ORIGINAL CONTRIBUTION

A Coded Block Adaptive Neural Network System With a Radical-Partitioned Structure for Large-Volume Chinese Characters Recognition

MARK W. MAO AND JAMES B. KUO

National Taiwan University

(Received 30 August 1991; revised and accepted 26 December 1991)

Abstract—This paper presents a coded block adaptive neural network system using a radical-partitioned structure for large-volume Chinese characters recognition. Using the coded block adaptive neural network system with a radical-partitioned structure, 1,000 frequently-used Chinese characters have been successfully trained in 139.2 hours using an 18 MIPS computer. According to the simulation results, the coded block system with a radical-partitioned structure provides an acceptable learning time, a good recognition rate, and an excellent expansion capability for large-volume Chinese characters recognition.

Keywords—Neural network, Radical, Chinese characters, Recognition, Coded, Block, Adaptive.

1. INTRODUCTION

A pattern recognition system, as shown in Figure 1, is composed of a translation-invariant network and a standard adaptive two-layer network. The translation-invariant network is used to map a retinal image into multibit outputs (Rumelhart & McClelland, 1986; Widrow, Winter & Baxter, 1988). The standard adaptive two-layer network (Mirchandani & Cao, 1989) made of $2n^2$ adaptive neurons and $2n^4$ connections between them, can be trained to provide output responses corresponding to the original image as required. Software implementation of the adaptive neural network using back propagation (Vogl, Mangis, & Rigler, 1988; Yamada, Kami, Tsukumo, & Temma, 1989) has been demonstrated. A structured neural network architecture to speed up learning for hand written digit recognition has been reported (Le Cun et al., 1989). In fact, for Western characters recognition, only about 36 alphanumerics are needed. However, for Chinese characters recognition, over 20,000 characters are necessary. For large-volume Chinese characters recognition, learning speed is determined by the number of neurons and connections between them. How to train large-volume Chinese characters using an efficient neural network structure has been a challenging effort. Re-

cently, a block adaptive network structure using n^2 local blocks for an $n \times n$ input array as shown in Figure 2 has been used to enhance the learning speed (Chen, Hsiao, & Kuo, 1990; Kuo, Chen, & Hsiao, in press) for alphanumerics recognition. However, the block adaptive network structure requires n^2 local blocks, which is still a bottleneck for training large-volume Chinese characters. A coded block neural network system, which requires an order-of-magnitude fewer local blocks, has been used to provide a faster learning process for the alphanumerics recognition (Kuo et al., 1991; Kuo & Mao, 1991). In this paper a coded block adaptive neural network system with a radical-partitioned structure for large-volume Chinese characters recognition is presented. In the following sections, the radical-partitioned coded block structure is introduced, followed by the algorithm used in the system and simulation results.

2. THE CODED BLOCK ADAPTIVE RADICAL-PARTITIONED STRUCTURE

Figure 2 (top) shows the block adaptive network structure for an $n \times n$ input array (Chen et al., 1990; Kuo et al., in press). In the block structure, the system is partitioned into n^2 blocks. Each block, as shown in Figure 2 (bottom), contains $S + 1$ neurons and $Sn^2 + S$ connections between them, which are grouped into two layers. Instead of $2n^4$ global connections in two layers of the standard structure, only Sn^2 local connec-

---

Requests for reprints should be sent to James B. Kuo, Room 526, Department of Electrical Engineering, National Taiwan University, Roosevelt Road, Section 4, #1, Taipei, Taiwan 107.

FIGURE 1. A pattern recognition system composed of a translation-invariant neural network and an adaptive two-layer network.

tions are needed in the first layer and S connections in the second layer of the local block structure. Compared to the standard structure, the second layer neurons in all local blocks are separated from one another. Therefore, not only the number of neurons can be minimized, but also the number of synapses can be greatly reduced. Consequently, the computation referred to each block during training can be shortened substantially. In addition, the local block structure can be easily concatenated to build a large adaptive network, which is especially useful for radical-partitioned Chinese characters recognition. Furthermore, each local block needs to consider convergence associated with only a portion of input patterns during training. Consequently, the block structure converges faster during training since it is much easier to identify different patterns in a multidimensional space. Compared to the standard adaptive neural network system, the local block system already shows a substantial improvement in enhancing the learning speed as a result of the property that each local block can be trained independently and the convergence can be more easily reached since a much smaller structure—a local block—is involved during training.

However, the block adaptive network structure requires n^2 local blocks, which is still a bottleneck for large-scale pattern recognition. A coded block system (Kuo et al., 1991; Kuo & Mao, 1991) as shown in Figure 3 has been used to further improve the block system.

