Report on Hugging Face Transformers: SAMSum Corpus Summarization

Harshit Kumar

Introduction

In this project, we explored the capabilities of Hugging Face Transformers for a text summarization task using the SAMSum corpus. This dataset comprises approximately 16,000 messenger-like conversations with corresponding summaries, crafted to simulate real-life dialogues with varied topics and styles. Our goal was to understand the effectiveness of pre-trained models for summarization, fine-tune a model on this specific dataset, and evaluate the improvements achieved through fine-tuning.

## 1. Exploratory Data Analysis (EDA)

Our initial EDA focused on understanding the dataset's structure, specifically the length distribution of dialogues and summaries, lexical diversity, sentiment distribution, and thematic visuals through word clouds. Key findings include:

- **Data Distribution:** The data distribution of dataset is as follows:

  train: Dataset({

      features: ['id', 'dialogue', 'summary'],

      num_rows: 14732

  })

  test: Dataset({

      features: ['id', 'dialogue', 'summary'],

      num_rows: 819

  })

  validation: Dataset({

      features: ['id', 'dialogue', 'summary'],

      num_rows: 818

  })

})

- **Length Distribution**: Dialogues and summaries exhibit a broad range of lengths, with dialogues generally being much longer than summaries. This variance underscores the summarization task's challenge: condensing verbose dialogues into concise summaries without losing essential information.
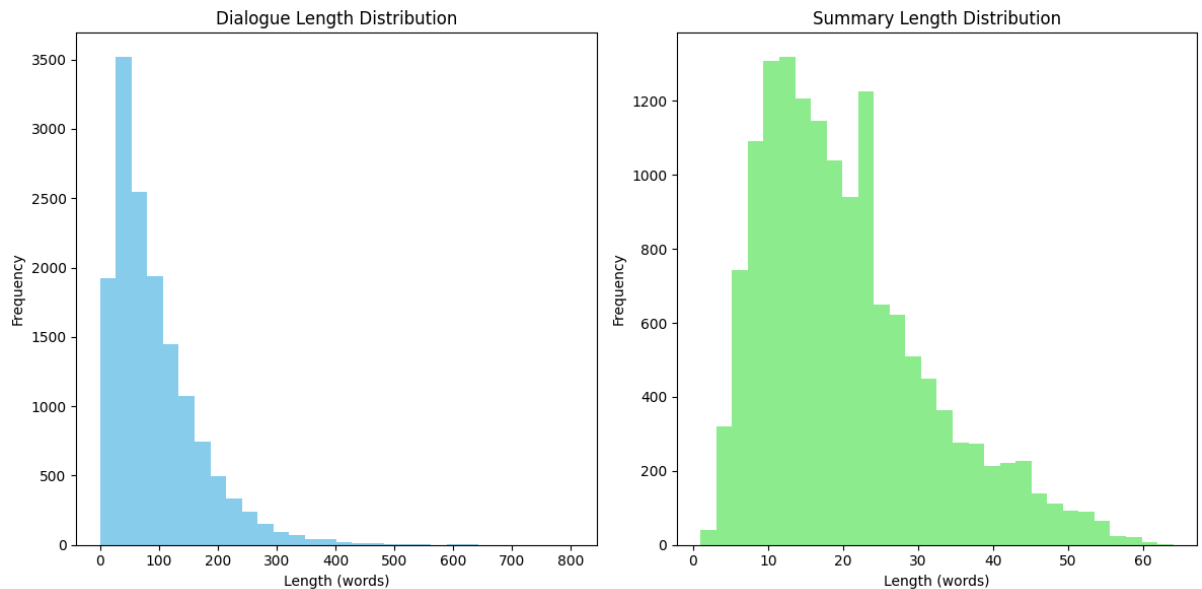
- **Lexical Diversity**: The dataset showcases a rich vocabulary across both dialogues and summaries, indicating the presence of diverse topics and conversational styles.

