Proceedings of the 6th WSEAS Int. Conf. on NEURAL NETWORKS, Lisbon, Portugal, June 16-18, 2005 (pp146-151)

Combination Methods for Ensembles of Multilayer Feedforward[1]

JOAQUÍN TORRES-SOSPEDRA
MERCEDES FERNÁNDEZ-REDONDO
CARLOS HERNÁNDEZ-ESPINOSA
Dept. de Ingeniería y Ciencia de los Computadores
Universidad Jaume I
Avda. Vicente Sos Baynat s/n 12071 Castellón
SPAIN

Abstract: - As shown in the bibliography, training an ensemble of networks is an interesting way to improve the performance with respect to a single network. The two key factors to design an ensemble are how to train the individual networks and how to combine the different outputs of the networks to give a single output class. In this paper, we focus in the combination methods. We study the performance of fourteen different combination methods for ensembles of the type "simple ensemble" and "decorrelated". In the case of the "simple ensemble" and low number of networks in the ensemble, the method Zimmermann gets the best performance. When the number of networks is in the range of 9 and 20 the weighted average is the best alternative. Finally, in the case of the ensemble "decorrelated" the best performing method is averaging over a wide spectrum of the number of networks in the ensemble.

Key-Words: - Neural Networks, Ensembles, Multilayer Feedforward, Combination methods.

1 Introduction

Probably, the most important property of a neural network is the generalization capability. The ability to correctly respond to inputs which were not used in the training set.

One technique to increase the generalization capability with respect to a single neural network consist on training an ensemble of neural networks, i.e., to train a set of neural network with different weight initialization or properties in the training process and combine the outputs of the different networks in a suitable manner to give a single output for the whole ensemble.

It is clear from the bibliography that this procedure in general increases the generalization capability [1,2] for the case of Multilayer Feedforward and other classifiers.

The two key factors to design an ensemble are how to train the individual networks and how to combine the different outputs to give a single output.

Among the methods of training the individual networks there is an important number of alternatives. Our research group has performed a comparison detailed in paper [3], which shows that the best performing method among the eleven methods analyzed is called "Decorrelated".

In paper [3], it is also shown that the "simple ensemble" also provide a reasonable performance with a lower computational cost, because in the case of "Decorrelated" we should tune a parameter by trial an error. In this paper, we focus on both methods of building an ensemble.

In the other aspect, (the combination methods for the outputs) there are also several different methods in the bibliography, and we can also find a comparison of twelve different methods in paper [4].

In paper [4], the authors conclude that the combination by the weighted average with data dependent weights is the best method.

However, the comparison of paper [4], under our point of view, lacks of two problems. The method CVC was used to construct the ensemble, and according to our results, this method performs worse than the "simple ensemble" in several situations. Beside that, the experimental results were obtained in only four databases, so the generality of the results might not be clear.

In this paper, we present a comparison of fourteen different method of combining, the outputs for a total of ten databases. So the results are more complete that the ones of paper [4].

Furthermore, we present results for two different methods of building the ensemble: "Simple ensemble" and "Decorrelated". According to our results of paper [3], "Simple Ensemble" is a reasonable alternative with a good performance and low computational cost and "Decorrelated" was the best performing method of the eleven methods analyzed in that paper, at the expense of a higher

computational cost because we should tune a parameter by trial and error.

2 Theory

In this section, first we briefly review the methods "Decorrelated" and "Simple Ensemble" of building the ensemble and then the fourteen methods of combination that are used to obtain experimental results.

2.1 Simple Ensemble and Decorrelated

Simple Ensemble: A simple ensemble can be constructed by training different networks with the same training set, but with different random weight initialization. We hope that the different networks will converge to different local minima and the errors of the networks will be uncorrelated.

