

Rapid and brief communication

# A study on the performances of dynamic classifier selection based on local accuracy estimation

Luca Didaci\*, Giorgio Giacinto, Fabio Roli, Gian Luca Marcialis

*Department of Electrical and Electronic Engineering, University of Cagliari, Piazza d'Armi, 09123 Cagliari, Italy*

Received 1 February 2005; accepted 11 February 2005

---

## Abstract

Dynamic classifier selection (DCS) plays a strategic role in the field of multiple classifier systems (MCS). This paper proposes a study on the performances of DCS by Local Accuracy estimation (DCS-LA). To this end, upper bounds against which the performances can be evaluated are proposed. The experimental results on five datasets clearly show the effectiveness of the selection methods based on local accuracy estimates.

© 2005 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.

**Keywords:** Multiple classifier systems; Dynamic classifier selection; Performance evaluation

---

## 1. Introduction

Dynamic classifier selection (DCS) could play a strategic role in the field of multiple classifier systems (MCS). In fact MCS take advantage of the strengths of the individual classifiers, avoid their weaknesses, and improve classification accuracy. Let us consider an *oracle* that for each pattern always predicts the classifier that gives the correct label. If such an oracle is available, for each pattern we can retain the decision from the selected classifier, and ignore those by others. This is the ideal case of dynamic classifier selection. Algorithms for DCS by local accuracy (DCS-LA) has been first proposed by Woods et al. [1], and Giacinto and Roli [2]. DCS-LA is based on estimating the accuracy of each classifier of the MCS in a local region surrounding the pattern to be classified, and selecting the classifier that exhibits the higher accuracy. While a number of experimental results showed the effectiveness of the DCS-LA approach,

its performances are usually much lower than that of the oracle. It is easy to see that the performances of the oracle are too optimistic to be considered as an useful upper bound for selection mechanisms based on local accuracy estimates. In this paper we present a study on the performances of DCS-LA mechanisms in terms of three upper bounds. The proposed upper bounds provide also some indication about future directions for improving DCS-LA algorithms.

## 2. Dynamic classifier selection by local accuracy estimation

Let us consider a set of  $N$  classifiers  $C = \{C_1, \dots, C_N\}$  which have already been trained to solve the  $M$ -class classification task at hand. For each unknown test pattern  $\mathbf{x}^*$ , let us consider a local region  $\mathcal{R}(\mathbf{x}^*)$  of the feature space surrounding pattern  $\mathbf{x}^*$ . Typically,  $\mathcal{R}(\mathbf{x}^*)$  is defined in terms of the  $k$ -nearest neighbours in the validation data. Validation data are extracted from the training set and are not used for classifier training. Let us define  $\mathcal{R}_k(\mathbf{x}^*)$  as the local region made up of the  $k$ -nn of pattern  $\mathbf{x}^*$ . Let  $LA_{j,k}(\mathbf{x}^*)$  be the local accuracy estimate for classifier  $C_j$

---

\* Corresponding author. Tel.: +39 070 675 5776;  
fax: +39 070 675 5782.

E-mail addresses: [luca.didaci@diee.unica.it](mailto:luca.didaci@diee.unica.it) (L. Didaci),  
[giacinto@diee.unica.it](mailto:giacinto@diee.unica.it) (G. Giacinto), [roli@diee.unica.it](mailto:roli@diee.unica.it) (F. Roli),  
[marcialis@diee.unica.it](mailto:marcialis@diee.unica.it) (G.L. Marcialis).

in the local region. The DCS-LA algorithm is defined as follows [1,2]:

- if all the classifiers  $C_j$  assign  $\mathbf{x}^*$  to the same class, then the pattern is assigned to this class; otherwise,
- compute  $LA_{j,k}(\mathbf{x}^*)$ ,  $j = 1, \dots, N$  (see below),
- identify the classifier  $C_j | LA_{j,k}(\mathbf{x}^*) = \max_i (LA_{i,k}(\mathbf{x}))$  and assign  $\mathbf{x}^*$  to the output class of  $C_j$ .

