

Contradiction Resolution of Competitive and Input Neurons to Improve Prediction and Visualization Performance

Ryotaro Kamimura

IT Education Center and Graduate School of Science and Technology
Tokai University, 1117 Kitakaname, Hiratsuka, Kanagawa, Japan
ryo@keyaki.cc.u-tokai.ac.jp

Abstract—In this paper, we propose a new type of information-theoretic method to resolve the contradiction observed in competitive and input neurons. For competitive neurons, contradiction between self-evaluation (individuality) and outer-evaluation (collectivity) exists, which is reduced to realize the self-organizing maps. For input neurons, there exists contradiction between the use of many and few input neurons. We try to realize a situation where as many input neurons as possible are used, and at the same time, another where only a few input neurons are used. This contradictory situation can be resolved by viewing input neurons on different levels, namely, the individual and average level. We applied contradiction resolution to two data sets, namely, the Japanese short term economy survey (Tankan) and Dollar-Yen exchange rates. In both data sets, we succeeded in improving the prediction performance. Many input neurons were used on average, but a few input neurons were only taken for each input pattern. In addition, connection weights were condensed into a small number of distinct groups for better prediction and interpretation performance.

Keywords—contradiction resolution; self- and outer-evaluation; visualization; self-organizing maps; dependent input neuron selection

I. INTRODUCTION

A. Contradiction Resolution

We have so far introduced contradiction resolution for neural networks [1], [2]. We believe that neural networks can be viewed from multiple points of view. If contradiction exists between the different points of view, it should be reduced as much as possible. An example of contradiction is two types of evaluation for a neuron [2], namely, self- and outer-evaluation. In self-evaluation, a neuron is evaluated for itself without considering the other neurons. On the other hand, in outer-evaluation, the neuron can be evaluated by all the other neurons. If contradiction between self- and outer-evaluation exists, this contradiction should be reduced as much as possible.

Contradiction resolution has previously been applied to the self-organizing maps (SOM) [3], [4], [5]. In particular, it has been used to improve the visualization and clarification of class structure. Roughly speaking, outer-evaluation corresponds to cooperation between neurons, which is not considered in self-evaluation. Thus, the self-organizing maps can be realized by minimizing contradiction between self- and outer-evaluation.

When we applied the method to the SOM, we focused on the extraction of clear class structure. The SOM is well-known

and has been established as one of the main techniques in the visualization and interpretation of neural networks. Though the SOM has a good reputation in visualization and interpretation, we have had serious difficulty in extracting useful information from the knowledge obtained. Thus, a number of different types of methods have been developed to clarify SOM knowledge, [6], [7], [8], [9], [10], [11], [12]. In addition, there have been many other methods to create more interpretable connection weights by changing the learning procedures [13], [14], [15], [16]. However, we cannot say that SOM knowledge can be easily interpreted with these methods. Contradiction resolution has been previously introduced to clarify class structure or to create more interpretable connection weights, because the characteristics shared by self- and outer-evaluation tend to be enhanced [2].

B. Contradiction in Input Neurons

The above example of contradiction is concerned with competitive (output) neurons in the SOM. We have found that in addition to competitive neurons, contradiction can be found in input neurons in terms of their types of responses and the number of input neurons.

First, we try to resolve contradiction between the actual and expected responses of input neurons. We have found several cases where only specific input neurons among many tend to respond to all input patterns. For example, in our experiments in this paper, the most important input neuron (winning neuron) tended to respond to specific data (period), irrespectively of different input patterns. Because these specific input neurons are only used in learning, we do not have any way to examine whether the other input neurons have a role in learning. To examine the roles of all available input neurons, we need to develop a method to force all input neurons to play a role in learning. Thus, we need to resolve contradiction between the use of specific and all input neurons.

However, we can immediately point out another contradiction in input neurons, namely, contradiction between many and few neurons. If we interpret input neurons as input variables, this contradiction is related to the variable selection method widely used in machine learning [17], [18], [19]. Using variable selection, we can more easily interpret data and internal representations, reducing the computational time and storage. In addition, it is widely believed that reducing the number of input variables (neurons) is effective in improving prediction

performance [17]. Thus, we have contradiction between a few and many neurons to be actually used in learning. We thus try to use many input neurons while at the same time using only a few. To resolve contradiction between many and few input neurons, we introduce dependent input neuron selection. In dependent input neuron selection, for a given input pattern, a few input neurons are taken, while for another input pattern, a few but different input neurons are used. For individual input patterns, the number of input neurons is small. On the other hand, many input neurons are used on the whole.

C. Outline

In Section 2, we first explain the concept of contradiction resolution in the order of competitive neuron, input neurons and the number of input neurons. By using the Kullback-Leibler divergence, we formulate contradiction between self- and outer-evaluation for competitive neurons. Then, contradiction between actual and expected responses of input neurons is formulated. Contradiction in the number of input neurons is resolved by introducing dependent input neuron selection. We applied the method to two data sets, namely, the short term economic survey and the Dollar-Yen exchange rates. In both data sets, prediction performance was improved, condensing connection weights into a few groups. In addition, the first, the second and the third winners responded to a variety of input patterns.

II. THEORY AND COMPUTATIONAL METHODS

A. Contradiction Resolution

We consider learning to be a process of contradiction resolution. We here propose three types of contradiction resolution processes, namely, contradiction resolution in terms of competitive neurons, input neurons and the number of input neurons.

1) *Competitive Neurons:* Contradiction resolution [1], [2] has been introduced to produce the self-organizing maps. In the self-organizing maps, neurons behave collectively and individual neurons must imitate those collective behaviors. Figure 1 shows the process of contradiction resolution. In Figure 1(a1), a neuron at the center is self-evaluated to fire; the firing rate of the neuron is determined by the neuron itself. On the other hand, in Figure 1(a2), a neuron at the center is outer-evaluated; the firing rate of the neuron is determined by all surrounding neurons. Because the firing rates obtained by the self- and outer-evaluation are low and high, respectively, we have contradiction between self- and outer-evaluation. Figure 1(b) shows an example of resolved contradiction. As shown in Figure 1(b1), the firing rate by the self-evaluation is forced to match the level by the outer-evaluation in Figure 1(b2). Thus, contradiction between self- and outer-evaluation is finally resolved.

