Learning Methods for Dynamic Neural Networks

Emmanuel Daucé[†], Hedi Soula[‡] and Guillaume Beslon[‡]

†Movement and Perception, University of the Mediterranean, Marseille, France
‡Artificial Life and Behavior, PRISMa, Villeurbanne, France
Email: edauce@egim-mrs.fr, hedi.soula@insa-lyon.fr, guillaume.beslon@insa-lyon.fr

Abstract—In the framework of dynamic neural networks, learning refers to the slow process by which a neural network modifies its own structure under the influence of environmental pressure. Our simulations take place on large random recurrent neural networks (RRNNs). We present several results obtained with the use of a TD (temporal difference) and STDP (Spike-Time Dependent Plasticity) rule. First, we show that under some conditions, those learning rules give rise to an increase of the neurons synchronization, which can be interpreted as the crossing of a bifurcation line between non-synchronized and synchronized regimes. Second, we present various results obtained in control, under a reinforcement learning paradigm: inverted pendulum control and obstacle avoidance.

1. Introduction

In neural modeling, learning is a path in the space of control parameters possibly driving the system toward phase transition and bifurcations [14, 5]. It is also seen as a slow process which interacts with the fast activation process. Starting from the study of several families of random recurrent neural networks with complex intrinsic dynamics, we present here some of the generic properties displayed by temporal Hebbian learning processes in various experiments. In section 2, various models of networks and neurons are first presented. The section 3 illustrates the richness of their dynamics under standard parametric changes. The section 4 shows how temporal Hebbian learning rules interfere with the intrinsic dynamics and produce regime transitions. At last, we present in section 5 some of the applications of those learning techniques for skill acquisition in robotic devices.

2. Neuron models

This section presents three of the most classical discrete neurons models, namely firing rate models, binary models and integrate and fire models.

2.1. Firing rate model

The firing rate models (i.e. models with continuous activation) own the lowest time precision. The output $\bar{S}_i(t)$ of a neuron represents the firing rate within a certain time window Δt. For the seek of clarity, the firing rate is set to take place within interval $[0, 1]$ and the sample interval Δt is assimilated to a unitary delay. We suppose we have a population of N neurons, and $\bar{\mathbf{S}}$ is the vector of the N current firing rates. The vector of membrane potentials is

$$\mathbf{V}(t+1) = \mathbf{W}\bar{\mathbf{S}}(t) + \mathbf{u}(t) \tag{1}$$

where $\mathbf{u}(t)$ is the incoming signal and $\mathbf{W}$ is the interaction matrix. A typical activation function is $f(\mathbf{V}, \theta, g) = \frac{1+\tanh(g(\mathbf{V}-\theta))}{2}$ where θ is the threshold and g is the "gain" of the activation function. This gives the network update,

$$\bar{\mathbf{S}}(t) = f\left(\mathbf{V}(t), \theta, g\right)$$

2.2. Spiking models

In spiking models, the state of a neuron i is given by its membrane potential V_i. When V_i reaches the threshold θ, a spike is emitted (the threshold is supposed positive). In the most simple case (Mac Cullogh and Pitts model [9]), spike emission is stored in variable S_i whose value is 1 when a spike is emitted, and 0 elsewhere. The activation dynamics of neuron i is thus formally

$$S_i(t) = H(V_i(t) - \theta) \tag{2}$$

where H is the Heaviside function, which is equal to 1 when $V_i > \theta$, and 0 elsewhere.

In the more elaborate *leaky integrate and fire* (I&F) model [15], the neuron owns a memory of its past potential (leaky integrator), and is governed by the discrete equation :

$$V_i(t+1) = \gamma V_i(t) + \mathcal{I}_i(t) \tag{3}$$

The input $\mathcal{I}_i$ is composed of external input current and the contribution from all the neurons according to the interaction matrix ($\mathbf{W}$) whose charge function is a pulse (δ) of the firing times. The activation also depend on a refractory period which sets the maximum firing rate.

