

State of the Art Summary of Low-Resource Neural Machine Translation

Brennan Dell

October 2019

The latest advancements to the state of the art in machine translation have been accomplished through Neural Machine Translation (NMT). However, NMT depends on large corpora to produce decent quality translations. This document will describe approaches to increase the quality of Neural Machine Translation in low-resource settings, to produce models which are competitive with or surpass the performance of Phrase-Based Statistical Machine Translation (PBSMT) systems, which typically outperform NMT systems in low-resource contexts. Developing NMT systems which achieve high performance despite small corpus size are especially crucial for a majority of the world’s languages, for which large corpora of parallel data do not exist. Additionally, developing NMT systems which require smaller corpora, and therefore less time and fewer training resources, may decrease the energy usage required to train Neural Models. [9]

A common approach used for neural machine learning application when there is little training data available is to make use of a transfer-method, namely training a model on a large dataset related to the problem of interest, and fine-tuning that model on a smaller dataset which is more pertinent to the problem being solved. Many such approaches have been developed for NMT, where an MT system is trained on a high-resource language pair, and later is fine-tuned to a low-resource language pair. Such transfer systems may even result in zero-shot translation, where a translation system can translate in a language pair which was not encountered in the dataset. Transfer methods have been implemented in a variety of ways, where various parameters are held constant while others are allowed to be trained during fine-tuning.

Zoph et al. [11] demonstrate a transfer-based approach to NMT which is able to improve BLEU scores of low-resource language pairs by 5.6 BLEU. Their method was to train a "parent" model using a large bilingual parallel corpus, and produce a "child" models, which would begin with the same parameters as the parent model, and were trained on smaller corpora for other language pairs. The parent model which Zoph et al. used was a French-English translator, trained on a parallel corpus of 300 million tokens (roughly 6 million sentences per Sennrich’s estimation of 50 tokens per sentence [8]). From this French-English parent model, child models were trained to translate Hausa, Turkish, Uzbek, and Urdu into English. These low resource languages varied in size

from 0.2 million tokens for Urdu (4,000 sentences) to 1.8 million sentences for Turkish (36,000 sentences). Zoph et al. compare child models trained with transferred parameters to models trained on the low-resource corpora alone, and found significant improvements across the board (Hausa +4.5 BLEU, Turkish +5.6 BLEU, Uzbek +3.7 BLEU, and Urdu +8.6 BLEU). Note the significant improvement of the Urdu-English model, which happened to be the smallest corpus Zoph et al. worked with. These BLEU scores were further refined with ensembling and unknown-word replacement (which could be further improved through the use of subword units [7]). Additionally, Zoph find that transfer learning performs better when translating between related language pairs. Zoph et al. found that allowing all parameters to be trained except for the target language embedding produced the optimal results.

Imankulova et al. [2] note that a frequent problem in low-resource NMT is that corpora are not available in a wide variety of domains. They suggest to solve this problem using a three step approach: training high-resource language pairs on out-of-domain data, fine-tuning these models to in-domain data on the same high-resource language pairs, and finally fine-tuning these models to the in-domain data in the low resource language pair. Imankulova et al. focus in particular on training sentences from the news domain from Russian to Japanese, using out-of-domain corpora for Russian-English and Japanese-English, as well as in domain data for all three language pairs. They found that using all three of these steps produces models with higher performance than models trained on any two steps, and find an improvement in 3.7 BLEU over a strong baseline. Additionally, they found that further increases in performance could be made through the use of back-translation [6]. Additionally, Imankulova et al. found that the ideal architecture to use was a multi-lingual Transformer based architecture.

Micher’s thesis proposal [5] addresses the issues that such a NMT needs to solve. Micher’s thesis is concerned with the Inuktitut language, a

References

- [1] Jiatao Gu, Hany Hassan, Jacob Devlin, and Victor OK Li. Universal neural machine translation for extremely low resource languages. *arXiv preprint arXiv:1802.05368*, 2018.
- [2] Aizhan Imankulova, Raj Dabre, Atsushi Fujita, and Kenji Imamura. Exploiting out-of-domain parallel data through multilingual transfer learning for low-resource neural machine translation. *arXiv preprint arXiv:1907.03060*, 2019.
- [3] Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander M Rush. Opennmt: Open-source toolkit for neural machine translation. *arXiv preprint arXiv:1701.02810*, 2017.

- [4] Surafel M Lakew, Mauro Cettolo, and Marcello Federico. A comparison of transformer and recurrent neural networks on multilingual neural machine translation. *arXiv preprint arXiv:1806.06957*, 2018.
- [5] Jeffrey C Micher. Addressing challenges of machine translation of inuit languages. Technical report, US Army Research Laboratory Adelphi United States, 2018.
- [6] Rico Sennrich, Barry Haddow, and Alexandra Birch. Improving neural machine translation models with monolingual data. *arXiv preprint arXiv:1511.06709*, 2015.
- [7] Rico Sennrich, Barry Haddow, and Alexandra Birch. Neural machine translation of rare words with subword units. *arXiv preprint arXiv:1508.07909*, 2015.
- [8] Rico Sennrich and Biao Zhang. Revisiting low-resource neural machine translation: A case study. *arXiv preprint arXiv:1905.11901*, 2019.
- [9] Emma Strubell, Ananya Ganesh, and Andrew McCallum. Energy and policy considerations for deep learning in nlp. *arXiv preprint arXiv:1906.02243*, 2019.
- [10] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.
- [11] Barret Zoph, Deniz Yuret, Jonathan May, and Kevin Knight. Transfer learning for low-resource neural machine translation. *arXiv preprint arXiv:1604.02201*, 2016.