Paper: *Achieving Open Vocabulary Neural Machine Translation with Hybrid Word-Character Models*

Mushfika Rahman

Date: 07/04/2022

Quote *Our hybrid architecture, leverages the power of both words and characters to achieve the goal of open vocabulary NMT.* pp.12.

Overview The authors in this paper addressed the problems with existing solutions that deal with vocabulary words. Furthermore, they have proposed a hybrid architecture to tackle the vocabulary words during machine translation. Neural machine translation is a simple architecture that has surpassed statistical machine translation in performance. However, most of the NMTs are trained on restricted vocabulary. As a result, when they encounter rare words, they translate to $< unk >$. There are many simple solutions that occur at the post-processing step to this issue. Firstly,[Luong et al. (2015)] proposed that involves tracking the alignment of target $< unk >$ and later resulting simple dictionary lookup. [Jean et al. (2014)] proposed similar techniques proposed in the first work but introduced an attention mechanism to obtain the positional alignment. However, these simple solutions have limitations. These solutions do not take care of languageś monolingual property by treating words as independent entities. Additionally, do not look after the cross-lingual property as language has different alphabets.

The authors proposed a hybrid model that consists of a word-based NMT that performs most of the translation job. Whenever encountered with rare words, the proposed model calculates the representation on the fly using deep neural networks separately. Then the components are learned jointly end-to-end removing the replacement for separate unk symbol. Their proposed hybrid architecture has exceeded the state-of-the-art translation tasks at the BLEU score level.

Intellectual Merit The research has advanced the knowledge in the natural language processing field. NMTs are trained on restricted vocabulary, and taking care of rare words is necessary. Since most of the solutions include character-level models, which are slow and do not consider various language properties, a new solution faster in training was required. The research is well reasoned. It introduced an additional component in the existing architecture thus exploring creative concepts. The research involved scientists from the Stanford NLP group. The group and the scientists have been working on language modeling problems for a while and have published many novel works, making them credible for conducting this research. They used the benchmark WMT'15 dataset for their experiments.

Broader Impact The research has improved machine translation quality, thus impacting the natural language tasks. The research has been disseminated into further publications since its publication in 2016. The research involved one Ph.D. student. They have publicly released all the codes and datasets.

Keywords Computing methodologies, Natural language processing, Machine learning, Neural Machine Translation.

Discussion Questions

- The experiment was based on one language translation, how the model works on other languages is not stated.
- Character modelś backpropogation performance with longer dependency is not stated properly.

Table 1: Grade deductions by section

Overview	Intellectual M.	B. Impact	Keywords	Questions	Is Online?

References

S. Jean, K. Cho, R. Memisevic, and Y. Bengio. On using very large target vocabulary for neural machine translation. *arXiv preprint arXiv:1412.2007*, 2014.

M.-T. Luong, H. Pham, and C. D. Manning. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*, 2015.