Neural Networks 34 (2012) 28–41

Contents lists available at SciVerse ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

From modulated Hebbian plasticity to simple behavior learning through noise and weight saturation

Andrea Soltoggio [a,*], Kenneth O. Stanley [b]

[a] Research Institute for Cognition and Robotics, University of Bielefeld, Germany

[b] Department of Electrical Engineering and Computer Science, University of Central Florida, USA

ARTICLE INFO

Article history:
Received 8 August 2011
Received in revised form 8 June 2012
Accepted 17 June 2012

Keywords:
Adaptive behavior
Computational models
Learning
Neural plasticity
Neuromodulation

ABSTRACT

Synaptic plasticity is a major mechanism for adaptation, learning, and memory. Yet current models struggle to link local synaptic changes to the acquisition of behaviors. The aim of this paper is to demonstrate a computational relationship between local Hebbian plasticity and behavior learning by exploiting two traditionally unwanted features: neural noise and synaptic weight saturation. A modulation signal is employed to arbitrate the sign of plasticity: when the modulation is positive, the synaptic weights saturate to express exploitative behavior; when it is negative, the weights converge to average values, and neural noise reconfigures the network's functionality. This process is demonstrated through simulating neural dynamics in the autonomous emergence of fearful and aggressive navigating behaviors and in the solution to reward-based problems. The neural model learns, memorizes, and modifies different behaviors that lead to positive modulation in a variety of settings. The algorithm establishes a simple relationship between local plasticity and behavior learning by demonstrating the utility of noise and weight saturation. Moreover, it provides a new tool to simulate adaptive behavior, and contributes to bridging the gap between synaptic changes and behavior in neural computation.

1. Introduction

This paper describes a novel, modulated Hebbian plasticity rule that makes productive use of features of Hebbian dynamics that in the past were thought undesirable. By utilizing noise and saturation, operant reward learning can emerge from the present learning rule, establishing an important link between local plasticity and macro-level behavioral adaptation.

The idea that adaptation, learning, and memory rely on synaptic change has gathered increasing consensus, beginning with the early studies of Hebb (1949) and the seminal work of Kandel and Tauc (1965). Early studies on the mollusk Aplysia proved that behavioral changes were precisely linked to the growth of particular pathways from sensory to motor systems (Carew, Walters, & Kandel, 1981; Kandel & Tauc, 1965). However, synaptic change follows rich dynamics that are often the product of different chemical signals (Clark, 2001) whose interaction and mechanisms are not completely understood. The Hebbian paradigm (Bi & Poo, 2001; Brown, Kairiss, & Keenan, 1990; Cooper, 2005; Gerstner & Kistler, 2002a; Hebb, 1949; Marr, 1969; Stent, 1973), which states that *neurons that fire together wire together*, is a ubiquitous paradigm in neuroscience that has been substantially validated through neural recordings (Kelso, Ganong, & Brown, 1986; Lisman, 1989; Markram, Lübke, Frotscher, & Sakmann, 1997; McNaughton, Barnes, Rao, & Rasmussen, 1986; Stent, 1973), corroborating detailed rate-based (Bienenstock, Cooper, & Munro, 1982; Gerstner & Kistler, 2002a; Grossberg, 1976; Oja, 1982; Rauschecker & Singer, 1981) and spiking neural (van Rossum, Bi, & Turrigiano, 2000) models.

The increasingly evident link between behavior learning and synaptic plasticity has encouraged researchers to propose numerous models whose overall behavior changes with the modification of synaptic weights; for reviews, see Bi and Poo (2001); Dayan and Abbott (2001); Gerstner and Kistler (2002a). However, one controversial and often unwanted feature of Hebbian models is that increasing firing leads to increasing synaptic strength, which in turn leads to further increasing of firing (Hasselmo, 1994; Miller & Mackay, 1994; Moldakarimov & Sejnowski, 2008). When a weight is larger than a certain threshold, a positive feedback will cause the weight to increase indefinitely. Such a model yields auto-correlation rather than cross-correlation of signals (Porr & Wörgötter, 2006). To prevent indefinite weight growth, various constraints can be imposed on the basic Hebbian plasticity (Bienenstock et al., 1982; Miller & Mackay, 1994; Oja, 1982). A second limitation of

* Corresponding author. Tel.: +49 5211 06 67062.
E-mail addresses: asoltogg@cor-lab.uni-bielefeld.de (A. Soltoggio), kstanley@eecs.ucf.edu (K.O. Stanley).

