# Review of "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling"

Riyad Bin Rafiq

## 1. Paper summary

Recurrent neural networks (RNN) have shown good performance in sequence modeling tasks such as language translation, speech processing and so on. There are three types of recurrent units in RNN including traditional *tanh* unit, long short-term memory (LSTM) unit and gated recurrent unit (GRU). In this paper [1], the authors evaluated these recurring units by applying them on the tasks of polyphonic music modeling and speech signal modeling. The research demonstrated that the recurrent units performed better than the traditional *tanh* unit.

## 2. Contribution

### 2.1 Similarities

As the paper evaluated GRU units and LSTM units, the authors provided the similarities between them with a good explanation and equations. The research explained the gating units and stated on how an important feature is remembered by the forget gate of the LSTM unit or the update gate of the GRU unit.

### 2.2 Differences

The most significant difference between two recurring units is controlling of the memory content. The memory content in the LSTM is controlled by the output gate whereas GRU exposes the full content without any control. Another difference is the location of the input gate of LSTM and the reset gate of GRU. The research explained how new information was added to the memory for the two recurring units.

### 2.3 Dataset and result

The authors considered four types of music dataset and two types of speech dataset for the evaluation. Moreover, they also provided the result and analysis based on both CPU time and number of iterations.

## 3. Critique

### 3.1 Why specific LSTM unit?

In the paper, the authors said that a lot of minor modifications have been made to the original LSTM. So they chose a specific LSTM unit [2] for the research but they didn't mention why they

selected that unit. Is there any particular advantage of using that LSTM unit? In my opinion, this concept should be clear in the research.

### 3.2 Analysis should be in-depth

In the evaluation, it has been seen that the recurrent units (GRU and LSTM) outperformed the traditional *tanh* unit. Furthermore, GRU performed better than LSTM in most cases. But it was not explained why this happened and how GRU did better according to the different datasets. I think there should be an in-depth analysis of the reasons. The authors only stated that the GRU unit did better than the LSTM unit and they were cryptic in explaining the reasons.

## Reference


1. Chung J, Gulcehre C, Cho K, Bengio Y. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. arXiv [cs.NE]. 2014. Available: http://arxiv.org/abs/1412.3555

2. Graves A, Mohamed A-R, Hinton G. Speech recognition with deep recurrent neural networks. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. 2013. pp. 6645–6649.