

ACCUMULATIVE INFORMATION ENHANCEMENT IN THE SELF-ORGANIZING MAPS AND ITS APPLICATION TO THE ANALYSIS OF MISSION STATEMENTS

Ryozo Kitajima¹, Ryotaro Kamimura²

¹*Graduate School of Science and Technology, Tokai University,
1117 Kitakaname Hiratsuka Kanagawa 259-1292, Japan*

²*IT Education Center and Graduate School of Science and Technology,
Tokai University, 1117 Kitakaname Hiratsuka Kanagawa 259-1292, Japan*

Abstract

This paper proposes a new information-theoretic method based on the information enhancement method to extract important input variables. The information enhancement method was developed to detect important components in neural systems. Previous methods have focused on the detection of only the most important components, and therefore have failed to fully incorporate the information contained in the components into learning processes. In addition, it has been observed that the information enhancement method cannot always extract input information from input patterns. Thus, in this paper a computational method is developed to accumulate information content in the process of information enhancement. The method was applied to an artificial data set and the analysis of mission statements. The results demonstrate that while we were able to explicitly extract the symmetric properties of the data from the artificial data set, only one main factor was able to be extracted from the mission statement, namely, “contribution to the society”. The companies with higher profits tend to have mission statements concerning the society. The results can be considered to be a first step toward the full clarification of the importance of mission statements in actual business activities.

1 Introduction

1.1 Necessity of Detailed Information

Information-theoretic methods have been widely used as an important learning technique in neural networks [1], [2], [3], [4]. Though information-theoretic methods sometimes need heavy computations to approximate information content, a number of simplified computational methods have been proposed, particularly for facilitating computational procedures [5], [6], [7]. In addition, more direct relations have been found between existing learning methods and information-theoretic approaches. Namely, competitive learning

as well as self-organizing maps [8], [9], [10] can be realized by mutual information maximization [11], [12], [13], [14], [15], [16]. When mutual information is defined for output neurons and input patterns, and is maximized, just one output neuron fires, while all the others cease to do so. By changing mutual information obtained by the neural networks, the degree of competition can easily be changed as well. Thus, the method is a more flexible type of competitive learning and can be applied to many problems.

Such methods have shown that the information-theoretic approach is a promising one for neural networks for almost every aspect of learning. How-

ever, one of the main problems of the approach is that the majority of the methods are based on mutual information, which is averaged over all components - information-theoretic methods should aim to analyze the behavior of neural networks in more detail. Specifically, information-theoretic methods have reached the stage where researchers should examine what kind of and how much information is stored in the components of neural networks. One of the main merits of such a detailed analysis of stored information content is that it allows for interpretation of how a component functions in response to input patterns. In addition, it can be inferred how important a component is. When a component contributes to an increase in mutual information, the component can be considered to be more important. In addition, a component with much mutual information naturally contains more information in input patterns. Thus, if it is possible to extract a few important and informative components, those components alone will be enough to fully understand the mechanisms of the neural network.

1.2 Accumulative Information Enhancement

To extend mutual information to provide a more detailed description of the components in neural networks, the information enhancement and information loss methods were introduced [17], [18], [19], [20], [21]. In information enhancement, a component of a neural network is more strongly activated while forcing other components to be inactive. In this way it is possible to know how much mutual information is increased; subsequently, the importance of the component can be determined. On the other hand, in information loss, a target component is forced to be inactive while all the others remain active. Here, we examine how a neural network behaves without the target component, and can actually measure how much mutual information is decreased. Consequently, the importance of the component can be determined.

Both methods have shown good performance for relatively simple problems. However, one of the main problems of the methods is that there have been some cases where the enhancement of a component does not affect the overall performance of the neural networks. More concretely, there have been cases where mutual information does not

change even when a component is enhanced. This means that it is impossible to extract important individual components using these methods. To overcome this problem, neural networks are re-trained, taking into account the importance of components, which becomes gradually clearer [20]. These results show that the importance of components can be gradually clarified by repeating the information enhancement procedures.

As discussed above, to make information enhancement function more explicitly, we need to augment the enhancement of the components by repeating the enhancement procedures. For this purpose, we propose a computational method for mutual information which tries to absorb as much information possible contained in input patterns. Because information in each processing state is gradually accumulated, we call it "accumulative information maximization".

The method is related to the multi-stages of information processing. For example, Linsker [1], [2], [3], [4] proposed the organizing principle called "infomax" by analyzing visual systems. In infomax, visual information is processed in multiple stages from simple to complex forms. In each processing stage, the amount of information is maximally preserved. By this successive information maximization, feature-analyzing functions are supposed to emerge spontaneously. In the same way, it is supposed that informational processing in our neural systems goes through multiple stages, from rough processing to more explicit processing. The multiple information processing with the maximum information preservation principle is applied to the extraction of a few important input neurons (variables), related to variable selection. With multiple stages of information maximization, it can be expected that important input neurons will gradually emerge.

1.3 Application to the Analysis of Mission Statements

To demonstrate the effectiveness of the present method, we applied it to the analysis of company creeds or mission statements. It is well known that mission statements play very important roles in business. For example, companies make decisions according to their mission statements, which specify business plans and philosophies are specified.

The company's mission statements should represent the main principles of business activities.

However, few attempts have been made to analyze relations between mission statements and profitability [22], [23]. For example, Oda and Mitsuhashi reported [24] three types of mission statements: those which stress employees' ideal vision, the contribution to society, and business expansion. They stated that companies with an employee-focused mission statement were the most profitable. However, the data in those papers were limited, and it may be impossible to explore further the validity of this hypothesis. The scope of the data is here extended more widely, using mission statements of companies listed in the second section of the Tokyo Stock Exchange (TSE) [25].

Because the datasets of mission statements are complex, the present method is well suited for this kind of problem. The method aims to condense the information in the complex data as much as possible, meaning in this case that information content is forced to be represented in a fewer number of components. Thus, it becomes easier to interpret the meaning of the complex data, because it is only necessary to examine a few number of condensed components.

1.4 Outline

In Section 2, the close relationship between competitive learning and mutual information maximization is explained. Then, we propose the computational method for the importance of input neurons by using mutual information. In particular, the learning processes of accumulative information maximization is fully explained. In Section 3, two experimental results are presented, namely, an artificial data set and the mission statements. In the artificial data set, it is intuitively shown that the accumulative information maximization can be used to accentuate the main parts of the input patterns and connection weights. The present method excels in extracting one the main factors behind the mission statements.

