Fine-tuning Large Language Models for Text Summarization

Submitted By - Ritvik Garg (ritik.anselm@gmail.com)

Objective

The objective of this assignment is to assess your skills in fine-tuning large language models for the task of text summarization. You will be working with a pre-trained language model and a dataset of news articles, and your task is to develop a fine-tuned model that can generate concise and accurate summaries of the articles.

Dataset

BBC news summary dataset: https://www.kaggle.com/datasets/pariza/bbc-news-summary

Dataset	# of samples
Train	1557
Test	445
Validation	223
Total	2225

Base-model Used

T5-small : https://huggingface.co/t5-small

Evaluation metric

ROUGE - https://huggingface.co/spaces/evaluate-metric/rouge

ROUGE (Recall-Oriented Understudy for Gisting Evaluation), is a set of metrics and a software package specifically designed for evaluating automatic summarization, but that can be also used for machine translation. The metrics compare an automatically produced summary or translation against reference (high-quality and human-produced) summaries or translations.

ROUGE compares the n-grams of the generated summary to the n-grams of the references. There are several types of ROUGE metrics, such as ROUGE-1, ROUGE-2, ROUGE-L, etc.

ROUGE-N measures the number of matching n-grams between the model-generated text and a human-produced reference.

ROUGE-L is based on the longest common subsequence (LCS) between our model output and reference, i.e. the longest sequence of words (not necessarily consecutive, but still in order) that is shared between both.

Steps followed to fine-tune pretrained model

1. Created csv file containing all articles and summaries at one place, using `create_df.py` file
2. Splitted the dataset into train, test and validation in the ratio of 70:20:10
3. Train with PyTorch Trainer
4. Save the best model
5. Load the fine-tuned model, and do inferencing over it.

Model Performance

```
# train on complete data
trainer.train()
```

[3900/3900 16:50, Epoch 10/10]

Epoch	Training Loss	Validation Loss	Rouge1	Rouge2	Rougel	Rougelsum	Gen Len
1	No log	0.752737	0.177500	0.118700	0.158100	0.157900	18.062800
2	1.110100	0.682200	0.178500	0.125200	0.162100	0.162600	17.466400
3	0.739200	0.653441	0.193100	0.138500	0.177100	0.177300	18.574000
4	0.676600	0.635428	0.198500	0.145600	0.184000	0.184400	18.914800
5	0.676600	0.624578	0.200100	0.148600	0.185900	0.185700	18.914800
6	0.640800	0.617297	0.207200	0.158400	0.194100	0.194000	19.000000
7	0.622500	0.612026	0.209500	0.160400	0.195900	0.196200	19.000000
8	0.617400	0.609243	0.211100	0.164000	0.197900	0.198200	19.000000
9	0.598800	0.607350	0.215600	0.169200	0.202600	0.202800	19.000000
10	0.598800	0.606775	0.215900	0.169300	0.202900	0.203100	19.000000

Evaluating model on Validation dataset :

{'eval_loss': 0.6067754030227661, 'eval_rouge1': 0.2159, 'eval_rouge2': 0.1693, 'eval_rougeL': 0.2029, 'eval_rougeLsum': 0.2031, 'eval_gen_len': 19.0, 'eval_runtime': 19.8844, 'eval_samples_per_second': 11.215, 'eval_steps_per_second': 2.816, 'epoch': 10.0}

Evaluating model on Testing dataset :

{'eval_loss': 0.5863471627235413, 'eval_rouge1': 0.2141, 'eval_rouge2': 0.1645, 'eval_rougeL': 0.2007, 'eval_rougeLsum': 0.2006, 'eval_gen_len': 19.0, 'eval_runtime': 48.0938, 'eval_samples_per_second': 9.253, 'eval_steps_per_second': 2.329, 'epoch': 10.0}
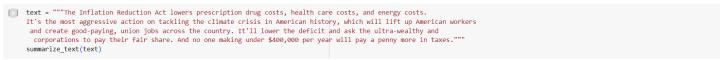
Challenges

1. Training on a big dataset takes a lot of time and computational resources of Colab.
2. While fine-tuning, some T5-based models were giving "No log" in Traiaing loss and "NaN" in Validation loss. **Solution :> make fp16 = False**
3. Cuda- OOM error. **Solution : Use a smaller batch size while training.**

Model Inferencing

- Pass your article text into `summarize_text` function .

Sample Result

Inferencing on Finetuned model

```
text = """The Inflation Reduction Act lowers prescription drug costs, health care costs, and energy costs.
It's the most aggressive action on tackling the climate crisis in American history, which will lift up American workers
 and create good-paying, union jobs across the country. It'll lower the deficit and ask the ultra-wealthy and
  corporations to pay their fair share. And no one making under $400,000 per year will pay a penny more in taxes."""
summarize_text(text)
```

'The Inflation Reduction Act lowers prescription drug costs, health care costs, and energy costs.It'll lower the deficit and ask the ultra-wealthy and corporations to pay their fair share.It'll ask the ultra-wealthy and corporations to pay their fair share.'