The *Similarity Index* lower and upper bounds: Theoretical Considerations and Experimental Verification

G. Pirlo, D. Impedovo and D. Barbuzzi

***Abstract*—**In this paper the *Similarity Index* variability range is investigated. Depending on the recognition rates of *abstract-level* classifiers, the lower and upper bounds of the of the *Similarity Index* variability range is theoretically analysed. The experimental tests, carried out in the field of handwritten numeral classification, confirm the theoretical findings.

***Keywords*—**Classifier Combination, Classifier System, Similarity Index

I. Introduction

The collective behaviour of classifiers is a topic which has recently attracted the interest of a continuously growing research community. In fact, it is well-known that many difficult classification problems can be solved effectively by combining weakly similar classifiers, whereas no useful result can be obtained from the combination of very similar classifiers. As matter of this fact, much research has been devoted to design weakly similar classifiers based different classification methods, random selection of feature sets and resampling techniques of the training data [2, 5, 7, 11, 14, 15].

Several measures of similarity (or dissimilarity) have been also considered so far, to investigate on the collective behaviour of classifiers [9]. They have been applied to the selection of the most valuable subset of classifiers to be combined [6] and to the prediction of the performance of combination methods, depending on the characteristics of the combined classifiers [1]. Some measures work on a pairwise basis and then average the results [1, 5] , others work on the whole set of classifiers [4, 8].

Although several similarity (or dissimilarity) measures have been proposed, little formal work has been done on theoretical analysis of similarity among classifiers and several important aspects must be investigated yet. Among the others, it is very important to determine to what extent the interval of possible values of similarity (or dissimilarity) depends on the recognition rates of the individual classifiers. In fact, any similarity (or dissimilarity) value must be interpreted in respect to the similarity (or dissimilarity) variability range as well as the comparison of different similarity (or dissimilarity) values is only possible on the basis of the theoretical limits of the corresponding ranges of variability [12].

This paper presents a theoretical analysis on similarity among *abstract-level* classifiers. For this purpose, the *Similarity Index* is used to estimate the similarity among *abstract-level* classifiers and the lower and upper bounds of the *Similarity Index* variability range is determined, depending on the recognition rates of the individual classifiers. The experimental results, which have been carried out in the field of hand-written numeral recognition, confirm the theoretical findings.

The paper is organised as follows. Section 2 describes the *Similarity Index*, as an estimator of similarity among classifiers. The theoretical analysis of the lower and upper bounds of the *Similarity Index* is reported in Section 3. Section 4 shows the experimental results. The conclusion of the paper is reported in Section 5.

II. The Similarity Index

The *Similarity Index* is an estimator of similarity between *abstract-level* classifiers, which measures the average agreements between their decisions [1].

Let $A=\{\varepsilon_i \mid i=1,2,...,K\}$ be a set of *abstract-level* classifiers and $P=\{p_t \mid t=1,2,...,N\}$ a set of patterns each one belonging to one of the *m* possible classes $\Omega=\{\omega_1,\omega_2,\ldots,\omega_m\}$. Moreover let $\varepsilon_i(p_t)=\omega_j$ ($\omega_j\in\Omega$) be the decision of $\varepsilon_i\in A$ for a pattern $p_t\in P$ (it is assumed that classifiers cannot reject).

The *Similarity Index* for A is defined as:

$$\rho_A = \frac{\sum_{\substack{i,j=1,\dots K \\ i<j}} \rho_{\{\varepsilon_i,\varepsilon_j\}}}{\binom{K}{2}} \tag{1}$$

where:

$$\rho_{\{\varepsilon_i,\varepsilon_j\}} = \frac{1}{N}\sum_{t=1}^{N} Q(\varepsilon_i(p_t),\varepsilon_j(p_t)) \tag{2}$$

G. Pirlo is with the Department of Computer Science, University of Bari, Italy (corresponding author to provide phone: +390805443295; fax: +390805443196; e-mail: giuseppe.pirlo@uniba.it).

D. Impedovo is with the Department of Computer Science, University of Bari, Italy (e-mail: impedovo@gmail.com).

D. Barbuzzi is with the Department of Computer Science, University of Bari, Italy (e-mail: donato.barbuzzi@uniba.it).

and

$$Q(\varepsilon_i(p_t),\varepsilon_j(p_t)) = \begin{cases}1 & if \quad \varepsilon_i(p_t)=\varepsilon_j(p_t) \\ 0 & if \quad \varepsilon_i(p_t)\neq\varepsilon_j(p_t)\end{cases} \quad (3)$$

Figure 1a shows the decisions of four classifiers ε_1, ε_2, ε_3, ε_4, for the patterns $p_1, p_2, \ldots p_{10}$. Recognitions are indicated by the symbol "R" in white cells, misclassifications by shaded cells. Different shading denotes misclassifications by different class labels. Figure 1b reports the Similarity Index values for each pair of classifiers of Figure 1a. In this case it results that $\rho_{\{\varepsilon1, \varepsilon2, \varepsilon3, \varepsilon4\}}$ = (0.7+0.4+0.5+0.6+0.7+0.7) /6=0.6.

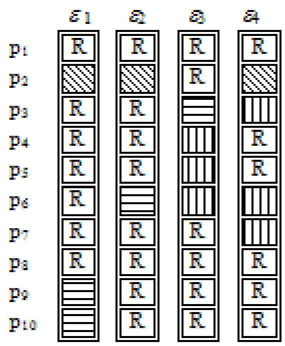


Fig. 1a Outputs of abstract-level classifiers.

ρ	ε1	ε2	ε3	ε4
ε1	1	0.7	0.4	0.5
ε2	0.7	1	0.6	0.7
ε3	0.4	0.6	1	0.7
ε4	0.5	0.7	0.7	1

Fig. 1b Similarity Index values.

III. On the Variability of the Similarity Index

In this section, the theoretical analysis on the variability interval of the *Similarity Index* is presented. In particular, the lower and upper bounds of the interval, in which the Similarity Index can range, are theoretically determined on the basis of the recognition rates of the classifiers.

Preliminarily, let A={ε_1, ε_2} be a set of two classifiers and P={p_t | t=1,2,...,N} the set of N input patterns. Moreover, let B_1 and B_2 be two subsets of P which contain the patterns recognised by ε_1 and ε_2, respectively (hence the recognition rate of ε_1 and ε_2 is R_1=card(B_1)/card(P), R_2=card(B_2)/card(P)). Depending on the agreement between the decisions of ε_1 and ε_2 in classifying the patterns $p_t \in P$, the following five conditions can occur [1]:

- p_t is misclassified by ε_1 and is recognised by ε_2;
- p_t is recognised by both ε_1 and ε_2;
- p_t is recognised by ε_1 and is misclassified by ε_2;
- p_t is misclassified by both ε_1 and ε_2, and furthermore $\varepsilon_1(p_t) \neq \varepsilon_2(p_t)$
- p_t is misclassified by both ε_1 and ε_2, and furthermore $\varepsilon_1(p_t) = \varepsilon_2(p_t)$.

