Controlling Competition by Structural Information

Ryotaro Kamimura and †Taeko Kamimura

Information Science Laboratory, Tokai University,
1117 Kitakaname Hiratsuka, Kanagawa 259-1292, Japan
ryo@cc.u-tokai.ac.jp
†Department of English, Senshu University,
2-1-1 Higashimita Tama-ku, Kawasaki, Kanagawa, 214-8580, Japan
taekok@isc.senshu-u.ac.jp

Abstract
In this paper, we propose a new information theoretic method called structural information control for flexible feature discovery. The new method has three distinctive characteristics. First, the new method can directly control competitive unit activation patterns. Second, competitive units compete with each other by maximizing their information content about input patterns. Consequently, this information maximization makes it possible to flexibly control competition processes. Third, in structural information control, it is possible to define many different kinds of information content, and we can choose a specific type of information according to a given objective. When applied to competitive learning, structural information can be used to control the number of dead or spare units, and to extract macro as well as micro features of input patterns in explicit ways. We first applied this method to simple pattern classification to demonstrate that information can be controlled and that different neuron firing patterns can be generated. Second, we applied the method to a language acquisition problem in which networks must flexibly discover some linguistic rules by changing structural information.

1 Introduction
In this paper, we propose a new information theoretic method for competitive learning. The new method can contribute to neural computing from three perspectives: (1) it is a competitive learning method that can directly control competitive unit activation patterns; (2) competition is controlled by adjusting structural information in competitive units; and (3) most importantly, the method makes it possible to shed a new perspective on the problems inherent in competitive learning.

First, our new method for competition is considerably different from the traditional competitive method [1], [2]. Our method does not imitate input patterns to produce internal representations. Instead, our method can directly control competitive unit activation patterns to which connections are adjusted. This permits considerable flexibility with which networks can respond to input patterns. For example, if some parts of input patterns are not necessary for producing final competitive unit activation patterns, those parts can be ignored, which is impossible when using traditional competitive learning methods.

Second, competition in the new method is realized by controlling information content in competitive units. Information theoretic methods applied to neural computing have so far given promising results on various aspects of neural computing. Since the concept of information is ambiguous and open to different interpretations, many different approaches have been proposed [3], [4], [5], [6], [7]. Parallel to these information theoretic approaches, practical information theoretic methods have been developed to control hidden unit activation patterns [8], [9], [10]. However, these information theoretic methods were based upon supervised learning, and they could not be applied to the simulation of self-organization processes. In this context, we propose a novel and practical approach called structural information control to unsupervised learning. In this new method, information content can be computed by using unit activations, and can produce appropriate activation patterns in unsupervised ways. In addition, the new method is easy to implement in any neural network architecture, and the process of information maximization can be intuitively interpreted.

Third, our method has been extended so that it can control several different kinds of information, which correspond to different neuron activation patterns. By controlling these kinds of information content, we can flexibly adjust competitive unit activation patterns. For example, by controlling structural information, we can create an activation pattern in which some neurons are always off and thus not used for classification. These neurons can be classified as so-called dead neurons [1],
In traditional competitive learning methods, dead neurons are considered to be unnecessary and are eliminated as much as possible [12]. However, the utility of dead neurons becomes clear when we deal with flexible feature discovery. If we can control dead neurons, the problems inherent in conventional competitive learning can be unified in a framework of dead neuron control. In this perspective, structural information control can be used to adjust the number of neurons and to choose the appropriate number of competitive units. In actual problems, it is impossible to determine the number of exact classes in input patterns in advance, and thus this control of neurons is of great importance in competitive learning. In brief, structural information control can provide the ability to examine these problems in competitive learning from a new perspective.

2 Structural Information

In this paper, we are concerned not with information to be transmitted through information channels but with stored information in systems [13]. Information is thus defined as a decrease in uncertainty, and as such, if we can see any kinds of uncertainty decrease, we call it information.

Here, we consider simple structural information with two random variables, which can be extended to the n random variables. Because information stored in a system is represented by a decrease in uncertainty, this uncertainty, independent of input signals, is the decrease from maximum theoretical uncertainty to actual uncertainty in the system. The maximum uncertainty \( H_0 \) is computed by \( \log M \), where \( M \) is the number of elements in a system and actual uncertainty \( H_1 \) is described by the first order entropy or uncertainty

\[
H_1 = - \sum_j p(j) \log p(j),
\]

where the probability \( p(j) \) denotes the probability with which the \( j \)th element occurs. Thus, information, independent of input patterns, that is, first order information is defined by

\[
D_1 = H_0 - H_1 = \log M + \sum_j p(j) \log p(j).
\]

