Generation of Sets of Synthetic Classifiers for the Evaluation of Abstract-Level Combination Methods

N. Greco, S. Impedovo, R.Modugno, G. Pirlo

Abstract - This paper presents a new technique for generating sets of synthetic classifiers to evaluate abstract-level combination methods. The sets differ in terms of both recognition rates of the individual classifiers and degree of similarity. For this purpose, each abstract-level classifier is considered as a random variable producing one class label as the output for an input pattern. From the initial set of classifiers, new slightly different sets are generated by applying specific operators, which are defined at the purpose. Finally, the sets of synthetic classifiers have been used to estimate the performance of combination methods for abstract-level classifiers. The experimental results demonstrate the effectiveness of the proposed approach.

Keywords - Abstract-level Classifier, Dempster-Shafer Rule, Multi-expert Systems, Similarity Index, System Evaluation.

I. Introduction

A diffuse paradigm for solving difficult classification problems is based on multi-classifier systems. A multi-classifier system determines the final classification decision by combining the decisions of the individual classifiers. Three major categories of combination methods are generally defined, depending on the level of decisions combined [1]: measurement-level, ranked-level, abstract-level. The *Measurement-level* combination methods combine values provided by individual classifiers as a measure of the degree of membership of the input pattern to each class. The *Ranked-level* combination methods combine ranked lists of class labels ordered according to the degree of membership of the input pattern. The *Abstract-level* combination methods combine simple class labels. Of the three categories, the *abstract-level* combination methods is the most general since any kind of classifier can supply at least decisions at the *abstract-level*.

Although multi-classifier systems have been applied in many applications, many problems related to classifier combination methods still remain open. Among the others, one of the most relevant problems concerns the evaluation of combination methods. In fact, theoretical analysis of combination methods is generally extremely complex and concrete results have been derived only for very simple combination schemes.

Manuscript received November 18, 2004. This paper has been supported by the Italian Ministry "Ministero dell'Istruzione,dell'Università e della Ricerca", MIUR. *S. Impedovo and G. Pirlo are with the Università degli Studi di Bari -Dipartimento di Informatica,Via Orabona 4-70126 Bari– Italy. N. Greco and R. Modugno are with the Università degli Studi di Bari- Centro Rete Puglia, Via G.Petroni 15/F.1 – 70100 Bari –Italy.*

Kittler et al. [2] use a Bayesian framework to evaluate the sensitivity of combination schemes based on "sum" and "product" rules. Kleinberg [3] studied the combination of a large number of automatically generated classifiers for a two-class problem. Srihari [4] provides a theoretical analysis of the performance of combining by majority voting a set of independent classifiers for a two-class problem. Lam [5] discussed some aspects of majority voting derived from the Condorcet Jury Theorem. Tumer and Ghosh [6] quantified the improvement in classification results due to linear combination of neural networks.

When the combination methods are too complex, theoretical analysis can be impracticable and the evaluation of performance can be estimated on experimental basis. In this case the result depends on the specific conditions of the test and no useful information is achieved on the performance of the combination method under different operative conditions.

In order to face this problem, the use of simulated classifiers has recently emerged to artificially determine different operative conditions under which the performance of the combination method can be estimated. Zouari et al.[7] propose a two-step algorithm to generate synthetic classifiers based on a specified confusion matrix. Parker [8] uses a confusion matrix to characterize the classifiers. Based on the confusion matrix he generates classifier outputs to evaluate ranked-level combination methods. Kuncheva and Kountchev [9] randomly generate sets of classifiers, with specified accuracies and dependencies between them, to evaluate abstract-level combination methods.

Despite of approaches previously proposed in the literature, this paper starts from the consideration that whatever combination method is used, the combined classifier significantly outperforms the individual classifiers, if classifiers are diverse enough from each other [2]. Therefore, the degree of similarity of the set of individual classifiers is a fundamental parameter that must be carefully monitored in the process of evaluating combination methods. As matter of this fact, the process of synthetic classifier generation is here carried out by controlling the recognition rate of each individual classifier of the set and the degree of similarity of the overall set of classifiers. The organization of this paper is the following. Section 2 describes the process of performance evaluation for abstract-level combination methods. Section 3 presents the new Classifier Generation Procedure (CGP). The experimental results are reported in Section 4. They show the capability of the new technique in producing synthetic sets of classifiers useful for accurate evaluation of combination methods.

