Ikhlas Attarwala
04-15-19 / Week #3
RNN Architectures / Paper 1

RNN-based Language Model by Mikolov, Tomas et al.

The term 'recurrent' stands for performing a task across each element in a sequence. While the inputs of other neural networks function independently of one another, inputs of an RNN are related to one another. This proposed model by Mikolov presents a 50% reduction in perplexity & avoids the problem of leveraging arbitrarily large contexts as it doesn't use a standard feed-forward network. Mikolov's model had great results and very fast training times. The desired vector subtracted from the softmax output yielded the loss function for training.

I don't have a great background in language modeling, but from what it appears, it seems impressive to create a model that has both phenomenal accuracy and trains much faster than previous models (compare 6 hours for basic implementation vs 26 hours with importance sampling by Bengio's feed-forward network). I had to agree with a point Mikolov made in the conclusion where he said that "it is possible that further investigation into backpropagation… will provide additional improvements". I would think the basis for this would be the comparison of Bengio's not being able to work around arbitrarily-larger sized data. I would suggest taking a stab at this with an RNTN (Recursive Neural Tensor Network). I would also be interested in learning how memory plays a role into this. Would an LSTM outperform the RNN on a different sized corpus?