Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Neurocomputing ∎ (∎∎∎∎) ∎∎∎–∎∎∎
NEUROCOMPUTING
www.elsevier.com/locate/neucom

Improved stability and convergence with three factor learning

Bernd Porr[a,*], Tomas Kulvicius[b], Florentin Wörgötter[b,c]

[a]Department of Electronics & Electrical Engineering, University of Glasgow, Glasgow, GT12 8LT, UK
[b]Bernstein Center of Computational Neuroscience, University Göttingen, Germany
[c]Computational Neuroscience, Psychology, University of Stirling, FK9 4LR Stirling, UK

Abstract

Donald Hebb postulated that if neurons fire together they wire together. However, Hebbian learning is inherently unstable because synaptic weights will self-amplify themselves: the more a synapse drives a postsynaptic cell the more the synaptic weight will grow. We present a new biologically realistic way of showing how to stabilise synaptic weights by introducing a third factor which switches learning on or off so that self-amplification is minimised. The third factor can be identified by the activity of dopaminergic neurons in ventral tegmental area which leads to a new interpretation of the dopamine signal which goes beyond the classical prediction error hypothesis.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Three factor learning; Hebbian learning; Reinforcement learning; Robotics; Closed loop

1. Introduction

Hebbian learning [2] is the most prominent paradigm in correlation based learning. However, Hebbian learning is inherently unstable because of its *autocorrelation* term: Briefly, a changing weight will alter the output which will lead to further weight change, and so on. In this study we present a novel learning rule which is an extension of our differential Hebbian learning rule (isotropic-sequence-order or ISO-learning [4]) which minimises the destabilising autocorrelation term by switching learning on when the autocorrelation term is minimal and which is performed by a third factor which acts like a neuromodulator [1]. Therefore, we call this learning rule ISO3 learning. We will demonstrate the applicability of the rule with a simulated robot that learns to retrieve food disks.

2. Three factor learning

We are going to demonstrate using the open loop case how to minimise the destabilising autocorrelation term of Hebbian learning. Fig. 1A shows the basic components of the neural circuit. The learner consists of three inputs x_0, x_1 and r which are filtered by low pass filters: $u_0 = x_0 * h_0$, $u_1 = x_1 * h_1$ and $u_r = \Theta((r * h_r)')$ where Θ is a threshold > 0 as depicted in Fig. 1. The circuit can easily be extended to a bank of filters with different resonators $h_j, j > 0$ and individual weights $\rho_j, j > 0$ to generate complex shaped responses. The learning rule for the weight change ρ_j is: $\rho_j' = \mu u_r u_j v', j > 0$ where we have added a third factor u_r to the classical differential Hebbian learning [3,4].

The input signals x_0, x_1, r to our open loop circuit are delta pulses which trigger damped filter responses (see Fig. 1B). Weight change is driven by two factors: the cross-correlation between u_1 with the derivative u_0' and the autocorrelation of u_1 with its own derivative u_1'. However, auto- and the cross-correlations happen at different moments in time. Consequently we can switch on learning when the autocorrelation is minimal and the cross-correlation is maximal. This can be achieved by switching on the third factor u_r at the same time as the signal x_0 is triggered. Fig. 1C shows the behaviour of ISO3 learning as compared to ISO-learning for a relatively high learning rate. To test the effect of the autocorrelation we switched off the signal x_0 after step 4000. As shown in [4], ideally the weight should stabilise after x_0 has been switched off. Instead, one can see clearly that ISO-learning contains an exponential instability,

[*] Corresponding author. Tel.: +44 141 3305237.
E-mail addresses: b.porr@elec.gla.ac.uk (B. Porr), tomas@chaos.gwdg.de (T. Kulvicius), worgott@chaos.gwdg.de (F. Wörgötter).

