Mohamed Sameh Saafan

# Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling

## Introduction

- Recurrent Neural Networks (RNNs) are effective for sequence modeling, especially with variable-length inputs and outputs.
- Traditional RNNs struggle with long-term dependencies due to vanishing/exploding gradients.
- This paper compares three types of RNN units: traditional tanh, Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU).

## Background

- RNNs process sequences by maintaining a hidden state over time.
- They are generative models that can output probabilities for sequences of variable length.
- Gradient issues (vanishing/exploding) hinder learning long-term dependencies in vanilla RNNs.

## Gated Recurrent Neural Networks

- LSTM uses memory cells and gates (input, forget, output) to maintain long-term dependencies.
- GRU simplifies LSTM by using update and reset gates without a separate memory cell.
- Both architectures allow additive updates, facilitating gradient flow and long-term learning.

## Experiments

- Tasks: polyphonic music modeling and speech signal modeling.
- Compared LSTM, GRU, and tanh RNNs with similar parameter counts.
- Used RMSProp with gradient clipping and early stopping on validation sets.

## Results and Analysis

- GRU generally performed best on polyphonic music datasets except Nottingham.
- LSTM and GRU significantly outperformed tanh RNNs on speech datasets.
- GRU showed faster convergence in both iterations and wall-clock time.

## Conclusion

- Gated units (LSTM and GRU) are superior to traditional tanh units for sequence modeling.
- GRU and LSTM performed comparably, with slight variations depending on the dataset.
- Further work is needed to isolate the contributions of individual components in gated units.