Increasing the Number of Classifiers in Multi-classifier Systems: A Complementarity-Based Analysis

L. Bovino[2], G. Dimauro[1], S. Impedovo[1], G. Pirlo[1], and A. Salzo[2]

[1] Dipartimento di Informatica, Università degli Studi di Bari
Via Orabona 4, 70126 Bari-Italy
impedovo@di.uniba.it
[2] Consorzio Interuniversitario Nazionale per l'Informatica (CINI),
Via G.Petroni 15/F.1, 70124 Bari– Italy

Abstract. Complementarity among classifiers is a crucial aspect in classifier combination. A combined classifier is significantly superior to the individual classifiers only if they strongly complement each other. In this paper a complementarity-based analysis of sets of classifier is proposed for investigating the behaviour of multi-classifier systems, as new classifiers are added to the set. The experimental results confirm the theoretical evidence and allow the prediction of the performance of a multi-classifier system, as the number of classifiers increases.

1 Introduction

Complementarity among classifiers is crucial in classifier combination. In fact, classifier combination significantly outperforms individual classifiers only if they are largely complementary each other. Complementarity among classifiers can be achieved by using different feature sets and classification strategies [1,2]. Alternatively, complementarity is also expected when different training sets and resampling strategies are used [3,4,5,6].

In this paper a complementarity-based analysis of sets of classifier is used for investigating the behaviour of multi-classifier systems, as new classifiers are added to the set. The result allows the prediction of the effect of increasing the number of classifiers on the performance of multi-classifier systems. The experimental tests, which have been carried out in the field of hand-written numeral recognition, confirm the expected performance of the combination method and validate the proposed approach.

The paper is organised as follows: Section 2 introduces an estimator of complementarity for *abstract-level* classifiers. Section 3 shows the complementarity of a set of classifiers, as the number of classifiers increases. Section 4 presents the methodology used for the analysis of combination methods. The experimental results are discussed in Section 5.

D. Lopresti, J. Hu, and R. Kashi (Eds.): DAS 2002, LNCS 2423, pp. 145–156, 2002.

2 Complementarity among Classifiers

In order to measure the degree of complementarity among *abstract-level* classifiers, the *Similarity Index* has been recently introduced [7]. Let A={A_1 , A_2} a set of two classifiers and P = {P_t | t=1,2,...,N} a set of patterns and let $A_i(P_t)$ be the class label produced by A_i for the input pattern P_t. The *Similarity Index* ρ_A for the set {A_1,A_2} is defined as:

$$\rho_{\{A_1, A_2\}} = \frac{1}{N} \sum_{t=1}^{N} Q\,(A_1(P_t),\, A_2(P_t)) \tag{1}$$

and

$$Q(A_1(P_t),A_2(P_t))=\begin{cases}1 & \text{if } A_1(P_t)=A_2(P_t)\\ 0 & \text{otherwise}\end{cases} \tag{2}$$

Of course, $\rho_A \in [0,1]$: when ρ_A is close to 0, classifiers are strongly complementary; when ρ_A is close to 1, classifiers are weakly complementary. Figure 1 shows the outputs of two classifiers A_1, A_2, for N=10 input patterns $P_1,P_2,\ldots,P_{10}$. Recognitions are indicated by R, substitutions by the labels S1, S2,S3 (with Si≠Sj, ∀ i≠j). In this case the recognition rates for A_1 and A_2 are R_1=0.7 and R_2=0.6, respectively. The degree of complementarity between A_1 and A_2 is ρ_{A_1,A_2}= 0.6. In fact:

- P_1,P_2,P_3,P_6,P_7 are recognised by both classifiers ($Q(A_1(P_t),A_2(P_t))=1$, t=1,2,3,6,7);
- P_4 is substituted by both classifiers which provide different responses: $A_1(P_4)$=S1, A_2 (P_4)=S3 ($Q(A_1(P_4),A_2(P_4))=0$);
- P_5 is substituted by both classifiers which provide the same response: $A_1(P_5)=A_2(P_5)$=S2 ($Q(A_1(P_5),A_2(P_5))=1$);
- P_8 and P_{10} are recognized by A_1 and substituted by A_2: $A_1(P_8)$=R and $A_2(P_8)$=S2, $A_1(P_{10})$=R and $A_2(P_{10})$=S1 ($Q(A_1(P_t),A_2(P_t))=0$, t=8,10);
- P_9 is substituted by A_1 and recognised by A_2: $A_1(P_9)$=S3, $A_2(P_9)$=R ($Q(A_1(P_9),A_2(P_9))=0$).

