Linear Combiners for Classifier Fusion: Some Theoretical and Experimental Results

Giorgio Fumera and Fabio Roli

Department of Electrical and Electronic Engineering, University of Cagliari,
Piazza d'Armi, 09123, Cagliari, Italy
{fumera, roli}@diee.unica.it

Abstract. In this paper, we continue the theoretical and experimental analysis of two widely used combining rules, namely, the simple and weighted average of classifier outputs, that we started in previous works. We analyse and compare the conditions which affect the performance improvement achievable by weighted average over simple average, and over individual classifiers, under the assumption of unbiased and uncorrelated estimation errors. Although our theoretical results have been obtained under strict assumptions, the reported experiments show that they can be useful in real applications, for designing multiple classifier systems based on linear combiners.

1 Introduction

Recently, some works started to analyze the theoretical foundations of techniques for combining multiple classifiers [7,10,6,11]. Some works also provided analytical comparisons between the performance of different techniques [4,5,3,1,9]. Because of the complexity of developing analytical models of combining rules, the focus was limited to the simplest and most used rules, like majority vote, linear combination of classifier outputs, and order statistics combiners.

In this paper, we focus on linear combiners. A theoretical framework for evaluating the performance improvement achievable by simple averaging the outputs of an ensemble of classifiers, was developed by Tumer and Ghosh [10,11]. In previous works, we extended this framework to weighted average [9,1], and to classification with reject option [2]. In these works, we focused on the comparison between the performance of simple and weighted average. In particular, we provided analytical results which showed how the difference between the accuracies of individual classifiers (performance "imbalance") affect the improvement achievable by weighted average over simple average, for classifier ensembles made up of three classifiers, and under the hypothesis of unbiased and uncorrelated estimation errors [9,1]. We also provided a preliminary analysis of the effects of correlated estimation errors.

In this work, we extend the above analysis of linear combiners. In particular, we determine the conditions on performance imbalance that affect the improvement of weighted average over simple average, and also over individual classifiers, for classifier ensembles of any size. We then analytically evaluate and compare such improvement. We also extend the experimental investigation with respect to our previous works. The behavior of simple and weighted average is evaluated on two real data sets, with the aim of assessing the usefulness of our theoretical results for the design of multiple classifier systems based on linear combiners.

Our previous works are summarised in section 2. The new theoretical results are presented in section 3, while section 4 reports the experimental results.

2 Summary of previous results

Tumer and Ghosh developed a theoretical framework which allows to evaluate the added error probability, over Bayes error, of individual and multiple classifiers combined by simple average (SA), as a function of bias, variance and pair-wise correlation of estimation errors that affect the a posteriori probability estimates [10,11]. It turns out that the added error, for a given class boundary, is proportional to the sum of the variance and the squared bias of the estimation errors. For classifiers combined by SA, the variance component is also affected by pair-wise correlation between errors of different classifiers. It was shown that simple averaging the outputs of N classifiers reduces the variance component of the added error, while the bias component is not necessarily reduced. In particular, the reduction factor of the variance component is equal to N, if the errors are unbiased and uncorrelated, while it is lower (higher) than N if the errors are positively (negatively) correlated [10,11].

In [9,1], we extended the above framework to analyze and compare the performance of simple and weighted average (WA). We considered the case in which the estimate $\hat{P}(\omega_i \mid \mathbf{x})$ of the a posteriori probability of the *i*-th class, $P(\omega_i \mid \mathbf{x})$, is computed as a linear combination of the estimates provided by N individual classifiers, $\hat{P}(\omega_i \mid \mathbf{x}) = \sum_{m=1}^{N} w_m \hat{P}_m(\omega_i \mid \mathbf{x})$, where the weights are positive and sum up to 1: $\sum_{m=1}^{N} w_m = 1, w_m \geq 0 \; m = 1,\ldots,N$. If the errors $\varepsilon_i^m(\mathbf{x}) = \hat{P}_m(\omega_i \mid \mathbf{x}) - P(\omega_i \mid \mathbf{x})$ are unbiased and uncorrelated[1], we showed that the added error of a linear combination of classifiers is $E_{add}^{ave} = \sum_{m=1}^{N} E_{add}^{m} w_m^2$. It turns out that the optimal weights w_m (i.e., the ones which minimise E_{add}^{ave}) are inversely proportional to the added error E_{add}^{m} of the corresponding classifiers. Therefore, SA ($w_m = 1/N$) provides the minimum of E_{add}^{ave} only if individual classifiers exhibit the same added error, or equivalently, the same overall error probability. Otherwise, WA provides a lower E_{add}^{ave}. Accordingly, it turned out that the minimum of E_{add}^{ave} is:

$$E_{add}^{WA} = \left(\sum_{m=1}^{N} 1 / E_{add}^{m}\right)^{-1} . \quad \textbf{(1)}$$

