Coreference-Aware Abstractive Dialogue Summarization

Arda Yüksel *Georgios Sigas* *Abdullah Orhun Aksoy*

Summarization

Text summarization in Natural Language Processing is the process of summarizing the information in large texts for quicker consumption. It is essential for the summary to be a fluent, continuous and depict the significant.

Abstractive Summarization

There are two types of summarization, extractive and abstractive. Extractive summarization considers the exact sentences as they appear in the text. On the other side, abstractive summarization does not simply copy important phrases from the source text but also potentially come up with new phrases that are relevant, which can be seen as paraphrasing

Abstractive Dialogue Summarization

Common abstractive summarization methods doesn't fit well on dialogues since they do not perform well on multi-party conversations. Therefore, abstractive dialog summarization stands as a roof for methods that are special for the dialogue texts such as meetings, calls, interviews etc.

Abstractive Dialogue Summary Examples

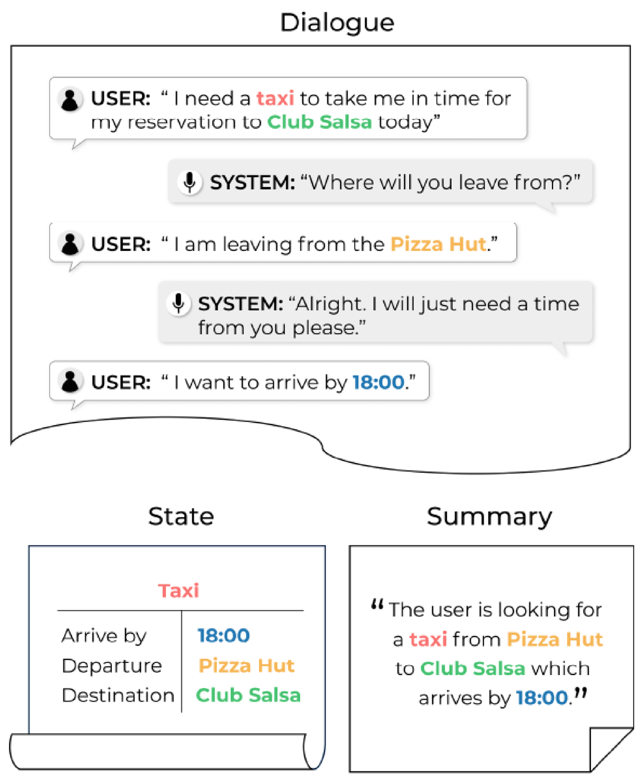

Figure 1. *An abstractive summary with dialog states from a call to a taxi station*

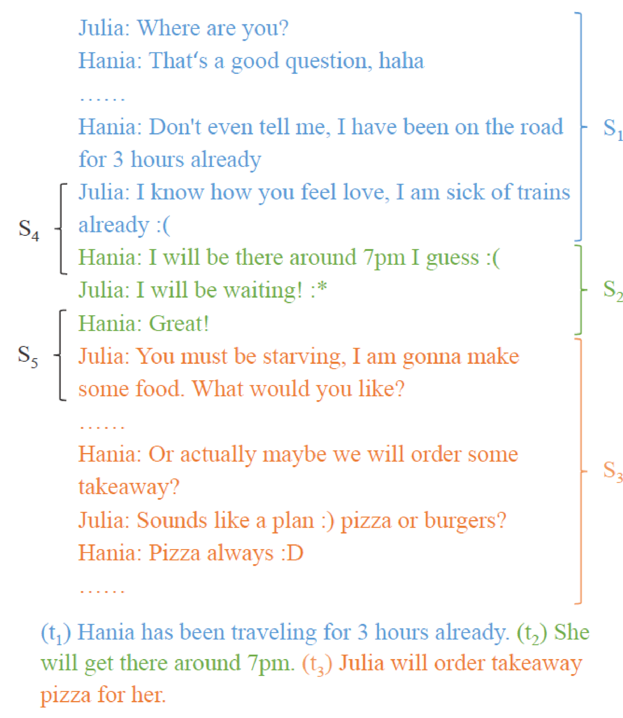

Figure 2. *An abstractive summary of a conversation in between two friends*

Methodology

BART

Bidirectional and Auto-Regressive Transformer (BART) is a sequence-to-sequence (Seq2Seq) model with a bidirectional encoder similar to BERT's and a left to right autoregressive decoder similar to GPT's.

