Proceedings of the 2007 IEEE Symposium on
Foundations of Computational Intelligence (FOCI 2007)

Combining Hard and Soft Competition in Information-Theoretic Learning

Ryotaro Kamimura
Information Science Laboratory
ryo@cc.u-tokai.ac.jp
Tokai University
1117 Kitakaname Hiratsuka Kanagawa 259-1292, Japan

Abstract

In this paper, we try to combine conventional competitive learning with information-theoretic methods to improve competitive performance. We have so far proposed a new type of information-theoretic method to simulate competitive processes. Though the information-theoretic method solves the dead neuron problem and shows the soft-type competition, the method is sometime slow in convergence. To solve this problem, we combine standard learning with information-theoretic learning. By this combination, we can shorten a learning process considerably.

1 Introduction

We have so far introduced information-theoretic competitive learning that realizes the very soft-type competition [Kamimura, 2003a], [Kamimura, 2003b], [Kamimura, 2006]. The method has also been proposed to solve the fundamental problems of competitive learning such the dead neuron problem [Rumelhart and McClelland, 1986], [Grossberg, 1987], [DeSieno, 1988], [Ahalt et al., 1990], [Xu, 1993], [Luk and Lien, 2000], [Hulle, 1997]. Compared with the rigid type of competition of the winner-take-all, better performance can be expected in many problems. In addition, in information processing in living systems, the soft-type competition seems to be a fundamental mechanism. The method is suitable for simulating information processing in living systems.

However, one of the main problems of the information-theoretic method is slow convergence for complex problems. Information is slowly increased, because the method must take into account all connection weights. In this context, we combine conventional competitive learning with our information-theoretic method. Because the standard method is computationally light, we use it as much as possible. However, when the standard method becomes ineffective in learning, the information-theoretic method should be applied.

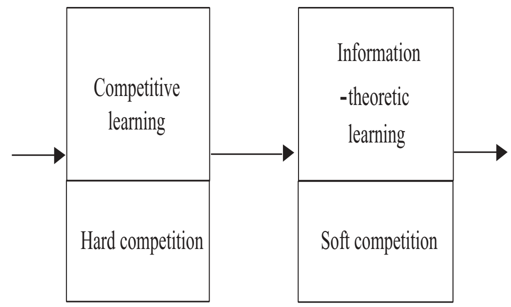

Figure 1. A concept of combining competitive learning and information-theoretic learning.

2 Theory and Computational Methods

2.1 Competitive Learning as a Process of Information Maximization

Figure 1 shows a concept of our combination. In conventional competitive learning, the winner-all-take algorithm, that is, hard competition is used. On the other hand, in information-theoretic learning, all neurons must be updated to simulate competition, meaning that soft competition is realized. Because all neurons must be taken into account, more sophisticated competition can be realized. However, as already mentioned, learning is slow due to soft competition. Conventional competitive learning directly choose just one winner and updates connection weights to the winner. This means that the method is computationally light. We have had so far many applications. For example, self-organizing map is one of most successful application of competitive learning. Thus, considering the effectiveness of conventional competitive learning, we try to use the conventional method as much as possible. When information cannot be increased by the standard learning, information maximization is applied.

In competitive learning, the winner-take-all algorithm is used to detect a winner, and to update connection weights into the winner. Update rules can be formulated as follows:

$$\Delta w_{jk} = \beta Q_j^s (x_k^s - w_{jk}), \quad (1)$$

where β is a learning parameter, and Q_j^s is set to one

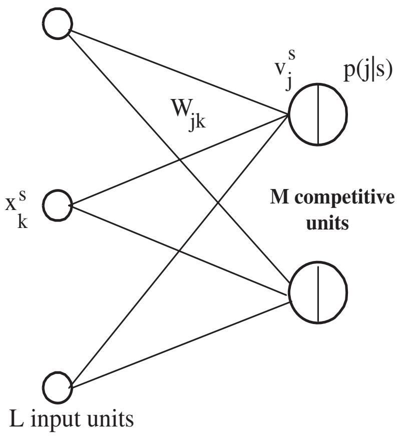

Figure 2. A network architecture for competitive learning.

only if the jth unit is a winner in competition. From an information-theoretical point of view, this is a method where information is supposed to be already maximized. Thus, the probability of the winner is set to one. Then, connection weights into the winner are updated. This is the so-called psudo-maximized information learning.

