Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Structural Enhanced Information to Detect Features in Competitive Learning

Ryotaro Kamimura
IT Education Center, Tokai University,
1117 Kitakaname Hiratsuka Kanagawa 259-1292, Japan
ryo@cc.u-tokai.ac.jp

***Abstract*— In this paper, we propose structural enhanced information for detecting and visualizing main features in input patterns. We have so far proposed information enhancement for feature detection, where, if we want to focus upon components such as units and connection weights and interpret the functions of the components, we have only to enhance competitive units with the components. Though this information enhancement has given favorable results in feature detection, we further refine the information enhancement and propose structural enhanced information. In structural enhanced information, three types of enhanced information can be differentiated, that is, first-, second- and third-order enhanced information. The first-order information is related to the enhancement of competitive units themselves in a competitive network, and the second-order information is dependent upon the enhancement of competitive units with input patterns. Then, the third-order information is obtained by subtracting the effect of the first-order information from the second-order information. Thus, the third-order information more explicitly represents information on input patterns. With this structural enhanced information, we can estimate more detailed features in input patterns. For demonstrating explicitly and intuitively the improved performance of our method, the conventional SOM was used, and we transformed competitive unit outputs so as to improve visualization. The method was applied to the Johns Hopkins University Ionosphere database. In the problem, we succeeded in visualizing the detailed and important features of input patterns by using the third-order information.**

I. Introduction

A. Information Enhancement

Mutual information has been widely used in neural computing, because it can deal with higher order statistics [1] and it can be related to information processing principles in living systems [2], [3], [4], [5], [6]. However, one of the major problems of mutual information is that it is the average of information content in all components. Thus, if we want to extract information content about specific components in a network, it is of no use to examine mutual information. To extract information specific to input variables, we have proposed conditional mutual information and a method called *information loss* [7], [8]. In conditional mutual information [9], [10], the average operation in mutual information is partially removed to obtain information specific to some components. The same type of conditional information has also been applied to information processing describing the dendrite [11]. On the other hand, the information loss has been defined by the change of mutual information between a full network and a network without a certain element [9]. If a large change in mutual information, in particular, large information loss, is observed, then the element surely plays a very important role in a network. We applied the information loss to actual problems and confirmed that the information loss is a good indicator to suggest the possibility of the importance of elements in a network [10].

Parallel to this information loss, we have introduced *information enhancement* of competitive units by enhancement and relaxation [12] to extract information specific to some components. In the information enhancement, competitive units are forced to respond to input patterns by enhancing some elements, such as input variables, in a network. With this enhancement, competitive units tend to respond explicitly to input patterns, and it becomes easier to interpret how competitive units respond to input patterns. This enhancement is certainly related to the well-known *attention*. For example, visual attention has been believed to speed up the reaction time and to improve visual perception [13], and there have been a number of models for modeling attention [13], [14], [15], [16].

For the procedure of enhancement, we have introduced *enhancement* and *relaxation* for competitive units. When a competitive unit is enhanced, it responds explicitly to all the input patterns, while when the competitive unit is relaxed, it responds uniformly to all the input patterns. Information change by enhancement or relaxation is called *enhanced information*. The enhancement is realized by changing the Gaussian width in computing outputs from neurons. As the Gaussian width becomes smaller, more explicit responses of neurons can be obtained. On the other hand, as the Gaussian width becomes larger, all neurons respond more equally to any input patterns. Because enhancement can be realized by decreasing the Gaussian width, it turns out to be easy to implement a concept of enhancement in neural networks.

B. Structural Information Enhancement

Enhanced information has been proved to be a powerful method to detect features [12]. However, we have found that, if we consider several types of enhanced information, we can more clearly detect important features in input patterns. In this context, we introduce structural enhanced information to take into account several types of information for obtaining more detailed information on input patterns. We have already introduced structural information [17], [18], [19]. In this structural information, we have two types of information content,