2 Theory and Computational Methods

2.1 Information-Theoretic Competitive Learning

Information-theoretic methods have been introduced to realize competitive processes in neural networks [26], [12], [13]. Competitive learning has been established as one of the fundamental methods of learning in neural networks [27]. In competitive learning, when an input pattern is given to a neural network, only one output neuron fires, while all the others cease to fire, as shown in Figure 1. In addition, on average, each neuron is expected to fire uniformly [28], [29], [30], [31], [32], [33]. There have been many attempts to improve competitive learning by solving such problems as dead neurons, initial conditions, and the number of neurons [34], [35], [36], [33], [37], [38], [39], [40]. It has been found that competitive learning can be realized by maximizing mutual information between output neurons and input patterns, as shown in Figure 1.

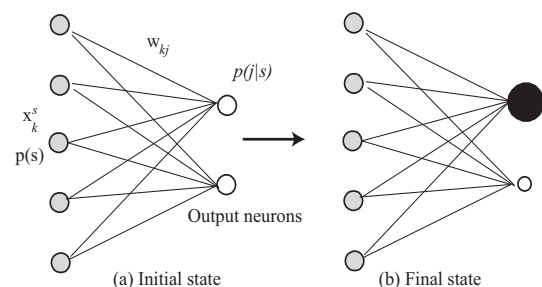


Figure 1. Concept of competitive learning for output neurons

Now, let $p(s)$ and $p(j|s)$ denote the probability of occurrence of the s th input pattern and the firing rate of the j th output neuron for the s th input pattern. Then, mutual information between output neurons and input patterns can be defined by

$$MI = \sum_s p(s) \sum_j p(j|s) \log \frac{p(j|s)}{p(j)}. \quad (1)$$

Mutual information is maximized when an output neuron fires for an input pattern while all other neu-

rons cease to do so, and when the average firing rate $p(j)$ is uniform. Thus, we can say that mutual information can be used to realize competitive learning.

If competitive learning is considered to be a realization of information maximization, the importance of individual components can be immediately determined. Specifically, components which increase information can be considered to be very important. This concept is applied to the detection of important variables (neurons). When an input neuron can increase mutual information between input patterns and output neurons, the input neuron can be considered to be important.

2.2 Information Enhancement Method

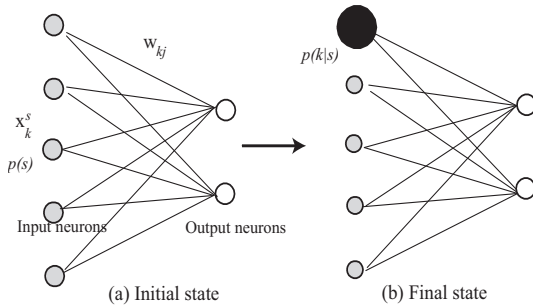


Figure 2. Concept of important input neurons

The present section shows how to enhance a specific input neuron (variable) and to force other neurons to be inactive. In this information enhancement, the importance of input neurons is determined with respect to information in output neurons, as shown in Figure 2.

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 output neuron are computed by $\mathbf{w}_j = [w_{1j}, w_{2j}, \dots, w_{Lj}]^T$, $j = 1, 2, \dots, M$. The output from the output neuron, when the t th input neuron is enhanced, is defined by

$$v_{j,k}^s = \exp \left(- \sum_{l=1}^L \frac{(x_l^s - w_{lj})^2}{2\sigma_{kl}^2} \right), \quad (2)$$

where σ denotes the spread parameter. The spread parameter is changed by using the parameter β ($\beta > 0$)

$$\sigma_{kl} = \begin{cases} 1/\beta, & k = l(\text{enhanced}) \\ \beta, & \text{otherwise.} \end{cases}$$

In enhancing the k th input neuron, the parameter $1/\beta$ is used. On the other hand, the remaining neurons are relaxed by the parameter β .

By normalizing this output, the firing probability is computed by

$$p(j | s; k) = \frac{v_{j,k}^s}{\sum_{m=1}^M v_{m,k}^s}. \quad (3)$$

By using this probability, the mutual information for the k th input neuron is defined as

$$MI(k) = \sum_{s=1}^S \sum_{j=1}^M p(s) p(j | s; k) \log \frac{p(j | s; k)}{p(j; k)}, \quad (4)$$

where

$$p(j; k) = \frac{1}{S} \sum_{s=1}^S p(j | s; k). \quad (5)$$

By using this mutual information, the firing probability for input neurons is

$$p(k) = \frac{MI(k)}{\sum_{l=1}^M MI(l)}. \quad (6)$$

When the firing probability becomes higher, mutual information becomes larger.

The input information obtained by input neurons is introduced, defined by the decrease from maximum uncertainty to observed uncertainty

$$I = \log L + \sum_{k=1}^L p(k) \log p(k). \quad (7)$$

When this input information increases, fewer input neurons fire.

2.3 Accumulative Information Maximization

In accumulative information maximization, connection weights are obtained through two steps of learning, namely, the inner and outer learning cycle in Figure 3. In the outer learning cycle, the spread parameter β is gradually increased using connection weights at the previous $\beta - 1$ th step, where $\beta = 1, 2, \dots, g$. In the inner learning cycle, the parameter β is fixed, and learning is forced to continue until no change in connection weights can be seen. The step in the inner learning cycle is denoted by $\theta = 1, 2, \dots, f$.

Let us now show how to compute the connection weights. At the β th stage of outer learning,

the firing probabilities of input neurons and connection weights at the $\beta - 1$ th stage of outer-learning are used. Then, the inner learning begins, where winners are determined and connection weights are updated until connection weights cease to change, namely, the learning cycle reaches the θ_f th step. Then, the $\beta + 1$ th outer learning cycle begins with the same procedures of inner learning.

Now, the parameter β is set in the outer learning cycle, and the inner learning cycle begins with the fixed value of the parameter. Let \mathbf{x}^s and $^{(\beta, \theta_f)}\mathbf{w}_j$ denote the input and weight column vectors at the β th outer learning cycle and the θ_f th final inner learning step. Then, the distance between input patterns and connection weights at the $(\beta, 1)$ th cycle, namely, at the β th outer cycle and the first inner cycle is

$$^{(\beta, 1)}\|\mathbf{x}^s - \mathbf{w}_j\|^2 = \sum_{k=1}^L {}^{(\beta-1, \theta_f)}p(k)(x_k^s - {}^{(\beta-1, \theta_f)}w_{kj})^2. \quad (8)$$

The $^{(\beta, 1)}c^s$ th winning neuron is computed by

$$^{(\beta, 1)}c^s = \operatorname{argmin}_j {}^{(\beta, 1)}\|\mathbf{x}^s - \mathbf{w}_j\|. \quad (9)$$

Let us consider the following neighborhood function usually used in self-organizing maps

$$h_{j^{(\beta, 1)}c^s} = \exp\left(-\frac{\|\mathbf{r}_j - \mathbf{r}_{^{(\beta, 1)}c^s}\|^2}{2\sigma_\gamma^2}\right), \quad (10)$$

where \mathbf{r}_j and $\mathbf{r}_{^{(\beta, 1)}c^s}$ denote the position of the j th and the $^{(\beta, 1)}c^s$ th neuron on the output space and σ_γ is a spread parameter. Then, the re-estimation equation in the batch mode becomes

$$^{(\beta, 1)}\mathbf{w}_j = \frac{\sum_{s=1}^S h_{j^{(\beta, 1)}c^s} \mathbf{x}^s}{\sum_{s=1}^S h_{j^{(\beta, 1)}c^s}}. \quad (11)$$

As mentioned, the inner learning cycle continues until a certain stopping criterion is met, namely, until the inner learning cycle reaches its final step of (β, θ_f) . The inner learning cycle is considered to be finished when distances between connection weights at the present and at the previous learning inner learning cycle are less than 0.001. Then, the value of the parameter β is increased and again a new inner learning cycle begins. The strength of this method lies in the accumulation of information obtained in learning. More specifically, the present learning process is based on the information obtained in the previous steps.

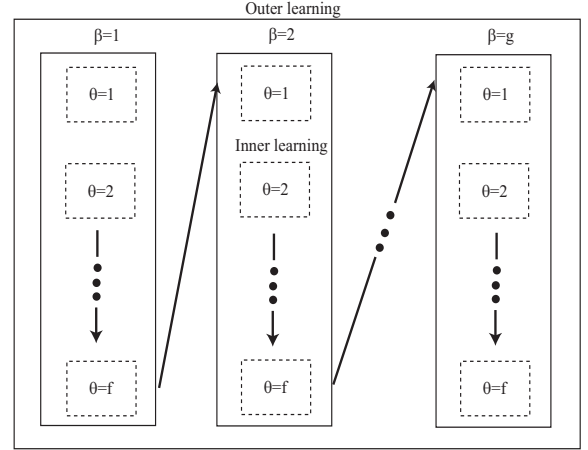


Figure 3. Concept of outer and inner learning

3 Results and Discussion

3.1 Symmetric Data

First, symmetric and artificial data was used in Figure 4(a) to intuitively show how information enhancement works. The number of input neurons was 16 as shown in Figure 4(a). The number of output neurons was 13 by 6 neurons. The number of output neurons was larger than what would be expected for this size of data to clearly demonstrate the performance of our method. Naturally, similar results could be obtained even if the network size was smaller. In the middle of Figure 4, the U-matrix was inserted by using self-organizing maps (SOM), a well-known method for visualizing class boundaries in SOM, representing distances between neurons. The U-matrix represents the distance between two neurons

$$U_{jm} = \|\mathbf{w}_j - \mathbf{w}_m\|. \quad (12)$$

Because focus was on the performance of our new information enhancement method, very conventional methods were used to evaluate its performance. Thus, it can be expected that the results presented here can be easily reproduced.

Figure 5 shows information as a function of the parameter β by standard information maximization (1), the inner learning cycles (2), and accumulative information maximization (3). As can be seen in the figures, information was increased almost up to two by accumulative information maximization in Figure 5(a3). This value of information was much larger than those by standard informa-

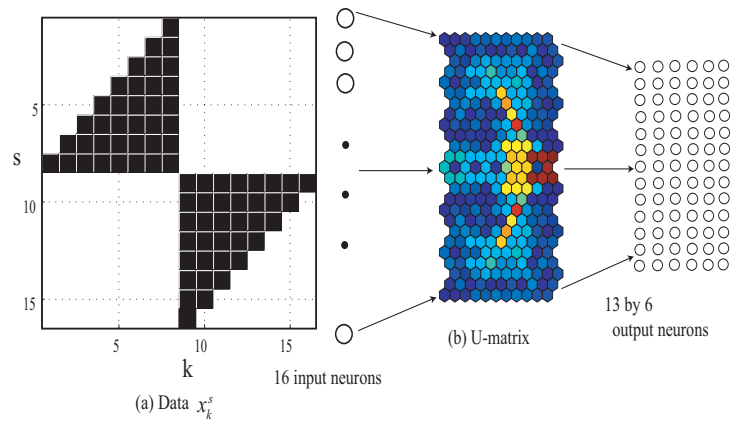


Figure 4. Data (a), network architecture with U-matrix (b) by SOM for the symmetric data

