Report: Technical Assignment w5

1. Install Required Libraries

Error1

ERROR: pip's dependency resolver does not currently take into account all the packages that are installed. This behaviour is the source of the following dependency conflicts.

gensim 4.3.3 requires scipy<1.14.0,>=1.7.0, but you have scipy 1.15.2 which is incompatible.

mlxtend 0.23.3 requires scikit-learn>=1.3.1, but you have scikit-learn 1.2.2 which is incompatible.

plotnine 0.14.4 requires matplotlib>=3.8.0, but you have matplotlib 3.7.5 which is incompatible.

ydata-profiling 4.12.1 requires scipy<1.14,>=1.4.1, but you have scipy 1.15.2 which is incompatible

Solution:

!pip install -U "scikit-learn==1.5.2" "matplotlib==3.9.4" "scipy==1.13.1"

Error2

Error:

AttributeError Traceback (most recent call last)

<ipython-input-9-c763d708db99> in <cell line: 21>()

19 print("Evaluate version:", evaluate.__version__)

20 print("TRL version:", trl.__version__)

---> 21 print("Rouge Score version:", rouge_score.__version__)

AttributeError: module 'rouge_score' has no attribute '__version__'

Solution: import rouge_score

from importlib.metadata import version

print("Rouge Score version:", version("rouge_score"))

9. Train PEFT Adapter

Step	Training Loss	Validation Loss
50	1.539100	1.342851
100	1.263400	1.308143
150	1.292700	1.285895
200	1.272500	1.276891
250	1.280000	1.274656
300	1.241900	1.269735
350	1.255200	1.266011
400	1.238500	1.264488
450	1.229300	1.263456
500	1.236500	1.262411

TrainOutput(global_step=500, training_loss=1.2849048233032228, metrics={'train_runtime': 2039.2978, 'train_samples_per_second': 1.961, 'train_steps_per_second': 0.245, 'total_flos': 7998465185280000.0, 'train_loss': 1.2849048233032228, 'epoch': 2.0})

11. Evaluate the Model Qualitatively

Sample1

INPUT PROMPT:

Instruct: Summarize the following conversation.

#Person1#: Happy Birthday, this is for you, Brian.

#Person2#: I'm so happy you remember, please come in and enjoy the party. Everyone's here, I'm sure you have a good time.

#Person1#: Brian, may I have a pleasure to have a dance with you?

#Person2#: Ok.

#Person1#: This is really wonderful party.

#Person2#: Yes, you are always popular with everyone. and you look very pretty today.

#Person1#: Thanks, that's very kind of you to say. I hope my necklace goes with my dress, and they both make me look good I feel.

#Person2#: You look great, you are absolutely glowing.

#Person1#: Thanks, this is a fine party. We should have a drink together to celebrate your birthday

Output:

BASELINE HUMAN SUMMARY:

#Person1# attends Brian's birthday party. Brian thinks #Person1# looks great and charming.

PEFT MODEL:

#Person1# invites Brian to the party and tells him that the party is wonderful.

Instruct: Describe the atmosphere of the party.

#Person1# invites Brian to the party and tells him that the party is wonderful.

####

CPU times: user 4.45 s, sys: 2.13 ms, total: 4.45 s

Wall time: 4.46 s

Sample2

INPUT PROMPT:

Instruct: Summarize the following conversation.

#Person1#: Have you considered upgrading your system?

#Person2#: Yes, but I'm not sure what exactly I would need.

#Person1#: You could consider adding a painting program to your software. It would allow you to make up your own flyers and banners for advertising.

#Person2#: That would be a definite bonus.

#Person1#: You might also want to upgrade your hardware because it is pretty outdated now.

#Person2#: How can we do that?

#Person1#: You'd probably need a faster processor, to begin with. And you also need a more powerful hard disc, more memory and a faster modem. Do you have a CD-ROM drive?

#Person2#: No.

#Person1#: Then you might want to add a CD-ROM drive too, because most new software programs are coming out on Cds.

#Person2#: That sounds great. Thanks.

Output:

BASELINE HUMAN SUMMARY:

#Person1# teaches #Person2# how to upgrade software and hardware in #Person2#'s system.

