

becomes more intense, and more severe competition processes are realized. On the other hand, when obtained information is smaller, competition between neurons becomes weaker and all neurons tend to fire equally. This means that just by adjusting the information to be obtained by learning, we can control the competition processes. In addition, the method is not winner-take-all, and many neurons can participate in competition and cooperation. By controlling the information content in neural networks, we can easily control its performance.

C. Necessity of Double Competition

The information-theoretic SOM has shown good performance in clarifying class structure. However, the method cannot always detect clear class structure, in particular when the problems are complex. To resolve this, we introduce the concept of competition into input neurons, as mentioned above. In our framework, the input neurons must compete with each other to represent input patterns. In addition, if an output neuron fires at the same time as an input neuron, the corresponding connection weights between the input and output neuron should be stronger. We call this competition "double competition", because input and output neurons compete with each other to represent input patterns.

We have so far tried to introduce competition in input neurons, which is called "information enhancement". In information enhancement, we tried to enhance competition between neurons by focusing on specific input neurons [48], [49]. On the other hand, we have combined information maximization in input neurons and output neurons, which are separately defined [50]. Those methods have shown improved performance for several problems. However, they are not always effective for taking into account the combined effect of input and output neurons. In this double competition, we suppose two types of actions, namely, competition and mutual interaction. In the competition, input as well as output neurons compete with each other. In addition, we suppose some interaction between input and output neurons. Concretely, when an input and output pattern fire in the same way, the interaction between them becomes stronger.

D. Outline

In Section 2, we first explain the correspondence between information maximization and competitive learning. We explain the concept of double competition to include input and output neurons. Then, we try to present the information-theoretic learning method to realize double competition by using the free energy. Finally, we explain how to estimate the firing probabilities of input neurons. In Section 3, we present two experimental results from the well-known machine learning database. Using a principal component analysis, we try to show that class structure can be clarified by using the present method. However, we point out that topological preservation may be sacrificed for this better visualization. Thus, it is important to more closely examine the relations between better visualization and topological preservation.

II. THEORY AND COMPUTATIONAL METHODS

A. Double Competition

Competitive learning has been considered to be one of the most important learning methods in neural networks [51],

[2], [52], [53], [4], [5], [3], [54], [55], [1], [3], [56], [7], [57], [58]. In particular, we have introduced information-theoretic competitive learning [59], [60], [61]. Contrary to the computational methods so far developed, we have supposed that competitive learning is a realization of mutual information maximization between output neurons and input neurons. In competitive learning, attention has been mainly to output neurons. However, we can imagine that any components in a neural network compute with each other and we try to apply the concept of competition to input neurons. In the input neurons, we focus on the importance of input neurons. Because input neurons correspond to input variables, the importance of input variables should be taken into account.

Let us explain how to produce self-organizing maps by using a network architecture in Figure 1. 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 denoted by $\mathbf{w}_j = [w_{1j}, w_{2j}, \dots, w_{Lj}]^T$, $j = 1, 2, \dots, M$. Supposing that the firing rate $p(k | s)$ of the k th input neuron for the s th input pattern can be computed, then the distance between input patterns and connection weights can be computed by

$$\|\mathbf{x}^s - \mathbf{w}_j\|^2 = \sum_{k=1}^L p(k | s)(x_k^s - w_{kj})^2. \quad (1)$$

The firing rate $p(k | s)$ is considered to be the importance of the k th input neuron for the s th input pattern. The output from an output neuron is computed by

$$v_j^s = \exp\left(-\frac{\|\mathbf{x}^s - \mathbf{w}_j\|^2}{2\sigma^2}\right), \quad (2)$$

where σ denotes a spread parameter and defined by

$$\sigma = \frac{1}{\beta}, \quad (3)$$

where β is larger than zero. By normalizing the output, we have the firing rate

$$p(j | s) = \frac{\exp\left(-\frac{\|\mathbf{x}^s - \mathbf{w}_j\|^2}{2\sigma^2}\right)}{\sum_{m=1}^M \exp\left(-\frac{\|\mathbf{x}^s - \mathbf{w}_m\|^2}{2\sigma^2}\right)}. \quad (4)$$

We should also compute the collective output from the neuron. In the self-organizing maps, the collective outputs are determined by distances from the winner. The winner c_1 is determined by

$$c_1 = \operatorname{argmax}_j v_j^s. \quad (5)$$

Following the formulation of SOM, we compute the distance between the winner and the other neurons by

$$\phi_{jc_1} = \exp\left(-\frac{\|\mathbf{r}_j - \mathbf{r}_{c_1}\|^2}{2\sigma_{ngh}^2}\right), \quad (6)$$

where \mathbf{r}_j denotes the position of the j th neuron on the output map and σ_{ngh} is the spread parameter. Thus, the expected output is approximated by this function

$$q(j | s) = \frac{\phi_{jc_1}}{\sum_{m=1}^M \phi_{mc_1}}. \quad (7)$$

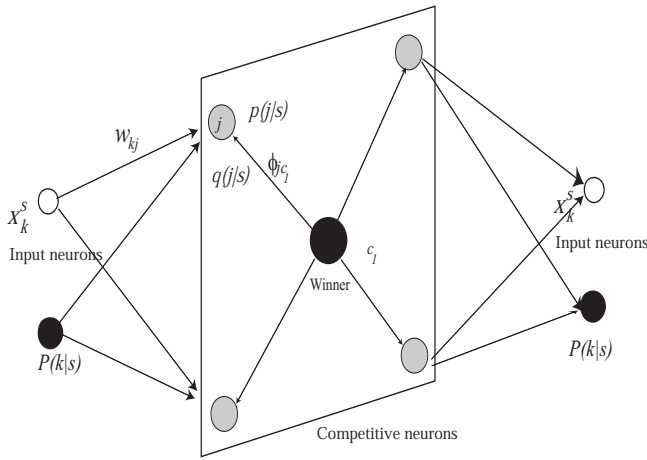


Fig. 1. Network architecture for the information-theoretic self-organizing maps where only connection weights from a winner at the center are plotted.

