TODO: finish paper

Neural Machine Translation By Jointly Learning To Align and Translate

1 Overview

Translation is equivalent to finding a target sentence y that maximizes the conditional probability of y given a source sentence x, i.e., $\arg\min_y p(y|x)$. Once the conditional distribution is learned by a translation model, given a source sentence a corresponding translation can be generated by searching for the sentence that maximizes the conditional probability. First techniques for learning this distribution is "*encoder-decoder mechanism*". Encoder reads the input sentence and converts it into a intermediate representing vector "c". Most common technique for it is RNN. [h=hidden states, f&g are nonlinearities]

$$h_t = f(x_t, h_{t-1}) \tag{1}$$

$$c = g(h_1, ..., h_T) \tag{2}$$

Decoder is trained to predict the next word y_t given the context vector c and all the previously predicted words.

$$p(y) = \prod_{t=1}^{T} p(y_t|y_1, ..., y_{t-1}, c) \tag{3}$$

This system has bottleneck problem which is encoder tries to put all information fixed sized "c". Moreover, it is hard to cope with long sentences with aforementioned method.

1.1 New Method: LEARNING TO ALIGN AND TRANSLATE

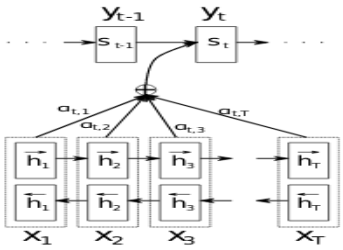

Figure 1: The graphical illustration of the proposed model trying to generate the t-th target word y_t given a source sentence $(x_1, x_2, \ldots, x_T)$.