Long short-term memory (LSTM) and gated recurrent units (GRU)

Week 6 Mini Survey

Luis Alberto Portilla López

LSTM and GRU Overview

Both emerged as solutions to the vanishing gradient problem in traditional RNNs

Have become the backbone for sequence-based tasks

Their architectures have proven to be effective in time-series forecasting, natural language processing, and predictive control, balancing complexity and convergence speed depending on the task.

Performance Insights Across Applications

LSTM

- Better control over long-term dependencies
- More complex architecture
- Ideal for tasks requiring fine-tuned long-term memory.

GRU

- More computationally efficient alternative
- Faster convergence while maintaining similar accuracy
- More beneficial in real-time applications

Search Methodology & Criteria

CITATION CHAINING AND FORWARD CITATION

KEYWORD SEARCH

BOOLEAN SEARCH

CRITERIA:

- Initial review of abstracts to assess relevance based on the title, publication venue, and year.
- Direct and indirect relevance to the paper being cross-referenced through the abstract.
- Consideration of the number of citations and field-weighted citation impact (fwci), a metric that measures the citation impact of a paper adjusted for disciplinary differences.

Preliminary Terms

Key terms identified during the week:

- Gated Recurrent Units (GRU)
- AutoRegressive Integrated Moving Average (ARIMA)
- Backpropagation Through Time (BPTT)
- Memory Cells

Document Comparison

"Long Short-Term Memory"

"Using LSTM and GRU neural network methods for traffic flow prediction"

"LSTM and GRU Neural Networks as Models of Dynamical Processes Used in Predictive Control: A Comparison of Models Developed for Two Chemical Reactors"

Characteristic	Long Short-Term Memory (1997)	Using LSTM and GRU (2016)	LSTM and GRU in Predictive Control of Industrial Processes (2021)
Primary Model	Long Short-Term Memory	Long Short-Term Memory and Gated Recurrent Units	LSTM and GRU
Objective	Solving long term dependencies in seq. data, various synthetic taks	Short-term traffic flow prediction	Industrial process control (chemical reactors)
Application	General purpose neural network for sequence taks	Intelligent Transportation Systems and traffic flow prediction	Model Predictive Control (MPC) for chemical reactors
Dataset	Simulated tasks, Reber grammar, 2-sequence problem, Adding problem, Multiplication problem	Caltrans PeMS dataset (real-world traffic data from California)	Simulated industrial processes (polymerization and pH control)
Performance Metrics	Qualitative performance (no specific MSE/MAE provided)	MSE (Mean Squared Error), MAE (Mean Absolute Error)	Model accuracy, control quality, computation time
MSE (Mean Squared Error)	N/A	LSTM: 710.05, GRU: 668.93, ARIMA: 841.00	N/A
MAE (Mean Absolute Error)	N/A	LSTM: 18.13, GRU: 17.21, ARIMA: 19.18	N/A
Convergence Speed	N/A	GRU converges faster than LSTM	GRU converges faster than LSTM

References

1. Long Short-Term Memory (LSTM Paper):

• Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

2. Using LSTM and GRU for Traffic Flow Prediction:

• Fu, R., Zhang, Z., & Li, L. (2016). Using LSTM and GRU neural network methods for traffic flow prediction. In IEEE Youth Academic Conference (pp. 7804912). https://doi.org/10.1109/YAC.2016.7804912

3. LSTM and GRU in Predictive Control of Industrial Processes:

• Zarzycki, K., & Ławryńczuk, M. (2021). LSTM and GRU Neural Networks as Models of Dynamical Processes Used in Predictive Control: A Comparison of Models Developed for Two Chemical Reactors. Sensors, 21(16), 5625. https://doi.org/10.3390/s21165625