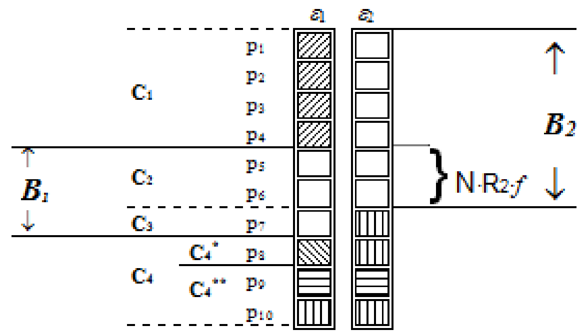


Fig. 2 Analysis of agreements between two classifiers.

Hence, the set P can be partitioned into the following five subsets, as fig. 2 shows:

- C_1={ $p_t \in P$ | $p_t \notin B_1$ and $p_t \in B_2$ } (i.e. $\forall p_t \in C_1$: p_t is misclassified by ε_1 and recognised by ε_2),
- C_2= {$p_t \in P$ | $p_t \in B_1$ and $p_t \in B_2$ } (i.e. $\forall p_t \in C_2$: p_t is recognised both by ε_1 and ε_2),
- C_3= {$P_t \in P$ | $p_t \in B_1$ and $p_t \notin B_2$ } (i.e. $\forall p_t \in C_3$: p_t is recognised by ε_1 and misclassified by ε_2),
- C_4={ $p_t \in P$ | $p_t \notin B_1$ and $p_t \notin B_2$} (i.e. $\forall p_t \in C_4$: p_t is misclassified both by ε_1 and ε_2). Of course, C_4 can be divided into two subsets C_4^* and C_4^{**} ($C_4 = C_4^* \cup C_4^{**}$), with:
  - C_4^* = { $p_t \in P$ | $\varepsilon_1(p_t) \neq \varepsilon_2(p_t)$ } (i.e. $\forall p_t \in C_4^*$: ε_1 and ε_2 misclassify p_t differently);
  - C_4^{**} = { $p_t \in P$ | $\varepsilon_1(p_t) = \varepsilon_2(p_t)$ } (i.e. $\forall p_t \in C_4^{**}$: ε_1 and ε_2 misclassify p_t with the same class label).

Now, if *f*=card(C_2)/card(B_2), it results that:

- Card(C_2)= Card(B_2) · *f* = N · R_2 · *f*;

and

- Card(C_1)=Card(B_2)-Card(C_2)= N·R_2 - N·R_2 · *f* = N·R_2·(1-*f*);
- Card(C_3)=Card(B_1)-Card(C_2)= N·R_1 - N·R_2 · *f* = N·(R_1-R_2·*f*).

1. Finally, from the consideration that
N=Card(P)=Card(C_1)+Card(C_2)+Card(C_3) +Card(C_4)
it follows that

- Card(C_4)=N-Card(C_1)-Card(C_2)-Card(C_3)=N-[N·R_2·(1-*f*)]–[N·R_2 · *f*]–[N·(R_1-R_2·*f*)] =
=N·[1- R_2·(1-*f*) – R_2 · *f* –(R_1-R_2·*f*)]=N· [1- R_2 – R_1 + R_2·*f*] =
=N· [1-(1-*f*)·R_2–R_1]=N· [(1-R_1)- R_2 · (1- *f*)].

Of course, the *Similarity Index* reported in eq. 1, for the set of classifiers A, is equal to:

$$\rho_A = \frac{\lfloor Card(C_2) + Card(C_4^{**}) \rfloor}{Card(P)}, \qquad (4)$$

and it results that:

❖ $$\rho_A = \frac{[Card(C_2)]}{Card(P)} = R_2 \cdot f \quad if \quad C_4^{**} = \varnothing, \qquad (5)$$

(this is the case in which $\forall p_t \in P$ so that ε_1 and ε_2 misclassify p_t, then $\varepsilon_1(p_t) \neq \varepsilon_2(p_t)$).

❖ $$\rho_A = \frac{[Card(C_2) + Card(C_4)]}{Card(P)} = R_2 \cdot f + (1 - R_1) - R_2 \cdot (1 - f) = 1 - (R_1 + R_2) + 2R_2 \cdot f$$

, if $C_4^{**} = C_4$ (6)

(this is the case in which $\forall p_t \in P$ so that ε_1 and ε_2 misclassify p_t, then $\varepsilon_1(p_t) = \varepsilon_2(p_t)$).

More in general, let $A=\{\varepsilon_i \mid i=1,2,...,K\}$ be a set of *abstract-level classifiers*, and R_i the recognition rate of ε_i, i=1,2,..,K (hereafter it is supposed that $R_i<1$, i=1,2,…,K, since, if there exists one individual classifier ε_i for which $R_i=1$, other classifiers are no longer necessary [15]), it results that

- *Similarity Index: Lower Bound*

$$\rho_A^{\min} = \frac{k' R' + \binom{k'}{2}}{\binom{K}{2}}, \text{ where } k' = \left\lfloor \sum_{i=1}^{K} R_i \right\rfloor \text{ and } R' = \sum_{i=1}^{K} R_i - \left\lfloor \sum_{i=1}^{K} R_i \right\rfloor = \sum_{i=1}^{K} R_i - k'.$$

This result is demonstrated in section 3.1;

- *Similarity Index: Upper Bound*

$$\rho_A^{Max} = 1 - \frac{\left[2\sum_{i=1}^{K} i \cdot R_i - (K+1)\sum_{i=1}^{K} R_i \right]}{\binom{K}{2}}$$

.

This result is demonstrated in section 3.2.

A. *The Similarity Index: Lower Bound*

Theorem 1 (*Similarity Index* Lower Bound)

Let $A=\{\varepsilon_i \mid i=1,2,...,K\}$ be a set of classifiers, R_i the recognition rate of ε_i, i=1,2,..,K, and let $P=\{p_t \mid t=1,2,...,N\}$ a set of N patterns. Furthermore, let k' and R' be respectively the integer part and the decimal part of the sum of the recognition rates of all classifiers included in A:

$$k' = \left\lfloor \sum_{i=1}^{K} R_i \right\rfloor, R' = \sum_{i=1}^{K} R_i - \left\lfloor \sum_{i=1}^{K} R_i \right\rfloor = \sum_{i=1}^{K} R_i - k'. \qquad (7)$$

It can be shown that the *Similarity Index* ρ_A is minimum iff a partition[1] $\{S'_0, S'_1\}$ of P exists for which it results that:

- $Card(S'_0)=N \cdot R'$ and $\forall p_t \in S'_0$: p_t is recognised by k'+1 classifiers out of K;
- $Card(S'_1)=N \cdot (1-R')$ and $\forall p_t \in S'_1$: p_t is recognised by k' classifiers out of K;

and $\forall p_t \in P$: if ε_i and ε_j misclassify p_t, then $\varepsilon_i(p_t) \neq \varepsilon_j(p_t)$, $\forall$i,j=1,2,…,K, i≠j.

Proof Theorem 1

Theorem 1 is proved by induction on K.