0925-2312/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2006.10.137

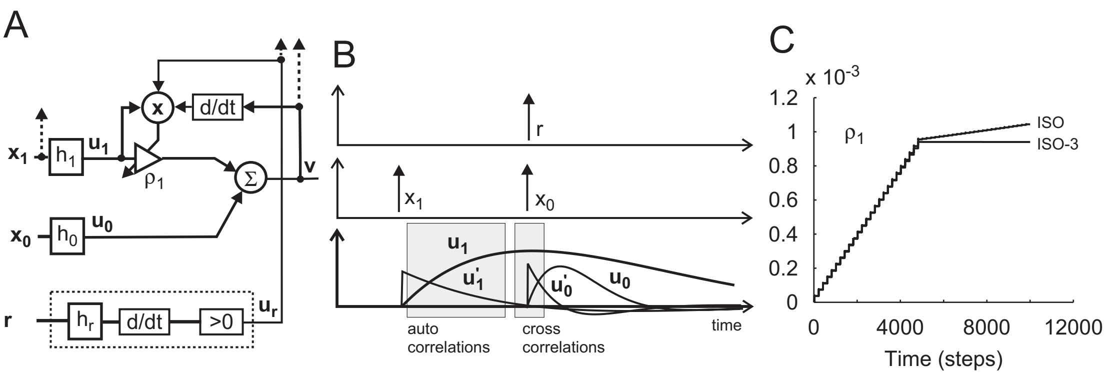

Fig. 1. (A) General form of the neural circuit. The inputs x_0, x_1, r are filtered by standard resonators (h_0, h_1, h_r which have frequency f and quality Q as parameters). u_0 and u_1 are summed at v with weights ρ_0 and ρ_1. The number of filters in the x_1 pathway can be extended to a filterbank with different resonators h_k and corresponding weights ρ_k which is indicated by the dotted lines. From the output of the filter h_r the derivative $\mathrm{d}/\mathrm{d}t$ is taken and then rectified (>0). The symbol $\otimes$ is a correlator and $\sum$ is a summation node. (B) Signals u_0, u_1 and their derivatives illustrate how learning works (see text for explanation). (C) Comparing ISO and ISO3 learning rules. System parameters: $f_{h_0,h_1,h_r} = 0.1$ and damping parameter $Q = 0.51$ were used to filter inputs x_0, x_1 and relevance signal r. Learning rate was $\mu = 0.005$ for ISO learning rule and $\mu = 0.07$ for ISO-3 rule. Time difference between x_1 and x_0 was $T = 10$ (x_1 always precedes x_0).

which leads to an upward bend. This is different for ISO3 learning which does not contain this instability. ISO3 learning is also stable when there is a bank of filters in the x_1 pathway and/or when the filter functions are not orthogonal to each other (data not shown).

In summary ISO3 learning uses the fact that auto- and cross-correlation happen at different moments in time. Consequently, we can stabilise differential Hebbian learning by switching learning on at the moment when the autocorrelation term is minimal.

3. Closed loop

The behavioural experiment of this section has two purposes: it will give the signals x_0, x_1 and r a behavioural meaning and it will demonstrate the superiority of ISO3 compared to ISO learning. Fig. 2A,B presents the task whereby a simulated robot has to learn to retrieve "food disks". The reflex x_0 is established by two light detectors (LD) which draw the robot into the centre of the white disks (Fig. 2A1). Learning uses the sound detectors (SD, Fig. 2A2) which feed into x_1 to generate an anticipatory reaction towards the "food disk" [7]. The reflex reaction is established by the *difference* of two light dependent resistors which causes a steering reaction towards the white disk (Fig. 2B). Hence x_0 is equal to zero if both LDs are not stimulated or when they are stimulated *at the same time* which happens during a straight encounter with a disk. The latter situation occurs after successful learning. The reflex has a constant weight ρ_0 which always guarantees a stable reaction. The predictive signal x_1 is generated by using two signals coming from the SD. The difference in the signals from the left and the right microphone is a measure of the azimuth of the sound source to the robot.

We quantify successful and unsuccessful learning for increasing learning rates μ. The learning rates have been chosen in a way that ensures in both cases that the contacts for successful learning are the same to make the failures comparable. Learning was considered successful when we received a sequence of five contacts with the disk at a subthreshold value of $| x_0 | < 1.1$ (which means that an alias of one pixel between robot and food disk is allowed). We recorded the actual number of contacts until this criterion was reached. The log–log plots of the number of contacts in Fig. 2C,D show that both rules follow a power law. The simulations demonstrate clearly that ISO3 learning is much more stable than the Hebbian ISO learning. ISO3 learning can therefore operate at more than 10 times higher learning rates than ISO learning. In addition this experiment also shows how to connect the learner from Fig. 1 with a behaving agent: the sensor signals feed into x_0 and x_1 and generate the steering angle v of the robot. While the sensor signals x_0, x_1 and v will change substantially during learning, the r-signal, however, is always triggered when the robot enters the food disk and stabilises learning by its correct timing but not by its amplitude which always remains the same.