Simple average provides instead the following value of E_{add}^{ave}:

[1] As in [T&G96, T&G99], we assume that only the errors of different classifiers on the same class, $\varepsilon_i^m(\mathbf{x})$ and $\varepsilon_i^n(\mathbf{x})$, can be correlated.

$$E_{add}^{SA} = \left(1/N^2\right)\sum_{m=1}^{N} E_{add}^{m} \quad . \tag{2}$$

If the estimation errors of individual classifiers are correlated, by assuming that the variance and pair-wise correlations do not depend on the classes, E_{add}^{ave} is:

$$E_{add}^{ave} = \sum_{m=1}^{N} E_{add}^{m} w_m^2 + \sum_{m=1}^{N} \sum_{m \neq n} \rho^{mn} \sqrt{E_{add}^{m} E_{add}^{n}} w_m w_n \quad , \tag{3}$$

where ρ^{mn} is the correlation between the errors of classifiers m and n, on the considered class boundary. Even if in this case the optimal weights can not be computed analytically, it turns out that SA minimises E_{add}^{ave} only if the individual classifiers exhibit the same performance (i.e., equal values of E_{add}^{m}), and equal values of $\rho^{mn}, \forall m,n$. Otherwise, WA is needed.

In [9,1], we exploited the above results to analyse the performance improvement achievable by WA over SA, i.e., $E_{add}^{SA} - E_{add}^{WA}$, for ensembles of three classifiers ($N = 3$). Note that the difference between the added errors coincides with the difference between the overall error probabilities. In particular, we analysed how $E_{add}^{SA} - E_{add}^{WA}$ is affected by the difference between the performance and the pair-wise correlations of individual classifiers. Without loss of generality, consider the individual classifiers ordered for decreasing values of E_{add}^{m}: $E_{add}^{1} \geq E_{add}^{2} \geq E_{add}^{3}$. In the case of unbiased and uncorrelated errors (Eqs. (1),(2)), we proved that $E_{add}^{SA} - E_{add}^{WA}$ is maximum, for a given value of the error range $E_{add}^{1} - E_{add}^{3}$, if $E_{add}^{2} = E_{add}^{1}$. In other words, given the performance of the best and worst individual classifiers, WA provides the maximum improvement, with respect to SA, if the second best classifier exhibit the same performance of the worst classifier. We qualitatively denoted this condition as the maximum performance "imbalance" between individual classifiers, with respect to $E_{add}^{SA} - E_{add}^{WA}$. For correlated estimation errors (Eq. (3)), only a numerical analysis was possible (as pointed out above, the weights which minimise Eq. (3) can not be computed analytically). Consider any given value of the error range $E_{add}^{1} - E_{add}^{3}$, and of the correlation range $\rho_{\max} - \rho_{\min}$ (note that $\rho_{\max}$ and $\rho_{\min}$ depend on the number *N* of classifiers). $E_{add}^{SA} - E_{add}^{WA}$ turns out to be maximum, if $E_{add}^{2} = E_{add}^{1}$, and if $\rho^{23} = \rho_{\min}, \rho^{12} = \rho^{13} = \rho_{\max}$. In other words, we found the same conditions of maximum performance imbalance as for uncorrelated errors, and analogous conditions of maximum correlation imbalance.

Finally, using Eqs. (1)-(3), we evaluated $E_{add}^{SA} - E_{add}^{WA}$ under the above conditions of maximum performance and correlation imbalance, as a function of the error range $E_{add}^{1} - E_{add}^{3}$. To sum up, our model predicts that WA can achieve a significant improvement over SA only for ensembles of classifiers exhibiting highly imbalanced performance (that is, large error range and pair-wise correlations). However, in this case, WA tends to perform very similarly to the best individual classifier, and, therefore, combining provides little advantage.

