ELSEVIER Neurocomputing 44–46 (2002) 97–101

NEUROCOMPUTING

www.elsevier.com/locate/neucom

Training neuron models with the Informax principle

Jianfeng Feng*

School of Cognitive and Computing Sciences, Sussex University, Brighton BN1 9QH, Sussex, UK

Abstract

In terms of the Informax principle, and the input–output relationship of the integrate-and-fire (IF) model, IF neuron learning rules are developed and applied to blind separation tasks. © 2002 Published by Elsevier Science B.V.

Keywords: The Informax principle; The integrate-and-fire model; LTP; LTD

1. Introduction

Neural computation is presumed to be a field to understand learning rules of realistic neuron systems (computational neuroscience) and then apply them to solving engineering problems (artificial neural networks). Nevertheless, majority, if not all, learning rules in artificial neural networks in the literature are basically based upon heuristic formulations of input–output relationship of a neuron: the sigmoidal function *depending only on the mean of its inputs*. Nowadays, we see that the gap between artificial neural networks and computational neuroscience are more wider than 10 years ago. On the one hand, people working on artificial neural networks are interested in developing methods which are directly applicable to solve engineering problems, somewhat a subdivision of traditional statistics. On the other hand, computational neuroscientists are working on more and more details of biophysical properties of a cell: from calcium to hundred different channels, to NO and hundreds or thousands compartments etc.

* Tel.: +44-1273-606755; fax: +44-1273-671320.

E-mail address: jianfeng@cogs.susx.ac.uk (J. Feng).

URL: http://www.cogs.susx.ac.uk/users/jianfeng

PII: S0925-2312(02)00368-5

In the present paper, we try to bring the yawning gap between two approaches into a unified framework: to develop a learning rule, which is applicable to solving engineering problems and is based upon (biophysical) models of a cell. The learning rule is derived under the principle of the maximization of the mutual information of input–output, which has been proposed and widely used in artificial neuron networks [1,11]. Due to the recent developments on modelling single neurons, we know exactly the input–output relationship of some neuron models such as the integrate-and-fire (IF) model [3,5,7,10,13] and IF-FHN model [6], etc. Combining these two approaches together, we are able to develop learning rules relied on input–output relationship of a neuron.

We first consider an ideal case where all synaptic strengths are identical. For supervised learning, by which we mean that the input and output firing rates of a neuron are fixed, we prove that there are stable states for the derived learning rule. Numerically, we also show that the stable state is unique. We then go further to consider unsupervised learning where the output firing rate is a function of inputs. We show that both long-term potentiation (LTP) and long-term depression (LTD) could occur, but LTD tends to be observable only when the inhibitory input is weak. This seems to be a reasonable conclusion: LTD is in a way equivalent to inhibitory inputs. For the situation where LTP and LTD are both observable, there is a critical point of the efferent firing rate. Above that it implies that when the efferent firing rate is faster than input firing rates, LTD occurs. Otherwise, LTP occurs. This is in general in agreement with the recent experimental data [2] which show that when the postsynaptic neuron fires faster than presynaptic neuron, LTD is observed, otherwise LTP happens. Very different from the previous applications of the Informax principle, where the anti-Hebbian learning rules are found, the derived learning rules are coincident with recent experimental data [12,14].

2. The IF model

Suppose that a cell receives excitatory postsynaptic potentials (EPSPs) at p synapses and inhibitory postsynaptic potentials (IPSPs) at q inhibitory synapses. When the membrane potential V_t is between the resting potential V_{rest} and the threshold V_{thre}, it is given by

$$\mathrm{d}V_t = -L(V_t - V_{\text{rest}})\,\mathrm{d}t + \mathrm{d}\bar{I}_{\text{syn}}(t), \tag{1}$$

where L is the decay rate and synaptic inputs

$$\bar{I}_{\text{syn}}(t) = \sum_{i=1}^{p} w_i^E E_i(t) - \sum_{j=1}^{q} w_j^I I_j(t)$$

with $E_i(t), I_i(t)$ as Poisson processes with rate λ_i^E and λ_i^I, respectively, $w_i^{\mathrm{E}} > 0, w_j^{\mathrm{I}} > 0$ being magnitude of each EPSP and IPSP. Once V_t crosses V_{thre} from below a spike is generated and V_t is reset to V_{rest}. This model is termed as the IF model [13,4].

Here we use the usual approximation to approximate the IF models, or more exactly the synaptic inputs of the models. We do not check the approximation accuracy since it has been done by many authors [13].