Instead of n^2 local blocks, only less than n local blocks with outputs coded corresponding to the input patterns are used in the system. Using the coded block system, the training time can be shortened further owing to the drastic reduction in the number of local blocks. The coded local block system is suitable for training the system with small-volume Chinese characters. However, for training over 200 Chinese characters, the coded block adaptive neural network using back propagation algorithm suffers from convergence difficulties.

LOCAL BLOCK STRUCTURE

Local Block Structure

FIGURE 2. A block adaptive neural network structure (top). The local block structure (bottom).

Coded Local Block Structure

FIGURE 3. The coded block adaptive neural network structure.

Although there are over 20,000 Chinese characters, they can be categorized into about 230 radical groups. Each radical group has 5–200 characters containing an identical radical portion as shown in Figure 4a. Taking advantage of the radical properties, the coded block adaptive neural network system can be reconfigured as shown in Figure 4b such that each coded block is assigned to be responsible for training Chinese characters having the same radical portion. Using a radical-partitioned structure, as shown in Figure 4b, the coded block adaptive neural network can overcome the convergence difficulties during training since training of the large-volume Chinese characters can be divided into small procedures associated with radical blocks. In fact, the training of each radical block can be divided further into steps associated with internal local blocks. Therefore, using the coded block system with a radical-partitioned structure, the overall training procedure is simplified from two approaches. First, Chinese characters as input training patterns are categorized into several radical groups. Chinese characters in each radical group having an identical radical portion is assigned to a radical block. Second, the training of each radical block is divided into steps associated with several internal local blocks. Training of each internal local block

Coded Block Adaptive Neural Network System with a Radical-Partitioned Structure

(b)

(c)

FIGURE 4 (a). The radical of Chinese characters, (b) the coded block adaptive neural network system with a radical-partitioned structure for large-volume Chinese characters recognition, (c) the normalized dot products method used in the radical selector.

is independent. Therefore, convergence during training is determined by a very small portion of the overall system—the internal local block in each radical block with a small portion of the Chinese characters to train. Therefore, convergence is guaranteed and training time is much shorter.

As shown in Figure 4b, above the coded radical blocks, there is a radical selector, which is used to identify the radical category of each input Chinese character during recognition. As shown in Figure 4c, during recognition in the radical selector block, dot products of each input Chinese character vector ($\vec{I}$) and all radical vectors ($\vec{R}$'s), ($\vec{I} \cdot \vec{R}$) are carried out, followed by normalization by the norm of the radical vector ($(\vec{I} \cdot \vec{R})/|\vec{R}|$). Then, according to the highest value of the normalized dot products, a decision is made on the input character about which radical group it belongs to. Then, the input character is transferred to the associated radical block for further processing. Since only one radical block handles the second-step processing during recognition, the recognition time is greatly shortened as compared to the coded block adaptive neural network system without the radical-partitioned structure. Consequently, the coded block adaptive neural network system with the radical-partitioned structure can be advantageous for both training and recognition of large-volume Chinese characters.

3. THE BACK PROPAGATION ALGORITHM

Back propagation algorithm (Jones & Hoskins, 1987; Vogl et al., 1988; Yamada et al., 1989) has been used in the coded block adaptive neural network structure. The objective of the back propagation algorithm is to establish the desired responses for the neurons by propagating back the output error signal through the network to modify the weight values. First, present a simple pattern to the inputs of the network primed with random initial weights. Change the second layer weights w_{kj} according to:

$$\Delta w_{kj}(n) = \eta \delta_j h_k + \alpha \Delta w_{kj}(n-1),$$

where η is the learning rate, α is the smoothing factor, the δ_j is defined as $\delta_j = u_j(1 - u_j)(t_j - u_j)$ with the second layer, and desired and actual outputs (t_j , u_j) and h_k will be defined below. After computing δ_j in the output layer, the first layer values δ_k^* can be obtained:

$$\delta_k^* = h_k(1 - h_k) \sum_j \delta_j w_{kj},$$

where

$$h_k = \frac{1}{1 + e^{-\sum_i v_{ik} s_i}}.$$

The weight v_{ik} associated with the synapse between input s_i and the first layer k -th neuron is:

$$\Delta v_{ik}(n) = \eta \delta_k^* s_i + \alpha \Delta v_{ik}(n-1).$$

Several training epoches of pattern activation and error feedback are performed with all the training patterns until the output of each local block for all input patterns is correct.