	A_1	A_2
Pattern 1	R	R
Pattern 2	R	R
Pattern 3	R	R
Pattern 4	S1	S3
Pattern 5	S2	S2
Pattern 6	R	R
Pattern 7	R	R
Pattern 8	R	S2
Pattern 9	S3	R
Pattern 10	R	S1

Fig. 1. List of output of two classifiers

In general, let A ={A_i | i=1,2,...,K} be a set of classifiers, P = {P_t | t=1,2,...,N} a set of patterns the *Similarity Index* ρ_A for the set A is defined as [7]:

$$\rho_A = \frac{\sum\limits_{\substack{i,\, j=1,\ldots K \\ i<j}} \rho_{\{A_i, A_j\}}}{\binom{K}{2}} . \tag{3}$$

3 Increasing the Number of Classifiers: Analysis of Complementarity

Let A ={A_i | i=1,2,...,K} be a set of classifiers with *Similarity Index* equal to ρ_A, and suppose that a new classifier A_{K+1} is added to the set. The *Similarity Index* of $A\cup\{A_{K+1}\}$ is (see eq. (3)):

$$\rho_{A\cup\{A_{K+1}\}} = \frac{\sum\limits_{\substack{i,\, j=1,\ldots K+1 \\ i<j}} \rho_{A_i,A_j}}{\binom{K+1}{2}} = \frac{\sum\limits_{\substack{i,\, j=1,\ldots K \\ i<j}} \rho_{A_i,A_j} + \sum\limits_{i=1,\ldots K} \rho_{A_i,A_{K+1}}}{\binom{K+1}{2}}$$

$$\frac{\sum\limits_{\substack{i,\, j=1,\ldots K \\ i<j}} \rho_{A_i,A_j}}{\binom{K+1}{2}} + \frac{\sum\limits_{i=1,\ldots K} \rho_{A_i,A_{K+1}}}{\binom{K+1}{2}} = \frac{\sum\limits_{\substack{i,\, j=1,\ldots K \\ i<j}} \rho_{A_i,A_j}}{\frac{(K+1)}{(K-1)}\binom{K}{2}} + \frac{\sum\limits_{i=1,\ldots K} \rho_{A_i,A_{K+1}}}{\binom{K+1}{2}}$$

$$= \frac{(K-1)}{(K+1)} \frac{\sum\limits_{\substack{i,\, j=1,\ldots K \\ i<j}} \rho_{A_i,A_j}}{\binom{K}{2}} + \frac{\sum\limits_{i=1,\ldots K} \rho_{A_i,A_{K+1}}}{\frac{(K+1)K}{2}} .$$

Hence:

$$\rho_{A\cup\{A_{K+1}\}} = \frac{K-1}{K+1}\rho_A + \frac{2}{K(K+1)} \sum_{i=1,\ldots,K} \rho_{A_i, A_{K+1}} \tag{4}$$

Of course, the variability of $\rho_{A\cup\{A_{K+1}\}}$ depends on

$$\sum_{i=1,\ldots,K} \rho_{A_i, A_{K+1}} \tag{5}$$

In order to estimate to what extent the quantity (5) can vary, the relationships between $\rho_{A_i A_j}$ (A_i, A_j $\in$ A) and $\rho_{A_i A_{K+1}}$, $\rho_{A_j A_{K+1}}$ (due to the extra classifier A_{K+1}) are determined in the following. For this purpose, from now on we suppose that all classifiers have similar performance, i.e. all of them have recognition rate equal to R.

3.1 Analysis of the Complementarity between A_i and A_j.

Let A_i and A_j be two classifiers of A with *Similarity Index* equal to ρ_{A_i,A_j} (Figure 2). The analysis of the outputs of A_i and A_j leads to the following cases:

[A] $A_i(t)=Si$, $A_j(t)=Sj$ ($Si=Sj$);
[B] $A_i(t)=Si$, $A_j(t)= Sj$ ($Si \neq Sj$);
[C] $A_i(t)= Si$, $A_j(t)= R$;
[D] $A_i(t)=R$, $A_j(t)= R$;
[E] $A_i(t)= R$, $A_j(t)= Sj$.

Now, let P_A, P_B, P_C, P_D, P_E be the percentage of patterns corresponding to the cases A,B,C,D,E respectively, the following equations hold:

- ❖ $P_A+P_B+P_C+P_D+P_E=1$; (6)
- ❖ $P_C+P_D=R$ and $P_D+P_E=R$ (7)
- ❖ $P_A+P_D= \rho_{Ai,Aj}$ (the quantity $\rho_{Ai,Aj}$ concerns all the cases in which the decisions of A_i and A_j agree, i.e. cases (A) and (D)). (8)

From eq. (8) it follows that:

$$P_A=\delta \tag{9}$$

$$P_D=\rho_{Ai,Aj} -\delta \tag{10}$$

where δ is a positive quantity ($\delta<\rho_{Ai,Aj}$).
Moreover, from (7), (9) and (10), it results:

$$P_C=P_E=R-P_D=R-(\rho_{Ai,Aj}-\delta), \tag{11}$$

while from (6),(9),(10) and (11):

$$P_B=1-P_A-P_C-P_D-P_E=1-\delta-(R-(\rho_{Ai,Aj} -\delta))-(\rho_{Ai,Aj} -\delta)-(R-(\rho_{Ai,Aj} -\delta))=1-2R+\rho_{Ai,Aj} -2\delta \,. \tag{12}$$