B. Free Energy Minimization

Learning should be performed to reduce the difference between actual and expected outputs. We represent this difference by using the Kullback-Leibler divergence

$$C = \sum_{s=1}^S p(s) \sum_{j=1}^M p(j | s) \log \frac{p(j | s)}{q(j | s)}. \quad (8)$$

In addition to this KL divergence, we have other errors which must be minimized, namely quantization errors between connection weights and input patterns

$$Q = \sum_{s=1}^S p(s) \sum_{j=1}^M p(j | s) \|\mathbf{x}^s - \mathbf{w}_j\|^2. \quad (9)$$

Fixing this quantization errors and minimizing the KL-divergence, we have the optimal firing rates

$$p^*(j | s) = \frac{q(j | s) \exp\left(-\frac{\|\mathbf{x}^s - \mathbf{w}_j\|^2}{2\sigma^2}\right)}{\sum_{m=1}^M q(m | s) \exp\left(-\frac{\|\mathbf{x}^s - \mathbf{w}_m\|^2}{2\sigma^2}\right)}. \quad (10)$$

We have the following equation called "free energy"

$$F = 2\sigma^2 \sum_{s=1}^S \log \sum_{j=1}^M q(j | s) \exp\left(-\frac{\|\mathbf{x}^s - \mathbf{w}_j\|^2}{2\sigma^2}\right). \quad (11)$$

Finally, by differentiating the free energy, we can obtain the re-estimation formula

$$\mathbf{w}_j = \frac{\sum_{s=1}^S p^*(j | s) p(k | s) \mathbf{x}^s}{\sum_{s=1}^S p^*(j | s) p(k | s)}. \quad (12)$$

As shown in the equation (10), connection weights are modified to make the actual outputs closer to expected outputs.

C. Estimating Firing Rates of Input Neurons

To obtain connection weights w_{kj} , we must estimate the firing rates of input neurons $p(k | s)$ in the equation (1). For

this purpose, we first compute the outputs from the j th neuron by

$$\hat{v}_j^s = \exp\left(-\frac{\sum_{k=1}^L (x_k^s - w_{kj})^2}{2\sigma^2}\right), \quad (13)$$

Normalizing this output, we have the estimated firing rates

$$\hat{p}(j | s) = \frac{\hat{v}_j^s}{\sum_{m=1}^M \hat{v}_m^s}. \quad (14)$$

By using $\hat{p}(j | s)$, we have the output from the k th input neuron

$$\hat{v}_k^s = \exp\left(-\frac{\sum_{j=1}^M \hat{p}(j | s) (x_k^s - w_{kj})^2}{2\sigma^2}\right), \quad (15)$$

Then, we have the firing rate of the k th input neuron

$$\hat{p}(k | s) = \frac{\hat{v}_k^s}{\sum_{l=1}^L \hat{v}_l^s}. \quad (16)$$

Putting this firing rate $\hat{p}(k | s)$ in the equation (1), (2), we have the distance considering input neurons,

$$\|\mathbf{x}^s - \mathbf{w}_j\|^2 = \sum_{k=1}^L \hat{p}(k | s) (x_k^s - w_{kj})^2. \quad (17)$$

III. RESULTS AND DISCUSSION

A. Experiment Outline

We applied the method two data sets, namely, the glass and dermatology data. Both were taken from the well-known machine learning database [62]. The number of input neurons and patterns for the glass data were 214 and 10, respectively. The number of input neurons and patterns for the dermatology data were 366 and 34, respectively. All the data were normalized to range between zero and one. For quantitative evaluation, we used the well-known quantization and topographic errors. There have been many attempts [12], [13], [14], [15], [16], [17], [18] to measure map quality quantitatively. Among them, both errors are very simple and easy to implement. For visual evaluation, we used the principal component analysis (PCA) to summarize connection weights. As mentioned in the introduction section, there is difficulty in interpreting SOM knowledge, a number of methods have been developed to clarify the knowledge [20], [21], [22], [23], [24], [25], [26], [27], [28], [29]. In this study, we used the PCA for clarification, in particular for simplifying the knowledge. It is easy to demonstrate the performance by using the techniques specific to the SOM, such as the U-matrix. However, we used the PCA so that the present results could be widely interpreted and reproduced.

B. Glass Data

1) *Firing Rates of Input Neurons*: First, we examine how the firing rates could be changed by increasing the parameter β or by increasing information content in input neurons. When the parameter β was increased from one in Figure 2(a) to three in Figure 2(c), little change in the firing rates could be seen. When the parameter β was increased from four in Figure 2(d) to eight in Figure 2(f), the firing rates became gradually differentiated. When the parameter β was increased from ten

in Figure 2(g) to 14 in Figure 2(i), higher and lower firing rates becomes clearer. Finally, when the parameter β was increased from 16 in Figure 2(j) to 20 in Figure 2(l), the clearest firing rates could be seen. Input neurons No.6 and No.8 had the highest firing rates, while input neuron No.5 had the lowest firing rate. The results show that when the parameter β was increased, the firing rates became gradually clearer.

2) *Results of PCA for Connection Weights:* Figure 3 shows the results of the PCA for connection weights by the SOM and the information-theoretic method with double competition. Figure 3(a) shows the results of PCA by using the conventional SOM. We can see that a condensed group could be seen on the right hand side of the map, and the remaining connection weights were scattered widely on the left hand side of the map. When the parameter β was two and three in Figures 3(b) and (c), connection weights were close to those by the conventional SOM in Figure 3(a). When the parameter β was increased from four in Figure 3(d) to eight in Figure 3(f), a group on the right side began to separate from the others. When the parameter β was increased from 10 in Figure 3(g) to 14 in Figure 3(i), two explicit groups of connection weights on the left and right hand sides began to form. Finally, when the parameter β was increased from 16 in Figure 3(j) to 20 in Figure 3(l), connection weights were separated into two groups on the left hand side and right hand side. In addition, another group could be seen in the middle of the map.

Figure 4 shows the results of PCA by the information-theoretic methods without considering input neurons. When the parameter β was increased from two in Figure 4(a) to 20 in Figure 4(c), a condensed group on the right hand side remained the same, but connection weights on the left hand side became more scattered. These results show that when the parameter β was increased, input patterns were separated into explicit groups by using the information on input neurons. On the other hand, without considering input neurons, explicit groups could not be expected.