4. Discussion

The third factor of our work can be related to the dopaminergic neurons in the ventral tegmental area (VTA) which respond strongly to primary rewards [5]. The VTA in turn is driven by the lateral hypothalamus (LH) which is the primary nucleus which becomes active while eating food. The VTA could have the task to switch on learning in a number of brain areas like the prefrontal cortex, the hippocampus, the nucleus accumbens and the striatum which could act as a global switch for learning. This means that the dopamine signal tells the target areas *when* to learn but not *what* to learn which is left to local processing in the target area. It is known that dopaminergic activity decreases at the primary reward and builds up at the location of the conditioned stimulus [5].

Please cite this article as: B. Porr, et al., Improved stability and convergence with three factor learning, Neurocomputing (2006), doi:10.1016/j.neucom.2006.10.137

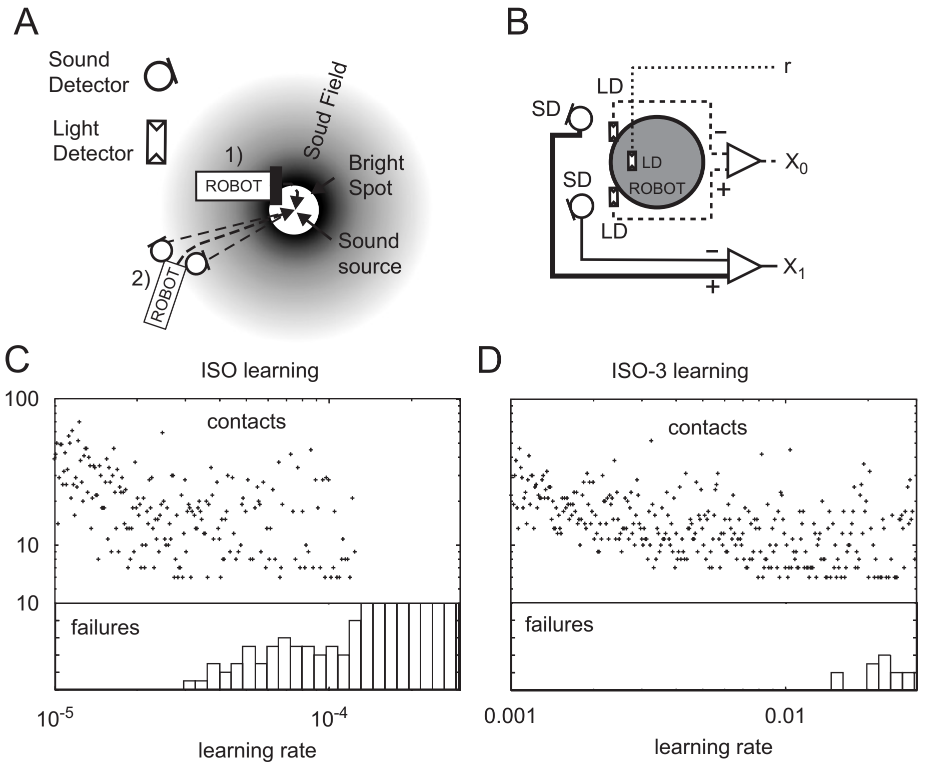

Fig. 2. The robot simulation. (A) The robot has two pairs of sensors: it has two light sensors which detect the food blob only in their direct proximity. In addition it has two sound detectors which are able to "hear" the food source from a distance. (B) The output v is the steering angle of the robot. The two light detectors (LD) establish the reflex reaction (x_0). The sound detectors (SD) establish the predictive loop (x_1). The weights $\rho_1 \dots \rho_N$ are variable and are changed either by ISO or ISO3 learning. The signal r is generated by a third light sensor and is triggered as soon as the robot enters the food blob. The robot also has a simple retraction mechanism when it collides with a wall ("retraction") which is not used for learning. The output v is the steering angle of the robot. Filters are set to $f_0 = 0.01$ for the reflex, $f_j = 0.1/j, j = 1 \dots 5$ for the filter bank where $Q = 0.51$. Reflex gain was $\rho_0 = 0.005$. (C) and (D) plot the number of contacts for both learning rules needed for successful learning against the learning rate. In addition the number of failures against the learning rate are plotted.

