Information Processing with neuro-inspired delay-based nonlinear systems

Claudio R. Mirasso

Instituto de Física Interdisciplinar y Sistemas Complejos

Palma de Mallorca, Spain

1st Summer School of “Interdisciplinary Research on Brain Network Dynamics”, Terzolas, June 24-28 2019.
Ingo Fischer

Miguel Cornelles

Apostolos Argirys

Daniel Brunner

Silvia Ortín

Miguel Escalona
Quiz: computer vs. brain
Question 1

\[ \sqrt{4} = ? \]

Options:
- 2
- 1.6
- 16
Question 2

\[ \sqrt{208849} = ? \]

\[ \boxed{457} \]

Options:
1. 443
2. 457
3. 917
Question 3
Brain to process information
The complex brain

• Network of $\sim 10^{11}$ neurons with $\sim 10^{14}$ links (synapses)

• Multi-scale structure:
  • microscopic to macroscopic: neurons - columns – areas/structures– whole brain
  • fast to slow: spiking - population activity – plasticity

The brain is “simply” a multiscale complex recurrent network of heterogeneous elements that (usually) self-organizes, delay coupled, performs computation
Artificial neural networks
Introduction

Delayed feedback systems

How similar behave delayed feedback systems and neural networks?
Can a delay system mimic information processing in the brain?

Started to develop within the FET Open Project PHOCUS (2010-2013)
The idea was initiated with the artificial neural networks

Recurrent networks: connections were trained, which complicated enormously the training procedure.

It was in 1995 when Buonomano & Merzenich proposed the idea of using a randomly (fixed) coupled I&F neural network connected to an output layer whose weights were trained for certain tasks.

This was the origin of the machine learning technique known as “Reservoir Computing”.

Reservoir Computing:

Neuro-inspired concept

Consider brain a “black-box“ complex recurrent network

Analyzes transient responses to (sensory) input

Echo State Networks
(H. Jaeger, 2002)

Liquid State Machines
(W. Maass et al 2003)
How does Reservoir Computing work?
Utilizes the projection of an input state (usually low dimensional) onto a high-dimensional feature space

- Nonlinear mapping onto a higher dimensional state space can make a classification problem linearly separable
- Linear separability becomes exponentially more likely with increasing state space dimension
Key Properties for RC

**Consistency:** responses to same inputs must be consistent

**Approximation:** two closed enough inputs yield the same output

**Separation:** Different enough inputs must be classified into different outputs

**Fading memory:** Processing in the context of previous states

**Network Connections:** network connections are kept fix

**Readout:** Apply training procedure; only the readout weights are adjusted

**Universality:** capability to perform any computation

Physiological Evidence


Classical attractor states cannot be realistically reached in typical neuronal phenomena timescales.

Transient dynamics as in “liquid-state machines” allow for computation over time without any need for a classical attractor state.

Since the transients are input-specific, they contain information about what caused them.

Experimental observations in the olfactory systems of locust and zebrafish support such framework.
Physiological Evidence

Stable transients are observed whether a stimulus is sustained or not.

When the responses to several stimuli are compared, the distances between the trajectories corresponding to each stimulus are greatest during the transients.

Because these states are read out (“decoded”) by yet other neuronal populations, stimulus identification should be more reliable with transients than with fixed-point states.

Schematic of an antennal lobe; trajectories, representing the succession of states visited by the neural network in response to one odor; dynamical model of transients: dissipative Saddles sequentially connected by unstable separatrices.
Physiological Evidence


Multielectrode recordings from cat primary visual cortex and analyzed the temporal evolution of stimulus-related information in the spiking activity of large ensembles (100 neurons)
When a single letter was shown for a duration 100 ms, the performance (% of correctly identified symbols) was very high.

Under multiple images, the response to an image contains as much information about the preceding image as about the current one.
Information Processing

![Image of a computer screen with a brain and circuit patterns]
The drawback of RC is its hardware implementation

Usually requires 100’s-1000’s nodes

Can the complex network forming the reservoir be replaced by a single nonlinear node?

Introducing delayed feedback / coupling makes the system high-dimensional.
\[ \dot{x}(t) = F(x(t), x(t - \tau); p) \]

**Math.** Phase space is infinite-dimensional space of continuous functions \( C \) on the interval \([-\tau, 0]\)

**Phys.** Degrees of freedom are distributed within the feedback loop (continuous function on the interval \([-\tau, 0]\))
In how far can we replace the complex network by the delay system?

Many degrees of freedom distributed within the delay loop → virtual nodes within the delay

Fading memory introduced by delay

Consistency → initial steady state

Facilitates hardware implementation
Why a hardware implementation is important?

- Era of BIG data → Era of HUGE Data → Processing speed is crucial
- Energy Consumption → More Efficient
- New kind of computation → Traditional computers
- It can be integrated with already existing devices
- It would allow for parallel (and architecture-adapted) implementations

But, overall, with our approach we expect to learn about the basic mechanisms once reducing the concept to the minimal ingredients.
Is the connectivity enough?


What is the minimal complexity of reservoir construction for obtaining competitive models? and what is the memory capacity of such simplified reservoirs?

