

Final Project Progress Report - Text Classification Competition

Name: Junzhe Sun

NetID: junzhes2

Captain: Junzhe Sun (individual)

Competition: text classification competition

Progress:

For the text classification competition, I looked at a few different options and decided to use a pre-trained BERT model with fine tuning.

1) Which tasks have been completed?

A: I looked at different shallow/deep learning methods, including FastText, LSTM and Transformer neural networks. I first trained a FastText model, a shallow neural network, using the training data set, but wasn't able to beat the baseline. Then I used a pre-trained BERT model from huggingface using PyTorch and fine tuned it using the training data set. The prediction using the BERT model was able to beat the baseline.

2) Which tasks are pending?

A: (1) Further fine-tuning the BERT model to improve F1 score. (2) Source code documentation. (3) Software code submission with documentation/report. (4) Record software usage tutorial presentation.

3) Are you facing any challenges?

A: No big challenge remains. A minor challenge is that, since the training data set is not very big, splitting it into training and validation sets can be tricky. In addition, I will probably need a more automated and systematic way for hyperparameter tuning (train-test split, number of epochs, rate for drop-out layer, etc.).