The structure of complex neural networks and its effects on learning

Pau Vilimelis Aceituno [1,3] Gang Yan [2] Yang-Yu Liu [3,4]

[1]Max Planck Institute for Mathematics in the Sciences

[2]School of Physics Science and Engineering, Tongji University

[3]Channing Center for Complex Medicine, Brigham and Womens' Hospital and Harvard Medical School

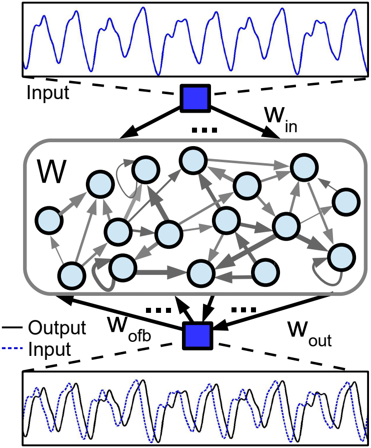
[4]Dana Faber Center for Cancer Research

Reservoir Computing[1,2]: Introduction

Reservoir is:

- A Recurrent Neural Network
- A dynamical system

Training: Linear regression from the dynamical system to the target



System equations

$$\mathbf{x}(t) = f(\mathbf{W}\mathbf{x}(t-1) + \mathbf{w}_{\text{in}} u(t))$$
$$y(t) = \mathbf{w}_{\text{out}} \mathbf{x}(t)$$
$$\mathbf{w}_{\text{out}} = \arg\min \sum (y(t) - s(t))^2$$

[1] Jaeger, H. and Haass, H. "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication." Science (2004)

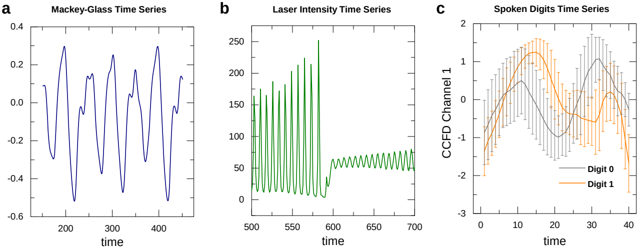
[2] Maass, Wolfgang, Thomas Natschläger, and Henry Markram. "Real-time computing without stable states: A new framework for neural computation based on perturbations." Neural computation (2002)

Approach

Method

- Machine Learning: Define error and problem
- Statistical Physics/Control Theory: Adapt System Dynamics
- Graph Theory: Find the right network features
- **Test**: Try many network structures/params and compare

Datasets



a: Mackey, and Glass. "Oscillation and chaos in physiological control systems". Science (1974)

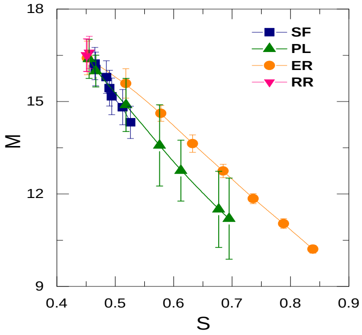
b: Huebner, et al. "On problems encountered with dimension calculations." Measures of Complexity and Chaos. (1989)

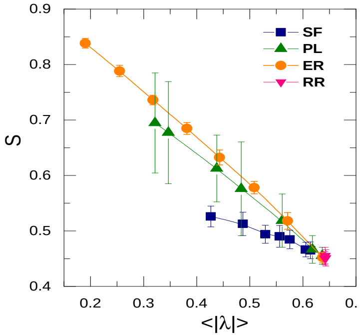
c: Hammami, et al. "Improved tree model for arabic speech recognition." ICCISIT 2010

Memory and Correlations: Theory

The main feature of Recurrent Neural Networks is **memory**[4]: M

- Memory $\iff$ Independent Variables $\iff$ low **correlations**: S
- Low Correlations $\iff$ Large **eigenvalues**: $\langle|\lambda|\rangle$
- $\langle|\lambda|\rangle$ is a proxy for M

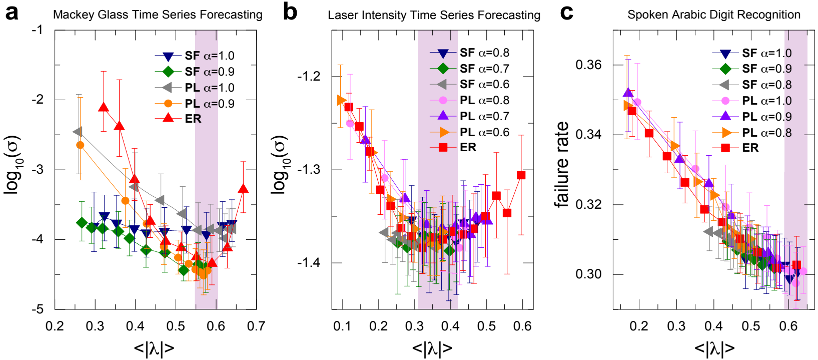




[2] Jaeger, H. "Short-Term Memory in Echo State Networks". Technical Report

Memory vs Lambda: Practical effects

Different tasks require different memories, but a single task requires the same memory for all reservoir instances ⇒ The performance for a single task should be maximal at a single $\langle|\lambda|\rangle$