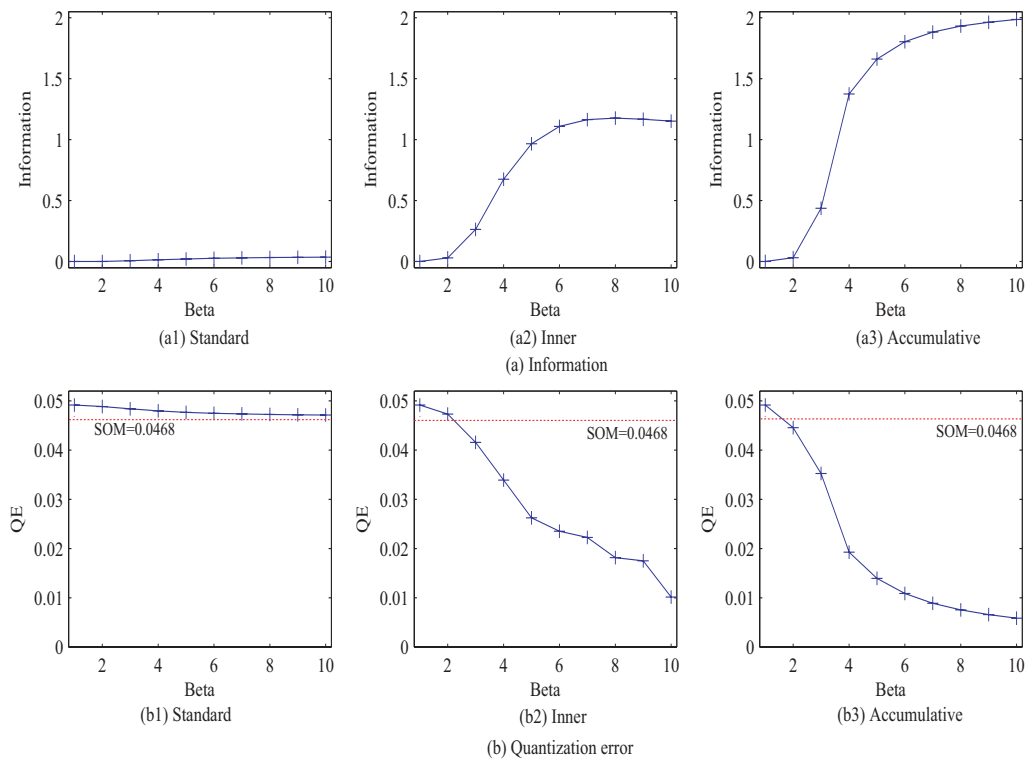


Figure 5. Input information (a) and quantization errors (b) by standard information maximization (1), with the inner learning cycle (2) and accumulative information maximization (3)

tion maximization (a1) and the inner learning cycle (a2). Figures 5(b1), (b2) and (b3) show the quantization errors by standard information maximization, the inner learning cycle and accumulative information maximization, respectively. By using the standard information maximization in Figure 5(b1), the quantization error obtained was larger than that by SOM. When the parameter β was increased from one to 10, the quantization error became closer to that by SOM. By using the inner learning cycle in Figure 5(b2) and accumulative information maximization in Figure 5(b3), the quantization error decreased far below the level obtained by SOM. When accumulative information maximization was used, the lowest error rate was obtained, as shown in Figure 5(b3).

Figure 6(a) shows the firing probabilities $p(k)$ for 15 input neurons using standard information maximization. The firing probabilities remained almost flat even if the parameter β was increased from two to 10 in Figures 6(a1)-(a5). By using the inner learning cycle in Figure 6(b), the firing probabilities at the center became gradually stronger when the parameter β was increased from two in Figure 6(b1) to 10 in Figure 6(b5). Finally, when using accumulative information maximization in Figure 6(c), only two neurons at the center became stronger, while all the other neurons became weaker when the parameter β was four in Figure 6(c2).

Figure 7(a) shows the U-matrices by standard information maximization. The class boundaries in warmer colors divided the input patterns into several parts. When the inner learning cycle was used, the class boundaries became clearer when the parameter β was increased from six in Figure 7(b3) to 10 in Figure 7(b5). When accumulative information maximization was used in Figure 7(c), the class boundary in the middle of the matrix became much stronger and clearer, in Figures 7(c2) to (c5).

3.2 Mission Statements

3.2.1 Data Preparation

The mission statements we analyzed were extracted from 123 manufacturing companies listed in the second section of the Tokyo Stock Exchange (TSE). They were grouped into four industry categories: chemical, machinery, food and electric appliances. The data were constructed by following

two steps. First, the mission statements were extracted from the web sites of the companies. If the mission statements contained plural messages, only one message among them was selected.

Second, text mining was used to analyze the mission statements. Text mining is a technique which analyzes words and sentences to extract the main features of written documents [41]. Text mining includes both morphological and syntax analysis. Morphological analysis aims to extract morphemes or minimum elements of words from sentences. Syntax analysis examines how words are connected with each other. The software "JUMAN" [41] is a well-known morphological analysis software which was used for this purpose.

For the analysis, only content words (noun, verb and adjective), the names of categories and the names of domains were extracted. For example, the category and the domain of a word "school" are "area - institution" and "education". The data set was composed of 44 variables (the three parts of speech, the 28 names of categories and the 13 names of domains). Though the number of original "JUMAN" categories and domains was 34, several new items were added to create a total of 41 domains.

The mission statements were analyzed in five steps, as shown in Figure 8. After determining the data sets to be analyzed (a), text mining technique was used to obtain the morphological data (b). Then, using accumulative information enhancement (AIE), a few important variables were extracted which corresponded to the domain and categories (d). Finally, words closely associated with these variables were selected for easy interpretation (e).

3.2.2 Experimental Results

To show how to analyze the mission statements concretely, we presented one example of the analyses using accumulative information maximization for simplification. Naturally, the same procedure can be applied to the other data sets. Figures 9(a) and (b) show the U-matrices and labels by the conventional SOM and accumulative information maximization (AIE) for the food industry. The U-matrix is a well-known method to visualize connection weights in SOM. The matrix represents the aver-