The above results are limited to ensembles made up of three classifiers. In the following section, we extend our analysis to classifier ensembles of any size *N*. In particular, we provide the conditions under which $E_{add}^{SA} - E_{add}^{WA}$ is maximum and minimum, for any N. Moreover, we determine the conditions which affect the improvement achievable by simple and WA over individual classifiers. We then compare these conditions, and provide a quantitative evaluation for the case of unbiased and uncorrelated errors, based on Eqs. (1), (2).

3 Some theoretical results for linear combiners

Without loss of generality, consider N individual classifiers ordered for decreasing values of their added errors, $E_{add}^1 \geq E_{add}^2 \geq ... \geq E_{add}^N$. For unbiased and uncorrelated estimation errors (Eqs. (1) and (2)), it can be shown that $E_{add}^{SA} - E_{add}^{WA}$ is maximum, for any given error range $E_{add}^1 - E_{add}^N$, if k classifiers ($k < N$) have an added error equal to that of the best one (E_{add}^N), and the other $N-k$ have an added error equal to that of the worst one (E_{add}^1). The value of k can be either $\lfloor k^* \rfloor$ or $\lceil k^* \rceil$, where $k^* = N\left(E_{add}^N - \sqrt{E_{add}^1 E_{add}^N}\right) / \left(E_{add}^N - E_{add}^1\right)$ (which one depends on the particular values of the added errors). Note that, for $N = 3$, one always obtains $k = 1$, which is the condition found in our previous works (section 2). Moreover, $E_{add}^{SA} - E_{add}^{WA}$ is minimum, for any given $E_{add}^1 - E_{add}^N$, if classifiers $2, 3, \ldots, N-1$ exhibit the same added error, equal to $2E_{add}^1 E_{add}^N / \left(E_{add}^1 + E_{add}^N\right)$. These two conditions, depicted in Fig. 1(a),(b), can be denoted respectively as maximum and minimum performance imbalance, with respect to $E_{add}^{SA} - E_{add}^{WA}$.

The proof of the above results can be summarised as follows. The partial derivative of $E_{add}^{SA} - E_{add}^{WA}$ with respect to any $E_{add}^m, 1 < m < N$, is $N^{-2} - \left(E_{add}^m\right)^{-2}\left(\sum_{n=1}^N \left(E_{add}^n\right)^{-1}\right)^{-2}$. By analyzing its sign, and noting that $E_{add}^1 \geq E_{add}^m \geq E_{add}^N$, it is easy to see that $E_{add}^{SA} - E_{add}^{WA}$ is convex with respect to E_{add}^m. In particular, the maximum of $E_{add}^{SA} - E_{add}^{WA}$ is achieved either when $E_{add}^m = E_{add}^1$, or when $E_{add}^m = E_{add}^N$. It follows that, for given E_{add}^1 and E_{add}^N, the maximum of $E_{add}^{SA} - E_{add}^{WA}$ with respect to $E_{add}^2, \ldots, E_{add}^{N-1}$ is achieved when some of the added errors (say, the first $N-k-1$) are equal to E_{add}^1, and the other $k-1$ are equal to E_{add}^N. The corresponding $E_{add}^{SA} - E_{add}^{WA}$ is:

$$E_{add}^{SA} - E_{add}^{WA} = \left((N-k)E_{add}^1 + kE_{add}^N\right)N^{-2} - \left((N-k) / E_{add}^1 + k / E_{add}^N\right)^{-1} . \quad \textbf{(4)}$$

By minimising this expression with respect to k, one obtains the value of k given above. Let us consider the minimum of $E_{add}^{SA} - E_{add}^{WA}$, with respect to $E_{add}^2, \ldots, E_{add}^{N-1}$, for given E_{add}^1 and E_{add}^N. First, we recall that $E_{add}^{SA} - E_{add}^{WA}$ is convex with respect to any $E_{add}^m, 1 < m < N$: it has therefore only one global minimum. By setting to zero the partial derivatives (given above) with respect to each $E_{add}^m, 1 < m < N$, we obtain $N\left(E_{add}^m\right)^{-1} = \sum_{n=1}^N \left(E_{add}^n\right)^{-1}$. It easily follows that $E_{add}^{SA} - E_{add}^{WA}$ is minimum when $E_{add}^2, \ldots, E_{add}^{N-1}$ are all equal to $2E_{add}^1 E_{add}^N / \left(E_{add}^1 + E_{add}^N\right)$.