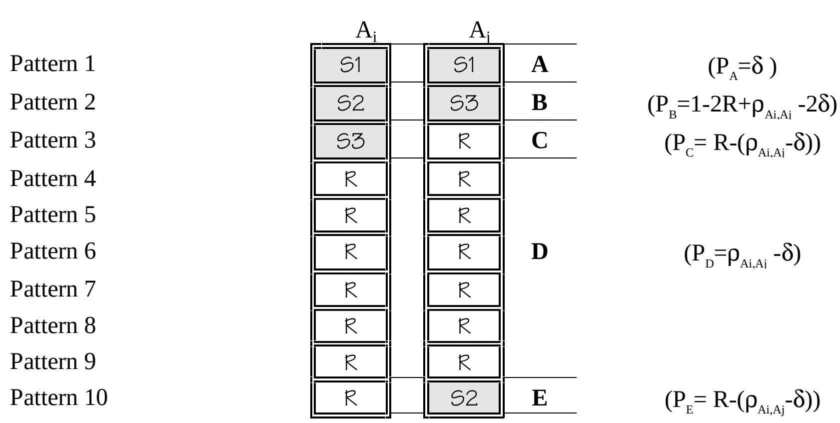


Fig. 2. Analysis of complementarity between the classifiers A_i and A_j

3.2 Analysis of the Complementarity between A_{K+1} and A_i , A_j.

When the new classifier A_{K+1} is considered, two cases must be examined concerning respectively the minimum (Case (a)) and the maximum (Case (b)) value of the quantity (5):

Case (a). In this case the outputs of A_{K+1} must be as complementary as possible to those of A_i and A_j. Hence, the recognitions of A_{K+1} must occur according to the following priorities (see Fig. 3):

a.1) both A_i and A_j substitute the patterns (cases A and B). For this case, the contribution of A_{K+1} to the *Similarity Index* is null since A_{K+1} disagrees both with A_i and A_j. The percentage of patterns concerning (a.1) is P_A+P_B at the best.

a.2) A_i or A_j substitute the patterns (cases C and E). In this case A_{K+1} agrees with A_i or A_j. Therefore the contribution due to each pattern recognized by A_{K+1} is weighted by 1. The percentage of patterns concerning (a.2) is P_C+P_E at the best.

a.3) both A_i and A_j recognise the patterns (case D). In this case A_{K+1} agrees both with A_i and A_j. Therefore the contribution due to each pattern recognized by A_{K+1} is weighted by 2. The percentage of patterns concerning (a.3) is P_{D_2} at the best, where P_{D_2}=R-P_A-P_B-P_C-P_E (if we assume the common condition: R>P_A-P_B-P_C-P_E).

Concerning substitutions, it must be assumed that A_{K+1} always provides substitutions as different as possible from those of A_i and A_j. Hence it results (see figure 3):

$$\rho_{A_i,A_{K+1}} + \rho_{A_j,A_{K+1}} = 0\cdot(P_A+P_B)+1\cdot(P_C+P_E)+2\cdot P_{D_2} = P_C+P_E+2\cdot P_{D_2} \qquad (13)$$

where

P_c is due to patterns recognised by A_{K+1} and A_j, and substituted by A_i;
P_{D_2} is due to patterns recognised by A_{K+1} , A_i and A_j;
P_E is due to patterns recognised by A_{K+1} and A_i, and substituted by and A_j.

	A_i	A_j	A_{K+1}		
Pattern 1	S1	S1	R	**A**	
Pattern 2	S2	S3	R	**B**	
Pattern 3	S3	R	R	**C**	
Pattern 4	R	R	S1		
Pattern 5	R	R	S2	**D_1**	
Pattern 6	R	R	S3		
Pattern 7	R	R	R		**D**
Pattern 8	R	R	R	**D_2**	
Pattern 9	R	R	R		
Pattern 10	R	S2	R	**E**	

Fig. 3. Analysis of complementarity among A_{K+1} and A_i , A_j- Case (a)

Substituting eqs. (9),(10),(11) and (12) in (13) it results:

$$\rho_{A_i,A_{K+1}}+\rho_{A_j,A_{K+1}}=2(R-(\rho_{Ai,Aj}-\delta))+2[R-\delta-(1-2R+\rho_{Ai,Aj}-2\delta)-2(R-(\rho_{Ai,Aj}-\delta))]=2(2R-1) \qquad (14)$$

Case (b). In this case the outputs of A_{K+1} must be as similar as possible to those of A_i and A_j.Hence,the recognitions of A_{K+1} must occur according to the following priorities (see Fig. 4):

b.1) both A_i and A_j recognise the patterns (case D). For this case the contribution of A_{K+1} to the *Similarity Index* is weighted by 2, since A_{K+1} agrees both with A_i and A_j. The percentage of patterns concerning (b.1) is P_D at the best.

b.2) A_i or A_j substitute the patterns (cases C and E). In this case A_{K+1} agrees with A_i or A_j. Therefore the contribution due to each pattern recognized by A_{K+1} is weighted by 1. The percentage of patterns concerning (b.2) is P_C+P_E at the best.

b.3) both A_i and A_j substitute the patterns (cases A and B). For these cases the contribution of A_{K+1} to the *Similarity Index* is null since A_{K+1} disagrees both with A_i and A_j. The percentage of patterns concerning (b.1) is P_A+P_B at the best.

