Human Language Technology Homework 4 README

Mou Zhang

November 25, 2019

1 Question 1 Reporting

0 The notebook address is https://drive.google.com/open?id=1FQ0fRZ5gF41fJ1h5HafDcRYu1Hh8dKkj

1 The program uses 4 params, they are
vocab_size=len(train_corpus.vocab),
embedding_size=1024,
hidden_size=1024,
num_layers=2

2 Training loss is 0.0564, vslidation loss is 1.1353

3 My implementation gets training and validation accuracy of 97.92% and 75.50%

4 The number of params used by the BERT based model is 0

5 My implementation attains 0.2206, 0.3239 for training and validation loss at the end of 3 epochs.

6 My model attains 91.35%, 86.57% training and validation accuracy at the end of 3 epochs.

7 Test Prediction Accuracy for the LSTM based model is 77.16%

8 Test Prediction Accuracy for the BERT based model is 80.08%

2 Question 1 Answering

1 Word2vec provides a vector for each token/word and those vectors encode the meaning of the word. But the problem with word2vec is that each word has only one vector but in the real world each word has different meaning depending on the context and sometimes the meaning can e totally different. Bert deals with this problem by consider the context and set a vector for every token.

2 I can barely get read a pre-trained model instead of training word2vec with the given data. Therefore ,when I need to use vector for representation of word, I gonna use pre-trained word2vec to do the job in LSTM.

3 No. Because that word embedding only consider 1 word as 1 meaning, and Bert is better than that. This will not do any help cause word embedding is worse than Bert, so it will only do harm to the BERT.

4 Yes. The first token of every input sequence is the special classification token – [CLS]. This token is used in classification tasks as an aggregate of the entire sequence representation. It is ignored in non-classification tasks. So even without the pre-trained BERT, I can still use [CLS] to build a sentence similarity classifier.