Oren Cohen
ID: 305164295
USER: cohenorx

Int this section, I trained 4 models on 2 train sets- NER and POS
each model was with this parameters:

Embedding word size was 100, which means that every repr(w_i) (for a,b, c and d) should get to this size.

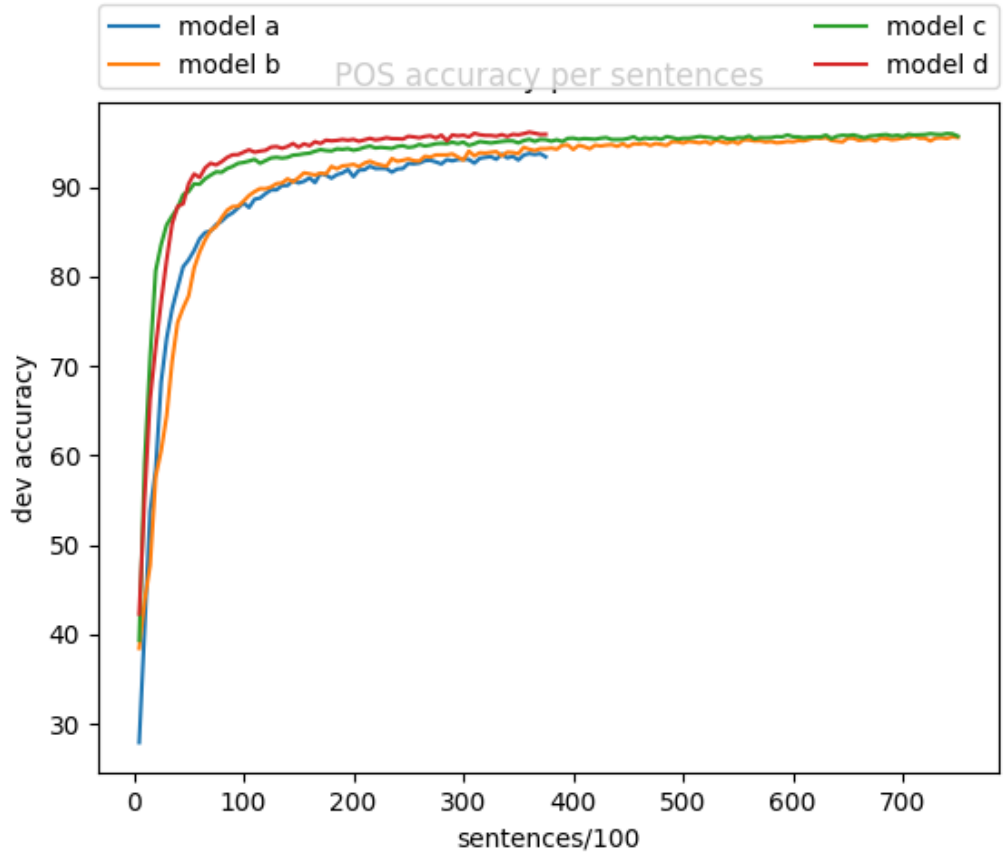
Also instead of Linear layer – I used MLP W1*tanh(W2)
where W1 dim is (#tags,32) and W2 dim is: (100,32)
learning rate was same for all models 0.001
Trained with dynet.AdamTrainer

Important:

some of the models converge quickly then others – so I stopped them at the peak dev accuracy (you can see that at the graphs:)

of course for the prediction I used model d, as you can see from the graphes its was very good

POS Graph:



Oren Cohen
ID: 305164295
USER: cohenorx

NER Graph: