HEBBIAN LEARNING IN LARGE RECURRENT NEURAL NETWORKS

Emmanuel Daucé
E.C.M
Technopôle de Chateau Gombert
38, rue Joliot Curie
Marseille, France
email: edauce@egim-mrs.fr

Frédéric Henry
Movement and Perception
University of the Mediterranean
163, av.de Luminy
Marseille, France
email: frederic.henry@etumel.univ-mrs.fr

ABSTRACT
This paper presents the guidelines of an ongoing project of the "Movement Dynamics" team in the "Movement and perception" Lab, UMR6152, Marseille.

We address the question of Hebbian learning in large recurrent networks. The aim of this research is to present new functional models of learning, through the use of well known methods in a context of high non-linearity and intricate neuronal dynamics.

KEY WORDS
Hebbian Learning, Motor Control, Recurrent Neural Networks, Reinforcement Learning, Chaotic Dynamics

1 Introduction

The Hebbian learning framework [1] is one of the foundations of the computational neurosciences domain, since it lies at the crossroad of neurobiology and optimization/parametric adjustment methods. Despite its computational limitations[1], increasing evidence of its biological plausibility comes from observations [2, 3, 4]. In order to overcome this apparent paradox (the tremendous efficacy of real brains is not to be proven), Hebbian approaches of learning should thus be tested in more realistic frameworks. In real life indeed, Hebbian learning processes take place all lifetime long, without notable saturation effects. The problem is thus to define a framework where realistic and long time running Hebbian processes could allow artificial agents to increase their ability to handle their environment, without (or with limited) saturation effects.

We couple this question with another fundamental aspect of brain organization: its intrinsic recurrent organization. A large majority of its activity is indeed an internal self-sustained activity [2]. The stimuli are somewhat "far" from the actions, i.e. the current action as much depends on the history of the internal signals than on the current stimulus.

We present in the following several tracks that relate to the application of a Hebbian process in networks where the internal activity can dominate the sensory signal. The idea is of course to give evidence that a local Hebbian process can be associated with a global behaviour improvement. Under that approach, formal or rigorous results are still lacking. Our study of the properties of the learning processes will thus mainly rely on intensive simulation work. Simple setups imply an agent (body and surroundings) and an internal controller made of a random recurrent neural network. In such a context, as the parameter space is vaste, a careful methodology must be followed for the results to be properly established.

2 Simulated environments

In order to model realistic learning conditions, we define here several constraints that make the simulations close to learning situations that a small animal could, for instance, encounter in an experimental setup. Those conditions seem natural to follow, but most of them are broken in classical learning paradigms:

- closed loop interaction: the actions of the agent have direct consequences on its perceptions. The agent internal processes and its environment evolve in parallel and continuously influence each other.
- realistic environments: real-valued coordinates, continuous movements, realistic processing times between perception and action (no discrete mazes or block environments).
- locality: the learning rule only uses informations available in the vicinity of the neuron cell.
- continually running simulation: there is no possibility to re-play the trajectories. No resetting or reinitialization of the system (even if the system goes in a dead-end).
- on-line learning : the adaptation of the system parameters takes place immediately every time a new reinforcement signal occurs. There is no distinction between an exploration phase and an exploitation phase.

[1] Well-known for their ability to capture the structure of spatial pattern or spatio-temporal patterns in auto-associative networks, the Hebbian rules are also known to have their capacity to linearly depend on the size of the network, where saturation leads to "catastrophic forgetting".

[2] Some figures say for instance that 80% of the signal treated by the thalamus comes from the cortex itself.

3 Building networks

The controller is composed of a neural network, and some interfaces that translate the external states to tractable sensory signals, and the internal activity to motor commands. The neural network is made of several interacting neuron populations with random couplings. The distribution of the synaptic weights and thresholds follow several Gaussian or uniform distributions whose mean and variance are the global parameters of the network.

The networks are built on simple architectures with one or two layers, and inhomogeneous (non symmetric) connections. As the size of the networks are taken sufficiently large for the mean level of activity to be predictable, we can choose different levels of spatial and temporal resolution: (a) the low resolution systems use simple binary neurons with homogeneous delays; the temporal resolution is of the order of 10 ms; (b) the high resolution systems use integrate and fire (I&F) neurons with inhomogeneous delays; the temporal resolution is of the order of 1 ms.

Our setup aims to establish a functional link between the variability of the internal dynamics and some perception/action processes. Three interface modules must thus be carefully defined: (1) the sensory module: one must check that: (i) the sensory signals significantly reflects the environment state; (ii) the sensory signal is somewhat "orthogonalized" through various filters (edge detection, feature extraction...); (iii) the level of the input signal is conveniently balanced with the level of the internal self-sustained signal; (2) the reinforcement module: (i) the reinforcement signals must rely on sensory features that are directly perceptible. One can not use global "fitness function"; (ii) as we use continuously running simulations, there is no final reward, but a series of reinforcements that punctually occur during the experiment; (iii) the balance between positive and negative rewards must be checked, for instance by giving more credit to a positive reinforcement occuring in a context of negative performance; (3) the action module: (i) the motor command may be extracted from a mean activity (mean over several neurons activity, mean of one neuron over time...) in order to avoid hectic outputs; (ii) the motor activity is closely related to the activity of the recurrent layer. In some cases, the controller can thus produce spontaneous movements without being directly stimulated. The recurrent layer can be functionnaly seen as a CPG (Central Pattern Generator);

4 Hebbian traces

Depending on the model, various Hebbian rules can be tested from simple co-activation rules toward more elaborate STDP rules. Once again, the use of a reinforcement learning paradigm (also called reward learning or operant conditionning) aims to make the simulations as close as possible to real-world situations. The reinforcement learning paradigm is indeed considered as one of the most primitive adaptation mechanism, as the reward mechanism is easy to realize (through dopamine or other neurotransmitter release). It has been observed, for instance, in very simple invertebrates [5].

The experimental setup we have defined favours the use of direct reinforcement mechanisms (direct policy learning) [6], without any explicit or implicit model of the environment. Bartlett and Baxter, in [7], give a plausible interpretation of direct policy learning in terms of local synaptic adaptation. Their main innovation (to our knowledge) is their interpretation of the classical TD(λ) trace in terms of local *synaptic traces* which store the most recent co-activations between the pre-synaptic and the post-synaptic neurons. This trace doesn't take effect immediately on the synaptic value. When a reward occurs, the synaptic weight is modified according to the trace in a positive or negative way, depending on the reward sign.

We are about to generalize the use of synaptic traces in the case of random recurrent neural networks controllers. Some succesful applications of that principle are given in [8]. The question of the plausibility of such traces remains, as their existence is not proven at the present time.

5 Learning at the edge of chaos

One of the starting point of this study was the question of the role of chaos in neural processing [9]. On the contrary to simple feed-forward networks, we must indeed take into account the nature of the *dynamical regime* taking place inside the network during the learning process. This regime may vary from a mere fixed point toward a random-like chaotic activity. The way the neuronal activities self organize in a context of highly frustrated non-symmetrical couplings is not yet fully understood, even if some interpretations in terms of linear response can be given [10]. The study of the mean field of random recurrent neural networks helps to define parametric domains where chaotic regimes can be found [11, 12, 13], knowing the characteristics of the sensory signals.

The most unanimously admitted point is the reduction of the variability of the initial dynamics through Hebbian processes. This point has been observed throughout various models of neurons and various implementations of the Hebb's rule [9, 14]. Those results are moreover compatible with the hypotheses of the neurophysiologist W. Freeman [15], who first suggested that a link could be established between a reduction of the variability of the dynamics and perception processes.

Moreover, in a context of reinforcement learning, it can be shown that the application of a positive and a negative Hebb's rule have opposite effects on the dynamics: the Hebb's rule tends to lower the complexity while the anti-Hebb rule tends to increase the complexity [16].

However, in a real learning context, keeping the balance between order and chaos is difficult, as some saturation effects tend to take place in the long term. A possible way of preventing this, or at least to reduce the consequences, is to use a regulation principle, inspired by synap-

tic scaling, which modify the ratio between synaptic potentiation and depression in order to keep the postsynaptic neuron's frequency in the desired window. Such a regulation mechanism is given in [16].

