Artif Life Robotics (2008) 12:223–226
DOI 10.1007/s10015-007-0471-7

ORIGINAL ARTICLE

Michael Gutmann · Kazuyuki Aihara

Toward data representation with spiking neurons

Received and accepted: July 13, 2007

Abstract Notable advances in the understanding of neural processing were made when sensory systems were investigated from the viewpoint of adaptation to the statistical structure of their input space. For this purpose, mathematical methods for data representation were used. Here, we point out that emphasis on the input structure has been at the cost of the biological plausibility of the corresponding neuron models which process the natural stimuli. The signal transformation of the data representation methods does not correspond well to the signal transformations happening at the single-cell level in neural systems. Hence, we now propose data representation by means of spiking neuron models. We formulate the data representation problem as an optimization problem and derive the fundamental quantities for an iterative learning scheme.

Key words Spiking neuron · Encoding · Decoding · Learning · Data representation

1 Introduction

Science is about exploring the structure and function of incompletely understood systems or phenomena. For the system "brain" or the phenomena "learning" and "vision", great advances have been made since the debates in the early 20th century about whether individual neurons are the basic elements of the nervous system or not (neuron doctrine). Since then, much emphasis has been on structure, i.e., on individual neurons or on how distinct classes of neurons are connected with each other. However, the functional aspects of these networks of neurons cannot be fully understood by its structure alone. How are the interconnected neurons marshaled to give rise to behavior? Why are the neurons as they are? Why are they connected the way they are?

These kinds of question were mostly addressed from the second half of the 20th century onwards. The brain was considered to be an information processing system. The principles and tools of signal processing and information theory were used to understand the function of some parts of the brain (e.g., the redundancy reduction hypothesis[1]). As information theory requires knowledge about the statistical structure of the information source, this approach triggered research into the properties of the sensory environment, especially with respect to vision,[2–4] and its links to neural processing.[5,6] For vision, in addition to the principles of information theory, other principles were used to explain its function in the form of "The early visual system might be optimized for . . . ," including energy expenditure[7] and minimal wiring among neurons.[8]

However, the data representation tools used which were from signal processing and information theory, were methods that were not developed with sensory neuroscience in mind. When these methods are used in a neuroscientific framework, the following assumptions were implicitly made:

— information is conveyed using a firing-rate code;
— neural processing is described by a linear filter.

Here, we might summarize them as linear rate-coding assumptions. The assumption are of course well justified as a first approximation to reality, especially when learning from natural stimuli is involved. However, we feel it is time to reconsider them. The neural system is nonlinear, and single spikes were found to be possible information carriers, at least in the early visual system of the fly.[9]

In this article, we propose data representation with a spiking neuron model. We formulate the problem in Sect. 2 as an optimization problem, and in Sect. 3.1, give a detailed

M. Gutmann (✉)
University of Tokyo, Room Ce605, Institute of Industrial Science, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan
Tel. +81-3-5452-6697; Fax +81-3-5452-6692
e-mail: gutmann@sat.t.u-tokyo.ac.jp

K. Aihara
Institute of Industrial Science, University of Tokyo, Aihara Complexity Modelling Project ERATO, JST, Tokyo, Japan

This work was presented in part at the 12th International Symposium on Artificial Life and Robotics, Oita, Japan, January 25–27, 2007

derivation of the key quantities for an iterative optimization rule. In Sect. 3.2 we give an interpretation, and a summary in Sect. 4 concludes the article. Here, we put emphasis on the mathematical derivation. A more detailed analysis of the resulting online rule and its properties, along with application examples, will be given elsewhere (in preparation).

2 Problem formulation

First, we specify the neuron model we are working with. Then, we put the aim "data representation with spiking neurons" into mathematical form. This is done by means of a cost functional, which needs to be minimized to accomplish data representation.

2.1 Neuron model

The assumed neural model is closely related to the SMR_0 model,[10] so that the equation for the membrane voltage u is

$$u(t) = \underbrace{\eta_0 \exp\left[-\frac{t-\hat{t}}{\tau}\right]}_{I_r(t)} + \underbrace{\int_0^{\min\{t,T_w\}} x(x-s)w(s)\mathrm{d}s}_{I(t)} + I_n(t) \tag{1}$$

where $I_n(t)$ is a noise current, $\hat{t}$ is the last spike timing before time t, and w is an unknown encoding filter of length T_w, to be learned for the minimization of the aforementioned cost functional. The convolution of input x with encoding filter w produces the input current I. Spike timings $\{t^f, f = 1, \ldots\}$ are defined by $u(t^f) = \theta$, where $\theta > 0$ is a fixed threshold. The remaining constants are the recovery time constant τ of the recovery current I_r and the reset amount $\eta_0 < 0$.

