CLASSIFIER COMBINATION

Lecce Di, V.; Dimauro, G.; Guerriero, A.; Impedovo, S.; Pirlo, G.; Salzo, A.

CLASSIFIER COMBINATION: THE ROLE OF A-PRIORI KNOWLEDGE

V.DI LECCE[1], G.DIMAURO[2], A.GUERRIERO[1], S.IMPEDOVO[2], G.PIRLO[2], A.SALZO[2]

(1) *Dipartimento di Ing. Elettronica -Politecnico di Bari-*
via Re David -70126 Bari- Italy
(2) *Dipartimento di Informatica - Università di Bari -*
Via Orabona, 4 - 70126 Bari – Italy

The aim of this paper is to investigate the role of the a-priori knowledge in the process of classifier combination. For this purpose three combination methods are compared which use different levels of a-priori knowledge. The performance of the methods are measured under different working conditions by simulating sets of classifier with different characteristics. For this purpose, a random variable is used to simulate each classifier and an estimator of stochastic correlation is used to measure the agreement among classifiers.

The experimental results, which clarify the conditions under which each combination method provides better performance, show to what extend the a-priori knowledge on the characteristics of the set of classifiers can improve the effectiveness of the process of classifier combination.

1   Introduction

Classifier combination is a diffuse strategy that has been widely used in complex classification problems for which very high performance is required [1]. For classifier combination, many methods have been proposed so far which are generally classified into three categories depending on the amount of information they combine [2,3]. *Abstract-level* combination methods use the top candidate provided by each classifier [4,5,6] ; *Ranked-level* combination methods use the entire ranked list of candidates [7,8] ; *Measurement-level* combination methods use also the confidence value of each candidate in the ranked list [9,10]. Among the others, classifier combination at *abstract-level* is the most general approach since every classifier is able at least to provide results at abstract level.

In the process of classifier combination, some kind of a-priori knowledge can also be used in order to achieve better performance. On the basis of the kind of a-priori knowledge the combination methods use, they can be classified into three categories. Methods of the first category do not require any kind of a-priori information on the combined classifiers [4,7,8]. Methods of the second category use information at the level of individual classifiers as a-priori knowledge [5]. Methods of the third category require information at the level of the entire set of combined classifiers [6,9,10].

In this paper, the role of a-priori knowledge in the process of classifier combination is investigated by comparing three combination methods: the Majority Vote Method (MV) which is of the first category [4]; the Dempster-Shafer Method (DS) which is of the second

category [5]; the Behavioural Knowledge Space Method (BKS) which is of the third category [6]. For this purpose, a recent methodology is considered for the evaluation of methods for classifier combination. The behaviour of a method is evaluated by using sets of classifiers with different characteristics. Each classifier is considered as a random variable and a suitable estimator of complementarity is used to measure the agreement in the set of combined classifiers. This paper is organised as follows. Section 2 presents the methodology used in evaluating the methods for classifier combination based on simulated data sets. Section 3 discusses the process of data set generation. The combination methods used in the experimental test are briefly illustrated in Section 4. Section 5 reports the experimental results. They clarify the conditions under which each combination method is the best, and show to what extend the use of a-priori knowledge on the characteristics of the set of classifiers is useful to improve the performance of classifier combination.

2   A Methodology to Evaluate *Abstract-Level* Combination Methods

In this paper a recent methodology is used for the evaluation of *abstract-level* combination methods. The performance of a combination method is measured by using several sets of classifiers which differ in terms of recognition rate and level of correlation. For this purpose, each classifier is considered as a random variable and it is simulated by its outputs (which are simple class labels) and a "Similarity Index" is used to estimate the stochastic correlation among the classifiers of each set [11]. Precisely, let be A_1 and A_2 two classifiers and T a database of N patterns. Let be $A_1(x)$ and $A_2(x)$ the top candidate provided by A_1 and A_2 for the pattern x, $x \in$ T, the agreement between A_1 and A_2 for x is evaluated by the function:

$$Q(A_i(t), A_j(t)) = \begin{cases} 1 & if\ A_i(t) = A_j(t) \\ 0 & otherwise \end{cases} \tag{1}$$

and the "Similarity Index" between A_1 and A_2 is defined as:

$$\rho_{A_i, A_j} = \frac{1}{N} \sum_{t=1}^{N} Q(A_i(t), A_j(t)) \tag{2}$$

For instance, figure 1 shows two sets of simulated classifiers. Each set consist of two classifiers. It is assumed that N=20 input patterns belonging to the class "0" are fed to the classifiers and therefore that the correct recognition result is "0". Although the recognition rate of the classifiers is 60% in both cases, in Fig. 1a it results $\rho=1$, while in Fig. 1b it results $\rho=0.2$. It should be noted that, in order to achieve $\rho=1$ the two classifiers must always produce the same response both in the case of correct recognition and in the case of misrecognition.