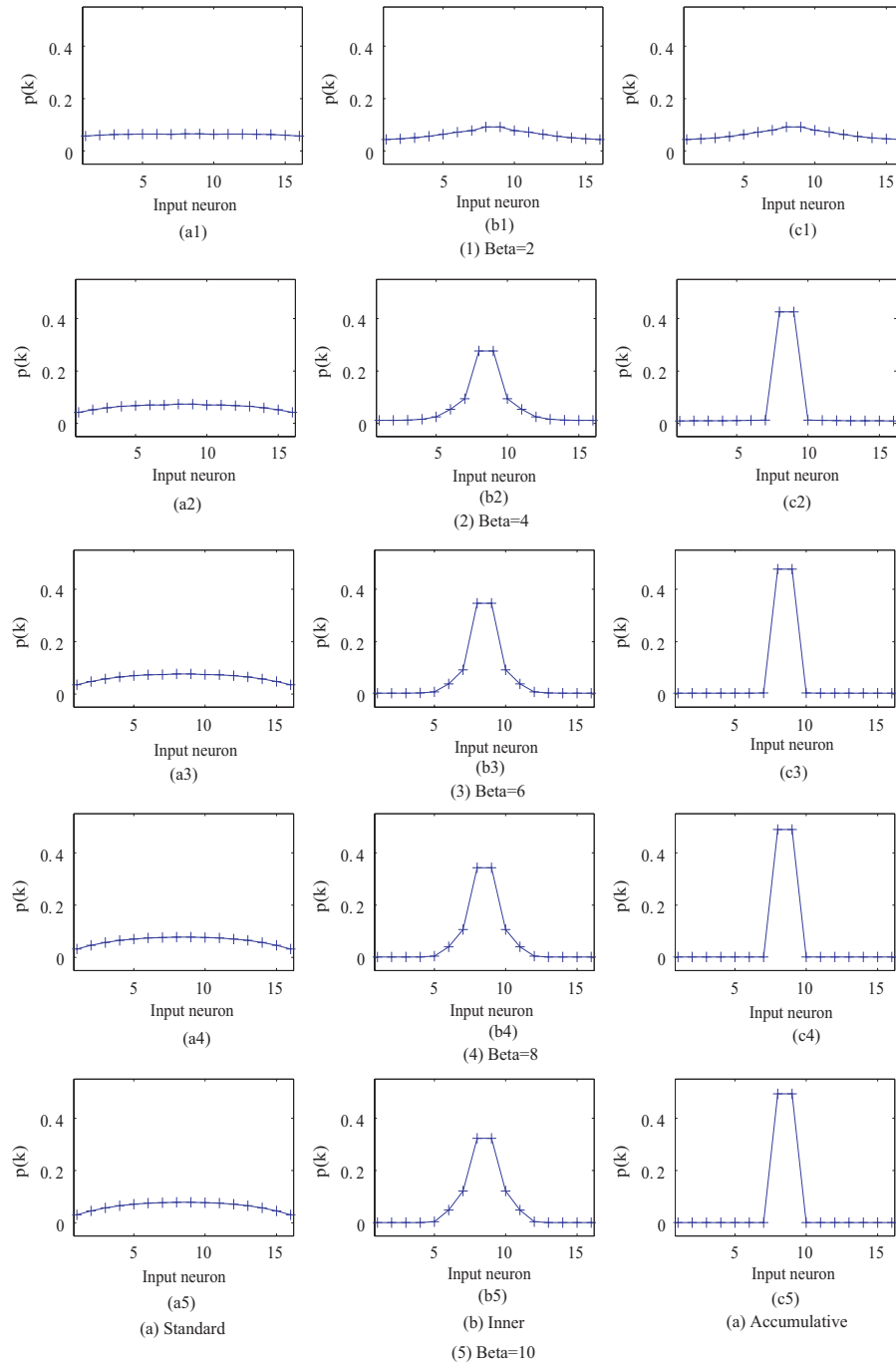


Figure 6. Firing rates when the parameter β was increased from two (1) to 10 (5) by standard information maximization (a), with inner learning (b) and accumulative information maximization (c) for the symmetric data

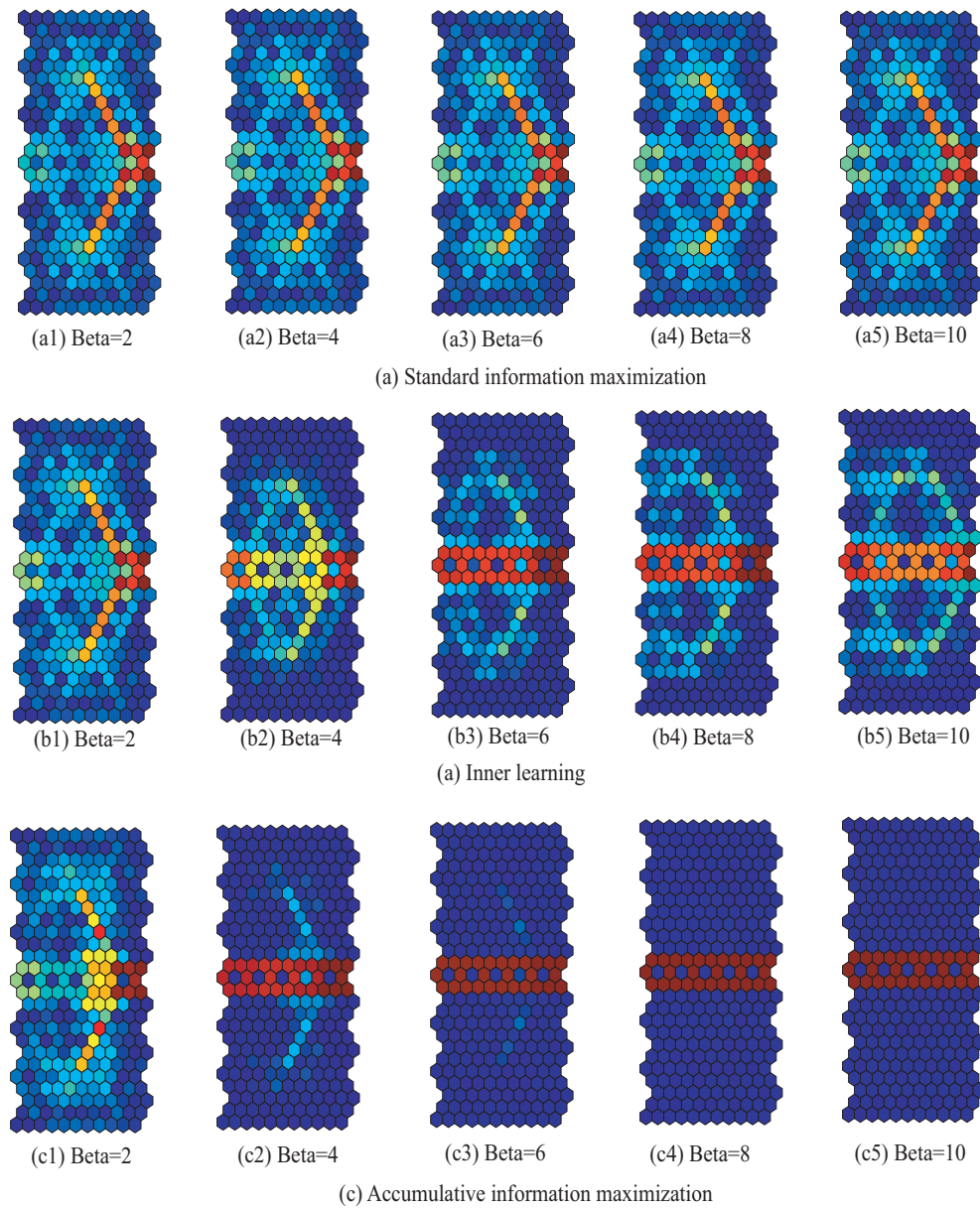


Figure 7. U-matrices when the parameter β was increased from two (1) to 10 (5) by standard information maximization (a), with inner learning (b) and accumulative information maximization (c) for the symmetric data set