3) *Quantization and Topographic Errors:* We have seen that class structure is clearer by using the information-theoretic method with double competition. The next step is to quantify the map quality obtained using this method. Figure 5(a) shows the quantization errors by the information-theoretic method with double competition in red, without considering input neurons in blue, and SOM in black. The information-theoretic method without considering input neurons showed a sharp decrease in quantization errors, while the quantization errors by SOM and the method with double competition had relatively higher errors. Topographic errors using the information-theoretic method without considering input neurons were quite large. On the other hand, topographic errors did not increase when using the information-theoretic method with double competition. The decrease in quantization errors and increase in the topographic errors by the information-theoretic method without considering input neurons can be inferred from free energy equation (18) (for more detailed discussion, see the discussion section). On the other hand, quantization and topographic errors did not change excessively with the double competition information-theoretic method, and were close to the errors obtained by the conventional SOM. Thus, it can be said that the introduction of information on input neurons attenuated the operation of the free energy.

C. Dermatology Data

1) *Firing Rates of Input Neurons:* We applied the information-theoretic method to the well-known data set of the dermatology from the machine learning database. Figure 6 shows the firing rates of input neurons when the parameter β was increased from one (a) to 15 (i). Even if we increased the parameter β beyond this point, little change could be seen in the firing rates. When the parameter β was one in Figure 6(a), the firing rates were almost uniform. When the parameter β was two and three in Figure 6(b) and (c), small changes in the firing rates appeared. When the parameter β was increased from five in Figure 6(d) to nine in Figure 6(f), differences between higher and lower rates became larger. When the parameter β was increased from 11 in Figure 6(g) to 15 in Figure 6(i), higher and lower firing rates were at their largest.

2) *Results of the PCA for Connection Weights:* Figure 7 shows the results of the PCA for connection weights by the SOM (a) and the information-theoretic method with double competition when the parameter β was increased from one (b) to 15 (i). By using the SOM, as shown in Figure 7(a), connection weights seemed to be divided into three groups with weak boundaries. When the parameter β was one and three in Figures 7(b) and (c), the results of the PCA were almost equivalent to that by the SOM in Figure 7(a). When the parameter β was increased from five in Figure 7(d) to nine in Figure 7(f), a distinct group became separated on the right hand side of the map. When the parameter β was further increased from 11 in Figure 7(g) to 15 in Figure 7(i), three groups were clearly separated.

Figure 8 shows the results of the PCA by the information-theoretic method without considering input neurons. When the parameter β was increased from one in Figure 8(a) to nine in Figure 8(b), three groups became more apparent. Then, even when the parameter β was increased from nine in Figure 8(b) to 15 in Figure 8(c), the results of the PCA remained almost the same. The results of the PCA by the information-theoretic method without considering input neurons were inferior to those by the information-theoretic method with double competition in terms of class structure. This shows that the information of input neurons is critical in clarifying class structure.

3) *Quantization and Topographic Errors:* Figure 9 shows the quantization and topographic errors when the parameter β was increased from one to 15. Figure 9(a) shows quantization errors by the SOM in black, the information-theoretic method with double competition in red, and without considering input neurons in blue. By using the information-theoretic method without considering input neurons, the quantization errors decreased sharply from the beginning onwards. On the other hand, by using the information-theoretic method with double competition, quantization errors increased and became larger than that by the conventional SOM. Figure 9(b) shows topographic errors when the parameter β was increased from one to 15. By using the information-theoretic method without considering input neurons, the topographic error increased sharply and eventually became much larger than the error obtained by the conventional SOM. On the other hand, by using the information-theoretic method with double competition, the

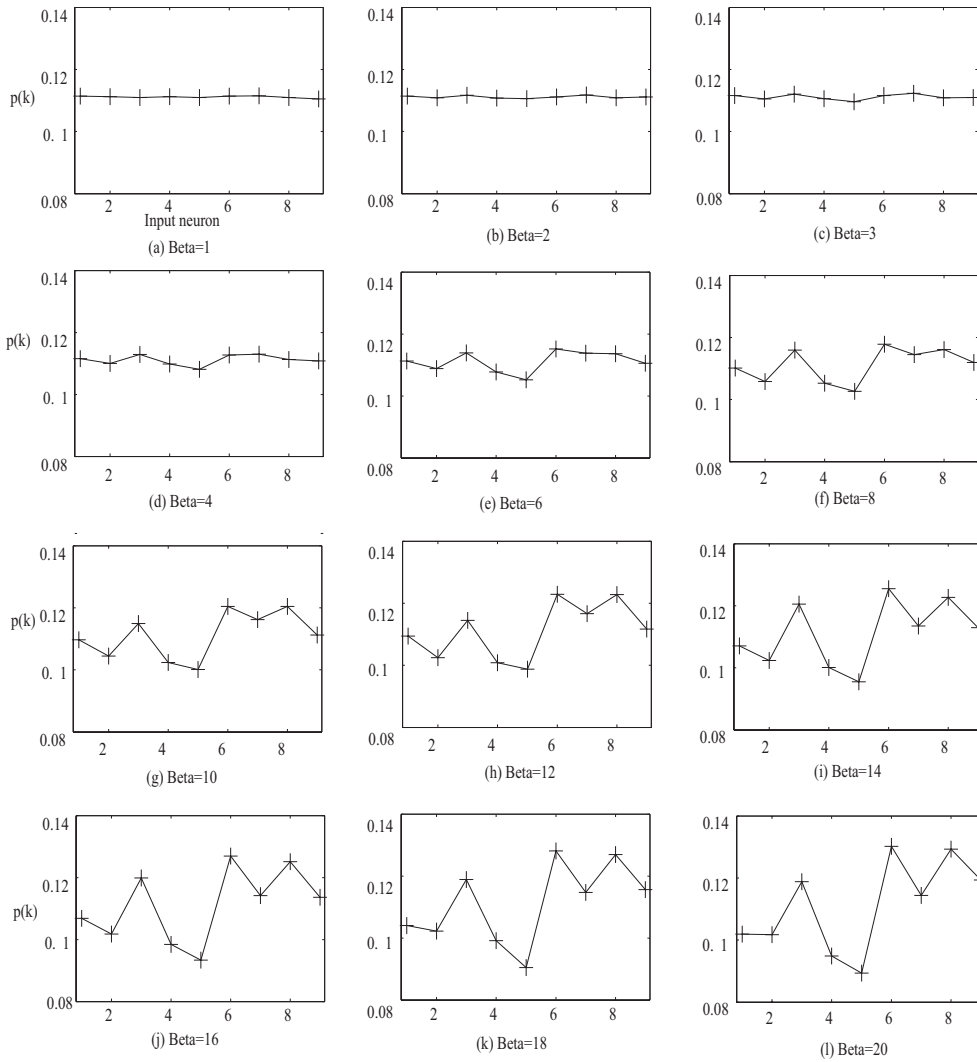


Fig. 2. Firing rates of input neurons by the information-theoretic method with double competition when the parameter β was increased from one (a) to 20 (t) for the glass data.