Lexical Diversity - Dialogues: 0.021, Summaries: 0.052

- **Sentiment Analysis**: Sentiment distribution across dialogues and summaries revealed a wide emotional range, with mostly positive sentiments.
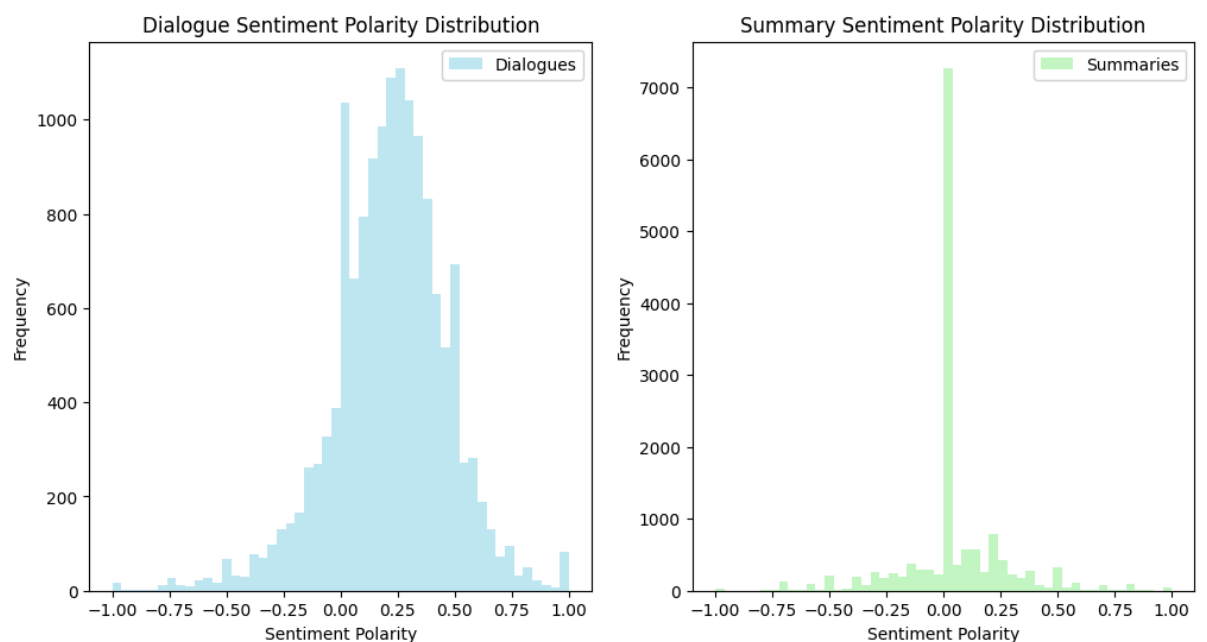
- **Word Clouds**: Visual analysis through word clouds highlighted frequent themes and terms used in dialogues and summaries, offering a glimpse into the common topics of conversation.

Dialogue WordCloud

Summary WordCloud

2. Inference with Pre-trained Models

We selected three pre-trained models from the Hugging Face model hub: facebook/bart-large-cnn, t5-large, and transformersbook/pegasus-samsum, to generate summaries for randomly selected dialogues.

Dialogue #1:

Cara: hey

Cara: are you at home

Celine: hey Cara

Celine: No i'm not

Cara: okay then, i just wanted to pass by

Celine: im sorry, i can drop by in the evening if you dont mind

Cara: its fine, call me then if you decide to come

Celine: ok

- **facebook/bart-large-cnn**: Cara calls her friend, Celine, and asks if she is at home. Celine says she is not, but can drop by in the evening. Cara says she will call her if she decides to come.
- **t5-large**: celine: i'm sorry, i can drop by in the evening if you dont mind Cara: its fine, call me then . cara: ok then, if i decide to come celin
- **transformersbook/pegasus-samsum:** Celine is at home. Cara wants to pass by in the evening. Celine will drop by in the evening if she doesn't mind. Cara will call her if she decides to come to her.

Dialogue #2

Dialogue: Tom: Ben. We've decided. 2pm in the Oval Room.

Ben: Ok, I'll be there

Tom: Take all your papers, it's going to be a fight! And remember: take no prisoners, shoot to kill!

Ben: hahaha, we have to win this battle.

Tom: We will, the justice is on our side.

- **facebook/bart-large-cnn**: Tom: Take all your papers, it's going to be a fight! And remember: take no prisoners, shoot to kill! Ben: hahaha, we have to win this battle. Tom: We will, the justice is
- **t5-large**: Ben: take no prisoners, shoot to kill! tom: we will, the justice is on our side . 'we have to win this battle!' - tom .
- **transformersbook/pegasus-samsum:** Tom and Ben are going to fight at 2 pm in the Oval Room. They will take Ben's papers and shoot to kill him. The justice is on Tom's side and they will win the battle.

Observations: Pegasus seems to be better consistently compared to other models, mostly because it's trained for summarization. However, as you can can se for dialogue #2, without

the context of video game, the model is considering the fight to be a real fight, which is probably expected. The BART and T5-large aren't always able to generate summary, they seem to be re-iterating the dialouges. This suggests that they are required to be fine-tuned. We'll see the effect of fine-tuning of bart-large-cnn next.