978-1-4244-2794-9/09/$25.00

namely, the first- and the second-order information. The first-order information is due only to the probability of the competitive units themselves, and the second-order information is information content due to given input patterns. By separating these two types of information, we have shown that more detailed interpretation is possible for a network. In addition, Gatlin [20] found that the separation of different types of information can be used to differentiate the different species of living systems. Following these experimental results showing the effectiveness of different types of information content, we introduce here structural enhanced information in which we can differentiate three types of information; the first-, second- and third-order enhanced information. The first-order information is an information change due only to competitive units. On the other hand, the second-order information is an information change due to competitive units with input patterns. The third-order information is obtained by subtracting the effect of the first-order information from the second-order information. The third-order information is used to produce information without the effect of the first-order information. Because we take into account the structure of information content, this information is called *structural enhanced information*. Though all types of information play important roles in interpreting final representations, we focus upon the third-order information in this paper and in particular focus upon competitive units for enhancement, because this type of third-order information has shown significantly improved performance for many problems.

Finally, we should briefly note the applications used in this paper. Structural enhanced information has been based upon competitive learning, and thus it is the best suited for interpreting internal representations created by competitive learning. However, we applied the method here to the conventional self-organizing maps or SOM, because it is easy to demonstrate how well the new method can improve the visualization performance of the SOM. In the conventional SOM, many techniques have bee developed, such as Sammon maps [21] and U-matrix [22] and some other techniques discussed in many pieces of literature [23], [24], [25], [26], [27]. However, these sophisticated techniques may be of no use for ambiguous connection weights. Our method can be used to visualize connection weights more explicitly.

II. Theory and Computational Methods

A. Mutual Information and Competitive Learning

Competitive learning [28] has been considered to be one of the most important learning methods in neural networks. Competitive learning has been developed by Rumelhart and Zipser [28] to detect regularity in input patterns, and the competitive units should equally be activated. However, as already mentioned in the introduction section, we have had the underutilized problem, and many methods have been proposed to overcome this problem [28], [29], [30], [31], [32], [33].

To overcome this problem and to realize the softer type of competitive learning, we have introduced mutual information between input patterns and competitive units in competitive

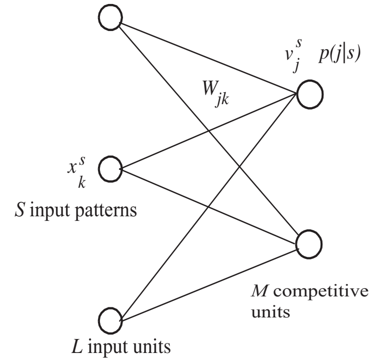


Fig. 1. A network architecture for competitive learning.

learning [18], [34], [35], [36], [37]. The use of mutual information for competitive learning is due to the fact that there is a strong similarity between competitive learning and mutual information maximization. As mutual information is maximized, just one neuron's activation is maximized, while all the other neurons are inactive. In addition, on average, all neurons are equally activated.

Now, let us compute mutual information for a network. Figure 1 shows a network architecture. A network is composed of the kth input unit x_k^s for the sth input pattern, and connection weights w_{jk} from the kth input unit to the jth competitive unit. The jth competitive unit receives a net input from input units, and an output v_j^s from the jth competitive unit can be computed by

$$v_j^s \propto \exp\left\{-\frac{1}{2}(\mathbf{x}^s - \mathbf{w}_j)^T \mathbf{\Sigma}^{-1}(\mathbf{x}^s - \mathbf{w}_j)\right\}, \quad (1)$$

where

$$\mathbf{\Sigma}_j = \sigma^2 \mathbf{I}, \quad (2)$$

where $\mathbf{\Sigma}$ represents a scaling matrix[1], and $\mathbf{I}$ denotes an identity matrix. The output is increased when connection weights are closer to input patterns. The conditional probability of the firing of the jth competitive unit, given the sth input pattern of the S input patterns, can be obtained by

$$p(j \mid s) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{x}^s - \mathbf{w}_j)^T \mathbf{\Sigma}^{-1}(\mathbf{x}^s - \mathbf{w}_j)\right\}}{\sum_{m=1}^{M} \exp\left\{-\frac{1}{2}(\mathbf{x}^s - \mathbf{w}_m)^T \mathbf{\Sigma}_j^{-1}(\mathbf{x}^s - \mathbf{w}_m)\right\}}. \quad (3)$$

The probability of the firing of the jth competitive unit is computed by

$$p(j) = \sum_{s=1}^{S} p(s) p(j \mid s), \quad (4)$$

where $p(s)$ denotes the probability of the sth input pattern and S is the number of input patterns. With these probabilities, we can compute mutual information

$$I = \sum_{s=1}^{S}\sum_{j=1}^{M} p(s) p(j \mid s) \log \frac{p(j \mid s)}{p(j)}, \quad (5)$$

[1]Instead of the covariance matrix, we used this term, because the competitive units do not follow the Gaussian distribution.

