

# Paper: NEURAL MACHINE TRANSLATION BY JOINTLY LEARNING TO ALIGN AND TRANSLATE

**Razwan Ahmed Tanvir**

Date: 02/23/2022

**Quote**      *“... The most important distinguishing feature of this approach from the basic encoder–decoder is that it does not attempt to encode a whole input sentence into a single fixed-length vector. Instead, it encodes the input sentence into a sequence of vectors and chooses a subset of these vectors adaptively while decoding the translation.”*

**Overview**      Neural Machine translation is rather a recent approach to translate source language to target language. However, these neural models belong to the encoder-decoder family where the encoder encodes the source sentences to a fixed-length vector, and from this produced vector, the decoder generates the target translation. The authors of this paper conjectured that the usage of this fixed-length vector from the source sentence acts as a bottleneck for the model’s performance and subsequently proposed a solution that uses a soft-search in the source sentence to find parts of it which is appropriate for the target word. This approach of choosing a subset of the vectors found from the source sentence keeps the model from accruing all the source sentence information into a fixed length vector. Moreover, this trick allows the model to perform better with long sentences. As an underlying model, the authors used an RNN Encoder-Decoder model proposed by Cho in 2014. The proposed architecture has a bidirectional RNN which works as an encoder and a decoder which searches the source sentence to produce translation.

The authors reported their results using the standard BLEU score. Proposed model RNNSearch performs better than the existing conventional RNNEncdec model. One notable result is that the proposed model performs as high as the phrase-based translation system when the known words are considered for the source sentence.

**Intellectual Merit**      The authors in this paper proposed a model with a Bidirectional RNN. Also, they did not generate a fixed length vector from the source sentence, rather their model searches a part of the source sentence to generate appropriate translation. The model description and the experiments of this research is well described and organized. The author measured their success by comparing their model with current conventional translation model. BLUE score was used to measure the translation quality of the model. The authors used WMT’14 English-French parallel corpora. They combined the corpus to have 348M words using the selection method described by Axelrod in 2011.

**Broader Impact**      The authors provided a model which uses a trick so that the model does not have to squash all the source sentence information into a fixed-length vector. The usage of the bidirectional RNN also helps generate better output translation. Dzmitry Bahdanau is currently working at McGill University. The data set is available for public use and there is no mention for the availability of the code for this model. However, the detail of the training procedure is mentioned in the paper to replicate this research.

Keywords    Natural Language Processing, Machine Translation, Bidirectional RNN, Encoder-Decoder

Discussion    • Is there any fixed length for the search of the parts in source sentence? Does it have  
Questions    any impact in the performance?

- Discussion in using soft-max should be more elaborated. Justification in using this method could bring other approaches to solve this problem.

Table 1: Grade deductions by section

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