Note: Laser Intensity Time series has two timescales ⇒ Multiple memory ranges

Frequency and Cycles

Main Idea

- If the reservoir resembles the target, the regression is better.
- Adapt the reservoir to the **frequencies** in the signal.

Proof

Linear regression: project N points in T-D space to a line

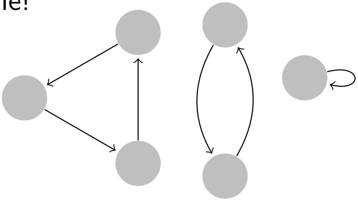
$\Rightarrow$If the points are close to the line it works better

In the frequency domain: The same!

How to

How to adapt frequency?

$\rightarrow$ Feedback loops (cycles)

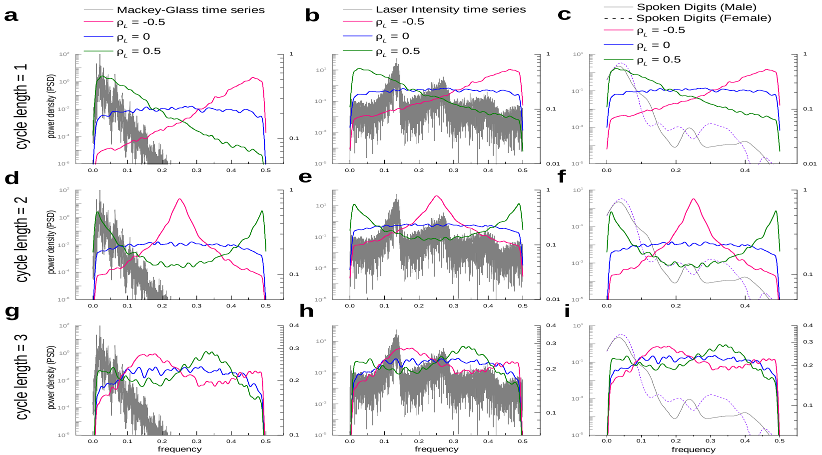


Proof

Frequencies to autocorrelations: Wiener-Khichin Theorem

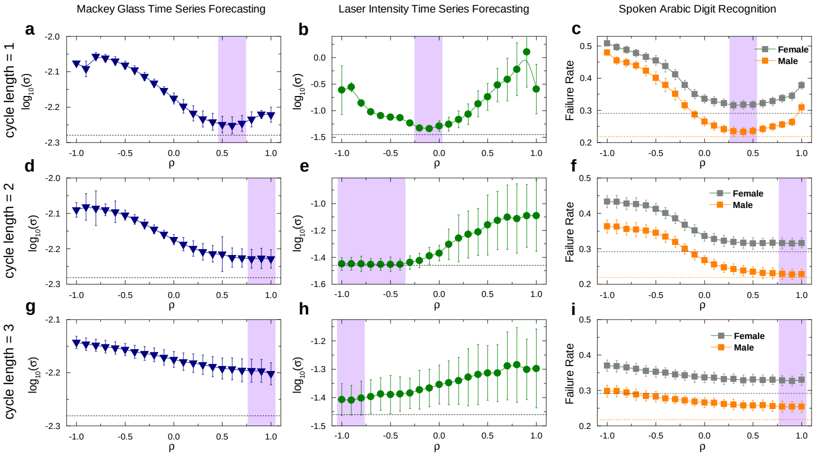
Autocorrelations to weights: Mean value theorem on the derivative

Controlling Dynamics: Adapt frequency



Performance improvement

- Heuristic: Independently optimize $\langle|\lambda|\rangle$ and ρ_L.



Conclusion

Summary:

- Correlations control memory: Use eigenvalues
- Adapting the frequency is useful: Use cycles

Open questions:

- Machine Learning: Does training give this structures?
- Neuroscience: Does synaptic plasticity enhance frequencies?
- Applications: Neuromorphic hardware, network-level training

Many Thanks to...

- Collaborators: Yang-Yu Liu, Gang Yan
- "la Caixa" foundation for paying
- Herbert Jaeger for discussion and code

Reference: "Tailoring Recurrent Neural Networks for Optimal Learning", arxiv.org: 1707.02469