First order entropy or uncertainty \( H_1 \) may be further decreased to second order uncertainty \( H_2 \), that is, uncertainty decrease after receiving input signals, that is,

\[
H_2 = - \sum_j \sum_s p(s) p(j \mid s) \log p(j \mid s),
\]

where \( p(s) \) represents the probability of input signals \( x_s \), and \( p(j \mid s) \) means the conditional probability of \( j \), given the \( s \)th input signal. Uncertainty decrease, that is, second order information is defined by

\[
D_2 = H_1 - H_2 = \sum_j p(j) \log p(j) - \sum_j \sum_s p(s) p(j \mid s) \log p(j \mid s).
\]

Using the structural parameter \( \alpha \), the second order structural information \( SI \) is defined by

\[
SI = \alpha D_1 + (1 - \alpha) D_2.
\]

where \( \alpha \) is the structural parameter, ranging between 0 and 1. When the structural parameter \( \alpha \) is zero, structural information is equivalent to second order information. On the other hand, when the structural parameter is one, structural information is equal to first order information.

3 Application to Neural Learning

![Figure 1: A network architecture for defining structural information. Structural information can directly control competitive unit activation patterns (a), while the conventional competitive learning methods must approximate input patterns (b).](image)

In this section, we attempt to apply the structural information discussed in the above section to neural network architecture. As shown in Figure 1, the network architecture is composed of input units \( x^s_k \) and competitive units \( v^s_j \). The bias is used only to break the symmetry of competitive unit activation patterns. Thus, in some applications, it is possible for networks to learn input patterns without the bias. Let us define information for competitive units and try to control competitive unit
activation. The $j$th competitive unit receives a net input from input units, and an output from the $j$th competitive unit can be computed by

$$v_j^t = f\left(\sum_k w_{jk} \xi_k^t\right),$$  

where $w_{jk}$ denotes a connection from the $k$th input unit to the $j$th competitive unit. In modeling competition among units, one of the easiest ways is to normalize outputs from the competitive units as follows:

$$p_j^t = \frac{v_j^t}{\sum_m v_m^t}. \quad (7)$$

The conditional probability $p(j \mid s)$ is approximated by this normalized competitive unit output, that is,

$$p(j \mid s) \approx p_j^t. \quad (8)$$

Since input patterns are supposed to be uniformly given to networks, the probability of the $j$th competitive unit is approximated by

$$p(j) = \sum_s p(s)p(j \mid s) \approx \frac{1}{S} \sum_s p_j^s = p_j. \quad (9)$$

Using these approximated probabilities, first order information is approximated by

$$D_1 = \log M + \sum_j p(j) \log p(j) \approx \log M + \sum_j p_j \log p_j. \quad (10)$$

By definition, first order information represents information on the distribution of competitive units from the average point of view. Second order information $D_2$ is approximated by

$$D_2 = -\sum_j p(j) \log p(j) + \sum_s \sum_j p(s)p(j \mid s) \log p(j \mid s) \approx -\sum_j p_j \log p_j + \frac{1}{S} \sum_s \sum_j p_j^s \log p_j^s. \quad (11)$$

As second order information is larger, specific pairs of input patterns and competitive units are strongly correlated. Second order structural information or simple structural information is approximated by

$$SI = \alpha D_1 + (1 - \alpha) D_2 \approx \alpha \log M + (2\alpha - 1) \sum_j p_j \log p_j + (1 - \alpha) \sum_s \frac{1}{S} \sum_j p_j^s \log p_j^s. \quad (12)$$

Differentiating structural information with respect to input-competitive connections $w_{jk}$, we have final update rules:

$$\Delta w_{jk} = \beta(2\alpha - 1) \sum_m \left(\log p_j - \sum_{m} p_m^s \log p_m\right) \times Q_{jk} + \beta(1 - \alpha) \sum_m \left(\log p_j^s - \sum_{m} p_m^s \log p_m^s\right) \times Q_{jk}. \quad (13)$$

where $\beta$ is the learning parameter.

4 Structural Information Control

Let us consider how structural information controls competitive unit activation patterns. Figure 2 shows four input patterns and competitive activation patterns for three different parameter values (a), (b) and (c), and activation patterns by the conventional competitive learning method (conscience learning) (d). In the figures, competitive unit activation levels $p_j^t$ close to one and close to zero are represented by black squares and white squares respectively. We can observe that the competitive unit activation levels $p_j^t$ were actually close to 1 (e.g., 0.99) and 0 (e.g., 0.01) for (a), (b) and (c). No intermediate levels of activation could be seen. As shown in (a), when the structural parameter is zero, four different competitive units respond to different input patterns, respectively. This means that we can obtain completely specialized competitive units. The number of dead units is zero.$^2$When the structural parameter is increased to 0.4, we have many different kinds of activation patterns, depending upon different initial conditions. Figure (b) shows one of the possible activation patterns for this parameter value in which the network can classify four input patterns into two groups. The number of dead units is two. This effect will be obtained by restricting the number of competitive units to two, as in the usual competitive learning method. However, by using structural information, we can obtain this pattern by changing the structural parameter $\alpha$. Figure (c) shows an activation pattern when the structural parameter is 1, that is, when only first order information is used, just one competitive unit responds to all four input patterns every time. The number of dead units becomes three. This experiment shows that we can control the number of dead units by adjusting the structural parameter $\alpha$. Finally, Figure (d) shows activation patterns by the conventional competitive learning (conscience method). In the figure, black squares represent winners in the competition. We

