Proceedings of 1993 International Joint Conference on Neural Networks

A Multiple Network Architecture Combined by Fuzzy Integral*

Sung-Bae Cho[†] and Jin H. Kim

Center for Artificial Intelligence Research and Computer Science Department
Korea Advanced Institute of Science and Technology
373-1, Koosung-dong, Yoosung-ku, Taejeon 305-701, Republic of Korea

April 29, 1993

Abstract

Recently, in the area of artificial neural network, the concept of combining multiple networks has been proposed as a new direction for the development of highly reliable neural network systems. In this paper we propose a method for multinetwork combination based on the fuzzy integral. This technique nonlinearly combines objective evidence, in the form of a fuzzy membership function, with subjective evaluation of the worth of the individual neural networks with respect to the decision. The experimental results with the recognition problem of on-line handwriting characters show that the performance of individual networks could be improved significantly.

1 Introduction

Multiple networks are desirable due to the basic fact that selection of the weights w is an optimization problem with many local minima. All global optimization methods in the face of many local minima yield "optimal" parameters (w) which differ greatly from one run of the algorithm to the next, i.e., which show a great deal of randomness stemming from different initial points (w^0) and sequencing of the training examples. This randomness tends to differentiate the errors of the networks, so that the networks will differ in the values of the weights w. These different weights correspond to different ways of forming generalizations about the patterns inherent in the training set. As each network makes generalization errors on different subsets of the input space, we shall argue that the collective decision produced by the multiple networks is less likely to be in error than the decision made by any of the individual networks [1].

In this paper, we present a multiple neural network architecture combined by a new evidence fusion technique, based on the notion of the fuzzy integral. In the fuzzy integral both objective evidence supplied by various sources and the expected worth of subsets of these sources are considered in the fusion process: It combines objective evidence for a hypothesis with the system's expectation of the importance of that evidence to the hypothesis. This approach may provide a possibility for incorporating any *a priori* knowledge regarding the underlying problem to improve the ability of the network to generalize.

2 Multiple Neural Networks Classifier

A neural network can be considered as a mapping device between a set of input and a set of output. Mathematically speaking, a neural network represents a function F that maps I into O: $F : I \to O$, or $y = f(x)$ where $y \in O$

*This work was supported in part by a grant from the Korea Science and Engineering Foundation (KOSEF).

[†]E-mail: sbcho@gorai.kaist.ac.kr.

and $x \in I$. By the way, since the classification problem is a mapping from the feature space to some set of output classes, we can formalize the neural network, especially two-layer feedforward neural network trained with the generalized delta rule, as a classifier.

This kind of network trains on a set of example patterns and discovers relationships that distinguish the patterns. A network of a finite size, however, does not often load a particular mapping completely or it generalizes poorly. Increasing the size and number of hidden layers most often does not lead to any improvements. Furthermore, in complex problems such as character recognition, both the number of available features and the number of classes are large. The features are neither statistically independent nor unimodally distributed. Therefore, if we can make the network consider the only specific part of the complete mapping, it will perform it's job better.

The basic idea of the multiple network scheme is to develop N independently trained neural networks with relevant features, and to classify a given input pattern by obtaining a classification from each copy of the network and then using a consensus scheme to decide the collective classification by utilizing combination methods [2, 3].

In this scheme, the output value of each neuron is taken to the estimated *a posteriori* probability of the training samples belonging to that class. (see Fig. 1.) Then it naturally raises the question of obtaining a consensus on the results of each individual network or expert.

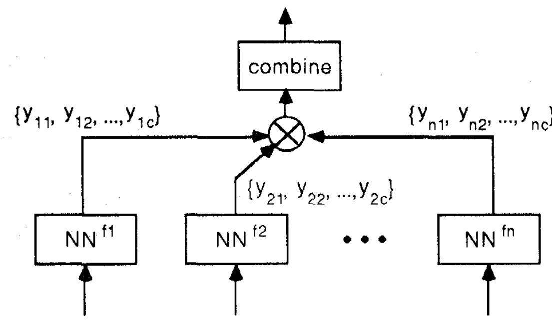

Figure 1: The multiple network architecture with consensus scheme.