Decorrelated (Deco): This ensemble method was proposed in [5]. It consists on introducing a penalty added to the usual error function of Backpropagation. The penalty term for network *j* is:

$$Penalty = \lambda \cdot d(i,j)(y - f_i)\cdot(y - f_j) \quad (1)$$

Where λ determines the strength of the penalty term and should be found by trial and error, *y* is the target of the training pattern and f_i and f_j are the outputs of networks number *i* and *j* in the ensemble. The term *d(i,j)* is in equation 2.

$$d(i,j) = \begin{cases} 1, & if\ \ i = j-1 \\ 0, & otherwise \end{cases} \quad (2)$$

2.2 Combination Methods

Average: This approach simply averages the individual classifier outputs across the different classifiers. The output yielding the maximum of the averaged values is chosen as the correct class.

Majority Vote: Each classifier provides a vote to a class, given by its highest output. The correct class is the one most often voted by the classifiers.

Winner Takes All (WTA): In this method, the class with overall maximum output across all classifier and outputs is selected as the correct class.

Borda Count:. For any class *q*, the Borda count is the sum of the number of classes ranked below *q* by each classifier. If $B_j(q)$ is the number of classes ranked below the class *q* by the *j*th classifier, then the Borda count for class *q* is in the following equation.

$$B(q) = \sum_{j=1}^{L} B_j(q) \quad (3)$$

Bayesian Combination: This combination method was proposed in reference [6]. According to this reference a belief value that the pattern *x* belongs to class *i* can be approximated by equation (4).

$$Bel(i) = \frac{\prod_{k=1}^{L} P(x \in q_i \mid \lambda_k(x) = j_k)}{\sum_{i=1}^{Q}\prod_{k=1}^{L} P(x \in q_i \mid \lambda_k(x) = j_k)}, \quad 1 \le i \le Q \quad (4)$$

Where the conditional probability that sample *x* actually belongs to class *i*, given that classifier *k* assign it to class *j* ($\lambda_k(x)=j_k$) can be estimated from the values of the confusion matrix [4].

Weighted Average: This method introduces weights to the outputs of the different networks prior to averaging. The weights try to minimize the difference between the output of the ensemble and the "desired or true" output. The weights can be estimated from the error correlation matrix. A full description of the method can be found in reference [7].

Choquet Integral: This method is based in the fuzzy integral and the Choquet integral. The method is complex and a full description can be found in reference [4].

Combination by Fuzzy Integral with Data Dependent Densities (Int. DD): It is another method based on the fuzzy integral and the Choquet integral. But in this case, prior to the application of the method it is performed a partition of the input space in regions by k-means clustering or frequency sensitive learning. The full description can be found in reference [4].

Weighted Average with Data Dependent weights (W.Ave DD): This method is the weighted average described above. But in this case, a partition of the space is performed by using k-means clustering and the weights are calculated for each partition. We have a different combination scheme for the different partitions of the space. The number of partitions of the space is determined by cross-validation.

BADD Defuzzification Strategy: It is another combination method based on fuzzy logic concepts. The method is complex and the description can be found in [4].

Zimmermann's Compensatory Operator: This combination method is based in the Zimmermann's compensatory operator described in [8]. The method is complex and can be found in [4].

Dynamically Averaged Networks (DAN), version 1 and 2: It is proposed in reference [9]. In this method instead of choosing static weights derived from the network performance on a sample of the input space, we allow the weights to adjust to be proportional to the certainties of the respective network output. In the reference two different versions of the method are described and we have included both.

Nash Vote: In this method each voter assigns a number between zero and one for each candidate output. The product of the voter's values is compared for all candidates. The higher is the winner. The method is reviewed in reference [10].

3 Experimental Results

We have applied the fourteen combination methods in ten different classification problems. They are from the UCI repository of machine learning databases. Their names are Cardiac Arrhythmia Database (Aritm), Dermatology Database (Derma), Protein Location Sites (Ecoli), Solar Flares Database (Flare), Image Segmentation Database (Image), Johns Hopkins University Ionosphere Database (Ionos), Pima Indians Diabetes (Pima), Haberman's survival data (Survi), Vowel Recognition (Vowel) and Wisconsin Breast Cancer Database (Wdbc).

