Dafna Magid 315635870 Oved Nagar 302824875

The parameters we used to train our models:

- All word embedding layers are pre-trained word2vec from assignment 2.

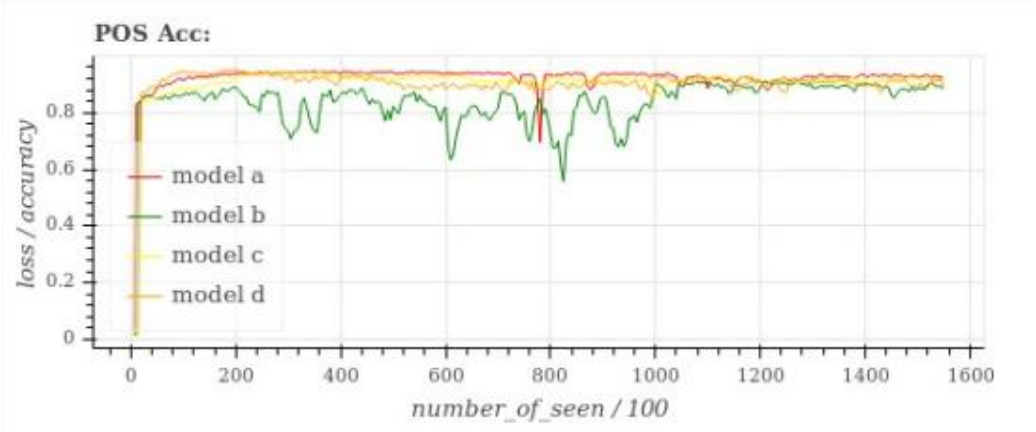
	a
layer sizes:	Pre-trained word embedding layer[50] LSTM layer hidden layer[100], linear layer [size of POS data] linear layer: out[60] + tanh linear layer + softmax
learning rate:	0.01
optimizers:	ADAM
dropout:	0.5
Batch:	124

	b
layer sizes:	Embedding layer ASCII: [10] Word embedding LSTM layer num. 1 : hidden layer [50] LSTM layer hidden layer[100], linear layer [size of POS data]
learning rate:	0.01
optimizers:	ADAM
dropout:	0.5
Batch:	124

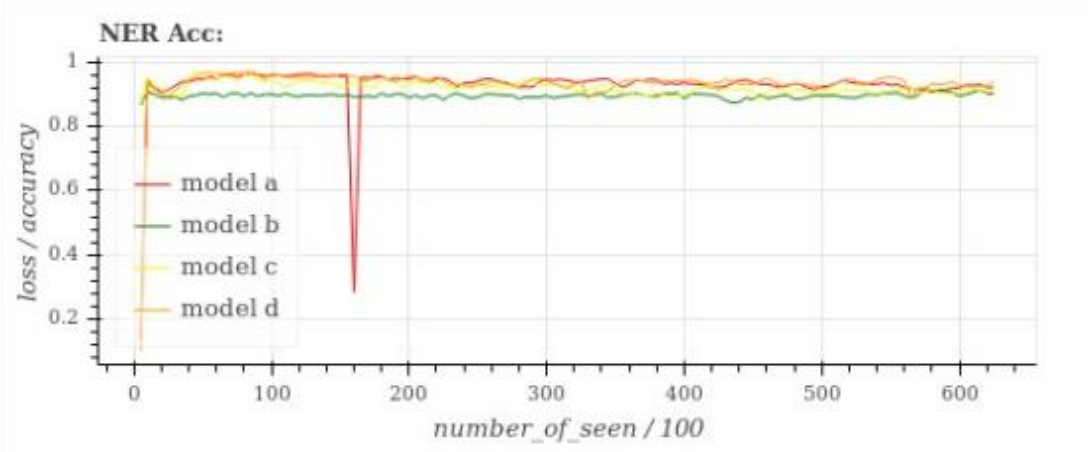
	c
layer sizes:	Embedding layer num. 1: prefix [50] + Embedding layer num. 2: suffix [50] + Embedding layer num. 3: pre-trained word embedding [50] (sum) Word embedding LSTM layer num. 1 hidden layer[50] LSTM layer hidden layer[100], linear layer [size of POS data]
learning rate:	0.01
optimizers:	ADAM
dropout:	0.5
Batch:	124

	d
layer sizes:	Embedding layer num. 1: prefix [50] Embedding layer num. 2: suffix [50] Embedding layer num. 3: pre-trained word embedding [50] (concat) Word embedding LSTM layer num. 1 hidden layer[50] LSTM layer hidden layer[100], linear layer [size of POS data]
learning rate:	0.01
optimizers:	ADAM
dropout:	0.5
Batch:	124

Graph 1 -learning curves for the **POS** data (the dev-set accuracies)
4 lines, corresponding to input representations (a), (b), (c), (d) above.



Graph 2 -learning curves for the **NER** data (the dev-set accuracies)
4 lines, corresponding to input representations (a), (b), (c), (d) above.



- accuracy (y-axis) vs. (number of sentences seen / 100) (x-axis).