HW3: Named Entity Recognition

Model inputs

1. word embeddings which is inicialized with fasttext embedding
2. Character level embeddings which is inicialized randomly
3. Capitalization features which is from the paper: Named Entity Recognition with Bidirectional LSTM-CNNs

Model architecture

I tried three models based paper Named Entity Recognition with Bidirectional LSTM-CNNs :

1. character level LSTM+character level cnn+capaitalizeation embedding+word embedding + BiLSTM
2. character level LSTM +capaitalizeation embedding+ word embedding + BiLSTM
3. character level cnn+ capaitalizeation embedding + word embedding + BiLSTM

Finally I choose the third one, since the third one works best on the dev set. The model architecture is the following:

Layer (type) Output Shape Param # Connected to

Character_input (InputLayer) (None, None, 61) 0

Character_embedding

(TimeDistributed) (None, None, 61, 30) 2910 Character_input[0][0]

Convolution (TimeDistributed) (None, None, 61, 53) 4823 Character_embedding[0][0]

---

max_pooling (TimeDistributed) (None, None, 1, 53) 0 Convolution[0][0]

---

words_input (InputLayer) (None, None) 0

---

casing_input (InputLayer) (None, None) 0

Flatten (TimeDistributed) (None, None, 53) 0 max_pooling[0][0]

embedding_1 (Embedding) (None, None, 50) 3061650 words_input[0][0]

embedding_2 (Embedding) (None, None, 8) 64 casing_input[0][0]

dropout_1 (Dropout) (None, None, 53) 0 Flatten[0][0]

concatenate_1 (Concatenate) (None, None, 111) 0 embedding_1[0][0]
embedding_2[0][0] dropout_1[0][0]

BiLSTM (Bidirectional) (None, None, 550) 851400 concatenate_1[0][0]

Softmax_layer (TimeDistributed) (None, None, 18) 9918 BiLSTM[0][0]

Total params: 3,930,765 Trainable params: 869,051 Non-trainable params: 3,061,714

Tuning

I use the same hyper-parameter as the paper Named Entity Recognition with Bidirectional LSTM-CNNs and get the similar results in the dev set which is 94.61% f1 score