2.2 Reconstruction

From the obtained spike timings $\{t^f\}$, we aim at linearly reconstructing the stimulus x via

$$\hat{x}(t) = \sum_{f:t-T_p<t^f<t+T_d} h\left(t-t^f\right) \tag{2}$$

where h is the decoding filter, also to be learned for the minimization of the cost functional, and T_d, the estimation time delay. For the spikes generated prior to t, only those within a time-window of length T_p before t are considered for the reconstruction.

From Eq. 2 we see that the arguments for h are in the range $[-T_d\ T_p]$. For a good decoder, $h(-T_d) = h(T_p) = 0$ should hold. The role of $h(s)$ is different for $s > 0$ and $s < 0$. For $s > 0$, the input at t is predicted from a spike event at $t^f < t$. On the other hand, for $s < 0$, the input is reconstructed from a later spike event at $t^f > t$.

2.3 Cost functional

Both the encoding filter w and the decoding filter h are unknown. They are determined in order to minimize the cost functional

$$J(w,h) = \underbrace{\frac{1}{2T}\int_0^T (\hat{x}(t)-x(t))^2\mathrm{d}t}_{\text{Reconstruction error}} + \underbrace{\frac{\alpha}{2}\int_0^{T_w} w(t)^2\mathrm{d}t}_{\text{Energy cost}} \tag{3}$$

T is a fixed time horizon, and α weights the energy cost. This optimization problem implements the aim of data representation with spiking neurons in mathematical form. It is a quadratic in h, and hence a standard problem, but on the other hand it is not trivial in w.

3 The functional derivative $\delta J/\delta w$

Gradient-based methods are often used for optimization. Here, we obtain the expression for the functional derivative $\delta J/\delta w$ for the optimization with respect to w when the decoder h is held fixed.

3.1 Derivation

We start by perturbating $w(s)$ to $w(s) + \delta w(s)$, where

$$\delta w(s) = \varepsilon\varphi(s) \tag{4}$$

for a small constant $\varepsilon > 0$ and a sufficiently smooth, but otherwise arbitrary, function $\varphi(s)$. The perturbation δw causes a perturbation δt^f in the spike timings, which in turn causes a perturbation $\delta\hat{x}(t)$ of the reconstruction $\hat{x}$. The resulting perturbation δJ of the cost functional J is

$$\begin{aligned}\delta J = &\frac{1}{T}\int_0^T (\hat{x}(t)-x(t))\delta\hat{x}(t)\mathrm{d}t + \alpha\int_0^{T_w} w(t)\delta w(t)\mathrm{d}t \\ &+ \frac{1}{2T}\int_0^T (\delta\hat{x}(t))^2\mathrm{d}t + \frac{\alpha}{2}\int_0^{T_w} (\delta w(t))^2\mathrm{d}t\end{aligned} \tag{5}$$

The perturbation $\delta\hat{x}(t)$. We use the chain rule to obtain

$$\delta\hat{x}(t) = \sum_f \frac{\partial\hat{x}(t)}{\partial t^f}\delta t^f \tag{6}$$

For a fixed spike index f, Eq. 2 leads to

$$\frac{\partial\hat{x}(t)}{\partial t^f} = \begin{cases} -\dot{h}\left(t-t^f\right) & t^f - T_d < t < t^f + T_p \\ 0 & \text{else}\end{cases} \tag{7}$$

In order to evaluate Eq. 6 and thus Eq. 5, we must know the perturbation δt^f.

The perturbation δt^f. The spike timing $t^f > T_w$ is defined by $u(t^f) = \theta$, i.e.,

$$\theta = \eta_0 \exp\left[-\frac{t^f - t^{f-1}}{\tau}\right] + \int_0^{T_w} x\left(t^f - s\right)w(s)\mathrm{d}s + I_n\left(t^f\right) \tag{8}$$

