Sequence to Sequence Learning with Neural Networks

December 9, 2023

1 Summary

0. This paper uses LSTMs to present a general end-to-end approach to sequence learning that makes minimal assumptions about sequenes. The main result is in English to French translation with good BLEU score that also captures long term dependencies.

1. **The model:** A recurrent neural network (RNN) can be captured in the following two equations for a time step t:

$$h_t = sigm(W^{hx}x_t + W^{hh}h_{t-1})$$
$$y_t = W^{yh}h_t$$

This cannot capture long term dependencies. LSTMs are designed to overcome this. The goal of LSTM is to capture conditional probability $p(y_1, \cdots, y'_T|x_1, \cdots, x_T)$. This is done by first computing v from the input using an LSTM and then using another LSTM to compute

$$p(y_1, \cdots, y'_T|x_1, \cdots, x_T) = \Pi_{t=1}^{T'} p(y_t|v, y_1, \cdots, y_{t-1}).$$

Three novelties are introduced in the model - using two LSTMS, using LSTMs with multiple layers, and reversing the order of the input which performed better.

2. **Experiments:** English to French translation is done maximizing the log probability

$$1/\mathcal{S} \sum_{(T,S)\in\mathcal{S}} \log p(T|S)$$

where $\mathcal{S}$ is the training set. Translations most likely according to the LSTM are produced.

Reversing inputs sentences and parallelization of computation using 8-GPUs techniques were used. Results were promising in that this algorithm beat all the existing ones on BLEU scores.

2 Analysis

This paper was the first successful one to use a large LSTM network in achieving sequence to sequence generation that respected long term dependencies. A few simple new techniques were introduced but the crux of contribution is in training the large LSTM network successfully.