Consider now the improvement achievable by simple and WA over the best individual classifier. From Eqs. (1) and (2), it is easy to see that the maximum of $E_{add}^{N} - E_{add}^{SA}$ and $E_{add}^{N} - E_{add}^{WA}$, for any given $E_{add}^{1} - E_{add}^{N}$, is achieved when classifiers $2,3,\ldots,N-1$ exhibit an added error equal to E_{add}^{N}, while the minimum is achieved when the added error of classifiers $2,3,\ldots,N-1$ is equal to E_{add}^{1}. Let us denote these conditions respectively as maximum and minimum performance imbalance, with respect to both $E_{add}^{N} - E_{add}^{SA}$ and $E_{add}^{N} - E_{add}^{WA}$. These conditions are depicted in Fig. 1(c),(d). By analysing Eq. (3), it is easy to see that these conditions also hold for correlated errors, when all correlation coefficients are non-negative.

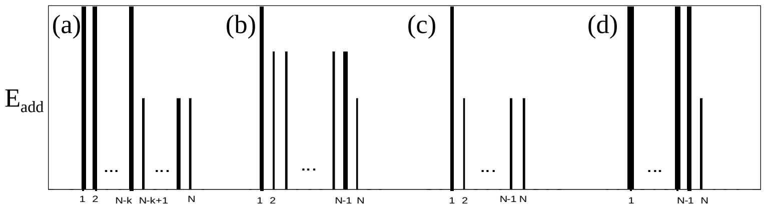
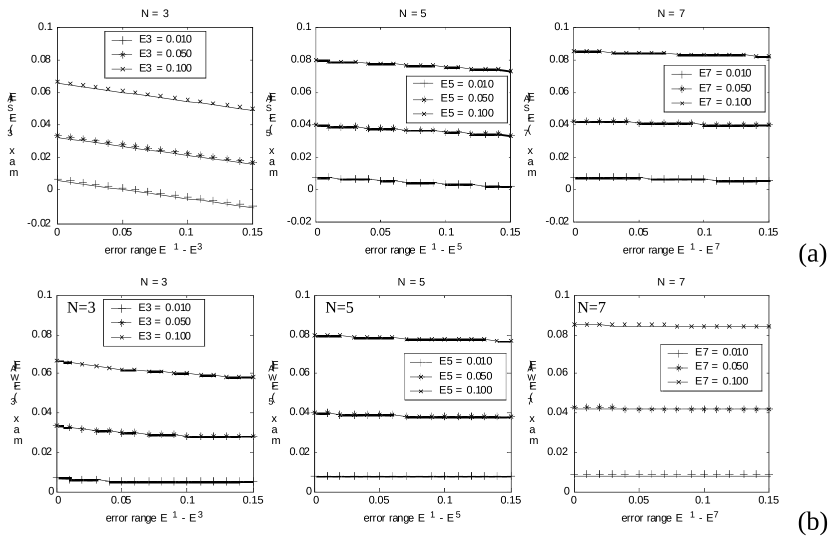


Fig. 1. Conditions of maximum and minimum performance imbalance, with respect to the added errors of individual classifiers, for $E_{add}^{SA} - E_{add}^{WA}$ (a), (b), and for $E_{add}^{N} - E_{add}^{SA}$ and $E_{add}^{N} - E_{add}^{WA}$ (c), (d).

We point out that the conditions of maximum performance imbalance for $E_{add}^{SA} - E_{add}^{WA}$, are different from the ones for $E_{add}^{N} - E_{add}^{SA}$ and $E_{add}^{N} - E_{add}^{WA}$ (see Figs. 1(a),(c)). This means that when simple and WA provide the maximum improvement over individual classifiers, the improvement of WA over SA is not maximum, and vice-versa. Let us now quantitatively evaluate and compare $E_{add}^{SA} - E_{add}^{WA}$, $E_{add}^{N} - E_{add}^{SA}$, and $E_{add}^{N} - E_{add}^{WA}$, for the case of unbiased and uncorrelated errors.