A simple *deterministically* constructed cycle reservoir is comparable to the standard echo state network methodology.
How do we feed the information into the virtual nodes?
How do we feed the information?

- **$u(t)$**: input states, time-varying scalar variable or vector of any dimension
- Feeding the virtual nodes: serializing the input using time-multiplexing
- Each state vector $u(t_0)$ is fed into the N virtual nodes during one delay interval $\tau$.
- Coupling weights from input layer to virtual nodes is done by (random) masking.
• Each virtual node receives:
  \[ J(t) = M(t) \times u(t_0) \text{ for } 0 < t < \tau \]
• After \( \tau \):
  • Input vector \( u(t) \) changes
  • the states of virtual nodes are updated \( \rightarrow \) new reservoir state
• reservoir states are read out for further processing
• weighted sum of the states to approximate the target is done off-line after demultiplexing
Reservoir Computing based on Delay-dynamical Systems

Lennert Appeltant

Joint PhD
Vrije Universiteit Brussel
Universitat de les Illes Balears
May 2012

Available @ https://ifisc.uib-csic.es/en/publications/
Information processing using a single dynamical node as complex system

L. Appeltant¹, M.C. Soriano², G. Van der Sande¹, J. Danckaert¹, S. Massar³, J. Dambre⁴, B. Schrauwen⁴, C.R. Mirasso² & I. Fischer²

Parallel photonic information processing at gigabyte per second data rates using transient states

Daniel Brunner¹, Miguel C. Soriano¹, Claudio R. Mirasso¹ & Ingo Fischer¹
Complex tasks

To prove the potential of the proposed system, several benchmark tasks need to be overcome.

- Pattern recognition tasks
- Classification tasks
- Time series prediction
- Dynamical System Modeling

Among these, voice & speaker recognition is one of the most challenging task!
Complex tasks

Seven

Ingo Fischer

Miguel C. Soriano

One

Jan Danckaert

Lennert Appeltant

Nine

Daniel Brunner

Laurent Larger
Spoken digit recognition
(Chaotic) time series prediction
How important is the nonlinearity?

**electronic**

Mackey-Glass nonlinearity


**opto-electronic**

Ikeda nonlinearity


**all optical**

Laser diode nonlinearity

Mackey-Glass Oscillator

- Mackey-Glass oscillator with delay

\[
\dot{X}(t) = -X(t) + \frac{\eta \cdot [X(t - \tau) + \gamma \cdot J(t)]}{1 + [X(t - \tau) + \gamma \cdot J(t)]^p}
\]

- \(\eta\): feedback strength
- \(\gamma\): Amplification factor
- \(\tau\): delay time
- \(p\): exponent
- Optimum of WER for certain input scaling. Good agreement between experiment and numerical simulations.
- Min(WER) ~ 0.2 %, meaning 1 misclassification in 500 words
- Better performance than for 1200-node RC and other approaches!
**Opto-electronic System**

Large delay condition, $\vartheta=52.18$ ns ($\tau_D \sim 20.87 \, \mu s$, $T_R=240$ ns)

Required input and read-out resolution: 40 MSamples/s.

**Spoken Digit Recognition**

**Time Series Prediction**
Can we process faster?

**Semiconductor lasers**

- **Size**: 1μm..... 1mm
- **Operation**: electrical & optical
- **Speed**: Hundredths of Gbit/s
- **Applications**: Optical Comm., Blu-ray, etc.
- **Info Processing**: potentially high speed

**Optical Fibers**

- **Size**: 1μm... 100μm diameter
- **Capacity**: Thousands of channels
- **Coupling**: easy to couple to the laser
- **Applications**: Optical Comm., medicine, etc.
All-optical Setup

- Tunable laser
  - Coupler
  - Attenuator
  - Circulator
  - Oscilloscope - 16 GHz
  - PD
  - Pol. Cont
  - τD
  - MZM

Input:
- $u^{(e)}(t)$
- $u^{(o)}(t)$
• Excellent performance with 0% WER biasing close to threshold
• Potentially 300,000 words/s can be classified
• Simultaneous classification of words and speakers
• Energy consumption of about 10 mJ compared with 2 J in computers
- Optical Injection
- Best performance 10.6%, 5% after averaging
- Traditional RC below 1% error (with external memory)
- Prediction rate of $1.3 \times 10^7$ points/second
From Reservoir Computing to Extreme Learning Machines

RC  ELM

Activating or deactivating a single connection (the feedback connection of the neuron) we can easily switch between both learning machines.

Our simple scheme enables a hardware implementation of ELMs and RC with an almost effortless mechanism.

Is a simplification of (one-layer) feedforward neural networks, suitable for pattern classification problems.

Correspond to the RC without inter-neuron connectivity.
Summary & Conclusions

• A simple nonlinear system with delayed feedback can process information!

• Neuro-inspired concept

• Replacing complex network by a delay system
  • Enormous simplification for hardware implementations

• Conceptually simple and potentially cheap system

• Classification and time series prediction tasks already demonstrated

• Enables new kind of computation

• Potentially energy efficient

• Electronic & Optical & Mixed implementations are feasible
THANK YOU
for your attention