New insights on homeostatic activity-dependent structural plasticity in rate based neural networks

Tanguy Cazalets[1], **Joni Dambre**[1]
1. Ugent University, Gent, Belgium

Summary. We introduce a novel rate-based homeostatic activity-dependent structural plasticity (HADSP) algorithm for echo state networks (ESNs). The algorithm employs solely homeostatic plasticity, yet enables the emergence of principles of Hebbian learning. Our analysis suggest that HADSP is able to generate networks that effectively recombine redundant inputs to improve performances and highlight the role of structural plasticity in influencing network function and organization.

Structural plasticity is a fundamental mechanism in the formation and maturation of biological neural circuits during development and has been the subject of significant scientific research [2] [1]. Although fundamental in biology, structural plasticity is totally absent in the context of ESN where connections between neurons are generated randomly [3] and remain fixed. Our algorithm aims to tackle this gap between ESN and biology.

Echo state networks. The ESN framework is a system of n neurons with vector states $\mathbf{x}$ determined by a combination of a connection matrix $\mathbf{W}$, input matrix $\mathbf{W_{in}}$, input $u[t]$, bias vector $\mathbf{b}$, and activation function σ. The evolution of our system is described by : $\mathbf{x}[t+1] = \sigma(\mathbf{W} \times \mathbf{x}[t] + \mathbf{W_{in}} \times u[t]) + \mathbf{b}$, where σ is the hyperbolic tangent function, $\mathbf{b}$ is a $1 \times n$ vector sampled from a normal distribution with mean 0.1 and standard deviation 0.1, and the dimensions of $\mathbf{W}$ and $\mathbf{W_{in}}$ are $n \times n$ and $1 \times n$, respectively. The desired output $y[t]$ is obtained from the learned matrices $\mathbf{W_{out}}$ and $\mathbf{b_{out}}$, which are obtained by ridge regression to approximate a target output $y^{target}[t]$, as given by : $y[t+1] = \mathbf{W_{out}} \times \mathbf{x}[t+1] + \mathbf{b_{out}}$. [3]

HADSP rule. Every Δt steps, a growth indicator, $\mathbf{\Delta z}$, is calculated for each neuron according to the following formula :

$$\mathbf{\Delta z} = -\frac{1}{\beta}(\langle \mathbf{x} \rangle - \rho) \quad (1)$$

where $\langle \mathbf{x} \rangle$ is the average of the state vector $\mathbf{x}$ over the increment period. For the j-th neuron if $\Delta z_j < -1$ one input connection is added ($w_{ij} = w_{ij} + 0.05$) or pruned ($w_{ij} = w_{ij} - 0.05$) if $\Delta z_j > +1$. Self-connections are not allowed in our model and the number of connections k is constrained by $k \leq \gamma$ with γ the maximum number of partners. We utilize only excitatory neurons to ensure that adding or removing inputs will have a positive or negative impact on the rate.

$\mathbf{\Delta z}$ depends on two hyperparameters : the growth parameter β and the target rate ρ that are choosen in order to obtain a given spectral radius while allowing the algorithm to converge. We indeed found that if the algorithm fails to converge, the network undergoes constant reorganization, leading to a changing dynamic over time and unstable representation of the inputs, thus making its training with Ridge regression inefficient. Similarly the value 0.05 of weight increment is kept low enough to allow the HADSP to recombine the inputs without creating unstable regimes.

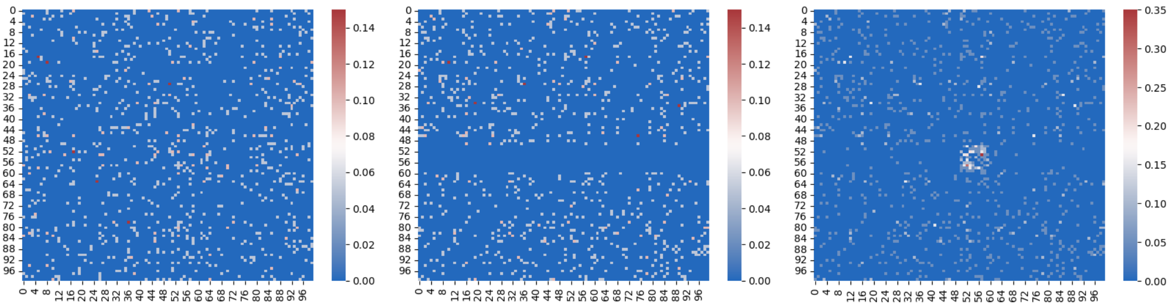
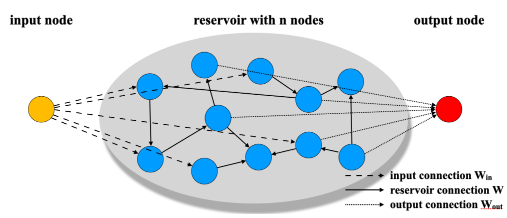
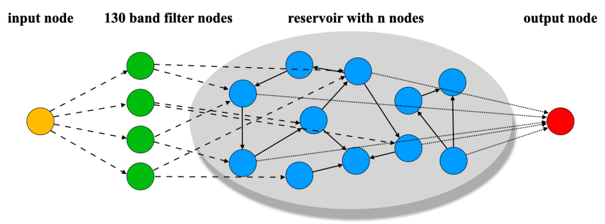