Applying the back propagation algorithm, the adaptive neural network with the local block structure has advantages in learning time. The intrinsic better properties of the local block system can be reasoned as follows. Assume that $\vec{x}_1, \dots, \vec{x}_n$ are n dimensional input patterns and $\vec{y}_1, \dots, \vec{y}_n$ are corresponding outputs. For a standard neural network structure without local blocks, after training, a common function, f , representing the trained weights can be obtained such that:

$$f: \vec{x}_1 \rightarrow \vec{y}_1$$

$$f: \vec{x}_2 \rightarrow \vec{y}_2$$

⋮

$$f: \vec{x}_n \rightarrow \vec{y}_n.$$

On the other hand, in the system with the local block structure, a simpler common function, g , can be identified for each block.

$$g: \vec{x}_1 \rightarrow (1 \text{ or } 0)$$

$$g: \vec{x} \rightarrow (1 \text{ or } 0)$$

⋮

$$g: \vec{x}_n \rightarrow (1 \text{ or } 0).$$

Based on the above analysis, obviously a common function, g , can be obtained much more easily for the local block systems. Therefore, the local block structure can have a quicker convergence during training.

4. SIMULATION RESULTS AND DISCUSSION

In order to investigate the potential of the coded block adaptive neural network system with a radical-partitioned structure for large-volume Chinese characters recognition, the performance in terms of learning time, reconfiguration based on simulation results has been obtained. One thousand Chinese characters composed of 24×24 lattice dots in 25 frequently-used radical groups as shown in Figure 5 have been used as the training patterns for the coded block adaptive neural network system with a radical-partitioned structure. Each radical block is trained with its 40 associated Chinese characters independently. Each of the radical blocks contains 16 internal local blocks. In the coded block adaptive neural network system with a radical-partitioned structure, a learning rate, η , which affects the learning speed of a block adaptive neural network system using back-propagation algorithm (Chen et al., 1990; Kuo et al., in press), of 0.2 has been used. In addition, the number of first layer neurons in the local block structure, which also affects the learning time, is defined as 12.

FIGURE 5. The 1,000 Chinese characters used as the training patterns.

The coded block adaptive neural network system with a radical-partitioned structure has an excellent expansion capability for large-volume Chinese characters recognition. Depending on applications, the radical block associated with Chinese characters belonging to the same radical group can be trained sequentially and glued together with other radical blocks. Conse-

TABLE 1
Numbers of Epoches Needed for Training Local Blocks in all Radical Blocks

Rad	Block															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Rad1	0	35	51	77	84	26	65	25	112	80	14	25	27	35	35	47
Rad2	0	23	23	28	32	43	55	23	30	37	18	40	70	11	23	22
Rad3	0	121	31	21	33	36	27	52	28	21	36	33	28	44	89	11
Rad4	0	63	155	142	69	144	326	176	47	75	33	35	23	19	31	46
Rad5	0	198	24	24	50	62	25	102	72	44	80	61	18	47	118	20
Rad6	0	475	168	113	85	35	24	116	305	213	77	259	24	81	36	17
Rad7	0	0	0	16	42	24	15	53	29	13	31	51	14	102	31	28
Rad8	0	38	33	45	43	210	258	33	86	166	69	35	41	31	14	56
Rad9	0	66	34	202	28	161	91	250	50	28	83	122	44	99	36	12
Rad10	0	21	27	133	28	162	79	12	15	18	121	29	25	91	33	20
Rad11	0	60	63	87	186	25	132	106	171	148	27	18	20	32	9	128
Rad12	0	132	193	187	73	285	101	258	40	56	32	40	31	58	29	35
Rad13	0	16	20	20	66	73	53	31	99	0	52	18	41	39	41	23
Rad14	0	81	52	154	34	28	202	69	57	30	20	18	11	74	50	60
Rad15	0	0	0	31	23	29	16	90	21	60	38	37	19	28	31	75
Rad16	0	11	8	11	12	119	49	44	73	14	39	60	14	36	25	24
Rad17	0	0	0	106	150	38	31	23	42	75	44	55	36	70	13	57
Rad18	0	31	430	110	93	52	67	31	41	40	33	160	50	72	19	91
Rad19	0	60	60	33	74	81	26	19	147	60	48	64	35	160	87	24
Rad20	0	0	0	70	34	432	109	97	55	34	117	48	64	16	180	22
Rad21	0	26	40	112	31	174	26	102	32	43	51	39	79	50	187	36
Rad22	0	0	0	20	155	22	15	27	58	274	57	39	12	33	32	55
Rad23	0	44	72	71	40	83	16	19	19	13	11	26	64	144	49	137
Rad24	0	86	90	26	42	26	58	42	25	132	20	155	165	33	41	149
Rad25	0	43	48	81	403	139	25	30	41	92	24	115	16	91	132	27
Ave.	0	65	65	77	76	100	76	73	68	72	47	63	39	60	55	49

quently, in building a large Chinese characters system, the training process does not have to start from the very beginning every time. Instead, the previous trained radical blocks, which relate to the Chinese characters

of identical radicals, can be used. Only the new untrained characters need to be trained. Consequently, the cost in training can be greatly reduced. With the radical-partitioned structure, radical coded blocks can

FIGURE 6. The average learning time for each of the 16 internal local blocks of all 25 radical blocks in the coded block adaptive neural network with a radical-partitioned structure using the back propagation algorithm. Twenty-five radical groups of 1,000 Chinese characters have been used as the training patterns. Each radical block has 40 Chinese characters as the training patterns.

FIGURE 7. The reconfigurability of the code block neural network system with a radical-partitioned structure. The average output percentage error versus input percentage error curves for the system using input samples taken randomly from several radicals. Each input sample is with random errors. The total bits of random errors out of the 24×24 lattice dots in each input sample are expressed as the percentage random error.

be concatenated to build a large coded block adaptive neural network system for large-volume Chinese characters recognition. In addition, the convergence of training the growing system is always guaranteed since each time only the newly-added characters belonging to the same radical group need to be trained. Furthermore, taking advantage of the expansion capability of the system, recognition of multifont printed Chinese characters or hand-written Chinese characters can be expedited via adding extra local radical blocks of various printed fonts and various possible hand-written styles.

5. CONCLUSION

In this paper, an efficient coded block adaptive neural network system with a radical-partitioned structure suitable for large-volume Chinese characters recognition, has been successfully trained with 1,000 Chinese characters in 139.2 hours CPU time on a 18 MIPS computer. According to simulation results, the coded block adaptive network with a radical-partitioned structure provides an acceptable learning time, a good reconfigurability, and an excellent expansion capability.

REFERENCES

Chen, C. C., Hsiao, C. C., & Kuo, J. B. (1990). A structured adaptive neural network for pattern recognition. Digest of 1990 International EDMS.

- Denker, J. S., & Wittner, B. S. (1987). Network generality, training required and precision required. *IEEE Conference Neural Information Processing Systems*. American Institute of Physics Press, New York.
- Feldman, J., & Ballard, D. (1982). Connectionist models and their properties. *Cognitive Science*, 6, 205-254.
- Kuo, J. B., Chen, C. C., & Hsiao, C. C. (in press). A block adaptive neural network structure for pattern recognition VLSI. *IEE Proceedings on Devices, Circuits and Systems*.
- Kuo, J. B., Chen, Y. K., Lu, Y. H., & Mao, W. C. (1991). A coded block adaptive block neural network structure for pattern recognition VLSI. Digest of 1991 International Symposium on VLSI Tech., Systems, and Applications.
- Kuo, J. B., & Mao, W. C. (1991). An adaptive neural network structure with coded local blocks for pattern recognition VLSI. Proc. of 1991 IJCNN.
- Le Cun, Y., Guyon, I., Jackel, L. D., Henderson, D., Boser, B., Howard, R. E., Denker, J. S., Hubbard, W., & Graf, H. P. (1989, November). Handwritten digit recognition: Application of neural network chips and automatic learning. *IEEE Communications Magazine*, 41-46.
- Mirchandani, G., & Cao, W. (1989). On hidden nodes for neural nets. *IEEE Transactions on Circuits and Systems*, 36, 661-664.
- Rumelhart, D. E., & McClelland, J. (1986). Parallel Distribution Processing. *Massachusetts Institute of Technology: Darpa Neural Network Study*. October 1987-February 1988.
- Vogl, T. P., Mangis, J. K., & Rigler, A. K. (1988) Accelerating the convergence of the back propagation method. *Biological Cybernetic* 59, 257-263.
- Widrow, B., Winter, R. G., & Baxter, R. A. (1988). Layered neural nets for pattern recognition. *IEEE Transactions on Acoustics Speech Signal Processing*, 36, 1109-1117.
- Yamada, K., Kami, H., Tsukumo, J., & Temma, T. (1989). Hand-written numeral recognition by multi-layered neural network with improved learning algorithm. *International Joint Conference of Neural Networks (IJCNN)*, II-259-II-266.