H-027

A Self-Organizing Neural Structure for Concept Formation

Noriyasu Homma† Madan M. Gupta‡ Masao Sakai† Makoto Yoshizawa† Kenichi Abe†

1. Introduction

Although the supervised backpropagation learning can be used in many successful neural applications, unsupervised learning or self-organization is an attractive function in biological neural networks [1]. For example, Hebbian learning is a biologically based unsupervised learning [2] and the self-organizing map (SOM) can learn a topological relation of the environment as its network structure [3].

On the other hand, cognition using incomplete information is an another attractive and intrinsic function in brain. The SOM, however, needs the complete observation of the input information.

In this paper we propose a self-organizing neural structure for acquiring feature space representation of logical concepts from incomplete observation by using a neuron model with dynamic and spatial changing weights (DSCWs)[4]. To form the complete informational structure of concepts, (i) a necessary connection structure is created by an extended Hebbian rule and (ii) unnecessary connections are deleted by a self-organizing competitive learning. An ability of the proposed neural model for acquiring the information structure is proven by using a concept formation problem.

2. Concept Formation Problem

Human cognition can be based on an incomplete information of a target concept. For example, a concept “apple” is described by its attributes such as “shape is round,” “color is red or green,” “taste is good,” and so on. Indeed it is very difficult to explain or discover all the attributes of “apple.”

Note that each of these attributes is also a *concept*. Thus, if we do not know about a concept of “shape,” we might not be *aware* of the attribute “shape is round” even if we *receive* this information of the shape. An interesting thing of our cognition is that even if we could not *recognize* all the attributes of a concept, we can still *understand* the concept in our way. That is, we can understand “this is an apple” even if we are not aware of its shape. In fact, we did not know a concept of “mass – energy equivalence $E = mc^2$” before Einstein had discovered. This equivalence is also an important attribute of the concepts “mass” and “energy.” We, however, *did* know and understand “mass” and “energy” before the discovery, though *deepness* of our understanding of these concepts was improved after the discovery.

According to the observation described above, let us define a concept formation problem as follows. For this problem, we use a vector representation of a *concept*, $\boldsymbol{x}$, defined by a set of feature variables $\{x_1, x_2, \dots, x_n\}$

$$\boldsymbol{x} = [x_1, x_2, \dots, x_n]^T \in \Re^n \tag{1}$$

Let us consider a set of N vector representations of N concepts, S, given as $S = \{\boldsymbol{x}^1, \boldsymbol{x}^2, \dots, \boldsymbol{x}^N\}$ where $\boldsymbol{x}^i$, $i = 1, 2, \dots, N$, are the vector representations of concepts i. A neuron *receives* the complete information of concept i, $\boldsymbol{x}^i$, but can *recognize* only an incomplete observation, $\hat{\boldsymbol{x}}^i = [\hat{x}^i_1, \hat{x}^i_2, \dots, \hat{x}^i_n]^T$, through its synaptic connection vector $\boldsymbol{w} = [w_1, w_2, \dots, w_n]^T$

$$\hat{x}^i_j = w_j x^i_j, \quad j = 1, 2, \dots, n \tag{2}$$

Let the initial weight vector be $\boldsymbol{w}(0) = 0$ and it will be developed to a connection structure $\boldsymbol{w} \neq 0$. For example, at the initial stage a neuron cannot *recognize* any information through the no connection structure.

Here the goal of the ith concept formation by the ith neuron is to satisfy the following condition

$$\hat{\boldsymbol{x}}^i \longrightarrow \boldsymbol{x}^i \tag{3}$$

In other words, the concept formation defined in this paper is an acquisition process of necessary *attention* or *awareness* structure in order to observe the complete information by developing connection weights $\boldsymbol{w}^i$. In this sense, the connection weights imply a strength of *attention* or *awareness* for the received information.

Here if we use unipolar binary values for the features, i.e, $x_i \in \{0, 1\}$, $i = 1, \dots, n$, then $x_i^2 = x_i$. This implies that the following weights vector is a solution of Eq. (3)

$$\boldsymbol{w}^i \longrightarrow \boldsymbol{x}^i \tag{4}$$

In this case, the neural connection structure will converge the *informational structure* of the concept.