3. Dynamics and transitions

In large recurrent neural networks with random connectivity and continuous activation (see the corresponding contribution [3]), the gain of the transfer function can be used as control parameter. A generic route to chaos by quasi-periodicity can be observed under g quasi-stationary increase [6] (see figure 1). Every transition, from fixed point to cycle, T2 torus, frequency locking and chaos modifies the behavior of the system by steps, from order (fixed point) to strong disorder (deep chaos).
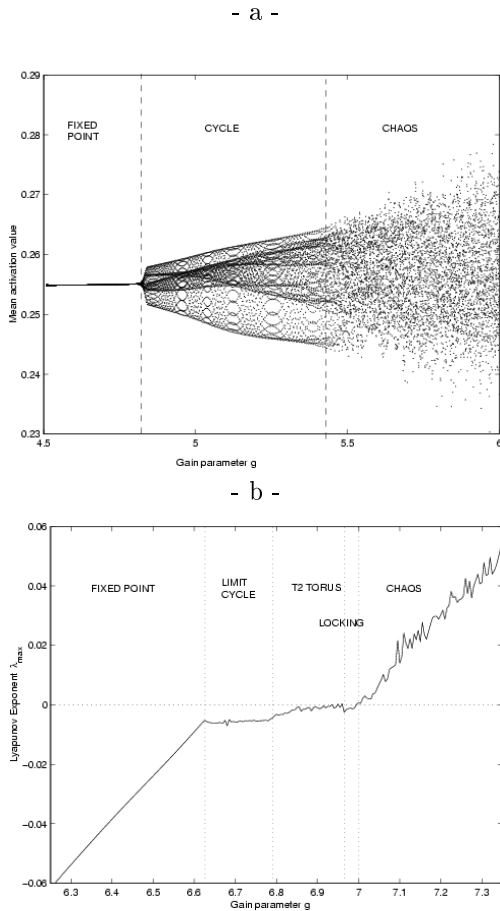
- a -

- b -

Figure 1: Examples of generic quasi periodicity route to chaos in continuous random network. Gain parameter g slowly increases from left to right. The figures - a - and - b - come from two distinct networks. - a - Mean activity - b - Biggest Lyapunov exponent.

Fewer results are available analytically for network of leaky Integrate and Fire neurons. Numerical simulations show that the standard deviation of a centered random recurrent network plays can play the role of a control parameter. Analogous to the gain in continuous model, its slow increase drives the network from a trivial state of neural death toward a chaotic state (see figure 2-a). The first bifurcation was proved an-alytically [12]. In case of refractory period different from the sampling rate, a further increase of the standard deviation leads toward "saturation", ultimately leading to a synchronous regime (see figure 2 -a). This diminution of chaos can be measured by a numerical estimation of the Kaplan-Yorke dimension of the attractor (see figure 2).
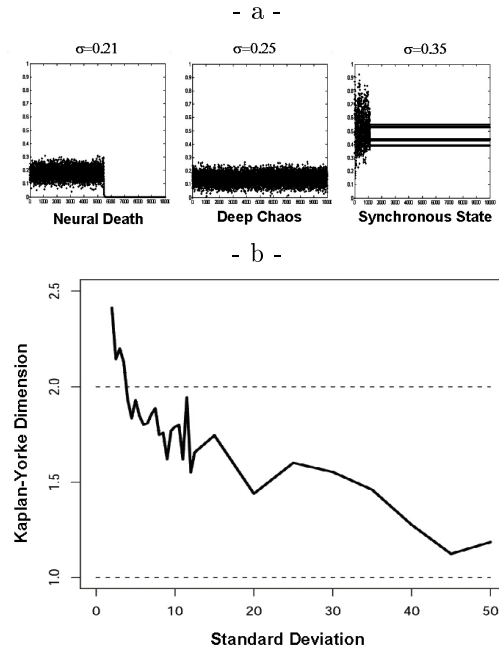
- a -

- b -

Figure 2: Examples of generic evolution diagram for I& F recurrent network. The standard deviation increases from left to right. - a - Mean potential (model with refractory period of 5 time steps) - b - Kaplan-Yorke Dimension of the attractor of the mean potential embedded in a space of dimension 3.

4. Learning rules

A good learning rule must rely on signals that are available in the vicinity of the neurons and be plausible, i.e. realizable at low cost at the level of the neurons. In the framework of Hebbian learning [8], the weight evolution relies on a product between pre-synaptic and post-synaptic activities of the form

$$\Delta W_{ij} = \alpha \sigma_i \sigma_j$$

where α is a learning parameter and σ_i and σ_j are respectively post-synaptic and pre-synaptic signals.

The concrete implementation depends on the neuron model and on the choice of the signals which are extracted from the neurons activity. Independently from the neuron model, a distinction can here be made between order 0 (static) and order 1 (temporal) rules. Order 0 rules only take into account the current neuron

activation. Such rules will reinforce the neurons whose activity is strong, and weaken the neurons whose activity is weak. On the contrary, first order Hebbian rules rules mostly take into account the difference (or derivative) of the neurons activities. In that case, the learning process will reinforce the order 1 characteristics of the neurons dynamics (their "dynamism").