$^2$Dead units in this case are defined as units whose activation levels are lower than 0.3 for all input patterns. This definition is used in the following experiments.
can see that four different units respond to four different input patterns. The traditional method can only produce a case where the structural parameter $\alpha$ is zero, as shown in Figure (a). In addition, differences between winners and losers in terms of activation levels are significantly small, which suggests some difficulty of classification when problems to be classified are more complex.

5 Experiment on Flexible Feature Discovery

We conducted a series of experiments to simulate linguistic rule acquisition, that is, acquisition of Japanese donatory verbs: *ageru* (give), *kureru* (give) and *momu* (receive). In these experiments, we attempt to show different types of competitive unit activation patterns and their corresponding connections by changing structural information. These patterns and connections are easy to interpret, which is not the case with the traditional competitive methods that fail to produce explicit internal representations.

Figure 3(a) shows information as a function of the number of epochs for $\alpha = 0$. Second order information increases rapidly and reaches the level of 90 percent, while first order information is kept at a small value. When $\alpha$ is increased to 0.2, first order information is more clearly increased and reaches the level of 40 percent, whereas second order information is moderately increased and reaches the level of 60 percent. When $\alpha$ is increased to 0.25, first order information surpasses second order information and reaches the level of 60 percent. Finally, when $\alpha$ is further increased to 0.3(Figure 3(b)), second order information maintains a small value, while first order information is increased rapidly and reaches the level of 80 percent.

Figure 4 shows final competitive unit activation patterns for four different values. Figure (a) shows competitive unit activation patterns for 66 input patterns when the structural parameter $\alpha$ is 0. The first, second and final 22 input patterns represent the patterns for the verbs *ageru*, *kureru* and *momu*. The size of squares in the figures represents the magnitude of activations. As shown in Figure (a), there is only one dead unit (No.7). Each unit tends to fire in response to a small group of input patterns. Though some linguistic knowledge is necessary, it is easy to interpret these activation patterns and the corresponding connections.

When $\alpha$ is increased to 0.2, the number of dead units increases to 5, and the remaining units tend to capture features common to the larger groups of input patterns (Figure 4(b)) than those obtained when the parameter $\alpha$ is 0. It is also possible to interpret all the competitive unit activation patterns for this case. When the parameter $\alpha$ is increased to 0.25, networks clearly tend to classify input patterns into two groups, that is, the giving
Figure 3: Information as a function of the number of epochs for four different parameter values: $\alpha = 0(a)$ and $0.3(b)$.

event and the receiving event (Figure 4(c)). The number of dead units is further increased to 7. Finally, as the parameter $\alpha$ is further increased to 0.3, the unit tends to capture all the input patterns (Figure 4(d)).

Figure 5 shows competitive unit activation patterns by the rival penalized method [14], which showed the best performance in traditional competitive learning methods. The learning parameters for a winner and a rival were set to 0.05 and 0.002, respectively, following Xu [14]. The number of learning epochs was 10,000. Even when the number of epochs was further increased, little difference could be observed. In Figure 5, if we closely examine two activation patterns, we can see that different competitive units classify input patterns into two or three groups. However, it is impossible to observe clear characteristics from these activation patterns.

6 Conclusion

In this paper, we have proposed a novel information theoretic method called structural information control to adjust information content in competitive units. The structural information method can directly control competitive unit activation patterns by adjusting information content in competitive units. We have classified information into two
different types of information content, that is, first order and second order information. First order information represents information gained from maximum uncertainty, while second order information represents information gained from first order information when receiving input patterns. By controlling these two types of information content, we can flexibly control competitive unit activation patterns. We applied the new method to two problems: artificial pattern recognition and flexible feature discovery. In the first problem, we have shown that by controlling structural information, we can flexibly produce different types of activation patterns. In the flexible feature discovery problem, we have attempted to show that as the parameter is controlled, the number of dead units is controlled. We have also demonstrated that as the structural parameter is increased, the network can detect more macroscopic features common to the large groups of input patterns.

For further development of the new method, we should take into account two points. First, the present study has been confined to pure competitive learning. However, if it is possible to extend our method to the self-organization maps such as Kohonen’s maps. In other words, if we can take spatial locations of neurons into account, more interesting neuron firing patterns may be created. Second, because structural information can be used to extract macro as well as micro features of input patterns, more exact relations between extracted features and the structural parameter should be explored. Though further study is needed for the extension of structural information, we have shown at least in this paper that structural information gives a new paradigm for unsupervised learning.

References