A PERTURBATION-BASED APPROACH FOR MULTI-CLASSIFIER SYSTEM DESIGN

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

Published in:
EPRINTS-BOOK-TITLE

Publication date:
2004

A PERTURBATION-BASED APPROACH FOR MULTI-CLASSIFIER SYSTEM DESIGN

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

This paper presents a perturbation-based approach useful to select the best combination method for a multi-classifier system. The basic idea is to simulate small variations in the performance of the set of classifiers and to evaluate to what extent they influence the performance of the combined classifier. In the experimental phase, the Behavioural Knowledge Space and the Dempster-Shafer combination methods have been considered. The experimental results, carried out in the field of hand-written numeral recognition, demonstrate the effectiveness of the new approach.

1 Introduction

The interest of researchers in multi-classifier systems is motivated by the consideration that they allow high performances also in the cases of difficult classification problems [1]. Since theoretical analysis of combination methods can be very difficult [2,3] the evaluation of complex combination methods is generally carried out on experimental basis [4,5]. The net result is that the design of multi-classifier systems still presents many unsolved issues [2,3].

An important issue is related to the selection of the best method to combine the individual classifier for a specific application. This aspect is generally evaluated by measuring the performance of the methods by using data sets derived from the real application. In this case, there is no assurance that similar results can be obtained in the working environment in which the performance of the individual classifiers can change [5].

In this paper a perturbation-based approach is presented for the selection of the best combination method for a multi-classifier system. Small variations in the performance of the set of classifiers are simulated and their effect to the performance of the combined classifier is evaluated. Some indexes are also provided to obtain information on the performance of the combined classifiers.

The organisation of the paper is the following: Section 2 presents the problem of selection of the combination method for multi-classifier system. The new perturbation-based approach is presented in Section 3. The experimental results are reported in Section 4.

In: L.R.B. Schomaker and L.G. Vuurpijl (Eds.), Proceedings of the Seventh International Workshop on Frontiers in Handwriting Recognition, September 11-13 2000, Amsterdam, ISBN 90-76942-01-3, Nijmegen: International Unipen Foundation, pp 553-558

2 Multi-Classifier System Design

In a multi-classifier system the final decision is obtained by combining the decisions of several classifiers [1]. When a *parallel-combination* topology is considered [6], the input pattern p_t is provided to each classifier $A_{i,}$ i=1,..,K which decides the membership of the pattern x_t to the pattern classes ω_1, ω_2,..., ω_m . Let $A_i(t)$ be the response of the i-th classifier when the pattern x_t is processed, the combination method M provides the final response by combining the responses of the individual classifiers. From the outputs of the individual classifiers, the method computes a confidence score $S(\omega_i)$ for each class ω_i, i=1,2,...,m. Successively a decision rule is used to produce the final classification response.

The performance of a combination method is generally evaluated on experimental basis by considering measures as the recognition rate R_M, the substitution rate E_M and the rejection rate L_M, or other indexes and cost functions generally defined as a linear combination of R_M , E_M and L_M [1,2].

Now, let be $C_M = L_M + a \cdot E_M$ the cost function considered for the evaluation of the combination methods [6], the selection of the best method among M_1 and M_2 for a multi-classifier system follows the rules:

- If $C_{M1} < C_{M2}$ then use M_1
- If $C_{M1} > C_{M2}$ then use M_2
- If $C_{M1} = C_{M2}$ then use M_1 or M_2 equivalently.

where C_{M1} and C_{M2} are evaluated using a test data set derived from the real working environment. Unfortunately, since the performance of the individual classifiers can change in time due to the learning capabilities of the classifiers or to the variability in input data quality, there is no assurance that the combination method will provide acceptable results for the specific application in the working environment. Therefore, it is important that the effectiveness of methods for classifier combination M_i is evaluated also when the characteristics of the set of individual classifiers change.

3 A perturbation-based approach to combination method selection

Let $A_1, A_2, \ldots, A_K$ be the individual classifiers combined by a combination method M. Let us consider the vector of responses obtained by inputting the patterns x_t belonging to a database T. If T contains N patterns, the vector of responses consists of N elements each having K responses (one for each classifier) $(A_1(t), A_2(t), \ldots, A_K(t))$, t=1,2,,N.

