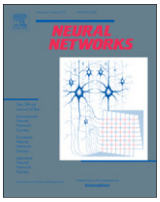
Neural Networks 105 (2018) 294–303

Contents lists available at ScienceDirect

Neural Networks

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



Representation learning using event-based STDP

Amirhossein Tavanaei [a,*], Timothée Masquelier [b], Anthony Maida [a]

[a] The School of Computing and Informatics, University of Louisiana at Lafayette, Lafayette, LA 70504, USA
[b] CERCO UMR 5549, CNRS-Université de Toulouse 3, F-31300, France

ARTICLE INFO

Article history:
Received 2 November 2017
Received in revised form 6 March 2018
Accepted 25 May 2018
Available online 1 June 2018

Keywords:
Representation learning
Spiking neural networks
Quantization
STDP
Bio-inspired model

ABSTRACT

Although representation learning methods developed within the framework of traditional neural networks are relatively mature, developing a spiking representation model remains a challenging problem. This paper proposes an event-based method to train a feedforward spiking neural network (SNN) layer for extracting visual features. The method introduces a novel spike-timing-dependent plasticity (STDP) learning rule and a threshold adjustment rule both derived from a vector quantization-like objective function subject to a sparsity constraint. The STDP rule is obtained by the gradient of a vector quantization criterion that is converted to spike-based, spatio-temporally local update rules in a spiking network of leaky, integrate-and-fire (LIF) neurons. Independence and sparsity of the model are achieved by the threshold adjustment rule and by a softmax function implementing inhibition in the representation layer consisting of WTA-thresholded spiking neurons. Together, these mechanisms implement a form of spike-based, competitive learning. Two sets of experiments are performed on the MNIST and natural image datasets. The results demonstrate a sparse spiking visual representation model with low reconstruction loss comparable with state-of-the-art visual coding approaches, yet our rule is local in both time and space, thus biologically plausible and hardware friendly.

1. Introduction

Unsupervised learning approaches using neural networks have frequently been used to extract features from visual inputs (Bhand, Mudur, Suresh, Saxe, & Ng, 2011; Lee, Ekanadham, & Ng, 2008). Single layer networks using distributed representations or autoencoder networks (Bengio, Courville, & Vincent, 2013; Coates, Ng, & Lee, 2011) have offered effective representation platforms. However, the robust, high level, and efficient representation that is obtained by networks in the brain is still not fully understood (Frégnac, Fournier, Gérard-Mercier, Monier, Pananceau, Carelli, & Troncoso, 2016; Landi & Freiwald, 2017; Logothetis & Sheinberg, 1996; Quiroga, Reddy, Kreiman, Koch, & Fried, 2005; Riesenhuber & Poggio, 2002; Wandell, 1995; Young & Yamane, 1992). Understanding the brain's functionality in representation learning can be accomplished by studying spike activity (Self et al., 2016) and bio-inspired spiking neural networks (SNNs) (Ghosh-Dastidar & Adeli, 2009; Izhikevich, 2004; Maass, 1997). SNNs provide a biologically plausible architecture, high computational power, and an efficient neural implementation (Maass, 1996, 2015; Neil, Pfeiffer, & Liu, 2016). The main challenge is to develop a spiking representation learning model that encodes input spike trains to uncorrelated, sparse, output spike trains using spatio-temporally local learning rules.

In this study, we seek to develop representation learning in a network of spiking neurons to address this challenge. Our contribution determines novel spatio-temporally local learning rules embedded in a single layer SNN to code independent features of visual stimuli received as spike trains. Synaptic weights in the proposed model are adjusted based on a novel spike-timing-dependent plasticity (STDP) rule which achieves spatio-temporal locality.