We have constructed ensembles of a wide number of networks, in particular 3, 9, 20 and 40 networks in the ensemble. In this case we can test the results in a wide set of situations.

First, we trained the ensembles of 3, 9, 20 and 40 networks. We repeated this process of training an ensemble ten times for different partitions of data in training, cross-validation and test sets. With this procedure we can obtain a mean performance of the ensemble for each database (the mean of the ten ensembles) and an error in the performance calculated by standard error theory.

The results of the performance for an ensemble of the type "Simple Ensemble" are in table 1 for the case of ensembles of three networks and in table 2 for the case of nine. For the ensemble of type "Decorrelated" the results are in table 3 for the case of three networks. The rest of results are omitted by the lack of space, but a resume of the performance can be found below.

Table 1. Results for the ensemble of three networks, different combination methods with an ensemble of the type "Simple Ensemble".

	ARITM	DERMA	ECOLI	FLARE
SingleNetwork	75.6 ± 0.7	96.7 ± 0.4	84.4 ± 0.7	82.1 ± 0.3
Average	73.5 ± 1.1	97.2 ± 0.7	86.6 ± 0.8	81.8 ± 0.5
Majority V.	73.1 ± 1.0	96.9 ± 0.8	86.0 ± 0.9	81.5 ± 0.5
WTA	73.6 ± 1.0	97.2 ± 0.7	86.3 ± 0.9	81.7 ± 0.5
Borda	73.1 ± 1.0	97.0 ± 0.7	86.5 ± 0.8	81.5 ± 0.5
Bayesian	73.6 ± 0.9	96.9 ± 0.8	86.3 ± 0.9	81.5 ± 0.5
W. Average	73.0 ± 0.9	96.3 ± 0.7	85.9 ± 0.9	81.3 ± 0.6
Choquet	74.1 ± 1.1	97.2 ± 0.7	86.3 ± 0.9	81.7 ± 0.5
Int. DD	74.1 ± 1.1	97.2 ± 0.7	85.9 ± 0.7	81.8 ± 0.5
W. Ave DD	73.5 ± 1.1	97.2 ± 0.7	86.6 ± 0.8	81.8 ± 0.5
BADD	73.5 ± 1.1	97.2 ± 0.7	86.6 ± 0.8	81.8 ± 0.5
Zimmermann	74.7 ± 1.4	97.3 ± 0.7	86.0 ± 1.2	81.5 ± 0.6
DAN	73.2 ± 1.1	96.9 ± 0.6	85.7 ± 1.0	81.4 ± 0.6
DANversion 2	73.2 ± 1.1	96.9 ± 0.6	85.4 ± 0.9	81.4 ± 0.6
Nash Vote	73.5 ± 1.1	97.3 ± 0.7	86.6 ± 0.8	81.8 ± 0.5

Table 1. (continuation 1) Results for the ensemble of three networks, different combination methods with an ensemble of the type "Simple Ensemble".

	IMAGEN	IONOS	PIMA
SingleNetwork	96.3 ± 0.2	87.9 ± 0.7	76.7 ± 0.6
Average	96.5 ± 0.2	91.1 ± 1.1	75.9 ± 1.2
Majority V.	96.2 ± 0.3	91.3± 1.0	75.9 ± 1.3
WTA	96.4 ± 0.2	91.1 ± 1.1	75.9 ± 1.2
Borda	95.7 ± 0.2	91.3± 1.0	75.9 ± 1.3
Bayesian	96.3 ± 0.3	91.4 ± 1.1	75.8 ± 1.3
W. Average	96.7 ± 0.3	91.3± 0.9	75.3± 1.3
Choquet	96.3 ± 0.2	91.3± 1.1	76.1 ± 1.2
Int. DD	96.3 ± 0.2	91.1 ± 1.1	75.6 ± 1.3
W. Ave DD	96.5 ± 0.2	91.1 ± 1.1	75.9 ± 1.2
BADD	96.5 ± 0.2	91.1 ± 1.1	75.9 ± 1.2
Zimmermann	96.6 ± 0.3	91.9± 1.1	76.0 ± 1.0
DAN	95.7 ± 0.2	90.0 ± 1.2	75.9± 1.2
DANversion 2	95.7 ± 0.2	90.0 ± 1.2	75.9± 1.2
Nash Vote	95.8 ± 0.2	91.3± 1.2	75.9 ± 1.2