2) *Input Neurons:* For input neurons, we think that there is contradiction between actual and expected responses. Let's look at an example to explain concretely this contradiction. We here explain a problem of the data used in our experiments in Section 3. We chose the time series data for demonstrating our method, because we found that the $t - 1$ th input neuron was mainly used for predicting the t th period, as in Figure 2(a1). This means that the input neuron responded mainly to the

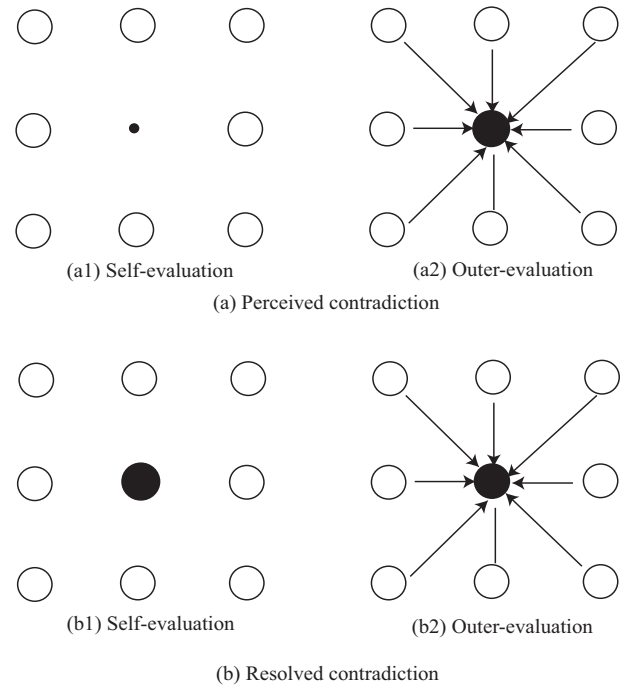


Fig. 1. Contradiction resolution between self- and outer-evaluation for competitive neurons.

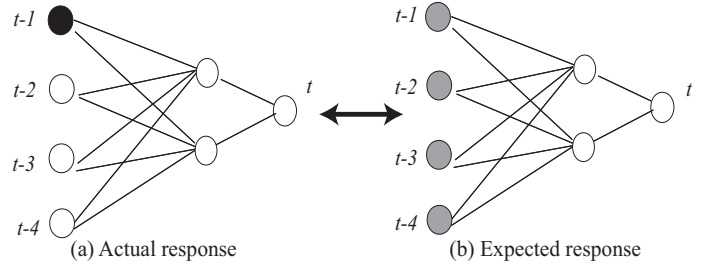


Fig. 2. Contradiction resolution between actual and expected responses of input neurons.

immediately previous step for predicting the present step. We found this fact in both data sets in this paper. If it is possible to take into account many input neurons or more previous steps, it may be possible to improve the prediction performance. More concretely, we should consider the $t - 2$, $t - 3$ and $t - 4$ as well as the $t - 1$ th step. Thus, we have a contradiction that the actual responses of input neurons are biased toward the immediately previous step, while the expected responses are more uniform responses of input neurons. This contradiction between actual (biased) responses and expected (uniform) ones must be resolved, as shown in Figure 2(b).

3) *Number of Input Neurons:* Finally, we have the other contradiction. As mentioned above, we expect that input neurons respond more uniformly to input patterns. However, for better prediction performance, one of the more conventional methods is to reduce the number of input variables (neurons), namely, variable subsection selection, as shown in Figure 3(b). This is contrary to the state we try to achieve in Figure 2(b). This contradiction is resolved by introducing dependent input neuron selection, as shown in Figure 3(c). In dependent input neuron selection, a smaller number of input neurons are used, but they are chosen depending on input patterns. As shown in

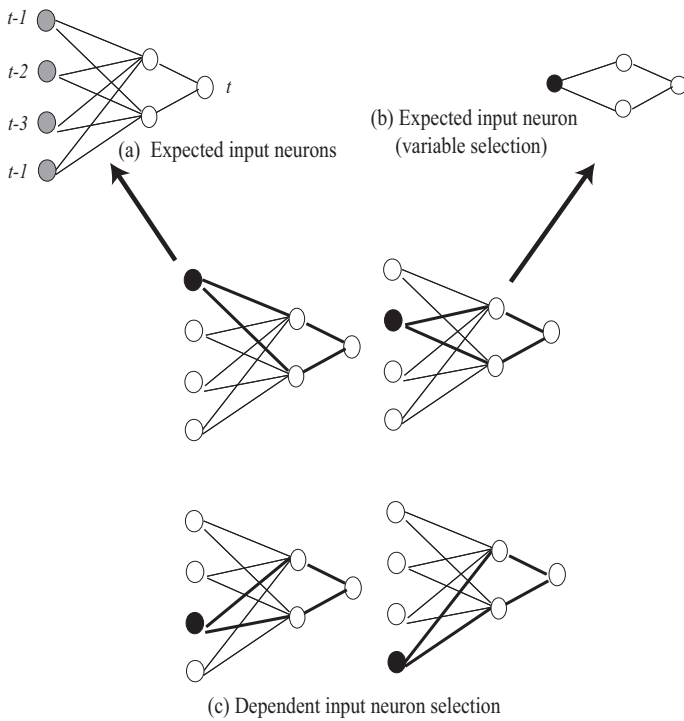


Fig. 3. Contradiction resolution by dependent input neuron selection.

Figure 3(c), for each neural network, only one input neuron is used. However, a different input neuron is used for each network in Figure 3(c). When we average these firing rates of four input neurons, they become the same for all input patterns in Figure 3(a). Thus, a smaller number of input neurons can be expected for some input neurons with averaged uniform firing rates.

B. Contradiction Resolution for Competitive Neurons

First, we consider contradiction resolution for competitive neurons in Figure 4. We suppose that a neuron, for example, a neuron in the middle of network in Figure 4, is evaluated by itself (self-evaluation) and evaluated by the other neurons (outer-evaluation). If any contradiction between self- and outer-evaluation exists, it should be reduced as much as possible. In terms of firing rates, if the firing rates by self-evaluation are different from those by outer-evaluation, this difference should be as small as possible.

1) *Self- and Outer-Evaluation*: Let us explain how to compute outputs from competitive neurons and input patterns in Figure 4. The s th input pattern can be represented by $\mathbf{x}^s = [x_1^s, x_2^s, \dots, x_L^s]^T$, $s = 1, 2, \dots, S$. Connection weights into the j th competitive neuron are computed by $\mathbf{w}_j = [w_{1j}, w_{2j}, \dots, w_{Lj}]^T$, $j = 1, 2, \dots, M$. Now, we can compute the firing rates by self- and outer-evaluation. A neuron's firing rates by self-evaluation can be defined by using the outputs from the neuron. Then, the j th competitive neuron output without considering the other neurons can be computed by

$$v_j^s = \exp \left(-\frac{\sum_{k=1}^L (x_k^s - w_{kj})^2}{2\sigma_{\beta_1}^2} \right). \quad (1)$$

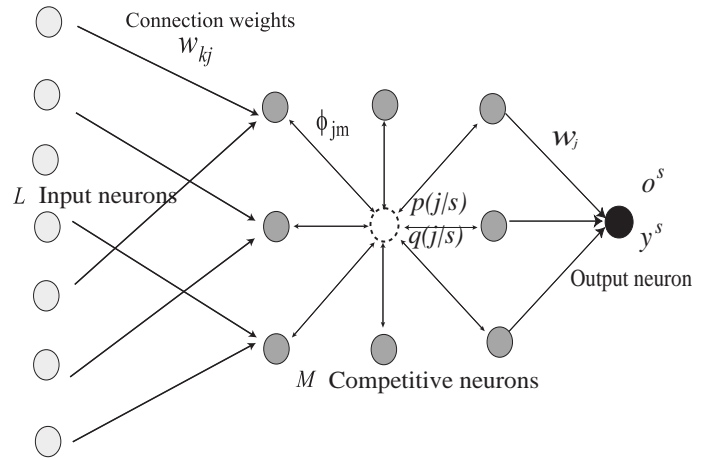


Fig. 4. Network architecture for contradiction resolution.