II. *Abstract-Level* Combination Method Evaluation

In this paper, the performance of an *abstract-level* combination method **C** is evaluated as function of the recognition rate of the individual classifiers and the degree of similarity among them [10]:

$$\mathbf{C}\,(K, \underline{R}, \rho) \rightarrow (R_{\mathbf{C}}, L_{\mathbf{C}}) \qquad (1)$$

where:

- K is the number of *abstract-level* classifiers that are combined;
- $\underline{R} = (R_1,R_2,\ldots,R_K)$ is the vector of recognition rates of each individual classifier;
- ρ is the *Similarity Index* of the set of classifiers. It is a measure of the degree of similarity among the individual classifiers of the set and it is computed as the average pairwise agreement between classifier decisions [11,12], and
- $R_{\mathbf{C}}$, $L_{\mathbf{C}}$ are respectively the recognition rate and the reliability rate of **C** ($L_C = \frac{R_C}{1\text{-Rejection Rate}}$) [1].

Moreover, it has been demonstrated recently that the degree of similarity, for a set of K abstract-level classifiers, can vary in a well-defined range. In particular, depending on the characteristics of the individual classifiers, we have that ρ ranges in [10,13]

$$[\rho^{min}, \rho^{Max}] \qquad (2)$$

where [13]:

- $\rho^{min} = \dfrac{k'R' + \binom{k'}{2}}{\binom{K}{2}}$,(k'= $\left\lfloor \sum_{i=1}^{K} R_i \right\rfloor$,R'= $\sum_{i=1}^{K} R_i - k'$) (3)
- $\rho^{Max} = 1 - \dfrac{\left[2\sum_{i=1}^{K} i \cdot R_i - (K+1)\sum_{i=1}^{K} R_i \right]}{\binom{K}{2}}$. (4)

The net result is that, for any vector of recognition rates $\underline{R} = (R_1,R_2,\ldots,R_K)$, the exhaustive analysis on the performance of combination method as a function of the degree of similarity among classifiers can be performed by generating sets of synthetic classifiers with similarity index ranging in the range (2).

III. Synthetic Classifiers Generation.

In this paper, each classifier A_i, i=1,2,...,K, is considered to be a discrete random variable producing a simple class label $A_i(r)$ as its decision corresponding to the r-th input pattern. More precisely, if a set of N input pattern for each class label ω_j is supposed to be input to the classifier A_i, with recognition rate R_i , A_i will generate a list of N class labels which simulate the classifier decisions. The list contains (in random order):

√ $N \cdot R_i$ recognitions (that are indicated by R);

√ $N \cdot (1-R_i)$ misclassifications (that are indicated by `S1`,`S2`,`S3`,…). Of course, misclassifications are obtained by uniformly picking in the set $\{\omega_1,\omega_2,\ldots,\omega_m\}-\{\omega_j\}$.

Figure 1 shows the list of outputs simulating a set of K=4 classifiers with $\underline{R}=(R_1,R_2,\ldots,R_K)=(0.6,\ 0.6,\ 0.6,\ 0.6)$ and ρ=2.3/6 (in this case we suppose that N=10).

	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 1. Lists of outputs of 4 classifiers

Of course, traditional approaches for producing sets of classifiers are based on random number generation procedures that are used iteratively (CGP_1). In this paper a different technique is proposed (CGP_2) that derives , from an initial list of outputs, several other sets of classifiers by well-suited operators named CHANGE+ , CHANGE- , SWAP+ , SWAP- . These operators have the aim to generate sets of synthetic classifier with the same individual characteristics (i.e. the same vector $\underline{R}=(R_1,R_2,\ldots,R_K)=(0.6,\ 0.6,\ 0.6,\ 0.6)$) and different degree of similarity. A detailed description of the operators is in the following.