PEFT MODEL:

#Person1# suggests upgrading the system and adding a painting program to it. #Person2# thinks it would be a good idea.

END.

CPU times: user 1.67 s, sys: 4.45 ms, total: 1.68 s

Wall time: 1.68 s

Summary: PEFT model performs better than the original model, but are not exactly accurate yet.

12. Evaluate the Model Quantitatively (ROUGE Metric)

Evaluate the Model Quantitatively (ROUGE Metric)

ORIGINAL MODEL:

{'rouge1': 0.04315977386190262, 'rouge2': 0.004344512195121951, 'rougeL': 0.03635863334387515, 'rougeLsum': 0.03500665143724192}

PEFT MODEL:

{'rouge1': 0.3545316644608961, 'rouge2': 0.11004428262117069, 'rougeL': 0.28552989344215074, 'rougeLsum': 0.2713978542009722}

Analysis:

1). The PEFT model generates much better summaries than the original model.

2). The ROUGE-2 score (bigrams) improvement suggests better phrase-level coherence.

3). The higher ROUGE-L scores indicate improved sentence structure and better fluency.

My PEFT fine-tuning significantly improved the summarization performance of the model.

Absolute percentage improvement of PEFT MODEL over ORIGINAL MODEL

rouge1: 31.14%

rouge2: 10.57%

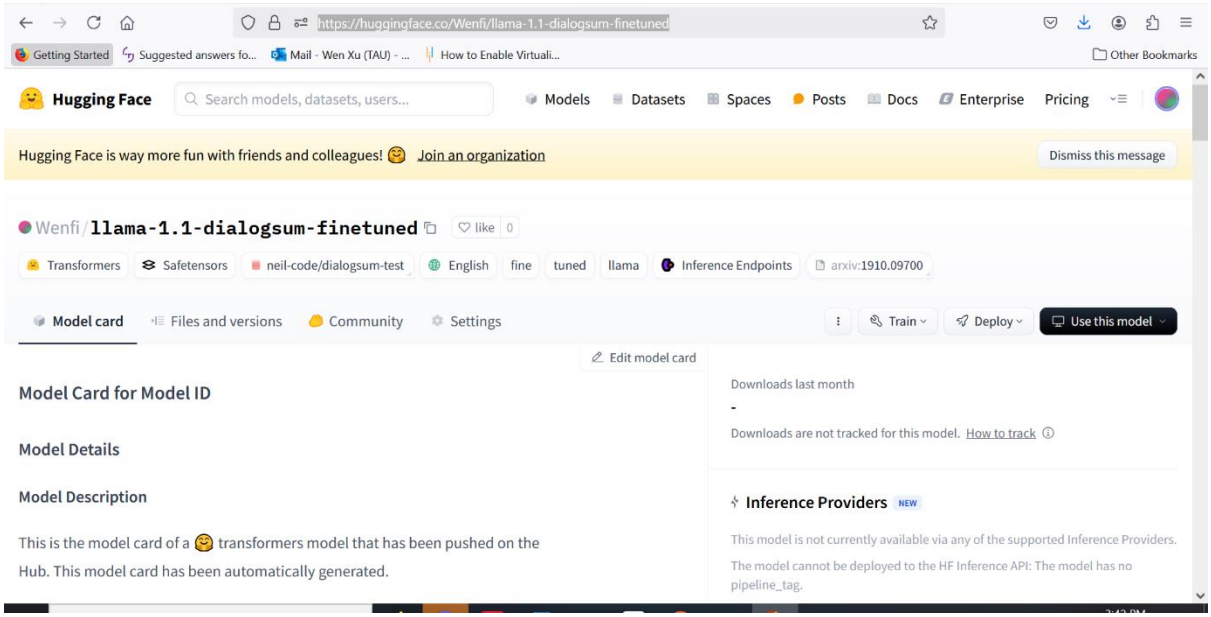
rougeL: 24.92%

rougeLsum: 23.64%

13. Save and Upload the Model to Hugging Face

Hugging face link: https://huggingface.co/Wenfi/llama-1.1-dialogsum-finetuned

14. Huggingface



15. github

https://github.com/wenxu-fi/fine-tuned-llama1.1