It is easy to see that DCS-LA accuracy strictly depends on the correctness of the LA estimate. Let us recall how LA can be computed.

### 2.1. Local accuracy “a priori”

We can define the local accuracy  $LA_{j,k}(\mathbf{x}^*)$  as the percentage of correctly classified patterns in the local region  $\mathfrak{R}_k(\mathbf{x}^*)$  of  $\mathbf{x}^*$ . Suppose that  $k_T$  out of the  $k$  samples in  $\mathfrak{R}_k(\mathbf{x}^*)$  are correctly classified by classifier  $C_j$ . An estimate of the LA in the vicinity of  $\mathbf{x}^*$  is  $k_T/k$ . This LA estimate has been named “a priori” LA, a.k.a. “overall LA”, because the class assigned by the classifier  $C_j$  to test pattern  $\mathbf{x}^*$  is not taken into account [1,2].

### 2.2. LA “a posteriori”

The estimation of the LA  $LA_{j,k}(\mathbf{x}^*)$  “a posteriori”—a.k.a. “local class accuracy” [1,2]—exploits the information on the class assigned by the classifier  $C_j$  to the test pattern  $\mathbf{x}^*$ . Let us suppose that the classifier  $C_j$  assigns the pattern  $\mathbf{x}^*$  to the class  $\omega_p$ , that is,  $C_j(\mathbf{x}^*) = \omega_p$ . Thus we can define

$$LA_{j,k}(\mathbf{x}^*) = N_{pp} / \sum_{i=1}^M N_{ip}, \quad (1)$$

where  $N_{pp}$  is the number of neighbourhood patterns of  $\mathbf{x}^*$  that are correctly assigned by  $C_j$  to class  $\omega_p$ , and  $\sum_{i=1}^M N_{ip}$  is the total number of neighbourhood patterns that are assigned by  $C_j$  to class  $\omega_p$ . In this case, LA is computed “a posteriori”, that is, after each classifier  $C_j$  produces the output on the test pattern  $\mathbf{x}^*$ . In order to handle the “uncertainty” in defining the appropriate neighbourhood size, different authors proposed to introduce some distance-weighted schemes (see, for example, [2]). In addition, if the classifier outputs can be regarded as estimates of the posterior probabilities, these outputs can be taken into account in order to improve the estimation of the above LA.

$$LA_j(\mathbf{x}^*) = \left( \sum_{x_n \in \omega_p} \hat{P}(\omega_p | \mathbf{x}_n) \frac{1}{d_n} \right) / \left( \sum_{m=1}^M \sum_{x_n \in \omega_m} \hat{P}(\omega_p | \mathbf{x}_n) \frac{1}{d_n} \right). \quad (2)$$

The term  $d_n$  is the distance of the “neighbouring” pattern  $\mathbf{x}_n$  from the test pattern  $\mathbf{x}^*$ . The term  $\hat{P}(\omega_p | \mathbf{x}_n)$  is the estimate of the posterior probability of the class  $\omega_p$  for the ‘neighbouring’ pattern  $\mathbf{x}_n$ . In the following we will refer to the implementations of the LA estimate defined in (1) and (2).

## 3. Upper bounds for the performances of DCS by LA

Usually, the effectiveness of selection mechanisms is evaluated by comparing the attained performances to those of the oracle. Unfortunately, the performances of the oracle are too optimistic to be considered as an useful upper bound for the DCS-LA. From a theoretical point of view, the DCS-LA aims to realize a Bayesian classifier. In fact, if the classifiers in the MCS make optimal Bayes decisions in complementary parts of the feature space, DCS-LA can output the decisions of the optimal classifier by choosing the classifier that locally exhibits the lowest classification error [3]. It is clear that the Bayesian classifier is the true upper bound for the DCS-LA accuracy. On the other hand, the performances of the oracle can outperform those of the Bayesian classifier. In fact, the oracle can correctly classify even those patterns that should be wrongly classified according to the Bayesian decision theory as it is sufficient that a classifier outputs the correct label for those patterns. Concerning the lower bound of the accuracy, the DCS-LA performs at least as good as the best classifier in the ensemble, in the hypothesis that we are able to choose the most accurate classifier in the vicinity of the test pattern  $\mathbf{x}^*$  [4].