144

A₁	A₂		A₁	A₂
0	0		0	1
0	0		0	4
0	0		0	7
0	0		0	7
0	0		0	8
0	0		0	5
0	0		0	4
0	0		0	1
0	0		0	0
0	0		0	0
0	0		0	0
1	1		1	0
4	4		4	0
7	7		7	0
7	7		7	0
8	8		8	0
5	5		5	0
4	4		4	0
1	1		1	0
(a)			(b)	

Figure 1: Variability range of "Similarity Index"

Now, if a set A of K classifiers is considered, $A = \{A_i \mid i=1,2,...K\}$, the Similarity Index ρ_A for A is defined as :

$$\rho_A = \frac{\displaystyle\sum_{\substack{i,j=1,...K \\ i<j}} \rho_{A_i,A_j}}{\binom{K}{2}} \quad (3)$$

Moreover, for a set of K classifiers each one with a recognition rate equal to R, the "Similarity Index" ρ_A ranges in $[\rho_{min},1]$, where ρ_{min} is equal to [11]:

$$\rho_{min} = \frac{k'R' + \binom{k'}{2}}{\binom{K}{2}} \quad (4)$$

145

where:

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

The procedure for the evaluation of a combination method follows three steps:

Selection of the characteristics of the sets of classifiers. This step defines the number of classifier of the set (K) and the recognition rate of each individual classifier (R) (for the sake of simplicity in this paper we suppose that all classifiers have the same recognition rate);

Classifiers Simulation. For each value ρ^* of the Similarity Index (starting from $\rho=\rho_{min}$ to $\rho=1$, and using a suitable step $\delta\rho$) generate m lists of outputs for K classifiers with Similarity Index equal to ρ^* (an effective procedure used for data set generation is discussed in Section 3);

Performance Evaluation. For each value R of the recognition rate and ρ^* of the Similarity Index, evaluate the performance of the combination methods for the m lists available.

3   Data Set Generation

In order to evaluate the performance of a classifier combination method in different working conditions, different sets of individual classifiers must be simulated. Since *abstract-level* classifiers output simple class labels, they can be simulated by generating suitable lists of outputs. The aim of the procedure described in this section is to generate automatically outputs of sets of classifiers with different characteristics to be used for the testing of the combination method. The input data of the procedure are:
- the number K of classifiers of the set;
- the recognition rate R_i of each classifier A_i, i=1,2,....,K;
- the number N of outputs that must be generated by each classifier.

In the first phase, the input data are used to generate by a random number generation routine an initial list of outputs, which simulates only one set of K classifiers. Figure 2 shows a list of outputs (N=10) simulating a set of 4 classifiers having the same recognition rate R_i=60%, i=1,2,3,4. Correct outputs are indicated by *R*, while substitutions are indicated by *S1*, *S2*, *S3* and *S4*, where $\forall$ i≠j , we have *Si≠Sj*. It is easy to verify that the Similarity Index of the set is $\rho=2.3/6$.

Starting from this initial set of classifiers, new sets are generated by modifying the list of outputs. The basic idea is to generate new sets of classifiers having different correlation values without changing the recognition rate of the individual classifiers. For instance, if we set $A_3(9)$=S2 the correlation for the new list of outputs is $\rho=2.4/6$. If we also set $A_4(3)$=R and $A_4(9)$=S2 it results $\rho=2.5/6$. This modification procedure continues until a pool of different sets of classifiers is obtained with fixed individual characteristics

(recognition and reliability rate of the individual classifiers), which have, however, a correlation spanning the entire range of possible values, from $\rho=\rho_{min}$ to $\rho=1$.

	A_1	A_2	A_3	A_4
Pattern 1	R	R	R	R
Pattern 2	R	R	R	S1
Pattern 3	R	R	S3	S2
Pattern 4	S1	S4	R	S3
Pattern 5	S4	S3	R	R
Pattern 6	R	R	R	R
Pattern 7	S3	R	S2	R
Pattern 8	S2	R	S1	S1
Pattern 9	R	S2	S4	R
Pattern 10	R	S4	R	R

Figure. 2: Lists of outputs of 4 classifiers: R→Recognition; S1,S2,S3,S4→Substitution.

4 Combination Methods

In order to evaluate the effect of the a-priori knowledge on the effectiveness of methods for classifiers combination, three *abstract-level* combination methods have been considered in this work.

