E0-334 Deep Learning for NLP

Assignment 3

(due by 1st Sept, 11:59 PM)

Note: Use the following link for submitting your results.

https://forms.office.com/r/eCiCn7gxst

Problem:

Consider the sentence polarity dataset available at the following site:

https://www.cs.cornell.edu/people/pabo/movie-review-data/rt-polaritydata.tar.gz

The dataset contains 5331 positive and 5331 negative examples in two different files. We will use the last 831 examples from each file as a test set (Total test set size 1662).

In this assignment, we will study the dynamic meta-embedding idea discussed in [1] for sentence representation. Given multiple types of word embeddings (word2vec, GLoVe and FastText), the aim of this work is to learn which embeddings the network prefers for a given problem by predicting a weight for each embedding type. Using this dynamic meta-embedding idea for sentence representation, design a sentence classifier (using RNN/GRU/LSTM) for the dataset mentioned above.

Report the average accuracy on the test set (in the scale $(0, 1)$).

Compare the performance of this approach with that which does not use meta-embedding idea. You can use any of the word embeddings for this purpose. Use RNN/GRU/LSTM for sentence representation. Report the average accuracy on the test set (in the scale $(0, 1)$).

Reference:

1. Kiela *et al*, Dynamic Meta-Embeddings for Improved Sentence Representations, EMNLP, 2018.