Figure 1: HADSP implements a form of Hebbian learning. Starting with a random connection matrix, the neurons create connections between them through a homeostasis mechanism. When they receive larger inputs, neurons 51 to 60 prune their weights to reach homeostasis. Once the larger input is removed, the other neurons are already at homeostasis, hence neurons 51 to 60 create connections mostly between themselves. Thus neurons developed intricate connections based on their previous correlated over-stimulation, demonstrating a form of Hebbian learning in our purely homeostatic algorithm.

Results. We postulate that HADSP algorithm can enhance the dynamics of a reservoir computing network by combining redundant inputs into more useful ones. To validate this idea we use the popular Mackey-Glass (MG) benchmark: $\frac{du}{dt} = \beta \frac{u(t-\tau)}{1+u^{10}(t-\tau)} - \gamma u(t)$ where $u(t)$ is the state of the system at time t, β and γ are positive constants, and $\tau = 17, 30$ to obtain mild and strong chaotic behavior. Performance are evaluated on T-steps ahead prediction using the Normalised Root Mean Square Error (NRMSE).



(a) Schematic overview of the ESN in the first experiment



(b) Schematic overview of the ESN in the second experiment

Model	T	HADSP	random
MG17	1	0.021	0.0075
	30	0.57	0.52
	100	0.73	0.67
MG30	1	0.0060	0.0043
	10	0.53	0.50
	30	0.41	0.40

(c) NRMSE averaged over 10 trials, for MG-17 and MG-30 for first experiment

Model	T	HADSP+band	random + band
MG17	1	0.0022	0.0040
	30	0.096	0.10
	100	0.44	0.47
MG30	1	0.035	0.038
	10	0.15	0.17
	30	0.24	0.24

(d) NRMSE averaged over 10 trials, for MG-17 and MG-30 for second experiment

Figure 2: Comparison of ESN and results for our two experiments

We aim to prove that the HADSP algorithm can exploit input correlation to improve performance in ESNs therefore we didn't systematically optimise the hyperparameters of our ESNs.
First experiment : The unsupervised learning process of the HADSP algorithm with $\rho = 0.8$ and $\beta = 0.1$ was used to evolve a network using a MG time series as input with a correlation captured every $\Delta t = 20$ steps. The output weight matrix W_{out} was calculated in a supervised manner to reproduce the signal T steps ahead. This resulting ESN is then compared against a network instantiated from a uniform distribution between 0 and 1 with same size, connectivity and spectral radius. The results in 2(c) showed no improvement in performance for the MG time series.
Second experiment : To address this, the input was transformed into 130 separate inputs using Butterworth band filters on a selected set of frequencies. A connection matrix was generated through the HADSP mechanism of size $n = 260$ with $\rho = 0.8$, $\beta = 0.1$, $\gamma = 12$ and $\Delta t = 20$. The output weight matrix W_{out} was calculated using the generated ESN and the 130 filtered inputs to reproduce the original MG signal τ steps ahead. The HADSP algorithm was again compared to a second matrix randomly generated with the same size, connectivity and spectral radius. The results in 2(d) showed that the HADSP algorithm outperformed the randomly instantiated network

Conclusions. Our study highlights the effectiveness of HADSP in synthesizing inputs that have similar dynamics, which in turn leads to more useful dynamics in the reservoir. The interplay between Hebbian learning and homeostasis, as showcased by HADSP, provides new insights into the generation of non-random ESN. Overall, our results emphasize the importance of considering input similarity in the design and optimization of reservoir computing systems.

References

[1] M. Butz, F. Wörgötter, and A. van Ooyen. Activity-dependent structural plasticity. *Brain Research Reviews*, 60(2):287–305, may 2009.

[2] J. Cang and D. A. Feldheim. Developmental mechanisms of topographic map formation and alignment. *Annual Review of Neuroscience*, 36(1):51–77, jul 2013.

[3] M. Lukoševičius. A practical guide to applying echo state networks. In *Lecture Notes in Computer Science*, pages 659–686. Springer Berlin Heidelberg, 2012.