From the multi-dimensional vector, two kinds of features are derived:

- <u>Features at the level of individual classifiers.</u>

  The recognition rate of each classifier:

$$R_{Ai} = \frac{1}{N} \sum_t Q_R(A_i(t), x_t)$$

where

$$Q_R(A_i(t), x_t) = \begin{cases} 1 & \text{if} \quad A_i(t) = \omega_j \text{ and } x_t \text{ belongs to the class } \omega_j \text{ (i.e. } A_i(t) \text{ is correct)} \\ 0 & \text{otherwise} \end{cases}$$

Features at the level of the set of classifiers.

The correlation between classifiers [5]:

$$\rho_A = \left[1 \middle/ \binom{K}{2} \right] \cdot \left[\sum_{\substack{i,\, j = 1,\ldots K \\ i < j}} \rho_{A_i, A_j} \right]$$

where

$$\rho_{A_i, A_j} = \frac{1}{N} \sum_{t=1}^{N} Q_\rho(A_i(t), A_j(t))$$

and

$$Q_\rho(A_i(t), A_j(t)) = \begin{cases} 1 & \text{if} \quad A_i(t) = A_j(t) \\ 0 & \text{otherwise} \end{cases}.$$

Therefore, the performance of the combination method M can be considered as depending on the recognition rates of the individual classifiers $R=(R_{A1},R_{A2},\ldots,R_{AK})$ and on the correlation ρ of the set of classifiers

$$C_E(R,\rho)= L_E(R,\rho)+a\cdot E_E(R,\rho) \qquad (1)$$

where $L_M(R,\rho)$ and $E_M(R,\rho)$ denote the rejection rate and the substitution rate of the combination method M as functions of R and ρ.

The combination method M can be evaluated in different working conditions by perturbing the characteristics of the classifiers by mean of an automatic procedure. For this purpose the vector of responses of the individual classifiers can be modified according to the following considerations:

- Modification of features at the level of individual classifiers: For each classifier a variation is assumed in its recognition rate. For instance, for A_i whose recognition rate is originally of R_{Ai} , we have that the new recognition rate R^*_{Ai} will range in $[R_{Ai} - \delta R_{Ai}, R_{Ai} + \delta R_{Ai}]$, where δR_{Ai} is suitably defined.
- Modification of features at the level of the set of classifiers: A variation is assumed for the correlation ρ of the entire set of classifiers. Specifically we have that the new correlation ρ^* will range in $[\rho-\delta\rho, \rho+\delta\rho]$where $\delta\rho$ is suitably defined.

For example, let us consider the vector of responses in Figure 1 for A_1,A_2,A_3,A_4 (N=10), where: R→Recognition; S1,S2,S3,S4→Substitution (rejections are not considered in this example). It results R=(0.7,0.8,0.7,0.6) (in fact: R_{A1}=0.7, R_{A2}=0.8, R_{A3}=0.7, R_{A4}=0.6), and ρ=3.2/6. In this case, if we combine A_1,A_2,A_3,A_4 by the combination method M, we will achieve the value $C_M(R, \rho)= C_M(0.7, 0.8, 0.7, 0.6, 3.2/6)$. In order to investigate the behaviour of M, a perturbation can be performed on the vector of responses according to the considerations (a) and (b). In Figure 2 a perturbed version of the vector of responses is shown (the modifications are marked with "^"). First, the output of A_4 for pattern 3 has been changed (this modification changes the recognition rate of A_4 and also the value of ρ for the entire set of classifiers). Second, the outputs of A_2 for patterns 7 and 8 have been swapped (this modification does not change the recognition rate of A_2 but only the value of ρ for the entire set of classifiers). It is easy to verify that, for the case in Figure 2 we have R^*=(0.7,0.8,0.7,0.7) (in fact: R^*_{A1}=0.6, R^*_{A2}=0.6, R^*_{A3}=0.6, R^*_{A4}=0.7), and ρ^*=3.8/6.