6 A simple example

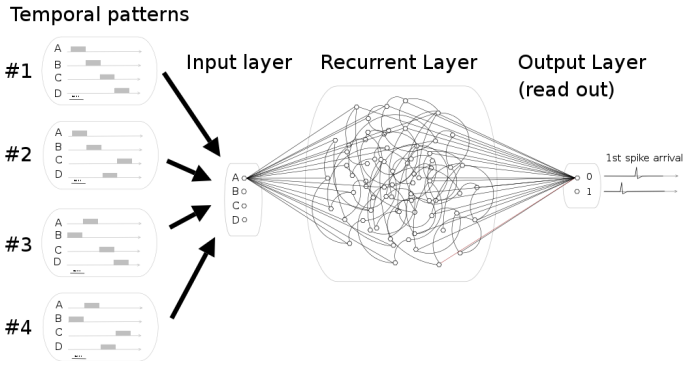
In order to give an intuitive view of the learning mechanisms we plan to implement, we present here a simple sequence classification experiment (see figure 1). The network is made of I&F neurons. The delays are inhomogeneous, and the mean transmission time is 10 ms. The temporal resolution is 0.5 ms. The network is composed of 3 layers : the first layer is composed of 4 neurons that receive a discontinuous input signal. The internal layer is composed of 200 neurons, whose interconnection pattern is a centered Normal law. The output layer is composed of 2 neurons. The output links are initially excitatory and homogeneous (while the delays are inhomogeneous), and the two output neurons are mutually inhibitive.

The input signal is composed of elementary steps that last 10 ms, whose amplitude is such that one spike is emitted for one step (see figure). The network is supposed to classify 4 periodic sequences in 2 categories, namely the sequences (A,B,C,D) and (B,A,D,C) belong to the category "0", and the sequences (A,B,D,C) and (B,A,C,D) belong to the category "1" (see figure 1a). This task has been named the "temporal XOR" [3]. Those sequences have been chosen for the response not to rely on one elementary step. At least one previous element must have been memorized for a correct response to be given. The network is forced to exploit its own internal activity to emit a proper answer.

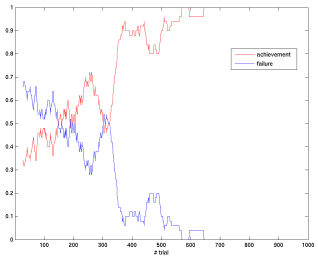
Starting from a quiet network, a given sequence is repeteadly presented until a first spike is emitted on the output neurons. If the answer is correct, a positive symmetrical STDP rule is applied on the internal and output links, with learning rate α_+. If the answer is wrong, the opposite rule is applied on the same links (anti-STDP), with learning rate α_-. The learning rates are adaptive and follow the right and wrong response statistics, namely $\alpha_+ \simeq 1-r$ and $\alpha_- \simeq r$, where r is the rate of correct answers according to the last 20 responses. The most "rare" events (success or failure) thus receive more credit. During a learning session, the sequences are chosen randomly (one for each trial) with equal probability.

Under this setup, the networks need about 500 trials to find the appropriate answers (see figure 1b). The time necessary to produce one answer is of the order of 60 ms (not shown), as this answer only relies on the internal activity (the network is initially quiescent).

[3] Indeed, if we reduce (A,B) and (C,D) to 0, and (B,A) and (D,C) to 1, then the four input sequences reduce to (0,0), (0,1), (1,0) and (1,1).



(a) Network



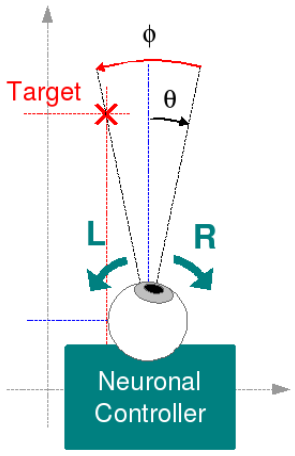
(b) Achievement rate

Figure 1. Temporal XOR task (see text).

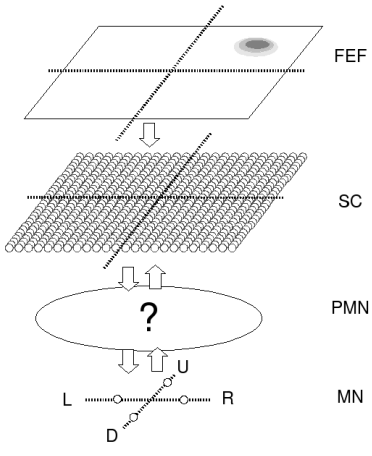
7 Done and to be done

A question underlying this presentation is the status of simulation work in computational neurosciences studies. The more intricate and realistic is the model, the less it can be tackled in a rigorous way. This remark, which holds even for simple Hebbian learning in random networks of continuous neurons, becomes drastically true in the case of STDP learning in networks of integrate and fire neurons.