topographic error increased less than by using the information-theoretic method without double competition. The behavior of the information-theoretic method without considering input neurons can be inferred from the free energy equation (18). By introducing the firing rates of input neurons, this tendency was attenuated. When using the information-theoretic method with double competition, the quantization and topographic errors did not increase or decrease to the extent observed when using the information-theoretic method without considering input neurons.

D. Discussion

1) *Validity of Methods and Experimental Results:* In this paper, we have proposed a new type of information-theoretic method which takes into account the firing rates of input neurons. We have so far shown that competitive learning as well as self-organizing maps aim to maximize mutual information between input patterns and output neurons [59], [60], [61]. However, little attention has been paid to information content in input neurons. In particular, we have not fully used any information on input neurons in learning processes. Thus, we

have introduced the firing rates of input neurons in the learning procedure of the self-organizing maps. We succeeded in determining the re-estimation formula for connection weights. We applied the method to two well-known data sets from the machine learning database, namely, the glass and dermatology data. In both data sets, we succeeded in extracting clearer class structure, particularly by detecting clear class boundaries for the both data sets.

In addition, we could see that quantization and topographic errors were inversely related when we used the method without considering input neurons. This inverse relation can be predicted by examining the free energy equation. The free energy equation in its expanded form appears as the following

$$F = \sum_{s=1}^S p(s) \sum_{j=1}^M p^*(j | s) \|\mathbf{x}^s - \mathbf{w}_j\|^2 + 2\sigma^2 \sum_{s=1}^S p(s) \sum_{j=1}^M p(j | s) \log \frac{p^*(j | s)}{q(j | s)}, \quad (18)$$

recalling that the spread parameter σ is defined by using the

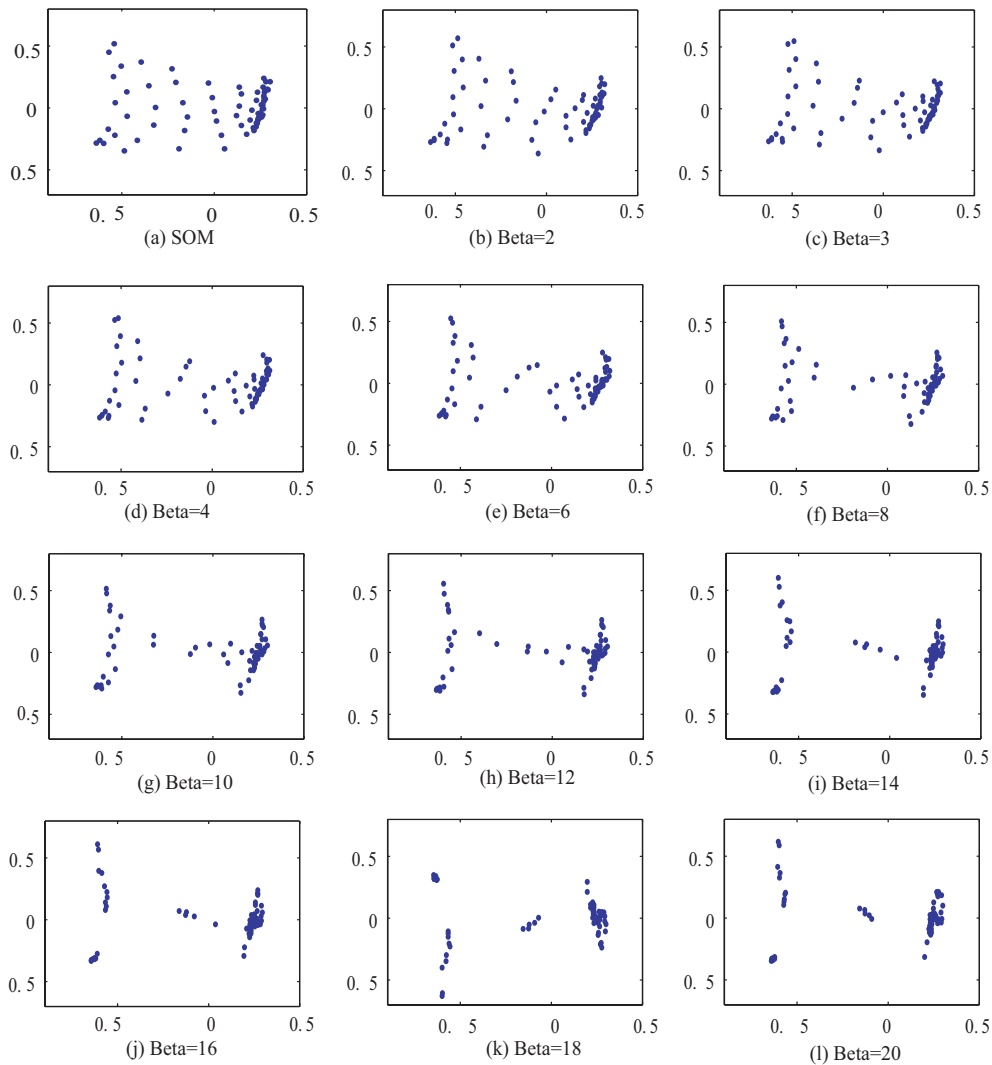


Fig. 3. Results by PCA for connection weights by double competition for the glass data.

