

Assignment 3: Pretrained Language Models (15 points)

CS 410/510 Natural Language Processing, Fall 2023

In this assignment, you will explore two pretrained language models (PLMs) - BERT [1] and GPT-2 [2], for text classification. The goal is to build yet another text classifier that will predict whether a piece of text is "positive" or "negative", but this time, instead of engineering the features (unigrams, bigrams, sentiment words, etc.), you will use pretrained language models.

1 Instructions

(same as last assignment) Data Preparation: You are provided with a dataset containing customer reviews of products. The dataset is the "Multilingual Amazon Reviews Corpus" [1] in json format. It includes several columns. For this assignment, we will be using a small subset of the original dataset and are concerned with the following two columns:

- (a) "review_title" (the title of the review), and
- (b) "stars" (an integer 1 or 5 indicating the number of stars, which is considered as the sentiment label, with 1 indicating "negative" and 5 indicating "positive").

Warning: *As is the case with most 'natural language' text, this data was collected from public websites and is mostly unfiltered. Therefore, it is possible that some text may be disturbing or you may not agree with it.*

- The dataset is pre-split into a training set and a test set.
- Load the dataset using Python and the appropriate library (e.g., pandas).
- **No preprocessing needed this time. Instead, you will use the PLM's tokenizer to prepare your data.**

Pretrained Language Models for Text Classification: As part of this assignment, you will experiment with two specific PLMs: BERT [2] and OpenAI's GPT-2 [3]. These can be accessed using the HuggingFace library¹. Prepare your dataset and fine-tune the PLMs using the training set. Evaluate the model's performance on the test set using appropriate evaluation metrics (e.g., precision, recall, F1-score).

¹ <https://huggingface.co/>

Results and Analysis: Provide a detailed analysis of your model's performance by comparing the output of the two PLMs, as well as how the results compare with your previous assignments. Include the following:

- F1-score and other relevant metrics.
- Any observed challenges or limitations.

2 Submission Guidelines and Grading

Submit a PDF of your Colab notebook on Canvas, under "Assignment 3". Include any additional files or resources used. Ensure your code is well-documented and organized.

Your assignment will be assessed based on the following criteria:

- implementation of the text classification model (7 points),
- model evaluation, analysis, and discussion (6 points),
- and overall clarity and organization of the assignment (2 points).

Good luck!

References

[1] Keung, Phillip, Yichao Lu, György Szarvas, and Noah A. Smith. "The multilingual amazon reviews corpus." arXiv preprint arXiv:2010.02573 (2020).

[2] Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

[3] Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. "Language models are unsupervised multitask learners." OpenAI blog 1, no. 8 (2019).