SI630 Assignment 4

Name: Junqi Chen

UMID: 03846505

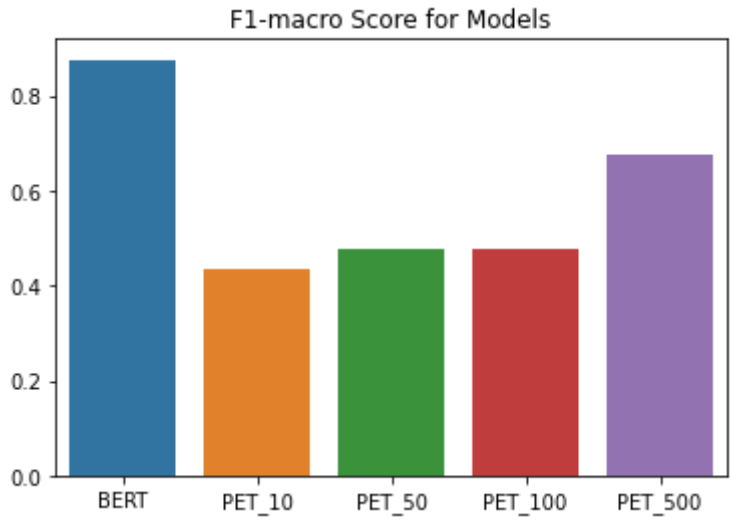
Uniqname: junqich

This is the write-up file for SI630 Assignment 4.

Problem 5

Score Plot

The f1-macro score plots for different models I use is shown below:



From the plot above, we can see that the performance of PET models becomes better as the training set becomes larger. However, it can still hardly have the equal performance comparing to the BERT model.

Instance Estimation

As the size training data becomes larger, the performance of the PET will greatly increase. From my point of view, I think about 2,000 entries of dataset will make PET model have a close performance comparing to the BERT model.

Problem 1-4

Please refer to `source_code_all_junqich.zip` for more details.

For previous Problems:

- Please see Problem 1, 3 in file `SI630_Assignment4_BERT_junqich.ipynb` under directory `\BERT`.

- Please see Problem 2, 4 in file `SI630_Assignment4_PET_junqich.ipynb` under directory `\PET`, and the modified `tasks.py`, `pvp.py` source code from the PET under directory `\PET\pet`.

For the training results:

- Please see the result under directories `\hw4_output_X`, where X is the size of training set.