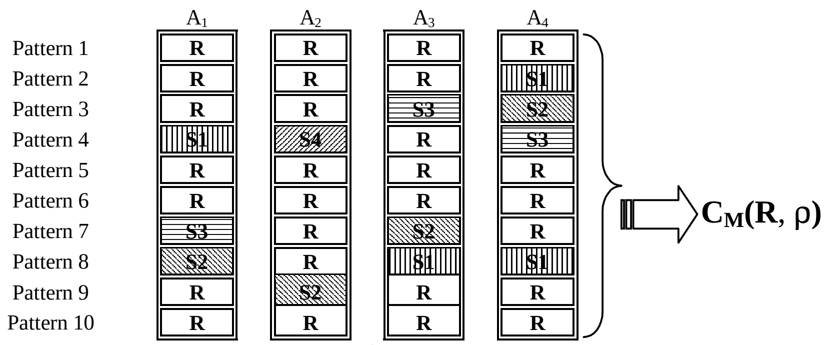

Figure 1: Vector of responses (0.7,0.8,0.7,0.6,3.2/6)

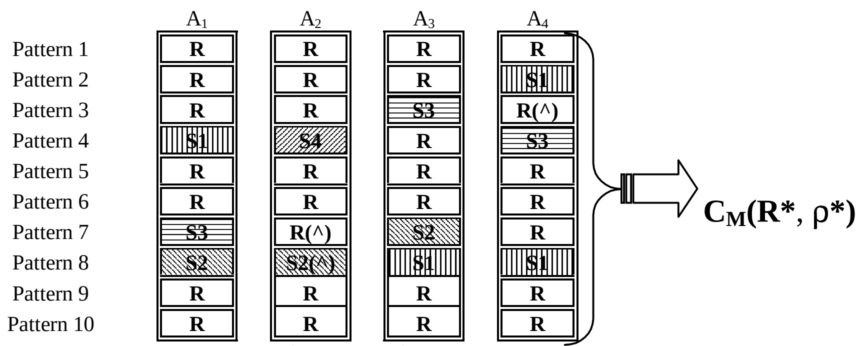

Figure 2: Vector of responses after perturbation (0.7,0.8,0.7,0.7,3.8/6).
Modifications: $A_4(3)$=R (Fig. 1: $A_4(3)$=S2);$A_2(7)$=R ; $A_2(8)$=S2 (Fig. 1: $A_2(7)$=S2 ; $A_2(8)$=R).

Now, let us consider the set of values assumed by the cost function for M when the entire set of perturbed vector is considered $\mathcal{C}_M=\{C_M(R^*, \rho^*) \mid (R^*, \rho^*)\in I^*\}$ where the perturbation range I^* is defined as $I^* =[R_{A1}-\delta R_{A1}, R_{A1}+\delta R_{A1}]\times[R_{A2}-\delta R_{A2}, R_{A2}+\delta R_{A2}]\times..\times[R_{AK}-\delta R_{AK}, R_{AK}+\delta R_{AK}]\times[\rho-\delta\rho,\rho+\delta\rho]$.

From the set $\mathcal{C}_M$, useful information on the effectiveness of M in different working conditions can be derived:

- **max**($\mathcal{C}_M$): the maximum value in $\mathcal{C}_M$;
- **min**($\mathcal{C}_M$): the minimum value in $\mathcal{C}_M$;
- **M**($\mathcal{C}_M$): the mean value of $\mathcal{C}_M$;
- **SD**($\mathcal{C}_M$): the standard deviation of $\mathcal{C}_M$.

These indexes allow a more accurate selection of the combination method for a multi-classifier system depending on the particular application field. Typical selection criteria can be:

- **max**($\mathcal{C}_M$) minimum (worst performance as good as possible);
- **max**($\mathcal{C}_M$) minimum (best performance as good as possible);
- **M**($\mathcal{C}_M$) minimum (best performance on average);
- **SD**($\mathcal{C}_M$) minimum (performance as stable as possible).

