OPTICAL REVIEW Vol. 1, No. 1 (1994) 41-43

Artificial Neural Network That Can Say "Unknown"

Norimichi TSUMURA, Kazuyoshi ITOH and Yoshiki ICHIOKA

Osaka University, Department of Applied Physics, Yamadaoka 2-1, Suita, 565 Japan

(Accepted July 18, 1994)

A new artificial neural network that can reject strange patterns is presented. The new network is an improved version of the RCE network. Each cell in the last layer of the present network has two thresholds. The new threshold produces the smallest boundary that encloses all examples associated with the cell in the pattern vector space. The present network can reject strange patterns by using this boundary. The rejection of strange patterns in the present and RCE networks was investigated by practical experiments using digits sampled from X-ray films of the human chest.

Key words: strange pattern, rejection, RCE network, digits, classification

1. Introduction

Most artificial neural networks can be considered as pattern-classification systems that can learn from training examples. Valid generalization for patterns that are not included in examples is desired in artificial neural networks. Valid or reliable generalization requires rejection of strange patterns that are not included in any trained classes.

An artificial neural network that can reject strange patterns appropriately is proposed in this paper. The new network is an improved version of the RCE (restricted coulomb energy)[1-3] networks. The RCE networks have well-known excellence in their training and processing abilities. The present network has two features that the RCE networks do not have. Firstly, each cell in the last layer of the two-layered architecture has two thresholds. Secondly, the transfer function of the cells in the present network is different in training and processing modes. One of the two thresholds is the same as that of the cells in the RCE network, while the other produces the smallest boundary that encloses all the examples associated with the cell in the pattern vector space. The network can reject strange patterns by using this boundary appropriately.

2. The Present Network

The present network consists of input and prototype layers, as shown in Fig. 1. The cells in the prototype layer and the weight vectors associated with these cells are called prototype cells and prototype vectors, respectively. Each cell has six parameters; prototype vectors $\boldsymbol{p}_i$ ($i=1\sim n$), class c_i, inter-class threshold $r^{(0)}{}_i$, intra-class threshold $r^{(1)}{}_i$, distance d_i and state s_i.

In the training mode, the pairs of training patterns and their classes (training classes) $\{(\boldsymbol{f}_j,t_j); j=1\sim N\}$ are used to train the network. These pairs are called training pairs, and may successively be fed to the network in random order. In the initial state of the network, there are no prototype cells. During the training of the network, prototype cells are committed, and their two thresholds are modified. In all the prototype cells, the rules of the modifications are the same.

In the case where the j-th training pair $(\boldsymbol{f}_j,t_j)$ is presented to the network, the modifications of the thresholds in the i-th prototype cell are as follows. The prototype cell determines the distance d_i and the state s_i, according to the transfer function below.

$$d_i=\|\boldsymbol{f}_j-\boldsymbol{p}_i\| \ , \tag{1}$$

$$s_i=\mathrm{sgn}(\ r^{(0)}{}_i-d_i) \ , \tag{2}$$

where $\|\bullet\|$ and sgn ($\bullet$) are the Euclidean norm and the signum function, respectively. When the state s_i is equal to 1, we say that the cell is active. If the cell is not active, no modifications are made in the cell. If the cell is active and the class c_i is equal to the training class t_j and the distance d_i of the cell is the smallest in all prototype cells, then the intra-class threshold $r^{(1)}{}_i$ is set to the larger value between $r^{(1)}{}_i$ and d_i. If the cell is active and the class c_i is not equal to a training class t_j, then the inter-class threshold $r^{(0)}{}_i$ is set to d_i. At this time, if the modified inter-class threshold $r^{(0)}{}_i$ becomes smaller than the intra-class threshold $r^{(1)}{}_i$, the intra-class threshold $r^{(1)}{}_i$ is reset to 0.

If no cells are active in the network after the above modifications of the thresholds, a new prototype cell is committed to the prototype layer. The parameters of the committed cell are set as follows; 1) the input vector $\boldsymbol{f}_j$ is substituted for the prototype vector $\boldsymbol{p}_i$, 2) the inter-class threshold $r^{(1)}{}_i$ is set to the smallest value of the distance parameter in all cells whose classes are not equal to the training class t_j, 3) the class c_i is set to the training class t_j, 4) the intra-class threshold $r^{(1)}{}_i$ is set to 0. At that time, if the present training pair is the first fed to the network, the inter-class threshold $r^{(0)}{}_i$ is set to be the largest possible value in the machine.