The input now reads $E_i(t) \sim \lambda_i^E t + \sqrt{\lambda_i^E} B_i^E(t)$ and similarly $I_i(t) \sim \lambda_i^I t + \sqrt{\lambda_i^I} B_i^I(t)$ where $B_i^E(t)$ and $B_i^I(t)$ are standard Brownian motions. Therefore, the IF model can be approximated by

$$\mathrm{d}v_t = -L(v_t - V_{\text{rest}})\,\mathrm{d}t + \mathrm{d}\bar{i}_{\text{syn}}(t),$$

where

$$\bar{i}_{\text{syn}}(t) = \sum_{i=1}^{p} w_i^E \lambda_i^E t - \sum_{j=1}^{q} w_j^I \lambda_j^I t + \sum_{i=1}^{p} \sqrt{(w_i^E)^2 \lambda_i^E} B_i^E(t) - \sum_{j=1}^{q} \sqrt{(w_i^I)^2 \lambda_i^I} B_j^I(t).$$

Since the summation of Brownian motions is again a Brownian motion we can rewrite the equation above as follows: $\bar{i}_{\text{syn}}(t) = \mu t + \sigma B(t)$ where $B(t)$ is a standard Brownian motion $\mu = \sum_{i=1}^{p} \lambda_i^E w_i^E - \sum_{j=1}^{q} \lambda_j^I w_j^I$, $\sigma^2 = \sum_{i=1}^{p} \lambda_i^E (w_i^E)^2 + \sum_{j=1}^{q} \lambda_j^I (w_j^I)^2$.

In the sequel we assume that $p = q, \lambda_j^I = r\lambda_j^E$ for $r \in [0,1]$. Therefore, when $r = 0$ the cell receives purely excitatory input and when $r = 1$ its inputs are exactly balanced. The interspike interval of efferent spikes is $T(r) = \inf\{t \colon V_t \geqslant V_{\text{thre}}\}$.

3. Uniform synapse case

We first assume that $w_i^E = w_i^I = w$, $i = 1, \ldots, p$, $\sum_i \lambda_i^E = \lambda$. The assumption of $w_i = w$ is certainly not true and only of theoretical interest. Nevertheless, under the circumstance, we are able to carry out a rigorous study on the learning dynamics and gain insights into the general case, discussed in the next sections:

For the IF model we have

$$\langle T(r)\rangle = \frac{2}{L} \int_{\frac{V_{\text{rest}}L - \lambda(1-r)w}{w\sqrt{L\lambda(1+r)}}}^{\frac{V_{\text{thre}}L - \lambda(1-r)w}{w\sqrt{L\lambda(1+r)}}} g(x)\,\mathrm{d}x, \tag{2}$$

where

$$g(x) = \left[\exp(x^2) \int_{-\infty}^{x} \exp(-u^2)\,\mathrm{d}u\right].$$

Therefore, the output firing rate (Hz) is

$$\gamma = s(\lambda) = \frac{1000}{T_{\text{ref}} + \langle T(r)\rangle}, \tag{3}$$

where T_{ref} is the refractory period.

Now, we apply the informax principle to the input–output relationship of the IF model. To this end, we have

$$\frac{\partial\langle T(r)\rangle}{\partial\lambda} = -\frac{2}{L}\, g\left(\frac{V_{\text{thre}}L - \lambda(1-r)w}{w\sqrt{\lambda L(1+r)}}\right) \frac{w(1-r)\lambda + V_{\text{thre}}L}{2w\lambda\sqrt{\lambda(1+r)L}}$$
$$+ \frac{2}{L}\, g\left(\frac{V_{\text{rest}}L - \lambda(1-r)w}{w\sqrt{\lambda L(1+r)}}\right) \frac{w(1-r)\lambda + V_{\text{rest}}L}{2w\lambda\sqrt{\lambda(1+r)L}}.$$

From now on we further assume that $V_{\rm rest} = 0$ mV, the equation above thus becomes

$$\frac{\partial\langle T(r)\rangle}{\partial\lambda} = -\frac{2}{L} g\left(\frac{V_{\rm thre}L - \lambda(1-r)w}{w\sqrt{\lambda L(1+r)}}\right)\frac{w(1-r)\lambda + V_{\rm thre}L}{2w\lambda\sqrt{\lambda(1+r)L}} + \frac{2}{L} g\left(\frac{-\lambda(1-r)}{\sqrt{\lambda L(1+r)}}\right)\frac{(1-r)}{2\sqrt{\lambda(1+r)L}}.$$

Note that the second term in the expression above is independent of w, learning independent.