4 Experimental Results

The proposed approach has been applied to a system for hand-written numeral recognition that implements both the Behavioural Knowledge Space Method (BKS) [7] and the Dempster-Shafer Method (DS) [8]. The system [9] combines four classifiers trained with about 18468 numerals of the CEDAR Database (BR directory) [10]: *Region* ($R_{A1}=0.91$), *Contour* ($R_{A2}=0.87$), *Loci* ($R_{A3}=0.90$), *Histogram* ($R_{A4}=0.87$).

Initially, from the vector of responses of the set of classifiers (tested with data of the BS directory of the CEDAR database) we derive the values R=(0.91,0.87,0.90,0.87) and ρ=0.84. From eq. (1), (with a=10), it results C_{BKS}=0,485 and C_{DS} =0,472. Hence DS outcomes BKS.

Now, when the perturbation-based approach is applied to investigate the behaviour of BKS(R*, ρ*) and DS(R*, ρ*) for $(R^*, \rho^*) \in I^*$, where $\delta R_{Ai}=0.01$ (a variation of 1% is considered for the recognition rate of the individual classifiers) and $\delta\rho=0.03$, we obtain the sets $\mathcal{C}_{BKS}$ and $\mathcal{C}_{DS}$ as Table 1 shows.

Table 1: Perturbation-based analysis of BKS and DS

	$\mathcal{C}_{BKS}$	$\mathcal{C}_{DS}$
max($\mathcal{C}_M$)	**0.525**	**0.690**
min($\mathcal{C}_M$)	**0.475**	**0.323**
M($\mathcal{C}_M$)	**0.493**	**0.491**
SD($\mathcal{C}_M$)	**0.019**	**0.156**

The result in Table 1 shows that, although DS and BKS have similar mean performance (**M**($\mathcal{C}_{DS}$)≅**M**($\mathcal{C}_{BKS}$)), DS outperforms BKS in terms of best result ((**min**($\mathcal{C}_{DS}$)<**min**($\mathcal{C}_{BKS}$)), while BKS results more stable than DS ((**SD**($\mathcal{C}_{DS}$)>**SD**($\mathcal{C}_{BKS}$)) and it is also superior to DS in terms of worst result ((**max**($\mathcal{C}_{DS}$)>**max**($\mathcal{C}_{BKS}$)).

5 Conclusions

This paper presents a new perturbation-based approach for the selection of the combination method best suited in a multi-classifier system for *abstract-level* classifiers. The approach provides useful information to evaluate the effectiveness of a combination method in real environments, where modifications of working conditions can occur due to variations in quality of input patterns or to learning capabilities of classifiers.

References

1. L.Xu, A.Krzyzak, C.-Y Suen, “Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition”, *IEEE SMC*, Vol. 22, N.3, 1992, pp. 418-435.
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. J. Kittler, M. Hatef, R.P.W. Duin, J. Matas, "On Combining Classifiers", IEEE PAMI, Vol. 20, n. 3, 1998, pp. 226-238.
4. C. Nadal, R. Legault, and C.Y. Suen, “Complementary algorithms for the recognition of totally un-constrained hand-written numerals”, *Proc. 10th ICPR.*, Vol. A, June 1990, pp. 434-449.
5. G. Dimauro, S. Impedovo, G. Pirlo, “Multiple Experts: a new methodology for the evaluation of the combination processes”, *Progress in Handwriting Recognition*, ed. by A.C.Downton and S. Impedovo, World Scientific Publishing Co. Pte. Ltd., Singapore, 1997, pp. 329-335.
6. A.F.R. Rahman, M.C. Fairhurst, “Introducing New Multiple Expert Decision Combination Topologies: A Case Study using Recognition of Hand-written Characters”, *Proc. ICDAR’97*, Ulm, Germany, 1997, pp. 886-891.
7. 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.
8. 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 and L.N.Kanal, North Holland, Amsterdam, 1988, pp. 381-393.
9. 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.
10. J.J. Hull, “A database for handwritten text recognition research", *IEEE PAMI*, Vol. 16, N. 5, 1994, pp. 550-554.