

CS447 Research Project Proposal:

Study of deep encoders in RNNs for machine translation from English to German and English to Chinese

Shivesh Pathak, sapatha2

October 13, 2020

Task description

I am interested in evaluating the relative benefits of using deep recurrent neural network (RNN) encoders for neural machine translation (NMT) from English to German and Chinese. First, I would like to quantitatively assess the benefits of deep RNN encoder architectures in machine translation: is the increased parameterization complexity and training cost counteracted by significantly higher accuracy, for example. Second, I would like to determine how the benefits vary in translation between similar languages (English and German) versus languages which are very dissimilar (English to Chinese): are the increases in accuracy higher for more dissimilar languages, for example.

Background

The problem of deep RNNs for NMT has been heavily studied, but a recent paper by Barone *et al.* [1] lays the groundwork for the questions that I want to answer. In this paper, they investigate the benefits of deep architectures for just the encoder portion, just the decoder portion, and both encoder and decoder portions of the RNN together in MT between English and German. They train on publicly available parallel news data and use BLEU on three different test sets to determine their model accuracy. They conclude that the deep architectures have positive impact on accuracy, with a simple deep transition encoder and shallow decoder architecture performing significantly better than the shallow encoder and decoder network.

While the results of this paper are very instructive to the benefits of deep encoder RNNs in NMT (and has a very clear explanation and definition of network types), it has the drawback that the project was conducted on English to German translation only. These languages are quite similar, and this similarity may bias the conclusions when speaking about the benefits for NMT in general. This was likely because the news data set being used in this paper, the WMT News Database for 2017 (<http://www.statmt.org/wmt17/>), did not have any non-European languages present until 2018. But with the inclusion of Chinese into the news data, it is possible to test whether the conclusions in this paper extend to NMT between two very dissimilar languages: English and Chinese.

Data and evaluation

The data required to train and develop my data set is readily available. I will be using the WMT datasets which are publicly available, following the procedure outlined in [1]. As described later, the WMT 2017 data set will be used to ensure that I can reproduce the results in the paper, and the WMT 2020 data sets

will be used to do a final comparison between the RNN performances on English to German and English to Chinese translation. Also following the paper above, I will measure the RNN accuracy using both cross-entropy on the development data set, and BLEU on test data sets from WMT 2018 and 2019 which both contain English, German and Chinese parallel data.

Your approach

My approach is broken into two parts: reproduction of some results in [1] and investigation of model performance on English to Chinese NMT. For the reproduction approach, I will code up the shallow-encoder shallow-decoder presented in the paper and the highest performing deep-encoder shallow-decoder RNN only, both in PyTorch. This choice is to make the project feasible, as I am concerned that I will not have enough time to develop and debug, as well as train and write up the report for the project if I tried to do all the RNNs presented in the document. I will follow the steps in the paper for training the model and testing it using the WMT 2017 data sets, and compare the CE on the dev data set and BLEU on the testing data tests which I compute to what the paper has. Once this is complete, I will move on to the next stage.

In the second stage, I will train the model independently for English-German and English-Chinese translation using the WMT 2020 data set. I will evaluate identical metrics to the reproduction task, and from this make inferences about the model performance. I hope to then answer whether the benefits seen in using the deep-encoder for English-German translation carry over to English-Chinese, and whether the benefits, if they exist, are larger or smaller.

Your to-do list

1. The data is available online, but I will have to load this in to PyTorch and into necessary structures like DataLoaders. **Deadline: October 20**
2. I will implement the shallow encoder-decoder RNN and the deep-transition encoder shallow-decoder RNN from [1] in PyTorch. If I get stuck, there is a reference code in TensorFlow which I can fall back on, either for help debugging *or* for direct usage (worst case). **Deadline: November 3**
3. I will train and test these RNNs for English-German translation using the WMT 2017 data to reproduce the results in [1]. I will also train and test English-German and English-Chinese on the WMT 2020 data. **Deadline: December 1**
4. I will be using BLEU and Cross entropy to measure the performance of the networks. Luckily, these are already present in PyTorch so writing this up should be quick. **Deadline: December 1**
5. Writeup! **Deadline: December 9**

References

- [1] Antonio Valerio Miceli Barone, Jindřich Helcl, Rico Sennrich, Barry Haddow, and Alexandra Birch. Deep architectures for neural machine translation. In *Proceedings of the Second Conference on Machine Translation*, pages 99–107, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.