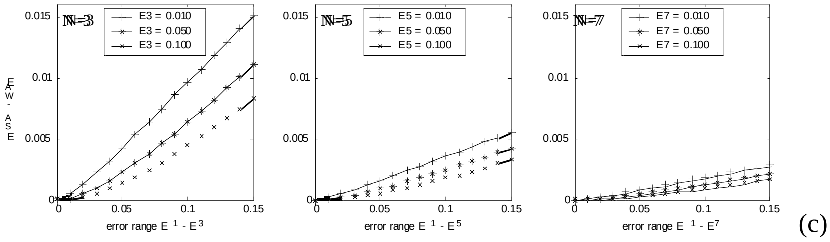
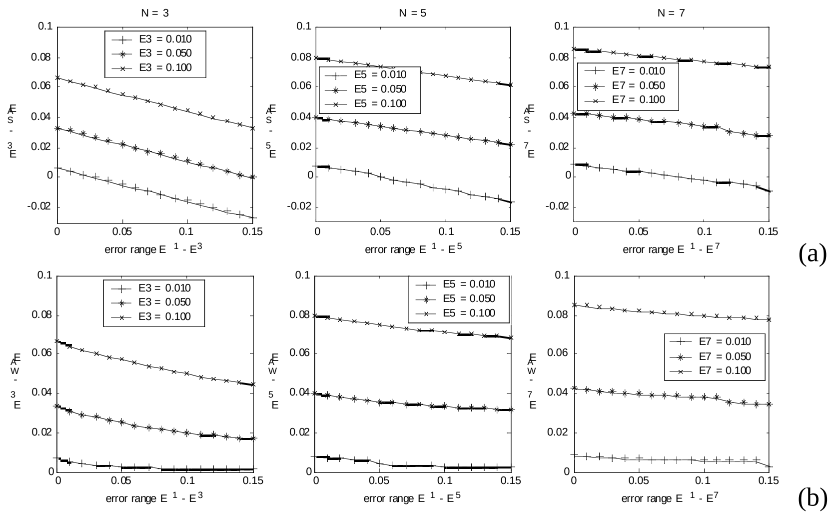


Fig. 2. Values of $E_{add}^{N} - E_{add}^{SA}$ (a), $E_{add}^{N} - E_{add}^{WA}$ (b), and $E_{add}^{SA} - E_{add}^{WA}$ (c), under the conditions of maximum performance imbalance for $E_{add}^{N} - E_{add}^{SA}$ and $E_{add}^{N} - E_{add}^{WA}$.

Fig. 2 shows their values, under the condition of maximum performance imbalance for $E_{add}^{N} - E_{add}^{SA}$ and $E_{add}^{N} - E_{add}^{WA}$ (i.e., $E_{add}^{1} E_{add}^{2} = \ldots = E_{add}^{N-1}$). We considered classifier ensembles of size $N = 3, 5, 7$, an error range between 0 and 0.15, and three different values of the added error of the best individual classifier, E_{add}^{N} (0.01, 0.05 and 0.10). Fig. 2(a) shows that SA can provide a remarkable improvement over the best individual classifier. In particular, the higher is the added error of the best individual classifier, E_{add}^{N}, the higher is the improvement achievable by SA. Such improvement decreases for increasing values of the error range, but this effect becomes negligible when the number of classifiers increases. This is reasonable, since $N-2$ classifiers exhibit the same added error of the best one, E_{add}^{N}. Fig. 2(b) shows that the improvement achievable by WA over E_{add}^{N} is quite similar to that achievable by SA. This is evident from Fig. 2(c), where the values of $E_{add}^{SA} - E_{add}^{WA}$ are reported. Note that for $N > 3$, the improvement of WA over SA is always below 0.5%. Moreover, for $N = 3$, the improvement exceeds 1% only if the best individual classifier performs very well ($E_{add}^{3} = 0.01$), and the others perform very poorly ($E_{add}^{1} = E_{add}^{2} > 0.10$).



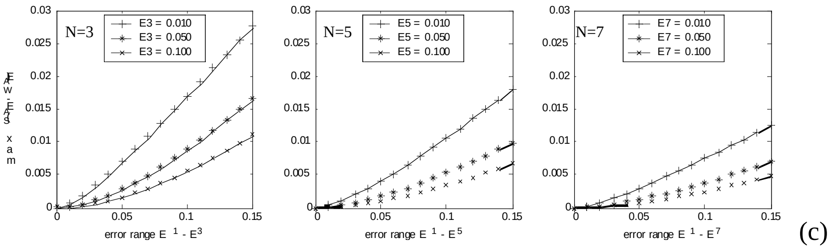


Fig. 3. Values of $E^N_{add} - E^{SA}_{add}$ (a), $E^N_{add} - E^{WA}_{add}$ (b), and $E^{SA}_{add} - E^{WA}_{add}$ (c), under the conditions of maximum performance imbalance for $E^{SA}_{add} - E^{WA}_{add}$.

