Notes on Homeostatic Adaptation and Error Driven Adaptation in ESNs

Fabian Schubert

September 29, 2019

1 Introduction

Strategies for the optimization of ESN hyperparameters can be divided in two categories: supervised and unsupervised methods, where the first one utilizes an error signal, while the latter only uses information contained within the network dynamics.

In the first part of our research, we investigated the possibility of defining an unsupervised homeostatic mechanism that controls the mean and variance of neuronal firing in such a way that the network acts in a regime that yields good performance in sequence learning tasks. This mechanism acts on two sets of parameters, biases b_i and neural gain factors a_i. It should be emphasized that we did not attempt to define an arbitrarily complex measure that would be most suitable for optimization, e.g. from a machine learning perspective. Rather, we restricted ourselves to adhere to ´biologically plausible' mechanisms. While no exact definition of this term exists, it embraced two aspects in our work:

- The dynamics of all variables must be local, i.e., they are bound to a specific neuron and may only access other variables that are locally accessible. In a strict sense, this means all other dynamic variables of the neuron itself and information about the activity of adjacent neurons.

- We use a time-discrete model where the state of a variable in the next step may only be determined by states of the previous step. This means that information about past states must be integrated dynamically.

Our approach was based on the assumption that network performance is optimal when the spectral radius of the effective recurrent connectivity, given by $a_i W_{ij}$, is close to, but slightly below 1. We attempted to transfer this non-local measure into a condition that could be implemented in a biologically plausible way.

2 Model

2.1 Network dynamics

$$y_i(t) = \tanh\left(a_i x_i(t) - b_i\right) \tag{1}$$

$$x_i(t) = \sum_{j=1}^{N} W_{ij} y_j(t-1) + \sum_{j=1}^{D_{\text{in}}} W_{ij}^{\text{u}} u_j(t) \tag{2}$$

$$o_i(t) = o_i^0 + \sum_{j=1}^{D_{\text{out}}} W_{ij}^{\text{o}} y_j(t) \tag{3}$$

where $\mathbf{y}, \mathbf{x}, \mathbf{a}, \mathbf{b} \in \mathbb{R}^N$, $W \in \mathbb{R}^{N \times N}$, $\mathbf{u} \in \mathbb{R}^{D_{\text{in}}}$, $W^{\text{u}} \in \mathbb{R}^{N \times D_{\text{in}}}$, $\mathbf{o}, \mathbf{o}^0 \in \mathbb{R}^{D_{\text{out}}}$ and $W^{\text{o}} \in \mathbb{R}^{D_{\text{out}} \times N}$.

Furthermore

$$\text{p}\left(W_{ij} = x\right) = \begin{cases} \delta(x) & i = j \\ p_{\text{r}} \mathcal{N}\left(x, \mu = 0, \sigma = \sigma_{\text{w}}/\sqrt{N p_{\text{r}}}\right) + (1-p_{\text{r}})\delta(x) & \text{else} \end{cases} \tag{4}$$

$$\text{p}\left(W_{ij}^{\text{u}} = x\right) = \mathcal{N}\left(x, \mu = 0, \sigma = 1\right) . \tag{5}$$

Initially, we chose $N = 1000$ as the network size, however, due to computational complexity, the results presented here are generated with a network of size $N = 500$, unless stated otherwise. See Table 1 for the standard network parameters.

Table 1: Standard network parameters.

N	D_{in}	D_{out}	p_{r}	σ_{w}
500	1	1	0.1	1

2.2 Homeostatic Adaptation

For our homeostatic update mechanism, we use the following dynamics:

$$b_i(t) = b_i(t-1) + \epsilon_{\text{b}} \left[y_i(t) - \mu_i^{\text{t}} \right] \tag{6}$$

$$\mu_i^{\text{y}}(t) = \left[1 - \epsilon_\mu \right] \mu_i^{\text{y}}(t-1) + \epsilon_\mu y_i(t) \tag{7}$$

$$a_i(t) = a_i(t-1) + \epsilon_{\text{a}} [\sigma_i^{\text{t}^2} - (y_i(t) - \mu_i^{\text{y}}(t))^2] . \tag{8}$$

See Table 2 for the standard values.

Table 2: Standard homeostasis parameters.

ϵ_{b}	ϵ_μ	ϵ_{a}	μ_i^{t}	σ_i^{t}
10^{-3}	10^{-4}	10^{-3}	0.05	variable

3 Theory

We stated in the model description that all recurrent weights were drawn independently from a sparse Gaussian distribution. In some sense, of course, this is an assumption that already fulfills one of the conditions that is known to make ESNs work, namely a balance between excitation and inhibition. However, since adjusting gains does not provide a means to dynamically achieve this property, we had to take it as given.

Furthermore, we assumed that all entries of the weight matrix were independently drawn from the same distribution. Under this assumption, using the circular law, it follows that setting all gain values to $1/\sigma_{\text{w}}$ will result in a uniform unit circle distribution of eigenvalues. The situation is less clear if we allow the rows or columns of the matrix to follow different distributions, in particular—since we assumed to have zero mean—different variances. Numerically, we found that if $\sum_i a_i^2 \sigma_{\text{w},i}^2 / N = 1$, where $\sigma_{\text{w},i}^2$ are the variances of individual rows of the matrix, the spectral radius will still be unity. This observation was also reported and proven in (Rajan and Abbott, 2006).

From this observation, two question emerge: first, can we find local gain dynamics that can tune the global measure $\sum_i a_i^2 \sigma_{\text{w},i}^2 / N$? And second, given the assumption that this constraint is fulfilled, does the particular distribution of gains affect the network performance?

References

Rajan, K. and L. F. Abbott
   2006. Eigenvalue spectra of random matrices for neural networks. *Physical Review Letters*, 97(18).