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/Test**: We are using a 98-2-1 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.

---

**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/Test**: We are using a 98-2-1 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.

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 3.58, 3.16, 3.66, and 2.63 for 1, 2, 3, and 4 grams respectively.
- **Model 2**: Adam Optimization, Gradient Clipping and L2 regularization successfully made training more efficient.
- **Model 3**: Got BLEU scores of 0.281, 0.153, 0.2456, 0.224.
- **Model 2**: Performed best with Nadam optimization and learning rate of 0.003.
- **Model 3**: Got BLEU scores of 0.26 without attention model and 0.48 with the attention model.

---

**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.

---

**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 corpora 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**