The **Majority Vote Method (MV)** does not require any kind of a-priori knowledge on the combined classifiers. MV assigns to each class ω_i, i=1,2,...,m, a score $S(\omega_i)$ equal to the number of classifiers for which the class ω_i is the top candidate [4]. The final response of the combined classifier is the class label ω_i for which the score is the maximum (i.e. $S(\omega_j)=\max\{S(\omega_i), i=1,2,...,m\}$).

The **Dempster-Shafer Method (DS)** uses as a-priori knowledge the performance of each individual classifier. DS combines different classifiers using their recognition and substitution rates as a-priori knowledge [5]. For an input pattern x, all classifiers having the same output are collected into a group E_k, k=1,...,K' (K' is the number of different outputs), which is equivalent to a new classifier with a new recognition and substitution rates. Successively, from the analysis of the set of equivalent classifiers E_k, k=1,...,K', two belief measures are computed (see [5] for details): the belief of correct output $Bel(A_j)$ and the belief of misrecognized output $Bel(\neg A_j)$. The final response of the combined classifier is the class label ω_j for which the difference is maximum between the belief measures for correct output and misrecognition (i.e. $Bel(A_j)- Bel(\neg A_j) =\max\{Bel(A_i)- Bel(\neg A_i) \mid i=0,1,...m\}$).

The **Behaviour Knowledge Space (BKS)** uses as a-priori knowledge the behaviour of the whole set of classifiers extracted in a suitable "learning" procedure (top–candidate vectors corresponding to the classification results of the whole set of classifiers). BKS is

147

based on two processing phases : the "learning" phase and the "operation" phase [6]. The "learning" allows to fill a suitable K-dimensional space. Each dimension of this space corresponds to the decision of a specific classifier and the K-tuple of decisions provided by the K classifiers defines a "Focal Unit". When a "Focal Unit" is addressed by the vector of recognition responses, the index I(j) corresponding to the class ω_j of the input pattern is incremented. This index counts the number of times in which a pattern belonging the class ω_j generates the specific K-tuple of decisions. In the "operation" phase, when a "Focal Unit" is addressed by the K-tuple of decisions, the final result of the combined classifier is the class label ω_j for which the corresponding index is maximum (i.e. $I(\omega_j) = \max \{I(\omega_i), i=1,2,...m\}$).

5   Experimental Results

The effectiveness of MV, DS and BKS has been evaluated by simulating 1000 different sets of classifiers for the training and 1000 for the test. Figure 3 shows the structure of a typical output generated by the simulation procedure. In this case it has been assumed that K=4, R=0.75 and N=20 input patterns belonging to the class of the numeral '0' were inputted to the classifiers (therefore "0" is the correct output). The values of the "Similarity Index" are $\rho_{12}=15/20$, $\rho_{23}=13/20$, $\rho_{34}=12/20$, $\rho_{13}=10/20$, $\rho_{14}=12/20$, $\rho_{24}=12/20$, and $\rho_{1234}=0.617$.

| A_1	A_2	A_3	A_4
0	0	0	3
0	0	0	0
0	0	0	0
0	0	0	0
0	0	5	3
8	8	0	0
0	0	0	0
0	0	0	0
0	0	3	0
3	0	0	0
3	8	0	3
0	8	1	3
0	0	0	0
0	0	0	0
0	5	5	0
7	0	0	0
0	0	0	1
4	4	0	0
0	0	6	0
0	0	0	0

Figure 3: Output of the simulation procedure

The performance of MV, DS and BKS are compared for the case of K=3,4 and 5 classifiers. The recognition rate of each classifier is equal to 90% and no rejection is allowed at the level of individual classifier. According to eq. (4), for all K=3,4 and 5, the "Similarity Index" changes within the range [0.8,1]. The results are shown in Figure 4 (obtained for $\delta\rho$=0.1) in which the performance of the methods are evaluated in terms of recognition rate and reliability rate (defined as: Reliability Rate=Recognition Rate/(1-Rejection Rate) [3]).

- When K=3, DS is the best method in terms of recognition rate when the combined classifiers are weakly correlated (ρ<0.93); for more correlated classifiers the best performance are achieved by BKS. Concerning reliability, MV is the most reliable method for very low correlated classifiers (ρ<0.86). As the correlation increases the most reliable method becomes DS first ($0.86 \leq \rho \leq 0.92$) and BKS successively (ρ>0.92).
- When K=4, the best method in terms of recognition rate is DS, for weakly correlated classifiers (ρ<0.85), and BKS for strongly correlated classifiers (ρ>0.85). In terms of reliability the best method is MV, for weakly correlated classifiers (ρ<0.92), and BKS when the correlation among classifiers increases (ρ>0.92).
- When K=5, concerning recognition the best method is MV if the classifiers are very weakly correlated (ρ<0.85). As the correlation increases the best method becomes DS first ($0.85 \leq \rho \leq 0.9$), and BKS successively (ρ>0.9). In terms of reliability rate the best method is MV, for weakly correlated classifiers (ρ<0.86), and BKS for more correlated classifiers (ρ>0.86).