Similarly, for $\partial\langle T(r)\rangle/\partial w$ we have

$$\frac{\partial\langle T(r)\rangle}{\partial w} = -\frac{2}{L} g\left(\frac{V_{\rm thre}L - \lambda(1-r)w}{w\sqrt{\lambda L(1+r)}}\right)\frac{\sqrt{L}V_{\rm thre}}{w^2\sqrt{\lambda(1+r)}}.$$

For simplicity of calculation we introduce two new variables

$$U = \frac{V_{\rm thre}L - \lambda(1-r)w}{w\sqrt{\lambda L(1+r)}}, \qquad V = \frac{V_{\rm thre}L + \lambda(1-r)w}{w\sqrt{\lambda L(1+r)}}$$

and note that from the relation $g'(x) = 2xg(x) + 1$ we obtain

$$\frac{\partial}{\partial w}\left(\frac{\partial\langle T(r)\rangle}{\partial\lambda}\right) = \frac{2}{L}[(2Ug(U)+1)V + g(U)]\frac{V_{\rm thre}\sqrt{L}}{2w^2\lambda\sqrt{\lambda(1+r)}}.$$

Hence,

$$l(w) = \frac{[(2Ug(U)+1)V + g(U)]\frac{V_{\rm thre}\sqrt{L}}{w^2\lambda\sqrt{\lambda(1+r)}}}{-g(U)V + g(\frac{-\lambda(1-r)}{\sqrt{\lambda L(1+r)}})\frac{(1-r)}{\sqrt{\lambda(1+r)L}}} + \frac{V_{\rm thre}}{250\sqrt{(1+r)L}}\,\frac{\gamma g(U))}{w^2\sqrt{\lambda}}.$$

Look at the second term in the learning rule above, we see that it gives us an anti-Hebb learning rule, as found out from the terms of the conventional input–output neuron relationship. Note that the anti-Hebb learning rule developed here takes the form of *output*/*input* rather than (*output*) (*input*). We will see the implication of the form later on.

We are particularly interested in two cases: $r=1$ (exactly balanced inputs) and $r=0$ (purely excitatory inputs). We refer the reader to [8,9] for details.

4. Discussion

We have presented a theoretical approach to derive novel learning rules based upon spiking neurons. In particular, for the IF model and both the supervised learning and unsupervised learning, we have proved that when the synapses are uniform, the learning rule obtained under the Informax principle is stable. Numerically, we have also demonstrated that the stable state is unique. For unsupervised learning, we conclude that both LTP and LTD are observable. Most interestingly, the derived learning rule quantitatively agrees with the recent experimental data. For the general case, supervised learning and unsupervised learning have been investigated and numerical results are included.

Acknowledgements

Partially supported by a grant from EPSRC (GR/R54569/01), a grant of the Royal Society and an exchange grant between UK and China of the Royal Society.

References

[1] A.J. Bell, T.J. Sejnowski, An information maximization approach to blind separation and blind deconvolution, Neural Comput. 7 (1995) 1129–1159.
[2] G.-Q. Bi, M.-M. Poo, Activity-induced synaptic modifications in hippocampal culture: dependence on spike timing, synaptic strength and cell type, J. Neurosci. 18 (1998) 10464–10472.
[3] D. Brown, J. Feng, S. Feerick, Variability of firing of Hodgkin-Huxley and FitzHugh-Nagumo neurons with stochastic synaptic input, Phys. Rev. Lett. 82 (1999) 4731–4734.
[4] J. Feng, Behaviours of spike output jitter in the integrate-and-fire model, Phys. Rev. Lett. 79 (1997) 4505–4508.
[5] J. Feng, D. Brown, Impact of temporal variation and the balance between excitation and inhibition on the output of the perfect integrate-and-fire model, Biol. Cybern. 78 (1998) 369–376.
[6] J. Feng, D. Brown, Integrate-and-fire models with nonlinear leakage, Bull. Math. Bio. 62 (2000a) 467–481.
[7] J. Feng, D. Brown, Impact of correlated inputs on the output of the integrate-and-fire model, Neural Comput. 12 (2000b) 671–692.
[8] J. Feng, H. Buxton, Training the IF model with the Informax principle I, J. Phys. A 35 (2002) 2379–2394.
[9] J. Feng, H. Buxton, Training the integrate-and-fine model with the Informax principle II, IEEE T. Neural Networks, submitted for publication.
[10] C. Koch, Biophysics of Computation, Oxford University Press, Oxford, 1999.
[11] R. Linsker, An application of the principle of maximum information preservation to linear systems, in: D.S. Touretzky (Ed.), Advances in Neural Information Processing Systems 1, Morgan-Kauffman, Los Altos, CA, 1989.
[12] S. Song, K.D. Miller, L.F. Abbott, Competitive Hebbian learning through spike-timing-dependent synaptic plasticity, Nature NeuroSci. 3 (2000) 919–926.
[13] H.C. Tuckwell, Introduction to Theoretical Neurobiology, Vol. 2, Cambridge University Press, Cambridge.
[14] L.I. Zhange, H.W. Tao, C.E. Holt, W.A. Harris, M.-M. Poo, A critical window for cooperation and competition among developing retinotectal synapses, Nature 395 (1998) 37–44.