2.2 Information Maximization

We have defined information content as mutual information between input patterns and competitive units [Kamimura, 2003b]. As shown in Figure 2, a network is composed of input units x_k^s and competitive units v_j^s. We used as the output function the inverse of the Euclidean distance between connection weights and input patterns. Thus, an output from the jth competitive unit can be computed by

$$v_j^s = \frac{1}{\sum_{k=1}^{L}(x_k^s - w_{jk})^2}, \qquad (2)$$

where L is the number of input units, and w_{jk} denote connections from the kth input unit to the jth competitive unit. The output is increased as connection weights are closer to input patterns.

The conditional probability of firing of the jth unit, given the sth input pattern $p(j \mid s)$ is computed by

$$p(j \mid s) = \frac{v_j^s}{\sum_{m=1}^{M} v_m^s}, \qquad (3)$$

where M denotes the number of competitive units. Since input patterns are supposed to be uniformly given to networks, the probability of the jth competitive unit is computed by

$$p(j) = \frac{1}{S}\sum_{s=1}^{S} p(j \mid s). \qquad (4)$$

By using these probabilities, information I is computed by

$$\begin{aligned} I &= -\sum_{j=1}^{M} p(j)\log p(j) \\ &\quad + \frac{1}{S}\sum_{s=1}^{S}\sum_{j=1}^{M} p(j \mid s)\log p(j \mid s), \end{aligned} \qquad (5)$$

where S is the number of input patterns. Differentiating information with respect to input-competitive connections w_{jk}, we have final update rules to increase information ([Kamimura, 2003b]).

Differentiating information with respect to input-competitive connections w_{jk}, we have

$$\Delta w_{jk} = \beta U_j^s (x_k^s - w_{jk}), \qquad (6)$$

where

$$\begin{aligned} U_j^s &= -\sum_{s=1}^{S}\left(\log p(j) - \sum_{m=1}^{M} p(m \mid s)\log p(m)\right) \\ &\quad \times p(j|s)v_j^s \\ &\quad + \sum_{s=1}^{S}\left(\log p(j \mid s) - \sum_{m=1}^{M} p(m \mid s)\log p(m \mid s)\right) \\ &\quad \times p(j|s)v_j^s \end{aligned} \qquad (7)$$

2.3 Combining Two Methods

We can easily combine two methods in one framework. In the first stage of learning, supposed maximized information learning should be used, because the method is computationally light. In the equation

$$\Delta w_{jk} = \beta R_j^s (x_k^s - w_{jk}), \qquad (8)$$

R_j^s should be changed according to a method taken.

3 Results and Discussion

In this experiment, we try to show that the hybrid method enhances the performance of competition. The artificial data was composed of patterns drawn from two normal distributions with three different variances. The number of input and competitive units are two, respectively. Figure 3(a) shows information as a function of the number of epochs by competitive learning and information-theoretic competitive learning. In this data(a), two groups are clearly separated with smaller variance. Information is increased by competitive learning until the number of epoch is 313, and then though information max is applied, information is not increased. Figure 4(a) shows data with the variance=1. As shown in Figure 4(b), information is increased until the number of epoch is 382, and then when information max is applied, information is further increased. However, as shown in Figure 4(c) and (d), training and generalization errors are not significantly decreased. Figure 5(a) shows data with the variance=1. As shown in Figure 5(b), information is increased until the number of epoch is 384, and then when information max is applied, information is further increased. Then, as shown in Figure 5(c) and (d), training and generalization errors are visibly decreased. These results show that the hybrid method has a possibility of improved performance when applied to complex problems. Figure 6 shows connection weights by competitive learning and information-theoretic learning. As can be seen in the figure, similar connection weights are obtained by both methods. The main difference is the magnitude of the connection weights. Thus, we can say that information-theoretic learning enhances connection weights by competitive learning.