Table 1. (continuation 2) Results for the ensemble of three networks, different combination methods with an ensemble of the type "Simple Ensemble".

	SURVI	VOWEL	WDBC
SingleNetwork	74.2 ± 0.8	83.4 ± 0.6	97.4 ± 0.3
Average	74.3 ± 1.3	88.0 ± 1.0	96.9 ± 0.5
Majority V.	74.4 ± 1.4	86.9 ± 0.9	96.9 ± 0.5
WTA	73.9 ± 1.4	86.7 ± 0.8	96.9 ± 0.5
Borda	74.4 ± 1.4	85.9 ± 1.0	96.9 ± 0.5
Bayesian	74.3 ± 1.4	86.4 ± 1.0	96.9 ± 0.5
W. Average	74.1 ± 1.3	87.7 ± 1.0	96.9 ± 0.5
Choquet	74.1 ± 1.4	86.4 ± 0.7	96.9 ± 0.5
Int. DD	74.1 ± 1.3	86.3 ± 0.7	96.9 ± 0.5
W. Ave DD	74.3 ± 1.3	88.0 ± 0.9	96.9 ± 0.5
BADD	74.3 ± 1.3	88.0 ± 0.9	96.9 ± 0.5
Zimmermann	74.3 ± 1.3	87.8 ± 1.0	96.9 ± 0.5
DAN	74.4 ± 1.4	84.6 ± 1.2	96.9 ± 0.5
DANversion 2	74.4 ± 1.4	84.5 ± 1.2	96.9 ± 0.5
Nash Vote	74.3± 1.3	86.2 ± 1.0	96.9 ± 0.5

Table 2. Results for the ensemble of nine networks, different combination methods with an ensemble of the type "Simple Ensemble".

	ARITM	DERMA	ECOLI	FLARE
SingleNetwork	75.6 ± 0.7	96.7 ± 0.4	84.4 ± 0.7	82.1 ± 0.3
Average	73.8± 1.1	97.5± 0.7	86.9 ± 0.8	81.6 ± 0.4
Majority V.	73.3 ± 1.0	97.5± 0.7	86.0 ± 0.9	81.5 ± 0.5
WTA	73.3 ± 1.1	96.9 ± 0.8	87.1± 0.9	81.5 ± 0.4
Borda	73.3 ± 0.9	97.5± 0.7	86.6 ± 0.9	81.5 ± 0.5
Bayesian	73.6± 0.9	96.3 ± 0.8	85.2± 0.9	81.6 ± 0.5
W. Average	74.8 ± 1.0	97.0 ± 0.7	86.2± 1.1	81.6 ± 0.4
Choquet	73.3 ± 1.1	97.2± 0.8	86.9 ± 1.0	81.5 ± 0.5
Int. DD	73.3 ± 1.1	97.2± 0.8	86.2± 1.0	81.5 ± 0.5
W. Ave DD	73.8± 1.1	97.5± 0.7	87.1± 0.9	81.6 ± 0.5
BADD	73.8± 1.1	97.5± 0.7	86.9 ± 0.8	81.6 ± 0.4
Zimmermann	74.3± 0.8	97.5± 0.6	85.4 ± 0.9	81.4 ± 0.6
DAN	73.6 ± 1.0	97.2± 0.6	85.2± 0.8	81.5 ± 0.5
DANversion 2	73.6 ± 1.0	97.2± 0.6	84.6± 0.8	81.5 ± 0.5
Nash Vote	73.7± 1.0	97.5± 0.7	87.2 ± 0.8	81.5 ± 0.4

Table 2. (continuation 1) Results for the ensemble of nine networks, different combination methods with an ensemble of the type "Simple Ensemble".

