Refael Shaked Greenfeld 305030868

Part 2:

Parameters:

1. Embedding dimension: 128
2. RNN type: LSTM
3. hidden dimension: 128 (* 2 because it is Bi-directional)
4. Directions: 2
5. Number Of Layers: 2
6. Batch size: 4
7. Optimizer: AdamW
8. Learning rate: 0.003
9. Dropout P: 0.3
10. Attention: 256*256 (Matrix W)

Score on the test: 91.44

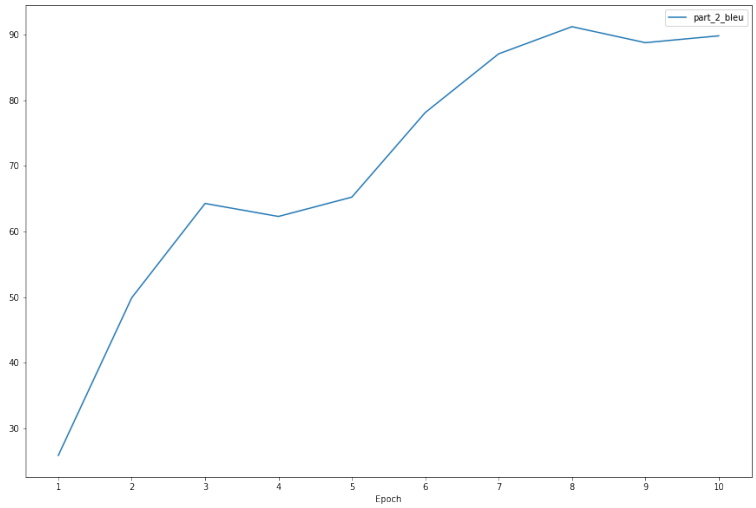
Best score on Dev: 91.27

General Questions:

The model with the attention preform significantly better than the “Vanilla” one. When it manage to outperform it in about 30 BLEU points. Both on the Dev set and both on the test set. Moreover, the attention model was able to get to over after only 3 apochs. Regarding the training time, my models were a bit different (I used for the Vanilla RNN a 3 layers bi-directional LSTM when in the Attention one I used 2 layers LSTM) but when I compared apple-to-apple i.e. using the same network exactly with the difference of attention, Attention Net took for the evaluation: 1.997 seconds and for training 50 seconds per epoch when the Vanilla one took 1.236 seconds for evaluation and 45 seconds per epoch (on CPU). So, as expected the Attention Net took more time in a about 10% in training time and significant 60% for evaluation.

Refael Shaked Greenfeld 305030868

BLUE score:



Evaluation Loss: