A SCHEME OF ON-LINE CHINESE CHARACTER RECOGNITION USING NEURAL NETWORKS

Hwei-Jen Lin and Shwu-Huey Yen
Department of Computer Science and Information Engineering
Tamkang University
Tamsui, Taiwan
R.O.C.
email: hjlin@cs.tku.edu.tw

ABSTRACT

This paper proposes a scheme of on-line Chinese character recognition, based on neural networks. The supervised backpropagation algorithm is used to train the network. The input character is converted as a sequence of virtual stroke segments as well as real stroke segments, which is a good feature exactly describing the complete structure of a character, and is to be extracted by our system.

In order to simplify recognition process and reduce the recognition time, the neural network is divided into several subnetworks, each of them is responsible for recognizing a group of about 75 character patterns. In other words, the huge set of Chinese characters is divided into several groups according to the numbers of stroke segments in the characters, and for each group of characters, a specific subnetwork is trained in order to recognize every character in the group.

Whenever the system accepts an input Chinese character, it will calculate the number of stroke segments, including virtual stroke segments as well as real stroke segments in that character, and then determine which subnets to enter for recognition process.

The system is allowed to accept and recognize some interconnected characters. The algorithm was experimentally implemented in a personal computer system, accepts interconnected Chinese characters written on an electronic tablet, and performs recognition in real time. Our experiment showed that recognition accuracy exceeded 96% on the test example.

Key words: on-line recognition, neural network, backpropagation, character pattern, feature extraction

1. INTRODUCTION

Since 1966, numerous methods for on-line handwritten English character recognition have been proposed [2,3,4,8,10,17,18,19]. Since then various methods of on-line handwritten Chinese character recognition have been also proposed [1,5,6,7,11,12,13,15], but they do not always have a recognition rate sufficient for a large set of characters. Error in recognition by these methods seems to be caused by the unstability of geometric features of characters.

Neural network models have a great degree of fault tolerance, this motivates us to use them as the base for our system. Most of character recognition systems based on neural networks take digit patterns as features of characters [9,14,16] (or apply the digit patterns to the neural nets’ input), for example, an image of 64 × 64 pixels for a Chinese character, which need large storage. To solve this problem of storage occupancy, we choose stroke segments (including both real and virtual stroke segments) as features of characters.
tion part. If the maximal value of the outputs for the neurons on the output layer is greater than a given threshold (= .5), the decision layer determines the character corresponding to the neuron with such a maximal output value as the answer for the recognition; otherwise, the resulting answer is "rejection".

4. EXPERIMENTAL RESULTS

The proposed scheme is composed of several neural subnetworks, each of them is responsible for recognizing about 75 character patterns.

Our experiment tested on a trained subnetwork, which is able to recognize 76 Chinese character patterns. The experimental results showed that the recognition rate is 100% on the training patterns and 96% on the test patterns.

The recognition was performed on a PC-486 DX22-66 with algorithms implemented in Visual Basic. The average recognition speed is 3 characters/sec.

5. CONCLUSIONS AND EXTENSIONS

We describe an on-line recognition system for handwritten Chinese characters. As soon as a character is entered through an electronic tablet, the preprocessor extracts stroke segments in the written character from the pen-tip movement loci on the tablet and normalizes the length and angle of each stroke segment in the character. Thus, the output of the preprocessor is a sequence of these normalized values, which forms a feature vector of the input character.

The system is composed of many multi-layered feed-forward neural networks, each of them is trained separately to recognize a group of character patterns with the same number of stroke segments.

Each pattern concerned by the system corresponds to a stroke order of a Chinese character. We only consider the frequently encountered stroke orders for a character, and regard them as different patterns for our system. However, the system is unable to recognize characters written with an unusual sequence (or order) of strokes, and an increment on number of patterns will enlarge our system. One alternative is to utilize a method of stroke segment ordering, which is in our further research upon this proposed scheme. Such a method provides stable order independent of the writing order of a character. On the basis of this method we may overcome the different writing order problem.

6. REFERENCES