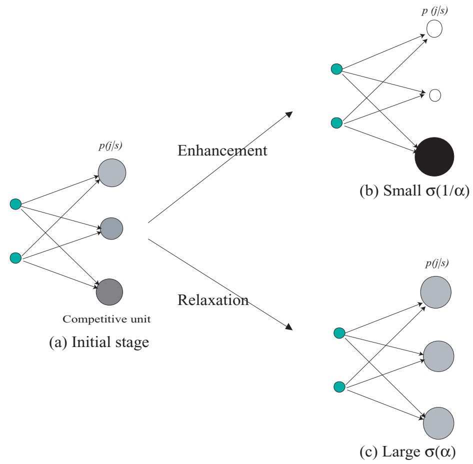


Fig. 2. Enhancement (b) and relaxation (c) of competitive units for a network itself.

where M denotes the number of competitive units. When this mutual information is maximized, the first term of entropy should be maximized, while the second term of conditional entropy should be minimized. When the entropy is maximized, all neurons are equally activated. On the other hand, when the conditional entropy is minimized, a neuron is only activated for specific input patterns. This finding of similarity between competitive learning and mutual information encourages us to use information-theoretic measures to describe competitive learning.

B. General Enhancement and Relaxation

Because mutual information is the average, it is of no use to extract information content specific to some components in a network. To overcome this problem, we have introduced a procedure of enhancement and relaxation to focus upon some components in a network [12]. These procedures are based upon the change in competitive unit activations by controlling the Gaussian width. We introduce here the most general case of enhancement and relaxation, namely, enhancement and relaxation for a network itself.

Figure 2 shows a process of enhancement and relaxation for a network itself. Figure 2(a) shows the original situation obtained by competitive learning, in which three neurons are differently activated for input units. With enhancement, as shown in Figure 2(b), the characteristics of competitive unit activations are enhanced, and only one competitive unit is strongly activated. This means that obtained information in competitive units is larger. On the other hand, Figure 2(c) shows a state obtained by relaxation, in which all competitive units respond uniformly to input units. Because competitive units cannot differentiate between input patterns, no information on input patterns is stored.

This enhancement can be realized by changing the Gaussian width. In a network shown in Figure 1, a neuron output can be defined by the Gaussian-like function:

$$v_{jk}^s \propto \exp\left\{-\frac{1}{2}(\mathbf{x}^s - \mathbf{w}_j)^T \mathbf{\Sigma}^{-1}(\mathbf{x}^s - \mathbf{w}_j)\right\}, \quad (6)$$

where $\mathbf{\Sigma}$ denotes the scaling matrix. Thus, enhancement is possible by decreasing the Gaussian width. In actual formulation in Section 2.4, the Gaussian width is replaced by the enhancement parameter α. The scaling matrix can be defined by

$$\begin{aligned} \mathbf{\Sigma} &= \mathbf{\Sigma}_{enh}^{(\alpha)} \\ &= \sigma^2 \mathbf{I} \\ &= (1/\alpha)^2 \mathbf{I}. \end{aligned} \quad (7)$$

When the relaxation is applied, we have

$$\begin{aligned} \mathbf{\Sigma} &= \mathbf{\Sigma}_{rel}^{(\alpha)} \\ &= (\alpha)^2 \mathbf{I}. \end{aligned} \quad (8)$$

For example, when a network itself is enhanced, we must compute the competitive outputs by

$$v_{j,enh}^s \propto \exp\left\{-\frac{1}{2}(\mathbf{x}^s - \mathbf{w}_j)^T (\mathbf{\Sigma}_{enh}^{(\alpha)})^{-1}(\mathbf{x}^s - \mathbf{w}_j)\right\}. \quad (9)$$

As the Gaussian width is decreased, the enhancement parameter is increased. This means that, as the enhancement parameter α is increased, more enhanced states can be produced.

