

Topic oriented Summarization using Transformer based Encoder Decoder model

Vipul Chauhan, Prasenjeet Roy, Lipika Dey and Tushar Goel

Tata Consultancy Services Limited, Delhi, India

(chauhan.vipul, r.prasenjeet2, lipika.dey, t.goel)@tcs.com

Introduction

Automatic Summarization is the task of generating a fluent, concise and coherent summary of the document.

Most of the work focused on single-speaker documents such as news, scientific publications etc. They contain a limited number of concepts around which summaries are generated. Dialogues are important forms of communication, which contain lot of information about ideas exchanged and nature of the participants. Dialogue summarization aims to condense a piece of content generated by multiple participants into a short passage. Dialogues are difficult to summarize since the underlying data contains diverse interactive patterns between speakers as well as inherent topic drifts.

We proposed a topic-oriented dialog summarizer such that same text can be summarized differently based on the topics given, by focusing on different portions of the text.

DialogSum Challenge

Dialogue – an important channel for achieving communicative intents, differs from monologic texts in nature and has received significantly less attention from the summarization research community.

A shared task on summarizing real-life scenario dialogues.

Challenge asks a model to generate a salient, concise, fluent, and coherent summary, given a piece of multi-turn dialogue text.

Both automatic and manual blind evaluations were conducted on the submitted model.

Dialogue Summary Overview

Dialogue Text:

#Person1#: Why didn't you tell me you had a girlfriend?

#Person2#: Sorry, I thought you knew.

#Person1#: But you should tell me you were in love with her.

#Person2#: Didn't I?

#Person1#: You know you didn't.

#Person2#: Well, I am telling you now.

#Person1#: Yes, but you might have told me before.

#Person2#: I didn't think you would be interested.

#Person1#: You can't be serious. How dare you not tell me you are going to marry her?

#Person2#: Sorry, I didn't think it mattered.

#Person1#: Oh, you men! You are all the same.

Summary from Dataset:

#Person1#'s angry because **#Person2#** didn't tell **#Person1#** that **#Person2#** had a girlfriend and would marry her.

Topic from Dataset: have a girlfriend

DialogSum Dataset



The public dataset consists of training, dev, test and hidden-test dataset



The training dataset (12,460) and dev dataset (500) consist of single human annotated summary and short topic for each dialogue



The test dataset (500) consists of 3 annotated summary and associated 3 short topic for each dialogue.



The hidden test (100) is not provided with any annotated summary but included a short topic for each dialogue.

Challenges

- Dialogue has diverse interactive patterns between speakers as well as the inherent drift of topics.
- It has frequent co-reference like referring same person or thing while in conversation – which makes it difficult to understand.
- Because of the special linguistic phenomena, the dialogue text can be difficult to encode using ordinary representation learning technologies.
- The dialogues are summarized from an observer's perspective – like “thinks” and “agrees” summarizes dialogue actions. Hence, it not only summarizes what the speaker is saying but also summarizes what the speaker is doing.

