Francesc J. Ferri

Combining adaptive vector quantization and prototype selection techniques to improve nearest neighbour classifiers

Kybernetika, Vol. 34 (1998), No. 4, [405]--410

Persistent URL: `http://dml.cz/dmlcz/135223`

KYBERNETIKA — VOLUME 34 (1998), NUMBER 4, PAGES 405 – 410

COMBINING ADAPTIVE VECTOR QUANTIZATION AND PROTOTYPE SELECTION TECHNIQUES TO IMPROVE NEAREST NEIGHBOUR CLASSIFIERS

FRANCESC J. FERRI[1]

Prototype Selection (PS) techniques have traditionally been applied prior to Nearest Neighbour (NN) classification rules both to improve its accuracy (editing) and to alleviate its computational burden (condensing). Methods based on selecting/discarding prototypes and methods based on adapting prototypes have been separately introduced to deal with this problem. Different approaches to this problem are considered in this paper and their main advantages and drawbacks are pointed out along with some suggestions for their joint application in some cases.

1. INTRODUCTION

Prototype Selection (PS) techniques have traditionally been applied prior to Nearest Neighbour (NN) classification rules both to improve its accuracy (editing) and to alleviate its computational burden (condensing). Even though this second goal has become less and less important considering the current computational systems, the first one can still be considered of extreme importance specially in the cases in which obtaining an appropriate Training Set (TS) is a difficult task in itself. Potentially, these techniques can be jointly applied with any type of (TS based) classifier [9] in order to improve or to speed its parameter estimation (or training) phase. Nevertheless, the most common practice consists of using selected prototypes to build a compact and accurate NN classifier.

Usually, two separate families of techniques called *Editing* and *Condensing* have been applied for PS purposes. Editing aims at selecting a subset from the TS which exhibits better performance in classification rate while Condensing tries to select a reduced subset with (approximately) the same behaviour in classification than the (usually edited) TS with respect to the 1-NN rule. These two processes can also be jointly applied to improve both the effectiveness and the computation cost [1].

A rather different approach to this problem was introduced by Kohonen with the Learning Vector Quantization (LVQ) method. LVQ aims at placing a given number of prototypes in the representation space in such a way that these prototypes reflect the probability distributions of the whole TS, so that they can be used to apply the

[1]This work was partially supported by the Spanish CICYT project TIC95-676-C02-01.

NN rule to approximate the performance of the Bayes classifier. This method and further extensions of it are usually referred to as *adaptive* methods because of their close relation to adaptive methods for vector quantization [7] and have been already compared to some classical PS techniques [10].

In the next section, several PS methods are briefly explained, while in Section 3, both adaptive and consistency-based methods are considered as a way to obtain appropriate NN classifiers. Comparative experiments are carried out and explained in Section 4 and then main conclusions are outlined in Section 5.

2. BASIC PROTOTYPE SELECTION METHODS

The main goal of the first Edited NN rule [11], consisted of discarding some prototypes from a given TS. Editing aims both at detecting outliers and also to "clean" the overlapping regions (close to the class boundaries). This gives rise to an increase in performance. The Edited NN has been deeply analysed in the asymptotic case and optimal procedures (in the Bayes sense) have been introduced [1]. The general idea below almost any editing procedure consists of estimating the true classification of prototypes to retain only those which are correctly classified. This estimation process can be performed in a leave-one-out way [11], in a Hold-out way [1] or even using cross-validation [3]. Repeating this process in a Hold-out way using the 1-NN rule gives rise to the well known *Multiedit* algorithm.

The condensing idea was first proposed by Hart [6] who introduced a straightforward algorithm based on the concept of *consistency*. A subset of labelled prototypes is said to be *consistent* with respect to another if the first set correctly classifies the last one using the 1-NN rule. Hart's algorithm presents some important drawbacks. First, the subset it selects is not minimal as it was shown by Gates [4] and it may depend on the order in which prototypes are taken into account. More importantly, the consistency property does not guarantee obtaining the same (or approximately the same) performance as with the original TS. In fact, when applied to a TS with some overlapping among classes, Hart's condensing tends to retain a lot of (undesirable) prototypes to maintain the consistency property and this generally leads to a very bad generalisation with respect to the behaviour of the original TS.