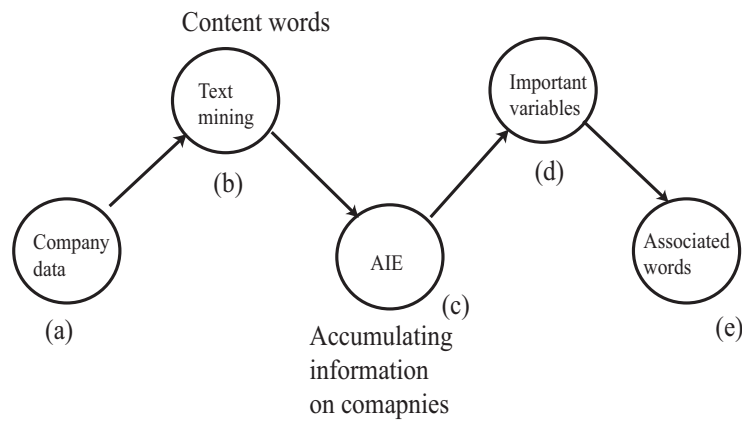
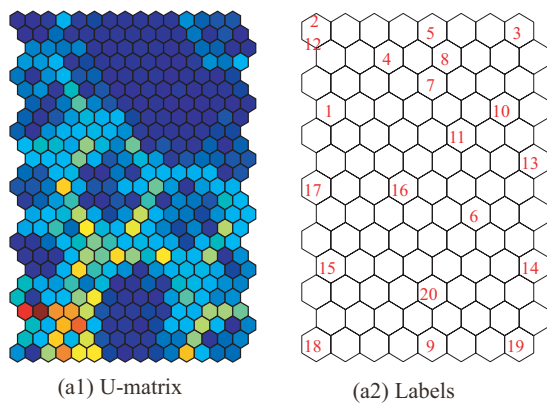
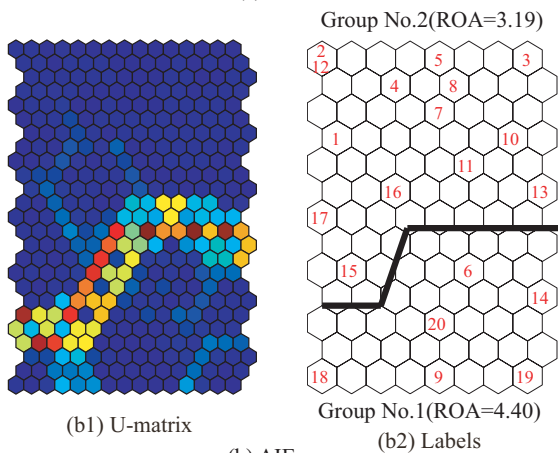


Figure 8. Analysis flow of the company data

age distance between output neurons. In the figure, when the color is warmer, the distance between neurons becomes larger. This means that warmer colors indicate the existence of class boundaries.



(a) SOM



(b) AIE

Figure 9. U-Matrix and labels for the food industry by the conventional SOM (a) and the accumulated information maximization (b)

Figure 9(a) shows the U-matrix and labels by the conventional SOM. The class boundary in warmer colors seems to be located on the lower left hand side of the matrix. In addition, several weak boundaries could be seen on the matrix. However, it was impossible to detect any clear class boundaries. As mentioned, the knowledge of SOM is sometimes very ambiguous, which prevents us from interpreting its meaning, especially for complex data sets.

On the other hand, using accumulative information enhancement in Figure 9(b), a strong class boundary in warmer colors could be detected immediately on the lower side of the matrix. By this class boundary, input patterns or companies could be divided into two classes or groups. This shows that accumulative information enhancement was able to condense information content in the mission statements, indicated by the extraction of clearer class structure.

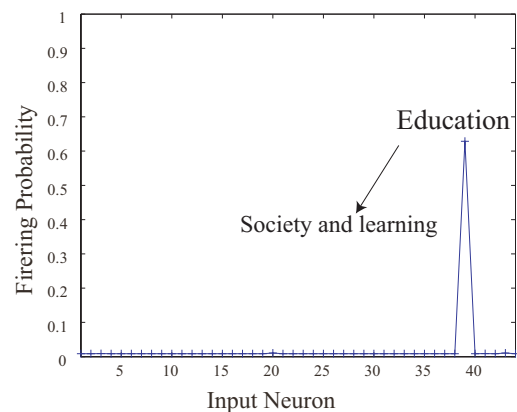


Figure 10. Firing probabilities of input neurons by the accumulated information maximization

Let us now show how accumulative information enhancement can condense the information into a smaller number of components. Figure 10 shows the firing probabilities $p(k)$ of input neurons, corresponding to their relative importance. When the firing probability of an input neuron becomes higher, mutual information of the neuron becomes larger. By increasing information content in input neurons, more mutual information can be accumulated. This means that information content is condensed into a smaller number of neurons by this accumulation. It is thus possible to condense all necessary information into one input neuron when information content is maximally accumulated.

The figure shows that the most important input neuron or variable is a neuron with the domain name "education". The domain name "education" is associated with words such as "society" and "learning". Group No.2 on the upper side of the matrix did not contain these words. On the other hand, these words were commonly found in group No.1 on the lower side.

It is necessary to attempt to explain this division. For this purpose, the average profitability of each company was computed in terms of "Return On Assets" (ROA). ROA was computed by using the ordinary profit and assets published by the Japanese financial services agency "EDINET" (Electronic Disclosure for Investors' NETWORK). When this ROA increases, the profit performance of the companies becomes stronger.

For each group, the average values of ROA were computed. The average ROA of group No.2 was 3.19, while that of group No.1 was 4.40. This means that group No.1, represented by the associated words "society" and "learning", had higher profitability.

3.2.3 Summary of Experimental Results

Table 1 shows the ROA results for the four industry groups. In the chemical industry, the values of group No.1 and No.2 were 3.55 and 1.34 ($\beta = 10$). In the machinery industry, the values of group No.1 and No.2 were 4.13 and 2.94 ($\beta = 10$). In the food industry, the values of group No.1 and No.2 were 4.40 and 3.19 ($\beta = 10$). In the electric industry, the values of group No.1 and No.2 were 6.40 and 1.70 ($\beta = 30$). With those parameter values, the

clearest representations were obtained. On average, the ROA value of group No.1 was 4.62, which was much higher than the 2.29 of group No.2. Across the four industry groups, it was possible to classify companies into two groups with higher and lower ROA values by the important words "contribution to the society".

Table 1. ROA values for two company groups

Industry	ROA	
	Group No.1	Group No.2
Chemical	3.55	1.34
Machinery	4.13	2.94
Food	4.40	3.19
Electric	6.40	1.70
Average	4.62	2.29

Figure 11 shows the important variables extracted and their associated words by accumulative information maximization. The chemical, food and electric industries can be represented by the words "contribution to the society". The "machinery" industry does not simply imply "contribution to the society". The machinery industry is related to the words "device" and "product", which seems to be different from the other industry groups. However, it was found that those words could be interpreted by "contributing to the society by supplying products or technology". Thus, in all cases, the most important factor was "contribution to the society".