	IMAGEN	IONOS	PIMA
SingleNetwork	96.3 ± 0.2	87.9 ± 0.7	76.7 ± 0.6
Average	96.7 ± 0.3	90.3 ± 1.1	75.9 ± 1.2
Majority V.	96.6 ± 0.3	90.6 ± 1.2	76.1 ± 1.2
WTA	96.5 ± 0.2	90.9 ± 1.3	75.8 ± 1.2
Borda	96.17 ± 0.18	90.6 ± 1.2	76.1 ± 1.2
Bayesian	96.1 ± 0.3	93.1 ± 1.4	75.9 ± 1.2
W. Average	96.8 ± 0.3	91.9 ± 1.0	76.1 ± 1.2
Choquet	96.3 ± 0.2	90.7 ± 1.2	75.7 ± 1.2
Int. DD	96.3 ± 0.2	90.4 ± 1.2	75.7 ± 1.3
W. Ave DD	96.7 ± 0.3	90.4 ± 1.1	75.9 ± 1.2
BADD	96.7 ± 0.3	90.3 ± 1.1	75.9 ± 1.2
Zimmermann	96.7 ± 0.3	92.0 ± 1.0	75.9 ± 1.0
DAN	96.2 ± 0.2	90.0 ± 1.1	75.7 ± 1.1
DANversion 2	96.1 ± 0.2	90.0 ± 1.1	75.7 ± 1.1
Nash Vote	96.24 ± 0.18	90.4 ± 1.2	75.9 ± 1.2

Table 2. (continuation 2) Results for the ensemble of nine networks, different combination methods with an ensemble of the type "Simple Ensemble".

	SURVI	VOWEL	WDBC
SingleNetwork	74.2 ± 0.8	83.4 ± 0.6	97.4 ± 0.3
Average	74.3 ± 1.3	91.0 ± 0.5	96.9 ± 0.5
Majority V.	74.1 ± 1.3	89.6 ± 0.6	96.9 ± 0.5
WTA	74.6 ± 1.3	90.3 ± 0.6	96.9 ± 0.5
Borda	74.1 ± 1.3	87.5 ± 0.9	96.9 ± 0.5
Bayesian	74.1 ± 1.3	82.5 ± 0.9	97.0 ± 0.4
W. Average	73.3 ± 1.2	92.0 ± 0.6	97.2 ± 0.5
Choquet	74.4 ± 1.4	88.2 ± 0.7	96.9 ± 0.5
Int. DD	74.1 ± 1.5	87.9 ± 0.8	96.9 ± 0.5
W. Ave DD	74.3 ± 1.3	91.0 ± 0.5	96.9 ± 0.5
BADD	74.3 ± 1.3	91.0 ± 0.5	96.9 ± 0.5
Zimmermann	74.4 ± 1.4	91.3 ± 0.4	96.9 ± 0.5
DAN	74.3 ± 1.3	85.8 ± 1.0	97.0 ± 0.4
DANversion 2	74.3 ± 1.3	85.9 ± 1.0	97.0 ± 0.4
Nash Vote	74.1 ± 1.4	87.5 ± 0.8	96.9 ± 0.5

Table 3. Results for the ensemble of three networks, different combination methods with an ensemble of the type "Decorrelated".

