A Supplementary Material

A.1 Features from S2S model

This is the list of features extracted from a S2S model including sentence log loss, the margin between the best and second best solutions, source frequency, source encoder last hidden state and target decoder last hidden state.

A.1.1 Sentence log loss

We extracted two features: the unnormalized sentence log loss which is described in equation (1) and the same score but normalized by by sentence length.

A.1.2 Margin of best and second best

We compute the margin of best (y^*) and second best (y^{**}) solution at each time step t:

$$\Delta_t = \mathsf{P}(y_t^*|x;\theta) - \mathsf{P}(y_t^{**}|x;\theta)$$

We extracted three features: $[\min\Delta, \max\Delta, \text{std} \quad \Delta]$. These features show how well the model disambiguate the best and second best solution.

A.1.3 Source frequency

The motivation is that if there are many low frequency words in the source utterance or LF, it is generally harder for the S2S model to generate the target. We extracted 5 log scale features $[\log(\text{freq}_i)]$ and 5 binary features $[\text{freq}_i > 0]$with $0 \leq i \leq 4$ and freq_i is the number of words having frequency i in the source according to the vocabulary. For example, freq_0 is the number of unknown words in the source. Note that we set $\log(\text{freq}_i) = 0$ if $\text{freq}_i = 0$

A.1.4 Encoder last hidden state

Since we use a bidirectional RNN for the S2S encoder, we extract two features $[h_{r\leftarrow l}, h_{r\rightarrow l}]$ where $h_{r\leftarrow l}$ and $h_{r\rightarrow l}$ are the last hidden state running from left to right and right to left respectively.

A.1.5 Decoder last hidden state

We extract the hidden state from a decoder as long as the decoder generates the end of sentence symbol (EOS).