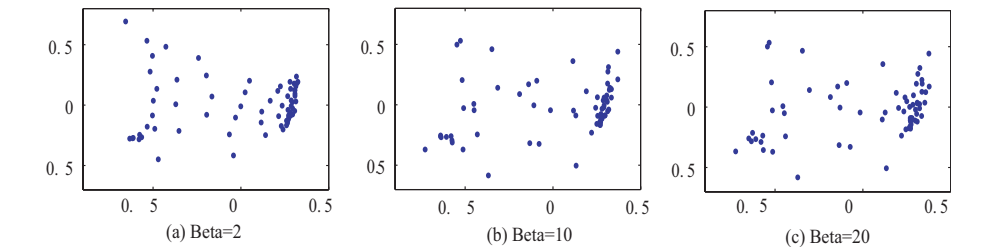


Fig. 4. Results by PCA for connection weights without considering input neurons for the glass data.

other parameter β .

$$\sigma = \frac{1}{\beta}. \quad (19)$$

When the parameter β was increased, and the spread parameter σ was decreased, the first term of the free energy became more effective. This means that quantization errors decreased, as shown in Figures 5 (a) and 9(a). On the other hand, when the parameter β is decreased and the spread parameter σ is increased, the effect of the second term of the free energy becomes dominant. The second term is the KL divergence is used to imitate the collective behavior of output neurons. Thus, when the parameter β is decreased, the topological

errors should decreased as well. This is shown in Figure 5(b) and 9(b). The introduction of input neuron firing rates in the learning processes attenuated this tendency.

2) *Problems of the Method:* There are two problems of this information-theoretic method, namely, the estimation of firing rates of input neurons and degradation in terms of quantization and topographic errors.

First, there is a problem with estimating the firing rates of input neurons, which must be computed in order to realize competitive processes. However, in the computation of competitive neurons, we must insert the firing rates of input neurons

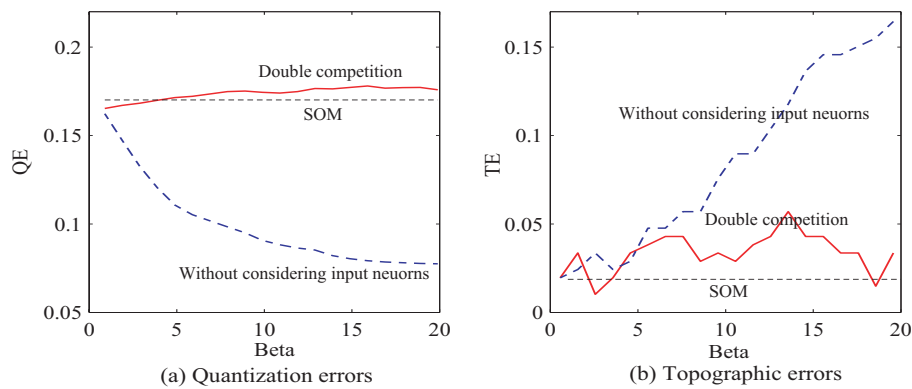


Fig. 5. Quantization and topographic errors by SOM in black, information-theoretic with double competition (in red) and without considering input neurons (in blue) for the glass data.

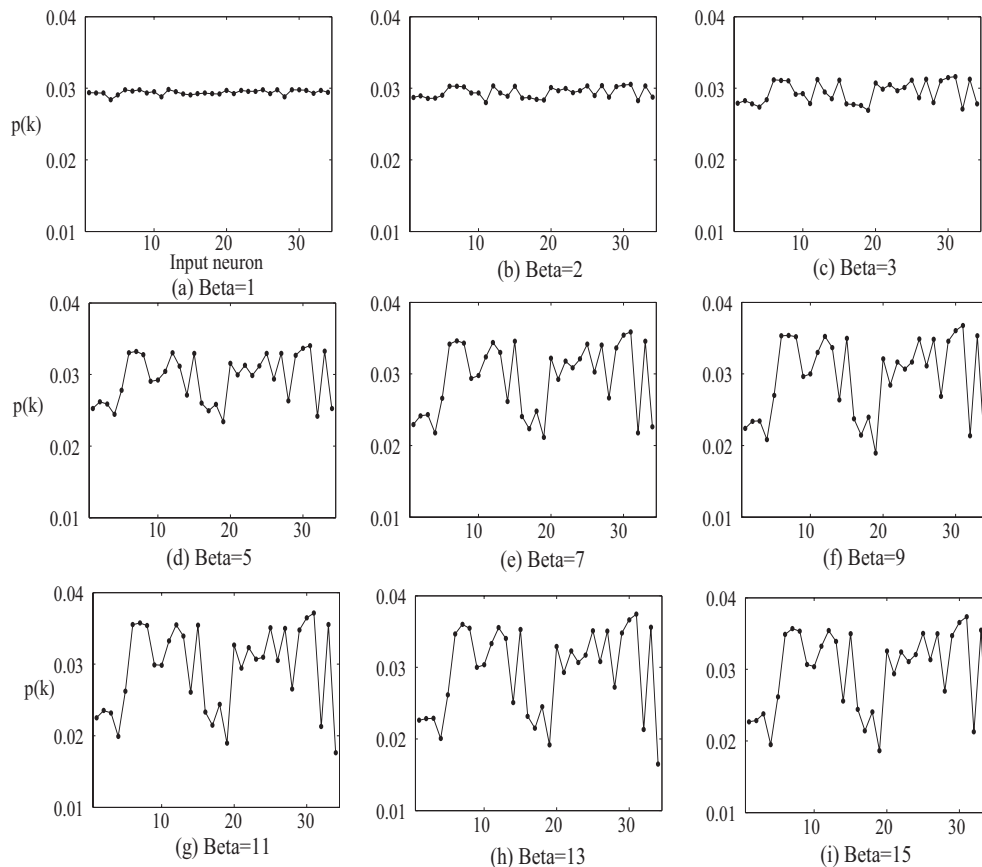


Fig. 6. Firing rates of input neurons when the parameter β was increased from one to 20 for the dermatology data.

into the equation (1). In Section II.C, we briefly presented how to estimate the firing rates of input neurons. However, in the estimation of the firing rates, we must insert the firing rates of competitive neurons into the equation (2). We should thus more carefully examine whether the firing rates of input neurons can be stabilized for the precise computation of the information content, and for producing stable self-organizing maps.

Second, we have a problem of degradation in terms of quantization and topographic errors. In Figures 5 and 9, quantization and topographic errors increased, though they did not reach extreme values as was the case with the method

without considering input neurons. We must explain why and how the degradation occurred and try to improve quantization and topographic errors.