where \mathbf{x}^s and \mathbf{w}_j are supposed to represent L -dimensional input and weight column vectors, where L denotes the number of input neurons. The parameter σ_{β_1} is computed by

$$\sigma_{\beta_1} = \frac{1}{\beta_1}, \quad (2)$$

where β_1 is larger than zero. Thus, the firing rate by self-evaluation is computed by

$$p(j | s) = \frac{\exp \left(-\frac{\sum_{k=1}^L (x_k^s - w_{kj})^2}{2\sigma_{\beta_1}^2} \right)}{\sum_{m=1}^M \exp \left(-\frac{\sum_{k=1}^L (x_k^s - w_{km})^2}{2\sigma_{\beta_1}^2} \right)}. \quad (3)$$

The output of a neuron by outer-evaluation is determined by the sum of all outputs of all the other neurons. We suppose that the result by outer-evaluation does not contain that by self-evaluation; given this, we have the final output by the outer-evaluation

$$z_j^s = \sum_{m=1}^M \phi_{jm} v_m^s - v_j^s, \quad (4)$$

where ϕ_{jm} denotes the relation between the j th and m th neuron. When we apply the method to the self-organizing maps, this corresponds to the neighborhood function. Then, the firing rate by the outer-evaluation is defined by

$$q(j | s) = \frac{z_j^s}{\sum_{m=1}^M z_m^s}. \quad (5)$$

2) *Reducing Contradiction*: Contradiction resolution aims to reduce contradiction between self- and outer-evaluation. We use the Kullback-Leibler divergence to represent this contradiction. Using the Kullback-Leibler divergence, contradiction is defined by

$$C_1 = \sum_{s=1}^S p(s) \sum_{j=1}^M p(j | s) \log \frac{p(j | s)}{q(j | s)}. \quad (6)$$

In addition to this contradiction, we have quantization errors between connection weights and input patterns

$$Q_1 = \sum_{s=1}^S p(s) \sum_{j=1}^M p(j | s) \sum_{k=1}^L (x_k^s - w_{kj})^2. \quad (7)$$

To realize minimum contradiction in terms of contradiction ratio and quantization errors, first, we try to minimize the KL divergence with fixed quantization errors. Then, we have the optimal firing rates

$$p^*(j | s) = \frac{q(j | s) \exp\left(-\frac{\sum_{k=1}^L (x_k^s - w_{kj})^2}{2\sigma_{\beta_1}^2}\right)}{\sum_{m=1}^M q(m | s) \exp\left(-\frac{\sum_{k=1}^L (x_k^s - w_{km})^2}{2\sigma_{\beta_1}^2}\right)}. \quad (8)$$

By substituting this optimal firing rate $p^*(j | s)$ for $p(j | s)$, we have the free energy:

$$F_1 = -2\sigma_{\beta_1}^2 \sum_{s=1}^S p(s) \log \sum_{j=1}^M q(j | s) \times \exp\left(-\frac{\sum_{k=1}^L (x_k^s - w_{kj})^2}{2\sigma_{\beta_1}^2}\right). \quad (9)$$

The free energy can be expanded into

$$F_2 = \sum_{s=1}^S p(s) \sum_{j=1}^M p^*(j | s) \sum_{k=1}^L (x_k^s - w_{kj})^2 + 2\sigma_{\beta_1}^2 \sum_{s=1}^S p(s) \sum_{j=1}^M p^*(j | s) \log \frac{p^*(j | s)}{q(j | s)}. \quad (10)$$

Thus, we actually minimize KL divergence as well as quantization errors. By differentiating the free energy, we have the re-estimation equation

$$\mathbf{w}_j = \frac{\sum_{s=1}^S p^*(j | s) \mathbf{x}^s}{\sum_{s=1}^S p^*(j | s)}. \quad (11)$$

C. Contradiction for Input Neurons

As mentioned above, we try to use as many input neurons as possible. On the other hand, only a specific input neuron responds to the input patterns in our data. This contradiction between the actual (biased) and expected (uniform) responses of input neurons must be resolved as much as possible. For measuring contradiction, we must compute the output from the k th input neuron.

Let us now compute the output from the k th input unit. The output from the k th input neuron is defined by

$$y_k^s = \exp\left(-\frac{\sum_{j=1}^M (x_k^s - w_{kj})^2}{2\sigma_{\beta_2}^2}\right) \quad (12)$$

Then, the firing rate is computed by

$$p(k | s) = \frac{\exp\left(-\frac{\sum_{j=1}^M (x_k^s - w_{kj})^2}{2\sigma_{\beta_2}^2}\right)}{\sum_{l=1}^L \exp\left(-\frac{\sum_{j=1}^M (x_l^s - w_{lj})^2}{2\sigma_{\beta_2}^2}\right)} \quad (13)$$

Because we try to achieve uniform responses of input neurons, the expected firing rate should be determined by $1/L$. Thus, we should maximize the entropy to achieve this uniform distribution

$$C_2 = \sum_{s=1}^S p(s) \sum_{j=1}^M p(k | s) \log p(k | s). \quad (14)$$

We can define two free energies for input and competitive neurons and we must minimize two types of free energy at the same time.

$$F = -2\sigma_{\beta_1}^2 \sum_{s=1}^S p(s) \log \sum_{j=1}^M q(j | s) \exp\left(-\frac{\|\mathbf{x}^s - \mathbf{w}_j\|^2}{2\sigma_{\beta_1}^2}\right) - 2\sigma_{\beta_2}^2 \sum_{s=1}^S p(s) \log \sum_{k=1}^L \exp\left(-\frac{\sum_{j=1}^M (x_k^s - w_{kj})^2}{2\sigma_{\beta_2}^2}\right) \quad (15)$$

For simplification, we suppose that two spread parameters β_1 and β_2 are equal. Then, by differentiating the free energy, we have the re-estimation rule

$$\mathbf{w}_j = \frac{\sum_{s=1}^S (p^*(j | s) + p(k | s)) \mathbf{x}^s}{\sum_{s=1}^S (p^*(j | s) + p(k | s))}. \quad (16)$$

D. Contradiction resolution by dependent input neuron selection

We have known that the reduction of the number of input neurons (variables) can be used to improve prediction performance [17]. This reduction is contrary to the previous contradiction resolution, which aimed to achieve the uniform responses of input neurons. We reduce this contradiction by introducing dependent input neuron selection. This method aims to choose an input neuron depending on a given input pattern. Though a small number of input neurons are used for the input patterns, they can be different for different input patterns. However, on average, they are uniformly used. Thus, the dependent input neuron selection can solve the contradiction between biased and uniform responses of input neurons. Figure 5 shows how to choose the input neurons according to the importance of the neurons.