Concerning substitutions, A_{K+1} must provides substitutions as similar as possible to those of A_i and A_j. Precisely:

b'.1) if $A_i(t)=A_j(t)=Si$ then it must results that $A_{K+1}(t)= Si$. For this case the contribution to the *Similarity Index* due to each pattern recognized by A_{K+1} is weighted by 2 since A_{K+1} agree both with A_i and A_j. The percentage of patterns concerning (b'.1) is P_A at the best.

b'.2) if $A_i(t)= Si$ and $A_j(t)= Sj$ then it must results that $A_{K+1}(t)= Si$ (or equivalently $A_{K+1}= Sj$). For this case the contribution to the *Similarity Index* due to each pattern recognized by A_{K+1} is weighted by 1 since A_{K+1} agrees with A_i (or A_j). The percentage of patterns concerning (b'.2) is P_B at the best (or equivalently P_E).

In this case we obtain (see figure 4):

$$\rho_{A_i,A_{K+1}} + \rho_{A_j,A_{K+1}}=2\cdot P_D+1\cdot(P_C+ P_E)+ 2\cdot P_A+1\cdot P_B=2\cdot P_A+P_B+P_C+2\cdot P_D+P_E \qquad (15)$$

where:

- P_A is due to patterns substituted by A_{K+1} , A_i and A_j with the same class label;
- P_B is due to patterns substituted by A_{K+1} and A_i with the same class label, and by A_j with a different class label;
- P_C is due to patterns recognised by A_{K+1} and A_i , and substituted by A_j;
- P_D is due to pattern recognised by A_{K+1} , A_i and A_j;
- P_E is due to pattern recognised by A_{K+1} and A_j, and substituted by and A_i.

	A_i	A_j	A_{K+1}	
Pattern 1	S1	S1	S1	**A**
Pattern 2	S2	S3	S2	**B**
Pattern 3	S3	R	S3	**C**
Pattern 4	R	R	R	
Pattern 5	R	R	R	
Pattern 6	R	R	R	**D**
Pattern 7	R	R	R	
Pattern 8	R	R	R	
Pattern 9	R	R	R	
Pattern 10	R	S2	R	**E**

Fig. 4. Analysis of complementarity among A_{K+1} and A_i , A_j- Case (b)

Substituting eqs. (9),(10),(11) and (12) in (15) it results:

$$\rho_{A_i,A_{K+1}}+\rho_{A_j,A_{K+1}}=2(\rho_{Ai,Aj}-\delta)+2\delta+(R-(\rho_{Ai,Aj}-\delta))+(1-2R+\rho_{Ai,Aj}-2\delta)=1+\rho_{Ai,Aj} \qquad (16)$$

3.3 Analysis of the Complementarity of $A \cup \{A_{K+1}\}$.

From eqs. (14) and (16) it follows that, $\forall$i,j=1,2,…,N:

$$2(2R-1) \le \rho_{A_i,A_{K+1}} + \rho_{A_j,A_{K+1}} \le 1+\rho_{Ai,Aj}. \tag{17}$$

Adding the inequalities (17), for i,j=1,2,…,N, i<j, it results:

$$\sum_{\substack{i,j=1\\i<j}}^{K} 2(2R-1) \le \sum_{\substack{i,j=1\\i<j}}^{K} (\rho_{A_i,A_{K+1}} + \rho_{A_j,A_{K+1}}) \le \sum_{\substack{i,j=1\\i<j}}^{K} (1+\rho_{A_i,A_j})$$

$$\binom{K}{2} 2(2R-1) \le (K-1)\sum_{i=1}^{K} \rho_{A_i,A_{K+1}} \le \binom{K}{2} + \sum_{\substack{i,j=1\\i<j}}^{K} \rho_{A_i,A_j}$$

$$\binom{K}{2} 2(2R-1) \le (K-1)\sum_{i=1}^{K} \rho_{A_i,A_{K+1}} \le \binom{K}{2} + \binom{K}{2}\rho_A$$

$$\frac{K(K-1)}{2} 2(2R-1) \le (K-1)\sum_{i=1}^{K} \rho_{A_i,A_{K+1}} \le \frac{K(K-1)}{2}(1+\rho_A)$$

$$(2R-1)K \le \sum_{i=1}^{K} \rho_{A_i,A_{K+1}} \le \frac{K}{2}(1+\rho_A) \tag{18}$$

Substituting expression (18) in (4) we obtain that the range of variability of the *Similarity Index,* when a new classifier is added to the set A, is given by:

$$\rho_{A\cup\{A_{K+1}\}} \in [\mathrm{Min}\ \rho_{A\cup\{A_{K+1}\}},\ \mathrm{Max}\ \rho_{A\cup\{A_{K+1}\}}] ,$$

where:

- ✓ Min $\rho_{A\cup\{A_{K+1}\}} = \frac{K-1}{K+1}\rho_A + \frac{2}{K(K+1)}(2R-1)K = \frac{K-1}{K+1}\rho_A + 2\frac{2R-1}{(K+1)}$; (19)
- ✓ Max $\rho_{A\cup\{A_{K+1}\}} = \frac{K-1}{K+1}\rho_A + \frac{2}{K(K+1)}\frac{K}{2}(1+\rho_A) = \frac{K}{K+1}\rho_A + \frac{1}{(K+1)}$. (20)

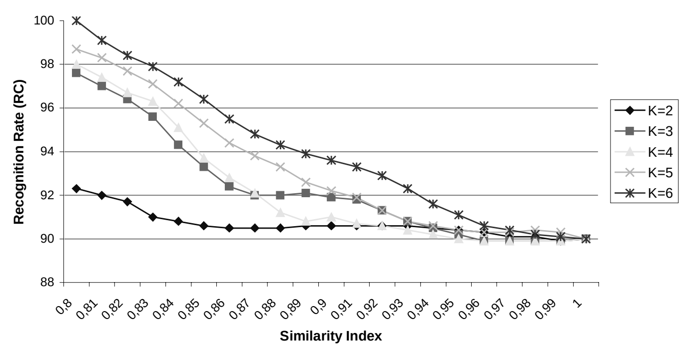
4 Analysis of Combination Methods

Although classifier combination is widely applied in many fields, theoretical analysis of combination schemes can be very difficult. The net result is that only simple combination have been explained up to now from a theoretical point of view [8]. In many cases the performance of a combination method cannot be estimated theoretically and it can be evaluated on experimental basis in specific working conditions (a specific set of classifiers, training data and sessions, etc.). In this case the result depends on the specific conditions of the test and no information can be derived on the performance of the combination method if the working conditions change. A different approach to estimate systematically the performance of a combination method **C** for abstract-level classifiers is based on the simulation of various sets of classifiers which are used to test the method under different conditions

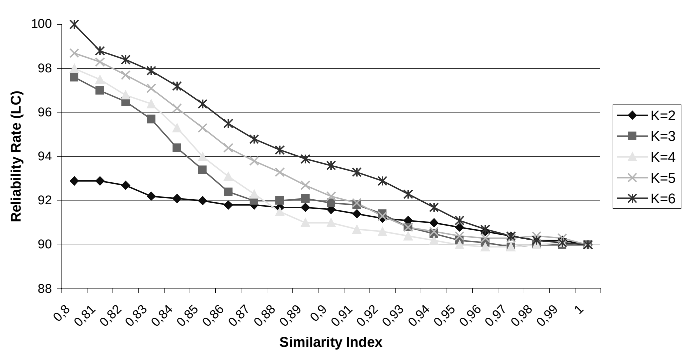
[9]. In this case, performance of **C,** which combines K abstract-level classifiers, is evaluated as a function of the recognition rate of the classifiers (R) and the degree of complementarity among them (ρ):

$$\mathbf{C}\ (K,R,\rho) \rightarrow (R_{\mathbf{C}}, L_{\mathbf{C}}) \tag{21}$$

where $R_{\mathbf{C}}$ and $L_{\mathbf{C}}$ are respectively the recognition rate and the reliability rate of **C** [1]. More precisely, since *abstract-level* classifiers are combined, each individual classifier is considered as a discrete random variable whose outputs are N class labels if N patterns are supposed to be input: N·R recognitions (labels equal to R) and N·(1-R) substitutions (labels equal to S1,S2,S3,…). Of course, for any 3-tuple (K,R,ρ), several sets (50 in out tests) of classifiers are simulated and used to test the combination method **C**, in order to estimate its mean performance in terms of R_C and L_C.



(a)



(b)

Fig. 5. Performance of DS as a function of ρ in combining K classifiers (R=90%)

In this work, the behaviour of the Dempster-Shafer (DS) combination method is analysed [10]. Specifically, we use the DS combination scheme and the decision rule proposed respectively in Section VI.C and Section VI.D (eq. [50], α=0) of ref. [1]. The performance of DS is reported in Figure 5 as a function of ρ, when sets of K classifiers are combined (K=2,3,4,5,6), each one with a recognition rate equal to R=90% (see ref. [11] for more details).

5 Experimental Results

This Section shows the analysis of complementarity of a set of classifiers, as the number of classifiers increases. Based on this result, the performance of the Dempster-Shafer (DS) method in combining classifiers is investigated. Two cases are discussed hereafter.

Case (a). In this case four initial sets of classifiers $A = \{A_i \mid i=1,2,...,K\}$, for K=2,3,4,5, are given. The recognition rate of the classifiers is R=90% and the degree of complementarity of each set is ρ_A=0.85.