However, in this case, the performance of WA is very similar to E^3_{add}, and therefore, combining is useless. It is also worth pointing out that, for any fixed error range, and for any *N*, when the improvement of WA over the best individual classifier increases, the corresponding improvement over SA decreases (see Fig. 2(b), (c)).

In Fig. 3 we report the same comparison of Fig. 2, under the conditions of maximum performance imbalance for $E^{SA}_{add} - E^{WA}_{add}$ (see Fig. 1(a)). As expected, Fig. 3(a),(b) shows that the improvement of simple and WA, with respect to the best individual classifier, E^N_{add}, are lower than those reported in Fig. 2(a),(b), while the value of $E^{SA}_{add} - E^{WA}_{add}$ increases with respect to Fig. 2(c). However, even under the condition of maximum performance imbalance for $E^{SA}_{add} - E^{WA}_{add}$, Fig. 3(c) shows that $E^{SA}_{add} - E^{WA}_{add}$ is quite low. For $N = 3$, $E^{SA}_{add} - E^{WA}_{add}$ exceeds 1% only for highly imbalanced classifiers; for instance, if $E^3_{add} = 0.01$, the error range must be higher than 5%; in this case, the corresponding improvement over E^3_{add} is near 0% (see Fig. 3(b)). Moreover, $E^{SA}_{add} - E^{WA}_{add}$ decreases when the number of classifiers increases. The results of the above comparison between simple and WA confirm the ones obtained in our previous works (see section 2), which were limited to ensembles of three classifiers. These results show that WA significantly outperforms both SA and individual classifiers, only for ensembles of few classifiers (say, $N = 3$), with highly imbalanced performances (for instance, one classifier with added error of about 5%, and two classifiers with added errors of at least 15%).

Let us now discuss briefly the practical relevance of the above results for the design of multiple classifier systems based on linear combiners. First of all, they show that the advantage of using WA, instead of SA, can be lower than one can expect. It is worth noting that, in real applications, the optimal weights of the linear combination can only be estimated from finite data. This means that, if the maximum theoretical improvement achievable by WA is low, it could be cancelled if a poor estimate from a small data set is done. Therefore, WA should be used only if large and representative data sets are available, in order to guarantee that its superiority over SA can be really exploited. For small data sets, WA can be used if the classifier ensemble at hand is likely to fit the conditions that guarantee the maximum advantage of WA over SA. This work pointed out such conditions. Finally, it should be noted that the above results have been obtained under the assumption of unbiased and uncorrelated errors, which is likely to be violated in real applications. Therefore, it is interesting to compare the performance of simple and WA on real data sets. Such an experimental comparison is provided in the next section.

4 Experimental results

With the experiments presented in this section, we evaluated the performances of weighted and SA for two real applications. The aim was to compare their performances with that predicted by the theoretical model of section 3, which is based on the strict assumption of unbiased and uncorrelated errors. According to results of our model, we focused on the effects of the difference between the performance of individual classifiers.

We carried out our experiments on a remote-sensing image classification problem (Feltwell data set [8]), and on a character recognition problem (Letter data set, available at http://www.ncc.up.pt/liacc/ML/statlog). The Feltwell data set has five classes, and consists of 5,124 training patterns and 5,820 test patterns, characterised by fifteen features. The Letter data set is made up of 15,000 training patterns and 5,000 test patterns, belonging to 26 classes, and characterised by sixteen features. In our experiments, we used MLP neural network classifiers, with one hidden layer, and a number of input and output units equal to the number of features and classes, respectively. For each data set, in order to obtain classifiers exhibiting an error range at least of 15%, we trained fifty MLPs, characterised by five different numbers of hidden units (15, 5, 4, 3 and 2 for Feltwell, 110, 60, 40, 35 and 30 for Letter). We then constructed sixteen ensembles of three classifiers, each one characterised by a different pattern of the performances of individual classifiers. More precisely, we selected the MLPs exhibiting a test set error rate nearest to predefined values, which are reported in the second column of Tables 1,2. We repeated these experiments by using ten different training sets, obtained by randomly selecting the 80% of patterns from the original training sets. Reported results are averaged over ten runs.