Let us compute the output from the k th input unit. The output from the k th input neuron is defined by

$$v_k^s = \exp\left(-\frac{\sum_{j=1}^M (x_k^s - w_{kj})^2}{2\sigma_{\beta_2}^2}\right) \quad (17)$$

When the s th input pattern is presented, we first determine the winner g_1

$$g_1 = \operatorname{argmax}_k v_k^s. \quad (18)$$

Then, based on this winner, we can obtain a winning ranking, as shown in Figure 5,

$$g_1 < g_2 < \dots < g_L. \quad (19)$$

Neurons of this ranking should keep the following relation:

$$v_{g_1}^s > v_{g_2}^s > \dots > v_{g_L}^s. \quad (20)$$

By using this ranking, we can define the output from the j th neuron with only r winning input neurons

$$v_j^s(r) = \exp\left(-\frac{\sum_{k=1}^r (x_{g_k}^s - w_{g_k j})^2}{2\sigma_{\beta_2}^2}\right), \quad (21)$$

where r is a specified number of input neurons and less than or equal to the number of input neurons. By using this output, we obtain the outputs by outer-evaluation for the r input neurons

$$z_j^s(r) = \sum_{m=1}^M \phi_{jm} v_m^s(r) - v_j^s(r), \quad (22)$$

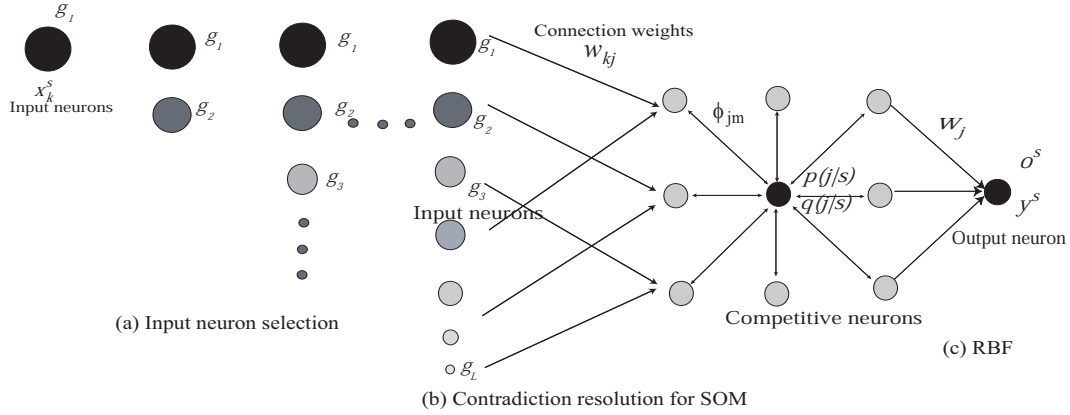


Fig. 5. Input neuron selection.

The firing rates are defined by

$$q(j | s; r) = \frac{z_j^s(r)}{\sum_{m=1}^M z_m^s(r)}. \quad (23)$$

By using these firing probabilities, we have

$$p^*(j | s) = \frac{q(j | s; r) \exp\left(-\frac{\sum_{k=1}^r (x_{g_k}^s - w_{g_k j})^2}{2\sigma_{\beta_2}^2}\right)}{\sum_{m=1}^M q(m | s; r) \exp\left(-\frac{\sum_{k=1}^r (x_{g_k}^s - w_{g_k m})^2}{2\sigma_{\beta_2}^2}\right)}. \quad (24)$$

We minimize two types of free energy at the same time:

$$\begin{aligned} F = & -2\sigma_{\beta_1}^2 \sum_{s=1}^S p(s) \log \sum_{j=1}^M q(j | s; r) \\ & \times \exp\left(-\frac{\sum_{k=1}^r (x_k^s - w_{kj})^2}{2\sigma_{\beta_1}^2}\right) \\ & -2\sigma_{\beta_2}^2 \sum_{s=1}^S p(s) \\ & \times \log \sum_{k=1}^L \exp\left(-\frac{\sum_{j=1}^M (x_k^s - w_{kj})^2}{2\sigma_{\beta_2}^2}\right). \quad (25) \end{aligned}$$

By differentiating the free energy with $\beta_1 = \beta_2$, the re-estimation rule can be computed by

$$\mathbf{w}_j = \frac{\sum_{s=1}^S (p^*(j | s; r) + p(k | s)) \mathbf{x}^s}{\sum_{s=1}^S (p^*(j | s; r) + p(k | s))}. \quad (26)$$

Finally, we should note a modification in the experiments. The parameter β_1 is supposed to be equivalent to the parameter β_2 for the easy implementation of our method. However, it is interesting to examine the relations between competitive and input neurons. For this, in the following experiments we slightly change the values of the two to examine the relation between competitive and input neurons. Specifically, the parameter β_2 is defined by

$$\beta_2 = \alpha \beta_1, \quad (27)$$

where α is larger than zero. When the parameter α is smaller, the effect of the input neurons becomes weaker.

III. RESULTS AND DISCUSSION

A. Experimental Outline

We here present two experimental results for demonstrating the good performance of our method. We used the mean squared error (MSE) for the testing data to measure the performance. In addition, we computed quantization and topographic errors [20] to measure the quality of the maps. The quantization error is the average error between input patterns and connection weights into the first winner. On the other hand, the topographic error is the percentage of input patterns that are not neighboring neurons.

B. Short-Term Economic Survey (Tankan)

We used the quarterly short-term economic survey of principal enterprise in Japan. This is called "Tankan." The data set ranged between March 1983 and December 2012. The training data was between March 1983 and December 2004, and the testing data was between March 2005 and December 2012.

1) *Quantitative Evaluation*: Table I shows the summary of experimental results for the Tankan data. When the parameter β was 48, the parameter α was 1, and the number of input winners was 4, the MSE was 70.215. When the parameter α was 0.7, the MSE decreased to 66.298. These MSE values were lower than the 85.022 by the SOM and much lower than the 188.319 by the feed-forward selection (FS) of RBF and the 424.849 by Ridge regression (RR) of RBF networks. The quantization error increased from 0.783 by SOM to 1.493 ($\alpha = 1.0$) and to 1.448 ($\alpha = 0.7$). The topographic error increased from 0.012 by SOM to 0.602 ($\alpha = 1.0$) and to 0.988 ($\alpha = 0.7$). Thus, good prediction performance was accompanied by a degradation in map quality.

2) *Visual Inspection*: Figure 6 (a) shows the Tankan rates by contradiction resolution with $\alpha = 0.7$. The actual rates in blue were close to the predicted ones in red. By SOM in Figure 6(b), differences between actual and predicted ones were slightly larger than those by contradiction resolution. By using the feed-forward selection of RBF networks, differences between actual and predicted rates were larger in the middle of the period in Figure 6(c). By using Ridge regression of RBF networks, Tankan rates fluctuated in the later period in Figure 6(d).

TABLE I. MSE FOR THE TESTING DATA, QUANTIZATION ERRORS (QE) AND TOPOGRAPHIC ERRORS (TE) FOR THE TANKAN DATA BY CONTRADICTION RESOLUTION, SOM, FEED-FORWARD SELECTION (FS) AND RIDGE REGRESSION (RR) OF RBF NETWORKS.