3 Network Integration with Fuzzy Integral

The fuzzy integral is a nonlinear functional that is defined with respect to a fuzzy measure, especially g_λ-fuzzy measure introduced by Sugeno [4, 5]. Let $Y = \{y_1, y_2, \ldots, y_n\}$ be a finite set and let $h : Y \rightarrow [0, 1]$ be a function. Suppose $h(y_1) \geq h(y_2) \geq \ldots \geq h(y_n)$, (if not, Y is rearranged so that this relation holds). Then a fuzzy integral, e, with respect to a fuzzy measure g over Y can be computed by

$$e = \max_{i=1}^{n} [\min (h(y_i), \ g(A_i))] \qquad (1)$$

where $A_i = \{y_1, y_2, \ldots, y_i\}$.

Note that when g is a g_λ-fuzzy measure, the values of $g(A_i)$ can be determined recursively as

$$\begin{aligned} g(A_1) &= g(\{y_1\}) = g^1 & (2) \\ g(A_i) &= g^i + g(A_{i-1}) + \lambda g^i g(A_{i-1}). & (3) \\ &\quad \text{for } 1 < i \leq n. & (4) \end{aligned}$$

λ is given by solving the equation

$$\lambda + 1 = \prod_{i=1}^{n} (1 + \lambda g^i) \qquad (5)$$

where $\lambda \in (-1, +\infty)$, and $\lambda \neq 0$. This can be easily calculated by solving an $(n-1)$st degree polynomial and finding the unique root greater than -1.

Thus the calculation of the fuzzy integral with respect to a g_λ-fuzzy measure would only require the knowledge of the density function, where ith density, g^i, is interpreted as the degree of importance of the source y_i towards the final evaluation. The value obtained from comparing the evidence and the importance in terms of the min operator is interpreted as the grade of agreement between real possibilities,

$h(y)$, and the expectations, g. Hence fuzzy integration is interpreted as searching for the maximal grade of agreement between the objective evidence and the expectation.

Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_c\}$ be a set of classes of interest. Note that each ω_i may, in fact, be a set of classes by itself. Let $Y = \{y_1, y_2, \ldots, y_n\}$ be a set of neural networks for the recognition of a particular class, ω_i, $1 \leq i \leq n$. Let A be the object under consideration for recognition. Let $h_k : Y \rightarrow [0, 1]$ be the partial evaluation of the object A for class ω_k, that is, $h_k(y_i)$ is an indication of how certain we are in the classification of object A to be in class ω_k using the group y_i where a 1 indicates absolute certainty that the object A is really in class ω_k and 0 implies absolute certainty that the object A is not in ω_k.

Now corresponding to each y_i the degree of importance, g^i, of how important y_i is in the recognition of the class ω_k must be given. These densities can be subjectively assigned by an expert, or can be generated from training data. The g^i's define the fuzzy density mapping. Hence λ can be calculated using (5) and the g_λ-fuzzy measure, g can be constructed. Now, using (1) to (5), the fuzzy integral can be calculated. Thus the following algorithm for network integration is given.

Algorithm : Network fusion by fuzzy integral

```
calculate λ; /* importance of each net */
for each class ω_k do
    for each neural network y_i do
        calculate h_k(y_i);
        determine g_k({y_i});
    end_for
    compute the fuzzy integral;
end_for
determine the final class;
```

In the final step, the class ω_k with the largest integral value can be chosen as the output class.

4 Experimental Results

In order to give an idea of the practical application of the presented method in pattern recognition, a data set of handwriting characters has been used as a source of training and test patterns. Handwriting characters were inputted to the computer (SUN workstation) by an LCD tablet of Photron FIOS-6440 which samples 80 dots per second. The tasks were to classify the Arabic numerals, the uppercase letters, and the lowercase letters which were collected from 13 writers.

