Large Language Models Report

Yuqing Qiao
Khoury College of Computer Sciences
Northeastern University
qiao.yuqi@northeastern.edu
July 1st, 2024

1 Introduction Model Choice

In this report, I explore three transformer models for the task of dialogue summarization and the process of fine-tuning large language models using the SAMSum dataset. The models tested are BART, T5-large, and PEGASUS. Summarization is a critical task in natural language processing (NLP) that aims to generate concise and coherent summaries of long text documents.

2 Dataset Preprocessing

The SAMSum dataset contains dialogues and their corresponding summaries. The dialogues and summaries were tokenized using the respective Hugging Face tokenizers, ensuring proper padding and truncation to fit the model's input requirements. The Trainer API from Hugging Face is used to facilitate the process of fine-tuning the model. This includes the training dataset, validation dataset, tokenizer, training arguments (hyperparameters), metrics, and the pretrained model.

4 Fine Tuning Process

Training arguments were defined, including the number of epochs, batch size, evaluation strategy, and the use of mixed precision training.

5 Summarization Performance

5.1 Before Fine-Tuning

Three models—BART, T5-large, and Pegasus—were tested on three dialogues. Their performance was human-evaluated based on coherence, fluency, key points, and clarity.

The BART summary captures multiple dialogue exchanges but lacks coherence and includes redundant phrases. It includes key points but lacks clarity and has repeated phrases. The T5 summary is brief and incomplete, missing critical details. It captures some key points but lacks context and nuances. The Pegasus summary is irrelevant and unrelated to the dialogue content.

```
Dialogue 1:
Claire: <file_photo>
Kim: Looks delicious...
Linda: No way... Look what I'm cooking right now:
Linda: <file_photo>
Claire: hahahaha
Kim: Curry dream team
Claire: Enjoy your dinner :*

BART Summary:
Linda: No way... Look what I'm cooking right now. Claire: hahahaha   Kim: Curry dream team. Linda: What's for dinner? Claire: Curry. Kim:

T5 Summary:
Linda: 'look what i'm cooking right now: no way... ... look what I'm preparing right now'

PEGASUS Summary:
Claire:
```

5.2 After Fine-Tuning BART

BART was selected for fine-tuning after 2 epochs using a 14k training dataset. It can capture the essence of the dialogue, accurately reflecting negotiation dialogues in a detailed and fluent manner.

```
Dialogue 1:
Claire: <file_photo>
Kim: Looks delicious...
Linda: No way... Look what I'm cooking right now:
Linda: <file_photo>
Claire: hahahaha
Kim: Curry dream team
Claire: Enjoy your dinner :*

Summary:
Linda is cooking curry for dinner right now. Claire, Kim and Linda think it looks delicious, but Linda thinks it's a waste of time as she's a
```

6. Conclusion

Before fine-tuning, models often missed crucial details and failed to capture the essence of the dialogues. Summaries lacked coherence, with models like Pegasus generating incomplete or irrelevant outputs. BART introduces redundant phrases. Summaries were often disjointed and lacked fluency, especially for T5 and Pegasus.

After fine-tuning, the BART model is effectively improved and captured key points, accurately reflecting the main content of the dialogues. The fine-tuned BART model produced more coherent summaries with a logical flow and narrative structure, effectively summarizing the dialogues. The fine-tuned BART model generated fluent summaries with smooth transitions between points, making the summaries easier to read and understand. Fine-tuning significantly enhanced the summarization capabilities of the BART model.