A. The CHANGE Operators.

The CHANGE operators act on substitutions. Two classifiers A_i , A_j and one position `r` in the lists of outputs are randomly selected.

- When the CHANGE+ operator is used, if A_i(`r`) and A_j(`r`) are substitutions, and $A_i(r) \neq A_j(r)$, they are made equal to increase the value of the *Similarity Index* (for instance $A_i(r) \leftarrow A_j(r)$) without varying the recognition rate of the two classifiers. The effect of the CHANGE+ operator on the classifiers of Fig. 1 is shown in Fig. 2a. In this case the classifiers A_2 and A_3 and the position corresponding to P_9 have been selected. The *Similarity Index* augments from ρ=2.3/6 (Fig.1) to 2.4/6 (Fig.2a).
- When the CHANGE- operator is used, if A_i(`r`) and A_j(`r`) are substitutions, and $A_i(r)=A_j(r)$, one of them is changed with a different wrong value to reduce the value of the *Similarity Index* (for instance $A_i(r) \leftarrow S_m$ with $A_j(r) \neq S_m$).The effect of the CHANGE- operator on the classifiers of Fig.1 is shown in Fig.2b.In this case the classifiers A_3 and A_4 and the position corresponding to P_8 have been selected. The *Similarity Index* diminishes from ρ=2.3/6(Fig.1) to 2.2/6(Fig. 2b).

B. The SWAP Operators.

The SWAP operators act on recognition and substitutions. Two classifiers A_i , A_j and two positions r and s are randomly selected.

	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	S2	R
Pattern 10	R	S4	R	R

(a) CHANGE+

Initial Condition (see Fig. 1): $A_2(9)$=S2 , $A_3(9)$=S4
Final Condition: $A_2(9)$=S2 , $A_3(9)$=S2

	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	S4
Pattern 9	R	S2	S4	R
Pattern 10	R	S4	R	R

(b) CHANGE-

Initial Condition (see Fig. 1): $A_3(8)$=S1 , $A_4(8)$=S1
Final Condition: $A_3(8)$=S1 , $A_4(8)$=S4

Figure 2. CHANGE Operators

	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	S2	S3
Pattern 5	S4	S3	R	R
Pattern 6	R	R	R	R
Pattern 7	S3	R	R	R
Pattern 8	S2	R	S1	S1
Pattern 9	R	S2	S4	R
Pattern 10	R	S4	R	R

(a) SWAP +

Initial Condition (see Fig. 1): $A_2(4)$=S4 , $A_2(7)$=R, $A_3(4)$=R , $A_3(7)$=S2
Final Condition: $A_2(4)$=S4 , $A_2(7)$=R, $A_3(4)$=S2, $A_3(7)$=R

	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	S2	S3
Pattern 5	S4	S3	R	R
Pattern 6	R	R	S4	R
Pattern 7	S3	R	R	R
Pattern 8	S2	R	S1	S1
Pattern 9	R	S2	R	R
Pattern 10	R	S4	R	R

(b) SWAP -

Initial Condition (see Fig. 1): $A_2(6)$=R, $A_2(9)$=S2, $A_3(6)$=R,$A_3(9)$=S4
Final Condition: $A_2(6)$=R , $A_2(9)$=S2, $A_3(6)$=S4, $A_3(9)$=R

Figure 3. SWAP Operators

- When the SWAP+ operator is used, if $A_i(s)$, $A_j(r)$ are recognitions, and $A_i(r)$, $A_j(s)$ are substitutions, $A_j(s)$ and $A_j(r)$ are swapped to increase the *Similarity Index.* The effect of the SWAP+ operator on the classifiers in Fig.1 is shown in Fig. 3a. In this case the classifiers A_2 and A_3 and the positions corresponding to P_4 and P_7 have been selected. The *Similarity Index* augments from ρ=2.3/6 (Fig. 1) to 2.5/6 (Fig. 3a).
- When SWAP- is used, if $A_i(s)$ and $A_j(s)$ are correct and $A_i(r)$ and $A_j(r)$ are substituted, $A_j(s)$ and $A_j(r)$ are swapped to decrease the *Similarity Index.* The effect of the SWAP- operator on the classifiers in Fig. 1 is shown in Fig. 3b. In this case the classifiers A_2 and A_3 and the positions corresponding to P_6 and P_9 have been selected. The *Similarity Index* diminishes from ρ=2.3/6 (Fig. 1) to 2.2/6 (Fig.3b).

