[SWE3052-41] Project #2: Large Language models

Release day: May 14nd, 2024

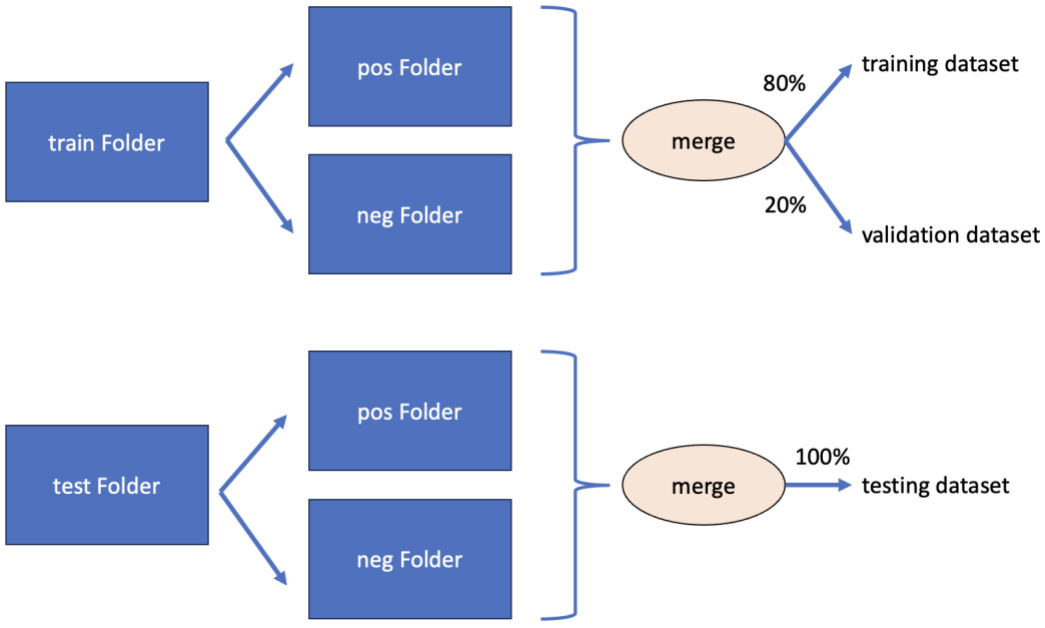
Due: 06/05/2024 (Hard deadline: no extension day is allowed)

Objective:
Your task is to generate a set of movie reviews using the pre-trained GPT-2 model and then classify their sentiment (positive or negative) using a fine-tuned BERT model.

Tasks:

0. Install the necessary libraries: Install HuggingFace Transformers, TensorFlow, and PyTorch, which are essential for this homework.
1. Generate movie reviews
   a. Load the pre-trained GPT-2 model ("heegyu/gpt2-emotion") from the HuggingFace and a tokenizer from the Hugging Face Transformers library like
   tokenizer = GPT2Tokenizer.from_pretrained('heegyu/gpt2-emotion')

   You should use the Dataloader of Pytorch (i.e., torch.utils.data.DataLoader) when loading your data. Please use a movie review dataset from the link below and randomly divide the data into these proportions: Training, Validation, and Testing.
   - Dataset Link: https://ai.stanford.edu/~amaas/data/sentiment/. Please follow the following procedure:

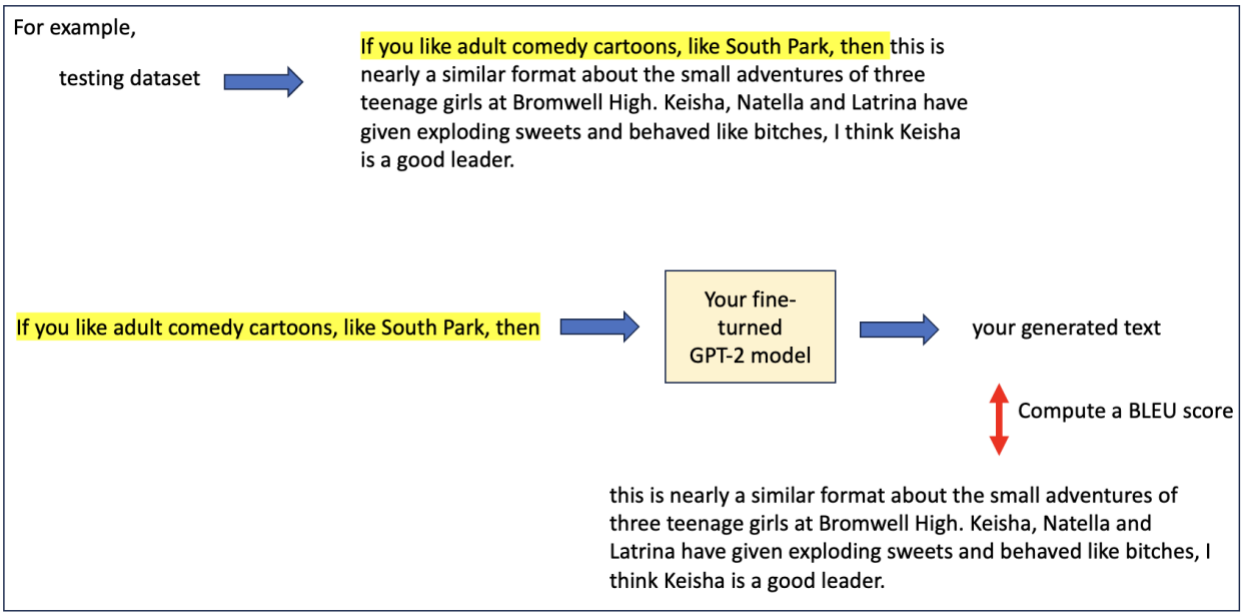


b. Fine-tune GPT-2 using a movie review dataset (IMDB) to generate more realistic movie reviews. Please define your own loss function in your report. You can use any metric or loss functions for leveraging training and validatation datasets.

c. Create a set of prompts (e.g., "This movie was", "The actors in the film") to guide the generation of movie reviews. You can choose any prompts from your own criteria.

d. Generate and report 30 movie reviews using the GPT-2 model and the prompts. Save the generated reviews as a list or a text file. **(10 points)**

e. Evaluate your model using a BLEU metric https://huggingface.co/spaces/evaluate-metric/bleu with your testing dataset. Please report the mean value of Bleu Scores from using the testing dataset. **(10 points)**
- Please use the first 10 words of a review as prompts for your model and compare the generated text with the true text to compute the BLEU score.



2. Sentiment classification with BERT:
a. Load the pre-trained BERT model ('textattack/bert-base-uncased-SST-2') for sentiment analysis and its tokenizer from the Hugging Face Transformers library.
- Please refer to https://huggingface.co/textattack/bert-base-uncased-SST-2
or https://github.com/QData/TextAttack
b. Simply use the pre-trained mode without additional training and classify the sentiment about movie review files generated by **Task 1-d**. In other words, BERT classifies the results generated from GPT-2. Please use accuracy for your evaluation metric, and true answers are from your own decision. **(10 points)**

c. Analysis and visualization:
   - Calculate the proportion of positive and negative reviews from **Task 2-b**. **(5 points)**
   - Discuss three interesting examples of generated reviews and their sentiment classification. **(5 points)**

Please submit your code as a Jupyter Notebook or a Python script, along with a report.

Reference

- Tutorial about how to use HuggingFace's **transformers** library(https://huggingface.co/docs/transformers/index)
- https://huggingface.co/gpt2?text=A+long+time+ago%2C
- https://huggingface.co/docs/transformers/main_classes/tokenizer
- https://pytorch.org/docs/stable/data.html