C. Structural Enhanced Information

In this paper, enhancement is measured in three different ways in order to see a more detailed change in neurons. For this purpose, we introduce structural enhanced information, in which we can define three types of enhanced information. First, the first-order enhanced information is defined by the cross entropy between the original probability $p(j)$ of firing of the jth competitive unit and the enhanced probability for competitive units. On the other hand, the second-order enhanced information is defined by the cross entropy between the original conditional probability $p(j \mid s)$ of competitive units for input patterns and the enhanced conditional probability. Finally, the third-order enhanced information can be defined by the difference between the second- and first-order enhanced information.

Let us introduce structural enhanced information. Suppose that the probability $p(j)$ of firing of the jth neuron is changed to $p^{enh}(j)$ by enhancement, and the conditional probability $p(j \mid s)$ to $p^{enh}(j \mid s)$. The first-order enhanced information for the probability $p(j)$ can be defined by using the cross entropy:

$$EI_1 = \sum_{j=1}^{M} p(j) \log \frac{p(j)}{p^{enh}(j)}. \quad (10)$$

The second-order enhanced information is defined by

$$EI_2 = \sum_{s=1}^{S} p(s) \sum_{j=1}^{M} p(j \mid s) \log \frac{p(j \mid s)}{p^{enh}(j \mid s)}. \quad (11)$$

Then, we have the third-order enhanced information

$$EI_3 = EI_2 - EI_1. \quad (12)$$

This third-order enhanced information is formulated by our expectation of the second-order enhanced information's being much larger than the first-order information. The first-order enhanced information is concerned with information change due to competitive units themselves. On the other hand, the second-order enhanced information is related to information change due to competitive units with input patterns. The third-order enhanced information is information change due to competitive units with input patterns without the effect due to the competitive units themselves.

D. Structural Enhanced Information for Competitive Units

Then, we consider a case where a competitive unit itself is used for enhancement, because we have had the most favorable experimental results with this enhancement. However, enhancement can be defined for any components of a network. For a more detailed discussion on the flexibility of the enhancement, see the discussion section. Competitive unit outputs when the rth competitive unit is used for enhancement are given by

$$v_{jr}^s \propto \exp\left\{-\frac{1}{2}(\mathbf{x}^s - \mathbf{w}_j)^T(\boldsymbol{\Sigma}_j^{(r,\alpha)})^{-1}(\mathbf{x}^s - \mathbf{w}_j)\right\}, \quad (13)$$

where

$$\boldsymbol{\Sigma}_j^{(r,\alpha)} = \delta_{jr}\boldsymbol{\Sigma}_{enh}^{(\alpha)} + (1-\delta_{jr})\boldsymbol{\Sigma}_{rel}^{(\alpha)}. \quad (14)$$

The scaling matrix means that, if the target is the rth competitive unit, the enhanced scaling matrix is used for the competitive unit, while for all the other competitive units, the relaxed scaling matrix is used. We can normalize these outputs for probabilities,

$$p^r(j \mid s) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{x}^s - \mathbf{w}_j)^T(\boldsymbol{\Sigma}_j^{(r,\alpha)})^{-1}(\mathbf{x}^s - \mathbf{w}_j)\right\}}{\sum_{m=1}^{M}\exp\left\{-\frac{1}{2}(\mathbf{x}^s - \mathbf{w}_m)^T(\boldsymbol{\Sigma}_j^{(r,\alpha)})^{-1}(\mathbf{x}^s - \mathbf{w}_m)\right\}}. \quad (15)$$

And we have

$$p^r(j) = \sum_{s=1}^{S} p(s)p^r(j \mid s). \quad (16)$$

Now, we can define the first-order enhanced information

$$EI_1^r = \sum_{j=1}^{M} p(j)\log\frac{p(j)}{p^r(j)}. \quad (17)$$

By using these probabilities, we have the second-order enhanced information for the rth competitive unit

$$EI_2^r = \sum_{s=1}^{S}\sum_{j=1}^{M} p(s)p(j \mid s)\log\frac{p(j \mid s)}{p^r(j \mid s)}. \quad (18)$$

The third-order enhanced information is defined by the difference between the two pieces of information, defined by

$$EI_3^{rt} = EI_2^{rt} - EI_1^{rt}. \quad (19)$$

The third-order enhanced information is the difference between the first- and the second-order enhanced information. Therefore, if the second-order enhanced information is overwhelmingly larger than the first, then the third-order enhanced information is of no use.