3) *Possibilities of the Method:* The method presented in this paper can be considered as a new input variable selection in SOM, and opens up the possibility of having competition in all components of neural networks. First, this method is an extension of the self-organizing maps which takes into account the importance of input neurons or input variables. The competition between input neurons can be considered as the introduction of the importance of input variables in the self-organizing maps. As is well known, variable se-

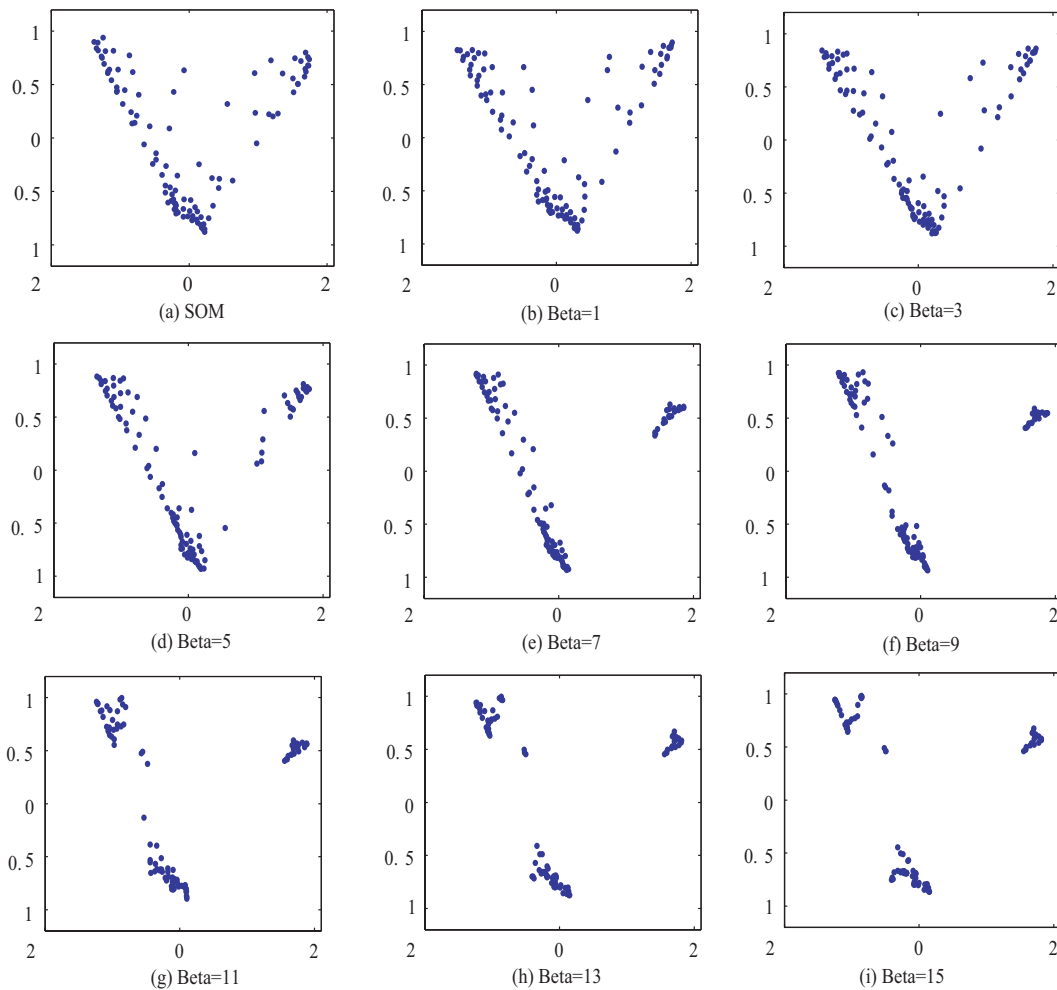


Fig. 7. Results by PCA for connection weights by double competition for the dermatology data.

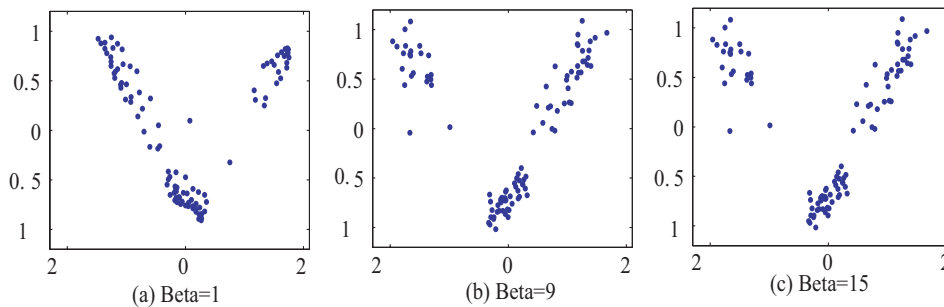


Fig. 8. Results of the PCA for connection weights without considering input neurons for the dermatology data.

lection has played important roles in learning, in particular in supervised learning [63], [64]. In unsupervised learning, such as SOM, the criteria to choose important variables have not been determined. However, in the information-theoretic method, the criteria to measure the importance of neurons is naturally introduced: the importance is measured in terms of information content in neurons. When this information increases, the importance of the neurons becomes larger. We use the importance of input neurons to visualize input patterns by SOM, as it plays an important role in this regard. Thus, it is important to examine relations between the importance of input neurons and the visualization of SOM.

Second, there is the possibility of having competition among all components in neural networks. In the present model of a neural network, in addition to input and output neurons, there are connection weights from the input neurons to output neurons. If it is possible to take into account the competition between all these connection weights, much better performance of a network can be expected. This means that in a neural network, every component competes with each other to most efficiently process outside stimuli.