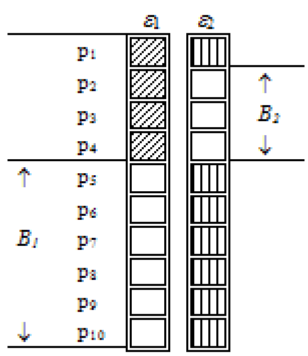
<u>*Base of induction*</u>

Let $A=\{\varepsilon_1, \varepsilon_2\}$ be a set of two classifiers, B_1 and B_2 the subsets of P containing the patterns recognised by ε_1 and ε_2, respectively (see Fig.2). If f=card($B_1 \cap B_2$)/card(B_2), eq.(4) provides the *Similarity Index* of A and the minimum occur for $C_4^{**}=\varnothing$ and f as small as possible (see eq. (5)). The following cases must be distinguished:

A) if $R_1+R_2<1$, then f minimum is equal to 0 and occurs for $B_2 \subset P-B_1$ (Fig. 3a). In this case the hypothesis of the theorem is satisfied for the partition $\{S''_o, S''_1\}$ of P and the parameters k'' and R'' defined as: $S''_0=B_1 \cup B_2$, $S''_1=P-S''_0$, and $k''=\lfloor R_1+R_2 \rfloor=0$, $R''=R_1+R_2$.

B) if $R_1+R_2=1$, then f minimum is equal to 0 and occurs for $B_2=P-B_1$ (Fig. 3b). In this case the hypothesis of the theorem is satisfied for the partition $\{S''_o, S''_1\}$ of P and the parameters k'' and R'' defined as: $S''_0=\varnothing$, $S''_1=P$, and $k''=\lfloor R_1+R_2 \rfloor=1$, $R''=R_1+R_2-1=0$.

C) if $R_1+R_2>1$, then f minimum is equal to $(R_1+R_2-1)/R_2$ and occurs for $P-B_1 \subset B_2$ (Fig. 3c). In this case the hypothesis of the theorem is satisfied for the partition $\{S''_o, S''_1\}$ of P and the parameters k'' and R'' defined as: $S''_0=B_1 \cap B_2$, $S''_1=P-S''_0$, and $k''=\lfloor R_1+R_2 \rfloor=1$, $R''=R_1+R_2-1$.



(a) $R_1+R_2<1$

[1] $\{S_1,S_2,…,S_N\}$ (N>1) is a partition of P iff: (a) ∀i,j=1,2,…,N: i≠j ⇒ $S_i \cap S_j=\varnothing$; (b) $\bigcup_{i=1}^{N} S_i = P$.

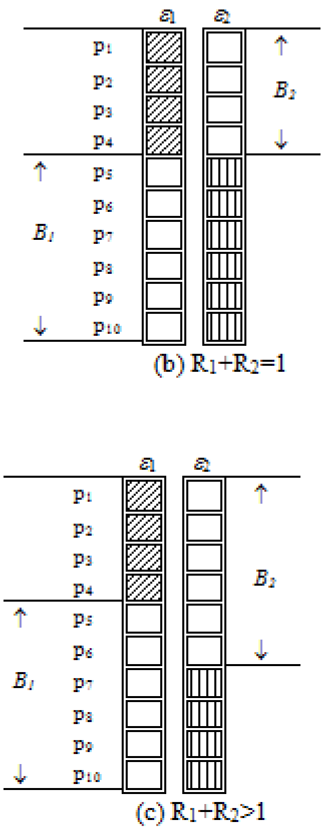


Fig. 3 Lower Bound of ρ_A: Base of Induction

Induction hypothesis

Let Theorem 1 be true for K=k; we have to verify it for K=k+1. For this purpose, let $A=\{\varepsilon_i \mid i=1,2,...,k\}$ be a set of k classifiers satisfying the hypothesis of Theorem 1 (with the partition $\{S'_0,S'_1\}$ of P and the parameters k' and R'). Let ε_{k+1} be an extra classifier (recognition rate R_{k+1}) joined to A, B_{k+1} the subset of P containing the patterns recognised by ε_{k+1}. If f=card($S'_0 \cap B_{k+1}$)/card (B_{k+1}), from similar considerations of those used for Fig. 2, we have that the contribution to the *Similarity Index* due to ε_{k+1} depends on the quantity

$$k'(1-f)R_{k+1}+(k'+1)fR_{k+1}=(k'+f)\,R_{k+1}, \qquad (8)$$

where:

- $k'(1-f)R_{k+1}$ derives from the patterns in S'_1 which are recognised by ε_{k+1}
- $(k'+1)fR_{k+1}$ derives from the patterns in S'_0 which are recognised by ε_{k+1}.

Note that no contribution to the *Similarity Index* is given by the patterns misclassified by ε_{k+1}. In fact, as in eq. (5), it must result that $\forall p_t \in P$ so that ε_i and ε_{k+1} misclassify p_t, then $\varepsilon_i(p_t)\neq\varepsilon_{k+1}(p_t)$.

Now, the minimum of eq. (8) occurs for f as small as possible. The following cases must be distinguished:

A) if $R'+R_{k+1}<1$, then f minimum is equal to 0 and occurs for $B_{k+1}\subset S'_1$ (Fig. 4a). In this case the hypothesis of the theorem is satisfied for the partition $\{S''_o,S''_1\}$ of P and the parameters k" and R" defined as: $S''_0=S'_0\cup B_{k+1}$, $S''_1=P-S''_0$, and $k''=k'$, $R''=R'+R_{k+1}$.

B) if $R'+R_{k+1}=1$, then f minimum is equal to 0 and occurs for $B_{k+1}=S'_1$ (Fig. 4b). In this case the hypothesis of the theorem is satisfied for the partition $\{S''_o,S''_1\}$ of P and the parameters k" and R" defined as: $S''_0=\emptyset$, $S''_1=P$, and $k''=k'+1$, $R''=0$.

C) if $R'+R_{k+1}>1$, then f minimum is equal to $(R'+R_{k+1}-1)/R_{k+1}$ and occurs for $S'_1\subset B_{k+1}$ (Fig. 4c). In this case the hypothesis of the theorem is satisfied for the partition $\{S''_o,S''_1\}$ of P and the parameters k" and R" defined as: $S''_0=S'_0\cap B_{k+1}$, $S''_1=P-S''_0$, and $k''=k'+1$, $R''=R'+R_{k+1}-1$.