III. Results and Discussion

In the following experiments, we try to show how well the new method extracts features in input patterns. More exactly, the third-order information is effective in detecting features, because the third-order information is obtained by subtracting the effect of the first-order information from the second-order information. For easy comparison, we use the conventional SOM [24]. The data in the following experiments were normalized, and the variance was unity. For comparison, we plotted the Sammon maps frequently used in the SOM applications and PCA projections with two principal components. All these methods, including the Sammon maps and PCA, were executed by using the SOM toolbox with all default values for easy comparison and reproduction[2].

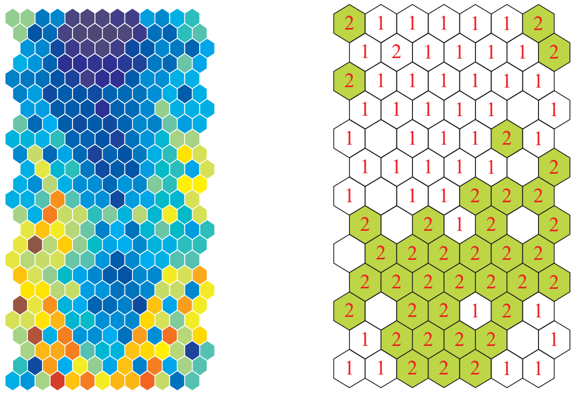
A. Ionosphere Database

We used the Johns Hopkins University Ionosphere database in UCI machine learning database[3]. The number of input patterns was 351, and the number of variables was 34. The data were normalized with zero mean and one variance. The data were described by the UCI machine learning database as follows. The data were collected by a system consisting of a phased array of 16 high-frequency antennas with a total transmitted power in the order of 6.4 kilowatts. The targets were free electrons in the ionosphere. The "good" and "bad" radar returns were those with and without evidence, respectively, of some type of structure in the ionosphere.

Figures 3(a) and (b) show a U-matrix and a map with labels obtained for the ionosphere database by the conventional SOM. In Figure 3(a), boundaries in warmer colors concentrate on the lower part of the map, and those boundaries are not necessarily those between two groups. On the other hand, by plotting labels on the map, in Figure 3(b), input patterns are clearly divided into two groups, namely, "good" and "bad" radar returns. Figure 4 shows enhanced information when the enhancement parameter is increased from 1.1 (a) to 2.0 (c). On the whole, as the order of information from the first to the third order is increased, the enhanced information of some neurons on the lower end of the map becomes higher. When the enhancement parameter is 1.1, in Figure 4(a), when the order of information is higher, the enhanced information of the lower part of the map is higher, in warmer colors, but the difference among the first- (a1), second- (a2) and third-order enhanced information (a3) is small. When the enhancement parameter is increased from 1.1 (a) to 1.5 (b), on the lower parts of the maps in the second- and the third-order information in Figure 4(b2)

[2]We used SOM Toolbox 2.0, February 11th, 2000 by Juha Vesanto http://www.cis.hut.fi/projects/somtoolbox/. No special options were used for easy reproduction.

[3]http://www1.ics.uci.edu/ mlearn/MLRepository.html



(a) U-matrix (b) Labels

Fig. 3. U-matrix (a) and labels (b) produced by the conventional SOM.