4 Conclusion

In this paper, we have tried to combine standard competitive learning and information-theoretic methods in one framework. To our point of view, standard competitive learning is a special case in which information is supposed to be maximized before learning (winner-take-all). Due to this supposition, the algorithm is greatly simplified with appropriate performance ever reported. Thus, we should use this conventional method so long as the method is effective in learning. When the method becomes ineffective in learning, new information-theoretic methods should be used for the practical problems. We have applied the method to an artificial data with different variances. We have found that as the variance is larger or the problem become more complex, the information-theoretic method is more effective in increasing information. For further study, we should apply this hybrid method to more complex and practical problem to see whether the main result obtained in this paper is valid for larger problems.

References

[Ahalt et al., 1990] Ahalt, S. C., Krishnamurthy, A. K., Chen, P., and Melton, D. E. (1990). Competitive learning algorithms for vector quantization. *Neural Networks*, 3:277–290.

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

[Grossberg, 1987] Grossberg, S. (1987). Competitive learning: from interactive activation to adaptive resonance. *Cognitive Science*, 11:23–63.

[Hulle, 1997] Hulle, M. M. V. (1997). 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.

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

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

[Kamimura, 2006] Kamimura, R. (2006). Improving information-theoretic competitive learning by accentuated information maximization. *International Journal of General Systems*, 34(3):219–233.

[Luk and Lien, 2000] Luk, A. and Lien, S. (2000). Properties of the generalized lotto-type competitive learning. In *Proceedings of International conference on neural information processing*, pages 1180–1185, San Mateo: CA. Morgan Kaufmann Publishers.

[Rumelhart and McClelland, 1986] Rumelhart, D. E. and McClelland, J. L. (1986). On learning the past tenses of English verbs. In Rumelhart, D. E., Hinton, G. E., and Williams, R. J., editors, *Parallel Distributed Processing*, volume 2, pages 216–271. MIT Press, Cambrige.

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

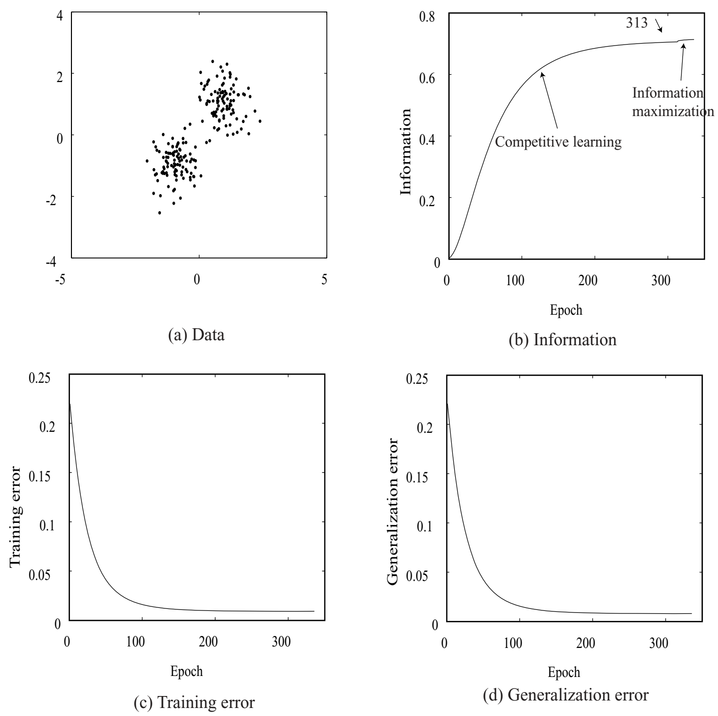

Figure 3. Data (a), information (b), training (c) and testing errors (d) for the variance=0.5.

