



1° Congresso Brasileiro de Redes Neurais

Escola Federal de Engenharia de Itajubá
Itajubá. 24 a 27 de outubro de 1994

Formulation of Reference Field in Neural Networks with External Input :Mechanism for Sensory Feature Integration

Kenichiro Mogi

Laboratory for Neural Networks
The Institute of Physical and Chemical
Research (RIKEN)
Hirosawa 2-1, Wako-shi, Saitama, 351-01
Japan
tel +81 484-62-1111 ext. 6444
FAX +81 484-62-4697
e-mail: kenmogi@murasaki.riken.go.jp

Abstract

I apply the graphic transformation method (K. Mogi, *Phys. Rev. E* 49, 4616 (1994)) to a neural network which receive input from external neurons. I show that the state distribution of the neural network receiving input from the external neurons can be expressed as a product of the intrinsic weight and the weight of the "reference field", which is formed by the interaction between the external and "internal" neurons. The visual system can be described as a direct product of the reference field and the feature space.

I. Introduction

The problem of the integration of the visual information is one of the most important unresolved problems in

neuroscience today.

The human visual system is a highly developed and complex system^[1]. The light signal falling on the retina is transmitted, via the retinal ganglion cells, to the LGN (lateral geniculate nucleus) of the thalamus. The visual information is then transmitted to the primary visual cortex, where various aspects of visual information is processed in different visual areas.

In the literature, three different processes of visual information integration is currently discussed^[1]. The first process involves enlarging the receptive fields of cells so that they are able to respond to, and collect information from, larger parts of the field of view. The second process occurs simultaneously with the first process and generates cells with more complex and specific properties. The third process involves the unification of signals from different sources, representing different visual attributes.

There are many theories and models aimed at explaining the mechanism of visual information integration [2]-[5].

It is one of the remarkable properties of the visual system that the third unification process occurs in such a way that the various features represented in the various "higher" visual areas are organized in the two-dimensional topographic map of the outside world. The retinotopic map of the outside world provides the frame of reference for the various visual features. In view of this aspect of visual information integration, we refer to the retinotopic map as the *reference field*. To elucidate the mechanism in which the various visual features are topographically organized

in reference to the reference field is one of the most interesting questions that can be asked about the visual system.

As the reference field for visual information integration is basically a two-dimensional retinotopic map, it is reasonable to assume that the locus where the reference field is represented in the visual area is the LGN (lateral geniculate nucleus) or the V1. Then, the question naturally arises, what is so special about these visual areas with retinotopic representation, the LGN and V1, that they should play the role of the reference field?

A special feature of LGN and V1 is that all visual information must pass through these regions, with an exception of the superior colliculus-pulvinar pathway. Another special feature of LGN and V1 is that these areas are densely connected with the reticular nucleus of thalamus, a area generally assumed to be involved in the modulation of visual attention [6]. These special features of the LGN and V1 are expected to be closely related to the fact that the locus of the reference field is the LGN or the V1.

Up to today, there has been no clear theoretical reasoning to backup the idea that the LGN and V1 play the role of retinotopic reference field in the visual information integration process. In this paper, we take note of the fact that the retinal ganglion cells, although a part of the visual system themselves, have a special role in the analysis of visual information. Namely, the retinal ganglion cells are special because

- (1) they receive light signals from the outside world and convert them into neural firings
- (2) they do not receive any back projection from the visual areas, so their activity does not depend on the internal connections and dynamics of the visual area; their activity is solely determined by the light signals falling on the retina.

It will be shown later that these properties give the retinal ganglion cells a special role in the integration of visual information.

II. Direct Product of Reference Field and Feature Space

Let us analyze what kind of mathematical structure we must have for describing the nature of visual information integration within the framework of the reference field.

Let us represent a point in the reference field as

$$\mathbf{R} = (x, y) \quad (1)$$

Here, the variables (x, y) are any system of coordinates representing the visual field.

On the other hand, let us express the various feature categories represented in the visual cortex as

$$\mathbf{F} = (f_1, f_2, \dots, f_p) \quad (2)$$

Here, it is implied that the suffices represent the categories of features represented in the visual system, such as color, shape, border, texture, etc. A particular value of the variables f_1, f_2, \dots, f_p would then describe a particular feature of these categories.