4 Discussion

4.1 Validity of Methods and Experimental Results

This paper proposed a new information-theoretic method called "accumulative information maximization", which aims to acquire more information content in input patterns by information enhancement. The information enhancement method was introduced to detect important components in neural networks. By focusing on or enhancing a component, it is possible to examine how much information is obtained by this component. When the

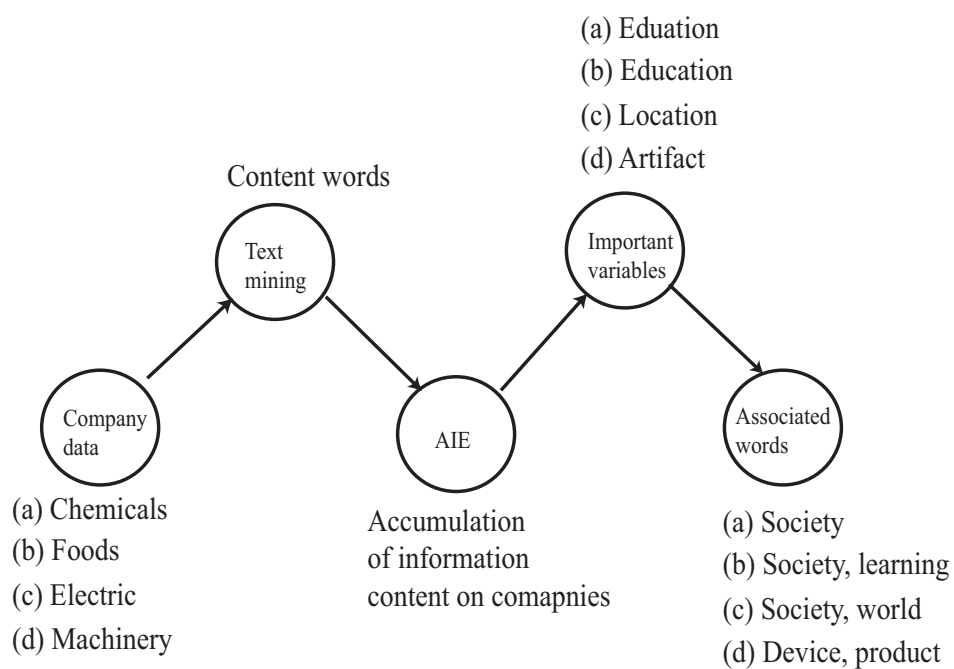


Figure 11. Extracted variables and associated words

obtained information content is larger, the component becomes more important. Though the method was successfully applied to several problems, several cases were found where the information enhancement couldn't be used to increase information content. This means that it is impossible to detect important components by using information enhancement alone.

To overcome this shortcoming, accumulative information maximization was introduced. The main principle of this method lies in the acquisition of information content at every stage of the learning processes. Accumulative information maximization is composed of two stages, namely, inner and outer learning. The learning starts with an initial value of the parameter β . In inner learning, with a fixed value of the parameter β , the learning processes are repeated to obtain stable firing probabilities $p(k)$. Then, the learning shifts to the outer learning, where the parameter β is slightly increased. The same procedures of inner and outer learning are repeated until the final condition is met and the same learning procedures have been conducted.

The method was applied to two data sets, namely, artificial and symmetric data and the analysis of mission statements. In the artificial data set, input information by accumulative information maximization became much larger than by the other methods. In addition, quantization errors were much smaller. By examining the firing probabilities and U-matrices, it was found that the accumulative information maximization focused only on the main parts of input patterns, completely ignoring the others. In the second data set, the mission statements, using accumulative information maximization detected only one important input variable: the words "contribution to the society". These two experimental results show that accumulative information maximization could produce intuitively and empirically reasonable results.

The main merit of this method is that the firing probabilities or the importance of input neurons is based on mutual information, more exactly, normalized mutual information. When this importance becomes stronger, mutual information becomes larger. Information-theoretic competitive learning has shown that mutual information between input patterns and output neurons realizes

the competitive processes of competitive learning. Thus, higher importance means that competition between neurons is highly realized. Because mutual information can be interpreted as a measure of the organization of a system, a neuron with higher importance contributes highly to this organization.

Second, the present method tends to detect only one important input neuron, completely ignoring others. Thus, the method is very convenient for detecting the main features of input patterns. When several features are needed to represent input patterns, obtained information content should only be decreased. As mentioned, when mutual information decreases, the number of active neurons tends to increase as well.

4.2 Limitations of the Method

There are two major problems with the method, namely, heavy computational procedures and poor topological preservation. First, one of the main problems of the method lies in the large amount of complex computations involved. As explained above, the information enhancement method is based on mutual information between output neurons and input patterns. In accumulating information, mutual information must be computed at every step of learning. Thus, it is necessary to simplify the computational procedures as much as possible. It should be noted, however, that in actual situations, heavy computational procedures can be easily reduced. Though many iterations are needed to reach the final informative state, in actual data analysis, this kind of extreme maximization is not necessary. Thus, reasonably small values of the parameter β are enough for solving many practical problems.

Second, topological preservation may be violated. The SOM has been developed to preserve topological consistency, which has received much attention in determining the quality of maps. Thus, it seems that the topological violation observed in the present work may be a serious problem. However, recent research on the quality of SOM shows that some violation of topological consistency is justified to a certain degree when dealing with the visualization of highly complex data [42], [43]. Thus, it is necessary to monitor topological preservation when applying the method to actual data. Feature extraction by accumulative information maximization is accompanied by an in-

crease in topological preservation. Thus, due attention should be paid to topological preservation when considering application to actual data.

4.3 Possibilities of the Method

The method introduced in this study can be considered as a new way to perform variable subset selection for self-organizing maps [44], [45], [46]. Usually, the criteria to select important variables are uncertain in unsupervised learning. However, in the present model, there is only one criterion: mutual information. When this mutual information becomes larger, the variable becomes more important. Selected input variables thus have much information content in input patterns.

The variable selection aspect of our method has two important properties for the self-organizing maps, namely, a new visualization method and the interpretation of input neurons. First, the method can be used to visualize SOM knowledge. As mentioned in the introduction section, it has been difficult to visualize and interpret SOM knowledge, and there have been many attempts to improve visualization and interpretation performance of SOM. Our method can improve the visualization of SOM knowledge by focusing on a small number of informative input neurons.

