

Paper: *Sequence to Sequence Learning with Neural Networks*

Razwan Ahmed Tanvir

Date: 02/18/2022

Quote *“... However, the Long Short-Term Memory (LSTM) is known to learn problems with long range temporal dependencies, so an LSTM may succeed in this setting.”*

Overview In this paper, the authors mentioned that Deep Neural Networks are not suitable to map sequence to sequence. The authors proposed an approach which uses multilayered Long Short-Term Memory (LSTM). Their model maps the input sequence, at first, to a vector than another LSTM decodes the target sequence using the mapped vector. This paper at first identifies the current issues with Deep Neural Networks in terms of variable dimensionality. They argued that, tasks like speech recognition, question answering and machine translations are sequential problems and in these cases a domain-independent approach could be successful. To tackle the dimensionality problem of DNNs, the author proposed an LSTM architecture. General idea of this architecture is to read input as a sequence with respect to time so that the fixed dimensional vector can be obtained. After that, Another LSTM produces output from the fixed dimensional vector representation. In addition, the second LSTM in this architecture is mostly a Recurrent Neural Network (RNN). However, the RNN is conditioned on input sentence.

There are several outcomes of this research. The authors reported their model's performance using BLEU scores. On the WMT'14 English To French dataset, their model scored 34.81 in BLEU metric. They claimed that this is the state-of-the-art result for this task as the SMT baseline score is 33.30. Moreover, the most surprising outcome of this paper as mentioned by the authors is that their model does not suffer with the longer sentences. It is achieved by just reversing the order of the words in the source sentence.

Intellectual Merit *This research produces several mentionable outputs and also the output was observed to be better than baseline scores. This research addresses some known issues with Deep Neural Networks in terms of fixed dimension of the input sequence. The approach to solve the issues were novel and also creative in a sense that a subtle modification which is reversing the order of the input sentence, surprisingly impacted the output. The results reported by the authors are well reasoned and compared with the current baseline scores. The researchers are from google and all the necessary resources were available to the researchers from data sets to hardware support to train their models.*

Broader Impact This research outlined and also addressed some of the issues of Deep Neural Networks. They have used multilayered LSTM to tackle the fixed dimension problem of DNNs. State-of-the-art results were produced from this research. This research is frequently followed in text processing research especially, machine translation tasks. The authors used WMT'14 English to French dataset to train their model which took 10 days regardless of using parallelization. The dataset is available public.

Keywords Natural Language Processing, Machine Translation, LSTM, RNN, DNN, Sequence to Sequence

- Discussion Questions
- Authors claimed that they reversed the order of the words in the source sentence. Is this trick can be used in general to obtain better result for all translation generation tasks?
 - There should be more research to identify why reversing the order of the input sequence makes the LSTM learning better.

Table 1: Grade deductions by section

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