Ikhlas Attarwala
04-15-19 / Week #3
RNN Architectures / Paper 2

Exploring the Limits of Language Modeling by Józefowicz, Rafal et al.

In the quest to study the limits of how well a computer can reason around and understand human language, Józefowicz et al. propose 4 language models in this paper in an effort to beat (the then) state-of-the-art language models (a standard LSTM, two models where the input & softmax embeddings are character CNNs where one has no input look-up table and the other has no look-up table of any kind, and one where the softmax is a character prediction via LSTM). They ended up suggesting the use of a large LSTM language model that used softmax approximation and was trained on a large corpus. Upon replacing the input look-up table, they reduced to the lowest number of parameters compared to previous related work (1.04 billion parameters), along with the lowest test perplexity for single and ensemble models (score of 30.0 and 23.7 respectively).

In order to develop high-performing neural language models, one should take note that the largest models performed best (more parameters), especially those that had greater memory units. Regularization across input connections was also important for results (note that dropout reduces overfitting). Additionally, character CNNs could replace word embeddings, and a combination of prediction models should yield better model performance. Their training time did seem considerably long, capped at 3 weeks for over 30 GPUs. Another issue I had with the paper was how the sample of words they provided under section 5.9 seemed quite formal, as if the text was from an article, and this doesn't match human speech. Otherwise, this paper was great at going into depth on how the differences in organization & structure of a model can effect or hinder performance.