In the following, we propose three upper bounds for DCS-LA designed to provide a more realistic limit to the performances that can be attained by DCS-LA approaches. The proposed upper bounds also provide some indications about future directions for improving DCS-LA mechanisms. Let  $\mathbf{x}^*$  be the test pattern,  $\omega_m$  its true label and  $C_j(\mathbf{x}^*)$  the label assigned by the classifier  $C_j$ .

### 3.1. Ideal selector for performance evaluation: oracle k-best

Let us assume that a classifier in the ensemble, for example  $C_i$ , correctly classify the pattern  $\mathbf{x}^*$ . The ‘oracle k-best’ selects  $C_i$  only if it is possible to find a  $k$  such that the local accuracy estimate for  $C_i$  in  $\mathbf{x}^*$  is greater than (or equal to) the LA estimate for other classifiers. Otherwise, the oracle k-best cannot select the classifier  $C_i$ .

*Oracle k-best:* if  $\exists(i, k) | C_i(\mathbf{x}^*) = \omega_m$  and  $LA_{i,k}(\mathbf{x}^*) \geq LA_{j,k}(\mathbf{x}^*) \forall j$  then select  $C_i$ .

This upper bound exhibits the maximum accuracy attainable using the considered selection mechanism, under the assumption that, for each pattern, the optimal size of the local region can be estimated correctly.

### 3.2. Ideal selector for performance evaluation: oracle $k$ -best<sub>AVE</sub>

Typically LA estimates are computed using a value of  $k$  computed by experiments on the training or validation set. In order to define an upper bound for the DCS-LA mechanisms, the optimal size of the neighbourhood can be estimated directly on the test set. Values of  $k$  in a predefined range are evaluated, and the value that allows attaining the maximum accuracy on the test set is selected. It is worth noting that this ideal selector exhibits less degrees of freedom with respect to the ‘oracle  $k$ -best’. Let  $\text{Acc}(\text{DCS}_k)$  be the accuracy attained by the DCS-LA on the test set using a local region of  $k$  patterns.

Oracle  $k$ -best<sub>AVE</sub> can be defined as follows:

select  $\hat{k} | \text{Acc}(\text{DCS}_{\hat{k}}) \geq \text{Acc}(\text{DCS}_k) \forall k$ ,  
 apply the DCS-LA method estimating the local accuracy using a local region of  $\hat{k}$  patterns.

Oracle  $k$ -best<sub>AVE</sub> exhibits the maximum accuracy attainable using the considered selection mechanism, under the assumption that we can choose the optimal  $k$  parameter for the LA estimate.