Consequently, formation of a concept set S using an M-neuron network, $(M \geq N)$, can be represented by

$$S \subset S_{NN} = \{\boldsymbol{w}^1, \boldsymbol{w}^2, \dots, \boldsymbol{w}^M\} \tag{5}$$

3. Self-Organizing Network for Concept Formation

Let us consider a discrete-time DSCWs neural model[4] in which the weight vector $\boldsymbol{w}(\boldsymbol{r}(k)) = [w_1(r_1(k)), \dots, w_n(r_n(k))]^T$ is defined as a function of its spatial distance vector $\boldsymbol{r}(k) = [r_1(k), \dots, r_n(k)]^T$ between the sensory devices or another neuron’s axon branches and the corresponding target dendrites at time k. In general if a spatial distance is sufficiently short, the synapse can be formed, e.g, $w_i(r_i \simeq 0) \neq 0$. On the other hand, the synapse cannot be formed if the distance is relatively large, e.g, $w_i(r_i \gg 0) = 0$. This relation can be represented by various monotonic decreasing functions such as an exponential function $w_i(r_i) = w_0 \exp(-r_i/r_0)$ and a sigmoid function $w_i(r_i) = w_0/\{1 + \exp(\frac{r_i - r_0}{\alpha})\}$ where w_0, r_0, and α are positive constants.

We use a competitive learning that selects a neuron to represent a concept according to the output value. Here we assume that the neural cognition is a process of matching its inner awareness structure $\boldsymbol{w}$ with an observed information structure $\hat{\boldsymbol{x}}$, not with the complete information structure $\boldsymbol{x}$. As described in Section 2, human cognition may be subjective, relative, and independent of the total amount of awareness $W = \sum_{i=1}^{n} |w_i|$. In this sense the neural outputs y_j, $j = 1, \dots, M$, as a result of the cognition process for concept i can be defined as

$$y_j(k) = \frac{1}{W^j(k)} \sum_{l=1}^{n} \hat{x}^i_l(k) = \frac{(\boldsymbol{w}^j(\boldsymbol{r}(k)))^T \boldsymbol{x}^i(k)}{W^j(k)} \tag{6}$$

Then the neural outputs are re-defined as $y_j(k) = 0$, $(j \neq c)$, and $y_c(k) = y_c(k)$ where the neuron c with the largest

†Tohoku University
‡University of Saskatchewan

output among the neural network satisfies

$$c = arg\max_j y_j(k) \tag{7}$$

The competitive learning rule with an incomplete observed information $\hat{\boldsymbol{x}}$, instead of the complete information $\boldsymbol{x}$, can be defined as

$$\begin{cases} w_j^c(k+1) &= w_j^c(k) + \gamma[\hat{x}_j(k) - w_j^c(k)] \\ &= w_j^c(k)[1 + \gamma(x_j(k) - 1)] \\ w_j^i(k+1) &= w_j^i(k), \quad (i \neq c) \end{cases} \tag{8}$$

where $\gamma > 0$ is a learning constant. The learning can be achieved by changing the distance. In the case of $\gamma = 1$ and $w_j^c(k) = 1$, note that if $\hat{x}_j(k) = 1$, that is $x_j(k) = 1$, then this feature is needed to be recognized and thus the connection structure will not be changed $w_j^c(k+1) = 1$. On the other hand, if $\hat{x}_j(k) = 0$, that is $x_j(k) = 0$, this feature is an unnecessary information and thus this connection should be disappear $w_j^c(k+1) = 0$. However, if $w_j^c(k) = 0$, this learning rule cannot change this weight.

The creation of a new connection can be achieved by the following extended Hebbian rule for structural adaptation given as

$$\begin{aligned} r_{ji}(k+1) &= r_{ji}(k) - \eta r_{ji}(k)x_i(k)y_j(k) \\ &= r_{ji}(k)(1 - \eta x_i(k)y_j(k)) \end{aligned} \tag{9}$$

where $\eta > 0$ is a growing constant and x_i and y_j are outputs of sensory and cognitive neurons, respectively. Here, if the two neurons i and j fire simultaneously, the distance $r_{ji}(k+1) = r_{ji}(k)(1-\eta)$. On the other hand, they do not fire simultaneously, the distance will not change. For simplicity, let $\eta = 1$, $x_i \in \{0,1\}$, $y_c = 1$, and $y_{j\neq c} = 0$ then

$$r_{ji}(k+1) = \begin{cases} 0, & (x_i(k)y_c(k) = 1) \\ r_{ji}(k), & (x_i(k)y_j(k) = 0) \end{cases} \tag{10}$$

Note that, using the monotonic decreasing functions $\boldsymbol{w}(\boldsymbol{r})$, the changes in distances affect the weights value.

To get the solutions given by Eqs. (4) and (5), a self-organizing algorithm for the DSCWs model is now proposed. Let us consider the unipolar binary case for features of the concept vector $\boldsymbol{x}$ and the weight vector $\boldsymbol{w}$; that is, $x_i, w_i \in \{0,1\}$.

1) Initialize iteration $k = 0$, the number of neurons $m(0) = 1$, and the connection vector $||\boldsymbol{w}^1(0)|| = 0$.

2) Select one concept vector $\boldsymbol{x}^{i(k)} \in S$, $i(k) \in \{1, 2, \ldots, N\}$ and input it to the neural network.