These results confirm the well-known concept that combination methods achieve best results when complementary classifiers are combined (i.e. ρ as close to ρ_{min} as possible). Moreover they demonstrate that when weakly correlated classifiers are combined, the a-priori knowledge is not necessary to achieve high-performance from the classifier combination process. In fact, MV works generally very well when weakly correlated classifiers are combined. Conversely, as the correlation increases, the a-priori knowledge becomes the key aspect for classifier combination. As matter of this fact, we observe that an increasing level of a-priori knowledge is necessary as correlation becomes close to the maximum (i.e. ρ=1). In fact, as the classifiers become more correlated the DS becomes very effective, while BKS achieve the best performance when classifiers are very strongly correlated.
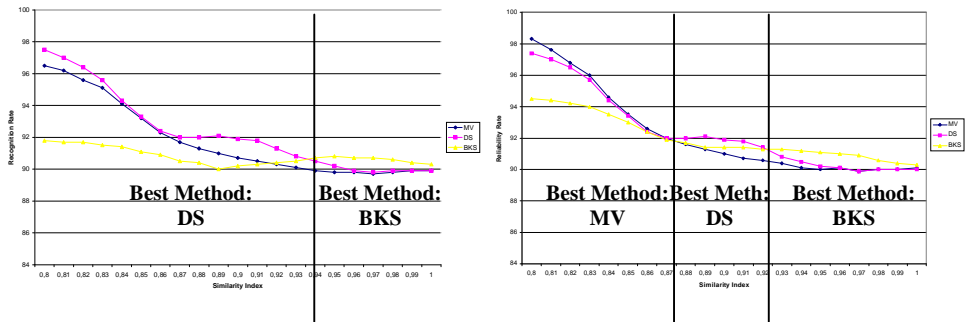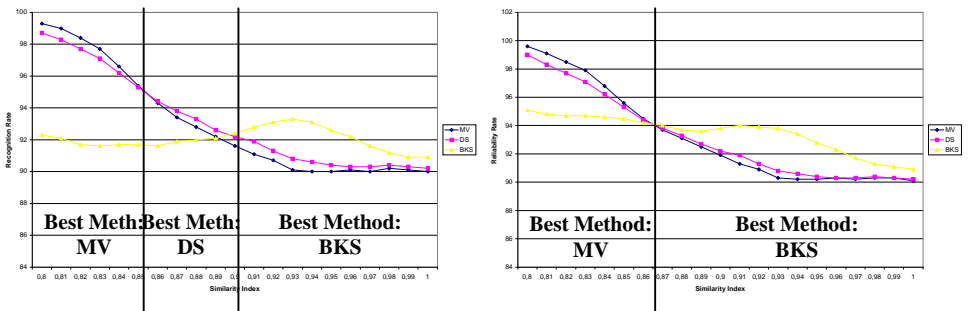
Number of Classifiers: K=3
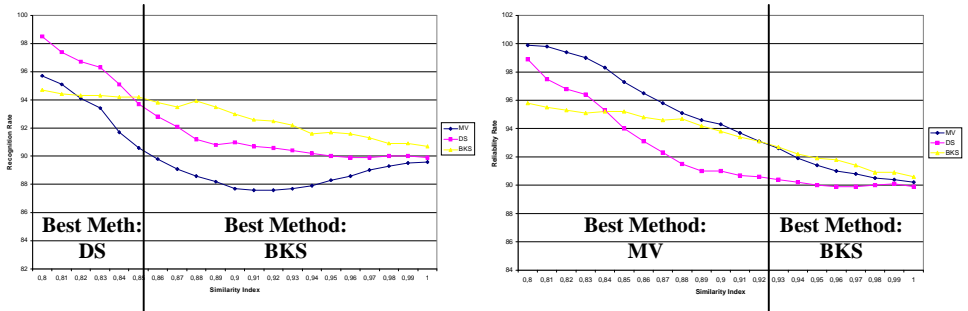
Number of Classifiers: K=4

Number of Classifiers: K=5

Figure 4: Experimental results

150

Another experimental test has been carried out with real data by using five different classification algorithms for handwritten numeral recognition [13] : A- *Region-Based Classifier, B- Contour-based Classifier, C- Enhanced-Loci Classifier, D- Histogram-based Classifier, E- Crossing-based Classifier.* Data sets from the CEDAR database (BR and BS directories) have been used for training and test. After training, the performance of each individual classifier is about 90%. Table 1 reports the results for differently correlated sets of K=3,4, and 5 classifiers. For each set of combined classifiers, the best results in terms of recognition rate and reliability are on grey background. These results confirm the considerations obtained in the previous experimental test (carried out by using simulated data) about the relevance of a-priori knowledge in classifiers combination when strongly correlated classifiers are considered.