0893-6080/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2012.06.005

simple Hebbian plasticity is that learning can be just as fast as unlearning. For such models, short-lived stimuli leave a short-lived trace in the network regardless of their relevance. This feature contrasts with long-term potentiation (LTP), in which certain conditions induce synapses to maintain the increased strength in the long term (Brown, Chapman, Kairiss, & Keenan, 1988; Gustafsson, Wigstroem, Abraham, & Huang, 1987; Kelso et al., 1986; Levy & Steward, 1979, 1983).

In effect, the dynamics of Hebbian plasticity in biology are often affected and substantially altered by additional homeostatic dynamics (Turrigiano, 2008) and neuromodulators (Bailey, Giustetto, Huang, Hawkins, & Kandel, 2000; Clark, 2001; Giocomo & Hasselmo, 2007; Harris-Warrick & Marder, 1991; Hasselmo, 1995). For example, when the Aplysia encounters noxious stimuli, additional modulatory activity is also triggered, resulting in longer-lasting synaptic changes (Bailey et al., 2000; Clark & Kandel, 1984). This observation suggests that additional modulatory chemicals act as selectors of relevant stimuli that require learning of long-lasting responses, as in the case of dangerous or pain-inducing conditions (Bailey et al., 2000). To date there is extensive evidence linking conditioning behavior and reward learning with neuromodulation. Modulatory activity appears to carry reward information across a surprisingly large spectrum of animals, from insects like the honeybee (Gil, DeMarco, & Menzel, 2007; Hammer, 1993), to mollusks like the Aplysia (Brembs, Lorenzetti, Reyes, Baxter, & Byrne, 2002; Walters & Byrne, 1983) and to mammals (Berridge & Robinson, 1998; Schultz, Apicella, & Ljungberg, 1993; Schultz, Dayan, & Montague, 1997; Wise & Rompre, 1989). Yet whether and why neuromodulation is computationally essential to achieve such long-lasting behavioral responses has not been clarified.

Driven by biological findings, researchers have augmented their models with modulatory signals (Fellous & Linster, 1998; Hasselmo & Schnell, 1994; Ludvig, Sutton, & Kehoe, 2008) or attempted to model biological modulatory activities (Baxter, Canavier, Clark, & Byrne, 1999; Cohen, 2008). The precise role of various modulatory chemicals (e.g. serotonin, acetylcholine, dopamine, and norepinephrine (Bear, Connors, & Paradiso, 2005; Hasselmo, 2006)) is still debated, in particular regarding the role of dopamine in reward learning (Berridge & Robinson, 1998; Montague, Hyman, & Cohen, 2004; Pennartz, 1996, 1997; Redgrave, Gurney, & Reynolds, 2008; Schultz, 2006). Moreover, modulation appears to regulate a large variety of behaviors such as arousal, attention, reward learning, and memory (Aston-Jones & Cohen, 2005; Harris-Warrick & Marder, 1991; Hasselmo, 1995), resulting in an accordingly large spectrum of dynamics and models that regulate synaptic efficacy, synaptic changes, and other neural variables (Cox & Krichmar, 2009; Doya, 2002; Fellous & Linster, 1998; Hasselmo & Schnell, 1994; Krichmar, 2008; Smith, Husbands, Philippides, & O'Shea, 2002). One promising computational aspect of modulation is the possibility of increasing, decreasing, or inverting the strength and sign of plasticity (Abbott, 1990; Florian, 2007; Izhikevich, 2007; Montague, Dayan, & Sejnowski, 1996; Pfeiffer, Nessler, Douglas, & Maass, 2010; Porr & Wörgötter, 2007), making neuromodulation particularly suitable for modeling and implementing learning processes (Cox & Krichmar, 2009; Doya, 2002; Doya & Uchibe, 2005; Farries & Fairhall, 2007; Krichmar, 2008; Soula, Alwan, & Beslon, 2005; Sporns & Alexander, 2002). The focus in this study is on this latter role of modulation as a gating mechanism for Hebbian synaptic plasticity.

