–25.

[3] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, "BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," *CoRR*, vol. abs/1910.13461, 2019. [Online]. Available: http://arxiv.org/abs/1910.13461

[4] S. Narayan, S. B. Cohen, and M. Lapata, "Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization," *ArXiv*, vol. abs/1808.08745, 2018.

[5] I. Mccowan, J. Carletta, W. Kraaij, S. Ashby, S. Bourban, M. Flynn, M. Guillemot, T. Hain, J. Kadlec, V. Karaiskos, M. Kronenthal, G. Lathoud, M. Lincoln, A. Lisowska Masson, W. Post, D. Reidsma, and P. Wellner, "The ami meeting corpus," *Int'l. Conf. on Methods and Techniques in Behavioral Research*, 01 2005.

[6] B. Gliwa, I. Mochol, M. Biesek, and A. Wawer, "SAMSum corpus: A human-annotated dialogue dataset for abstractive summarization," in *Proceedings of the 2nd Workshop on New Frontiers in Summarization*. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 70–79. [Online]. Available: https://www.aclweb.org/anthology/D19-5409

[7] Y. Chen, Y. Liu, L. Chen, and Y. Zhang, "DialogSum: A real-life scenario dialogue summarization dataset," in *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*. Online: Association for Computational Linguistics, Aug. 2021, pp. 5062–5074. [Online]. Available: https://aclanthology.org/2021.findings-acl.449

[8] A. See, P. J. Liu, and C. D. Manning, "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)*. Vancouver, Canada: Association for Computational Linguistics, Jul. 2017, pp. 1073–1083. [Online]. Available: https://www.aclweb.org/anthology/P17-1099

[9] R. Mihalcea and P. Tarau, "TextRank: Bringing order into text," in *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*. Barcelona, Spain: Association for Computational Linguistics, Jul. 2004, pp. 404–411. [Online]. Available: https://aclanthology.org/W04-3252

[10] M. Zhong, Y. Liu, Y. Xu, C. Zhu, and M. Zeng, "Dialoglm: Pre-trained model for long dialogue understanding and summarization," 2022.