The increasing computational power gives access to



(a) Environment model



(b) Neural Network

Figure 2. Saccade adaptation project. FEF: Frontal Eye Field; SC: Superior Colliculus (Neural Field model); PMN: Pre-Motor Neurons; MN: Motoneurons.

more realistic simulators, allowing to design models that take into account the full chain of learning, from the body/environment constraints to the characteristics of the particular neurons. A realistic model of the saccade adaptation is for instance currently under development (see figure 2). This "holistic" approach, even growing on weak theoretical ground, may however help to validate (or invalidate) some functional hypotheses, and thus, in the very end, help to understand the functions (or misfunctions) of the brain memory mechanisms. This approach is however delicate, as the success of a particular simulated mechanism does not give account for its biological plausibility. It may obviously be validated by real observations.

The several points which have been outlined here have already been partially realized. The origin of this study lies in the "Chaotic Neural Network" group of ONERA in Toulouse, composed of Bruno Cessac, Bernard Doyon, Mathias Quoy and Manuel Samuelides. This group studied in the years 1992-1995 both practically and theoretically the properties of large random neural networks under both dynamic and statistic approaches. The properties of Hebbian learning mechanisms have been studied later on, in the case of a generalization to multi-populations models by Emmanuel Daucé, Olivier Moynot, Bernard Doyon and Manuel Samuelides. The application of random recurrent neural networks to motor control has been tested in several ways. First in year 2000 on a real robotic platform in the ETIS team (Cergy Pontoise) [17]. In parallel, some advances have been obtained by the PRISMA team of INSA-Lyon with I&F neurons in simulations and on real robotic platforms. We owe them the "learning at the edge of chaos" motto [18]. The Hebbian trace approach of reinforcement learning is given in [8], and new developments are under redaction.

This study was supported from 2002 to 2006 by the DYNN ACI (Dynamical Neural Networks).

References

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

[2] T.V. Bliss and T. Lomo. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J. Physiol*, 232:331–356, 1973.

[3] H. Markram, J. Lubke, M. Frotscher, and B. Sakmann. Regulation of synaptic efficacy by coincidence of aps and epsps. *Science*, 275:213–215, 1997.

[4] Guo-Qiang Bi and Mu-Ming Poo. Synaptic modifications in cultured hippocampal neurons : Dependence on spike timing, synaptic strength, and postsynaptic cell type. *The Journal of Neuroscience*, 1998.

[5] Björn Brembs. Operant conditioning in invertebrates. *Current Opinion in Neurobiology*, 13:710–717, 2003.

[6] R. J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8:229–256, 1992.

[7] Peter L. Bartlett and Jonathan Baxter. Hebbian synaptic modifications in spiking neurons that learn. Technical report, Research School of Information Sciences and Engineering, Australian National University, 1999.

[8] E. Daucé. Hebbian reinforcement learning in a modular dynamic network. In *Proceedings of the Eighth International Conference on Simulation of Adaptive Behavior (SAB'04)*, pages 305–314, 2004.

[9] 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.

[10] B. Cessac and J.A. Sepulchre. Stable resonances and signal propagation in a chaotic network of coupled units. *Phys. Rev. E*, 70, 2004.

[11] H. Sompolinsky, A. Crisanti, and H.J. Sommers. Chaos in random neural networks. *Phys. Rev. Lett.*, 61:259–262, 1988.

[12] B. Cessac. Increase in complexity in random neural networks. *Journal de Physique I*, 5:409–432, 1995.

[13] E. Daucé, O. Moynot, O. Pinaud, and M. Samuelides. Mean-field theory and synchronization in random recurrent neural networks. *Neural Processing Letters*, 14:115–126, 2001.

[14] S. M. Bohte and M. C. Mozer. Reducing spike train variability : A computational theory of spike-timing dependent plasticity. In Lawrence K. Saul, Yair Weiss, , and Leon Bottou, editors, *Proceedings of Advances in Neural information processing (NIPS'05)*, pages 270–279. Cambridge, MA, MIT Press, 2004.

[15] C.A. Skarda and W.J. Freeman. How brains make chaos in order to make sense of the world. *Behav. Brain Sci.*, 10:161–195, 1987.

[16] F. Henry and E. Daucé. Temporal pattern identification using spike-timing dependent plasticity. In *Proceedings of the Computational NeuroSciences meeting (CNS'06), July 16th - 20th*, Edinburgh, U.K., accepted.

[17] E. Daucé and M. Quoy. Random recurrent neural networks for autonomous systems design. In *Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior (SAB'2000)*, pages 31–40. Meyer, J.-A., 2000.

[18] H. Soula, A. Alwan, and G. Beslon. Learning at the edge of chaos: temporal coupling of spiking neuron controller of autonomous robotics. In *AAAI Spring Symposium on Development Robotics*, Stanford, CA, 2005.