Method	Beta	Alpha	Win	MSE	QE	TE
Contradiction	41	0.7	3	66.298	1.448	0.988
SOM	48	1.0	4	70.215	1.493	0.602
RBF(FS)				188.319		
RBF(RR)				424.849		

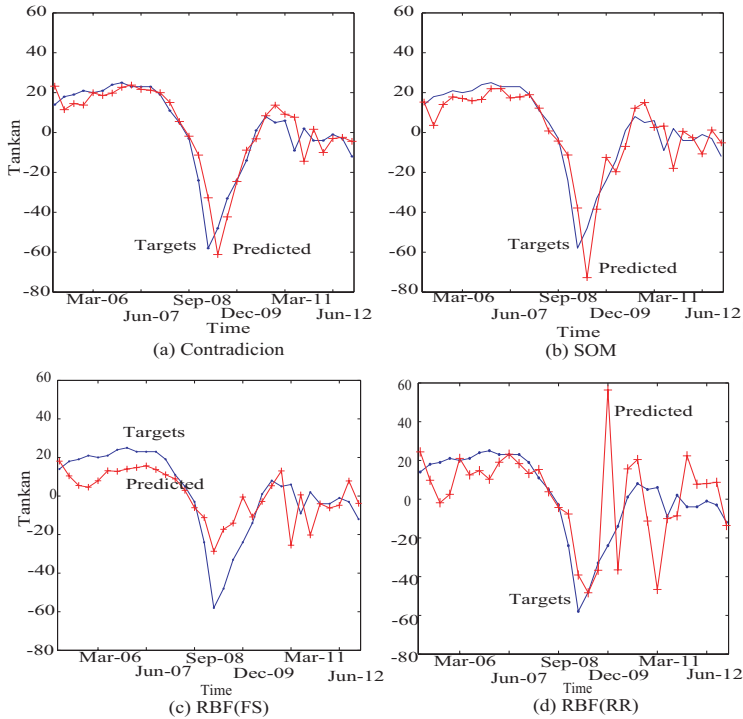


Fig. 6. Tankan rates in blue and predicted values in red by contradiction resolution (a), SOM (b), RBF(FS) (c) and RBF (RR) (d) for 5 by 3 sized map.

Figure 7(b) shows the results of PCA for connection weights by contradiction resolution. Connection weights were condensed into four distinct groups. On the other hand, by using the SOM, no regularity could be seen on the results of PCA in Figure 7(a).

Figure 8(a1) shows to which time lag the first input winner responded by the conventional SOM. The first input winner responded mainly to the $t - 1$ th and $t - 5$ th time lag. On the other hand, by contradiction resolution in Figure 8(a2), the first winner tended to respond more uniformly to the time lag. Figure 8(b) shows the response to the time lags by the second winner. The second winner by the SOM responded to the $t - 2$ th and $t - 4$ th time lag in Figure 8(b1). On the other hand, the second winner by contradiction resolution responded to the time lags almost uniformly in Figure 8(b2). Figure 8(c) shows the responses by the third winner. The third winner by the SOM responded to the $t - 3$ th time lag in Figure 8(c1), while the third winner by the contradiction resolution responded mainly to the $t - 4$ th lag but responded to many time lags (input neurons) in Figure 8(c2). These results show that by contradiction resolution, many input neurons were used to respond to input patterns.

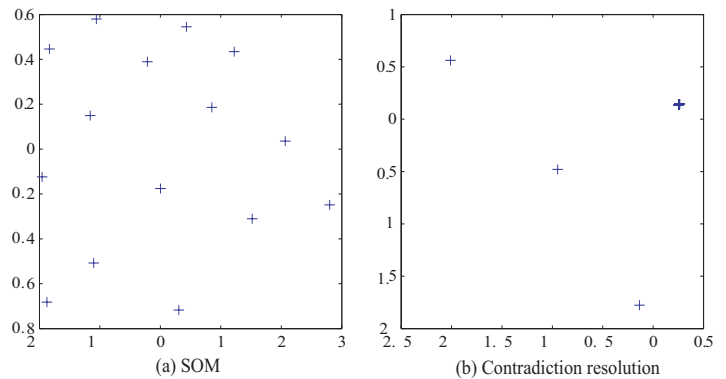


Fig. 7. Results by the principal component analysis (PCA) for connection weights by the SOM (a) and the contradiction resolution (b) for the Tankan data.

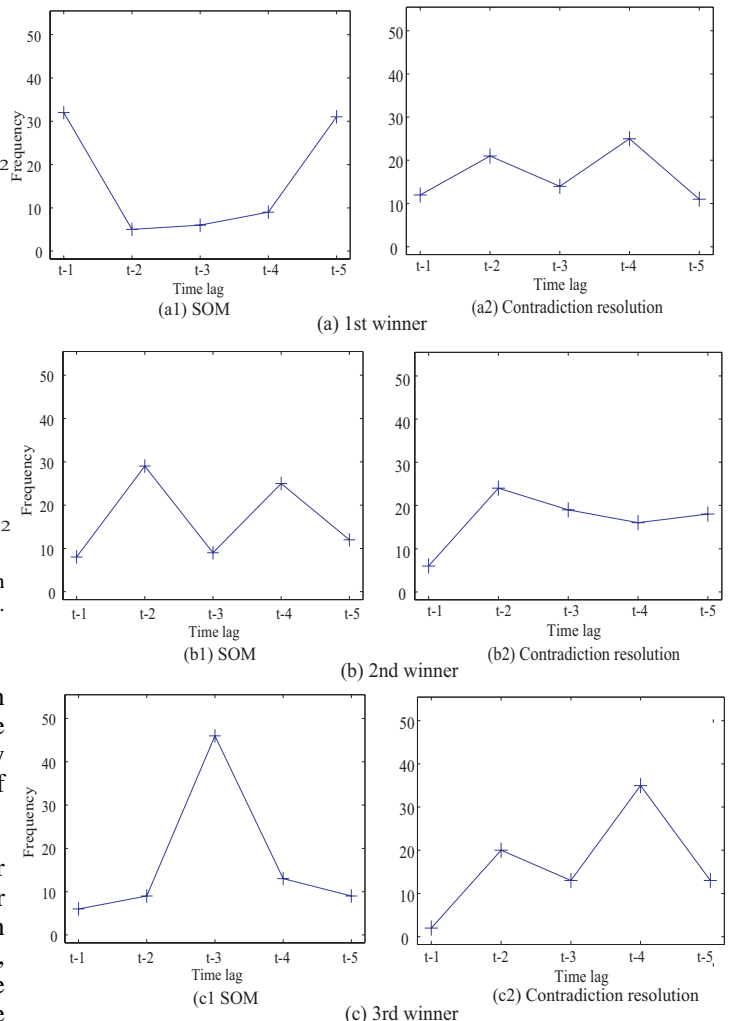


Fig. 8. Frequency of responses by the winners for each time lag for the Tankan data.

