

# Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling

## Objective:

The paper compares the performance of three types of RNNs: Traditional tanh units, LSTM, GRUs

The focus is on evaluating the effectiveness of gated units (LSTM and GRU) in capturing long-term dependencies

Comparison of Architecture:

- LSTM: Uses three gates input, forget, and output gates.
- GRU: Simpler than LSTM, with reset and update gates to modulate information flow but lacks a separate memory cell.
- Both units employ additive updates, mitigating vanishing gradients by creating shortcut paths for error backpropagation.

## Results:

- Gated units (LSTM, GRU) outperform better than RNNs, especially on speech tasks.
- GRU:
  - Fastest convergence (updates and CPU time)
  - Best overall performance on most music datasets and one speech dataset
- LSTM:
  - Best performance on one speech dataset
- Tanh RNNs: Significantly worse, especially on complex tasks like speech modeling

## Conclusions:

- Gated units (LSTM/GRU) consistently outperformed tanh units, especially in complex tasks like speech modeling.
- No clear winner between LSTM and GRU: Performance depended on the dataset/task.
- GRUs showed computational advantages, often converging faster with comparable accuracy.