Nonlinear Hebbian learning has played a key role in the development of a unified unsupervised learning approach to represent receptive fields (Brito & Gerstner, 2016). Földiák (1990), influenced by Barlow (1989), was one of the early designers of sparse, weakly distributed representations having low redundancy. Földiák's model introduced a set of three learning rules (Hebbian, anti-Hebbian, and homeostatic) to work in concert to achieve these representations. Zylberberg, Murphy, and DeWeese (2011) showed that Földiák's plasticity rules, in a spiking platform, could be derived from the constraints of reconstructive accuracy, sparsity, and decorrelation. Furthermore, the acquired receptive fields of the representation cells in their model (named SAILnet) qualitatively matched those in primate visual cortex. The representation kernels determining the synaptic weight sets have been successfully utilized by our recent study (Tavanaei & Maida, 2017) for a spiking convolutional neural network to extract primary

---

* Corresponding author.

E-mail addresses: tavanaei@louisiana.edu (A. Tavanaei), timothee.masquelier@cnrs.fr (T. Masquelier), maida@louisiana.edu (A. Maida).

https://doi.org/10.1016/j.neunet.2018.05.018

visual features of the MNIST dataset. Additionally, the learning rules only used information which was locally available at the relevant synapse. Although SAILnet utilized spiking neurons in the representation layer and the plasticity rules were spatially local, the learning rules were not temporally local. The SAILnet plasticity rules use spike counts accumulated over the duration of a stimulus presentation interval. Since the SAILnet rules do not use spike times, the question of training the spiking representation network using a spatio-temporally local, spike-based approach like spike-timing-dependent plasticity (STDP) (Markram, Gerstner, & Sjöström, 2012), which needs neural spike times, remains unresolved. Later work, King, Zylberberg, and DeWeese (2013), extends (Zylberberg et al., 2011) to use both excitatory and inhibitory neurons (obeying Dale's law), but the learning rules still use temporal windows of varying duration to estimate spike rates, rather than the timing of spike events. Our work seeks to develop a learning rule which matches this performance but remains local in both time and space.

In another line of research based on cost functions, Bell and Sejnowski (1997) and Olshausen and Field (1996) showed that the constraints of reconstructive fidelity and sparseness, when applied to natural images, could account for many of the qualitative receptive field (RF) properties of primary visual cortex (area 17, V1). These works were agnostic about the possible learning mechanisms used in visual cortex to achieve these representations. Following Olshausen and Field (1996) and Rehn and Sommer (2007) developed the sparse-set coding (SSC) network which minimizes the number of active neurons instead of the average activity measure. Later, Olshausen, Cadieu, and Warland (2009) introduced an L_1-norm minimization criterion embedded in a highly overcomplete neural framework. Although these models offer great insight into what might be computed when receptive fields are acquired, they do not offer insight into details of the learning rules used to achieve these representations.

Early works that proposed a learning mechanism to explain the emergence of orientation selectivity in visual cortex are those of Bienenstock, Cooper, and Munro (1892) and von der Malsburg (1973). A state-of-the-art model is that of Masquelier (2012). This model blends strong biological detail with signal processing analysis and simulation to establish a proof-of-concept demonstration of the original (Hubel & Wiesel, 1962) feedforward model of orientation selectivity. A key feature of that model, relevant to the present paper, is the use of STDP to account for RF acquisition. STDP is the most popular learning rule in SNNs in which the synaptic weights are adapted according to the relative pre- and postsynaptic spike times (Caporale & Dan, 2008; Markram et al., 2012). Different variations of STDP have shown successful visual feature extraction in layer-wise training of SNNs (Kheradpisheh, Ganjtabesh, & Masquelier, 2016; Kheradpisheh, Ganjtabesh, Thorpe, & Masquelier, 2017; Masquelier & Thorpe, 2007; Tavanaei, Masquelier, & Maida, 2016). In a similar vein, Burbank (2015) has also proposed an STDP-based autoencoder. This autoencoder uses a mirrored pair of Hebbian and anti-Hebbian STDP rules. Its goal is to account for the emergence of symmetric, but physically separate, connections for encoding weights (W) and decoding weights (W^T).