After perturbation of $w(s)$, we make the following ansatz for δt^f:

$$\delta t^f = \varepsilon a_f + o\left(\varepsilon^2\right) \tag{9}$$

where a_f needs to be determined. The implicit equation for a_f is given by

$$\begin{aligned}\theta = \eta_0 \exp\left[-\frac{t^f - t^{f-1}}{\tau}\right]\exp\left[-\frac{\varepsilon(a_f - a_{f-1}) + o(\varepsilon^2)}{\tau}\right] \\ + \int_0^{T_w} x\left(t^f + \varepsilon a_f + o(\varepsilon^2) - s\right)(w(s) + \varepsilon\varphi(s))\mathrm{d}s \\ + I_n\left(t^f + \varepsilon a_f + o(\varepsilon^2)\right)\end{aligned} \tag{10}$$

The constant $\varepsilon > 0$ can be made arbitrarily small, so that via the Taylor series and Eq. 8 we obtain

$$\begin{aligned}0 = \varepsilon\Bigg\{a_f\Bigg(\frac{-\eta_0}{\tau}\exp\left[-\frac{t^f - t^{f-1}}{\tau}\right] + \dot{I}_n(t^f) \\ + \int_0^{T_w} \dot{x}(t^f - s)w(s)\mathrm{d}s\Bigg) + \frac{a_{f-1}\eta_0}{\tau}\exp\left[-\frac{t^f - t^{f-1}}{\tau}\right] \\ + \int_0^{T_w} x(t^f - s)\varphi(s)\mathrm{d}s\Bigg\} + o(\varepsilon^2)\end{aligned} \tag{11}$$

From Eq. 11, we see that a_f must satisfy

$$a_f = -\frac{\int_0^{T_w} x(t^f - s)\varphi(s)\mathrm{d}s}{\dot{u}(t^f)} + \gamma(t^f, t^{f-1})a_{f-1} \tag{12}$$

where

$$\gamma(t^f, t^{f-1}) = \frac{-\eta_0}{\tau\dot{u}(t^f)}\exp\left[-\frac{t^f - t^{f-1}}{\tau}\right] \tag{13}$$

$$\begin{aligned}\dot{u}(t^f) = \frac{-\eta_0}{\tau}\exp\left[-\frac{t^f - t^{f-1}}{\tau}\right] + \dot{I}_n(t^f) \\ + \int_0^{T_w} \dot{x}(t^f - s)w(s)\mathrm{d}s\end{aligned} \tag{14}$$

Finally, we see from Eq. 12 that a_f has the form

$$a_f = \int_0^{T_w} y_f(s)\varphi(s)\mathrm{d}s \tag{15}$$

so that we obtain the update rule for y_f

$$y_f(s) = \frac{-x(t^f - s)}{\dot{u}(t^f)} + \gamma(t^f, t^{f-1})y_{f-1}(s) \tag{16}$$

and δt^f is given by

$$\delta t^f = \varepsilon\int_0^{T_w} y_f(s)\varphi(s)\mathrm{d}s + o(\varepsilon^2) \tag{17}$$

Calculation of the functional derivative $\delta J/\delta w$. Equation 17 along with Eq. 7 allows us to evaluate Eq. 6, and hence to obtain δJ via Eq. 5.

Equation 6 for $\delta\hat{x}(t)$ introduces a sum over the spike-timings t^f into the integral

$$M = \int_0^T (\hat{x}(t) - x(t))\delta\hat{x}(t)\mathrm{d}t \tag{18}$$

of Eq. 5. Since there are only a finite number of spikes during the time-interval [0 T], the sum can only have a finite number of terms. Therefore we interchange the summation and the integration to obtain, with Eq. 7,

$$M = -\sum_f \underbrace{\int_{t^f - T_d}^{t^f + T_p} (\hat{x}(t) - x(t))\dot{h}(t - t^f)\mathrm{d}t}_{\bar{e}(t^f)} \cdot \int_0^{T_w} \varepsilon y_f(s)\varphi(s)\mathrm{d}s + o(\varepsilon^2) \tag{19}$$

The quadratic terms in Eq. 5 yield terms of the order of ε^2, so that with the previous equation for M we obtain

$$\delta J = \varepsilon\left(-\frac{1}{T}\sum_f \bar{e}(t^f)\int_0^{T_w} y_f(s)\varphi(s)\mathrm{d}s\right) + \alpha\int_0^{T_w} w(s)\varphi(s)\mathrm{d}s + o(\varepsilon^2) \tag{20}$$

Taking the principal linear part,[11] we obtain the final expression for the functional derivative of J with respect to w

$$\frac{\delta J}{\delta w(s)} = -\frac{1}{T}\sum_f \bar{e}(t^f)y_f(s) + \alpha w(s) \tag{21}$$

