SUMBot: Summarizing Context in Open-Domain Dialogue Systems

Rui Ribeiro and Luísa Coheur
INESC-ID, Instituto Superior Técnico,
Universidade de Lisboa

CMU PORTUGAL SUMMIT 2022
NEW FRONTIERS IN TECH

Introduction

Most models struggle to identify and incorporate important knowledge from dialogues and simply use the entire turns as context, which increases the size of the input fed to the model with unnecessary information.

Results

Complete Turns	Includes Summary?	BLEU-4 (%)	ROUGE-1 (%)	ROUGE-2 (%)	ROUGE-L (%)	Avg. Length	Max. Length
0	No	3.70	18.4	4.23	17.6	71	115
2	No	3.94	19.2	4.55	18.3	96	291
4	No	3.86	19.4	4.62	18.5	118	309
6	No	4.03	19.6	4.30	18.6	136	366
8	No	3.32	19.3	4.50	18.4	150	274
10	No	3.89	18.0	3.66	17.2	160	434
0	Yes	3.76	18.7	4.23	17.9	86	115
2	Yes	3.95	19.5	4.72	18.5	107	305
4	Yes	3.95	19.1	4.19	18.2	127	349
6	Yes	3.73	18.9	4.28	18.0	140	376
8	Yes	4.11	19.5	4.44	18.6	153	380
10	Yes	4.05	19.3	4.13	18.3	162	386

Tab. 1: Results for BLEU and ROUGE metrics on the Persona-Chat dataset.

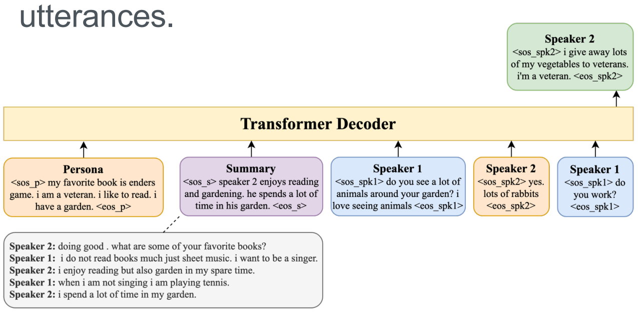
- First, we fine-tune BART in an abstractive summarization corpus and use it to generate summaries for the dialogue context.
- Then, we fine-tune a GPT decoder with the summaries from the previous stage by incorporating them with the dialogue between both speakers.

Conclusions

- We show that it is possible to improve dialogue generation and reduce the size of the input.
- However, the weak quality of the summaries may influence the overall performance of the system.

Methods

We propose a simple method that only includes a few complete speaker turns as input, and the remaining turns are compiled into a summary that describes succinctly the omitted utterances.



Speaker 2
<sos_spk2> i give away lots of my vegetables to veterans. i'm a veteran. <eos_spk2>

Transformer Decoder

Persona
<sos_p> my favorite book is enders game. i am a veteran. i like to read. i have a garden. <eos_p>

Summary
<sos_s> speaker 2 enjoys reading and gardening. he spends a lot of time in his garden. <eos_s>

Speaker 1
<sos_spk1> do you see a lot of animals around your garden? i love seeing animals <eos_spk1>

Speaker 2
<sos_spk2> yes. lots of rabbits <eos_spk2>

Speaker 1
<sos_spk1> do you work? <eos_spk1>

Speaker 2: doing good . what are some of your favorite books?
Speaker 1: i do not read books much just sheet music. i want to be a singer.
Speaker 2: i enjoy reading but also garden in my spare time.
Speaker 1: when i am not singing i am playing tennis.
Speaker 2: i spend a lot of time in my garden.

Fig. 1: Example of an input fed to the decoder.

- The results show that, in general, the inclusion of a summary improves the generation results when the number of complete turns are the same.
- The errors from the summaries are propagated to the decoder, which may contribute to a performance decrease.
- In other scenarios, the summary focuses on irrelevant information such as greetings: “Speaker 2 wants to know how are you doing”.

Acknowledgements

This work was partially funded by Fundação para a Ciência e a Tecnologia under the scope of the CMU Portugal (project UIDB/50021/2020).