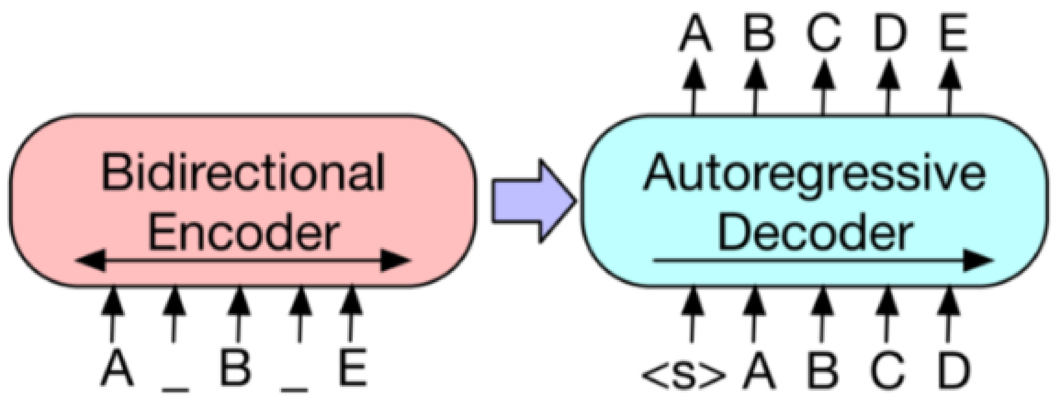

Figure 3. *BART brief model visualization*

Transfer Learning

- Training a model can be difficult with limited resources
- We used an already **pre-trained** model on dialogue summarizations
- We further trained **a part** of that model on an other dataset of our task

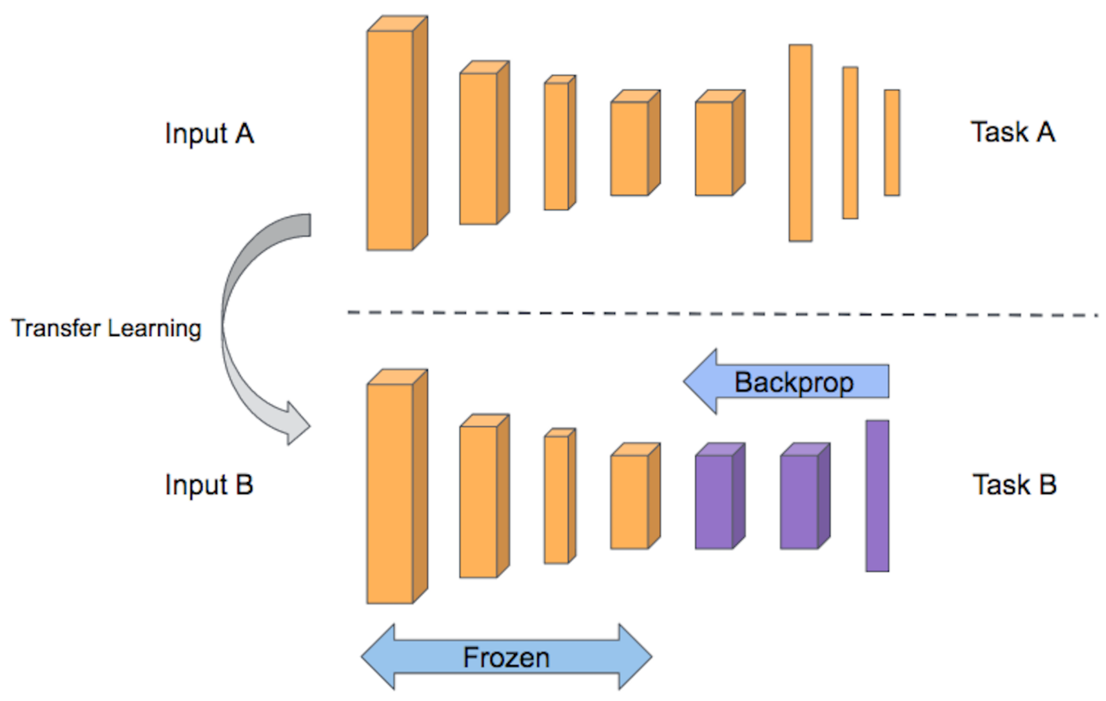

Figure 4. *Logic of Transfer Learning*

Coreference-Aware Dialogue Sums

Max: Know any good ***sites*** to buy clothes from?
Payton: Sure :) <file_other><file_other><file_other>
Max: That's a lot of ***them***!
Payton: Yeah, but ***they*** have different things so **I** usually buy things from 2 or 3 of ***them***.
Max: **I**'ll check ***them*** out. Thanks.