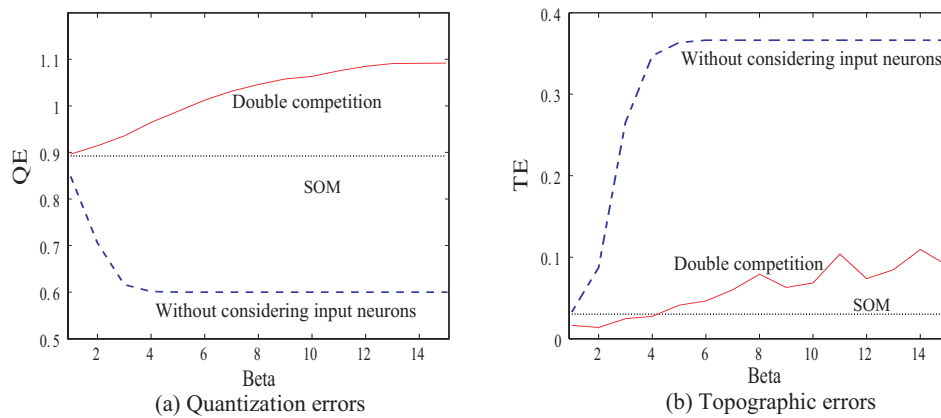


Fig. 9. Quantization and topographic errors by the SOM, and the information-theoretic method with double competition and without considering input neurons for the dermatology data.

IV. CONCLUSION

In this paper, we have introduced an information-theoretic method considering information in input neurons to realize competitive learning as well as the self-organizing maps. When mutual information is maximized between neurons and input patterns, just one neuron wins the competition. Namely, mutual information maximization corresponds to competitive learning. However, we can imagine that any component in a neural network should contain information on input patterns. Thus, we tried to take into account input neurons in addition to the output or competitive neurons usually used in competitive learning. We applied the information-theoretic method to the self-organizing maps by adding cooperation processes to competitive learning. Then, we applied the information-theoretic methods to two well-known data sets, namely, glass and dermatology data sets from the machine learning database. We found that by increasing information in input neurons, connection weights tended to be divided into clear groups. In addition, the inverse relation between quantization and topographic errors which was observed in the information-theoretic competitive learning without considering input neurons, was neutralized by considering these input neurons. However, quantization and topographic errors tended to degrade map quality when using the information-theoretic method. Thus, we should examine how and why this deterioration occurred in terms of quantization and topographic errors to realize the information-theoretic method with better quantization and visualization performance.

