Model design

Xiaonan Zhao

February 7, 2017

1 Matching LSTM

Matching LSTM is given,

$$i_k^m = \delta(W^{mi} m_k + U^{mi} h_{k-1}^m + b^{mi}) \tag{1}$$
$$f_k^m = \delta(W^{mf} m_k + U^{mf} h_{k-1}^m + b^{mf}) \tag{2}$$
$$o_k^m = \delta(W^{mo} m_k + U^{mo} h_{k-1}^m + b^{mo}) \tag{3}$$
$$c_k^m = f_k^m \odot c_{k-1}^m + i_k^m \odot \tanh(W^{mc} m_k + U^{mc} h_{k-1}^m + b^{mc}) \tag{4}$$
$$h_k^m = o_k^m \odot \tanh(c_k^m) \tag{5}$$

where m_k is the matching vector,

$$m_k = \begin{bmatrix} a_k \cdot \hat{P} \\ \hat{h}_k \end{bmatrix} \tag{6}$$

Let's use $\hat{P} = (\hat{p}_1, \dots \hat{p}_M)$ to denote the matrix of SPINN stack of premise sentence, $\hat{H} = (\hat{h}_1, \dots, \hat{h}_k, \dots, \hat{h}_N)$ to denote the matrix of SPINN stack of hypothesis sentence. $a_k \in \mathcal{R}^{M-1}$ is the alignment weight vector over SPINN stack of premise sentence.

2 Attention model

The attention model is used for generating the soft alignments.

$$a_k = softmax(\begin{bmatrix} \vdots \\ e_k j \\ \vdots \end{bmatrix}) \tag{7}$$

$$e_k j = w^e \cdot \tanh(W^p \hat{p}_j + W^h \hat{h}_k + W^m h_{k-1}^m) \tag{8}$$

here h_k^m is the hidden vector in matching LSTM predefined.

3 SPINN

Both $\hat{P}$ and $\hat{H}$ comes from the SPINN model.

$$\hat{P} = \left[\hat{p}_1, \ldots, \begin{bmatrix} h_i \\ c_i \end{bmatrix} \ldots, \hat{p}_{2M-1}\right] \tag{9}$$

where h_i is the hidden output from LSTM and c_i is the cell unit from LSTM. For shift operation,

$$\begin{bmatrix} h_i \\ c_i \end{bmatrix} = W_{wd}\vec{x} + \vec{b} \tag{10}$$

As for reduce operations,

$$\begin{bmatrix} h_i \\ c_i \end{bmatrix} = \begin{bmatrix} o_i \odot \tanh(c_i) \\ f_i^l \odot c_i^l + f_i^r \odot c_i^r + i_i \odot g_i \end{bmatrix} \tag{11}$$