Table 1. DS Performance as the number of classifier increases: Case (a)

K=2		
ρ	R_C	L_C
0,85	**90,6**	**92,0**

K=3		
ρ	R_C	L_C
0,82	96,4	96,5
0,83	95,6	95,7
0,84	94,3	94,4
0,85	93,3	93,4
0,86	92,4	92,4
0,87	**92,0**	**92,0**
0,88	**92,0**	**92,0**
0,89	92,1	92,1
0,90	91,9	91,9

(a) 2→3

K=3		
ρ	R_C	L_C
0,85	**93,3**	**93,4**

K=4		
ρ	R_C	L_C
0,83	96,3	96,4
0,84	95,1	95,3
0,85	93,7	94,0
0,86	92,8	93,1
0,87	92,1	92,3
0,88	91,2	91,5
0,89	90,8	91,0

(b) 3→4

K=4		
ρ	R_C	L_C
0,85	**93,7**	**94,0**

K=5		
ρ	R_C	L_C
0,83	97,1	97,1
0,84	96,2	96,2
0,85	95,3	95,3
0,86	94,4	94,4
0,87	93,8	93,8
0,88	93,3	93,3

(c) 4→5

K=5		
ρ	R_C	L_C
0,85	**95,3**	**95,3**

K=6		
ρ	R_C	L_C
0,83	97,9	97,9
0,84	97,2	97,2
0,85	96,4	96,4
0,86	95,5	95,5
0,87	94,8	94,8
0,88	94,3	94,3

(d) 5→6

Table 1 reports the effect of adding one extra classifier to each set. Eqs. (19) and (20) are used to determine the range of variability of the degree of complementarity , while the results in Fig.5 allows the prediction of the performance of the DS method:

- for K=2 (Table 1a), DS performance is equal to $\mathbf{R}_C$= 90.6, $\mathbf{L}_C$= 92.0 (Fig. 5). If an extra classifier A_{K+1} is added to A, the complementarity of $A\cup\{A_{K+1}\}$ is in the range [0.82, 0.90] (eqs.(19),(20)). Hence it results that the expected performance for $A\cup\{A_{K+1}\}$ ranges from $\mathbf{R}_C$=91.9, $\mathbf{L}_C$=91.9 (Fig.5, for $\rho_{A\cup\{A_{K+1}\}}$=0.90) to $\mathbf{R}_C$= 96.4, $\mathbf{L}_C$=96.5 (Fig.5, for $\rho_{A\cup\{A_{K+1}\}}$=0.82).
- for K=3 (Table 1b), DS performance is equal to $\mathbf{R}_C$= 93.3, $\mathbf{L}_C$= 93.4 (Fig. 5). If an extra classifier A_{K+1} is added to A, the complementarity of $A\cup\{A_{K+1}\}$ is in the range [0.83, 0.89] (eqs.(19),(20)) and from Fig. 5 it results that the expected

performance for $A \cup \{A_{K+1}\}$ ranges from $\mathbf{R_C}$=90.8, $\mathbf{L_C}$=91.0 (Fig. 5, for $\rho_{A\cup\{A_{K+1}\}}$=0.89) to $\mathbf{R_C}$=96.3, $\mathbf{L_C}$=96.4 (Fig. 5, for $\rho_{A\cup\{A_{K+1}\}}$=0.83).

Similar considerations lead to the results in Table 1c,d.

Case (b). In this case the initial set of K=2 classifiers A ={A_1 , A_2 } is given, with R=90% and ρ_A=0.85, and four extra classifiers A_3, A_4, A_5 and A_6 are added to the set A, one after the other (es. 2→3→4→5→6). In this case eqs (19) and (20) must be applied by an iterative scheme, in order to predict the range of variability of the enlarged sets of classifiers:

- when A_3 is added to the set A, the degree of complementarity $\rho_{A\cup\{A3\}}$ can varies in the range $\rho_{A\cup\{A3\}} \in [0.82,0.90]$ (eqs. (19),(20)) and from Fig.5 the performance of DS ranges from $\mathbf{R_C}$=91.9, $\mathbf{L_C}$=91.9 (for $\rho_{A\cup\{A_{K+1}\}}$=0.90) to $\mathbf{R_C}$= 96.4, $\mathbf{L_C}$=96.5 (for $\rho_{A\cup\{A_{K+1}\}}$=0.82).
- when A_4 is added to $A\cup\{A_3\}$, the degree of complementarity of the set $A\cup\{A_3\} \cup\{A_4\}$ can varies in the range $\rho_{A\cup\{A3\}\cup\{A4\}} \in [0.81,0.93]$ (where, of course, the lower bound of $\rho_{A\cup\{A3\}\cup\{A4\}}$ is obtained by applying eq. (19) to the lower bound of $\rho_{A\cup\{A3\}}$, and the upper bound of $\rho_{A\cup\{A3\}\cup\{A4\}}$ is obtained by applying eq. (20) to the upper bound of $\rho_{A\cup\{A3\}}$). For the set $A\cup\{A_3\} \cup\{A_4\}$, Fig. 5 shows that the expected performance of DS ranges from $\mathbf{R_c}$=90.4, $\mathbf{L_C}$=90.4 (for $\rho_{A\cup\{A3\}\cup\{A4\}}$=0.93) to $\mathbf{R_c}$= 97.4, $\mathbf{L_c}$=97.5 (for $\rho_{A\cup\{A3\}\cup\{A4\}}$=0.82).