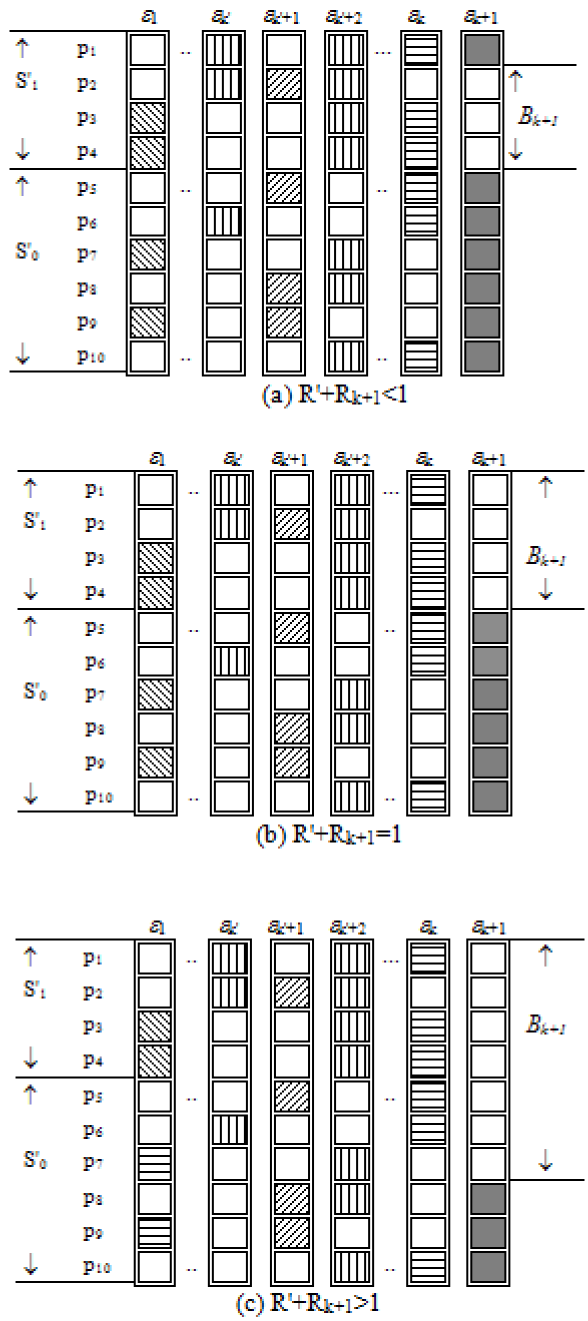


Fig.4. Lower Bound of ρ_A: Induction Hypothesis

Q.E.D.

Lemma 1
Let $A=\{\varepsilon_i \mid i=1,2,...,K\}$ be a set of *abstract-level* classifiers, R_i the recognition rate of ε_i, i=1,2,..,K, and let $P=\{p_t \mid t=1,2,...,N\}$ be a set of N patterns. The lower bound of the *Similarity Index* ρ_A^{min} for the set A is given by:

$$\rho_A^{min} = \frac{k'R' + \binom{k'}{2}}{\binom{K}{2}} \qquad (9)$$

where k' and R' are the same as those in eq. (7).

Proof Lemma 1
Substituting eq. (2) in eq. (1) and considering a set A satisfying the conditions of Theorem 1, it follows that the Similarity Index ρ_A^{min} for A is equal to:

$$\rho_A^{min} = \frac{\sum_{\substack{i,j=1,...K\\ i<j}} \frac{1}{N}\sum_{t=1}^{N} Q(\varepsilon_i(p_t),\varepsilon_j(p_t))}{\binom{K}{2}} =$$

$$= \frac{\frac{1}{N}\sum_{t=1}^{N}\sum_{\substack{i,j=1,...K\\ i<j}} Q(\varepsilon_i(p_t),\varepsilon_j(p_t))}{\binom{K}{2}} =$$

$$= \frac{\frac{1}{N}\sum_{P_t\in S'_0\cup S'_1}\left[\sum_{\substack{i,j=1,...K\\ i<j}} Q(\varepsilon_i(p_t),\varepsilon_j(p_t))\right]}{\binom{K}{2}}. \qquad (10)$$

Moreover, Theorem 1 states that if we let p_t be an input pattern $p_t\in S'_0$, p_t is recognised by k'+1 classifiers out of K while the remaining K-(k'+1) classifiers misclassify p_t with different class labels. Hence, for a pattern $p_t\in S'_0$ it results that:

$$\sum_{\substack{i,j=1,...K\\ i<j}} Q(\varepsilon_i(p_t),\varepsilon_j(p_t)) = \binom{k'+1}{2}, \qquad (11)$$

where $\binom{k'+1}{2}$ is due to the k'+1 classifiers that recognise p_t;

Similarly, for a pattern $p_t\in S'_1$ it results that:

$$\sum_{\substack{i,j=1,...K\\ i<j}} Q(\varepsilon_i(p_t),\varepsilon_j(p_t)) = \binom{k'}{2}, \qquad (12)$$

where $\binom{k'}{2}$ is due to the k' classifiers that recognise p_t;

Substituting eqs. (11) and (12) in eq.(10) it results that:

$$\rho_A^{min} = \frac{\frac{1}{N}\sum_{P_t\in S'_0}\left[\binom{k'+1}{2}\right] + \frac{1}{N}\sum_{P_t\in S'_1}\left[\binom{k'}{2}\right]}{\binom{K}{2}} =$$

$$= \frac{\frac{1}{N}\cdot Card(S'_0)\cdot\left[\binom{k'+1}{2}\right] + \frac{1}{N}\cdot Card(S'_1)\cdot\left[\binom{k'}{2}\right]}{\binom{K}{2}} =$$

$$= \frac{\frac{1}{N}NR'\binom{k'+1}{2} + \frac{1}{N}N(1-R')\binom{k'}{2}}{\binom{K}{2}} =$$

$$= \frac{R'\frac{k'+1}{k'-1}\binom{k'}{2} + (1-R')\binom{k'}{2}}{\binom{K}{2}} =$$

$$= \frac{R'\frac{2}{k'-1}\binom{k'}{2} + \binom{k'}{2}}{\binom{K}{2}} = \frac{R'\frac{2}{k'-1}\frac{k'(k'-1)}{2} + \binom{k'}{2}}{\binom{K}{2}} =$$

$$\frac{k'R' + \binom{k'}{2}}{\binom{K}{2}}$$

Q.E.D.

B. The Similarity Index: Upper Bound

Theorem 2 (*Similarity Index* Upper Bound)
Let $A=\{\varepsilon_i \mid i=1,2,...,K\}$ be a set of classifiers, R_i the recognition rate of ε_i, i=1,2,..,K. Without loss in generality, let $R_i\le R_{i+1}$, i=1,2,…,K-1. The *Similarity Index* for A is maximum iff a partition $\{S'_0, S'_1, S'_2, \dots, S'_{p-1}, S'_p, \dots, S'_K\}$ of P exists for which it results that:

- $card(S'_0)=N\cdot R_1$ and $\forall\, p_t \in S'_0$: p_t is recognised by K classifiers out of K;
- $card(S'_1)=N\cdot (R_2-R_1)$ and $\forall\, p_t \in S'_1$: p_t is recognised by K-1 classifiers out of K;
- $card(S'_2)=N\cdot (R_3-R_2)$ and $\forall\, p_t \in S'_2$: p_t is recognised by K-2 classifiers out of K;
- …
- $card(S'_{p-1})=N\cdot(R_p-R_{p-1})$ and $\forall p_t\in S'_{p-1}$:p_t is recognised by K-(p-1) classifiers out of K;
- $card(S'_p)=N\cdot (R_{p+1}-R_p)$ and $\forall\, p_t \in S'_p$: p_t is recognised by K-p classifiers out of K;
- …
- $card(S'_{K-1})=N\cdot (R_K-R_{K-1})$ and $\forall\, p_t \in S'_{K-1}$: p_t is recognised by 1 classifier out of K;
- $card(S'_K)=N\cdot (1-R_K)$ and $\forall\, p_t \in S'_K$: p_t is recognised by 0 classifiers out of K;

and $\forall p_t \in P$: if ε_i and ε_j misclassify p_t , then $\varepsilon_i(p_t) =\varepsilon_j(p_t)$, $\forall i,j=1,2,\dots,K$, $i\ne j$.

Proof Theorem 2

Theorem 2 is proved by induction on K.