We present here two temporal Hebbian learning rules. First, the family of temporal difference (TD) rules [13, 10] apply on binary or continuous models. A TD rule can be implemented as follows :

$$\Delta W_{ij}(t) = \alpha \left(S_i(t) - S_i(t-1) \right) S_j(t-1) \qquad (4)$$

so that the weight is reinforced when $S_i(t) > S_i(t-1)$, i.e. when the pre-synaptic signal arrives *before* a spike is emitted at time t. The weight decreases when $S_i(t) < S_i(t-1)$, i.e. when the pre-synaptic signal arrives at time t *after* a spike has been emitted at time $t-1$.

Spike-Timing Dependent Plasticity (STDP) [2, 1] apply on spiking models. The STDP rule relies on an asymmetric learning window based on the delay of firing times of the two neurons involved. It takes classicaly the form :

$$\frac{dW}{dt} = W \, \mathrm{F}(t^{\mathrm{post}} - t^{\mathrm{pre}}) \qquad (5)$$

using F as :

$$\mathrm{F}(t) = \begin{cases} \alpha_+ e^{-\frac{t}{\tau}} & t < 0 \\ -\alpha_- e^{-\frac{t}{\tau}} & t > 0 \\ 0 & t = 0 \end{cases} \qquad (6)$$

Some authors [10] have shown that STDP can be viewed as a TD rule under some assumptions.

Starting from a chaotic regime, the application of a Hebbian rule is found to drive the neural systems toward ordered cyclic and/or synchronized regimes. For instance, with order 0 rules, this regularizing behavior has been qualified as Hebbian driven "inverse route by quasi-periodicity" [5]. Here, using a TD rule in the continuous model, the dynamics converges toward a cycle with a high level of activity (and *do not* relax on a fixed point). In the spiking model, STDP also increases the coupling inside the network creating therefore high synchrony between neurons. Most importantly, in the two cases, the response of the system becomes stereotypic, i.e. any input pattern will lead the system to the same response.

To avoid this effect of saturation, one needs to introduce a competing learning rule that allows the network to remain in a "safe" zone. The basic idea is to reverse the sign of the correlation sign in the correlation rule : anti-Hebbian rule. Contrary to the Hebbian rules, the anti-Hebbian rules favor the transmissions

between neurons which are *not* correlated (and diminish the transmission between correlated neurons). The alternation between Hebb and anti-Hebb rules thus allows to control the degree of disorder. Figure 3 illustrates the mirror effect of the successive application of Hebb and anti-Hebb rule on the binary and the spiking model. The combined application of both rules allows the network to remain in a disordered state.
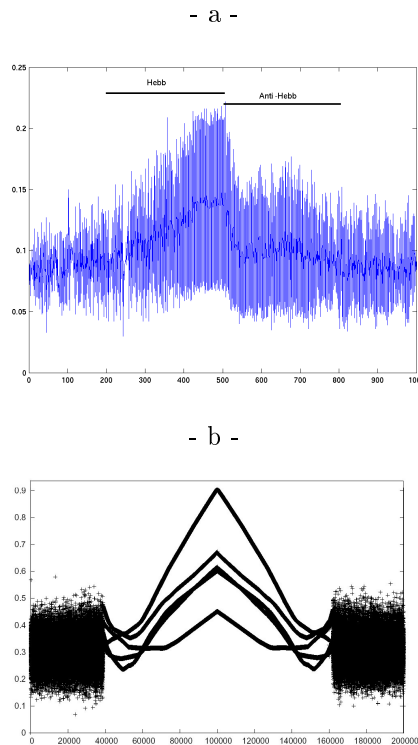
- a -

- b -

Figure 3: Successive application of a temporal Hebb and anti-Hebb rules. -a- application of TD and anti-TD in a binary network of 1000 neurons. This figure only shows the mean activity. -b- application of STDP and anti-STDP for the Integrate and Fire model.

5. Applications to Control

Under the classical TD learning approach [13], an "actor" process is responsible for the choice of a relevant action. The tuning of the action is thus under the control of the "critic" process, which owns the estimation of the value function. On the contrary to standard reinforcement methods, we present here alternative and biologically plausible reinforcement learning methods which discard explicit evaluation processes. The *selection* of a proper action out of several possible responses will rely on the versatility of the system.