This procedure is bring to the end, in order to obtain the results in Table 2.

Table 2. DS Performance as the number of classifier increases: Case (b).

	K=2 (A={A_1,A_2})		K=3 ($A\cup A_3$)		K=4 ($A\cup A_3\cup A_4$)		K=5 ($A\cup A_3\cup A_4\cup A_5$)		K=6 ($A\cup A_3\cup A_4\cup A_5\cup A_6$)	
ρ	$\mathbf{R_C}$	$\mathbf{L_C}$	$\mathbf{R_C}$	$\mathbf{L_C}$	$\mathbf{R_C}$	$\mathbf{L_C}$	$\mathbf{R_C}$	$\mathbf{L_C}$	$\mathbf{R_C}$	$\mathbf{L_C}$
0,80									100	100
0,81					97,4	97,5	98,3	98,3	99,1	99,1
0,82			96,4	96,5	96,7	96,8	97,7	97,7	98,4	98,4
0,83			95,6	95,7	96,3	96,4	97,1	97,1	97,9	97,9
0,84			94,3	94,4	95,1	95,3	96,2	96,2	97,2	97,2
0,85	**90,6**	**92,0**	93,3	93,4	93,7	94,0	95,3	95,3	96,4	96,4
0,86			92,4	92,4	92,8	93,1	94,4	94,4	95,5	95,5
0,87			**92,0**	**92,0**	**92,1**	**92,3**	93,8	93,8	94,8	94,8
0,88			**92,0**	**92,0**	**91,2**	**91,5**	**93,3**	**93,3**	94,3	94,3
0,89			92,1	92,1	**90,8**	**91,0**	**92,6**	**92,7**	**93,9**	**93,9**
0,90			91,9	91,9	91,0	91,0	92,2	92,2	93,6	93,6
0,91					90,7	90,7	91,9	91,9	93,3	93,3
0,92					90,6	90,6	91,3	91,3	92,9	92,9
0,93					90,4	90,4	90,8	90,8	92,3	92,3
0,94							90,6	90,6	91,6	91,7
0,95									91,1	91,1

Finally, a multi-classifier system for hand-written numeral recognition has been considered. The system combines by DS up to six classifiers trained on 12.000 digits extracted from courtesy amounts on bank-checks [12]: $\mathbf{A_1}$-Region, $\mathbf{A_2}$-Crossing, $\mathbf{A_3}$-Contour Slope, $\mathbf{A_4}$-Enhanced Loci, $\mathbf{A_5}$-Histogram, $\mathbf{A_6}$-Local Contour. Each classifiers outputs a single class label and no rejection is allowed at the level of individual classifiers. Moreover the recognition rate of each classifier is about 90% (differences are less than 0.4%). Table 3 reports the values of ρ for each subset (K=2,3,4,5,6) of classifiers. It is easy to verify from Tables 1 and 2 that the complementarity measured on real sets of classifiers, as the number of classifiers increases, is consistent with the results determined in eqs. (19), (20) (for instance, the particular case of adding new classifiers to the set $\{A_4, A_6\}$, for which ρ_A =0.85, is reported in bold type in Table 3 and in Tables 1,2). Finally, the effect of increasing the number of classifier on the performance of the multi-classifier system has been evaluated. It results that the differences between predicted and real recognition rate is less than 1.0%, while it is less than 1.3% in terms of reliability rate.

