Towards Syntax-Aware Dialogue Summarization using Multi-task Learning

Seolhwa Lee[1], **Kisu Yang**[2], **Chanjun Park**[3], **João Sedoc**[4], **Heuiseok Lim**[3,*]
[1]University of Copenhagen, Denmark
[2]VAIV Corp, South Korea
[3]Korea University, South Korea
[4]New York University, United States
sele@di.ku.dk
{bcj1210, limhseok}@korea.ac.kr, ksyang@vaiv.kr, jsedoc@stern.nyu.edu

Abstract

Abstractive dialogue summarization is challenging for several reasons: firstly, multiple speakers from different textual styles participate in dialogue, and secondly, informal dialogue structures (*e.g.,* slang, colloquial representation). We constructed a syntax-aware model by leveraging linguistic information (*i.e.,* POS tagging), which alleviates the above issues by inherently distinguishing sentences uttered from individual speakers. We employed multi-task learning of both syntax-aware information and dialogue summarization. Our approach is the first method to apply multi-task learning to the dialogue summarization task. Experiments on a SAMSum corpus (a large-scale dialogue summarization corpus) demonstrated that our method improved upon the vanilla model.

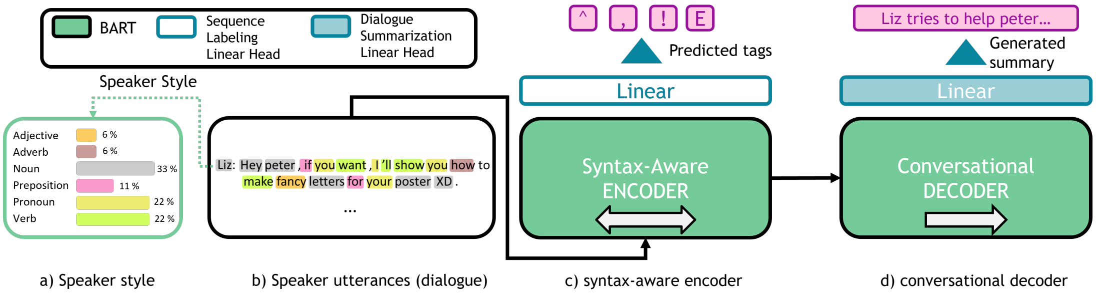
1 Background

During the COVID-19 pandemic, a virtual conversation tool like Zoom is inevitable. With this much demand, dialogue summarization has emerged as a means to summarize the dialogues. There are two challenges in dialogue summarization aforementioned. To address these challenges, we investigated the relationship between textual styles and representative attributes of utterances. (Kübler et al., 2010) proposed that the types (*e.g.,* intent or role of a speaker) of sentences from speakers are associated with different syntactic structures, such as part-of-speech (POS) tagging. This is derived from the fact that different speaker roles are characterized by different syntactic structures. In essence, the uttered text has a unique representation from each speaker, like a voiceprint (*i.e.,* identity information from the human voice) (Guo et al., 2021). Based on this prior research, we began our study with the assumption that because syntactic structures tend to be associated with a representative of a sentence uttered from speakers, these structures would help distinguish the different styles of utterances. In this work, we propose a novel abstractive dialogue summarization model for use in a daily conversation setting, characterized by an informal style of text that employs multi-task learning to learn linguistic information and dialogue summarization simultaneously.

2 Methodology

Approach Inspired by the success of the BART model (Lewis et al., 2020) on text summarization, we address two different tasks simultaneously: sequence labeling and summary generation. BART consists of a bidirectional encoder and an autoregressive decoder. Therefore, we conducted the sequence labeling task in the encoder (*i.e.,* syntax-aware encoder) and the summary generation task in the decoder (*i.e.,* conversational decoder). That is, task-specific linear heads were trained through multi-task learning, which performs the main task as a dialogue summarization task and the POS sequence labeling task as an auxiliary task.

* Corresponding author.


BART
Sequence Labeling Linear Head
Dialogue Summarization Linear Head
Speaker Style
Adjective 6 %
Adverb 6 %
Noun 33 %
Preposition 11 %
Pronoun 22 %
Verb 22 %
Liz: Hey peter , if you want , I 'll show you how to make fancy letters for your poster XD .
...
^ , ! E
Predicted tags
Linear
Syntax-Aware ENCODER
Liz tries to help peter...
Generated summary
Linear
Conversational DECODER

a) Speaker style

b) Speaker utterances (dialogue)

c) syntax-aware encoder

d) conversational decoder

Figure 1: (a) Speaker style example of frequency of POS tags. (b) Speaker utterances example. (c) The syntax-aware encoder with a task-specific linear head learns the sequence labeling task given the dialogue. (d) The conversational decoder learns the dialogue summarization task through the linear head.

Proposed Model The architecture in Figure 1 shows (a) the example of a uttering style from speaker based on POS frequencies, and (b) the input sequence to the utterances with separation token, and also (c) the application of syntax-aware information conducting sequence labeling, and finally (d) the decoder received syntax-aware encoder representation to generate summaries. Finally, we employed the joint training manner by using the λ parameter to adjust the strength in sequence labeling task for multi-task learning.

