Joe Gelsomini
Scott Gordon

NLP Project Contributions

Submitted are several Jupyter notebooks each containing the individual tests we ran. Each notebook contains mostly duplicate code. A lot of the code from the notebooks is taken from what we wrote and what was given to us in homework 4.

Joe Gelsomini:

Wrote the original train_full.ipynb. A lot of this code as stated was taken from our implementations in homework 4. We used my HW4 submission as the basis. The code at the beginning was also written by me to extract the relevant columns of the csv files to get the data into lists. I also borrowed a method to shuffle a slice of a list from stack overflow. This was the code we used to shuffle only half of the endings in the training set. Also rewrote the implementation of the GRU into an LSTM. There is code that can be uncommented out to run as bidirectional. I cloned the Bert directory onto my directory in the dandeneau labs, and created the tsv files found in the Bert data folder in the github. there is also another notebook which runs the evaluation on the bert output.

Scott Gordon:

Used train_full.ipynb as a basis. modified the notebook in train_end.ipynb so that the model would only train on story endings. I downloaded the 300 dimensional GLOVE embeddings from the following link:
https://github.com/stanfordnlp/GloVe
From there you need to download the glove.6b.zip, and extract the 300 dimensional glove embeddings to the same directory as the notebooks labeled GLOVE. Also wrote the code to initialize the embedding matrix in the original train_full_glove.ipynb. This code was reused for all of our other models.