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

Razwan Ahmed Tanvir

Date: 04/7/2022

Quote *"... we propose a novel hybrid architecture for NMT that translates mostly at the word level and consults the character components for rare words when necessary. "*

Overview Recent machine translation systems mostly rely on the confined and pre-defined vocabulary and hence the translation models produce unknown words. The author of this paper proposed a novel approach to achieve open vocabulary neural machine translation. The authors claimed that their model combines both word level translation and also consult character level recurrent neural network to recover unseen word. The advantages of this novel hybrid model is that it can be trained faster compared to the character-based models. Moreover, this hybrid model eliminates the possibility of generating unknown words. In this paper, the authors at first identified the shortcomings of the traditional approaches which are mainly phrase-based and have small memory to work with. Furthermore, these models have restricted vocabulary and the words outside of the vocabulary are treated as $< unk >$ symbol. The traditional NMT systems cannot learn a good representation of rare words. To address these issues, the authors model came up with their hybrid model which tackle the unknown words using character-based model. This proposed model is a hierarchical seq2seq model, however, it operates in a fine-grained word-character level. In this model, the core component is a deep Long Short Term Memory encoder-decoder which is trained on a fixed $|V|$ frequent words for each language. Moreover, there is an LSTM model for character level learning which is used to generated character level representation of the rare words. However, the character-based model took 3 months to train with a vocabulary size of 50K.

The authors measured the performance of their model with BLEU score. They compared the results with the state-of-the-art models and showed that their model performed well on translation tasks and also their models work with the unknown words. The authors claimed to have a boost of +2.1 to 11.4 BLEU score. Their best model achieved 20.7 BLEU score on WMT'15 dataset.

Intellectual Merit This study has shown that the hybrid approach is better for handling translation tasks which also tackles the unknown words. This authors implemented a well reasoned hypothesis and the study was structured in terms of model explanations and training to showcasing the result and model analysis. The authors are from Stanford and they have been working in this field for a long time. Manning is a well known name for the NLP community for his important contributions.

Broader Impact This research proposed a novel approach to machine translation tasks which addresses the issues with restricted vocabulary. This research is frequently cited in the field of machine translation. The authors published the code, data and their models publicly. Both of the authors are academics from Stanford and are well known for their work.

Keywords Natural Language Processing, LSTM, NMT, BLEU

Discussion Questions

- How is their model faster, if they are using separate LSTM models for word level and character level representation?
- The training time is huge, 3 months for a character based model. Are there any ways to significantly reduce the amount of time required to train a character-based model?

Table 1: Grade deductions by section

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