C. Dollar-Yen Exchange Rate Estimation

1) *Experiment Outline:* We used the dollar-yen exchange rates during 2012 (for the results of 2011, see [1]). The time lag was ten, meaning that the exchange rate at the t th period was determined by the rates of $t - 1, t - 2, \dots, t - 10$. The training data ranged between January to September, while the testing data ranged from October to December. For comparison, we

TABLE II. THE PARAMETER β AND α , THE NUMBER OF INPUT WINNING NEURONS, MSE FOR THE TESTING DATA, QUANTIZATION ERRORS (QE) AND TOPOGRAPHIC ERRORS (TE) FOR THE DOLLAR-YEN EXCHANGE RATES.

Method	Beta	Alpha	Win	MSE	QE	TE
Contradiction	36	0.5	4	0.122	1.601	0.145
SOM	32	1.0	1	0.182	1.579	0.436
RBF(FS)				0.270	0.986	0.073
RBF(RR)				5.870		

used several conventional methods such as the RBF networks with feed-forward selection and Ridge regression with GCV model selection criteria.

2) *Quantitative Evaluation:* Table II shows the summary of experimental results when the network size was 5 by 3 for the dollar-yen exchange rates. When the parameter β was 32, the parameter α was 1.0, the number of input winning neurons was one, the MSE was 0.182. When the parameter α was decreased to 0.5, the MSE further decreased to 0.122. The MSE values were much smaller than the 0.270 by SOM, 0.233 by the feed-forward selection of RBF and 5.870 by the Ridge regression of RBF networks. However, quantization errors increased from 0.986 by SOM to 1.579 ($\alpha = 1.0$) and to 1.601 ($\alpha = 0.5$). The topographical errors also increased from 0.073 by SOM to 0.436 ($\alpha = 1.0$) and to 0.145 ($\alpha = 0.5$). The prediction performance was improved at the expense of map quality in terms of quantization and topographic errors.

3) *Visual Inspection:* Figure 9(a) shows the actual and predicted rates by contradiction resolution with $\alpha = 0.5$. Actual and predicted rates were almost equal for the entire range of the period. When we used the SOM in Figure 9(b), differences between actual and predicted rates became larger as the time went on. By using the feed-forward selection of RBF networks in Figure 9(c), the predicted rates were almost flat for the entire range of the period. By the Ridge regression of RBF networks in Figure 9(d), the differences were relatively small, but they were larger than those by contradiction resolution for the entire range of the period. The experimental results showed that the prediction performance for the entire period was much improved by contradiction resolution.

Figure 10(a) shows the results of PCA for connection weights by contradiction resolution. We can see that connection weights were condensed into one major group with two distinct groups by contradiction resolution. On the other hand, by SOM in Figure 10(b), no regularity could be seen on the results of PCA. We think that the condensation of connection weights into a small number of groups was one of the main reasons for better performance.

Figures 11 (a1), (b1) and (c1) show the frequency of winning input neurons by the conventional SOM. As can be seen in the figures, the first, the second and third winning input neurons responded to the $t-1$ th, $t-2$ and $t-3$ th time lag by the conventional SOM. On the other hand, the first winning neuron by contradiction resolution in Figure 11(b1) responded to the $t-1$ th and $t-5$ th time lag mainly. In addition, the second and third winning neuron respond almost uniformly to the time lags in Figure 11 (b2) and (b3). Experimental results showed that the specific responses of input neurons were attenuated and

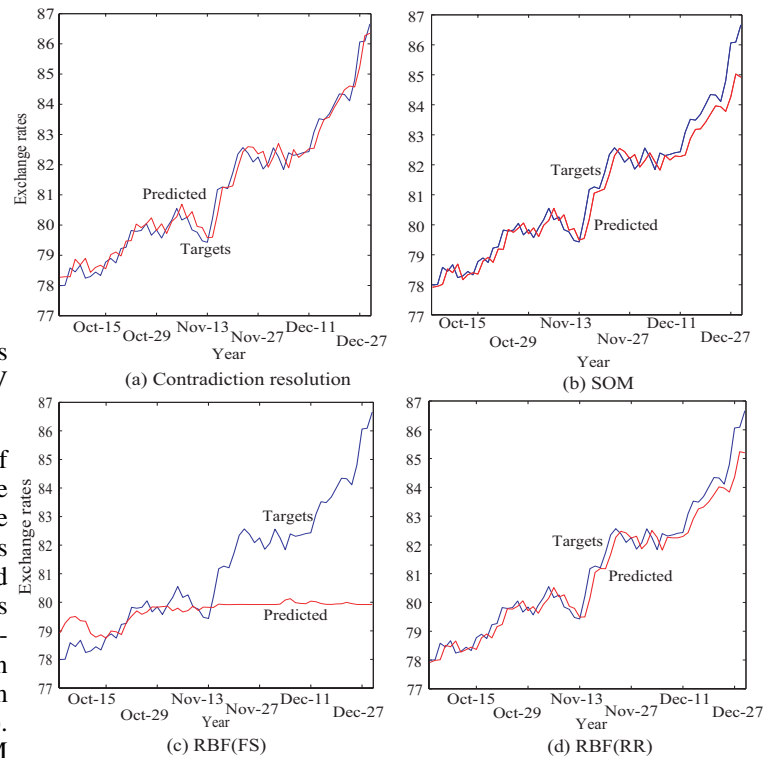


Fig. 9. Dollar-yen exchange rates and predicted values by contradiction resolution (a), SOM (b), RBF(FS) (c) and RBF (RR) (d) for 5 by 3 sized map.

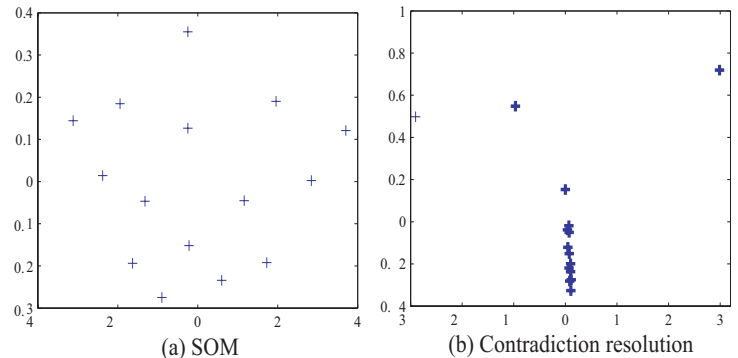


Fig. 10. Results by the principal component analysis (PCA) for connection weights by contradiction resolution (a), SOM (b) and data itself (c) for Dollar-Yen exchange rates.

input neurons tended to respond to many different time lags. This property was certainly related to improved performance.