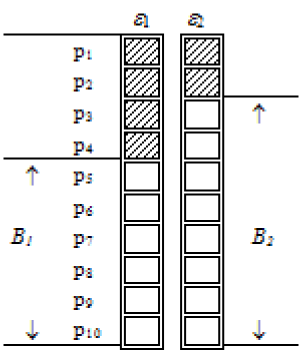
<u>*Base of induction*</u>

Let A={ε_1, ε_2 } be a set of two classifiers, B_1 and B_2 be the subsets of P containing the patterns recognised by ε_1 and ε_2, respectively (see Fig.2). If f=card($B_1 \cap B_2$)/card(B_2), eq.(4) provides the *Similarity Index* of A and the maximum of ρ_A occurs for $C_4^*=\emptyset$ and f as large as possible (see eq. (6)). The following cases must be distinguished:

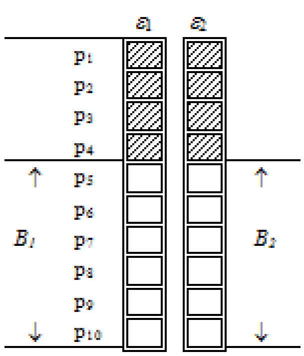
A) if $R_1<R_2$, then f maximum is equal to card(B_1)/card(B_2) and occurs for $B_1 \subset B_2$ (Fig. 5a). In this case the hypothesis of the theorem is satisfied for the partition {S''_0,S''_1, S''_2} of P defined as: $S''_0=B_1$,$S''_1=B_2-B_1$, $S''_2=P-B_2$.

B) if $R_1=R_2$, then f maximum is equal to card(B_2)/card(B_2)=1 and occurs for $B_1=B_2$ (Fig. 5b). In this case the hypothesis of the theorem is satisfied for the partition {S''_0,S''_1, S''_2} of P defined as: $S''_0=B_1$, $S''_1=\emptyset$, $S''_2=P-B_1$.

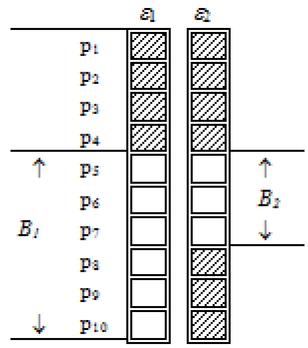
C) if $R_2<R_1$, then f maximum is equal to card(B_2)/card(B_2)=1 and occurs for $B_2 \subset B_1$(Fig. 5c). In this case the hypothesis of the theorem is satisfied for the partition {S''_0,S''_1, S''_2} of P defined as: $S''_0=B_2$, $S''_1=B_1-B_2$, $S''_2=P-B_1$.



(a) $R_1<R_2$



(b) $R_1=R_2$



(c) $R_2<R_1$

Fig5. Upper Bound of ρA: Base of Induction

<u>*Induction hypothesis*</u>

Let Theorem 2 be true for K=k; we have to verify it for K=k+1. Let A={ε_i | i=1,2,...,k} be a set of k classifiers (without loss of generality we assume that $R_i \leq R_{i+1}$, i=1,2,...,k-1) satisfying the hypothesis of Theorem 1 (with the partition {S'_0, S'_1, S'_2, … , S'_{p-1}, S'_p , …, S'_K} of P) and let ε_{k+1} be an extra classifier (recognition rate R_{k+1}) that is joined to A. Moreover, let B_i be the subset of P containing the patterns recognised by ε_i, i=1,2,…,k+1. If f_i=card($B_i \cap B_{k+1}$)/card(B_{k+1}), i=1,2,...,k, from similar considerations to those used in Fig.2 it results that the contribution to the *Similarity Index* due to ε_{k+1} depends on the quantity:

$$\sum_{i=1}^{k}[(R_{k+1} f_i)+(1-R_i)-R_{k+1}(1-f_i)]=\sum_{i=1}^{k}[1-(R_i+R_{k+1})+2R_{k+1} f_i], \qquad (13)$$

where:

- R_{k+1} f_i derives from the patterns recognised both by ε_i and ε_{k+1}
- (1-R_i)-R_{k+1}(1-f_i) derives from the patterns misclassified both by ε_i and ε_{k+1} (as for eq.(6), it must result that $\forall p_t \in P$ so that ε_i and ε_{k+1} misclassify p_t, then $\varepsilon_i(p_t)=\varepsilon_{k+1}(p_t)$).

The maximum of the quantity in eq. (13) occurs for f_i as large as possible, i=1,2,…,k. The following cases must be distinguished:

A) if $R_{k+1} \leq R_1$, then f_i maximum occurs for $B_{k+1} \subseteq B_i$, i=1,…,k (Fig. 6a). In this case the hypothesis of the theorem is satisfied for the partition {S''_0, S''_1, S''_2, … , S''_{p-1}, S''_p , …, S''_K, S''_{K+1}} of P defined as: $S''_0=B_{k+1}$, $S''_1=B_1-B_{k+1}$, $S''_2=B_2-B_1$, … , $S''_{p-1}=B_{p-1}-B_{p-2}$, $S''_p=B_p-B_{p-1}$, …, $S''_{k-1}=B_{k-1}-B_{k-2}$, $S''_k=B_k-B_{k-1}$, $S''_{k+1}=P-B_k$.

B) if an index p exists so that $R_{p-1} \leq R_{k+1} \leq R_p$, then f_i maximum occurs for (Fig. 6b):

❖ $B_i \subseteq B_{k+1}$, for i=1,2,…,p-1

❖ $B_{k+1} \subseteq B_i$, for i=p,…,k.

In this case the hypothesis of the theorem is satisfied for the partition {S''_0, S''_1, S''_2, … , S''_{p-1}, S''_p , …, S''_K, S''_{K+1}} of P defined as: $S''_0=B_1$, $S''_1=B_2-B_1$, $S''_2=B_3-B_2$, … , $S''_{p-1}=B_{k+1}-B_{p-2}$, $S''_p=B_p-B_{k+1}$, …, $S''_{k-1}= B_{k-1}-B_{k-2}$, $S''_k=B_k-B_{k-1}$, $S''_{k+1}=P-B_k$.

C) if $R_k \leq R_{k+1}$, then f_i maximum occurs for $B_{k+1} \subseteq B_i$, i=1,…,k (Fig. 6c). In this case the hypothesis of the theorem is satisfied for the partition {S"$_0$, S"$_1$, S"$_2$, … , S"$_{p-1}$, S"$_p$,…, S"$_K$, S"$_{K+1}$} of P defined as: S"$_0$=B$_1$, S"$_1$=B$_2$-B$_1$, S"$_2$=B$_3$-B$_2$,… ,S"$_{p-1}$=B$_p$-B$_{p-1}$, S"$_p$=B$_{p+1}$-B$_p$,…,S"$_{k-1}$=B$_k$-B$_{k-1}$, S"$_k$=B$_{k+1}$-B$_k$, S"$_{k+1}$=P-B$_{k+1}$.