This behaviour can be re-interpreted: it helps to stabilise behaviour associated with the primary reward because learning is happening then at the moment of the secondary reward.

Reinforcement learning is usually implemented as an actor/critic architecture where the actor has the task of manoeuvring the agent to the reward while the critic is trying to predict the reward [6]. If the critic has been able to anticipate the reward the critic issues an error signal which in turn then modifies the actor which then eventually leads to goal directed behaviour towards the reward. In other words: the error signal actively decides which actions will be chosen. However, in ISO3 the signal u_r does not choose actions. ISO3 rather switches learning of an ISO-learner [4] on or off but does not force ISO learning towards a certain behaviour. Instead ISO learning decides by itself which behaviour will be learned.

References

[1] C.H. Bailey, M. Giustetto, Y.Y. Huang, R.D. Hawkins, E.R. Kandel, Is heterosynaptic modulation essential for stabilizing Hebbian plasticity and memory?, Nat. Rev. Neurosci. 1 (1) (2000) 11–20.

[2] D.O. Hebb, The Organization of Behavior: A Neurophychological Study, Wiley-Interscience, New York, 1949.

[3] B. Kosco, Differential hebbian learning, in: J. S. Denker (Ed.), Neural Networks for Computing: Snowbird, Utah, AIP Conference Proceedings, vol. 151, American Institute of Physics, New York, 1986, pp. 277–282.

[4] B. Porr, F. Wörgötter, Isotropic sequence order learning, Neural Comput. 15 (2003) 831–864.

[5] W. Schultz, Dopamine neurons and their role in reward mechanisms, Curr. Opin. Neurobiol. 7 (2) (1997) 191–197.

[6] R.S. Sutton, A.G. Barto, Reinforcement Learning: An Introduction, Second ed., Bradford Books, MIT Press, Cambridge, MA, 1998.

[7] P.F.M.J. Verschure, T. Voegtlin, R.J. Douglas, Environmentally mediated synergy between perception and behaviour in mobile robots, Nature 425 (2003) 620–624.

Bernd Porr has a diploma in physics (1997) and a master in journalism (1999) from the University of Bochum. After a short stay in Stockholm at the KTH, Bernd Porr took up a job as RA at Stirling University in 2000 where he also finished his Ph.D. in sequence learning and predictive control. From 1 May 2003 Bernd Porr took up a post as lecturer at the University of Glasgow at the department of Electronics and Electrical Engineering. Bernd Porr is pursuing research in the field of biologically inspired intelligent systems and robotics.

Please cite this article as: B. Porr, et al., Improved stability and convergence with three factor learning, Neurocomputing (2006), doi:10.1016/j.neucom.2006.10.137

Tomas Kulvicius was born in Kaunas, Lithuania in 1978. He received his M.S. degree in Computer Science from the Vytautas Magnus University, Kaunas, Lithuania. Currently he is doing his Ph.D. in Bernstein Center for Computational Neuroscience, University of Göttingen. His research interests include closed loop behavioral systems, receptive fields, learning, robotics and biosignal analysis.

Florentin Woergoetter has studied Biology and Mathematics at the University of Duesseldorf, Germany. He received his Ph.D. in Essen working on the neurophysiology of the visual cortex in 1988. He was Postdoc at the CALTECH in the lab of Christof Koch between 1988 and 1990, where he started modeling work. After short stays in Beijing and Stockholm he returned to Bochum working in computational and experimental neuroscience until 2000. Between 2000 and 2005 he was professor of Psychology at the University of Stirling in Scotland and since 2005 he is leading the Computational Neuroscience group at the Bernstein Center of the University of Goettingen. Florentin Woergoetter is pursuing research in the field visual perception, learning and plasticity.

Please cite this article as: B. Porr, et al., Improved stability and convergence with three factor learning, Neurocomputing (2006), doi:10.1016/j.neucom.2006.10.137