Experimental Setup We trained and evaluated our model on a large-scale dialogue summary dataset SAMSum (Gliwa et al., 2019) based on ROUGE (Lin, 2004) and BertScore (Zhang et al., 2019) metrics. We automatically annotated seqeunce labels (Gimpel et al., 2010) for all the utterances as these are not included in the SAMSum corpus. Also, we tested the different setting input type as LONGEST-10 in our proposed model to retain the utterance format. LONGEST-10 takes ten lengthy utterances of the dialogue inspired by LEAD-3 (See et al., 2017). We used the `BART-base` model to initialize the backbone of the encoder/decoder frame and followed the default settings. The learning rate was set to 3e-4. We trained the model for 20 epochs. Also, we set λ as 0.1 in the final model. The training was conducted on a single RTX 8000 GPU with 48 GB memory.

3 Results & Discussion

Results In Table 1, we present the ROUGE-1, ROUGE-2, and ROUGE-L scores between our model and other, comparison models. Our proposed model improved the other baselines with respect to F1 for all ROUGE scores. *As hypothesized previously, our experiments demonstrate that the usage of linguistic information is worthwhile to enhance the model performance.* The key factor related to the overall lower performance of the baseline models seems to be that the baseline models fundamentally are not based on the language model; however, the DynamicConv model with the GPT-2 embeddings is based on the usage of pretrained embeddings from the language model GPT-2, which is trained on a large corpus.

Model	**Type**	**ROUGE-1**			**ROUGE-2**			**ROUGE-L**		
		F	P	R	F	P	R	F	P	R
Pointer Generator (See et al., 2017)*	-	0.401	-	-	0.153	-	-	0.366	-	-
DynamicConv + GPT-2 (Wu et al., 2019)*	-	0.418	-	-	0.164	-	-	0.376	-	-
Fast Abs RL Enhanced (Chen and Bansal, 2018)*	-	0.420	-	-	0.181	-	-	0.392	-	-
BART †	LONG-10	0.426	0.488	0.419	0.188	0.220	0.184	0.419	0.464	0.415
Syntax-aware BART † (λ =0.1)	LONG-10	**0.431**	0.486	0.426	**0.189**	0.216	0.186	**0.420**	0.460	0.418

Table 1: Performance comparison of the proposed method with different models on the test set. * denotes the results from (Chen and Yang, 2020), and † corresponds to our proposed method model, which shows the best performance (LONGEST-10). Note that F, P, and R indicate F1, precision, and recall scores, respectively.

In Table 2, we find our proposed method slight improvement than baseline model although the score margin is not high. This result could support the hypothesis that our approach has a positive influence on model performance.

The benefits of our approach are firmly shown in Table 3. It is unknown whether 'uni' is the name of a store or a specific place in the conversation, but the proposed model generates the 'uni' into 'the university', unlike the baseline model. Although it is different from the reference's intention, our proposed model can be interpreted as having syntax-aware characteristics by completing a word as related to the place. The second example also shows the complete grammar as to added 'a' in our proposed method.

Model	BertScore
BART	0.90
Syntax-aware BART (ours)	0.91

Table 2: Performance of BertScore.

REF	Ali left his wallet at Mohammad's place. Mohammad'll bring it to uni tomorrow.
Syntax-aware BART (ours)	Mohammad found Ali's wallet yesterday. He will bring it to the university tomorrow.
BART	Ali found his wallet. Mohammad will bring it to uni tomorrow.
REF	Maddie will buy a white bread and apples on John's request.
Syntax-aware BART (ours)	Maddie is in Asda. John will buy a white bread and apples for Maddie.
BART	Maddie is in Asda. John will buy white bread and some apples in Gala.

Table 3: Examples for model generated from ours and BART model (baseline). REF– reference summary, Blue– ours, Red–baseline.

Conclusions In this study, we proposed a novel syntax-aware sequence-to-sequence model that leverages syntactic information, considering the informal daily chat structure constraints, and implicitly distinguishes the different textual styles from multiple speakers for dialogue summarization. We benchmarked the experiments using the SAMSum corpus, and the experimental results demonstrate that the proposed method improves comparison models for all ROUGE scores. We concluded that the proposed approach has a positive impact on syntax awareness than baseline.

Acknowledgements

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2018-0-01405) supervised by the IITP(Institute for Information Communications Technology Planning Evaluation). This work was supported by the National Research Foundation(NRF), Korea, under project BK21 FOUR.

References

Yen-Chun Chen and Mohit Bansal. 2018. Fast abstractive summarization with reinforce-selected sentence rewriting. *arXiv preprint arXiv:1805.11080*.

Jiaao Chen and Diyi Yang. 2020. Multi-view sequence-to-sequence models with conversational structure for abstractive dialogue summarization. *arXiv preprint arXiv:2010.01672*.

Kevin Gimpel, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A Smith. 2010. Part-of-speech tagging for twitter: Annotation, features, and experiments. Technical report, Carnegie-Mellon Univ Pittsburgh Pa School of Computer Science.

Bogdan Gliwa, Iwona Mochol, Maciej Biesek, and Aleksander Wawer. 2019. Samsum corpus: A human-annotated dialogue dataset for abstractive summarization. *arXiv preprint arXiv:1911.12237*.

Miao Guo, Jiaxiong Yang, and Shu Gao. 2021. Speaker recognition method for short utterance. In *Journal of Physics: Conference Series*, volume 1827, page 012158. IOP Publishing.

Sandra Kübler, Matthias Scheutz, Eric Baucom, and Ross Israel. 2010. Adding context information to part of speech tagging for dialogues. In *Ninth International Workshop on Treebanks and Linguistic Theories*, page 115.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online, July. Association for Computational Linguistics.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81.

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.

Felix Wu, Angela Fan, Alexei Baevski, Yann N Dauphin, and Michael Auli. 2019. Pay less attention with lightweight and dynamic convolutions. *arXiv preprint arXiv:1901.10430*.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.