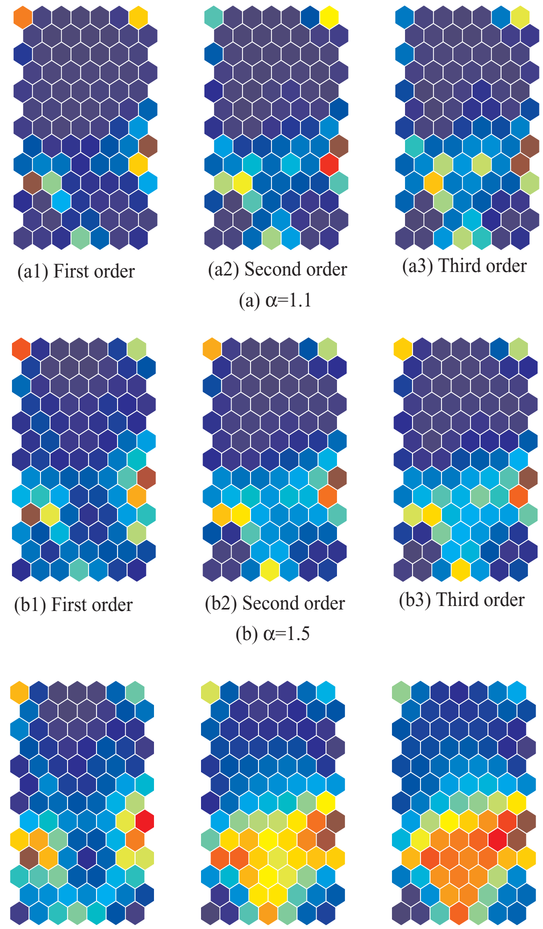
and (b3), the enhanced information becomes slightly higher, in lighter blue. Finally, when the enhancement parameter is increased to 2.0, in Figure 4(c), an upper region and a lower region for the second and third-order enhanced information in Figure 4(c2) and (c3) are more clearly separated. In particular, for the third-order information in Figure 4(c3), the lower part of the map significantly corresponds to the second group in the map with labels in Figure 3(b).

Finally, we compare the results obtained by the structural enhanced information with those by the Sammon map and the principal component analysis. Figures 4 (a) and (b) show a Sammon map and a PCA projection with two principal components for the ionosphere database. No clear boundary separating two groups can be detected by the Sammon map in Figure 5. On the other hand, in Figure 5(b), the results of the principal component analysis seem to have produced some clusters of input patterns on the left and right side of the map. However, we can say that patterns detected by the principal component analysis are more ambiguous than those obtained by the structural enhanced information. In addition, in the principal component analysis, the cumulative contribution rate with the first two principal components is only 0.396. Thus, with two principal components, it is difficult to explain input patterns.

These experimental results show that the third-order enhanced information can most clearly divide input patterns into two groups, while all the other methods failed to divide the input into appropriate groups.

B. Discussion

We have proposed a new information-theoretic method called *structural enhanced information* to detect features or to improve visualization performance in the conventional SOM. Information enhancement is a procedure to enhance competitive units and to force competitive units to respond explicitly to input patterns. We have so far defined and used enhanced information for feature detection for several problems [12]. However, we have found that more detailed structure inside information content reveals some important features in input patterns [18]. For this purpose, we have introduced structural



(a1) First order (a2) Second order (a3) Third order

(a) α=1.1

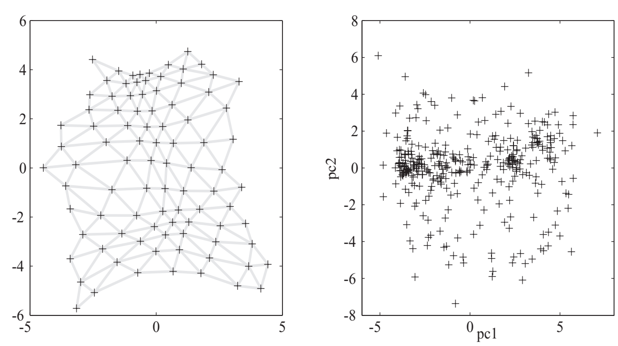
(b1) First order (b2) Second order (b3) Third order

(b) α=1.5

(c1) First order (c2) Second order (c3) Third order

(c) α=2.0

Fig. 4. The first-order (1), second-order (2) and third-order (3) enhanced information for competitive units when the enhancement parameter is increased from 1.1 (a) to two (c). Warmer and cooler colors represent larger and smaller values, respectively.