REFERENCES

- [1] A. Luk and S. Lien, "Properties of the generalized lotto-type competitive learning," in *Proceedings of International conference on neural information processing*, (San Mateo: CA), pp. 1180–1185, Morgan Kaufmann Publishers, 2000.
- [2] M. Van Hulle, "Topographic map formation by maximizing unconditional entropy: a plausible strategy for 'on-line' unsupervised competitive learning and nonparametric density estimation," *IEEE Transactions on Neural Networks*, vol. 7, no. 5, pp. 1299–1305, 1996.
- [3] M. M. V. Hulle, "The formation of topographic maps that maximize the average mutual information of the output responses to noiseless input signals," *Neural Computation*, vol. 9, no. 3, pp. 595–606, 1997.
- [4] M. M. Van Hulle, "Faithful representations with topographic maps," *Neural Networks*, vol. 12, no. 6, pp. 803–823, 1999.
- [5] M. M. Van Hulle, "Entropy-based kernel modeling for topographic map formation," *IEEE Transactions on Neural Networks*, vol. 15, no. 4, pp. 850–858, 2004.
- [6] I. Katsavounidis, C. C. J. Kuo, and Z. Zhang, "A new initialization technique for generalized lloyd iteration," *IEEE Signal Processing Letters*, vol. 1.
- [7] H. Xiong, M. N. S. Swamy, and M. O. Ahmad, "Competitive splitting for codebook initialization," *IEEE Signal Processing Letters*, vol. 11, pp. 474–477, 2004.
- [8] H. Xiong and M. N. S. Swamy, "Branching competitive learning network: a novel self-creating model," *IEEE Transactions on Neural Networks*, vol. 15, no. 2, pp. 417–429, 2004.
- [9] S. Wu, W. Liew, H. Yuan, and M. Yang, "Cluster analysis of gene expression data based on self-splitting and merging competitive learning," *IEEE transactions on information technology in biomedicine*, vol. 8, no. 1, pp. 5–15, 2004.
- [10] T. Kohonen, *Self-Organization and Associative Memory*. New York: Springer-Verlag, 1988.
- [11] T. Kohonen, *Self-Organizing Maps*. Springer-Verlag, 1995.
- [12] K. Kiviluoto, "Topology preservation in self-organizing maps," in *In Proceedings of the IEEE International Conference on Neural Networks*, pp. 294–299, 1996.
- [13] S. Kaski, J. Nikkila, M. Oja, J. Venna, P. Toronen, and E. Castren, "Trustworthiness and metrics in visualizing similarity of gene expression," *BMC Bioinformatics*, vol. 4, no. 48, 2003.
- [14] J. Venna and S. Kaski, "Neighborhood preservation in nonlinear projection methods: an experimental study," in *Lecture Notes in Computer Science*, vol. 2130, pp. 485–491, 2001.
- [15] G. Polzlbauer, "Survey and comparison of quality measures for self-organizing maps," in *Proceedings of the fifth workshop on Data Analysis (WDA04)*, pp. 67–82, 2004.
- [16] J. A. Lee and M. Verleysen, "Quality assessment of nonlinear dimensionality reduction based on K-ary neighborhoods," in *JMLR: Workshop and conference proceedings*, vol. 4, pp. 21–35, 2008.
- [17] T. Villmann, R. D. M. Herrmann, and T. Martinez, "Topology preservation in self-organizing feature maps: exact definition and measurement," *IEEE Transactions on Neural Networks*, vol. 8, no. 2, pp. 256–266, 1997.
- [18] H.-U. Bauer and K. Pawelzik, "Quantifying the neighborhood preservation of self-organizing maps," *IEEE Transactions on Neural Networks*, vol. 3, no. 4, pp. 570–578, 1992.
- [19] J. Vesanto, "SOM-based data visualization methods," *Intelligent Data Analysis*, vol. 3, pp. 111–126, 1999.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] S. Shieh and I. Liao, “A new approach for data clustering and visualization using self-organizing maps,” *Expert Systems with Applications*, 2012.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] L. Xu, Y. Xu, and T. W. Chow, “PolSOM—a new method for multidimensional data visualization,” *Pattern Recognition*, vol. 43, pp. 1668–1675, 2010.
- [28] 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.
- [29] 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.
- [30] E. Merényi, K. Tasdemir, and L. Zhang, “Learning highly structured manifolds: harnessing the power of soms,” in *Similarity-Based Clustering*, pp. 138–168, Springer, 2009.
- [31] K. Tasdemir and E. Merényi, “Exploiting data topology in visualization and clustering of self-organizing maps,” *Neural Networks, IEEE Transactions on*, vol. 20, no. 4, pp. 549–562, 2009.
- [32] R. Kamimura, “Self-enhancement learning: target-creating learning and its application to self-organizing maps,” *Biological cybernetics*, pp. 1–34, 2011.
- [33] R. Kamimura, “Constrained information maximization by free energy minimization,” *International Journal of General Systems*, vol. 40, no. 7, pp. 701–725, 2011.
- [34] R. Linsker, “Self-organization in a perceptual network,” *Computer*, vol. 21, pp. 105–117, 1988.
- [35] R. Linsker, “How to generate ordered maps by maximizing the mutual information between input and output,” *Neural Computation*, vol. 1, pp. 402–411, 1989.
- [36] R. Linsker, “Local synaptic rules suffice to maximize mutual information in a linear network,” *Neural Computation*, vol. 4, pp. 691–702, 1992.
- [37] R. Linsker, “Improved local learning rule for information maximization and related applications,” *Neural Networks*, vol. 18, pp. 261–265, 2005.
- [38] H. B. Barlow, “Unsupervised learning,” *Neural Computation*, vol. 1, pp. 295–311, 1989.
- [39] H. B. Barlow, T. P. Kaushal, and G. J. Mitchison, “Finding minimum entropy code,” *Neural Computation*, vol. 1, pp. 412–423, 1989.
- [40] G. Deco, W. Finnof, and H. G. Zimmermann, “Unsupervised mutual information criterion for elimination of overtraining in supervised multiplayer networks,” *Neural Computation*, vol. 7, pp. 86–107, 1995.
- [41] G. Deco and d. Obradovic, “Linear redundancy reduction learning,” *Neural Networks*, vol. 8, no. 5, pp. 751–755, 1995.
- [42] G. Deco and L. Parra, “Non-feature extraction by redundancy reduction in an unsupervised stochastic neural networks,” *Neural Networks*, vol. 10, no. 4, pp. 683–691, 1997.
- [43] R. Battiti, “Using mutual information for selecting features in supervised neural net learning,” *IEEE Transactions on Neural Networks*, vol. 5, no. 4, pp. 537–550, 1994.
- [44] J. W. Fisher and J. C. Principe, “A methodology for information theoretic feature extraction,” in *IEEE World congress on computational intelligence*, pp. 1712–1716, 1998.
- [45] K. Torkkola, “Nonlinear feature transform using maximum mutual information,” in *Proceedings of International Joint Conference on Neural Networks*, pp. 2756–2761, 2001.
- [46] K. Torkkola, “Feature extraction by non-parametric mutual information maximization,” *Journal of Machine Learning Research*, vol. 3, pp. 1415–1438, 2003.
- [47] V. Sindhvani, S. Raskshit, D. Deodhare, D. Erdogmus, and J. C. Principe, “Feature selection in mlps and svms based on maximum output information,” *IEEE Transactions on Neural Networks*, vol. 15, no. 4, pp. 937–948, 2004.
- [48] R. Kamimura, “Information-theoretic enhancement learning and its application to visualization of self-organizing maps,” *Neurocomputing*, vol. 73, no. 13–15, pp. 2642–2664, 2010.
- [49] R. Kamimura, “Double enhancement learning for explicit internal representations: unifying self-enhancement and information enhancement to incorporate information on input variables,” *Applied Intelligence*, pp. 1–23, 2011.
- [50] R. Kamimura, “Dependent input neuron selection in contradiction resolution,” in *Computational Intelligence & Applications (IWCIA), 2013 IEEE Sixth International Workshop on*, pp. 45–50, IEEE, 2013.
- [51] D. E. Rumelhart and D. Zipser, “Feature discovery by competitive learning,” *Cognitive Science*, vol. 9, pp. 75–112, 1985.
- [52] M. M. Van Hulle, “The formation of topographic maps that maximize the average mutual information of the output responses to noiseless input signals,” *Neural Computation*, vol. 9, no. 3, pp. 595–606, 1997.
- [53] M. M. Van Hulle, “Topology-preserving map formation achieved with a purely local unsupervised competitive learning rule,” *Neural Networks*, vol. 10, no. 3, pp. 431–446, 1997.
- [54] S. C. Ahalt, A. K. Krishnamurthy, P. Chen, and D. E. Melton, “Competitive learning algorithms for vector quantization,” *Neural Networks*, vol. 3, pp. 277–290, 1990.
- [55] L. Xu, “Rival penalized competitive learning for clustering analysis, RBF net, and curve detection,” *IEEE Transaction on Neural Networks*, vol. 4, no. 4, pp. 636–649, 1993.
- [56] Y. J. Zhang and Z. Q. Liu, “Self-splitting competitive learning: a new on-line clustering paradigm,” *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 369–380, 2002.
- [57] J. C. Yen, J. I. Guo, and H. C. Chen, “A new k-winners-take-all neural networks and its array architecture,” *IEEE Transactions on Neural Networks*, vol. 9, no. 5, pp. 901–912, 1998.
- [58] S. Ridella, S. Rovetta, and R. Zunino, “K-winner machines for pattern classification,” *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 371–385, 2001.
- [59] R. Kamimura, T. Kamimura, and T. R. Shultz, “Information theoretic competitive learning and linguistic rule acquisition,” *Transactions of the Japanese Society for Artificial Intelligence*, vol. 16, no. 2, pp. 287–298, 2001.
- [60] R. Kamimura, T. Kamimura, and O. Uchida, “Flexible feature discovery and structural information control,” *Connection Science*, vol. 13, no. 4, pp. 323–347, 2001.
- [61] R. Kamimura, “Information-theoretic competitive learning with inverse Euclidean distance output units,” *Neural Processing Letters*, vol. 18, pp. 163–184, 2003.
- [62] K. Bache and M. Lichman, “UCI machine learning repository,” 2013.
- [63] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.
- [64] E. P. Xing and R. M. Karp, “Cliff: clustering of high-dimensional microarray data via iterative feature filtering using normalized cuts,” *Bioinformatics*, vol. 17, no. suppl 1, pp. S306–S315, 2001.

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 main research interest is information-theoretic approaches to neural computing.