Table 3. Degree of Complementarity of sets of classifiers

K=2		K=3		K=4		K=5		K=6	
A	ρ_A	**A**	ρ_A	**A**	ρ_A	**A**	ρ_A	**A**	ρ_A
A_4,A_6	**0,85**	A_1,A_2,A_6	0,87	**A_2,A_3,A_4,A_6**	**0,87**	**A_1,A_2,A_3,A_4,A_6**	**0,88**	**A_1,A_2,A_3,A_4,A_5,A_6**	**0,89**
A_1,A_6	0,86	A_2,A_3,A_6	0,87	**A_1,A_2,A_4,A_6**	**0,88**	A_1,A_2,A_3,A_5,A_6	0,89		
A_2,A_6	0,87	**A_3,A_4,A_6**	**0,87**	**A_1,A_3,A_4,A_6**	**0,88**	**A_1,A_2,A_4,A_5,A_6**	**0,89**		
A_3,A_6	0,87	**A_2,A_4,A_6**	**0,87**	**A_2,A_4,A_5,A_6**	**0,88**	**A_1,A_3,A_4,A_5,A_6**	**0,89**		
A_1,A_2	0,88	A_2,A_3,A_4	0,88	A_1,A_2,A_3,A_4	0,89	**A_2,A_3,A_4,A_5,A_6**	**0,89**		
A_2,A_3	0,88	**A_4,A_5,A_6**	**0,88**	A_1,A_2,A_3,A_6	0,89	A_1,A_2,A_3,A_4,A_5	0,90		
A_2,A_4	0,88	A_1,A_3,A_6	0,88	A_1,A_2,A_5,A_6	0,89				
A_5,A_6	0,88	**A_1,A_4,A_6**	**0,88**	A_1,A_3,A_5,A_6	0,89				
A_3,A_4	0,89	A_1,A_5,A_6	0,88	**A_1,A_4,A_5,A_6**	**0,89**				
A_1,A_3	0,90	A_1,A_2,A_3	0,89	A_2,A_3,A_5,A_6	0,89				
A_4,A_5	0,90	A_1,A_2,A_4	0,89	**A_3,A_4,A_5,A_6**	**0,89**				
A_1,A_5	0,91	A_2,A_5,A_6	0,89	A_1,A_2,A_4,A_5	0,90				
A_1,A_4	0,92	A_1,A_3,A_4	0,90	A_2,A_3,A_4,A_5	0,90				
A_2,A_5	0,92	A_2,A_4,A_5	0,90	A_1,A_2,A_3,A_5	0,91				
A_3,A_5	0,95	A_3,A_5,A_6	0,90	A_1,A_3,A_4,A_5	0,91				
		A_1,A_2,A_5	0,91						
		A_2,A_3,A_5	0,91						
		A_3,A_4,A_5	0,91						
		A_1,A_4,A_5	0,91						
		A_1,A_3,A_5	0,92						

6 Conclusion

This paper presents a complementarity-based analysis of sets of abstract-level classifiers and uses the results to investigate the performance of multi-classifier systems, as the number of classifiers increases. This work clarifies important aspects of the collective behaviour of multiple classifiers systems, based on the analysis of complementarity among them.

Acknowledgements. This paper has been supported by the Italian Ministry "Ministero dell'Istruzione, dell'Università e della Ricerca ", MIUR under law 488, Project "Rete Puglia" CINI-BA - D.M. n.600, 11-11-1999.

References

1. Ley Xu, Adam Krzyzak, Ching Y-Suen, "Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition", *IEEE Transaction on Systems, Man and Cybernetics*, Vol. 22, N. 3, 1992, pp. 418–435.

2. L. Lam, Y.-S. Huang, C.Y. Suen, “Combination of Multiple Classifier Decision for Optical Character Recognition”, in *Handbook of Character Recognition and Document Image Analysis*, Eds. H. Bunke and P.S.P.Wang, World Scient. Publ., Singapore,1997,pp.79-101.
3. D. Wolpert, “Stacked Generalization”, *Neural Networks*, Vol. 5, pp. 241–259, 1992.
4. T.K.Ho, “Random Decision Forests”, *Proc.ICDAR ‘95*,Montreal,Canada,1995, pp.278-282.
5. L. Breiman, “Bagging Predictors”, *Machine Learning*, Vol. 24, no. 2, pp. 123–140, 1996.
6. R.E.Schapire,“The strength of weak Learnability”, *Mach. Learning*,Vol.5,pp.197–227, 1990.
7. G. Dimauro, S. Impedovo, G. Pirlo, A. Salzo, “Multiple Classifiers: a new methodology for the evaluation of the combination processes”, *Progress in Handwriting Recognition*, A.C.Downton and S. Impedovo (eds.), World Scientific, Singapore, pp. 329–335, 1995.
8. J. Kittler, M. Hatef, R.P.W. Duin and J. Matias, “On combining classifiers”, *IEEE T-PAMI*, vol. 20, no. 3, pp. 226–239, 1998.
9. V.Di Lecce, G.Dimauro, A.Guerriero, S.Impedovo, G.Pirlo, A.Salzo, “Classifier Combination: The role of a-priori knowledge", IWFHR VII, eds. L.R.B. Schomaker et al., 2000, Nijmegen: International Unipen Foundation Publishing, pp. 143–152.
10. E. Mandler and J. Schuermann, “Combining the Classification Results of independent classifiers based on the Dempster/Shafer theory of evidence”, in *Pattern Recognition and Artificial Intelligence*, eds. E.S.Gelsema et al., North Holland,Amsterdam,1988,pp.381–393.
11. S. Impedovo and A. Salzo, "A new methodology for expert combination in multi-expert system designing". In Lecture Notes in Computer Science, (vol. 1857), J. Kittler and F.Roli Eds., MCS2000, Cagliari, Italy, pp.230–239.
12. G.Dimauro, S.Impedovo, G.Pirlo, A.Salzo, “Automatic Bankchecks Processing: A New Engineered System”, *Int. Journal of Pattern Recognition and Artificial Intelligence*, Vol.11, N.4, World Scientific Publ.,Singapore, 1997, pp.1–38.