Table 1: Performance of the combination methods

Set	ρ	Classification Algorithm	MV		DS		BKS	
			Recogn.	Reliab.	Recogn.	Reliab.	Recogn.	Reliab.
K=3	0.82	A-B-C	94,1	94,7	94,3	94,4	91,3	93,5
	0.87	A-B-D	91,7	92,2	92.1	92.1	90,3	92,0
K=4	0.85	A-B-C-E	90,6	97,3	93,7	94.2	94,2	95,2
	0.86	A-B-C-D	89,7	96,5	92,5	93,1	93,7	94,8
K=5	0.84	A-B-C-D-E	96,4	96,7	96,1	96,4	91,8	94,6

6   Conclusion

This paper presents an investigation on the role of a-priori knowledge in the process of classifier combination. For this purpose, a recent methodology for the analysis of *abstract-level* methods for classifier combination is applied to evaluate the effectiveness of three combination methods: Majority Vote, Dempster-Shafer and Behaviour Knowledge Space. The results point out the relevance of using the a-priori knowledge in the process of classifier combination specially when strongly correlated classifiers are combined.

References

1. H. Bunke, S. Impedovo and P.S.P. Wang (ed.) "Bankcheck Processing Systems", World Scientific, Singapore, 1997.

2. L. Lam, Y.S. Huang, C.Y. Suen, "Combination of Multiple Classifier Decisions for Optical Character Recognition", in *Handbook of Character Recognition and Document Image Analysis,* Ed. H. Bunke and P.S.P. Wang, World Scientific, Singapore, 1997, pp. 79-101.

3. 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.

4. C.Y. Suen, C. Nadal, T.A. Mai, R. Legault, L. Lam, "Recognition of totally unconstrained handwritten numerals based on the concept of multiple experts", *Proc. Of Frontiers in Handwriting Recognition*, CENPARMI, Montreal, Canada, 1990, pp. 131-143.

5. Lu, F. Yamaoka, "Integration of Handwritten Digit Recognition results using Evidential Reasoning", Proc. Of IWFHR-4, 1994, pp. 456-463.

6. Huang, C.Y. Suen, "An Optimal Method of Combining Multiple Classifiers for Unconstrained Handwritten Numeral Recognition", Proc. Of IWFHR-3, Buffalo, NY, 1993, pp. 11-20.

7. T.K. Ho, J.J. Hull, S.N. Srihari, "Decision Combination in Multiple Classifier Systems" IEEE PAMI, Vol. 16, No. 1, Jan. 1994, pp. 66-75.

8. J.Hull,T.K.Ho,J.Favata,V.Govindaraju,S.Srihari,"Combination of Segmentation -based and Wholistic Handwritten Word Recognition Algorithms", in *From Pixels to Features III-Frontiers in Handwriting Recognition*, S. Impedovo and J.C. Simon eds., Elsevier Publ., 1992, pp. 261-272.

9. N. Gorsky, "Pratical Combination of Multiple Classifiers", in *Progress in Handwriting Recognition*, Eds. A.C. Downton and S. Impedovo, World Scientific Publ., 1997, pp. 277-284.

10. Y.S. Huang, K. Liu, C.Y. Suen, "A Neural Network Approach for Multi-classifier recognition systems", Proc. Of IWFHR-4, 1994, pp. 235-244.

11. G. Dimauro, S. Impedovo, G. Pirlo, "Multiple Expert : A New Methodology for the Evaluation of the Combination Processes", Proc. Of IWFHR-5, Colchester, Uk, 1996, pp. 131-136.

12. S. Impedovo, A. Salzo, "Evaluation of Combination Methods", Proc. ICDAR'99, Bangalore, Sept. 1999, pp. 394-397.

13. G. Dimauro, S. Impedovo, G. Pirlo, A. Salzo, "Automatic Bankchecks Processing : A New Engineered System", *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 11, N.4, World Scientific Publ., Singapore, 1997, pp. 1-38.