Second, the method can provide us with a new method for interpreting input neurons. In this paper, we focused on the extraction of only a small number of important input neurons. This is because we considered it important to examine how much information could be obtained by accumulating information content. However, when the information is decreased from its maximum state, the number of input neurons becomes larger.

5 Conclusion

The present paper improved information enhancement for detecting important input neurons by accumulating information content obtained in each learning step. Accumulative information maximization was used to store information content in input patterns as much as possible at every step of learning. This method was applied to two data sets, namely, artificial data and the analysis of mission statements. In both data sets, the experimental re-

sults confirmed that only a small number of important input neurons (variables) were obtained. Thus, the information enhancement method can be used to extract a small number of input neurons (variables) by accumulating information content in learning processes.

One of the main problems is that the accumulated information maximization method requires heavy computational procedures, because it is necessary to compute mutual information for each input neuron. Thus, future work should focus on simplifying these computational procedures as much as possible. Though several problems should be taken into consideration when applying the method to more practical problems, it nevertheless shows great potential for use in extracting a small number of input neurons (variables) for complex and actual data sets.

References

- [1] R. Linsker, "Self-organization in a perceptual network," *Computer*, vol. 21, pp. 105–117, 1988.
- [2] R. Linsker, "How to generate ordered maps by maximizing the mutual information between input and output," *Neural Computation*, vol. 1, pp. 402–411, 1989.
- [3] R. Linsker, "Local synaptic rules suffice to maximize mutual information in a linear network," *Neural Computation*, vol. 4, pp. 691–702, 1992.
- [4] R. Linsker, "Improved local learning rule for information maximization and related applications," *Neural Networks*, vol. 18, pp. 261–265, 2005.
- [5] Z. Nenadic, "Information discriminant analysis: Feature extraction with an information-theoretic objective," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 8, pp. 1394–1407, 2007.
- [6] K. Torkkola, "Feature extraction by non-parametric mutual information maximization," *Journal of Machine Learning Research*, vol. 3, pp. 1415–1438, 2003.
- [7] J. M. Leiva-Murillo and A. Artes-Rodriguez, "Maximization of mutual information for supervised linear feature extraction," *IEEE Transactions on Neural Networks*, vol. 18, no. 5, pp. 1433–1441, 2007.
- [8] D. E. Rumelhart and D. Zipser, "Feature discovery by competitive learning," in *Parallel Distributed Processing* (D. E. Rumelhart and G. E. H. et al.,

- eds.), vol. 1, pp. 151–193, Cambridge: MIT Press, 1986.
- [9] T. Kohonen, *Self-Organization and Associative Memory*. New York: Springer-Verlag, 1988.
- [10] T. Kohonen, *Self-Organizing Maps*. Springer-Verlag, 1995.
- [11] R. Kamimura and T. Kamimura, “Structural information and linguistic rule extraction,” in *Proceedings of ICONIP-2000*, pp. 720–726, 2000.
- [12] R. Kamimura, T. Kamimura, and O. Uchida, “Flexible feature discovery and structural information control,” *Connection Science*, vol. 13, no. 4, pp. 323–347, 2001.
- [13] R. Kamimura, “Information-theoretic competitive learning with inverse Euclidean distance output units,” *Neural Processing Letters*, vol. 18, pp. 163–184, 2003.
- [14] R. Kamimura, “Teacher-directed learning: information-theoretic competitive learning in supervised multi-layered networks,” *Connection Science*, vol. 15, pp. 117–140, 2003.
- [15] R. Kamimura, “Progressive feature extraction by greedy network-growing algorithm,” *Complex Systems*, vol. 14, no. 2, pp. 127–153, 2003.
- [16] R. Kamimura, “Information theoretic competitive learning in self-adaptive multi-layered networks,” *Connection Science*, vol. 13, no. 4, pp. 323–347, 2003.
- [17] R. Kamimura, “Feature discovery by enhancement and relaxation of competitive units,” in *Intelligent data engineering and automated learning-IDEAL2008(LNCS)*, vol. LNCS5326, pp. 148–155, Springer, 2008.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] R. Kamimura, “Selective information enhancement learning for creating interpretable representations in competitive learning,” *Neural Networks*, vol. 24, no. 4, pp. 387–405, 2011.
- [22] B. Bartkus, M. Glassman, and B. McAfee, “Mission statement quality and financial performance,” *European Management Journal*, vol. 24, no. 1, pp. 86–94, 2006.
- [23] B. R. Bartkus, M. Glassman, and R. B. McAfee, “A comparison of the quality of european, japanese and us mission statements:: A content analysis,” *European Management Journal*, vol. 22, no. 4, pp. 393–401, 2004.
- [24] E. Oda and H. Mitsuhashi, “Experimental study of management principle and company performance by text mining(in japanese),” *Management philosophy*, vol. 7, no. 2, pp. 22–37, 2010.
- [25] K. Ryozo and K. Ryotaro, “Company policy analysis by information theoretical neural networks,” in *Proceedings of the 40th fuzzy workshop*, pp. 13–14, 2014.
- [26] 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.
- [27] D. E. Rumelhart and D. Zipser, “Feature discovery by competitive learning,” *Cognitive Science*, vol. 9, pp. 75–112, 1985.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] M. M. Van Hulle, “Faithful representations with topographic maps,” *Neural Networks*, vol. 12, no. 6, pp. 803–823, 1999.
- [32] 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.
- [33] 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.

- [34] 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.
- [35] 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.
- [36] 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.
- [37] 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.
- [38] 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.
- [39] 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.
- [40] 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.
- [41] S. Kurohashi and D. Kawata, "<http://nlp.ist.i.kyoto-u.ac.jp/index.php?juman>,"
- [42] 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.
- [43] 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.
- [44] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.
- [45] A. Rakotomamonjy, "Variable selection using SVM-based criteria," *Journal of Machine Learning Research*, vol. 3, pp. 1357–1370, 2003.
- [46] 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.



Ryozo Kitajima received his B.A. and M.A. from Tokai University, Japan. He is a graduate student in the doctoral course at Graduate School of Science and Technology in Tokai University. His research interests are neural computational approaches to human behaviors in business.



Ryotaro Kamimura received his B.A. and M.A. from the University of Tsukuba, Japan, and Ph. D. in electric engineering from Tokai University, Japan. He is a Professor at IT Education Center and School of Science and Technology in Tokai University. His research interests focus on information-theoretic approaches to neural computing, including information theoretic self-organizing maps and information maximization methods for supervised learning.