Project Proposal

Alex Shah

9/25/17

0.1 Introduction

In neural networks, the Encoder Decoder method has recently seen increasingly efficient strides in translating languages with contextual awareness. A simple word by word translation would be ineffective at nuanced language dynamics such as gender agreement and pronoun agreement. As such, examining the words surrounding a given input word can lead to more accurate translation. Encoder Decoder methods such as sequence to sequence (seq2seq) and tensor2tensor (t2t) with the "Attention is All You Need" mechanism are the most promising at maximizing accuracy for computer translation. I plan to create a neural network translation implementation which first uses a language classifier to detect whether the input language is English or Spanish, then uses the methods and mechanisms mentioned to translate the given text to the opposing language, Spanish from English, or English from Spanish.

0.2 Methodology

The Sequence to Sequence method to Encoder Decoder is particularly effective at handling translation. Using two Recurrent Neural Networks, an encoder processes input and a decoder generates output. The attention mechanism allows the decoder to look at the input with each step. Other methods of translation do not focus on the sequence of words, seq2seq's strong suit is in contextual translation by means of considering previous words in a phrase or sentence.

0.3 Subject Matter

I chose translation because I feel it's a great example of what RNN's are capable of optimizing compared to older methods of computation. To be able to recognize context clues and nuances such as grammatical correctness are amazing strides that machine learning has brought to computation. Specifically I chose Spanish since I have experience enough in Spanish to troubleshoot and recognize correct or incorrect output while testing the project.

0.4 Novel Implementation

The novel concept to this approach is to utilize methods which were constructed to maximize translation accuracy, and combine several methods to create a working product. Namely, to combine a language classifier and two Neural Machine Translation networks into one finished product capable of identifying and handling translation in either direction between languages.