The coreference graph encoding layer is used to model the coreference connections between all mentions.

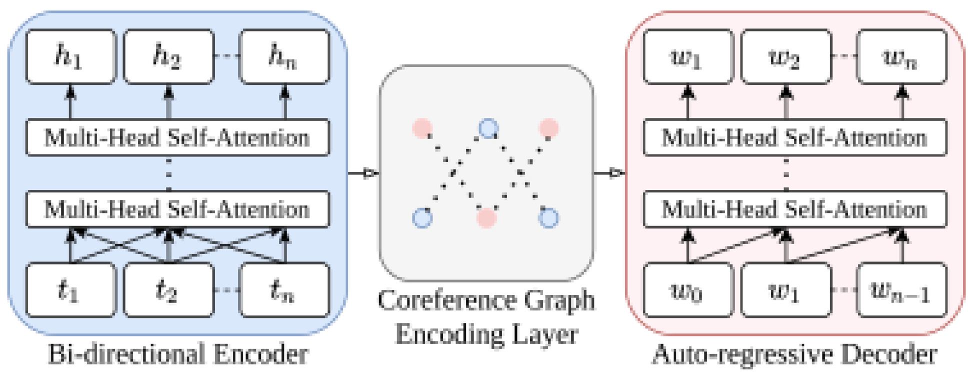

Figure 5. *Coreference graph encoding layer placement*

Dialogue Summarization Corpora

Name	*Domain*
SAMSum	Chat
DIALOGSUM	Spoken
MEDIASUM	Interview
XSUM*	Article

Table 1. *Some of the common datasets that are being used in dialogue summarization*
**XSum is not a dialog dataset*

Corpora Used

- SamSum
  - 16k dialogues
  - Messenger-like conversations
- XSum
  - 226k articles from BBC
  - Wide variety of domains
- DialogSum
  - 13k dialogues
  - Face-to-face conversations

Examples from SAMSum

Chat	Ernest: hey Mike, did you park your car on our street? Mike: no, took it into garage today Ernest: ok good Mike: why? Ernest: someone just crashed into a red honda looking just like yours Mike: lol lucky me	Anne: You were right, he was lying to me :/ Irene: Oh no, what happened? Jane: who? that Mark guy? Anne: yeah, he told me he's 30, today I saw his passport - he's 40 Irene: You sure it's so important? Anne: he lied to me Irene
Golden Summary	Mike took his car into garage today. Ernest is relieved as someone had just crashed into a red Honda which looks like Mike's.	Mark lied to Anne about his age. Mark is 40.
Predicted Summary (Test Results)	Mike took his car to the garage. Someone crashed into a red honda looking just like Mike's.	Mark lied to Anne. His passport shows he's 40.