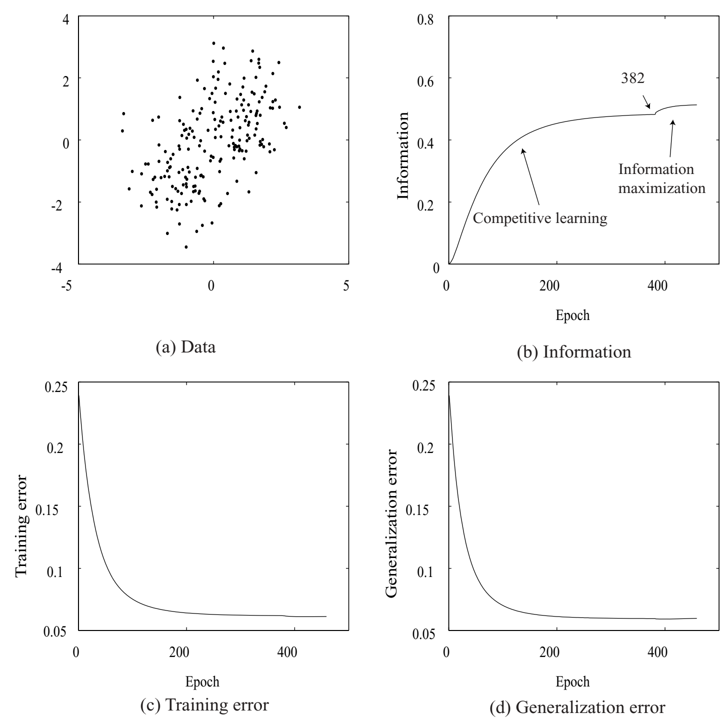

Figure 4. Data (a), information (b), training (c) and testing errors (d) for the variance=1.0.

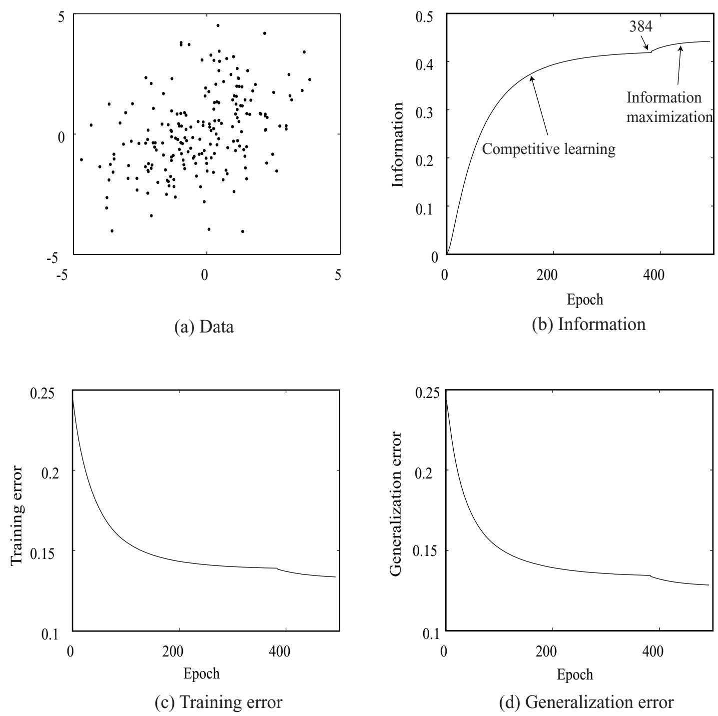

Figure 5. Data (a), information (b), training (c) and testing errors (d) for the variance=1.5.

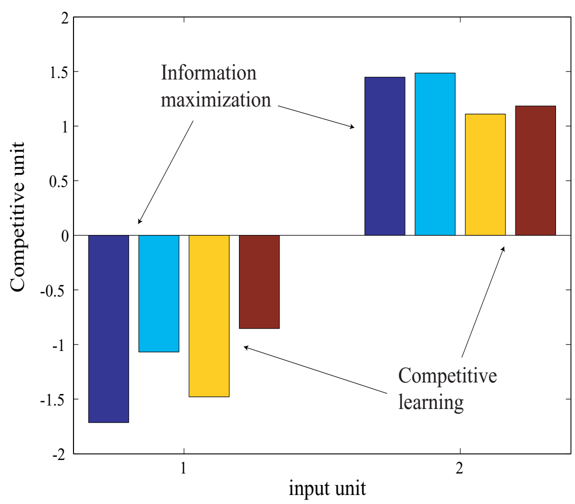

Figure 6. Connection weights by competitive learning and information-theoretic learning.