One epoch is defined as the term during which all the training patterns are fed to the network. The epochs are repeated in training, and the order of feeding the training pairs is randomized at the beginning of each epoch. If there are no modifications of thresholds and no commitment of cells in one epoch, the training may be finished. By this training process, the parameters of the network converge to a state in which all training pairs are correctly classified.

In the processing mode, an input pattern to be processed

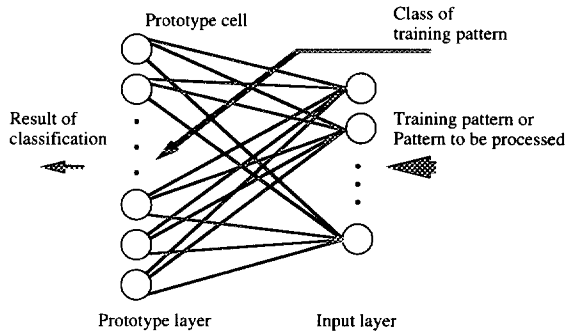

Fig. 1. Architecture of the present network.

is fed to the input layer. All the prototype cells in the network determine the distance d_i and the state s_i. The transfer function is different from that of processing mode as follows,

$$s_i = \mathrm{sgn}(r^{(b)}_i - d_i) \ , \tag{3}$$

where $r^{(b)}_i$ is called the boundary threshold. The boundary threshold is given by

$$r^{(b)}_i = (1-e)r^{(0)}_i + e r^{(1)}_i \ . \tag{4}$$

The parameter e ($0 \leq e \leq 1$) governs the degree of rejection of strange patterns. The value of the parameter e may be set through trial and error. The class c_i of the cell that is active and whose distance d_i is the shortest in all the cells is selected to be the result of classification. If no cells become active, the input pattern is rejected as a strange pattern.

In Fig. 2(a)-(d), middle stages of the training process of a simple network with two input cells are shown in a 2-D pattern vector space. White squares and white triangles in the figures indicate training vectors in the first and the second classes, respectively, and black squares and black triangles indicate the labeled prototype vectors for the first and the second classes, respectively. Solid and broken circles indicate boundaries produced by inter-class and intra-class thresholds, respectively. The cross marks in the figures indicate the presented training vector in each process.

As shown in Fig. 2(a), the first prototype cell is committed to the network when the first training pair is fed to the network. We omitted indicating the solid and broken circles in this state, because the radii of these circles are the largest possible value and zero, respectively. If the next input pattern is correctly classified, then the intra-class threshold is modified as in Fig. 2(b). Figure 2(c) shows the modification of the inter-class threshold of the triangular prototype cell and commitment of a new square prototype cell. Figure 2(d) shows the final state of the network. It is important to note that each radius of the broken circles is the smallest under conditions in which the broken circle surrounds as many training vectors in the same class as possible.

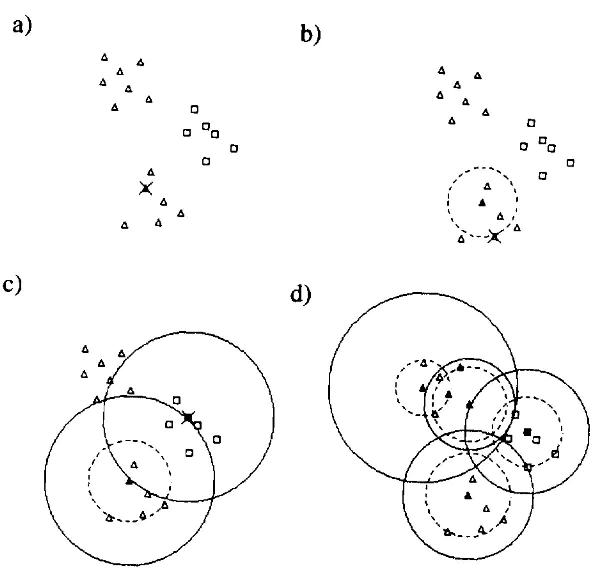

Fig. 2. Middle stages of the training process.

