Farsi handwritten digit recognition based on mixture of RBF experts

Reza Ebrahimpour\textsuperscript{1,2a}, Alireza Esmkhani\textsuperscript{2}, and Soheil Faridi\textsuperscript{2,3}
\textsuperscript{1} School of Cognitive Sciences (SCS), Institute for Research in Fundamental Sciences (IPM), Niavaran, P.O.Box:19395-5746, Tehran, Iran
\textsuperscript{2} Brain and Intelligent Systems Research Lab, Department of Electrical Engineering, Shahid Rajaee Teacher Training University Tehran, P.O.Box 16785-163, Iran
\textsuperscript{3} Islamic Azad University, South Tehran Branch, No. 3, 9th Nayestan, Pasdaran Avenue, Tehran, Iran
\textsuperscript{a}) ebrahimpour@ipm.ir

Abstract: In this paper, a new classifier combination model is presented for Farsi handwritten digit recognition. The model is consisted of four RBF neural networks as the experts and another RBF network as the gating network which learns to split the input space between the experts. Considering the input data, which is an 81-element vector extracted using the loci characterization method, the gating network assigns a competence coefficient to each expert. The final output is computed as the weighted sum of the outputs of the experts. The recognition rate of the proposed model is 93.5\% which is 3.75\% more than the rate of the mixture of MLPs experts previously ran on the same database.

Keywords: mixture of experts, handwritten digit recognition, loci characterization method

Classification: Science and engineering for electronics

References

1 Introduction

It is now about 50 years that research on English handwritten characters and digits is being performed. Regarding Farsi handwritten digit recognition, numerous researches have been done during the recent years which are mainly based on trainable classifiers and feature extraction methods. Different methods have been proposed for character and digit recognition among which zonal features, geometric moments, zernic moments, Fourier descriptors, invariant moments and loci characterization are the most famous ones. Generally, the type of the feature to be used is application-dependent and an experimental analysis of the input data can determine the appropriate features. In this paper, the loci characterization method is used for handwritten digit recognition [1].

In pattern recognition, combining classifiers is one of the popular approaches. The reason is the final goal to have a better classification task, increase the recognition rate and improve the system reliability. Combining classifiers is usually a good solution to complex problems which can be due to the small sample size, class overlapping, dimensionality, and substantial noise in the input samples. Previous experimental and theoretic results show that combining some classifiers is more useful when the base classifiers are of small error rates and also their errors are different [2]. In other words, the base classifiers should make uncorrelated decisions. In this paper, a combining classifiers scheme based on mixture of MLPs experts [3] is employed to increase the recognition rate of Farsi handwritten digits.

2 Loci characterization feature

Employing loci characterization method has yielded acceptable results for digit recognition task. This feature is usually defined toward horizontal and vertical or 45 and 135 degree directions. The loci characterization feature vector for each image is defined by assigning a number to each background point in the image. This number is computed using the number of the points intersecting the body of the digit, and the vertical and horizontal lines drawn from the background point toward four directions (up, down, right and left). If we limit the number of intersections with the body to two, then we can express each background point using a four-digit number of base three. As the result, the loci characterization feature vector consists of 81 elements each of which shows the relative frequency of their corresponding number in the input image. To achieve a normalized feature vector, its values are divided by the total number of white points of the background. Figure 1 (a) shows the way to calculate the loci characterization feature vector for a point around digit “seven”.

3 Mixture of RBF experts

The idea of the mixture of experts method is based on the divide and conquer principle in which a complex problem is split into some simple problems so that the final result is the mixture of the small simple problems’ solutions.
Mixture of experts, is one of the most well-known combining methods in which the input signal directly involves in a mechanism that converts the base classifiers outputs into a single output. Considering Figure 2, we can say that the learning process here, is a combination of both supervised and self-organized methods. The experts are involved in a supervised model in which the outputs are later combined to form the desired output.

The base classifiers engage in a self-organized learning process. That is, they find an acceptable partition of the feature space themselves so that they become expert in that small subspace. This way, the base classifiers, cover all the input problem space. In the original form of mixture of experts (firstly proposed by Jacobs et al. in [3]), the base classifiers and also the gating network were linear networks, while to solve more complicated problems, the simple linear networks would not suffice. In this paper, to improve the efficiency of the base classifiers and as the resultant the performance of the whole network, RBF is used as the structure of the base classifiers and the gating network.

In the first stage, using the K-means clustering method, we have split the training data into four clusters and the centers of these clusters are selected as the centers of the gating network’s hidden layer neurons. Each base classifier is an RBF network with one hidden layer and the output of each classifier $O_i$ is determined considering the input vector $x$, number of hidden layer neurons,
weights of the output layer and a sigmoidal activation function. The gating network is made up of two different parts: an RBF network and a softmax nonlinear operator.

