Lyapunov-Guided Representation of RNN Performance

Ryan Vogt[1], Yang Zheng[2], Eli Shlizerman[1,2]

[1]Department of Applied Mathematics, University of Washington, USA

[2]Department of Electrical and Computer Engineering, University of Washington, USA

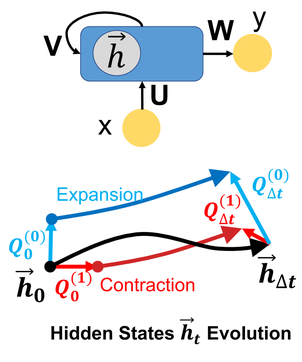
NSF Award Number: 2117997

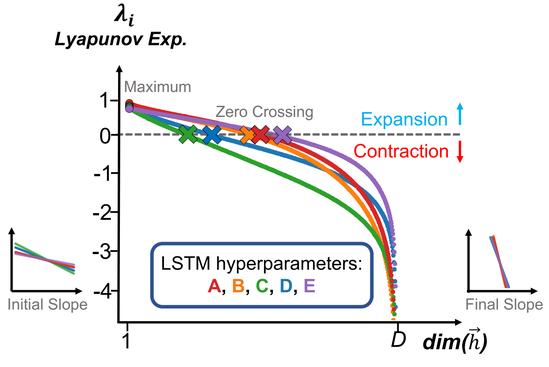
Motivation

- **RNN** are the backbone of **time series and sequences modeling**
- Architectural variants of RNNs only represent **singular points** in the **space of RNN**
- **RNN accuracy** depends on **hyperparameter configurations**
- Given RNN variant, predicting its accuracy before training is limited.

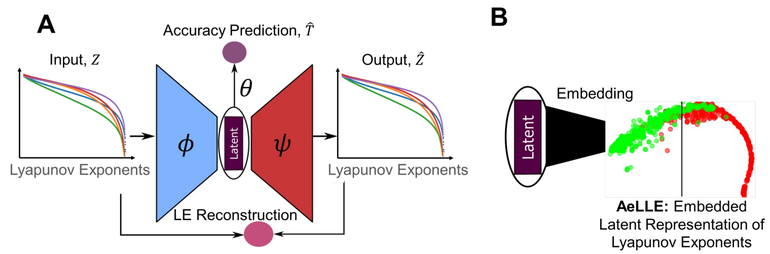
Background

- A powerful dynamical system method for characterization and predictability of dynamical systems is ***Lyapunov Exponents (LEs).***
- *LEs* capture the information generation by a system's dynamics through measurement of the separation rate of infinitesimally close trajectories.
- However, the connection between LE and network performance has not been explored extensively.





Method



AeLLE (***A***uto***e***ncoder ***L***atent ***L***yapunov ***E***xponents)

- Computation of LE spectrum

$$[\mathbf{J}_t]_{ij} = \frac{\partial \mathbf{h}_t^j}{\partial \mathbf{h}_{t-1}^i}. \quad \mathbf{Q}_{t+1}, \mathbf{R}_{t+1} = QR(\mathbf{J}_t \mathbf{Q}_t). \quad \lambda_k = \frac{1}{T}\sum_{t=1}^{T} \log(r_t^k)$$

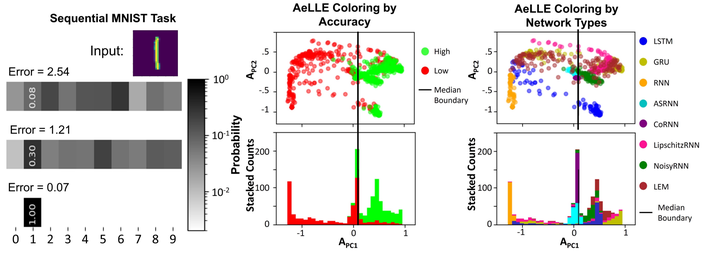
- Autoencoder for LE spectrum

$$\hat{Z} = (\psi \circ \phi)Z, \qquad \phi, \psi, \theta = \arg\min_{\phi,\psi,\theta} (\|Z - \hat{Z}\|^2 + \alpha \cdot \|T - \hat{T}\|_l).$$
$$\hat{T} = (\theta \circ \phi)Z,$$

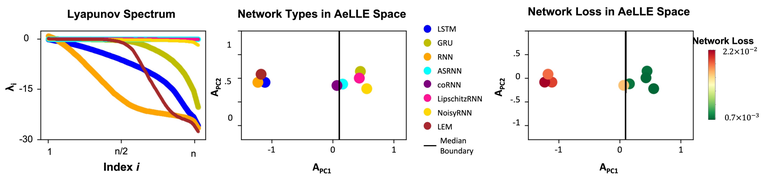
- Embedding of Autoencoder Latent representation

Experiments

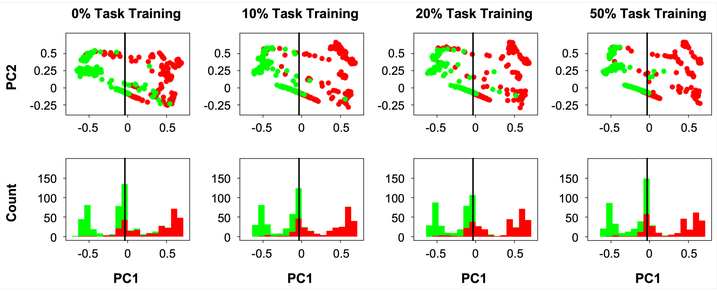
Sequential MNIST Classification with Different Network Type



LE Features Visualization in AeLLE Space



Pre-Trained AeLLE for Accuracy Prediction Across Training Epoch



Training %	**AeLLE** vs. [Loss] Recall		Precision	F1
0%	92.2%	[–]	81.9% [–]	0.87 [–]
10%	**96.9%**	[21.4%]	**83.6%** [100%]	**0.90** [0.35]
20%	**96.7%**	[49.0%]	**82.8%** [100%]	**0.89** [0.66]
50%	**95.3%**	[70.4%]	**81.0%** [99%]	**0.88** [0.82]

Table. Pre-trained AeLLE classifier vs. loss-based classifier evaluated at different stages of training

References

[1]Vogt, R. et al. On Lyapunov exponents for RNNs: Understanding information propagation using dynamical systems tools. Frontiers in Applied Mathematics and Statistics 8 (2022). https://doi.org/10.3389/fams.2022.818799

[2] Engelken, Rainer. et al. Lyapunov spectra of chaotic recurrent neural networks." arXiv preprint arXiv:2006.02427 (2020).

[3] LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324.

[4] Ryan, et al. "Lyapunov-Guided Embedding for Hyperparameter Selection in Recurrent Neural Networks." arXiv preprint arXiv:2204.04876 (2022)