Language Translation using RNNs: A Comparison and Analysis of Various Neural Machine Translation Models

Jackson Eilers, Gita Krishna, Faraz Abbasi

THE PROBLEM

Language Translation is a task with the potential for global impact. The inability to communicate due to a language barrier is one of the largest impediments we face today, in an ever globalized society. In order to overcome this barrier, we chose to create a language translation model.

DATA / TASK

- **EMNLP 2011 Sixth Workshop on Machine Translation**: This dataset is primarily intended for translation tasks between various European languages. Large portions of this dataset come from the European Parliament Proceedings Parallel Corpus, containing sentence aligned text for translation systems in 21 languages
- **Size**: This dataset contains over 45 million words of training data. We plan to use a small fraction of this for training feasibility
- **Train/Dev/Test**: We are using a 98-1-1 train/dev/test split
- **Format**: The examples come in source-target sentence pairs, in the following format

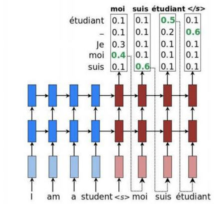
  Source sentence: *À cette époque, l'astronomie a une fonction essentiellement pratique.*
  Target translation: *During the time of the early settlers, the function of astronomy was primarily practical.*

APPROACH

Model 1: Sequence to Sequence Recurrent Neural Network (RNN) with Adam Optimization, Gradient Clipping, Dropout, and L2 regularization

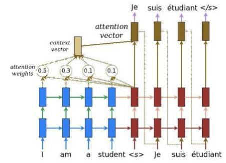
Model 2: Sequential Long Short-Term Memory (LSTM) RNN with Nadam Optimization

Model 3: Seq2seq RNN with LSTM cells and Stochastic Gradient Descent (SGD) as the method of optimization, incorporating a self-attention mechanism



ANALYSIS

Model 1 was able to perform the best with a standard sequence model, but adding self-attention to Model 3 improved the accuracy by a large amount. The self-attention mechanism we used is as follows:

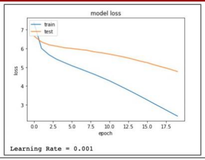


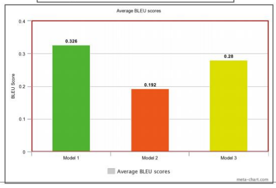
RESULTS

Model 1:
- Was able to get BLEU scores of .358, .316,.366 and .263 for 1,2,3 and 4-gram respectively
- Adam Optimization, Gradient Clipping and L2 regularization successfully made training more efficient

Model 2:
- Got BLEU scores of 0.281, 0.153, 0.2436, 0.224
- Performed best with Nadam optimization and learning rate of 0.001

Model 3:
- Got BLEU scores of 0.26 without attention model and 0.48 with the attention model
- Used SGD, gradient clipping, and dropout. Given long sentences in corpus, attention model improved translations significantly





CONCLUSION / FUTURE WORK

Our first model was the highest performing on average due to the combination of using an Adam optimizer and L2 normalization. However, the highest BLEU score came from model 3 which used the attention mechanism. Since this model used SGD normally, we concluded that this is the reason why it did not perform as well as model 1 and that Adam was a preferable optimizer. If we had more time, we would have loved to explore translating other corpus' on the same models to see if it was a limitation of the data. Further, we would have loved to test all of the different elements of all the models to find some combination that was most effective at translating languages.

REFERENCES

Chansung, Park.Seq2Seq model in TensorFlow, https://towardsdatascience.com/seq2seq-model-in-tensorflow-ec0c557e560f., Sutskever, Ilya, Vinyals, Oriol. & Quoc V. Le (2014) Sequence to Sequence Learning with Neural Networks System. https://papers.nips.cc/paper/5346-sequence-to-sequence-learning-with-neural-networks.pdf., Gehring, Jonas & Michael Auli (2017 A novel approach to neural machine translation. https://code.fb.com/ml-applications/a-novel-approach-to-neural-machine-translation/., ffective approaches to attention-based neural machine translation, Minh-Thang Luong, Hieu Pham, and Christopher D Manning. EMNLP, 2015., https://github.com/bihor98/Neural-Machine-Translation-using-Kears, https://github.com/tesnorflow/nmt