A Novel Recurrence Truly for Long-Sequence Processing

MAXIMAL EXPRESSIVITY | LONG-TERM MEMORY | HIGH PARALLELISABILITY | GUARANTEED STABILITY | LONG-RANGE SENSITIVITY

Motivation

Properties for truly long-sequence processing

Strong expressivity

Long-term memory, capturing star-free up to all regular.

Efficient Evaluation

Sequential: time $O(T)$.
Parallel: depth $O(\log T)$, work $O(T)$.

Stability

Bounded activations and gradients, independently of sequence length.

Long-range Sensitivity

Sensitive to early inputs, for long-range dependencies.

Existing Architectures

Transformers

- ✗ Expressivity excludes basic languages
- ✗ Quadratic work
- ✓ Stable
- ✗ Limited-range sensitivity

Contractive Linear RNNs • Mamba, mLSTM, Gated DeltaNet

- ✗ Limited expressivity, no long-term memory
- ✓ Computationally efficient
- ✓ Stable
- ✗ Limited-range sensitivity

Non-expansive Linear RNNs • DeltaNet, DeltaProduct

- ✓ Strongly expressive
- ✓ Computationally efficient
- ✗ Unstable
- ✓ Long-range sensitivity

References

Transformers: *Attention is All you Need*, Vaswani et al., NeurIPS 2017
Mamba: *Mamba: Linear-Time Sequence Modeling with Selective State Spaces*, Gu and Dao, COLM 2024
mLSTM: *xLSTM: Extended Long Short-Term Memory*, Beck et al., NeurIPS 2024
DeltaNet: *Parallelizing Linear Transformers with the Delta Rule over Sequence Length*, Yang et al., NeurIPS 2024
Gated DeltaNet: *Gated Delta Networks: Improving Mamba2 with Delta Rule*, Yang, Kautz, and Hatamizadeh, ICLR 2025
DeltaProduct: *DeltaProduct: Improving State-Tracking in Linear RNNs via Householder Products*, Siems et al., NeurIPS 2025

MinMax RNCs

Replacing addition and multiplication with max and min at the level of scalars gives the **MinMax algebra**: $(+,\times) \rightsquigarrow (\oplus,\otimes)$

MinMax Recurrence

MinMax Recurrence is the affine recurrence in the MinMax algebra

$$x_t = \big(R(u_t) \otimes x_{t-1}\big) \oplus s(u_t)$$

Example. For $R(a,b) = a$ and $s(a,b) = b$, on state $x_0 = 0$ and input $u = (0,2),(7,0),(5,0),(0,1)$,
$x_1 = (0 \otimes 0) \oplus 2 = 2, \quad x_2 = (7 \otimes 2) \oplus 0 = 2, \quad x_3 = (5 \otimes 2) \oplus 0 = 2, \quad x_4 = (0 \otimes 2) \oplus 1 = 1.$

MinMax RNCs are cascades of layers of neurons that employ MinMax Recurrence with R, s given by feedforward networks.

Properties of MinMax RNCs

✓ Expressivity of Regular Languages

- MinMax neurons can realise all fundamental semiautomata $\rightsquigarrow$ Regular languages.
- The state-dimension of neurons matches the degree of the permutation semiautomata they can realise $\rightsquigarrow$ State-dimension one suffices for the star-free languages.

✓ Efficient Evaluation • Recurrence and Parallel Scan

- Sequential evaluation via recurrence, and parallel via associative MinMax composition.

✓ Stability

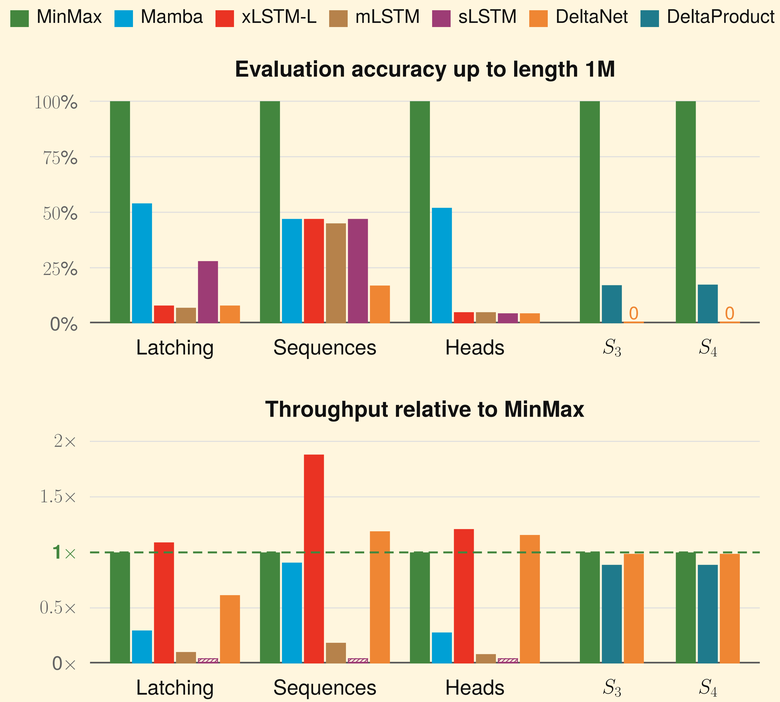
- Uniformly bounded states/activations are by MinMax operations simply selecting values.
- Uniformly bounded pseudo-gradients that coincide with gradients almost everywhere.

✓ Long-range Sensitivity

- Over arbitrarily-long ranges, MinMax recurrence can propagate state coordinates unchanged or permuted $\rightsquigarrow$ State-to-state pseudo-gradients of unit norm.

Experiments

Only MinMax solves every task: 100%.



Accuracy: evaluation accuracy (%), means of 99% CIs with max margin 1.7.
Throughput: avg. train-eval tok/s relative to MinMax, means of 99% CIs with max margin 0.4.

On OWT, **MinMax RNC** of 112M parameters with scalar-state neurons reaches the same validation loss of **3.12** as GPT-2 of 124M parameters.

Read the paper!

MinMax Recurrent Neural Cascades

Alessandro Ronca • ronca@iris-ai.org

Download the code!