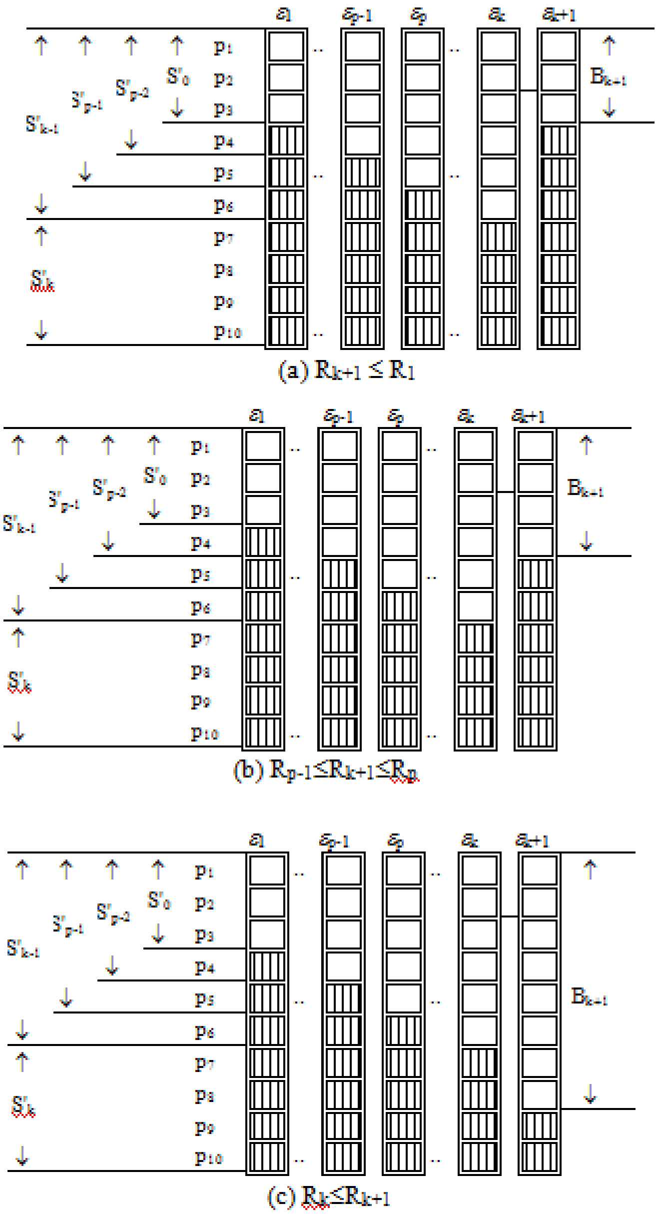


Fig. 6. Upper Bound of ρ_A: Induction Hypothesis

Q.E.D.

Lemma 2

Let A={ε_i | i=1,2,...,K} be a set of *abstract-level* classifiers, R_i the recognition rate of ε_i, $R_i \le R_{i+1}$, i=1,2,…,K-1, and let P={p_t | t=1,2,...,N} be a set of N patterns. The upper bound of the *Similarity Index* ρ_A^{Max} for the set A is given by:

$$\rho_A^{Max} = 1 - \frac{\left[2\sum_{i=1}^{K} i\cdot R_i - (K+1)\sum_{i=1}^{K} R_i\right]}{\binom{K}{2}} \qquad (14)$$

Proof Lemma 2

Substituting eq. (2) in eq. (1) and considering a set A satisfying the conditions of Theorem 1, it follows that the Similarity Index ρ_A^{Max} for A is equal to:

$$\rho_A^{Max} = \frac{\sum_{\substack{i,j=1,\dots K\\ i<j}} \frac{1}{N}\sum_{t=1}^{N} Q(\varepsilon_i(p_t),\varepsilon_j(p_t))}{\binom{K}{2}} =$$

$$\frac{\frac{1}{N}\sum_{t=1}^{N} \sum_{\substack{i,j=1,\dots K\\ i<j}} Q(\varepsilon_i(p_t),\varepsilon_j(p_t))}{\binom{K}{2}} =$$

$$= \frac{\frac{1}{N}\sum_{P_t\in S_0\cup S_1\cup\ldots\cup S_K}\left[\sum_{\substack{i,j=1,\dots K\\ i<j}} Q(\varepsilon_i(p_t),\varepsilon_j(p_t))\right]}{\binom{K}{2}} .(15)$$

Moreover Theorem 2 states that if we let p_t be an input pattern $p_t \in S'_{K-p}$, p_t is recognised by p classifiers out of K, while the remaining K-p classifiers misclassify p_t with the same class label. Hence, for the pattern p_t it results:

$$\sum_{\substack{i,j=1,\dots K\\ i<j}} Q(\varepsilon_i(p_t),\varepsilon_j(p_t)) = \binom{p}{2} + \binom{K-p}{2}, (16)$$

where

- ❖ $\binom{p}{2}$ is due to the p classifiers that recognise p_t;
- ❖ $\binom{K-p}{2}$ is due to the K-p classifiers that misclassify p_t.

Substituting eq.(16) in eq.(15) it results that:

$$\frac{1}{N}\left[\frac{\sum_{P_t\in S'_0}\binom{K}{2}+\sum_{P_t\in S'_1}\binom{K-1}{2}+..+\sum_{P_t\in S'_{i-1}}\left[\binom{K-(i-1)}{2}+\binom{i-1}{2}\right]+}{\binom{K}{2}}\ldots\right.$$

$$\left.\ldots\frac{\sum_{P_t\in S'_i}\left[\binom{K-i}{2}+\binom{i}{2}\right]+..+\sum_{P_t\in S'_K}\binom{K}{2}}{\binom{K}{2}}\right]=$$

$$\frac{1}{N}\left[\frac{Card(S'_0)\cdot\binom{K}{2}+Card(S'_1)\cdot\binom{K-1}{2}+..+Card(S'_{i-1})\cdot\left[\binom{K-(i-1)}{2}+\binom{i-1}{2}\right]+}{\binom{K}{2}}\right..$$

$$\left...\frac{Card(S'_i)\cdot\left[\binom{K-i)}{2}+\binom{i}{2}\right]+..+Card(S'_K)\cdot\binom{K}{2}}{\binom{K}{2}}\right]=$$

$$\frac{1}{N}\left[\frac{N\cdot R_1\cdot\binom{K}{2}+N\cdot(R_2-R_1)\cdot\binom{K-1}{2}+..+N\cdot(R_i-R_{i-1})\cdot\left[\binom{K-(i-1)}{2}+\binom{i-1}{2}\right]+}{\binom{K}{2}}\right...$$

$$\left...\frac{N\cdot(R_{i+1}-R_i)\cdot\left[\binom{K-i)}{2}+\binom{i}{2}\right]+..+N\cdot(1-R_K)\cdot\binom{K}{2}}{\binom{K}{2}}\right]=$$

$$\frac{1}{N}\left[\frac{N\cdot\binom{K}{2}+NR_1\cdot\left[\binom{K}{2}-\binom{K-1}{2}\right]+N\cdot R_2\cdot\left[\binom{K-1}{2}-\binom{K-2}{2}-\binom{2}{2}\right]+}{\binom{K}{2}}\right...$$

$$...\frac{+\sum_{i=3}^{K-2}N\cdot R_i\cdot\left[\binom{K-(i-1)}{2}+\binom{i-1}{2}-\binom{K-1}{2}-\binom{i}{2}\right]+}{\binom{K}{2}}...$$