	ARITM	DERMA	ECOLI	FLARE
SingleNetwork	75.6 ± 0.7	96.7 ± 0.4	84.4 ± 0.7	82.1 ± 0.3
Average	74.9 ± 1.3	97.2 ± 0.7	86.6 ± 0.6	81.7 ± 0.4
Majority V.	74.9 ± 1.1	97.5 ± 0.6	86.3 ± 0.7	81.4 ± 0.5
WTA	74.9 ± 1.1	97.0 ± 0.7	86.0 ± 0.6	81.7 ± 0.5
Borda	74.9 ± 1.1	97.0 ± 0.9	86.6 ± 0.8	81.4 ± 0.5
Bayesian	74.8 ± 1.2	96.9 ± 0.8	86.8 ± 0.5	81.3 ± 0.5
W. Average	74.4 ± 1.2	97.3 ± 0.5	86.6 ± 0.7	81.6 ± 0.5
Choquet	74.9 ± 1.2	97.0 ± 0.7	86.0 ± 0.6	81.7 ± 0.5
Int. DD	74.9 ± 1.2	97.0 ± 0.7	86.2 ± 0.8	81.5 ± 0.4
W. Ave DD	75.1 ± 1.3	97.2 ± 0.7	86.6 ± 0.7	81.5 ± 0.4
BADD	74.9 ± 1.3	97.2 ± 0.7	86.6 ± 0.6	81.7 ± 0.4
Zimmermann	74.6 ± 1.2	97.3 ± 0.5	86.0 ± 0.9	81.4 ± 0.6
DAN	72.9 ± 1.1	96.8 ± 1.1	85.2 ± 0.7	81.3 ± 0.5
DANversion 2	72.9 ± 1.1	96.6 ± 1.2	85.0 ± 0.8	81.3 ± 0.5
Nash Vote	75.3 ± 1.3	97.2 ± 0.7	86.3 ± 0.5	81.6 ± 0.5

Table 3. (continuation 1) Results for the ensemble of three networks, different combination methods with an ensemble of the type "Decorrelated".

	IMAGEN	IONOS	PIMA
SingleNetwork	96.3 ± 0.2	87.9 ± 0.7	76.7 ± 0.6
Average	96.7 ± 0.3	90.9 ± 0.9	76.4 ± 1.2
Majority V.	96.5 ± 0.3	90.7 ± 1.2	75.8 ± 1.1
WTA	96.7 ± 0.2	91.4 ± 0.9	76.1 ± 1.1
Borda	96.0 ± 0.4	90.7 ± 1.2	75.8 ± 1.1
Bayesian	96.7 ± 0.3	92.3 ± 1.0	75.7 ± 1.1
W. Average	96.7 ± 0.3	91.6 ± 0.8	76.1 ± 0.9
Choquet	96.58 ± 0.19	91.1 ± 1.0	75.9 ± 1.1
Int. DD	96.56 ± 0.19	91.0 ± 0.9	75.5 ± 1.1
W. Ave DD	96.7 ± 0.3	91.1 ± 1.0	76.4 ± 1.1
BADD	96.7 ± 0.3	90.9 ± 0.9	76.4 ± 1.2
Zimmermann	96.5 ± 0.3	91.4 ± 1.0	76.6 ± 1.0
DAN	95.9 ± 0.3	89.9 ± 1.2	75.2 ± 1.0
DANversion 2	96.0 ± 0.3	89.9 ± 1.2	75.2 ± 1.0
Nash Vote	96.2 ± 0.3	91.0 ± 0.9	76.4 ± 1.2

Table 3. (continuation 2) Results for the ensemble of three networks, different combination methods with an ensemble of the type "Decorrelated".