Pre-trained PEGASUS model

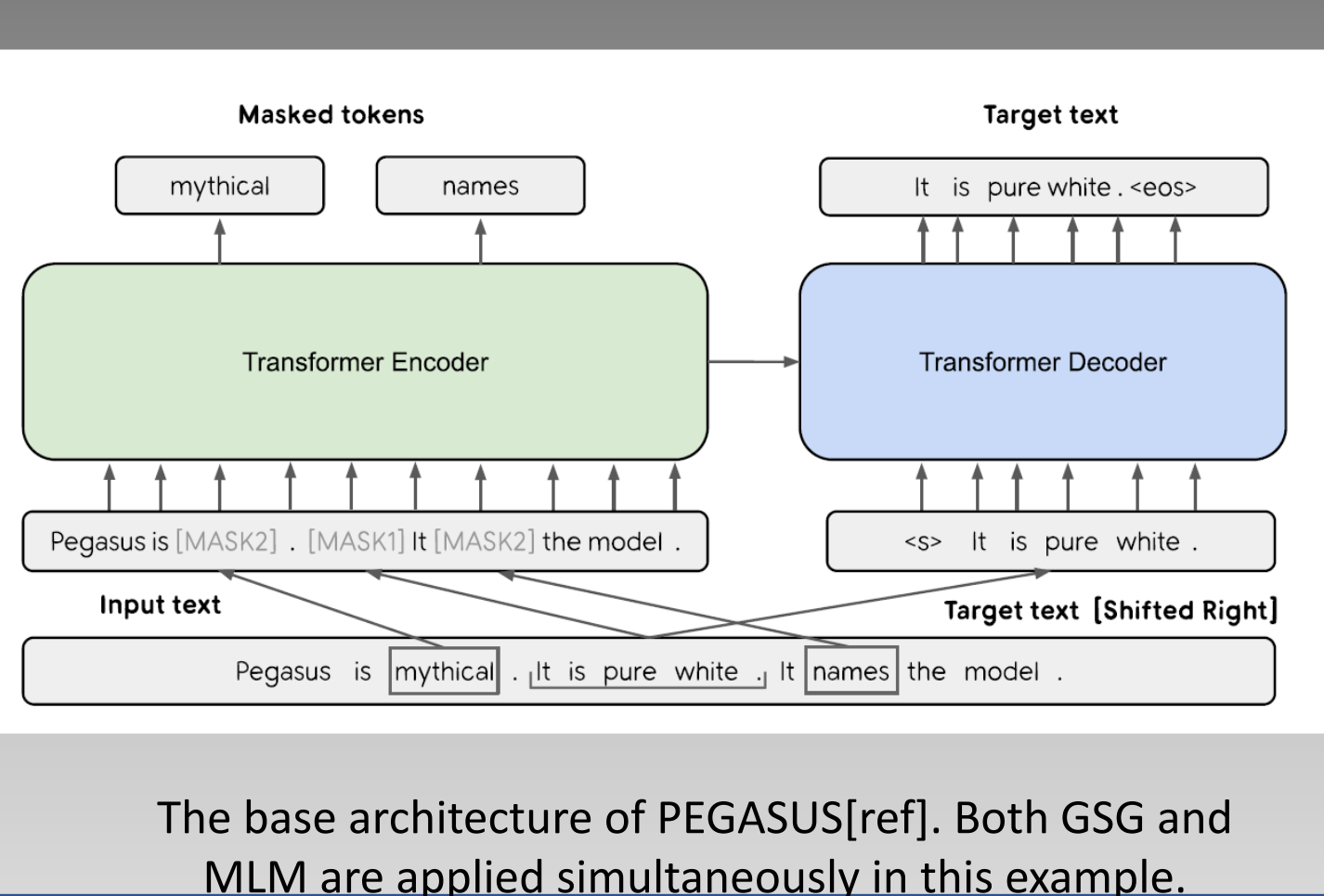
PEGASUS (Pre-training with Extracted Gap-sentence for Abstractive Summarization) is a Transformer encoder-decoder model trained with a self-supervised pre-training objective. It achieved state-of-the-art performance on 12 downstream summarization tasks spanning news, science, stories, emails etc.

Variants –
PEGASUS_{BASE} – 12 layers and H=768
PEGASUS_{LARGE} – 16 layers and H=1024

Pre-training Corpora – C4 [Colossal and cleaned version of Common Crawl] & HugeNews

Task –

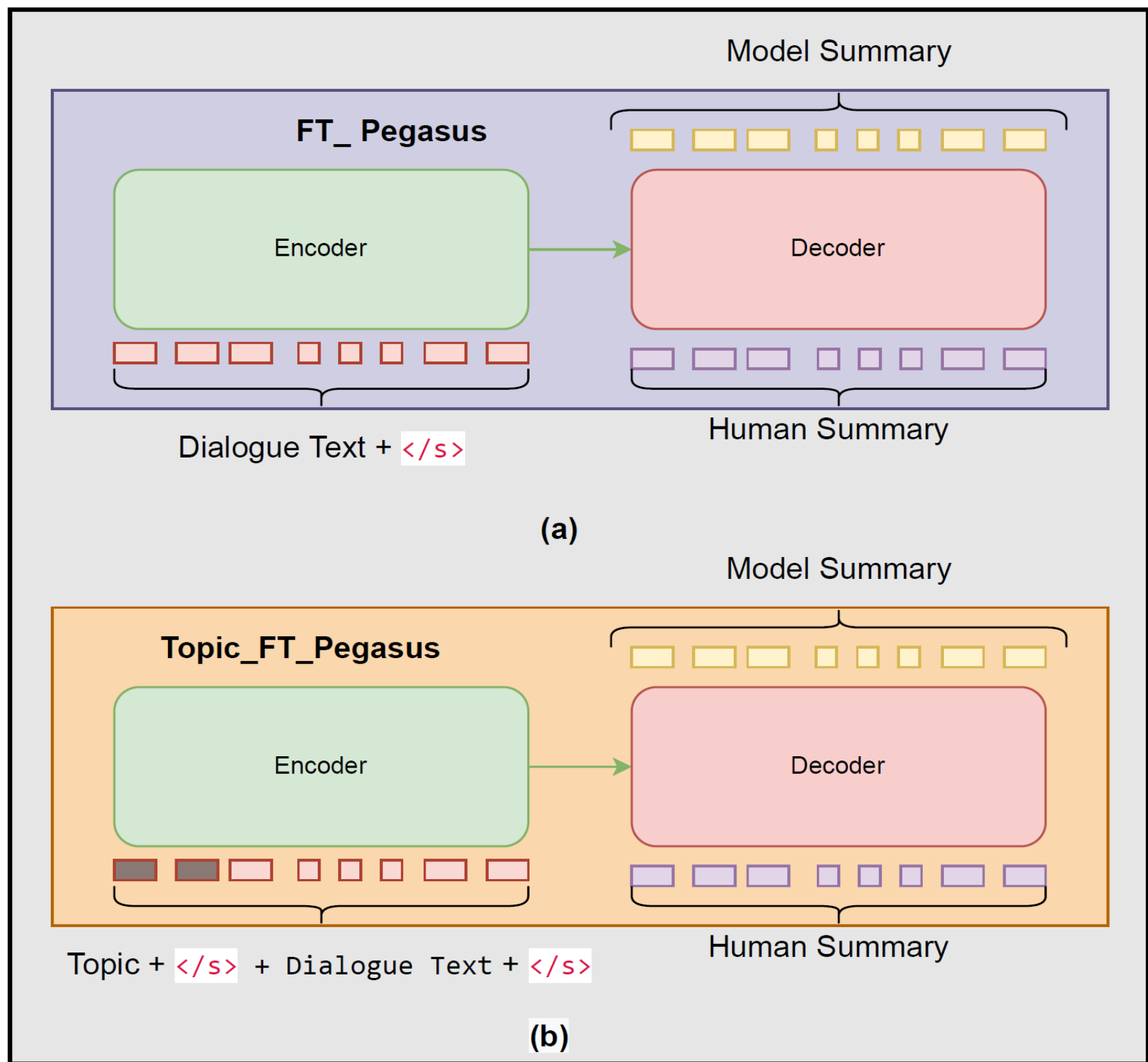
- Gap Sentence Generation (GSG)**– select top m scored sentences according to importance [Rouge1-F1 between the sentence and rest of the document]
- Masked Language Model (MLM)**