The results are reported in Tables 1, 2. Note that ensembles 1-4 are “balanced” (i.e., the error range of the corresponding classifiers is low), while the others are “imbalanced”, with error ranges of about 5%, 10% and 15%. For imbalanced ensembles, we considered the condition of maximum performance imbalance for $E_{add}^{SA} - E_{add}^{WA}$ (ensembles 6,8,10,12,14,16), and the one of maximum performance imbalance for $E_{add}^{N} - E_{add}^{SA}$ and $E_{add}^{N} - E_{add}^{WA}$ (ensembles 5,7,9,11,13,15).

Let us consider first whether the conditions of maximum performance imbalance, given in section 3, are verified. We recall that these conditions determine when $E_{add}^{SA} - E_{add}^{WA}$ is maximum, and when $E_{add}^{N} - E_{add}^{SA}$ and $E_{add}^{N} - E_{add}^{WA}$ are maximum, for fixed values of the error rate of the best and worst individual classifiers. Accordingly, we compared the ensembles of Tables 1 and 2, characterised by the same values of the error rate of the best and worst individual classifiers (for instance, ensembles 5 and 6). For all such ensembles, it is easy to see that $E_{add}^{SA} - E_{add}^{WA}$ is maximum when the second best individual classifier has the same error rate of the worst classifier. This is the condition of maximum performance imbalance for $E_{add}^{SA} - E_{add}^{WA}$, for $N = 3$, predicted by our theoretical model.

Table 1. Results for the Feltwell data set. Each row correspond to a different classifier ensemble, whose required percentage error rates are reported in the second column. The achieved error range is denoted as ΔE, while E^b, E^{sa} and E^{wa} denote, respectively, the test set error rate of the best individual classifier, and of simple and WA.

Ensemble		ΔE	E^{sa}	E^{wa}	$E^b - E^{sa}$	$E^b - E^{wa}$	$E^{sa} - E^{wa}$
1	10 10 10	0.51	9.91	9.67	0.49	0.73	0,24
2	15 15 15	0.16	12.67	11.94	2.26	3.00	0,73
3	20 20 20	1.33	15.98	14.43	1.9	3.45	1,55
4	25 25 25	0.97	25.11	24.23	-0.45	0.44	0,88
5	10 10 15	4.54	10.32	9.65	0.08	0.75	0,67
6	10 15 15	4.70	11.43	9.90	-1.03	0.51	1,53
7	15 15 20	4.27	13.45	12.42	1.49	2.52	1,03
8	15 20 20	4.27	14.71	12.83	0.23	2.11	1,88
9	20 20 25	6.13	16.44	15.13	2.09	3.40	1,31
10	20 25 25	6.15	21.82	17.77	-2.62	1.44	4,05
11	10 10 20	8.81	10.74	9.69	-0.34	0.71	1,05
12	10 20 20	8.81	12.33	9.80	-1.93	0.60	2,53
13	15 15 25	9.72	13.91	12.98	1.03	1.96	0,93
14	15 25 25	10.42	21.39	14.65	-6.45	0.29	6,74
15	10 10 25	14.26	11.21	9.97	-0.81	0.43	1,24
16	10 25 25	14.96	20.52	10.26	-10.11	0.14	10,26

Table 2. Results for the Letter data set.

Ensemble		ΔE	E^{sa}	E^{wa}	$E^b - E^{sa}$	$E^b - E^{wa}$	$E^{sa} - E^{wa}$
1	15 15 15	0.09	10.19	10.02	4.73	4.90	0,17
2	20 20 20	0.06	14.60	14.37	5.28	5.51	0,23
3	25 25 25	0.44	19.67	19.45	5.52	5.74	0,22
4	30 30 30	0.16	23.64	23.39	6.32	6.57	0,25
5	15 15 20	4.99	11.31	10.99	3.61	3.93	0,32
6	15 20 20	4.93	12.85	12.41	2.16	2.60	0,44
7	20 20 25	5.28	15.87	15.53	4.03	4.37	0,34
8	20 25 25	5.51	17.50	16.95	2.4	2.95	0,55
9	25 25 30	4.83	20.77	20.34	4.42	4.84	0,43
10	25 30 30	4.83	21.79	21.24	3.4	3.94	0,55
11	15 15 25	10.26	12.01	11.37	2.92	3.55	0,64
12	15 25 25	10.41	14.95	13.74	0.06	1.27	1,21
13	20 20 30	10.10	16.51	15.88	3.39	4.03	0,63
14	20 30 30	10.10	19.43	18.23	0.48	1.67	1,20
15	15 15 30	15.09	12.27	11.42	2.65	3.50	0,85
16	15 30 30	15.00	16.47	14.24	-1.46	0.77	2,23