D. Discussion

1) *Validity of Methods and Experimental Results:* In this paper, we have introduced contradiction resolution between self- and outer-evaluation, and between expected and actual responses of input neurons. First, we compute two types of firing rates of a neuron. The first type is computed by self-evaluation. The firing rate is determined by its own responses to input patterns. On the other hand, the firing rate of a neuron is determined by outer-evaluation, namely, other neurons. If contradiction or difference between self- and outer-evaluation exists, this should be reduced as much as possible. The outer-evaluation corresponds to cooperation between competitive

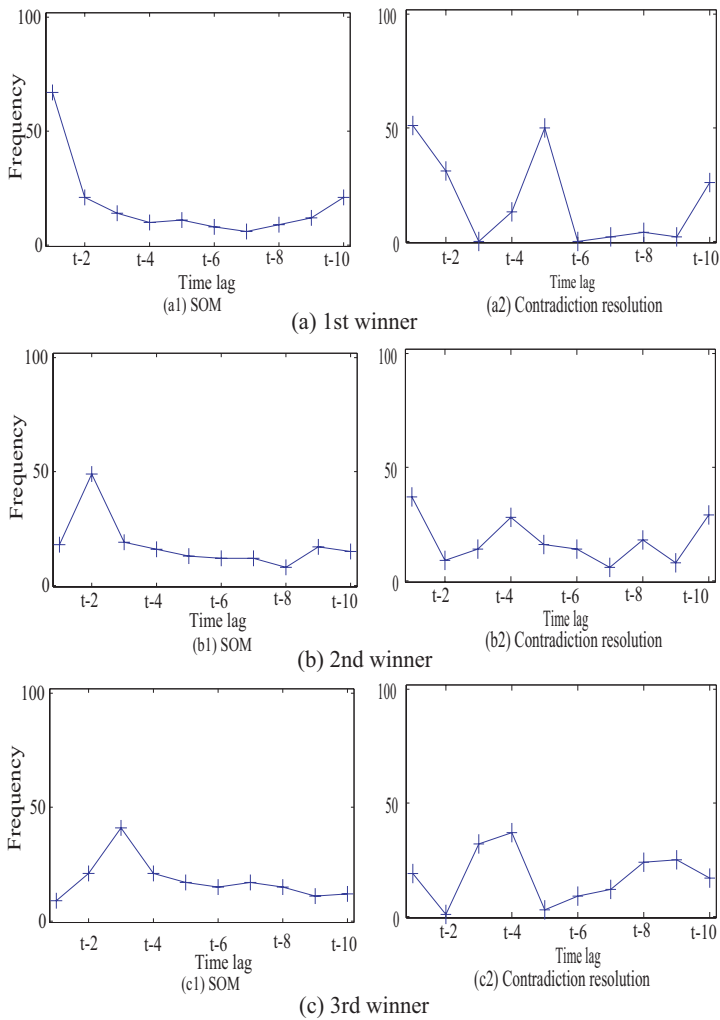


Fig. 11. Frequency of the responses by the winners for each time lag for dollar-yen exchange rates.

neurons. Thus, contradiction between self- and outer-evaluation can be used to realize self-organizing maps.

The second type of contradiction is between actual and expected responses of input neurons. We have found that only specific input neurons respond to input patterns for some problems. The other input neurons are of no use in learning. We tried to use as many input neurons as possible.

However, we have known that to improve the prediction performance, the number of input neurons (variables) should be reduced. This is contradictory to the expected responses we try to achieve. The contradiction can be resolved by introducing dependent input neuron selection. In dependent input neuron selection, a few input neurons are chosen as in the variable selection. However, chosen input neurons can be different, depending on input patterns. Thus, when we see the responses of input neurons for a specific input neuron, few input neurons are used. However, on average, all input neurons can be used.

We applied the method to two data sets, namely, the short term economic survey (Tankan) and dollar-yen exchange rates. In both data sets, when contradiction resolution was introduced, much better performance in terms of MSE for testing

data was observed, as in Tables I and II. Better performance was intuitively observed by plotting actual and estimated values for testing data in Figures 6 and 9. However, in these data sets, the map resolution in terms of quantization errors, and topological preservation in terms of topographic errors tended to degrade, as in Tables I and II. The better prediction performance was obtained at the expense of resolution and topological preservation.

The better performance of our method can be explained by the use of inactive input neurons and the condensation of connection weights. First, the good performance was due to the use of many inactive neurons. In both data sets, important input neurons tended to respond mainly to the specific input patterns. For example, the first winner tended to respond to the immediately previous step ($t - 1$ th step) for predicting the present step (t th step), as in Figures 8 (a1) and 11 (a1). By introducing contradiction resolution, many different input neurons tended to respond to input patterns, as in Figures 8 (a2) and 11 (a2). In addition, almost uniform responses could be detected in 8 (b2) and 11 (b2). Our method tried to use many input neurons actively. On the other hand, other conventional methods used, such as variable selection methods dealt passively with input neurons or input variables. The active use of input neurons or variables was one of the main reasons for better prediction performance.

Second, we could observe the condensation of connection weights. Our method detected four groups in the Tankan data, as in Figure 7(b). In the dollar-yen exchange rate data, one major group and two minor ones were detected, as in Figure 10(b). On the other hand, no explicit regularity could be distinguished by the conventional SOM, as shown in 7(a) and 10(a). The condensation of connection weights into a small number of groups is another reason for better prediction performance.

2) Limitation and Problems of Method: Though our method produced better results across all data sets, we should point out three problems or limitations of our method, namely, the parameters β and α , and the resolution and topological preservation of the maps.

First, contradiction between self- and outer-evaluation is determined by the parameter β . We define the spread parameter σ_β by using the parameter β

$$\sigma_\beta = \frac{1}{\beta}. \quad (28)$$

Thus, when the parameter β is increased, competition becomes more of the winner-take-all type. When the parameter β is decreased, the soft type of competition emerges. When the parameter β is too high, contradiction between self- and outer-evaluation becomes intense and produces maps that cannot well represent input patterns. Additionally, the characteristics shared by self- and outer-evaluation become clearer. Because of the focus on shared characteristics, the final maps do not necessarily represent input patterns accurately. On the other hand, when the parameter β is smaller, contradiction between self- and outer-evaluation becomes weaker and shared characteristics become weaker as well. At the present state of research, no explicit rules exist to determine the parameter β . Future studies should thus clarify the relationship between the parameter β and network performance.

Second, we have a problem determining the parameter α . Though competitive and input neurons are supposed to be governed by the same parameter, β in our formulation of contradiction resolution, we have empirically found that input and competitive neurons should be controlled by different parameter rules for better prediction performance. This means that in order to improve prediction performance, contradiction in competitive and input neurons should be more carefully treated. More concretely, by changing the values of the parameter α , we can change the characteristics of input and competitive neurons. For example, when α is small, an input neuron is more weakly evaluated, while competitive neurons are more strongly evaluated. We can control the relation between input and competitive neurons. Thus, we need to develop a method to take into account the different properties of competitive and input neurons.

Third, we have the problem of poorer resolution and topological preservation in terms of quantization and topographic errors. In all experimental results, we found that the best prediction performance was not accompanied by the better resolution and topological preservation. This means that better prediction performance is contradictory to better resolution and topological preservation. By the results of the PCA in Figures 7(b) and 10(b), we could see that connection weights were condensed into several groups by contradiction resolution. This condensation is related to better prediction and interpretation at the expense of quantization and topographic errors. Thus, we need to develop a method to provide better prediction performance, keeping higher resolution and topological preservation.