Proposed Fine-Tuning Methodology

Since the pre-trained PEGASUS was providing the salient information as summary from the dialogue text but it was unable to handle linguistic phenomena and coreferences of the dialogues. Hence there was a need of fine-tuning the model on DialogSum data. Thankfully we had sufficient data for fine-tuning.

Then standard method of finetuning is shown in Figure (a). The dialogue text is passed as the input and the summary is generated as the output. To generate topic driven summary- we proposed a method of finetuning as shown in Figure (b). Here dialogue text is passed along with the topic – separated by a special character – as an input to the model. This way the model learns to focus on different text segments centred around the given topic.



Model Output

Dialogue Text
#Person1#: Who stands out in your mind as a man or woman of sound character?
#Person2#: If I think of famous people, I think of Abraham Lincoln.
#Person1#: He's the US president, who walked five miles just to give a lady her change, isn't he?
#Person2#: That's the one. He also was famous for never giving up on his goals.
#Person1#: That's right. He ran for office quite a few times before he was finally elected.
#Person2#: And I also admire him for his courage in fighting for equal rights.
#Person1#: He had great vision, didn't he?
#Person2#: And humility. I would have liked to meet him personally.

Methodology Significance – the same dialogue text is summarized differently based on the given topic, by focusing on different portion of the text.

This is a sample from the test set of dataset – here 3 different topics were given – like sound character, famous people and Abraham Lincoln. Our model has generated different summaries considering topic.

Experiment and Results

Model	Average Score				Best Score			
	R1	R2	RL	B-S	R1	R2	RL	B-S
PT_PEGASUS	25.99	6.41	20.97	87.77	37.63	9.63	26.48	88.15
FT_PEGASUS	43.36	18.36	36.23	92.19	51.59	26.58	45.54	92.64
Topic_FT_PEGASUS	49.42	21.81	40.85	92.22	54.53	32.00	51.47	93.22

Evaluation results over the public test dataset.

S.No.	Team	R1	R2	RL	B-S
1.	GoodBai	49.66	26.03	48.44	91.69
2.	UoT	49.75	25.15	46.50	91.76
3.	IITP-CUNI	45.89	21.88	43.16	91.13
4.	TCS_WITM_2022	50.32	25.59	47.40	91.81

Evaluation results on the hidden test set evaluated by the challenge organisers.

Baselines

- PT_PEGASUS – pretrained PEGASUS_LARGE generate the summary using dialogue text as an input.
- FT_PEGASUS – fine-tuned PEGASUS_LARGE using DialogSum with input dialogue text as input.

Observations

- Fine-tuning helped the model to learn the linguistic phenomena of dialogues.
- Incorporating topics while fine-tuning allows the model to focus on different text segments centered around a given topic.

Evaluation Measures

1. ROUGE (Recall Oriented Understudy for Gisting Evaluation) automatically measures the quality of generated summary by counting the overlapping units like n-grams with reference summary. ROUGE-1, ROUGE-2 and ROUGE-L have been used for the evaluation.

2. BERTScore leverages the pre-trained contextual embeddings from BERT and matches the conceptual similarity between the model-generated and human summaries.

Conclusion and Future work

- Generate a topic-oriented summary using pre-trained abstractive model
- Shown that the PEGASUS can be fine-tuned using the proposed methodology.
- Proposed methodology performed significantly better than baselines.

Future work

- Incorporating nuances of dialogue, speech act theory etc.
- Propose a model that generate topic from dialogue