A fundamental issue is that a weight change that follows local rules does not always have a straightforward relationship with the system-level input–output mapping. This disconnect makes it difficult to apply local unsupervised plasticity rules to the fields of simulated adaptive behavior, artificial life (Langton, 1990; Sporns & Alexander, 2002), and robotics (Arkin, 1998). In these areas, the use of closed-loop controllers, in which the relationships between local and system-level dynamics are continuously tested, can provide the ultimate verification of the learning properties of a model. The model presented in this paper aims to establish a simple relationship between modulated Hebbian plasticity and operant reward learning, thereby connecting models of plasticity more closely to the learning of behaviors.

Instead of focusing on precise weight tuning, the unique position of this paper is to search for behavioral responses by allowing the weights to saturate, expressing either highly excitatory or inhibitory responses. By intentionally allowing weights to saturate, a network can express a marked and stable response to inputs, which can be interpreted as behavioral *exploitation*. On the other hand, by inverting this process at times, i.e. by inverting the sign of Hebbian plasticity (Lisman, 1989; Stent, 1973), pathways can be depressed to allow noisy neural transmission to implement behavioral *exploration*. The alternation of these two regimes of Hebbian and anti-Hebbian plasticity produces the key dynamics of alternating exploitation and exploration observed in operant reward learning. The change in modulatory activity has in fact been suggested to regulate the alternation of exploration and exploitation in Krichmar (2008). Thus, while the dynamics of modulated Hebbian plasticity and modulated spike-timing-dependent plasticity (STDP) have been extensively investigated (Abbott, 1990; Florian, 2007; Frémaux, Sprekeler, & Gerstner, 2010; Montague et al., 1996; Pfeiffer et al., 2010; Porr & Wörgötter, 2007), the novelty of this work is their extension by means of saturation and noise, resulting in a simpler and more fundamental connection between local changes and higher-level simulated behavior. The fundamental properties of the new plasticity model are tested in behavioral tasks employing first a single-neuron model, and later they are extended to multi-neuron networks.

As opposed to the algorithms proposed by Frémaux et al. (2010), Legenstein, Chase, Schwartz, and Maass (2010), and Pfeiffer et al. (2010), the present work neither devises a learning rule for optimal weight tuning nor proposes a new reinforcement learning algorithm. In fact, while reinforcement learning by means of modulated spike-timing-dependent plasticity (STDP) was demonstrated in Florian (2007), Frémaux et al. (2010) and Soula et al. (2005), the primary aim of this work is the exploitation of saturated weights and neural noise to achieve a simple bottom-up implementation of operant reward learning. Furthermore, in contrast to Pfeiffer et al. (2010), the current algorithm does not require a decay function, input signal preprocessing, nor winner-take-all action selection. Crucially, the neural noise in the present implementation is not used to *improve* exploration, as in Legenstein et al. (2010), but rather serves as the only and fundamental driving mechanism to reconfigure the network connectivity, thereby achieving behavioral exploration under the anti-Hebbian regime. Additionally, as opposed to Legenstein et al. (2010), where slow variation of input values and continuity in the task are required for recent activity averages, inputs and outputs in the proposed method can change state arbitrarily according to sudden changes of the task or environmental conditions. The insight that operant reward learning can emerge naturally and without additional engineering from Hebbian dynamics is a fundamental contribution of this study.

The plasticity mechanism, described in Section 2, is tested in several simulations reported in Section 3. Section 4 discusses the results, and Section 5 presents the conclusion. Appendix A shows that the plasticity rule behaves similarly on a simple spiking-neuron model. Further implementation details and how to reproduce the results with the Matlab code are reported in Appendices B–D.