### 3.3. Ideal selector for performance evaluation: oracle $\theta$ -best<sub>AVE</sub>

The correctness of the local accuracy estimation depends also on the variability of the true accuracy of classifiers in  $\mathfrak{R}_k(\mathbf{x}^*)$ . The smaller the variations of accuracy in  $\mathfrak{R}_k(\mathbf{x}^*)$ , the smaller the error of the estimate. To take this effect into account, we propose a modification of the ‘oracle  $k$ -best<sub>AVE</sub>’ where the shape of  $\mathfrak{R}_k(\mathbf{x}^*)$  is adapted so that the variations of LA are smaller than those in a spherical neighbourhood of the same size. This adaptive-neighbourhood DCS-LA is an upper bound for DCS-LA algorithms, as they are based on spherical neighbourhoods. We define an adaptive neighbourhood according to the DANN (discriminant adaptive nearest neighbourhood) metric proposed for the  $k$ -nn classification task [5]. The DANN metric exploits the information about the *between* and *within* scatter matrixes,  $\mathbf{B}$  and  $\mathbf{W}$ , respectively, computed in a  $K_M$ -nearest neighbour of the test pattern  $\mathbf{x}^*$ , and uses them to define a metric  $\Sigma = \mathbf{W}^{-1}\mathbf{B}\mathbf{W}^{-1}$ . The  $k$ -nn of  $\mathbf{x}^*$  is then made up of the  $k$  nearest patterns according to the distance metric  $d^2(\mathbf{x}_1, \mathbf{x}_2) = (\mathbf{x}_1 - \mathbf{x}_2)^T \Sigma (\mathbf{x}_1 - \mathbf{x}_2)$ . The ‘unit sphere’ according to  $\Sigma$  is thus an ‘hyper ellipsoid’ whose orientation and eccentricity depend on the local density of the classes. In the case of DCS-LA, we re-label validation patterns as ‘correctly classified’ or ‘incorrectly classified’, and compute the DANN metric as in a two-class classification problem. The resulting metric ‘weights’ the distances according to the position of the correctly and incorrectly classified validation patterns in the feature space, so that local accuracy values exhibit a small variability in this ‘adaptive’ neighbourhood.

It is worth noting that this metric depends both on the test pattern at hand and on the value of the  $K_M$  parameter used to estimate the scatter matrixes.

As in the ‘oracle  $k$ -best<sub>AVE</sub>’, the values of the parameters  $\{K_M, k\}$  are chosen optimizing the DCS-LA performances on the entire test set. Let  $\text{Acc}(\text{DCS}_\theta)$  be the accuracy attained by the DCS-LA on the test set using  $\theta = \{K_M, k\}$  in order to compute the adaptive metric and the corresponding adaptive local region. Then we define

Oracle  $\theta$ -best<sub>AVE</sub>

select  $\hat{\theta} | \text{Acc}(\text{DCS}_{\hat{\theta}}) \geq \text{Acc}(\text{DCS}_\theta) \forall \theta$ ,  
 apply the DCS-LA method estimating the local accuracy in a local region defined by  $\hat{\theta}$ .

## 4. Experimental results

The aim of the experiments is to compare the accuracy attained by DCS-LA with the proposed upper bounds. We use an MCS made up of three base classifiers, namely linear Bayes, quadratic Bayes, and  $k$ -nn. These classifiers have been chosen because their design requires the tuning of few parameters. In particular, the value of  $k$  in the  $k$ -nn classifier is chosen by the leave-one-out procedure on the training set, while default parameters have been used for all the other classifiers. The five data sets used in the experiments and the accuracy attained by the best single classifier are reported in Table 1. Reported accuracies have been computed as follows. Each data set has been first partitioned into two equal halves, one partition is used for training, while the other for testing. Experiments have been carried out by the 5-fold cross-validation procedure, and exchanging the role of the two partitions. Finally, the reported values of accuracy are computed as the average over the 10 trials.

For evaluating the performances of the DCS-LA algorithms [1,2] we used the value of the  $k$  parameter as the one that maximize the performances on the validation set. The accuracies of DCS-LA, those of the upper bounds defined in Section 3, and those of the oracle are reported in Table 2.

Reported experiments show that the performances of DCS-LA are close to those of the ‘oracle  $k$ -best<sub>AVE</sub>’ and ‘oracle  $\theta$ -best<sub>AVE</sub>’ upper bounds, therefore the considered DCS-LA implementations attain optimal performances.