Another component playing a key role in representing uncorrelated visual features in a bio-inspired SNN pertains to the inhibition circuits embedded within a layer. For instance, Savin, Joshi, and Triesch (2010) developed an independent component analysis (ICA) computation within an SNN using STDP and synaptic scaling in which independent neural activities in the representation layer were controlled by lateral inhibition. Lateral inhibition established a winner(s)-take-all (WTA) neural circuit to maintain the independence and sparsity of the neural representation layer. More recent work (Diehl & Cook, 2015) has combined a layer of unsupervised STDP with an explicit layer of non-learning inhibitory neurons. The inhibitory neurons impose a WTA discipline. Their representations were tested on the handwritten MNIST dataset and have been shown to be effective for recognition of such digits. The acquired representations tended to resemble MNIST prototypes, although their reconstructive properties were not directly studied. Shrestha, Ahmed, Wang, and Qiu (2017) also studied a spiking network with stochastic neurons that performs MNIST classification and acquires MNIST prototype representations. Their architecture is a 3-layer network where the hidden layer uses a soft WTA to implement inhibition. Since there is no functional need to introduce an explicit inhibitory layer if there is no learning, our work uses a softmax function (Bishop, 1995; Goodfellow, Bengio, & Courville, 2016) to achieve WTA inhibition. In our work, the standard softmax is adapted to a spiking network. Our acquired representations, when trained on the MNIST dataset, acquires representations resembling V1-like receptive fields, in contrast to the MNIST prototypes of the research described above.

Other works related to spike-based clustering and vector quantization are the evolving SNNs (eSNNs and deSNNs) of Kasabov, Dhoble, Nuntalid, and Indiveri (2013), Schliebs and Kasabov (2013), Soltic and Kasabov (2010), Wysoski, Benuskova, and Kasabov (2008) and Wysoski, Benuskova, and Kasabov (2010) which acquire representations via a recruitment learning paradigm (Grossberg, 2012) where neurons are recruited to participate in the representation of the new pattern (based on similarity or dissimilarity to preexisting representations). In the deSNN framework, if a new online pattern is sufficiently similar to an already represented pattern, the representations are merged to form a cluster. This later work uses a number of bio-plausible mechanisms, including spiking neurons, rank-order coding (Thorpe & Gautrais, 1998), a variant of STDP, and dynamic synapses (Maass & Markram, 2002).

The present research proposes event-based, STDP-type rules embedded in a single layer SNN for spatial feature coding. Specifically, this paper proposes a novel STDP-based representation learning method in the spirit of Burbank (2015), Masquelier (2012) and Zylberberg et al. (2011). Its learning rules are local in time and space and implement an approximation to clustering-based, vector quantization (Coates & Ng, 2012) using the SNN while controlling the sparseness and independence of visual codes. Local in time means that the information to modify the synapse is recent, say within at most a couple of membrane time constant of the postsynaptic spike that triggers the STDP. By local in space, we mean that the information used to modify the synaptic weight is, in principle, available at the presynaptic terminal and the postsynaptic cell membrane. Our derivation uses a continuous-time formulation and takes the limit as the length of the stimulus presentation interval tends to one time step. This leads to STDP-type learning rules, although they differ from the classic rules found in Caporale and Dan (2008) and Masquelier (2012). In this sense, the rules and resulting visual coding model are novel. Independence and sparsity are also maintained by an implicit inhibition and a new threshold adjustment rule implementing a WTA circuit.

2. Background

Földiák (1989) developed a feedforward network with anti-Hebbian interconnections for visual feature extraction. The Hebbian rule in his model, shown in Eq. (1), is inspired from Oja's learning rule (Oja, 1982) that extracts the largest principal component from an input sequence,

$$\Delta w_{ji} \propto (y_j x_i - w_{ji} y_j^2) \tag{1}$$

$$y_j = \sum_i x_i w_{ji} \tag{2}$$