Training

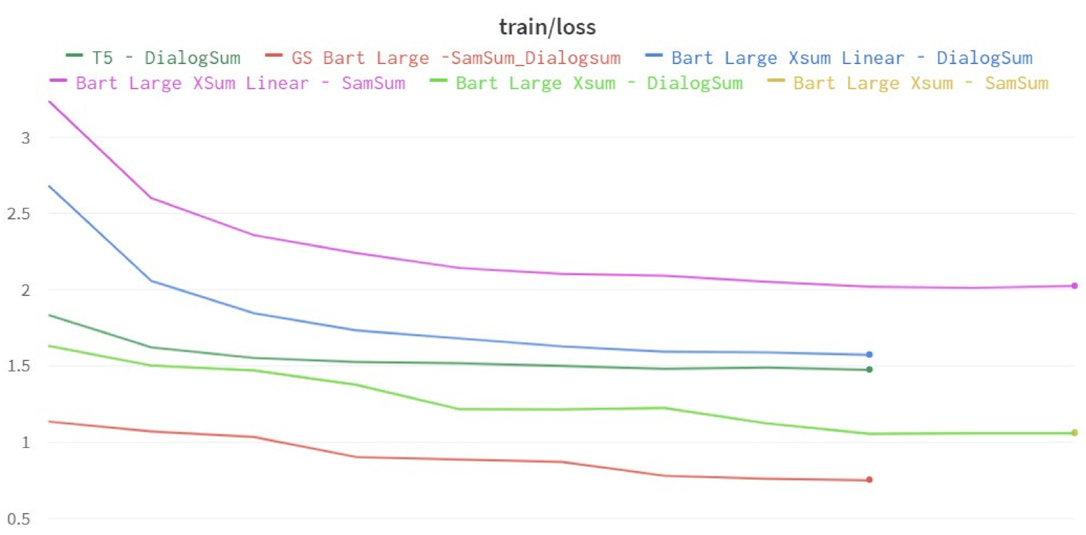

Figure 6. *Training Loss of BART LARGE and T5 based models for 3 Epochs*

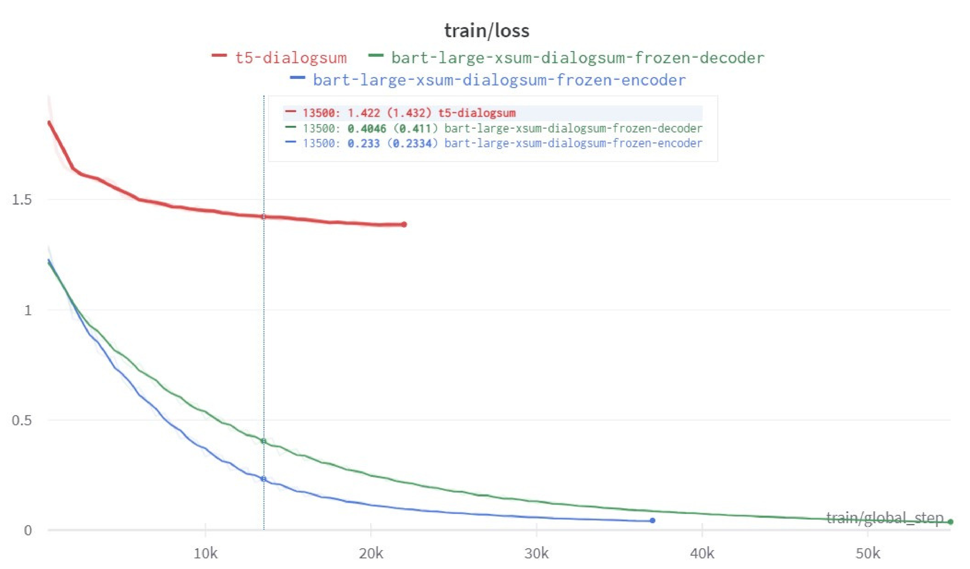

Figure 7. *Training Loss of BART LARGE and T5 models for longer epochs*

Validation

MODELS	ROUGE 1	ROUGE 2	ROUGE L
Different Learning Rate	**50,90**	23,83	46,19
Coreference Head	50,79	23,34	45,47
Replace Coref Head	50,14	22,87	45,02
Graph Only Training	50,31	**24,20**	**46,69**
Larger Hidden Size	49,30	23,66	45,86
Graph Only Training	49,30	23,66	45,86

Table 2. *BART LARGE XSUM based Graph Model Validation Scores*

Model	Corpus	Frozen	ROUGE 1	ROUGE 2	ROUGE L
T5	DialogSum	Encoder	34,99	12,18	29,14
Bart Large Samsum	DialogSum	Encoder	48,43	24,06	40,23
Bart Large Xsum	DialogSum	Encoder + Decoder	37,58	14,86	31,76
Bart Large Xsum	SamSum	Encoder + Decoder	38,92	15,49	31,60
Bart Large Xsum	DialogSum	Encoder	47,78	23,25	39,65
Bart Large Xsum	SamSum	Encoder	52,25	26,44	41,91

Table 3. *Common Transformer Models Validation Scores for Different Corpus and Setups*

Test

Model	ROUGE 1	ROUGE 2	ROUGE L
Frozen Decoder	0,50	0,24	0,46
Graph Only Training	0,48	0,23	0,45

Table 4. *Graph Based Models SamSum Test Rouge ScoresT*

Pretrained	Training Corpus	SAMSUM			DIALOGSUM		
		R - 1	R - 2	R - LSUM	R - 1	R - 2	R - LSUM
XSum	SamSum	**51,01**	**25,41**	**46,52**	35,35	13,71	31,57
XSum	DialogSum	45,95	19,26	40,38	44,52	19,01	39,12
SamSum	DialogSum	47,41	20,88	42,31	**44,98**	**19,62**	**39,92**

Table 5. *BART Models Test Rouge Scores for SamSum and DialogSumT*