Table 1

Datasets used in the experiments and performances attained by the best classifier in the MCS

| Dataset  | Patterns | Features | Classes | Best single classifier (%) |
|----------|----------|----------|---------|----------------------------|
| Phoneme  | 5404     | 5        | 2       | 86.55                      |
| Vehicle  | 846      | 18       | 4       | 89.29                      |
| Letter   | 20000    | 16       | 26      | 86.99                      |
| Satimage | 6435     | 36       | 6       | 83.00                      |
| Feltwell | 10944    | 15       | 5       | 92.90                      |

Table 2

Accuracy attained by the DCS-LA methods, the proposed upper bounds, and the oracle

| Data set | DCS using<br>non probabilistic<br>LA (%) | DCS using<br>probabilistic<br>LA (%) | Oracle<br>$k$ -best <sub>AVE</sub><br>(%) | Oracle<br>$\theta$ -best <sub>AVE</sub><br>(%) | Oracle $k$ -best<br>(%) | Oracle<br>(%) |
|----------|--|--------------------------------------|---|--|-------------------------|---------------|
| Phoneme  | 85.14                                    | 85.90                                | 86.47                                     | 87.05  | 92.91                   | 94.42         |
| Satimage | 89.10                                    | 89.18                                | 89.51                                     | 89.79  | 93.54                   | 95.11         |
| Feltwell | 83.89                                    | 86.38                                | 86.62                                     | 87.85  | 89.93                   | 93.13         |
| Vehicle  | 77.14                                    | 78.16                                | 80.06                                     | 80.54  | 88.18                   | 91.70         |
| Letter   | 91.27                                    | 91.44                                | 91.72                                     | 92.07  | 96.21                   | 97.10         |

This result confirms the correctness of the use of a spherical local region, and the validity of the estimation of the size of  $k$  on the validation set as proposed in literature [1,2]. As far as the ‘oracle  $k$ -best<sub>AVE</sub>’ is concerned, these results show that the sub-optimal choice of parameters on the validation set does not degrade the performances compared with the optimal choice of the parameters, that is, the choice of the parameters is not critical. Moreover, the performances of the ‘oracle  $\theta$ -best<sub>AVE</sub>’ show that the use of a more complex, adaptive shape for the local accuracy estimate do not provide significant improvements in accuracy compared to those attained using a spherical region. However, it is worth noting that these upper bounds are still far from those of the oracle.

Finally, the upper bound ‘oracle  $k$ -best’, that choose dynamically the size of the local region, provides substantial performance improvements compared to those of DCS-LA. In particular, performances are close to those of the oracle and outperform the best single classifier in the ensemble. These results clearly show that a size-adaptive local region can boost the performances of DCS-LA mechanisms.

## 5. Conclusions

The upper bounds for DCS-LA proposed in this work are more realistic than the so-called ‘oracle’, and they can be

attained in real cases by accurately tuning the parameters of DCS-LA. Thus the reported results confirm the effectiveness of the DCS-LA algorithms proposed in the literature [1,2].

Despite the fact that the oracle is a too optimistic upper bound, reported results also show that the difference between the performances of the considered DCS-LA algorithms and those of the oracle can be reduced by dynamically choosing the optimum size of the local region.

## References

- [1] K. Woods, W.P. Kegelmeyer, K. Bowyer, Combination of multiple classifiers using local accuracy estimates, *IEEE Trans. Pattern Anal. Mach. Intell.* 19 (4) (1997) 405–410.
- [2] G. Giacinto, F. Roli, Adaptive selection of image classifiers, *Proceedings of ICIAP '97, Lecture Notes in Computer Science*, vol. 1310, 1997, pp. 38–45.
- [3] G. Giacinto, F. Roli, Methods for dynamic classifier selection, *Proceedings of ICIAP '99, Italy*, 1999, pp. 659–664.
- [4] L.I. Kuncheva, Switching between selection and fusion in combining classifiers: an experiment, *IEEE Trans. Syst. Man Cybernet. Part B* 32 (2) (2002) 146–156.
- [5] T. Hastie, R. Tibshirani, Discriminant adaptive nearest neighbour classification, *IEEE Trans. Pattern Anal. Mach. Intell.* 18 (6) (1996) 607–615.