IV. Synthetic Classifiers Set Generation

On the basis of the operators defined in the previous section, sets of synthetic classifiers are generated and suitably organized, as described in the following.

A. Synthetic classifier Sets: Generation Procedure.

The aim of this stage is the generation of the artificial data sets simulating sets of K classifiers (A={A_1,A_2,…, A_K}), with various recognition rates (<u>R</u>=(R_1,R_2,…,R_K)) and degrees of similarity (ρ).

1. *For each class label do:*
2. *For different values of K (K=2,3,4,5,6) and <u>R</u> do:*
3. *Generate the initial list of (K·N) outputs simulating K classifiers with recognition rate equal to R and compute ρ;*
4. *Repeat until the number of list of output generated is greater than M (being M the number of sets of classifiers to be generated);*
5. *Repeat until $\rho=\rho_{MAX}$:*
   Modify the previous list of output by applying randomly CHANGE+ and SWAP+ (Increase ρ)
   Compute the new value of ρ
   Store the new list of output
   End Repeat;
6. *Repeat until $\rho=\rho_{MIN}$*
   Modify the previous list of output by applying randomly CHANGE- and SWAP- (Reduce ρ)
   Compute the new value of ρ
   Store the new list of output
   End Repeat;

End Repeat;
End do;
End do.

B. Database of Sets of Synthetic classifiers.

More specifically, the database containing the artificial data simulating sets of K classifiers is considered as a K dimension discrete space in which each axis reports the recognition rate of an individual classifier. In this space, each point corresponds to a vector <u>R</u>=(R_1,R_2,…,R_K) and collects the artificial data simulating sets of K classifiers A_1, A_2,…, A_K, with recognition rate R_1,R_2,…,R_K, respectively. At each point, the artificial data are organized into categories, on the basis of the degree of similarity ρ ranging from ρ^{min} to ρ^{Max}, according to eqs. (3) and (4).

V. Experimental results

For the experimental test, numerous sets of K synthetic classifiers have been generated, for K=2,3,4,5,6. We have compared the performance obtained by the traditional procedure based on random number generation (CGP_1) and the procedure for synthetic classifier generator described in the previous Section (CGP_2). The result is that unlike CGP_1, CGP_2 allows the generation of sets of classifiers with degrees of similarity spreading all over the variability range. Fig.4 compares the number of generated sets of synthetic classifiers, as function of the degree of similarity, for the case of K=4 classifiers, $\underline{R}=(R_1,R_2,R_3,R_4)=(0.9,0.9,0.9,0.9)$. It is worth noting that CGP_1 allows the development of sets of classifiers with degree of similarity varying over a small part of the range of variability (i.e. [0.85, 0,90]). Conversely, when CGP_2 is used, we obtain sets of classifiers with degree of similarity as variable as possible, and quite uniformly distributed over the entire range of variability (i.e. [ρ^{Max} , ρ^{Max}]= [0.8, 1.0]). Using the database of sets of synthetic classifiers, the behavior of classifier combination methods has been analyzed. Specifically, for each (K, $\underline{R}$, ρ), the performance **C**(K,$\underline{R}$,ρ) has been estimated by averaging the performance of the combination method obtained using the artificial data sets stored into the database and corresponding to the parameters (K,$\underline{R}$, ρ). Fig.5 shows the performance of Dempster-Shafer (DS) combination method [14], estimated by the artificial data sets of CGP_1 and CGP_2. Specifically, the DS scheme and decision rule proposed respectively in Section VI.C and Section VI.D (eq. [50], α=0) of ref. [1] have been considered for the test. The results point out two major aspects. The first aspect is that the results obtained using the artificial data sets from CGP_1 and CGP_2 are quite similar (in the domain in which both results have been derived). The second aspect is that CGP_2 allows the estimation of the performance for DS over the entire range of variability of the degree of similarity. It must be pointed out that, as discussed elsewhere [10], the performance estimated on artificial data set are very close to those obtained by using real classifiers. The difference between the results obtained by artificial and real data sets is less than 1.8% for reliability rate and less than 1.5% for recognition rate.