Then it is clear that representing the various features \mathbf{F} within the frame work of reference field \mathbf{R} means that we have a direct product of these fields $\mathbf{R} \times \mathbf{F}$, namely

$$\mathbf{R} \times \mathbf{F} = (x, y) \times (f_1, f_2, \dots, f_p) = (x, y, f_1, f_2, \dots, f_p)$$

It is to be noted that the weight distribution of the product space $\mathbf{R} \times \mathbf{F}$ can be formally written as the product of the weights for the space \mathbf{R} and \mathbf{F} as

$$W(\mathbf{R} \times \mathbf{F}) = W(\mathbf{R})W(\mathbf{F}) \quad (3)$$

Here, it should be noted that as the features in the feature space \mathbf{F} are generated as

a function of the reference field R , the weight expression for these areas are not completely independent. This fact will be expressed mathematically that there are some common elements in the expressions of F and R .

The question arises how we can establish a direct product of the kind depicted above in a neural network receiving and analyzing information from the outside world, as in the case of the visual cortex.

In the next section, we show that the existence of "external neurons" (which only project to the visual neural network, and do not receive a back projection) leads to a representation of the visual information in terms of the direct product of the reference field R and feature space F .

III. Neural Networks with External Input

In this section, we study the properties of neural networks with external neurons. We assume that the network has stochastic dynamics (the Boltzmann machines^[7]).

The network consists of N "internal" neurons, which are mutually connected *via* synapses. At a given instant, each neuron takes a value of either 0 (non-firing state) or 1 (firing state).

We represent by w_{ij} the strength of input from the j th neuron to the i th neuron.

We assume that

$$w_{ii} = 0$$

In addition to the above network of the "internal" neurons, we assume that there are T "external" neurons r_1, r_2, \dots, r_T which represent sensory signal coming from the outside world.

In the case of visual perception, the "internal" neurons can be considered to be the visual cortical cells (including cells in the LGN of the thalamus) and "external" neurons can be considered to represent the retinal

ganglion cells.

The state of the total ("internal" + "external") population of neurons would be represented by a vertex in the $N+T$ -dimensional hypercube.

$$(S, R) = (s_1, \dots, s_N, r_1, \dots, r_T) \in \{0, 1\}^{N+T}$$

It is assumed that the connections between the "external" neurons and "internal" neurons are uni-directional. Specifically, although "external" neurons project to the "internal" neurons with non-zero weight, there are no back projection from the "internal" neurons to the "external" neurons. At the same time, we assume for simplicity that there are no mutual connections between the "external" neurons. We express the connection from the j th "external" neuron to the i th "internal" neuron as α_{ij} ($i=1, 2, \dots, N; j=1, 2, \dots, T$).

It is to be noted that because of the above conditions, the states of the "external" neurons are determined solely by the nature of the signal from outside (*i.e.*, light signal falling on the retina), and not by the properties of the "internal" network of neurons.

In following discussions, we regard the states of the "external neurons" as given, and study the properties of the network of "internal" neurons under that constraint. We indicate a particular state of the "internal" neurons in $\{0, 1\}^N$ as S^k , where $k=1, 2, \dots, 2^N$.

Let us assume that the neurons are randomly chosen and called to update^[7] with rates of activation $F(i)$, where i denotes the index of the neuron that is called to update. The rate of transition from the state S^q to the state S^p can be written as

$$K(S^p, S^q) = F(i_c) \frac{1}{1 + e^{-\beta(1-2s_i^q)U_i(S^q)}} \tag{4}$$

where i_c denotes the index of the neuron that is called to update. $U_i(S)$ is the input signal on the i th neuron given as

$$U_i(S) = \sum_{j=1}^N w_{ij} s_j + \sum_{j=1}^T \alpha_{ij} r_j - \theta_i \quad (5)$$

where θ_i represents the threshold.