3) *Possibility of the Method*: The possibility of our method can be summarized by three points, namely, interpretation, application to time-series analysis and new types of evaluation. First, the method can be applied to the procedure where neural networks produce easily interpretable internal representations. In neural networks, a critical problem is how to interpret internal representations. Though many methods have been developed, we have had still serious problems with interpretation. Our method aims to produce explicit and interpretable representations by stressing the characteristics shared by self- and outer-evaluation.

Second, our method can be applied to the time series analysis. In the time series analysis, one key problem is how to take into account the previous states behind the present state. It has been difficult to consider long-term correlation between the present and the previous states. This is because conventional methods have tried to describe the time-series without taking into account the properties of input neurons. Our method can force some inactive input neurons to respond to input patterns through as many input neurons as possible. This property is very different from the conventional approach to time series analysis. Our method actively uses the previous steps, while conventional methods receive the previous steps very passively.

Third, we aim principally to describe and use many different types of evaluation for neurons. We have introduced self- and outer-evaluation, but these are only one realization of social interaction to be observed in neurons. We can imagine many different kinds of interaction between neurons. If it is possible to take into account different types of interactions between neurons, in particular, social interaction observed in actual human societies, then it will be possible to model human

societies and simulate how individual and collective human being behave.

IV. CONCLUSION

In the present paper, we introduced contradiction resolution to improve the prediction and interpretation performance of neural networks. Contradiction is realized in terms of competitive neurons, input neurons and the number of input neurons. For competitive neurons, there was contradiction between self- and outer-evaluation. In self-evaluation, the firing rate of a neuron is determined only by the neuron itself, while in outer-evaluation, the firing rates of a neuron are determined by all surrounding neurons. By resolving contradiction between self- and outer-evaluation, competitive neurons tend to acquire the collective behaviors realized by outer-evaluation.

For input neurons, there was contradiction between actual and expected responses. In our data, only a small number of specific input neurons tended to respond to input patterns. For example, the first winning input neuron tended to respond to the input neuron representing one period before ($t-1$ th period). We expect an input neuron to respond to many types of input patterns.

For the number of input neurons, we resolved the contradiction by introducing dependent input neuron selection. We know that a small number of neurons (variables) are effective in improving prediction performance. In variable selection, the number of input variables is reduced to improve prediction performance. This reduction in the number of input neurons is contrary to the objective we achieved, namely, the diversity of input neurons. For resolving this contradiction, we introduced dependent input neuron selection. The number of input neurons was small for an input pattern. However, many input neurons were used on average. Thus, this dependent input neuron selection aims to resolve contradiction between a small number of neurons and the use of many input neurons.

We applied this contradiction resolution to two data sets, namely, the short term economic survey and dollar-yen exchange rates. In both sets, the first, the second and the third winners tended to respond to specific input neurons, the representing $t-1$ th, $t-2$ th and $t-3$ th periods. By using contradiction resolution, the input neuron tended to respond to many different types of input neurons. For example, the first neuron did not necessarily respond to the first input neuron, but many different input neurons. In addition, connection weights were condensed into a small number of groups. This condensation is related to improved prediction performance.

However, we observed that quantization and topographic errors did not decrease by using contradiction resolution. Thus, when we try to interpret final connection weights, we should interpret them with due consideration of the quality of internal representations.

Though several problems should be solved for practical application, contradiction resolution can be used to improve prediction as well as interpretation performance.

REFERENCES

- [1] R. Kamimura, "Contradiction resolution between self and outer evaluation for supervised multi-layered neural networks," *Journal of Advanced Research in Artificial Intelligence*, vol. 2, no. 7, pp. 29–38, 2013.

- [2] R. Kamimura, "Separation and unification of individuality and collectivity and its application to explicit class structure in self-organizing maps," *Artificial Neural Networks and Machine Learning-ICANN 2012*, pp. 387-394, 2012.
- [3] T. Kohonen, "The self-organizing maps," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464-1480, 1990.
- [4] T. Kohonen, *Self-Organization and Associative Memory*. New York: Springer-Verlag, 1988.
- [5] T. Kohonen, *Self-Organizing Maps*. Springer-Verlag, 1995.
- [6] J. Vesanto, "SOM-based data visualization methods," *Intelligent Data Analysis*, vol. 3, pp. 111-126, 1999.
- [7] S. Kaski, J. Nikkila, and T. Kohonen, "Methods for interpreting a self-organized map in data analysis," in *Proceedings of European Symposium on Artificial Neural Networks*, (Bruges, Belgium), 1998.
- [8] I. Mao and A. K. Jain, "Artificial neural networks for feature extraction and multivariate data projection," *IEEE Transactions on Neural Networks*, vol. 6, no. 2, pp. 296-317, 1995.
- [9] C. De Runz, E. Desjardin, and M. Herbin, "Unsupervised visual data mining using self-organizing maps and a data-driven color mapping," in *Information Visualisation (IV), 2012 16th International Conference on*, pp. 241-245, IEEE, 2012.
- [10] S. Shieh and I. Liao, "A new approach for data clustering and visualization using self-organizing maps," *Expert Systems with Applications*, 2012.
- [11] H. Yin, "ViSOM-a novel method for multivariate data projection and structure visualization," *IEEE Transactions on Neural Networks*, vol. 13, no. 1, pp. 237-243, 2002.
- [12] M.-C. Su and H.-T. Chang, "A new model of self-organizing neural networks and its application in data projection," *IEEE Transactions on Neural Networks*, vol. 123, no. 1, pp. 153-158, 2001.
- [13] S. Wu and T. Chow, "Prsom: A new visualization method by hybridizing multidimensional scaling and self-organizing map," *Neural Networks, IEEE Transactions on*, vol. 16, no. 6, pp. 1362-1380, 2005.
- [14] L. Xu, Y. Xu, and T. W. Chow, "PolSOM-a new method for multidimensional data visualization," *Pattern Recognition*, vol. 43, pp. 1668-1675, 2010.
- [15] Y. Xu, L. Xu, and T. Chow, "Pposom: A new variant of polsom by using probabilistic assignment for multidimensional data visualization," *Neurocomputing*, vol. 74, no. 11, pp. 2018-2027, 2011.
- [16] L. Xu and T. Chow, "Multivariate data classification using polsom," in *Prognostics and System Health Management Conference (PHM-Shenzhen), 2011*, pp. 1-4, IEEE, 2011.
- [17] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157-1182, 2003.
- [18] A. Rakotomamonjy, "Variable selection using SVM-based criteria," *Journal of Machine Learning Research*, vol. 3, pp. 1357-1370, 2003.
- [19] S. Perkins, K. Lacker, and J. Theiler, "Grafting: Fast, incremental feature selection by gradient descent in function space," *Journal of Machine Learning Research*, vol. 3, pp. 1333-1356, 2003.
- [20] K. Kiviluoto, "Topology preservation in self-organizing maps," in *In Proceedings of the IEEE International Conference on Neural Networks*, pp. 294-299, 1996.

AUTHOR PROFILE

Ryotaro Kamimura is currently a professor of IT Education Center and Graduate School of Science and Technology of Tokai University in Japan. His research interests are information-theoretic approach to neural computing.