	SURVI	VOWEL	WDBC
SingleNetwork	74.2 ± 0.8	83.4 ± 0.6	97.4 ± 0.3
Average	74.6 ± 1.5	91.5 ± 0.6	97.0 ± 0.5
Majority V.	74.1 ± 1.5	89.4 ± 0.5	97.0 ± 0.5
WTA	74.6 ± 1.5	91.2 ± 0.7	96.8 ± 0.4
Borda	74.1 ± 1.5	87.8 ± 0.8	97.0 ± 0.5
Bayesian	73.8 ± 1.3	88.0 ± 0.4	97.0 ± 0.5
W. Average	73.4 ± 1.2	91.1 ± 0.4	96.9 ± 0.4
Choquet	74.8 ± 1.3	90.1 ± 0.5	96.7 ± 0.4
Int. DD	74.6 ± 1.3	90.0 ± 0.5	96.7 ± 0.4
W. Ave DD	74.6 ± 1.5	91.7 ± 0.6	97.0 ± 0.5
BADD	74.6 ± 1.5	91.5 ± 0.6	97.0 ± 0.5
Zimmermann	74.1 ± 1.3	90.6 ± 0.6	96.7 ± 0.4
DAN	74.4 ± 1.4	85.8 ± 0.7	97.1 ± 0.4
DANversion 2	74.4 ± 1.4	86.3 ± 0.8	97.1 ± 0.4
Nash Vote	74.6 ± 1.5	88.1 ± 0.8	96.9 ± 0.4

In tables 1, 2 and 3, the mean performance of a single network is included for comparison with the performance of an ensemble.

As the results show the performance difference among the different combination methods is low. For example, for the case of table 1 ("simple ensemble", three networks in the ensemble) the largest difference between two combination methods is only 1.9% in the case of database Ionos.

However, the computational cost of the different combination methods is very low in comparison to the computational cost of training the ensemble. In this sense, it can be worthy to select an appropriate combination method because it provides an improvement in the performance at no extra computational cost.

We can obtain further conclusions and insights in the performance of the different methods by calculating the percentage of error reduction of the ensemble with respect to a single network. We have used equation 5 for this calculation.

$$PorError_{reduction} = 100 \cdot \frac{PorError_{\text{single network}} - PorError_{ensemble}}{PorError_{\text{single network}}} \quad (5)$$

The value of the percentage of error reduction ranges from 0%, where there is no improvement by the use of a particular ensemble method with respect to a single network, to 100%. There can also be negative values, which means that the performance of the ensemble is worse than the performance of the single network.

Table 4. Mean percentage of error reduction for the different ensembles, "Simple Ensemble".

	Ensambles of			
	3 Nets	**9 Nets**	**20 Nets**	**40 Nets**
Average	5,49	8,25	8,13	9,73
Majority V.	2,73	6,61	7,52	8,11
WTA	4,16	6,01	6,65	6,14
Borda	1,55	4,63	5,35	6,39
Bayesian	3,20	-0,52	-9,05	-16,58
W. Average	2,22	9,77	10,82	6,38
Choquet	4,35	4,87	--	--
Int. DD	3,74	3,75	--	--
W. Ave DD	5,54	8,52	--	--
BADD	5,49	8,25	8,13	9,73
Zimmermann	6,80	9,18	4,98	-16,36
DAN	-1,27	1,72	-1,38	-0,71
DANversion 2	-1,50	1,15	-1,46	-0,83
Nash Vote	3,30	5,13	6,34	7,08

This new measurement is relative and can be used to compare more clearly the different methods. Furthermore we can calculate the mean performance of error reduction across all databases; this value is in table 4 for ensembles of 3, 9, 20 and 40 networks, in the case of ensembles of the type "simple ensemble". For the case of "Decorrelated" and 3, 9, 20 and 40 networks, they are in table 5.

As the results of table 4 show, the average is quite appropriate for the simple ensemble. In fact it provides the best performance for the case of 40 networks.

However, the method Zimmermann get the best performance and should be used for ensembles of low number of networks (3 networks). As the number of networks increases (20 and 40) the method does not work.

There is another method that should be taken into account, weighted average gets the best performance when an intermediate number of networks in the ensemble is used (9 and 20).

In the case of "decorrelated", as the results of table 5 show, the best performing methods are clearly the simple average and BADD over a wide spectrum in the number of networks in the ensemble. In fact they are only outperformed in the case of three networks by the method W. Average DD. In this case, considering the simplicity and the performance the best alternative is average of outputs.

Table 5. Mean percentage of error reduction for the different ensembles, "Decorrelated".

