Available online at www.sciencedirect.com
ScienceDirect
Neurocomputing 71 (2008) 3379–3386
NEUROCOMPUTING
www.elsevier.com/locate/neucom

Reinforcement learning of recurrent neural network for temporal coding

Daichi Kimura*, Yoshinori Hayakawa

Department of Physics, Tohoku University, Sendai 980-8578, Japan

Received 8 May 2007; received in revised form 28 September 2007; accepted 9 November 2007
Communicated by V. Jirsa
Available online 28 December 2007

Abstract

We study a reinforcement learning for temporal coding with neural network consisting of stochastic spiking neurons. In neural networks, information can be coded by characteristics of the timing of each neuronal firing, including the order of firing or the relative phase differences of firing. We derive the learning rule for this network and show that the network consisting of Hodgkin–Huxley neurons with the dynamical synaptic kinetics can learn the appropriate timing of each neuronal firing. We also investigate the system size dependence of learning efficiency.

Keywords: Temporal coding; Reinforcement learning; Hodgkin–Huxley neuron; Order coding; Phase coding

1. Introduction

Many studies have assumed that neurons transmit information by their firing rate. The McCulloch–Pits unit is a typical model and networks of these units have been investigated. On the other hand, recent experiments suggest that the timing of neuronal firing may also contribute to the information representation function in the brain and the synaptic modification [2,7,17,24]. For example, it seems that local and global synchronization play a significant role in integration of information which is distributed across the brain. Another example shows that the order of timing of neuronal firings can encode the information of stimuli on fingertips, and this encoding by sequence can transmit information faster than coding from the firing rate directly [12].

To capture the dynamical aspects of neural networks, networks consisting of various model neurons other than the McCulloch–Pitts unit have been investigated, because McCulloch–Pitts units cannot describe the temporal behavior of neurons over short time scales. In this context, an associative memory for neural networks of oscillator neurons or spiking neurons has been studied [1,8,11,13,15,16,28]. In these systems, the relative phase differences, i.e., the timing of firings, are used to represent the memory.

There are few studies of learning in pulse neuron models such as those consisting of Hodgkin–Huxley (HH) neurons because of difficulty in deriving the learning rule. Although several studies have been made of learning in networks that consist of integrate-and-fire (IF) model neurons [3,20,27], most of these studies focus only on coding in terms of the firing rate.

However, it would be useful to combine temporal coding and learning because it has been shown that temporal coding can deal with more information and process it faster than coding from just the firing rate [23]. As an example of this, Delorme et al. [4] show that a neural network consisting of IF neurons can learn to identify human faces by using "rank order coding", i.e., coding by the order of timing of each neuronal firing, where neurons are allowed to spike once only.

In this paper, we study a reinforcement learning for temporal coding with neural network consisting of stochastic spiking neurons. After defining a network of coupled stochastic HH neurons and some quantities in Section 2, we train the network to learn an XOR operation, where the output information is coded by the order of firing in Section 3. In Section 4, we investigate how the

*Corresponding author. Tel.: +81 22 795 6445; fax: +81 22 795 6447.
E-mail address: daichi@cmpt.phys.tohoku.ac.jp (D. Kimura).

0925-2312/$ - see front matter
doi:10.1016/j.neucom.2007.11.014

result or performance of learning depends on the system size and the strength of noise, and conclusions follow.

2. The model

To illustrate an example of the learning process of spiking neurons, we consider a neural network consisting of HH neurons. Since HH neurons show excitability, they can code information by the timing of firing. The complete dynamics for a network of coupled HH neurons may be expressed as

$$C_\mathrm{m}\frac{\mathrm{d}V_i}{\mathrm{d}t} = g_\mathrm{Na}m_i^3h_i(V_\mathrm{Na}-V_i)+g_\mathrm{K}n_i^4(V_\mathrm{K}-V_i) + g_\mathrm{L}(V_\mathrm{L}-V_i)+\sum_j w_{ij}I_j^\mathrm{s}(t)+\xi_i(t), \tag{1}$$

$$\frac{\mathrm{d}m_i}{\mathrm{d}t} = \alpha_m(V_i)(1-m_i)-\beta_m(V_i)m_i, \tag{2}$$

$$\frac{\mathrm{d}h_i}{\mathrm{d}t} = \alpha_h(V_i)(1-h_i)-\beta_h(V_i)h_i, \tag{3}$$

$$\frac{\mathrm{d}n_i}{\mathrm{d}t} = \alpha_n(V_i)(1-n_i)-\beta_n(V_i)n_i, \tag{4}$$

where V_i is the membrane potential of neuron i, C_m the membrane capacitance, V_r $(r=\mathrm{Na,K,L})$ are the equilibrium potentials, g_r $(r=\mathrm{Na,K,L})$ the conductance, m_i,h_i,n_i the voltage dependent activating or inactivating variables, α_x and β_x $(x=m,h,n)$ the functions of voltage V_i [10], w_{ij} is the synaptic weight from neuron j to i ($w_{ij}\neq w_{ji}$ in general), $I_j^\mathrm{s}(t)$ the synaptic current and $\xi_i(t)$ is the Gaussian white noise which obeys