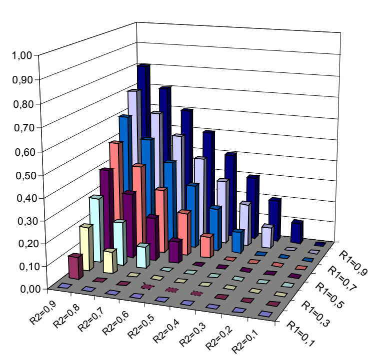
$$\left...\frac{N\cdot R_{K+1}\cdot\left[\binom{2}{2}+\binom{K-2}{2}-\binom{K-1}{2}\right]+N\cdot R_K\left[\binom{K-1}{2}-\binom{K}{2}\right]}{\binom{K}{2}}\right]$$

$$=\frac{\left[\binom{K}{2}+\sum_{i=1}^{K}R_i\cdot(K-2i+1)\right]}{\binom{K}{2}}=$$

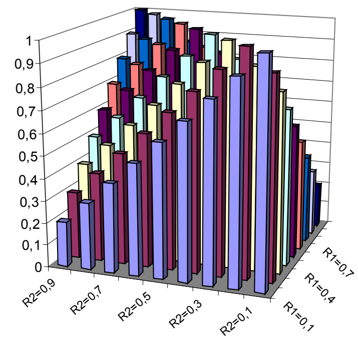
$$1-\frac{\left[2\sum_{i=1}^{K}i\cdot R_i-(K+1)\sum_{i=1}^{K}R_i\right]}{\binom{K}{2}}.$$

Q.E.D.

Figure 7 shows the lower and the upper bounds (obtained by eq. (9) and (14), respectively) of the *Similarity Index* variability range for two classifiers, depending on the recognition rates.



(a) Lower Bound



(b) Upper Bound

Fig.7. Upper and Lower boundary for the Similarity Index

IV. Experimental Results

The experimental results have been carried out in the field of hand-written numeral classifiers. Table 1 reports the set A={ε_1, ε_2, ε_3, ε_4, ε_5 , ε_6 } of the distance-based classifiers used for the tests, whose complete description can be found in ref. [3]. The classifiers were trained and tested using the patterns from the CEDAR database (training patterns: 18468 hand-written numerals; test patterns: 2711 hand-written numerals). Table 1 also reports the recognition rates of the individual classifiers at zero rejections.

Table 2 reports the *Similarity Index* for each subset of classifiers K classifiers picked up from A, K=2,3,4,5,6. It results that, for K=2, the most complementary sets of classifiers are A={ε_1, ε_3} and A={ε_1, ε_4} (ρ_A =0.76); the least complementary set is A={ε_5, ε_6} (ρ_A =0.86). For K=3, the most complementary set is A={ε_1, ε_2, ε_4} (ρ_A =0.78); the least complementary set is A={ε_3, ε_5, ε_6} (ρ_A =0.86). For K=4, the most complementary set is A={ε_1, ε_2, ε_3, ε_4} (ρ_A =0.80); the least complementary set is A={ε_3, ε_4, ε_5, ε_6} (ρ_A=0.85). For K=5, the most complementary set is A={ε_1, ε_2, ε_3, ε_4, ε_6} (ρ_A=0.81); the least complementary set is A={ε_2, ε_3, ε_4, ε_5, ε_6}(ρ_A=0.85).

When the *Similarity Index* values are compared to the variability range, determined by eqs.(9) and (14), the result is reported in Figure 8a. The subsets are ordered along the x axis for increasing values of the *Similarity Index*. Figure 8b shows the *Similarity Index* values for the case in which the range of variability is normalized to [0,1]. This results, which allows the comparison among *Similarity Index* values belonging to different variability ranges, makes evident that even though classifiers use features of various types, the *Similarity Index* ranges for sets of real classifiers ranges in a very reduced interval and no set among those available has a degree of similarity very close to the minimum.

Classifier		Recognition Rate (R_i)
ε_1	template matching	81.7%
ε_2	slope of the contour profile	86.3%
ε_3	projection histograms in the four main directions	89.7%
ε_4	characteristic loci technique	89.8%
ε_5	distribution of foreground pixels in different zones of the pattern image	90.4%,
ε_6	distribution in the pattern image of 3x3 templates of foreground pixels	90.6%

Table 1: Experimental Results: Numeral Classifiers

K=2

A	ρ_A
A_4,A_6	0,85
A_1,A_6	0,86
A_2,A_6	0,87
A_3,A_6	0,87
A_1,A_2	0,88
A_2,A_3	0,88
A_2,A_4	0,88
A_5,A_6	0,88
A_3,A_4	0,89
A_1,A_3	0,90
A_4,A_5	0,90
A_1,A_5	0,91
A_1,A_4	0,92
A_2,A_5	0,92
A_3,A_5	0,95

K=3

A	ρ_A
A_1,A_2,A_6	0,87
A_2,A_3,A_6	0,87
A_3,A_4,A_6	0,87
A_2,A_4,A_6	0,87
A_2,A_3,A_4	0,88
A_4,A_5,A_6	0,88
A_1,A_3,A_6	0,88
A_1,A_4,A_6	0,88
A_1,A_5,A_6	0,88
A_1,A_2,A_3	0,89
A_1,A_2,A_4	0,89
A_2,A_5,A_6	0,89
A_1,A_3,A_4	0,90
A_2,A_4,A_5	0,90
A_3,A_5,A_6	0,90
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

K=4

A	ρ_A
A_2,A_3,A_4,A_6	0,87
A_1,A_2,A_4,A_6	0,88
A_1,A_3,A_4,A_6	0,88
A_2,A_4,A_5,A_6	0,88
A_1,A_2,A_3,A_4	0,89
A_1,A_2,A_3,A_6	0,89
A_1,A_2,A_5,A_6	0,89
A_1,A_3,A_5,A_6	0,89
A_1,A_4,A_5,A_6	0,89
A_2,A_3,A_5,A_6	0,89
A_3,A_4,A_5,A_6	0,89
A_1,A_2,A_4,A_5	0,90
A_2,A_3,A_4,A_5	0,90
A_1,A_2,A_3,A_5	0,91
A_1,A_3,A_4,A_5	0,91

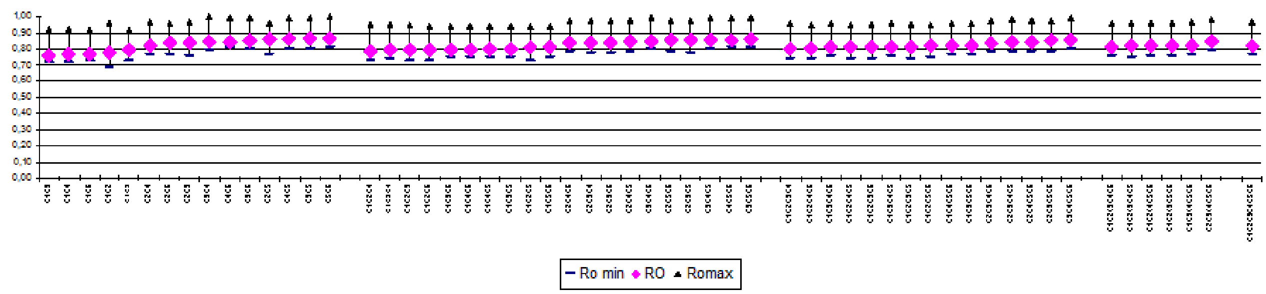
K=5

A	ρ_A
A_1,A_2,A_3,A_4,A_6	0,88
A_1,A_2,A_3,A_5,A_6	0,89
A_1,A_2,A_4,A_5,A_6	0,89
A_1,A_3,A_4,A_5,A_6	0,89
A_2,A_3,A_4,A_5,A_6	0,89
A_1,A_2,A_3,A_4,A_5	0,90

