

Intro to NLP

Devesh Marwah (2020115005)

February 2023

Pages: 2

1 Analysis

1.1 Results

The results for all the scores are as follows:

Accuracy: 0.9336723815329112

F1 Score: 0.9330802475371537

Precision: 0.9339073677442152

Recall: 0.9444480488699769

All of the following are done using the average measure as 'macro'. If used another measure like weighted, the f1 scores increase but on reading here, I observed macro to be a better measure as micro will give same values overall.

1.2 Hyper Parameters

in_dim=len(sentence_vocab)

mid_dim=100

lstm_dim=128

out_dim=len(tag_vocab)

n_layers=4

dropout_num=0.15

PAD_IDX=0

Epochs:10

Lr: 0.01

weight_decay: 1e-4

1.3 Analysis

On trying various hyper parameters, the results altered greatly. For increasing learning rate 0.02, the accuracy dropped to 87% and increasing it drastically reduced the accuracy. This is as expected because the LSTM will not be able to descent properly on a higher learning rate.

Given the smaller changes to the weights made with each update, smaller learning rates require more training epochs, whereas larger learning rates provide quick changes and require fewer training epochs. As a result, as we transitioned from version 1 to version 3, the number of epochs was raised from 5 to 10.

Similarly the layer parameters were decided, on increasing them the model complexity increased and it started over fitting hence I added a small dropout for it to work. Dropout works by randomly dropping layer weights with a specific probability and moves for few nodes only, preventing overfitting.

While the embedding size informs you of the size of your feature vector, the hidden dimension is essentially the number of nodes in each layer (the model uses embedded words as input). Better if these dimensions were smaller (but require more computing resources).

The given dataset is not so good hence performance of 93.4% accuracy on test set can also be considered good for a multiclassification type kind of model.