$$\overline{\xi_i(t)}=0, \tag{5}$$

$$\overline{\xi_i(t)\xi_j(t')}=Q\delta_{ij}\delta(t-t'), \tag{6}$$

where $\overline{A}$ is the average of A over time and Q the variance of noise. The synaptic current $I_j^\mathrm{s}(t)$ is given by

$$I_j^\mathrm{s}(t)=r_j(t)[V_\mathrm{syn}-V_i], \tag{7}$$

where V_syn is the synaptic reversal potential (here, $V_\mathrm{syn}=0.0\,\mathrm{mV}$) and $r_j(t)$ the fraction of bound receptors [5] described by

$$\frac{\mathrm{d}r_j}{\mathrm{d}t}=\alpha T(t)(1-r_j)-\beta r_j, \tag{8}$$

$$T(t)=\begin{cases}1, & t_j^0\leqslant t<t_j^0+\tau,\\ 0 & \text{otherwise,}\end{cases} \tag{9}$$

where $\alpha=0.94\,\mathrm{m\,s^{-1}}$, $\beta=0.18\,\mathrm{m\,s^{-1}}$, t_j^0 the time when the presynaptic neuron j fires (membrane potential over 27 mV) and $\tau=1.5\,\mathrm{ms}$ [14]. Fig. 1 shows the behavior of $V_i(t)$ and $r_i(t)$ of single neuron added the external current whose amplitude is 10 mA. We used the fourth order Runge–Kutta method with the time step $\Delta t=0.01\,\mathrm{ms}$ to solve Eqs. (1)–(4).

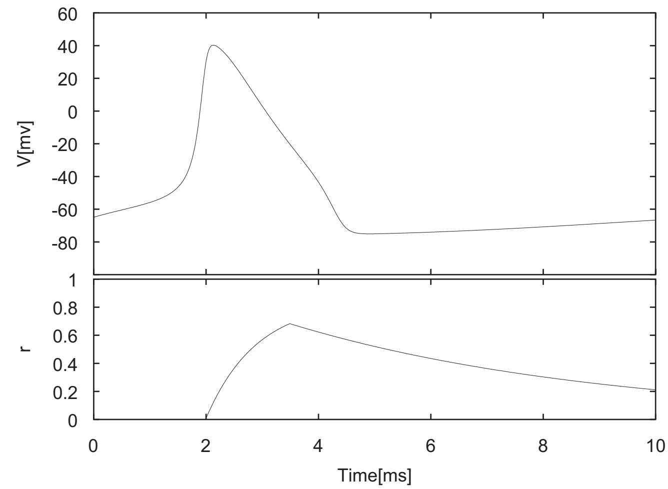

Fig. 1. The behavior of $V_i(t)$ and $r_j(t)$ of single HH neuron added the external current. The neuron fires at $t=2\,\mathrm{ms}$, then the value of $r_i(t)$ starts to increase. After a lapse of $\tau=1.5\,\mathrm{ms}$, $r_i(t)$ turns into decline.

To train the neural network, we use a reinforcement learning algorithm. Let us consider time sequences of states of neurons; $\sigma\equiv(\boldsymbol{V}(0),\boldsymbol{V}(1),\boldsymbol{V}(2),\ldots,\boldsymbol{V}(T))$, where $\boldsymbol{V}(t)$ denotes the vector $(V_1(t),\ldots,V_N(t))$. We assign a scalar value ("reward") to each time sequence σ according to the signal from the network [22]. We give a high reward R to the desirable time sequence σ in each episode. Here we consider episodic learning. Since Eq. (1) includes the Gaussian white noise, we calculate the expected value of the reward $\langle R\rangle$, where $\langle\cdots\rangle$ signifies the average over all possible time sequences σ. Then the goal of learning is to maximize $\langle R\rangle$ by adjusting w_{ij}. We use an ascending gradient strategy:

$$w_{ij}^\mathrm{New}=w_{ij}^\mathrm{Old}+\delta w_{ij}, \tag{10}$$

$$\delta w_{ij}=\varepsilon\frac{\partial\langle R\rangle}{\partial w_{ij}}, \tag{11}$$

where ε is the learning coefficient. We can calculate the gradient of $\langle R\rangle$ with respect to w_{ij} [6,9],

$$\frac{\partial\langle R\rangle}{\partial w_{ij}}=\frac{1}{Q}\left\langle R(\sigma)\int_0^T \mathrm{d}t\,\xi_i(t)I_j^\mathrm{s}(t)\right\rangle. \tag{12}$$

For details of the derivation, see appendix.

3. Learning procedure for temporal coding

The present learning rule Eqs. (10)–(12) can be effective for any information coding including the order coding. To show an example, we consider a neural network consisting of two input neurons, 15 output neurons and hidden neurons. We divide the set of output neurons into three disjoint subsets, $\mathcal{O}_1$, $\mathcal{O}_2$ and $\mathcal{O}_3$, each containing five output neurons.