Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling

Authors: Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, Yoshua Bengio
arXiv: 1412.3555v1

- Vanilla RNNs suffer from vanishing/exploding gradients.
- Two solutions:
  1. Improved optimization (e.g., gradient clipping).
  2. Gated recurrent units (LSTM, GRU).
- LSTM: Introduced in 1997, includes memory cell and multiple gates.
- GRU: Simpler unit proposed in 2014, fewer gates, no separate memory cell.

Units Evaluated

Traditional Tanh RNN

- Simple recurrence: h_t = tanh(Wx_t + Uh_{t-1})
- Always overwrites memory.
- Struggles with long-term dependencies.

LSTM

- Keeps a separate memory cell (c_t).
- Uses input, forget, and output gates.
- Can selectively remember or forget information.

GRU

- No separate memory cell.
- Uses update and reset gates.
- Combines input and memory in a simpler structure.

Results

Gated Units vs Tanh:

- LSTM and GRU outperformed tanh-RNNs.
- Faster convergence and better accuracy, especially on speech data.

LSTM vs GRU:

- Comparable performance.
- GRU slightly better on music datasets and Ubisoft B and GRU learns faster initially.
- LSTM best on Ubisoft A.