Adaptive PS methods can be considered as a refining process because they start from an appropriate set of prototypes that are now called *codebook vectors* (CV). The general method consists then of an iterative procedure to move CVs around in which their position (in a given iteration) is set using a reward-punishment criterion. The displacement of a CV is computed in each iteration depending on how it contributes to the (correct) classification of samples, its position in the previous iteration, and a certain scalar gain factor (that must be a monotonically decreasing function in the number of iterations). Therefore, the final position of CVs depends of their number and initial position (initialisation), the reward-punishment method, and the behaviour of the scalar gain factor.

Kohonen [7] showed that the LVQ procedure moves the CV in the representation space so that they approximate the probability distribution functions in each class. As a consequence they tend to uniformly fit the class acceptance regions in the best possible way and, correspondingly, if the number of CV is large enough, they also approximate the Bayes decision boundaries. Several modifications and improvements of this basic procedure have also been proposed [8].

A convenient way to improve the results obtained with LVQ consists of appropriately modifying the reward-punishment rules to directly fit the *boundaries* between classes with pairs of prototypes instead of approximating the probability distributions. This led to the so-called Decision Surface Mapping (DSM) method [5] in which pairs of prototypes are moved only when one of them is misclassified.

3. COMBINED APPROACHES TO SELECTING PROTOTYPES

There exists a high similarity between LVQ and the combined Multiedit-Condensing (MEC) from the point of view of the final result obtained by both methods. Both attempt to obtain a reduced set of prototypes whose 1-NN induced decision boundary approximates that of the corresponding Bayes classifier.

MEC is a composite approach whose first step consists of "cleaning" the boundaries between the Bayes acceptance regions. Then, the condensing phase selects a reduced set of prototypes consistent with the edited set. The effectiveness of the method mainly depends on the Editing phase [1] which can be arbitrarily bad in the small sample size case [3]. The LVQ method on the other hand, tries to represent the underlying statistics in the TS with a *fixed* number of CVs. As a sub-product, the 1-NN classification rule using these prototypes approximates the Bayes rule [7]. This method does not exhibit the same problems for small TS than the MEC approach due to its adaptive nature but suffers from some initialisation problems [7, 8].

The DSM method looks for a solution as close as possible to the 1-NN boundary induced by the sample set. In this sense, its behaviour is quite close to that of the consistency-based condensing methods despite its adaptive nature. In fact, the DSM method stops if a consistent solution is found. An important drawback of DSM comes from the fact that it is not applicable if there is overlapping among classes.

To make the DSM usable for practical problems it is possible to share the benefits of two opposite approaches: Multiedit (which is Bayes optimal) followed by consistency-based condensing (needs no initialisation and serves as very good initialisation procedure), and DSM (accurately modifies condensed prototypes to fit the 1-NN frontiers induced by the edited prototypes). The problem is that DSM cannot further improve an already consistent set. Even if using alternative initial CVs, there is no advantage in using DSM compared to plain condensing (the method stops when the first consistent solution is found). It is possible to modify this behaviour by using *all the original prototypes* instead of the edited ones only, but relabelling those discarded by the editing procedure according to the classification labels yielded by the edited TS. This fact contributes to tightly define the 1-NN classification boundaries induced by the *edited prototypes* and results in a better behaviour of the DSM.

For illustration purposes, the Multiedit algorithm has been applied to a problem introduced by Hart [6]. In Figure 1 (a), the DSM and condensed decision boundaries using similar number of prototypes are shown. Figure 1 (b) shows that further improvement can be obtained by relabelling the prototypes discarded by the Multiedit algorithm and applying the DSM method as discussed above. It is important to note that this hybrid approach improves the results of consistency-based condensing by approximating the performance of the Edited set. The condensed boundaries are in fact moved to the edited ones. This implies that the results obtained in this way will be (in principle) bounded by the performance of the Editing method itself. Nevertheless, the boundary simplification due to the use of the DSM result may improve the classification accuracy of the edited set specially with small TSs.