	Ensambles of			
	3 Nets	**9 Nets**	**20 Nets**	**40 Nets**
Average	9,29	12,15	12,33	12,47
Majority V.	7,46	10,15	11,15	11,63
WTA	7,94	5,18	8,92	7,67
Borda	3,90	6,68	8,26	8,99
Bayesian	6,60	-1,59	-9,87	-16,52
W. Average	8,68	7,44	9,49	3,21
Choquet	6,28	3,56	--	--
Int. DD	5,85	3,36	--	--
W. Ave DD	9,64	11,88	--	--
BADD	9,29	12,15	12,33	12,47
Zimmermann	7,22	9,50	3,39	-15,90
DAN	-0,62	0,26	1,88	0,60
DANversion 2	-0,60	0,10	1,73	-0,10
Nash Vote	5,46	6,84	10,04	9,69

4 Conclusion

In this paper, we have focused on the different alternatives of combination methods to obtain a unique output in an ensemble of neural networks. We have performed experiment with a total of fourteen different combination methods for ensembles of the type "simple ensemble" and "decorrelated". The experiments are performed with ten different databases from the UCI repository. Furthermore, the ensembles are trained with 3, 9, 20 and 40 networks; giving a wide spectrum in the number of networks in

the ensemble. The results show that in the case of the "simple ensemble" and low number of networks in the ensemble, the method Zimmermann gets the best performance. When the number of networks is in the range of 9 and 20 the weighted average is the best alternative. Finally, in the case of the ensemble "decorrelated" the best performing method is averaging over a wide spectrum of the number of networks in the ensemble.

Acknowledgments

This research was supported by the project MAPACI TIC2002-02273 of CICYT in Spain.

References:

[1] Tumer, K., Ghosh, J., "Error correlation and error reduction in ensemble classifiers", *Connection Science*, Vol. 8, Nos. 3 & 4, 1996, pp. 385-404.

[2] Raviv, Y., Intrator, N., "Bootstrapping with Noise: An Effective Regularization Technique", *Connection Science*, Vol. 8, No. 3 & 4, 1996, pp. 355-372.

[3] Fernandez-Redondo, Mercedes, Hernández-Espinosa, Carlos, Torres-Sospedra, Joaquín, "Classification by Multilayer Feedforward ensambles", *Internacional Symposium on Neural Networks, Lecture Notes in Computer Science*, Vol. 3173, 2004, pp. 852-857.

[4] Verikas, A., Lipnickas, A., Malmqvist, K., Bacauskiene, M., Gelzinis, A., "Soft Combination of neural classifiers: A comparative study", *Pattern Recognition Letters*, Vol. 20, 1999, pp 429-444.

[5] Rosen, B., "Ensemble Learning Using Decorrelated *Neural Networks", Connection Science*, Vol. 8, No. 3 & 4, 1996, pp. 373-383.

[6] Xu, L., Krzyzak, A., SUen, C.Y., "Methods of combining multiple classifiers and their applications to handwriting recognition", *IEEE Trans. On Systems, Man and Cybernetics*, Vol. 22, No. 3, 1992, pp. 418-435.

[7] Krogh, A., Vedelsby, J., "Neural network ensembles, cross validation, and active learning", *Advances in Neural Information Processing Systems 7*, 1995, pp. 231-238.

[8] Zymmermann, H.J., Zysno, P., "Decision and evaluations by hierarchical aggregation of information", *Fuzzy Sets and Systems*, Vol. 10, No. 3, 1984, pp. 243-260.

[9] Jimenez, D., "Dynamically Weighted Ensemble Neural Network for Classification", *IEEE World Congress on Computational Intelligence*, Vol. 1, 1998, pp. 753-756.

[10] Wanas, N.M., Kamel, M.S., "Decision Fusion in Neural Network Ensembles", *International Joint Conference on Neural Networks*, Vol. 4, 2001, pp. 2952-2957.