We assume that only one neuron changes its state at a particular time of transition (the serial update method). Namely, we have the relations

$$\begin{aligned} s_{i_c}^q &= 1 - s_{i_c}^p \\ s_i^p &= s_i^q \quad (i \neq i_c) \end{aligned} \quad (6)$$

The evolution of the system is then described by the Master equation

$$\frac{\partial \rho(S^p, t)}{\partial t} = \sum_{S^q} (K(S^p, S^q) \rho(S^q) - K(S^q, S^p) \rho(S^p)) \quad (7)$$

where $r(S)$ is the probability distribution for the state S .

Using the relations

$$\sum_{j=1}^T \alpha_{i_c, j} (s_{i_c}^q - s_{i_c}^p) r_j = \sum_{i=1}^N \sum_{j=1}^T \alpha_{ij} (s_i^q - s_i^p) r_j$$

We can then show that the rate constants satisfy the relations^[9]

$$\frac{K(S^p, S^q)}{K(S^q, S^p)} = e^{-\beta(E'(S^p) - E'(S^q) + d(S^p, S^q))} \quad (8)$$

where $E'(S^p)$, $E'(S^q)$ are the effective equilibrium energy values defined as the sum of the equilibrium energy value and a term representing the contribution of the external neurons

$$E'(S) = E(S) - \sum_{i=1}^N \sum_{j=1}^T \alpha_{ij} s_i r_j \quad (9-A)$$

The equilibrium energy $E(S)$ for the internal neurons is given as

$$E(S) = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N w_{ij} s_i s_j + \sum_{i=1}^N s_i \theta_i \quad (9-B)$$

The "asymmetric energy term"

$$d(S^p, S^q) = -\frac{1}{4} \sum_i \sum_j (w_{ij} - w_{ji}) (s_i^p - s_i^q) (s_j^p + s_j^q) \quad (10)$$

is a measure of the deviation from the equilibrium.

In a steady state, the state distribution $r(S)$ is given by the balance equation

$$\sum_{S^q} (K(S^p, S^q) \rho(S^q) - K(S^q, S^p) \rho(S^p)) = 0 \quad (11)$$

The normalization condition is

$$\sum_S \rho(S) = 1 \quad (12)$$

The steady state distribution of the Boltzmann machine can be obtained as the solution for equations (9), (11), and (12).

In order to see the effect of the input of external neurons on the state distribution of the neural network, we apply the graphic transformation method^{[8],[9]}.

We express the spanning in-trees which have a particular state S as the sink as $G_S(m)$ ($m=1,2,3 \dots n_g$), where n_g is the number of spanning in-trees of the hypercube $\{0,1\}^N$

The solution for the balance equation can be given by the graphic method^[10] as

$$W(S) = \sum_m \prod_{(S^p, S^q) \in G_S(m)} K(S^p, S^q) \quad (13)$$

where

$$\prod_{(S^p, S^q) \in G_S(m)} K(S^p, S^q) \quad (14)$$

represents the product of the rate constants corresponding to the directed edges of the spanning in-tree $G_S(m)$. The term (S^p, S^q) represents the ordered pair of states corresponding to the directed edges of the spanning in-tree $G_S(m)$.

By applying the graphic transformation method, we can transform the weight $W(S)$ as

$$W(S) = e^{-\beta E'(S)} D(S) \quad (15)$$

where S^0 is the standard state for normalization taken arbitrarily, and $P_{SS^0}(m)$ is the path in $G_S(m)$ connecting the states S^0 and S . $D(S)$ is a measure of deviation from the equilibrium given as

$$D(S) = \frac{\sum_m e^{-\beta \left(\sum_{(S^p, S^q) \in P_{SS^0}(m)} d(S^p, S^q) \right)} \prod_{(S^p, S^q) \in G_{S^0}(m)} K(S^p, S^q)}{\sum_m \prod_{(S^p, S^q) \in G_{S^0}(m)} K(S^p, S^q)} \quad (16)$$

For a symmetric ($w_{ij} = w_{ji}$) network, we have

$$D(S) = 1$$

We now note an important property of the weight representation (15). Namely, we can write (15) alternatively as

$$W(S) = e^{-\beta \left(-\sum_{i=1}^N \sum_{j=1}^T \alpha_{ij} s_i r_j \right)} w(S) \quad (17)$$

where

$$w(S) = e^{-\beta E(S)} D(S) \quad (18)$$

is the weight representing the internal neurons.

