We proposed interactive-matching-lstm model to deal with sentence relationship recognition problem. We used two LSTM networks that interact with each other to represent sentence features and sentence relationship respectively. We used, say, $LSTM_s$ to represent both of the sentences, and $LSTM_r$ to represent sentence relationship. Our intuition was that $LSTM_s$' output at each step will affect our opinion about the relationship between sentences, that is to say, affect $LSTM_r$'s inputs at next step; and $LSTM_r$'s output at each step may affect $LSTM_s$' inputs as well, because it may affect our opinion about what we should pay attention to in the following steps.

Given a short text $\left[x^{(k)}{}_1, x^{(k)}{}_2, \cdots x^{(k)}{}_{N_{(k)}}\right] (k \in \{1,2\})$,

At time step t, first we calculate the attention-based representation of the sentence TILL NOW,

$$a^{s^{(k)}}{}_t = \sum_{j=1}^{t} \alpha_{tj} h_j^{s^{(k)}}$$

where

$$\alpha_{tj} = \frac{\exp(e_{tj})}{\sum_{j'} \exp(e_{tj'})}$$

$$e_{tj} = w^e \cdot \tanh\left(W^{self} h^{s^{(k)}}{}_j + W^{other} h^{s^{(3-k)}}{}_t + W^a h^r{}_{t-1}\right)$$

where w^e、$W^{(1)}$、$W^{(2)}$、W^a are parameters to learn.

The operations inside $LSTM_s$ are same as normal LSTM, while its inputs are somewhat special. At step t, for $LSTM_s$ we have

$$s^{(k)}{}_t = \begin{bmatrix} x^{(k)}{}_t \\ h^{s^{(k)}}{}_{t-1} \\ h^r{}_{t-1} \end{bmatrix}$$

$$i^{s^{(k)}}{}_t = \sigma(W^{(si)} s^{(k)}{}_t + b^{(si)})$$

$$o^{s^{(k)}}{}_t = \sigma(W^{(so)} s^{(k)}{}_t + b^{(so)})$$

$$f^{s^{(k)}}{}_t = \sigma(W^{(sf)} s^{(k)}{}_t + b^{(sf)})$$

$$\widetilde{c^{s^{(k)}}}_t = \tanh\left(\widetilde{W^{sc}} s^{(k)}{}_t + b^{(sc)}\right)$$

$$c^{s^{(k)}}{}_t = f^{s^{(k)}}{}_t \otimes c^{s^{(k)}}{}_{t-1} + i^{s^{(k)}}{}_t \otimes \widetilde{c^{s^{(k)}}}_t$$

$$h^{s^{(k)}}{}_t = o^{s^{(k)}}{}_t \otimes \tanh\left(c^{s^{(k)}}{}_t\right)$$

then for $LSTM_r$ we have

$$r_t = \begin{bmatrix} a^{s^{(1)}}{}_t \\ h^r{}_{t-1} \\ a^{s^{(2)}}{}_t \end{bmatrix}$$

$$i^r{}_t = \sigma(W^{(ri)} \square_t + b^{(ri)})$$

$$o^r{}_t = \sigma(W^{(ro)} r_t + b^{(ro)})$$

$$f^r{}_t = \sigma(W^{(rf)} r_t + b^{(rf)})$$

$$\widetilde{c^r}_t = \tanh\left(\widetilde{W^{rc}} r_t + b^{(rc)}\right)$$

$$c^r{}_t = f^r{}_t \otimes c^r{}_{t-1} + i^r{}_t \otimes \widetilde{c^r}_t$$

$$h^r{}_t = o^r{}_t \otimes \tanh\left(c^r{}_t\right)$$

We feed the final state of $LSTM_r$ to a 3-way classifier for sentence entailment recognition on SNLI

and a 2-way classifier for question duplicate recognition on Quora Question Pairs.
Our model achieved 83.66% accuracy on SNLI and 82.19% accuracy on Quora(We randomly selected 4000 pairs from train.csv for test).