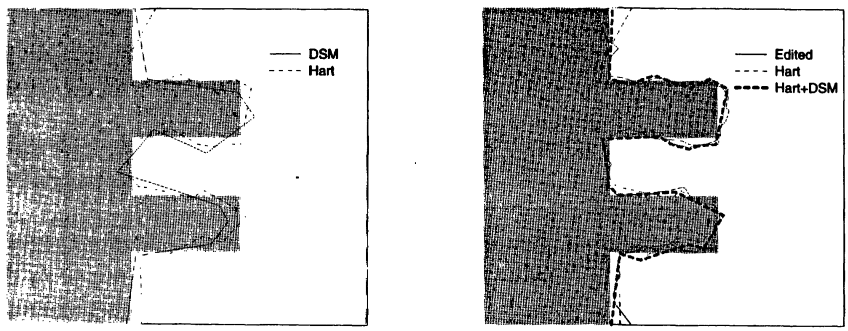

Fig. 1. Decision boundaries obtained with different condensing methods applied after Editing. (a) Adaptive (DSM) and Hart's condensings, (b) Hart's condensing, relabelled DSM (Cond+DSM) and the original (Edited) TS.

4. COMPARATIVE EXPERIMENTS

Several synthetic experiments have been designed to compare the LVQ method with MEC, and also DSM with consistency-based condensing both after Editing. The reported results correspond to the average over 8 repeated trials with randomly generated sets using typical parameter settings for each algorithm.

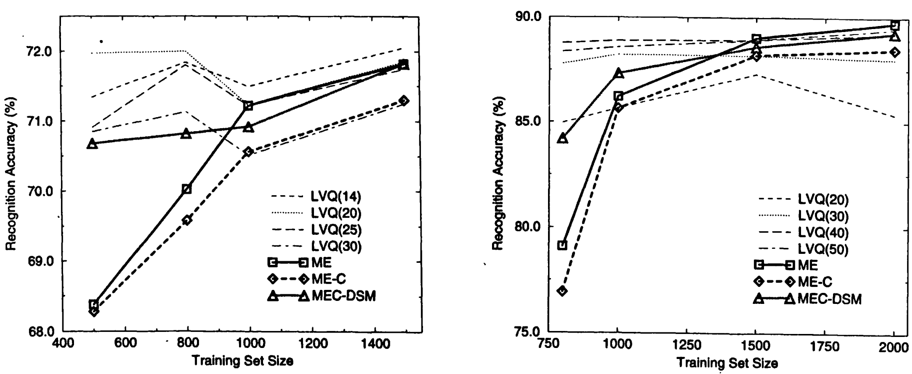

Fig. 2. Recognition accuracy for different TS sizes; (a) two Gaussian problem, and (b) two spirals problem.

First, a two-class bivariate normal distribution with the same (zero) mean and different diagonal covariance matrices ($\sigma_{11} = \sigma_{22} = 1$ for class 1 and $\sigma_{11} = \sigma_{22} = 4$ for class 2) with TSs of 500, 800, 1000 and 1500 samples was considered. Also, a two-class problem corresponding to two embedded spirals ($x_1(t) = 10t\sin(t); y_1(t) = 10\cos(t)$, and $x_2(t) = -x_1(t); y_2(t) = -x_1(t)$, $t \in [0, 10]$) with additive bivariate Gaussian noise ($\sigma_{11}(t) = \sigma_{22}(t) = t + 3.5$) with TSs of 800, 1000, 1500 and 2000 samples was considered. Improved versions of the Multiedit algorithm [3] were used to obtain better results for small TS sizes.

