Lea Setruk
Yoel Benabou

Deep Learning - 89687

Ass3 – Part 3

Our parameters:

On NER the loss on 'O' tag was weighted down to be 30% of the training loss over this tag.

Transducer a:

Pos:
-hidden dimension = 256
-embedding_size = 2000
-batch_size = 100 (and 1 if predict)
-learning_rate = 0.0003

Ner:
-hidden dimension = 500
-embedding_size = 200
-batch_size = 128 (and 1 if predict)
-learning_rate = 0.0007

Transducer b:

Pos:
-hidden dimension = 120
-embedding_size = 80
-batch_size = 8 (and 1 if predict)
-learning_rate = 0.003
-btw_rnns = 200

Ner:
-hidden dimension = 250
-embedding_size = 80
-batch_size = 80 (and 1 if predict)
-learning_rate = 0.003
-btw_rnns = 300

Transducer c:

Pos:
-hidden dimension = 128
-embedding_size = 30
-batch_size = 16 (and 1 if predict)
-learning_rate = 0.007

Ner:

Lea Setruk
Yoel Benabou

-hidden dimension = 250
-embedding_size = 2000
-batch_size = 128 (and 1 if predict)
-learning_rate = 0.0007

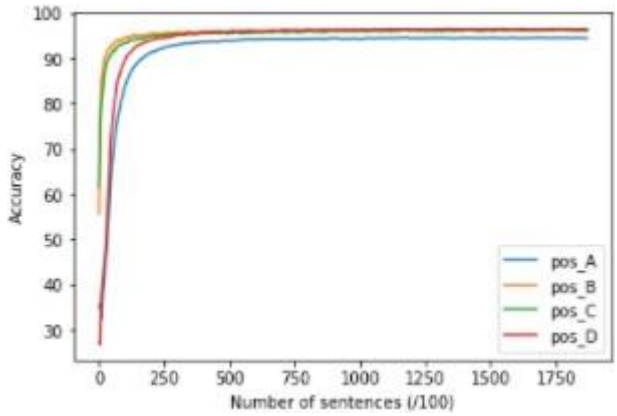
Transducer d:

Pos:
-hidden dimension = (256, 120)
-embedding_size = (2000, 80)
-batch_size = 64 (and 1 if predict)
-learning_rate = 0.0003
-btw_rnns = 200

Ner:
-hidden dimension = (200, 250)
-embedding_size = (1000, 80)
-batch_size = 64 (and 1 if predict)
-learning_rate = 0.0003
-btw_rnns = 500

Graphs

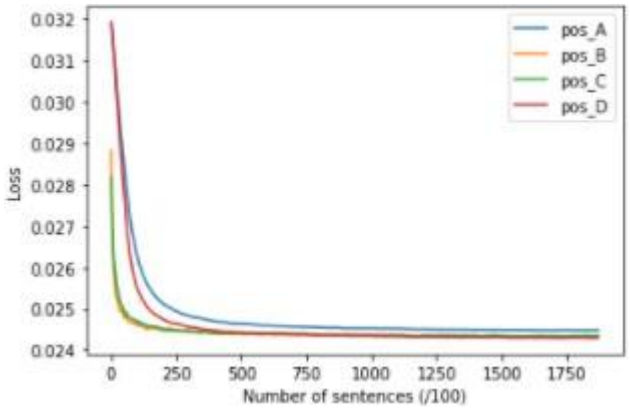
Pos - Accuracy:



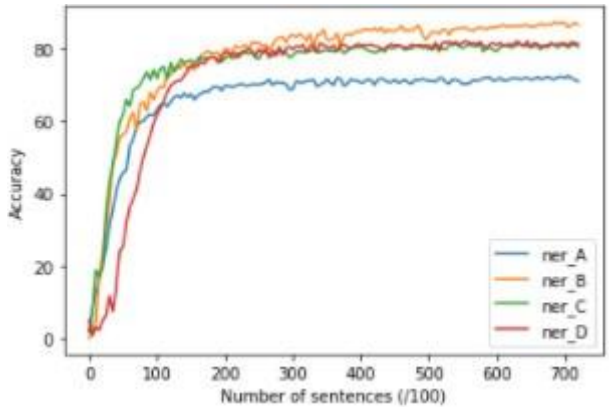
The best model seems to be D.

Pos – Loss:

Lea Setruk
Yoel Benabou



Ner – Accuracy:



The best model is B.

Ner – Loss: