

Title	Incorporating syntactical information to improve Neural Machine Translation
Author(s)	TRIEU, Hoang An
Citation	
Issue Date	2023-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/18346
Rights	
Description	Supervisor: NGUYEN, Minh Le, 先端科学技術研究科, 修士(情報科学)

Incorporating syntactical information to improve Neural Machine Translation

1910446 Trieu Hoang An

Recently, researches about machine translation have reached impressive results and can be applied in practical, thank to the development of deep learning model, especially the Transformer model. The Transformer model allows us to build an end-to-end translation system straight from the datasets. However, there are some issues that still exist even with the participation of the Transformer. Those issues are the lack of dataset in minor language and the loss of information while translating long sentences. There is one approach to face with this issue which is incorporating syntactical information into the Transformer model to enrich information from the small dataset in minor language and also help the model to get more information in translation step.

In this research, we proposed the methods to inject the syntactical information such as TF-IDF score for pair of words and Part of speech tags into the Transformer model. With each type of information, we have designed different type of attention mechanism to incorporate those information into Transformer effectively. Our target is to highlight the relationship between the tokens inside the input sequence, since then, enhance the quality of translation process.

To evaluate the methods, we did experiments on two benchmark dataset which are IWSLT 2015 English-Vietnamese and IWSLT 2014 English-Germany and the results is quite positive.