- The action production and exploratory processes rely on the self-generated chaotic activity.

- Learning is the selection process through which the better configurations are to be stabilized.

- Learning is based on punctual applications of "positive" and "negative" Hebbian rules.

The learning process will be conducted by selectively assigning Hebbian or anti-Hebbian rule in order to maintain the network in a disordered, versatile and reactive state.

This approach has been first tested on a binary model, on the inverted pendulum benchmark [4]. The learning process relies on a selective application of a positive Hebbian learning rule on feedforward excitatory and lateral inhibitory links (positive reinforcement) and local inhibitory links (negative reinforcement). An efficient (but not optimal) control is found to take place after few learning steps.

A STDP/anti-STDP learning rule was applied to a recurrent spiking neural network that control a real Khepera robot [11]. The task to be learned was a collision-free movement with the sole use of the robot's visual input (linear camera) in an arena whose walls were composed of black and white stripes of random size (a similar environment that in [7]). STDP was applied when the robot moved forward above a certain speed and anti-STDP was applied when it hit the walls. After the learning process, the robots were able to avoid obstacles while moving.

6. Discussion

As argued throughout this article, competing learning rules for competing dynamics can be a powerful way to develop neural architecture that learns temporal tasks. As the system spontaneously displays a great variety of responses, the arising of a reward at a given time helps to favor a particular response out of a set of possible responses. The choice is however limited, and the behavior of the system is not necessarily the the optimal behavior, but only a *viable* one. The learning paradigm lies, in that case, more on a regulation process than on an optimizing tool. This fine regulation is obtained via dynamical synapses preventing them from saturating. However, this straightforward use of Hebbian principles may not be enough to extract more than simple sensori-motor skills. Orienting the learning toward more complex tasks implying active memory and delayed responses is still the prospect for future development and work.

References

[1] L. F. Abbott and S. B. Nelson. Synaptic plasticity: taming the beast. *Nature America*, 3:1178–1182, December 2000.

[2] G. Bi and M. Poo. Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *The Journal of Neuroscience*, 18(24):10464–10472, December 1998.

[3] B. Cessac, O. Mazet, M. Samuelides, and H. Soula. Mean field theory for random recurrent spiking neural networks. In *NOLTA Conference*, Bruges, 2005.

[4] E. Daucé. Hebbian reinforcement learning in a modular dynamic network. In *SAB 04*, 2004.

[5] E. Daucé, M. Quoy, B. Cessac, B. Doyon, and M. Samuelides. Self-organization and dynamics reduction in recurrent networks: stimulus presentation and learning. *Neural Networks*, 11:521–533, 1998.

[6] B. Doyon, B. Cessac, M. Quoy, and M. Samuelides. Control of the transition to chaos in neural networks with random connectivity. *Int. J. of Bif. and Chaos*, 3(2):279–291, 1993.

[7] D. Floreano and C. Mattiusi. Evolution of spiking neural controllers for autonomous vision-based robots. In T. Gomi, editor, *Evolutionnary Robotics*, Berlin, Germany, 2001. Springer-Verlag.

[8] D. Hebb. *The Organization of behavior*. Wiley, New York, 1949.

[9] W. S. Mac Cullogh and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.*, 5:115–133, 1943.

[10] R. P. N. Rao and T. J. Sejnowski. Spike-timing-dependent hebbian plasticity as temporal difference learning. *Neural Computation*, 13:2221–2237, 2001.

[11] H. Soula, A. Alwan, and G. Beslon. Learning at the edge of chaos : Temporal coupling of spiking neuron controller of autonomous robotic. In *AAAI Spring Symposia on Developmental Robotic*, page 6, Stanford, USA, 2005.

[12] H. Soula, G. Beslon, and O. Mazet. Spontaneous dynamics of assymmetric random recurrent spiking neural networks. *Neural Computation*, to appear, 2005.

[13] R.S. Sutton. Learning to predict by the method of temporal differences. *Machine learning*, 3:9–44, 1988.

[14] J. Tani. Model-based learning for mobile robot navigation from the dynamical system perspective. *IEEE Trans. System, Man and Cybernetics part B*, 26(3):421–436, 1996.

[15] H. C. Tuckwell. *Introduction to theoretical neurobiology: Vol.2:Non linear and stochastic theories*. Cambridge University Press, Cambridge, USA, 1988.