Machine Translation Project

Yifan Zhao
yifanzhao@umass.edu

1 Introduction

Neural machine translation has made impressive progress since deep learning becomes a dominant approach in Artificial Intelligence. Machine translation aims to take one sequence of words in one language as input and output the same sequence of words in another language. Machine translation has many practical applications and plays a more important role in cultural exchange, education and scientific research with the development of internet and globalization.

In our project, we will extend the seq2seq machine translation model to have a better performance in German to English translation. We are given IWSLT 2016 corpus as our default dataset and use BLUE metric to evaluate the result of our model.

2 Related work

In 2014, a seq2seq machine translation model(Ilya Sutskever and Le, 2014) was introduced and drawn attention in many NLP researchers. The seq2seq model is an end-to-end model that consists of two stacked LSTM networks: the encoder and the decoder. The encoder takes the input sequence as input and transforms them into a fixed-size vector which compresses the information about the input content. The decoder uses the vector produced by the encoder as a seed to generate the output sequence.

In order to get better translation results in long sentences, some attention mechanism approach(Pham and Manning, 2015) is proposed to make use of the observation that different parts of the input sentence may have different weight when generating different parts of the output sentence. Another major issue in machine translation is large output vocabulary that may produce lots of unknown words. To address the problem, a paper(Luong and Manning, 2016) suggests a hybrid model that make up with word and character-based approaches. The system translates common words using word-level approach and switches to the character-level approach when encountering rare words.

3 Approach

In our project, we will elaborate the basic seq2seq model. while our model is based on the model, we have three major modifications aiming to improve it.

First, we will use bidirectional RNNs(Schuster and Paliwal, 1997) instead of one direction RNN cell. In the original seq2seq paper, the input sentence will be processed in reverse order which makes it easier for the output sequence to get started. Based on the experiments, the word may have a dependency on the word before or after it in different languages. The bidirectional RNNs can solve the issue since it can produce and concatenate the hidden state from both directions. Furthermore, we choose to use GRU(Junyoung Chung and Bengio, 2014) as our RNN cell instead of Vanilla RNN cell. The GRU is newly introduced, the performance is as good as the LSTM cell but computationally more efficient.

Second, We will add a global attention mechanism. In our attention layer, every decoder step will consider the entire input sequence and produce the hidden unit.

Third, we use beam search(Freitag and Al-Onaizan, 2017) instead of exhaustive search or greedy search as the sequence model decoders.

Our baseline is stacked LSTM seq2seq model.

3.1 Milestones & Schedule

Since our group only include one person, I have to do all the work. Tentative schedule:

1. Acquire and preprocess data (1 week)
2. Build models for task (3 weeks)
3. Write progress report! (due Apr 1)
4. Analyze the output of the model, do an error analysis (2 weeks)
5. Work on final report and presentation (2 weeks)

4 Data

The dataset consists of four different components: The German and the English training files and the German and English validation flies.

5 Tools

We will preprocessing our data using standard tokenization. We choose to use pytorch to implement our algorithms and we will train our model in google cloud.

References

Freitag, M. and Al-Onaizan, Y. (2017). Beam search strategies for neural machine translation. In *First Workshop on Neural Machine Translation*.

Ilya Sutskever, O. V. and Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems 27*.

Junyoung Chung, Caglar Gulcehre, K. C. and Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. In *NIPS 2014 Deep Learning and Representation Learning Workshop*.

Luong, M.-T. and Manning, C. D. (2016). Achieving open vocabulary neural machine translation with hybrid word-character models. In *ACL 2016 camera-ready versionn*.

Pham, M.-T. L. H. and Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. In *EMNLP 2015 camera-ready version*.

Schuster, M. and Paliwal, K. (1997). Bidirectional recurrent neural networks. In *IEEE Transactions on Signal Processing*.