We have seen in section II that in order to integrate the various features represented in the cortex within the framework of the retinotopic reference field, we would need a direct product of the reference field \mathbf{R} and feature space \mathbf{F} . We have now demonstrated that the state distribution of the visual system can be described by the product of two elements, one representing the reference field \mathbf{R}

$$W(\mathbf{R}) = e^{-\beta \left(-\sum_{i=1}^N \sum_{j=1}^T \alpha_{ij} r_i r_j \right)}$$

and one representing the feature space \mathbf{F}

$$W(\mathbf{F}) = w(S)$$

Therefore, we conclude that the weight distribution of the visual system can indeed be written as a product of weights for the reference field \mathbf{R} and the feature space \mathbf{F} .

IV Conclusions

In our theory, the external neurons play a significant role in formulating the reference field. In biological visual systems, the retinal ganglion cells correspond to the "external" neurons. The role played by these "external" neurons is interesting. Although they provide the input to the visual system, they do not receive corresponding projection back from the visual system. Therefore, the state of the external neurons are determined solely by the visual environment, and are not influenced by the internal conditions of the visual network. The state of the external neurons represent the "constraint" imposed upon the visual system by the outside world.

It is interesting to note that the "reference field", the frame of reference for the various features represented in the visual system, is formed through the interaction between the "external" neurons and the "internal" neurons. In this way, the space structure of the outside world is "copied" into the visual cortex, providing us with an internal "model" of the outside world.

From our model, we can conclude that the geometrical properties of the "reference field" is determined by the geometrical structure of the "external" neurons (the retinal ganglion cells) and those neurons receiving input from them (the LGN cells). It is known that in the LGN and V1, the retinotopic map of the retina is represented faithfully, with the modification that the central area of vision around fovea are over-represented. In our conscious perception, we do see the world as a collection of various visual features organized in a retinotopic framework. Therefore, the prediction of our model that the reference field is defined by the interaction between the retinal ganglion cells and the LGN cells is consistent with our experience.

It is to be noted that the disembodiment of the visual system into two parts **R** and **F** is unique. Due to the existence of the asymmetric term $D(S)$ in (18), it is not possible to express the weight for the internal visual system as a product of two subsystems.

We did not discuss the role of attention. In the literature, attention is usually held responsible for the integration of visual features. We proposed an alternative mechanism, in which the visual features are integrated by assignment of the various features to a particular locus of the retinotopic field. Compared to the attentional mechanism of integration, such a integration mechanism is more stable and can be conducted in a parallel manner. We consider such a pre-attentive mechanism of visual information integration as an essential feature of visual cognition.

Acknowledgements

We would like to thank Horace Barlow, Tomoki Fukai, Yoshihide Tamori, Shigeru Tanaka for helpful discussions. We are grateful to Masao Ito for valuable advises and encouragement.

References

- [1] Zeki, S. in *A Vision of the Brain* (Blackwell Scientific Publications) (1993).
- [2] H.B. Barlow & J.D. Mollon in *The Senses* (Cambridge University Press) (1982)
- [3] Damasio, A. R. *Seminars in The NeuroSciences* 2, 287-296 (1990)
- [4] Gray, C. M., König, P., Engel, A. K. & Singer, W. *Nature* 338, 334-337. (1989)
- [5] Malsburg, C. von der & Buhmann, J. *Biol. Cybern.* 67, 233-242. (1992)
- [6] Crick, F. *Proc. Natl. Acad. Sci. USA* 81, 4586-4590. (1984)
- [7] D. H. Ackley, G. E. Hinton and T. J. Sejnowski, *Cognitive Sci.* 9, 147 (1985).
- [8] K. Mogi *J. theor. Biol.* 162, 337-352 (1993).
- [9] K. Mogi *Physical Review E* 49, 4616-4626 (1994)
- [10] E.L. King, and C. Altman, *J. Phys. Chem.* 60, 1375 (1956).