Applied Deep Learning Assignment 3

Tung-Chun, Chiang R05922027

November 29, 2016

1 Bidirectional Recurrent Neuron Network

In this assignment, we have two task: slot filling and intent prediction. For slot filling, we have a sequence of input words and a sequence of output tags, and the task is to predict the output tags given input words. For intent prediction, we have a sequence of input words and a label, and the task is to predict the output label given input words.
In both the tasks, we use single layer bidirectional RNN with aligned inputs as encoder, while the output of decoder is tag sequence for slot filling and label for intent prediction and both use cross entropy as loss function. We pad each sequence to a fixed length. Figure 1 shows the model structure. Note that "aligned inputs" means we use the output of a word as the corresponding input of decoder.

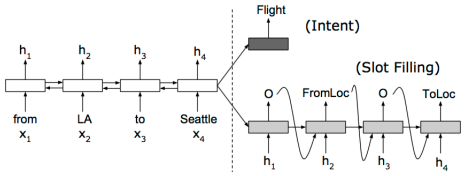

Figure 1: bidirectional RNN with aligned inputs

2 Improvement

To improve the model, we add attention mask in the model as Figure 2. In Eq 1, the c is attention value and it is a linear combination of the outputs of input word sequence(length T) and the weights α are also learned in training. And then, concatenate c and h as decoder inputs. Table 1 shows the improvement after adding attention mask. The attention method improves intent prediction task significantly. To do more, we use Google word2vec pre-trained model(300 dims) as initial value, and improve the intent predication score from 0.9585 to 0.9642.

$$c_i = \sum_{j=1}^{T} \alpha_{i,j} h_j \tag{1}$$

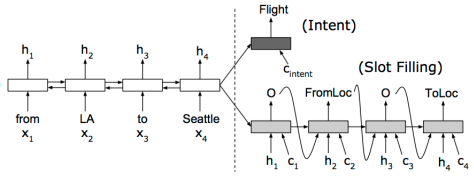

Figure 2: bidirectional RNN with aligned inputs and attention

f1/acc	with attention	without attention
slot filling	0.9579	0.9507
intent prediction	0.9642	0.9057

Table 1: attention improvement of two tasks

3 Learned

- Because of gradient explosion problem, we should clip the gradients.
- For gradient vanish problem, we use LSTM to control the back propagation flow.
- We can also use drop out to avoid overfitting.

References

[1] Bing Liu, Ian Lane, *Attention-Based Recurrent Neural Network Models for Joint Intent Detection and Slot Filling*, Carnegie Mellon University, 2016.