3.2 Interpretation

For the interpretation of Eq. 21, we first transform $\bar{e}(t^f)$ by a change of variables and partial integration, using $h(-T_d) = h(T_p) = 0$, into

$$\bar{e}(t^f) = -\int_{-T_d}^{T_p} \dot{e}(t^f + s)h(s)\mathrm{d}s \tag{22}$$

The rate of the error $e = \hat{x} - x$ is averaged with a weighting given by the reconstruction filter h. Equation 2 shows that the spike-timing t^f contributes via $h(s)$ to the reconstruction at $t^f + s$. Hence, the weighting is such that $\bar{e}(t^f)$ indicates the reconstruction error caused by spike-timing t^f on the time-interval $[t^f - \mathrm{Td}\ t^f + T_p]$.

Via $y_f(s)$ and Eq. 16, there is a recursion inherent in Eq. 21. We solve this recursion to allow for a better interpretation of the functional derivative $\delta J/\delta w(s)$. Grouping the terms with the common factor $x(t^f - s)$ together, we obtain

$$\frac{\delta J}{\delta w(s)} = -\frac{1}{T}\sum_f x(t^f - s)\frac{\tilde{e}(t^f)}{\dot{u}(t^f)} + \alpha w(s) \tag{23}$$

with

$$\tilde{e}(t^f) = \bar{e}(t^f) + \sum_{p \geq f+1} \gamma(t^{f+1}, t^f)\cdots\gamma(t^p, t^{p-1})\bar{e}(t^p) \tag{24}$$

The expression $\tilde{e}(t^f)$ can be interpreted as the total reconstruction error caused by the spike-timing t^f. Equation 23 then shows that the functional derivative $\delta J/\delta w(s)$ is the difference between αw and the correlation between the total reconstruction error caused by spike-timing t^f and the normalized input $x(t^f - s)/\dot{u}(t^f)$ at time s before the spike.

4 Summary

In the ongoing search to understand early sensory systems, notable advances have been made through data representation methods applied to natural stimuli. We pointed out, however, that the corresponding neuron models which process the natural stimuli tend to be abstract, and might be a limiting factor for further advances. This is especially so when the aim is to connect to experimental results at the single-cell level. Here, we made a small step toward biologically more plausible models, and considered data representation by means of a spiking neuron. Requiring linear reconstructability of the input from the spike train, we derived in detail the essential quantities for an iterative learning rule, and discussed their meaning.

Acknowledgments This research was partially supported by Grant-in-Aid for Scientific Research on Priority Areas – System study on higher-order brain functions – from the Ministry of Education, Culture, Sports, Science and Technology of Japan (17022012). MG was further supported by MEXT grant No. 040680.

References

1. Barlow HB (1961) Possible principles underlying the transformations of sensory messages. In: Rosenblith W (ed) Sensory communication. MIT Press, Cambridge, pp 217–234
2. Dong D, Atick J (1995) Statistics of natural time-varying images. Network Comp Neural Syst 6:345–358
3. Frazor RA, Geisler WS (2006) Local luminance and contrast in natural images. Vision Res 46:1585–1598
4. Ruderman D (1994) The statistics of natural images. Network Comp Neural Syst 5:517–548
5. Hyvärinen A, Hoyer PO, Hurri J, et al. (2005) Statistical models of images and early vision. In: Honkela T, Könönen V, Pöllä M, Simula O (eds) Proceedings of AKRR 05, International and Interdisciplinary Conference on Adaptive Knowledge Representation and Reasoning, Espoo, Finland, Jun. 2005, pp 1–14
6. Simoncelli EP, Olshausen BA (2001) Natural image statistics and neural representation. Anna Rev Neurosci 24:1193–1216
7. Vincent B, Baddeley R, Troscianko T, et al. (2005) Is the early visual system optimised to be energy-efficient? Network Comp Neural Syst 16:175–190
8. Graham DJ, Chandler DM, Field DJ (2006) Can the theory of "whitening" explain the center-surround properties of retinal ganglion cell receptive fields? Vision Res 46:2901–2913
9. Rieke F, Warland D, van Steveninck R de R, et al. (1997) Spikes: exploring the neural code. MIT Press, Cambridge
10. Gerstner W, Kistler WK (2002) Spiking neuron models. Cambridge University Press, Cambridge
11. Gelfand IM, Fomin SV (2000) Calculus of variations. Dover, New York