From the results shown in Figure 2, it is possible to conclude that the LVQ method yields the best results in both experiments, regardless of the TS size. However, the performance gets worse depending on the number of CVs. This confirms the high dependence of this method on the initialisation parameters. It is also interesting to note that the DSM always improves the results obtained by the MEC approach and even the Multiedit itself for small TS sizes.

Table 1. Error rate and set sizes for the colour pixel classification problem.

	LVQ	Multiedit	Condensing	DSM
Prototypes retained	15	2693	13	13
% error	5.15	3.97	5.05	4.25

Finally, another experiment involving real data concerning colour image segmentation has also been considered. The problem is stated in [2] and consists of performing pixel classification according to a 11-dimensional vector representation of colour and contextual information for each pixel of an image. A relatively large TS consisting of 3500 pre-labelled pixels obtained from significant zones of three training images is processed to obtain a reduced set of prototypes to segment seven different test images. The approximate number of prototypes in the considered test set is 100,000. The LVQ method and the hybrid approach discussed here have been applied and compared to the MEC approach and the results are shown in Table 1. It is possible to observe that in this single experiment (only one TS) the Multiedit algorithm gives the best result. And, even when processed by condensing and DSM, the result is better than the one obtained with LVQ.

5. CONCLUDING REMARKS

A comparative study about practical application of different PS techniques has been presented. The joint use of adaptive and consistency-based methods tries to overcome one of the main drawbacks of the first methods, i.e. the initialisation problem and also the drop in performance suffered by condensing methods when compared to Editing. Moreover, it is possible to share the benefits of both families of methods and, in particular, applying DSM to a multiedited, condensed and relabelled set allows us to obtain a final condensed set with classification performance very close to the (optimally) edited one for large TS sizes.

As a conclusion, the experimentation carried out suggest that adaptive methods for PS are able to improve the results from consistency-based ones. When appropriately initialised, LVQ-like methods can outperform the MEC approach for the small sample case. This fact could be exploited in a number of ways. Further work could be directed to designing alternative Adaptive PS methods directly benefiting from the properties of edited sets.

(Received December 18, 1997.)

REFERENCES

[1] P. A. Devijver and J. Kittler: Pattern Recognition. A Statistical Approach. Prentice Hall, Englewood Cliffs, NJ 1982.
[2] F. Ferri and E. Vidal: Colour image segmentation and labelling through multiedit-condensing. Pattern Recognition Lett. *13* (1992), 8, 561–568.
[3] F. Ferri and E. Vidal: Small sample size effects in the use of editing techniques. In: Proc. 11th Internat. Conf. on Pattern Recognition 1992, pp. 607–610,
[4] G. W. Gates: The reduced nearest neighbor rule. IEEE Trans. Inform. Theory *18* (1972), 431–433.
[5] S. Geva and J. Sitte: Adaptive nearest neighbour pattern classification. IEEE Trans. Neural Networks *2* (1991), 2, 318–322.
[6] P. E. Hart: The condensed nearest neighbor rule. IEEE Trans. Inform. Theory *14* (1968), 515-516.
[7] T. Kohonen: Self-Organization and Associative Memory. Second edition. 1988.
[8] T. Kohonen, J. Kangas, J. Laaksonen and K. Torkkola: Lvq_pak: The Learning Vector Quantization Program Package. Technical Report, Helsinki Univ. of Tech., 1992.
[9] M. A. Kraaijveld and R. P. W. Duin: On backpropagation learning of edited data sets. In: Proc. of the Int. Neural Network Conf., 1990, pp. 741–744.
[10] A. E. Lucas and J. Kittler: A comparative study of the kohonen and multiedit neural net learning algorithms. In: 1st IEE Int. Conf. Artificial Neural Networks, 1991, pp. 7–11.
[11] D. L. Wilson: Asymptotic properties of nearest neighbor rules using edited data. IEEE Trans. Systems Man Cybernet. *2* (1972), 3, 408–421.

Dr. Francesc J. Ferri, Dept. Informàtica i Electrònica (Facultat de Física) Universitat de València. Dr Moliner, 50 46100 Burjassot (València). Spain.
e-mail: ferri@uv.es