3. Fine-Tuning

The facebook/bart-large-cnn model was chosen for fine-tuning due to its promising balance between performance and computational efficiency. We preprocessed the SAMSum dataset to fit the input format required by BART, involving tokenization and padding. The fine-tuning process was monitored using the ROUGE metric to evaluate summarization quality against reference summaries.

Fine-tuning was performed over three epochs, with a learning rate of 2e-5 and a batch size of 4, adjusted based on the computational resources available. The model was trained with Mixed Precision (fp16). This process significantly improved the model's ability to generate accurate and relevant summaries, as evidenced by improved ROUGE scores on the validation set and test set.

Fine-tuning statistics

Epoch	Training Loss	Validation Loss	Rouge1	Rouge2	Rougel	Rougelsum
1	1.373600	1.404603	40.818900	20.921200	31.640100	38.161700
2	1.010000	1.408124	41.251100	21.182300	31.981200	38.277600
3	0.748200	1.502978	41.137400	20.917000	31.562600	38.163800

4. Evaluation and Analysis

Post fine-tuning, the model's performance was evaluated on the test set, showcasing notable improvements in ROUGE scores compared to its pre-fine-tuning state. This improvement was particularly evident in the model's enhanced ability to capture key points, coherence, and

fluency in generated summaries. Through fine-tuning, the model became better attuned to the dataset's nuances, demonstrating the value of task-specific model optimization.

Results on test-set:

- **Before fine-tuning**:

  'eval_rouge1': 30.8301, 'eval_rouge2': 10.2689, 'eval_rougeL': 23.1225, 'eval_rougeLsum': 28.7233

- **After fine-tuning:**

  'eval_rouge1': 40.1821, 'eval_rouge2': 20.0727, 'eval_rougeL': 30.99, 'eval_rougeLsum': 37.4897

Dialogue #1:

Cara: hey

Cara: are you at home

Celine: hey Cara

Celine: No i'm not

Cara: okay then, i just wanted to pass by

Celine: im sorry, i can drop by in the evening if you dont mind

Cara: its fine, call me then if you decide to come

Celine: ok

- **Before fine-tuning**: Cara calls her friend, Celine, and asks if she is at home. Celine says she is not, but can drop by in the evening. Cara says she will call her if she decides to come.
- **After fine-tuning:** Cara wants to visit Celine, but Celine is not at home. Celine will call Cara if she decides to come over in the evening.   She will let Cara know.

Dialogue #2

Dialogue: Tom: Ben. We've decided. 2pm in the Oval Room.

Ben: Ok, I'll be there

Tom: Take all your papers, it's going to be a fight! And remember: take no prisoners, shoot to kill!

Ben: hahaha, we have to win this battle.

Tom: We will, the justice is on our side.

- **Before fine-tuning**: Tom: Take all your papers, it's going to be a fight! And remember: take no prisoners, shoot to kill! Ben: hahaha, we have to win this battle. Tom: We will, the justice is
- **After fine-tuning:** Tom and Ben are going to fight at 2 pm in the Oval Room. Ben will take all his papers and remember to take no prisoners, shoot to kill.   Tom and Ben will win the fight.

Dialogue #3

Dialogue: Josh: Going to the pub tonight?

Sean: sure, pick up some chicks!

Logan: Please, behave Sean, I actually would like to meet some girls

Sean: ??

Logan: don't bullshit around with you sexist comments, it's counterproductive

- **Before fine-tuning**: Sean: Sure, pick up some chicks! Logan: Please, behave Sean, I actually would like to meet some girls. Josh: Going to the pub tonight? Sean: sure,Pick up some girls! Josh: Go to the
- **After fine-tuning:** Josh and Sean are going to the pub tonight. Sean wants to pick up some girls. Logan doesn't want to go with them because he would like to meet some girls himself.

Observations: The summary results have improved significantly after fine-tuning. For some sentences, before fine-tuning, the model wasn't able to summarize, it was just reiterating the dialogues. However, after fine-tuning, it can produce summary, which is coherent.

Conclusion

The project underscores the effectiveness of Hugging Face Transformers in handling complex NLP tasks like summarization. Through careful selection, fine-tuning, and evaluation of pre-trained models, we achieved significant improvements in summary quality. This experiment highlights the importance of dataset-specific tuning in unlocking the full potential of pre-trained NLP models.

Future Work

Future exploration could include experimenting with different models, further hyperparameter optimization, and investigating the impact of additional pre-processing steps. Additionally, exploring other metrics for summarization quality and incorporating feedback loops for model improvement could yield even better results.