3) For the concept $\boldsymbol{x}^{i(k)}$, select a candidate neuron c according to Eq. (7) and re-define the outputs.

4) Self-organizing the weights according to the the following condition

Case 1: If the candidate output y_c is equal to 1, create a new *attention* weight w_i^c so that $\hat{x}_i(k) = 1$ by the extended Hebbian rule in Eq. (10) if possible.

Case 2: If $0 < y_c < 1$, create new neuron $m(k)+1$ with the same weight vector $\boldsymbol{w}^{m(k)+1} = \boldsymbol{w}^c$ and $m(k)+1 \longrightarrow m(k)$. Then $c = m(k)$ and learn the weight to delete unnecessary connections by using Eq. (8).

Case 3: In the case of $y_c = 0$ if there is a neuron without any connections, then select this neuron for a candidate. Otherwise, create a new candidate neuron and $m(k)+1 \longrightarrow m(k)$. Then create a randomly selected new *attention* weight for this candidate by Eq. (10).

5) $m(k+1) = m(k)$, $k \to k+1$ and return to Step 2).

4. Discussions

A concept formation ability of the proposed network is summarized in the following theorem.

Theorem 1 *For a unipolar binary concept set* $S = \{\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^N\}$*, there exists a self-organizing network with* M *neurons,* $(M \geq N)$*, such that*

$$\boldsymbol{x}^i = \boldsymbol{w}^{j_i}, \quad i = 1, 2, \ldots, N, \quad j_i \in \{1, 2, \ldots, M\}$$

Proof: In Step 3), a candidate neuron c is selected for a concept vector $\boldsymbol{x}^i$. Then even if $y_c(\boldsymbol{x}^i, \boldsymbol{w}^c(k)) < 1$,

$$y_c(\boldsymbol{x}^i, \boldsymbol{w}^c(k+1)) = 1 \tag{11}$$

in Step 4). This is because we can delete only the wrong connections for the target concept by using the competitive learning rule. Also if $y_c(k) = 1$, then $||\boldsymbol{w}^c(k+1)|| > ||\boldsymbol{w}^c(k)||$ except for $\boldsymbol{w}^c(k) = \boldsymbol{x}^i$ by using the Hebbian rule that can create only the correct connection; there is no wrong connection weight for the candidate. That is,

$$||\boldsymbol{x}^i - \boldsymbol{w}^c(k+1)|| < ||\boldsymbol{x}^i - \boldsymbol{w}^c(k)|| \tag{12}$$

except for $\boldsymbol{w}^c(k) = \boldsymbol{x}^i$. Thus if this neuron is a candidate only for concept i

$$\boldsymbol{w}^c \longrightarrow \boldsymbol{x}^i \tag{13}$$

Otherwise, if this neuron c will also be a candidate (i.e. $y_c = 1$) for another concept $\boldsymbol{x}^j$, $j \neq i$ and $||\boldsymbol{x}^j|| > ||\boldsymbol{w}^c||$, then $\boldsymbol{w}^c \longrightarrow \boldsymbol{x}^j$. Once $\boldsymbol{w}^c$ converges to $\boldsymbol{x}^j$, an another candidate neuron $c'(\neq c)$ for concept i will not be a candidate for concept j since neuron c is the only candidate for concept j. Thus after all the concepts like concept j have their own candidates, the weight of a candidate neuron c' will converge to concept i

$$\boldsymbol{w}^{c'} \longrightarrow \boldsymbol{x}^i \tag{14}$$

Therefore if M is sufficiently large

$$S \subset S_{NN} = \{\boldsymbol{w}^1, \boldsymbol{w}^2, \ldots, \boldsymbol{w}^M\} \qquad \blacksquare$$

In addition to this ability, a major difference between the Kohonen's self-organizing map (SOM)[3] and the proposed method given in Eqs. (7) and (8) is completeness of the input information. SOM supposes that the complete information of the input $\boldsymbol{x}$ can be given, while only an incomplete observation of the input, $\hat{\boldsymbol{x}}$, is available to use in our method. In the sense of a hypothesis that human cognition is based on an incomplete information, the proposed network provides a better model of human cognition.

References

[1] M. M. Gupta, L. Jin, and N. Homma, *Static and Dynamic Neural Networks: From Fundamentals to Advanced Theory*, Wiley & IEEE Press, Hoboken, NJ, 2003.

[2] D. O. Hebb, *The Organization of Behavior*, Wiley, New York, 1949.

[3] T. Kohonen, *Self-Organization and Associative Memory*, 3rd ed., Springer-Verlag, 1989.

[4] N. Homma and M. M. Gupta, "Superimposing Neural Learning by Dynamic and Spatial Changing Weights," *J. Artificial Life and Robotics*, vol. 7, 2003 (Accepted).