3. Experiments by Practical Data

Practical data consist of images of numerals. Each numeral is sampled from ID numbers that are recorded on X-ray films of the human chest. The numerals are binary images of 12 by 14 pixels. Examples are shown in Fig. 3. Deformation and positional shift have been introduced in these images during the processes of recording, digitizing and segmentation. There are 3,184 samples which were sampled from 139 sheets of X-ray film. We compared the RCE, RCE-2[3] and present networks using this practical data. The RCE-2 network is an improved RCE network and usually demonstrates a higher rate of correct classification than the original. A set of 960 images of even numbers was used as a set of training patterns and the rest of 952 images of even numbers as a set of input patterns to be processed. To estimate the degree of rejection against strange patterns, a set of 1,272 images of odd numbers was used as a set of strange patterns.

The results of the experiments on the RCE, RCE-2 and present networks are shown in Table 1. In the present network, the parameter e was changed from 0.0 to 1.0 in steps of 0.5. The results show the rate of correct classification and incorrect classification against even digits, the rate of rejection against strange patterns, the number of prototype cells and the number of epochs required for training. The averages and standard deviations are obtained from 100 different initializations of the networks.

The RCE-2 network had the highest rate of correct

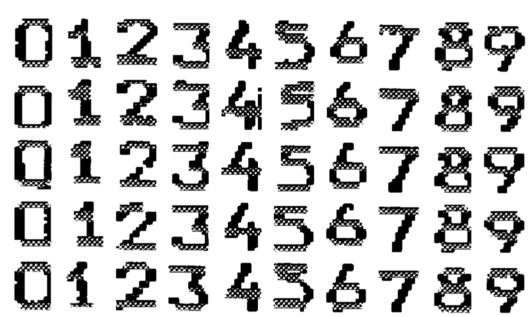

Fig. 3. Examples of images sampled from X-ray films.

Table 1. Results of experiments in RCE, RCE-2, present networks.

	e	Correct classification rate (%)	Incorrect classification rate (%)	Rejection rate against strange patterns (%)	Number of training epochs	Number of phototype cells
RCE network		98.85±0.36	0.36±0.22	67.04±5.77	3.7±0.5	35.4±2.3
RCE-2 network		99.17±0.35	0.55±0.33	32.73±10.42	2.8±0.5	29.6±2.5
Present network	0.0	98.94±0.39	0.32±0.22	67.27±6.06	4.0±0.5	35.4±2.0
	0.5	98.20±0.49	0.22±0.19	89.73±5.76		
	1.0	96.74±0.69	0.20±0.18	93.47±5.24		

classification, the highest speed of learning and the smallest number of prototype cells of the three networks. However, the rate of rejection against strange patterns was the lowest and the rate of incorrect classification was the highest of the three networks. This means that the RCE-2 network has the lowest reliability of classification of the three networks. The RCE network had a higher rate of rejection against strange patterns and lower rate of incorrect classification than the RCE-2 network. Nevertheless, approximately 32.9 percent of strange patterns could not be rejected. On the other hand, the present network raised the rate of rejection against strange patterns and suppressed the rate of incorrect classification rate by changing parameter e. Note that the high rate of rejection against the strange patterns sacrifices the high rate of correct classification.

4. Discussion and Conclusions

We have presented a new network that can reject strange patterns. The network has two thresholds in each cell in the last layer. The rate of rejection against strange patterns can be easily controlled by changing one parameter of the network. In an experiment involving the practical data, we showed the improvement of rejection against strange patterns and suppression of incorrect classification in the present network compared with the RCE networks. This improvement sacrifices the high rate of correct classification. For the reliability of classification in practical situations, the low rate of incorrect classification and high rate of rejection against strange patterns are more important than the high rate of correct classification. The present network may be used in practical applications that require reliable classification.

Acknowledgments

This research was supported in part by a grant-in-aid from the Ministry of Education, Science and Culture, Japan.

References

1) D.L. Reilly, L.N. Cooper and C. Elbaum: Biol. Cybern. 45 (1982) 35.
2) C.L. Scofield, D.L. Reilly, C. Elbaum and L.N. Cooper: *Neural Information Processing Systems*, eds. D.Z. Anderson (American Institute of Physics, New York, 1988) p. 674.
3) M.J. Hudak: Cybern. Syst. 23 (1992) 483.