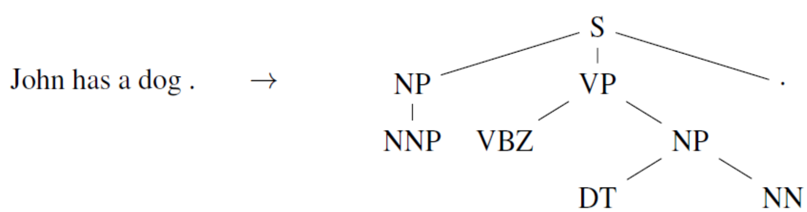
Colton Cookerly

Grammar as a Foreign Language (Capstone Abstract)

Project Concept

My project is based on the paper "Grammar as a Foreign Language" [3]. The core idea is a machine translation seq2seq model that, rather than translating a sentence from one natural language to another, instead translates a sentence to the parse tree representing the grammatical structure of that sentence. I trained my model on the Wall Street Journal section of the Penn Treebank.

Here is a diagram[1] which shows the parse tree for the sentence, "John has a dog":



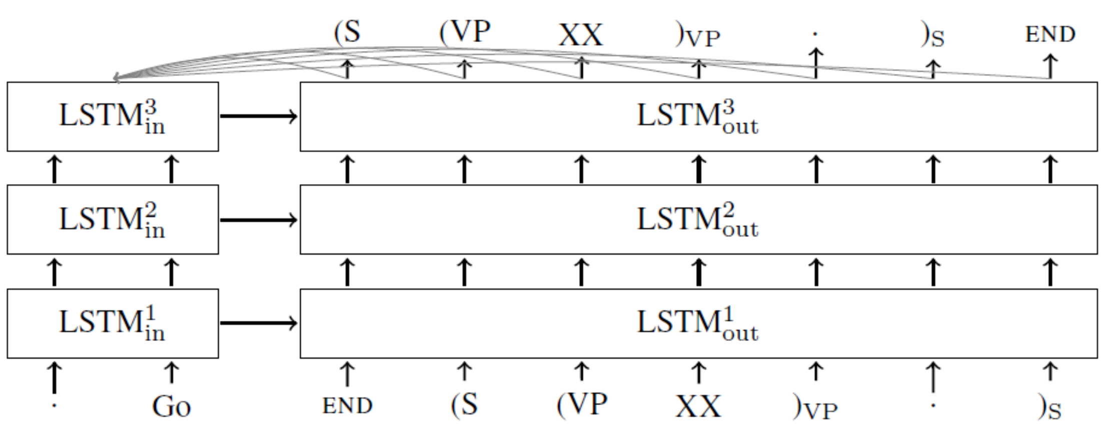
In order for parse trees to be something we can translate to, they are represented as a sequence of tokens (similarly to how sentences in natural language are represented as a series of tokens). This is called a linearization of a parse tree. Here is a linearization of the parse tree for the above sentence:

John has a dog . $\rightarrow$ (S (NP NNP $)_{NP}$ (VP VBZ (NP DT NN $)_{NP}$ $)_{VP}$. $)_{S}$

For example, '(S' is a token, as are 'NNP' and '$)_{VP}$'.

Architecture

Here is a diagram representing the model run on the input "Go .":



---

[1] The diagrams in this abstract are from [3].

The model consists of an "encoder" and a "decoder." Both use a multi-RNN cell which is comprised of three ordinary RNN cells (specifically, LSTMs). For the encoder, these are the boxes labeled '$LSTM_{in}$', and for the decoder, '$LSTM_{out}$'. (The horizontal length of these boxes represents different timesteps.)

In a multi-RNN, the output of the i^{th}-layer RNN cell becomes the input of the $(i+1)^{th}$-layer RNN cell at the same timestep. (This is represented by the vertical arrows between boxes.) The initial state of each RNN cell in the decoder is the final state of the RNN cell at the corresponding layer in the encoder. (This is represented by the horizontal arrows between the encoder and the decoder.)

The curved arrows, going from each timestep of the decoder back to the encoder, represent Bahdanau attention. The key idea here is that there are a learnable set of weights which the model uses to compare the current state of the decoder's RNN with the previous states of the encoder's RNN.[2]

References

[1] Bahdanau, et al. "Neural Machine Translation by Jointly Learning to Align and Translate." 19 May 2016.

[2] "Neural Machine Translation (seq2seq) Tutorial." *TensorFlow*, www.tensorflow.org/tutorials/seq2seq. 27 January 2018.

[3] Vinyals and Kaiser, et al. "Grammar as a Foreign Language." 9 June 2015.

[2] For more information on Bahdanau attention, see [1], which introduced the idea, and [2], which provides a more concise explanation.