(a) Sammon map (b) PCAprojections

Fig. 5. Sammon map (a), PCA projection (b) (cumulative contribution ratio=0.396).

enhanced information in which the first-, second- and third-order enhanced information can be differentiated. The first-order enhanced information is due to the information by competitive units themselves, while the second-order information contains information on input patterns. Thus, the third-order enhanced information is the difference between the second- and first-order information. The third-order information, therefore, represents information exclusively on input patterns. In this paper, we have focused upon the third-order enhanced information. We have applied the third-order enhanced information to the ionosphere database. In the data, the third-order structural information has shown more explicit maps than the U-matrices by the conventional SOM, the Sammon maps and the principal component analysis. Thus, experimental results have shown that, in addition to the first- and second-order information, we need to examine the third-order information in complex problems.

The limitations of this method can be summarized by two points: parameter setting and computational complexity. The first problem is related to the enhancement parameter. As already discussed in Section 3, the difference between the first- and the second-order information becomes larger as the enhancement parameter becomes larger. When the difference is larger, in particular, when the second-order information is much larger than the first-order information, the difference between the two types of information is of no use, because the second-order information becomes dominant. Thus, we need extensive studies on tuning the enhancement parameter α.

Then, the time complexity becomes serious when we must deal with larger problems, because for each competitive unit to be enhanced, we need to compute mutual information. Thus, for the case of this paper, we had to repeat the computation of mutual information M times (M is the number of competitive units). This can be a time-consuming procedure when the number of competitive units is larger. Enhanced information is flexible enough to be applied to any components and any combination of the components. For example, when we try to examine relations between L input units and M competitive units, the complexity becomes more serious (computation of mutual information by $M \times L$ times). We need to develop a computational method to simplify the computation of mutual information.

We have focused upon one type of structural information in which specific competitive units are directly enhanced by the enhancement parameter. However, we can use any components, such as input, competitive units and their combination, in a network for this kind of enhancement. For example, we use input units for enhancement, and we can estimate the importance of input variables. In this way, we can define enhanced information for any components and any combination of these components in a network. Thus, though our method has some problems, such as parameter setting and time complexity, we think that the flexibility of the method seems to surpass those problems.

IV. Conclusion

In this paper, we have introduced a new type of information called *enhanced information* as well as *structural enhanced information*. Enhanced information is obtained by enhancing competitive units through some elements in a network, while all the other competitive units are forced to be relaxed. If this enhancement causes a drastic change in information for competitive units, the elements surely play a very important role in information processing in competitive learning. To see the role of this enhanced information in more detail, we have introduced structural enhanced information, with three types of information, that is, first-, second- and third-order enhanced information. The first-order enhanced information consists of change only in competitive units. The second-order enhanced information consists of change in competitive units, given input patterns. The third-order enhanced information is the difference between the first- and the second-order enhanced information. We have applied the method to an ionosphere database to show easily how well the new method discovers features in input patterns and how well it visualizes complex input patterns. Experimental results have shown that features extracted by the new method are clearer than those extracted by the conventional SOM, and results correspond to our intuition on input patterns. Though much effort remains for the method to be made practically applicable, it is certain that the flexibility of the method can be used to interpret the functions of specific components in neural networks.

Acknowledgment

The author is very grateful to Mitali Das for their valuable comments and suggestions.

References

[1] K. Torkkola. Feature extraction by non-parametric mutual information maximization. *Journal of Machine Learning Research*, 3:1415–1438, 2003.

[2] R. Linsker. Self-organization in a perceptual network. *Computer*, 21:105–117, 1988.

[3] R. Linsker. How to generate ordered maps by maximizing the mutual information between input and output. *Neural Computation*, 1:402–411, 1989.

[4] R. Linsker. Local synaptic rules suffice to maximize mutual information in a linear network. *Neural Computation*, 4:691–702, 1992.

[5] H. B. Barlow. Unsupervised learning. *Neural Computation*, 1:295–311, 1989.

[6] H. B. Barlow, T. P. Kaushal, and G. J. Mitchison. Finding minimum entropy code. *Neural Computation*, 1:412–423, 1989.

[7] Ryotaro Kamimura. Feature detection and information loss in competitive learning. In *Proceedings of SCIS and ISIS*, pages 1144–1148, 2008.

[8] R. Kamimura. An information-theoretic approach to feature extraction in competitive learning (to appear). *Neurocomputing*, 2008.

[9] Ryotaro Kamimura. Conditional information and information loss for flexible feature extraction. In *Proceedings of the international joint conference on neural networks(IJCNN2008)*, pages 2047–2083, 2008.