The output of this softmax layer $g_i$ is computed using $O_g$, which is the output of the gating network’s RBF layer (Figure 2). The output of classifier $i$, the input vector $x$, and some of the parameters like the weights of the output layer and also the sigmoidal activation function, all affect the weight $g_i$. In fact $g_i$ is the probability of producing the desired output $y$ by the base classifier $i$ in the softmax layer:

$$g_i = \frac{e^{O_{gi}}}{\sum_{i=1}^{N} e^{O_{gi}}}, \quad i = 1, \ldots, N$$

in which $N$ is the number of base classifiers. $g_i$ are non-negative an sum to one. The total output of the network is calculated using the outputs of the base classifiers and the gating network:

$$O_T = \sum_i O_i g_i, \quad i = 1, \ldots, N$$

The weights of the output layer of the RBF network are updated using the back propagation algorithm. For the base classifier $i$ and the gating network we have:

$$\Delta W_y = \eta_e h_i (y - O_i) (O_i (1 - O_i)) \varphi_e(x_i)$$

$$\Delta W_{yg} = \eta_g (h - g) (O_g (1 - O_g)) \varphi_g(x_i)$$

in which $\eta_e$ and $\eta_g$ are the learning rates of the output layer of the base classifiers and the gating network, respectively and $W_y$ and $W_{yg}$ are the weights of the output layer of the base classifiers and the gating network. $h_i$ is an estimation which can be interpreted as the error probability of base classifier $i$ in computing the desired target $y$ for the next iteration.

$$h_i = \frac{g_i e^{-\frac{1}{2} (y - o_i)^T (y - o_i)}}{\sum_j g_j e^{-\frac{1}{2} (y - o_j)^T (y - o_j)}}$$

$\varphi_e(x_i)$ and $\varphi_g(x_i)$ are the hidden layer outputs of the base classifiers and the gating network, respectively and are produced using the following Gaussian function:

$$d(x_i, c_j) = ||x_i - c_j|| = \left( \sum_{u=1}^{p} |x_{iu} - c_{ju}|^2 \right)^{1/2}$$

$$\varphi_j(x_i) = \exp \left( -\frac{d(x_i, c_j)^2}{2\sigma_j^2} \right), \quad j = 1, 2, \ldots, K; \quad i = 1, 2, \ldots, Q$$

in which $x_i$ is the $i$th input vector, $p$ is the input dimension and $c_j$ is the center of the $j$th neuron in the hidden layer which are obtained using the K-means clustering method. $\sigma_j$ is the Euclidean distance between the center of the $j$th neuron in the hidden layer and the nearest center of its neighborhood.
which is later multiplied by a constant $\beta$. The optimal value for $\beta$ is larger than one and is determined during experiments.

$$
\sigma_j = \beta \times \min ||c_j - c_n||, \quad n, j = 1, 2, 3, \ldots, K \text{ and } n \neq j
$$

$$
O_{ik} = \sum_{j=0}^{K} [\varphi_j(x_i)w_{kj}], \quad k = 1, 2, 3, \ldots, m
$$

in which $w_{kj}$ is the weight between the hidden layer’s $j^{th}$ unit and the output layer’s $k^{th}$ unit. $\varphi_0(x_i)$ is the bias which is equal to one and $w_{k0}$ is the weight between the bias neuron and the $k^{th}$ output neuron. $M$, is also the total number of the base classifiers neurons. In learning, the base classifiers compete to learn the inputs and the gating network lets the winner to update its weights more than the others.

4 Experimental results

In order to provide a benchmark for evaluating the proposed method for Farsi handwritten digit recognition, we also implemented the mixture of MLPs experts on the same database. Our database consists of 8000 samples which are the hard samples of the Hoda database [4] with 80,000 samples. Figure 1 (b) shows some of the digits of the utilized database. As the Hoda database is gathered using the universities’ entrance exam forms, we consider it as an easy database (because students pay so much attention when completing such forms). So, in order to extract the harder samples, using the K-nearest neighbors method with K = 5, we selected the hard data which were classified into more than three classes. This way we gathered 8000 hard samples. For the training stage, 6000 samples were used and the rest 2000 were employed for test.

Table I shows the best topologies for the proposed method as well as the mixture of MLPs experts. It is to be mentioned that backpropagation algorithm is used to update the weights in the proposed model as well as the mixture of MLPs experts. The recognition rate of the proposed model is 95.3% which is 3.75% more than the recognition rate of the mixture of MLPs experts applied to the same database.

<table>
<thead>
<tr>
<th>Network Topology</th>
<th>Mixture of RBF Experts</th>
<th>Mixture of MLPs Experts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Base Classifiers</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Learning Rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experts</td>
<td>0.28</td>
<td>0.11</td>
</tr>
<tr>
<td>Gating</td>
<td>0.1</td>
<td>0.07</td>
</tr>
<tr>
<td>Recognition Rate</td>
<td>95.3 %</td>
<td>91.45 %</td>
</tr>
</tbody>
</table>
5 Conclusions

Considering the mentioned results in the previous section, we can claim that the proposed mixture of RBF experts method has more reliability and better performance in comparison with the mixture of MLPs method. Thus, utilizing a nonlinear sigmoidal function, the RBF neural network maps the input feature space to a fairly separable feature space which is linear to some extent. All in all, using the gating network to break the problem space into smaller partitions gives the base classifiers more generalization power in their specified feature space.

Acknowledgements

The authors wish to thank Iran Telecommunication Research Center (ITRC) for funding this project under contract number 500/7031.