Comparison of LSTM and GRU for Sequence Modeling Tasks

This paper compares three types of recurrent units used in Recurrent Neural Networks (RNNs) for sequence modeling tasks:

1. Tanh units (traditional RNNs)
2. Long Short-Term Memory (LSTM) units
3. Gated Recurrent Units (GRU)

The evaluation was done on two types of datasets:

- Polyphonic music datasets (Nottingham, MuseData, etc.)
- Raw speech datasets from Ubisoft

Main Experiments & Findings:

- GRU and LSTM performed significantly better than traditional tanh units.
- GRU performed similarly or better than LSTM, especially in training speed (convergence) and generalization in some datasets.
- On music datasets, all models had close performance, but GRU was usually faster.
- On speech datasets, GRU and LSTM outperformed the tanh model.
- GRU was faster and required fewer updates to converge.

Conclusion:

- Gating mechanisms (LSTM & GRU) improve RNN performance in sequence modeling tasks.
- No clear winner between LSTM and GRU — the best choice may depend on the dataset and task.