An input character consists of a set of strokes, each of which begins with a pen-down movement and ends with pen-up movements. Several preprocessing algorithms were applied to successive data points in a stroke to reduce quantization noises and fluctuations of the writer's pen motion. A sequence of preprocessed data points is approximated by a sequence of 8-directional straight-line segments.

To evaluate the performance of the proposed network scheme, we implemented three different networks, each of which is a two-layered neural network having different number of input neurons and 20 hidden neurons. NN_1, NN_2 and NN_3 have 10 input neurons, 15 input neurons, and 20 input neurons, respectively. In this fashion each network makes the decision through its own resolution; NN_1 using sparsely sampled input produces the result by means of coarser view of input image, while NN_3 uses finer view. Thus, NN_1 has large possibility to overcome the variation or noise of input image though it utilizes rather blurred input.

Each of the three networks was trained with 40 samples per class, validated with another 500 samples, and tested on ten sets of samples collected from different ten writers: the recognition rate on the validation set was monitored in order to stop the training process. For all of the following experiments, each consisted of ten trials in which the different data were made from different writers.

We assigned the fuzzy densities g^i, the degree of importance of each network, based on how good these networks performed on valida-

Table 1: Fuzzy densities (recognition rates on the training data) and the corresponding λ.

Subject	g^1	g^2	g^3	λ
Numeral	0.3430	0.3330	0.3230	−0.0149
Large	0.3447	0.3312	0.3240	0.0003
Small	0.3370	0.3321	0.3312	−0.0009

Table 2: Average recognition rates of the three individual networks, and fusion methods (average, maximum, and the fuzzy integral:%).

Networks	Numeral	Large	Small
NN_1	82.69	68.10	73.95
NN_2	81.25	68.63	71.80
NN_3	81.05	66.26	72.15
Avg	86.94	75.27	78.25
Max	85.72	71.37	76.90
Fuzzy	88.11	76.16	80.35

tion data. We computed these values as follows:

$$g^i = \frac{p_i}{\sum_j p_j} \cdot d\text{sum}. \tag{6}$$

where p_i is the performance of network NN_i for the validation data and $d$sum is the desired sum of fuzzy densities. The real values of these densities with the corresponding λ are shown in table 1.

Table 2 shows the recognition rates of numerals, uppercase letters, and lowercase letters with respect to the three individual networks and their combinations by utilizing the fusion methods.

In this table we can learn that the recognition rates of the consensus methods outperformed those of the individual networks. It is also seen from the table that the mean recognition rates of the proposed method are higher than those of the other conventional methods. From the statistical comparisons based on a t test with 9 degrees of freedom, "no-improvement" hypothesis is rejected at a 5% level of significance for all the cases.

5 Concluding Remarks

In this paper, we have presented a design method of the multilayer neural network, called the multiple network scheme, and proposed a fusion method based on the fuzzy integral. One of the important advantages of this method is that not only is the classification results combined but that the relative importance of the different networks is also considered.

Initial trials to use the method for classifying a large set of on-line handwriting characters were promising, but several works are remained for further research. The relatively easy ones are to increase the recognition rate of each base neural network for practical usage and to try the same experiments with the increased number of networks. Furthermore, another fusion methods such as Dempster-Shafer method may be applied to combine the decisions of the networks.

References

[1] F. Fogelman Soulie, "Neural network architectures and algorithms: a perspective," *Artificial Neural Networks*, 605–615 (Netherlands, Elsevier Science Publishers B.V.), 1991.

[2] L.K. Hansen, and P. Salamon, "Neural network ensembles," *IEEE Trans. Pattern Analysis and Machine Intelligence*, **12**, 993–1001, 1990.

[3] S. Shlien, "Multiple binary decision tree classifiers," *Pattern Recognition*, **23**, 757–763, 1990.

[4] M. Sugeno, "Fuzzy measures and fuzzy integrals: A survey," *Fuzzy Automata and Decision Processes*, Amsterdam: North Holland, 89–102, 1977.

[5] H. Tahani, and J.M. Keller, "Information fusion in computer vision using the fuzzy integral," *IEEE Trans. on Systems, Man, and Cybernetics*, **20**, 733–741, 1990.