

Checkpoint for final project

James Agresto, Lorenzo Ashurst

On First Download:

Unzip the folder that was submitted to canvas. This should contain all the code necessary for running/training our model as well as testing it

How to train the model:

Modify the following hyperparameters to your choosing located at the top of the “embedding_model.py” file.

- BATCH_SIZE
- CONTEXT_SIZE
- D_EMBEDDINGS
- DATA_SET_SIZE

After you initialize the hyperparameters to your liking run the following script in the directory that you unzipped the file into

- ./run.sh

How to download pretrained models

Go to the following link and click download:

<https://drive.google.com/file/d/1stM30496SNIPcPVaXG-9234Weu-soRvb/view?usp=sharing>

Once the .pt file has been downloaded, put that file into the same directory that you unzipped.

How to Test:

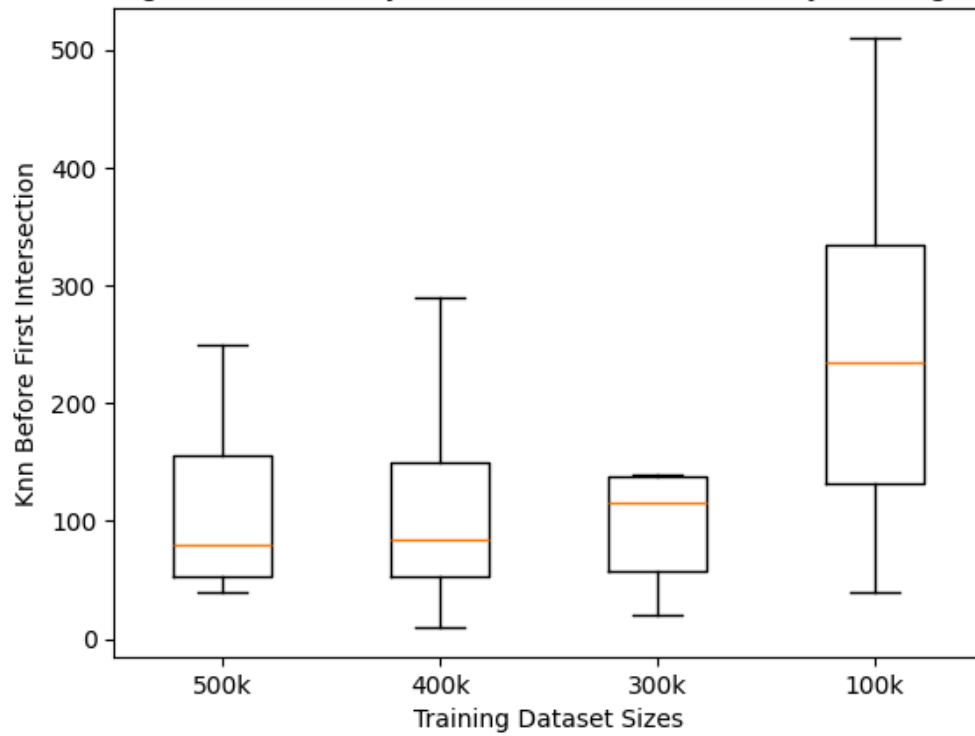
Simply run the following bash file

- ./runtests.sh

If you wish to change what model to test on you can change the MODEL_FILE_NAME global variable in testing_embeddings.py

The outputs for the tests should automatically be created in a text file called tests-outputs.txt

K Nearest Neighbors Necessary for Non-Zero Intersection by Training Dataset Sizes



Addendum: We were unable to finish running a 200k line dataset model, thus the gap in hyperparameters.