Analogously, the maximum of both $E_{add}^{N} - E_{add}^{SA}$ and $E_{add}^{N} - E_{add}^{WA}$ is achieved when the second best classifier exhibits the same error rate of the best one. Therefore, the theoretical results of section 3, concerning the conditions of maximum performance imbalance, showed to apply to the two real problems considered.

Let us now compare the performance of simple and WA. First, as expected, note that WA always outperformed the best individual classifier. However, consider the improvement achievable by WA over SA, $E_{add}^{SA} - E_{add}^{WA}$, and over the best individual classifier, $E_{add}^{N} - E_{add}^{WA}$, for the same classifier ensemble. Table 1 shows that, for the Feltwell data set, both values are higher than 1% only for five ensembles out of 16, while for the Letter data set (Table 2), this happens only for two ensembles. Moreover, for Letter, both simple and WA almost always outperform the best individual classifier, and exhibit a very similar error rate. By taking into account that the performance of WA can be worsened by weight estimation, these results show that it can be difficult to obtain a significant improvement over both SA and individual classifiers, as argued in section 3.

To sum up, the above experimental results show that the theoretical results of section 3 can be useful to predict the qualitative behaviour of linear combiners in real applications. Moreover, they also seem to confirm that the advantage of using the WA rule, instead of the SA, can be lower than one can expect.

References

1. Fumera, G., Roli, F.: Performance analysis and comparison of linear combiners for classifier fusion. In: Caelli, T., Amin, A., W. Duin, R.P., Kamel, M., de Ridder, D. (eds.): Proc. Joint Int. Workshops on Structural, Syntactic, and Statistical Pattern Recognition. LNCS, Vol. 2396. Springer-Verlag (2002) 424-432
2. Fumera, G., Roli, F., Vernazza, G.: Analysis of error-reject trade-off in linearly combined classifiers. In: Kasturi, R., Laurendeau, D., Suen, C. (eds.): Proc. 16th Int. Conf. on Pattern Recognition. Vol. 2 (2002) 120-123
3. Kittler, J., Alkoot, F. M.: Sum versus Vote Fusion in Multiple Classifier Systems. IEEE Trans. on Pattern Analysis and Machine Intelligence **25** (2003) 110-115
4. Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On combining classifiers. IEEE Trans. on Pattern Analysis and Machine Intelligence **20** (1998) 226-239
5. Kuncheva, L. I.: A theoretical study on six classifier fusion strategies. IEEE Trans. on Pattern Analysis and Machine Intelligence **24** (2002) 281-286
6. Lam, L., Suen, C.Y.: Application of majority voting to pattern recognition: an analysis of its behavior and performance. IEEE Trans. on Systems, Man and Cybernetics - Part A **27** (1997) 553-568
7. Perrone, M.P., Cooper, L.N.: When Networks Disagree: Ensemble Methods for Hybrid Neural Networks. In: Mammone, R.J. (ed.): Neural Networks for Speech and Image Processing. Chapman-Hall, New York (1993)
8. Roli, F.: Multisensor image recognition by neural networks with understandable behaviour. Int. Journal of Pattern Recognition and Artificial Intelligence **10** (1996) 887-917
9. Roli, F., Fumera, G.: Analysis of Linear and Order Statistics Combiners for Fusion of Imbalanced Classifiers. In: Kittler, J., Roli, F. (eds.): Proc. of the 3rd Int. Workshop on Multiple Classifier Systems. LNCS, Vol. 2364. Springer-Verlag (2002) 252-261
10. Tumer, K., Ghosh, J.: Analysis of decision boundaries in linearly combined neural classifiers. Pattern Recognition **29** (1996) 341-348
11. Tumer, K., Ghosh, J.: Linear and order statistics combiners for pattern classification. In: Sharkey, A.J.C. (ed.): Combining Artificial Neural Nets. Springer, London (1999) 127-155