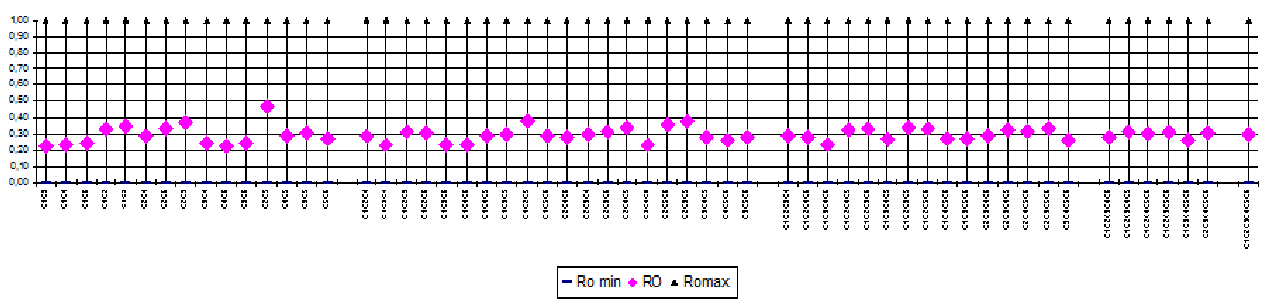
K=6

A	ρ_A
A_1,A_2,A_3,A_4,A_5,A_6	0,89

Table 2. Similarity Index value for sets of classifiers



(a) absolute values



(b) normalized values

Figure 8: Similarity Index value vs variability range

V. Conclusions

In this paper the lower and upper bounds of the Similarity Index are theoretically determined, depending on the recognition rates of the individual classifiers. The experimental tests, carried out in the field of handwritten numerals recognition, confirm the theoretical findings.

The results, which offer new insights to the analysis of similarity among *abstract-level* classifiers, can allow a deeper comprehension of other open questions in the area of classifier combination and multi-expert system design.

References

[1] L. Bovino, G. Dimauro, S. Impedovo, M.G. Lucchese, R. Modugno, G. Pirlo, A. Salzo, L. Sarcinella, "On the combination of Abstract-Level classifiers", IJDAR, to appear.

[2] L. Breiman, "Bagging Predictors", *Machine Learning*, Vol. 24, no. 2, pp. 123-140, 1996.

[3] G. Dimauro, S. Impedovo, G. Pirlo, A. Salzo, "Bankcheck recognition systems: re-engineering the design process". In *Progress in Handwriting Recognition*, A.C. Downton and S. Impedovo (eds.), *World Scientific Publ.*, Singapore, 1997, pp. 419-425.

[4] J.L. Fleiss, "Statistical Methods for Rates and Proportions", Wiley, New York, 1981.

[5] T. K. Ho, "The Random Subspace Method for Constructing Decision Forests", *IEEE Trans. on Pattern Analysis Machine Intelligence*, Vol. 20, No. 8, pp. 832-844, 1998.

[6] J. Kim, K. Seo, K. Chung, "A Systematic Approach to Classifier Selection on Combining Multiple Classifiers for Hand-written Digit Recognition", *Proc. ICDAR'97*, Ulm,Germany,1997, pp.459-462.

[7] J. Kittler, M. Hatef, R.P.W. Duin, J. Matias, "On combining classifiers", *IEEE Trans. on Pattern Analysis Machine Intelligence*, Vol.20, no.3, pp.226-239, 1998.

[8] R. Kohavi, D.H. Wolpert, "Bias plus variance decomposition for zero-one loss functions", Proc. 13th Int. Conference on Machine Learning, Morgan Kaufmann, 1996, pp. 275-283.

[9] L.I. Kuncheva, M. Skurichina, R.P.W. Duin, "An Experimental study on diversity for bagging and boosting with linear classifiers", Information Fusion, 2002, Vol. 3, pp. 245-258.

[10] D. Partridge, W.J. Krzanowski, "Software diversity: practical statistics for its measurement and exploitation", Information and Software Technology, 1997, 39, pp. 707-717.

[11] R.E. Schapire, "The strength of weak Learnability", *Machine Learning*, Vol. 5, pp. 197-227, 1990.

[12] C. A. Shipp, L.I. Kuncheva, "Relationships between combination methods and measures of diversity in combining classifiers", Information Fusion, 2002, Vol. 3, pp. 135-148.

[13] P.H.A. Sneath, R.R. Sokal, "Numerical Taxonomy", W.H. Freeman & Co, 1973.

[14] D. Wolpert, "Stacked Generalization", *Neural Networks*, Vol. 5, pp. 241-259, 1992.

[15] L. Xu, A. Krzyzak, C.Y. Suen, "Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition", *IEEE Trans. on Syst. Man and Cybern.*, Vol. 22, N.3, 1992, pp. 418-435.

G. Pirlo received the Computer Science degree cum laude in 1986 at the Department of Computer Science of the University of Bari, Italy. Since then he has been carrying out research in the field of pattern recognition and image analysis. He received a fellowship from IBM in 1988. Since 1991 he has been Assistant Professor at the Department of Computer Science of the University of Bari, where he is currently Associate Professor. His interests cover the areas of pattern recognition and biometry, image analysis, intelligent systems, computer arithmetic, communication and multimedia technologies. He has developed several scientific projects and published over one-hundred fifty papers in the field of handwriting recognition, automatic signature verification, document analysis and processing, parallel architectures for computing, multimedia technologies for collaborative work and distance learning. Prof. Pirlo is reviewer for many international journals including IEEE T-PAMI, IEEE T-SMC, Pattern Recognition, IJDAR, Information Processing Letters, etc. . He has been in the scientific committee of many International Conferences and has served as reviewer of ICPR, ICDAR, ICFHR, IWFHR, ICIAP, VECIMS, CISMA, etc. . He was the general co-chair of the International Conference on Frontiers in Handwriting Recognition (ICFHR 2012). He is IEEE member and member of the IAPR - Technical Committee on "Reading Systems" (TC-11). He serves as member of the SIe-L Head Committee and is member of the e-learning Committee of the University of Bari.

D. Impedovo received the MEng degree cum laude in Computer Engineering in 2005 and the PhD degree in Computer Engineering in 2009 both from the Polytechnic of Bari (Italy). In 2011 he received the M.Sc. (II Level italian Master degree) on Remote Science Technologies from the University of Bari. He is, currently, with the Department of Computer Science (University of Bari). His research interests are in the field of pattern recognition and biometrics. He is co-author of more than 20 articles on these fields in both international journals and conference proceedings. He received 'The Distinction' for the best young student presentation in May 2009 at the International Conference on Computer Recognition Systems (CORES – endorsed by IAPR). He serves as reviewer for the Elsevier Pattern Recognition journal, IET Journal on Signal Processing and IET Journal on Image Processing and for many International Conferences including ICPR and ICASSP. He is IAPR and IEEE member.

D. Barbuzzi received the Computer Science degree "cum laude" in 2011 from University of Bari "Aldo Moro". He worked from September to December 2011 as collaborator in the Center "Rete Puglia". Since 2012 he is a Ph.D. student for the Department of Computer Science at the University of Bari. His current research interest is in the field of multi-expert systems for patter recognition.