24rd Italian Workshop on Neural Networks (Vietri sul Mare)

Significance-based Pruning for Reservoir's Neurons in Echo State Networks

Scardapane S., Comminiello D., Scarpiniti M. and Uncini A.

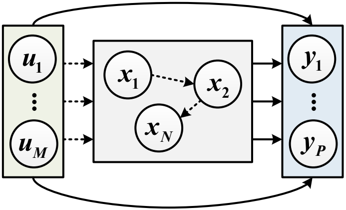
intelligent signal processing
and multimedia lab

Table of Contents

What is an Echo State Network?

An *Echo State Network* (ESN) is a *Recurrent Neural Network* (RNN), whose processing is partitioned in two components:

- A recurrent reservoir, which is *fixed in the beginning*.
- A *static* readout, trained using linear regression techniques.



ESN Update

Given an input $\mathbf{u}(n)$, the internal state $\mathbf{x}(n)$ of the reservoir's neurons is computed as:

$$\mathbf{x}(n) = \tanh(\mathbf{W}_r^r \mathbf{x}(n-1) + \mathbf{W}_i^r \mathbf{u}(n)) \tag{1}$$

where $\mathbf{W}_r^r$ and $\mathbf{W}_i^r$ are *randomly* generated (albeit with some caveats [LJ09]). Similarly, the output is computed as:

$$y(n) = \mathbf{w}_r^o \mathbf{x}(n) + \mathbf{w}_i^o \mathbf{u}(n) \tag{2}$$

where $\mathbf{w}_r^o$ and $\mathbf{w}_i^o$ must be learned. We also define the "extended" state $\mathbf{s}(n) = \left[\mathbf{u}(n)^T \; \mathbf{x}(n)^T\right]^T$.

ESN Training

1. We fed the network with a sequence of inputs $\{\mathbf{u}(1), \dots, \mathbf{u}(S)\}$ and compute the corresponding states.
2. We concatenate the states in the matrix $\mathbf{A} = [\mathbf{s}(1), \dots, \mathbf{s}(S)]$, and the desired outputs in $\mathbf{d} = [d(1), \dots, d(S)]$.
3. We compute the optimal output weights as:

$$\mathbf{w} = \left(\mathbf{A}^T\mathbf{A} + \lambda\mathbf{I}\right)^{-1}\mathbf{A}^T\mathbf{d} \tag{3}$$

where $\mathbf{I}$ is the identity matrix and λ a suitable positive scalar term.

Characteristics of a Reservoir

A reservoir must possess the following properties:

1. It must be *stable*. This is expressed in term of the so-called *echo state property* [LJ09].
2. It should be "big enough" to ensure optimal generalization capabilities.
3. Additionally, its connections (*synapses*) should be sparse to have heterogeneous states.

Few research has been devoted to the last problem, of obtaining optimally sparse reservoirs in an efficient manner.

Significance of a Synapse

We define the *significance* of a synapse at time-instant n as [SNC+14]:

$$s_{ij}(n) = \frac{1}{T} \sum_{z=n-T}^{n} \frac{(x_i(z-1) - \hat{\mu}_x)(x_j(z) - \hat{\mu}_x)}{\hat{\sigma}_x^2} \tag{4}$$

Using it, the probability of removing a given synapse is provided by:

$$p_{ij}(n) = \exp\left\{ -\frac{|s_{ij}(n)|}{t(n)} \right\} \tag{5}$$

where $t(n)$ is monotonically decreasing in n. This is inspired to the *Simulated Annealing* optimization algorithm.

Significance of a Neuron

Can we extend this to the direct pruning of neurons? We define the significance of a neuron as:

$$s_j(n) = \frac{1}{2M} \sum_{z \in \mathcal{I}_j(n)} s_{jz}(n) + \frac{1}{2N} \sum_{z \in \mathcal{O}_j(n)} s_{zj}(n) \tag{6}$$

where $\mathcal{I}_j(n)$ and $\mathcal{O}_j(n)$ are the incoming and outgoing connections of the j-th neuron at time-instant n. We use this to prune neurons in a similar way with respect to before.

In this way, "significant" clusters of neurons should be preserved, while neurons in uncorrelated groups should be removed.

Extendend Polynomial

We consider the extended polynomial detailed in [BVS+13]. The input is a random number extracted from an uniform distribution over $[-1, +1]$. The output is given by:

$$y(n) = \sum_{i=0}^{p} \sum_{j=0}^{p-i} c_{ij} u^i(n) u^j(n-d) \tag{7}$$

where the coefficients c_{ij} are randomly distributed over the same distribution as the input data, and the two parameters p and d control the requirements of the task in term of memory and non-linearity.

Average MSE

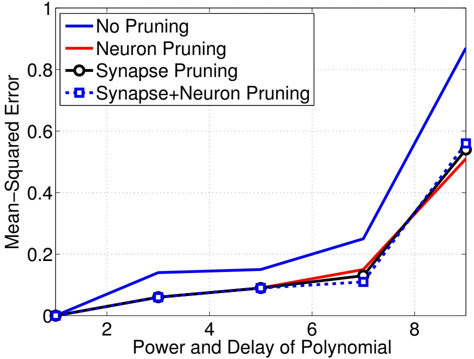


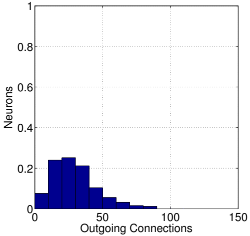
Figure : Average MSE when increasing simultaneously the delay and power of the polynomial from 1 to 9.

Original and Pruned ESNs

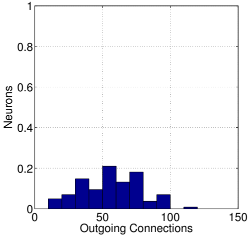
Case	Neurons	Synapses
Original	250	62500
Synapse pruning	250	7700
Neuron pruning	110	12000
Full pruning	50	1700

Putting together the two strategies, we obtain an optimally sparse ESN in an unsupervised way.

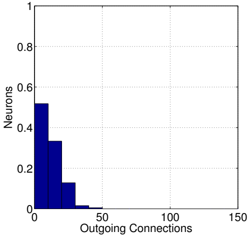
Analysis of the Reservoirs



(a) Synapse pruning



(b) Neuron pruning



(c) Simultaneous pruning of neurons and synapses

Figure : Histogram of the average number of outgoing connections in the ESN, after applying the three pruning strategies.

References

J.B. Butcher, D. Verstraeten, B. Schrauwen, C.R. Day, and P.W. Haycock.
Reservoir computing and extreme learning machines for non-linear time-series data analysis.
Neural networks, 38:76–89, February 2013.

Mantas LukošEvičIus and Herbert Jaeger.
Reservoir computing approaches to recurrent neural network training.
Computer Science Review, 3(3):127–149, 2009.

S. Scardapane, G. Nocco, D. Comminiello, M. Scarpiniti, and A. Uncini.
An effective criterion for pruning reservoir's connections in echo state networks.
2014 International Joint Conference in Neural Networks (Accepted for Presentation), 2014.