[10] Ryotaro Kamimura. Feature detection and information loss in competitive learning. In *Proceedings of the international conference on soft computing and intelligent systems and the international symposium on advanced intelligent systems(SCIS and ISIS2008)*, pages 1144–1148, 2008.

[11] K. N. Gurney. Information processing in dendrites: Ii information theoretic complexity. *Neural Networks*, 14.

[12] Ryotaro Kamimura. Feature discovery by enhancement and relaxation of competitive units. In *Intelligent data engineering and automated learning-IDEAL2008(LNCS)*, volume LNCS5326, pages 148–155. Springer, 2008.

[13] F. H. Hamker and M. Zirnsak. V4 receptive field dynamics as predicted by a systems-level model of visual attention using feedback from the frontal eye field. *Neural Networks*, 19:1371–1382, 2006.

[14] E. T. Rolls and G. Deco. Attention in natural scenes; neurophysiological and computational bases. *Neural Networks*, 19:1383–1394, 2006.

[15] L. J. Lanyon and S. L. Denham. A model of active visual search with object-based attention guiding scan paths. *Neural Networks*, 19:873–897, 2006.

[16] N. J. H. Korsten, N. Fragopanagos, M. Hartle, N. Taylor, and J. G. Taylor. Attention as a controller. *Neural Networks*, 19:1408–1421, 2006.

[17] Ryotaro Kamimura and Taeko Kamimura. Structural information and linguistic rule extraction. In *Proceedings of ICONIP*, pages 720–726, 2000.

[18] Ryotaro Kamimura, Taeko Kamimura, and Osamu Uchida. Flexible feature discovery and structural information. *Connection Science*, 13(4):323–347, 2001.

[19] R. Kamimura, T. Kamimura, and O. Uchida. Flexible feature discovery and structural information control. *Connection Science*, 13(4):323–347, 2001.

[20] L. L. Gatlin. *Information Theory and Living Systems*. Columbia University Press, 1972.

[21] J. W. Sammon. A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers*, C-18(5):401–409, 1969.

[22] A. Ultsch and H. P. Siemon. Kohonen self-organization feature maps for exploratory data analysis. In *Proceedings of International Neural Network Conference*, pages 305–308, Dordrecht, 1990. Kulwere Academic Publisher.

[23] T. Kohonen. *Self-Organization and Associative Memory*. Springer-Verlag, New York, 1988.

[24] T. Kohonen. *Self-Organizing Maps*. Springer-Verlag, 1995.

[25] J. Versanto. SOM-based data visualization methods. *Intelligent Data Analysis*, 3:111–126, 1999.

[26] 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.

[27] I. Mao and A. K. Jain. Artificial neural networks for feature extraction and multivariate data projection. *IEEE Transactions on Neural Networks*, 6(2):296–317, 1995.

[28] D. E. Rumelhart and D. Zipser. Feature discovery by competitive learning. *Cognitive Science*, 9:75–112, 1985.

[29] D. DeSieno. Adding a conscience to competitive learning. In *Proceedings of IEEE International Conference on Neural Networks*, pages 117–124, San Diego, 1988. IEEE.

[30] Lei Xu. Rival penalized competitive learning for clustering analysis, RBF net, and curve detection. *IEEE Transaction on Neural Networks*, 4(4):636–649, 1993.

[31] Marc 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*, 7(5):1299–1305, 1996.

[32] 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*, 9(3):595–606, 1997.

[33] M. M. Van Hulle. Faithful representations with topographic maps. *Neural Networks*, 12(6):803–823, 1999.

[34] Ryotaro Kamimura, Taeko Kamimura, and Haruhiko Takeuchi. Greedy information acquisition algorithm: A new information theoretic approach to dynamic information acquisition in neural networks. *Connection Science*, 14(2):137–162, 2002.

[35] R. Kamimura. Information theoretic competitive learning in self-adaptive multi-layered networks. *Connection Science*, 13(4):323–347, 2003.

[36] R. Kamimura. Information-theoretic competitive learning with inverse euclidean distance. *Neural Processing Letters*, 18:163–184, 2003.

[37] Ryotaro Kamimura. Unifying cost and information in information-theoretic competitive learning. *Neural Networks*, 18:711–718, 2006.