VI. Conclusion

This paper presents a new technique for the generation of sets of synthetic classifiers for the evaluation of decision combination methods. The technique is well-suited for the generation of data sets simulating sets of abstract-level classifiers which differ both in terms of individual characteristics (recognition rate) and collective behavior (degree of similarity of the set of classifiers).

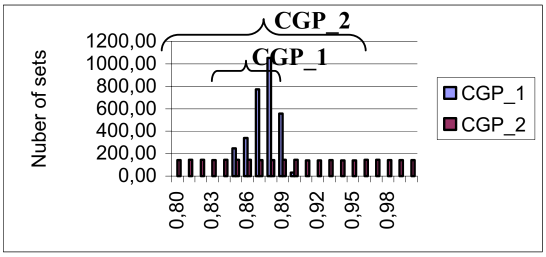

Figure 4. Distributions of sets of synthetic classifiers.

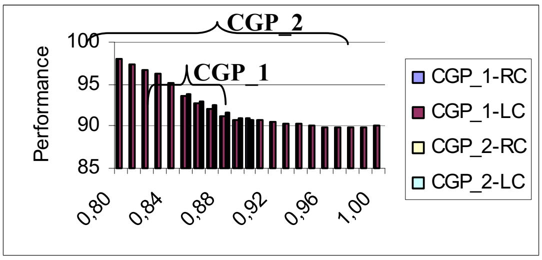

Figure 5. Performance Analysis by CGP_1 and CGP_2.

References

[1] L. Xu, A. Krzyzak, C. Y-Suen, “Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition”,*IEEE T-SMC*,Vol.22, N.3,1992, pp. 418-435.
[2] 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.
[3] E.M.Kleinberg,"Stochastic Discimination",Annals of Math. ematics and Artificial Intelligence, Vol. 1, pp. 207-239, 1990.
[4] S.N. Srihari, "Reliability Analysis of Majority Vote Systems", Information Sciences, Vol. 26, pp. 243-256, 1982.
[5] L.Lam, "Classifier Cotions: Implementations and Theoretical Issues", in *Multiple Classifier Systems*, J. Kittler an F. Roli (eds.), LNCS, vol. 1857, Springer, 2000, pp. 77-86.
[6] K. Tumer and J. Ghosh, "Analysis of Decision Boundaries in inearly Combined Neural Classifiers", Pattern Recognition, Vol. 29, pp. 341-348, 1996.
[7] H. Zouari, L. Heutte, Y. Lecourtier, A. Alimi, “ A New Classifier Simulator for Evaluating Parallel Combination Methods”, Proc. ICDAR 2003, Edinburgh, UK, *2003*, pp.26-30.
[8] J.R. Parker,”Rank and response combination from confusion matrix data”, Information Fusion, Vol. 2, 2001, pp. 113-120.
[9] L.I.Kuncheva and R.K. Kuntchev, “Generating classifier outputs of fixed accuracy and diversity”, PRL, Vol. 23, 2002, pp. 593-600.
[10] 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,* 2003, Vol. 6, pp. 42-54.
[11] 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 Scient.,Singapore, pp.329-335, 1995.
[12] V. Di Lecce,G.Dimauro,A.Guerriero,S.Impedovo,G.Pirlo,A. Salzo, "Knowledge-based Methods for Classifier Combination: an Experimental Investigation", Proc.ICIAP-7,Italy,1999, pp.562-565.
[13] S. Impedovo, G. Pirlo, “The Similarity Index”, Università degli Studi di Bari, Technical Report *TC. 1*, 2002.
[14] 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*, North Holland, Amsterdam, 1988, pp. 381-393.
[15] 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., 1997, pp. 419-425.