Using Proximity and Spatial Homogeneity in Neighbourhood-Based Classifiers *

J.S. Sánchez[1], F. Pla[1] and F.J. Ferri[2]

[1]Departament d'Informàtica. Universitat Jaume I
Campus Penyeta Roja, E-12071 Castelló. SPAIN.
{sanchez,pla}@inf.uji.es, +34 64 345676
[2]Departament d'Informàtica i Electrònica. Universitat de València
Dr. Moliner 50, E-46100 Burjassot (València). SPAIN.
ferri@uv.es, +34 6 3864768

Abstract

In this paper, a set of neighbourhood-based classifiers are jointly used in order to select a more reliable neighbourhood of a given sample and take an appropriate decision about its class membership. The approaches introduced here make use of two concepts: proximity and symmetric placement of the samples.

1 Introduction

The ultimate aim of any pattern recognition system is to achieve the best possible classification accuracy for the problem to be solved. In practical applications, due to the availability of a wide variety of classifiers, it is logical to test only a number of them and choose the algorithm with highest reliability as a final solution to the problem. Nevertheless, although one of the classifiers would yield the best performance, the other tested schemes can contribute complementary information about the samples to be classified and therefore, it seems obvious that those could be efficiently used to improve the performance of the selected classifier.

A traditional strategy to obtain an increase in performance consists of using some form of voting with the results of several single classifiers in order to arrive at a combined decision [4] [5]. In addition to simple majority vote (in which all votes have equal weight), the votes can be weighted for each class or even, according to the performance of each classifier on each class [8]. A common theoretical framework for combining classifiers which use distinct pattern representations is given in [7].

In a similar way, it is also possible to improve the classification rate by means of some kind of cooperation between different classifiers. In this case, there is not a combination of the decisions of various classifiers but, on the contrary, all the information given by a set of classifiers is used to arrive at a unique decision. In this paper, several "cooperating" neighbourhood-based classifiers are introduced and implemented in order to apply them to some real problems. The single classifiers used to obtain the proposed schemes are the well-known *k-Nearest Neighbours* decision

* This work was partially supported by grants P1A94-23, P1B96-13 (Fundació Caixa-Castelló-Bancaixa), GV2110/94 (Conselleria d'Educació i Ciència), AGF95-0712-C03-01 and TIC95-676-C02-01 (Spanish CICYT).

rule [3] (or k-NN, in which the k closest neighbours vote for the label of the sample to classify) as well as two novel non-parametric approaches [10], namely the *k-Nearest Centroid Neighbours* (k-NCN) and the *Graph Neighbours* (GN) rules, whose general idea consists of classifying a sample by the prototypes placed around it (instead of close to it). The reason to use these neighbourhood-based classifiers is that, while the k-NN provides a suitable notion of the statistical distribution in the proximity of a sample, the k-NCN and GN rules mainly take into account geometric information, such as the spatially homogeneous placement of prototypes around a sample.

2 Neighbourhood-based classifiers

Classification in pattern recognition has traditionally been tackled through two alternative approaches; namely, *parametric* and *non-parametric* methods [3]. The parametric classifiers assume a functional distribution of given samples. On the other hand, the non-parametric do not assume any functional distribution of the set of prototypes. While the parametric approach has theoretically been shown to be potentially capable of yielding optimal results, in practice, it often tends to actually fail because of inappropriate assumptions of a priori distributions.

Among non-parametric methods, those which are based on sample-to-sample distances are particularly remarkable; namely, k-NN techniques. When applied to classification, these schemes require the classes to be represented by appropriate sets of prototypes and the decision rule is generally reduced to label each given sample with the class that contains most of its k nearest neighbours. It is the conceptual simplicity of such a rule, along with its asymptotical tendency towards the *Bayes* rule in terms of minimum classification error, what makes the k-NN approach particularly appealing in many practical situations. Nevertheless, when the number of prototypes in the training set is not large enough, the k-NN rule is no longer optimal. This problem becomes more relevant when having few prototypes compared to intrinsic dimensionality of the feature space, which is a very common practical situation.

Some alternative definitions of neighbourhood have been used to obtain other non-parametric classifiers, trying to partially overcome the practical drawback pointed out above for the k-NN rule. In particular, the recently introduced concept of *Nearest Centroid Neighbourhood* [2] along with the neighbourhood relation derived from the *Gabriel Graph* (GG) and the *Relative Neighbourhood Graph* (RNG) [6], have been used to obtain the so-called k-NCN and GN rules [10], respectively.

As mentioned before, the k-NN rule consists of estimating the class of a given sample through its k closest prototypes. In other words, this classifier considers that all the information required to classify a new sample can be obtained from a small subset of prototypes close to the sample. However, it does not take into account the geometrical distribution of those k prototypes with respect to the given sample (that is, in general the nearest prototypes do not completely surround the sample since the k-NN rule considers the neighbourhood only in terms of Euclidean distance).

The non-parametric k-NCN and GN approaches are also based on the general idea of estimating the class of a sample from its neighbours, but considering a different kind of neighbourhood which allows to inspect a sufficiently small area *around* the

sample, in such a way that all prototypes surrounding that sample take part in the classification. As already pointed out, this is accomplished by using two different concepts about surrounding a sample with nearby prototypes: firstly, the Nearest Centroid Neighbourhood [2], which tries to surround a sample by taking prototypes in such a way that *(a)* they are as near as possible to the sample, and *(b)* their centroid is also as close as possible to the sample. Secondly, the Graph Neighbourhood (i.e., GG and RNG-based neighbourhoods) of a sample, defined as the union of all its graph neighbours. Taking into account that two points are graph neighbours if no other point lies inside a certain region of influence between them [6], it is possible to surround completely a sample by means of its graph neighbours.

Bearing this in mind, the Nearest Centroid Neighbourhood and Graph Neighbourhood concepts can be used to obtain two alternative non-parametric classifiers; namely, the k-NCN and GN rules, respectively. Both of them have in common the fact of considering a number of prototypes *around* (instead of *close to*) a sample to estimate its class. Given a set of labelled prototypes $X = \{x_1, ..., x_n\}$ and a new unknown sample q, the k-NCN and GN classification rules assign to q the class with majority of votes among its k nearest centroid neighbours or its graph neighbours, respectively. In the case of the GN rule, considering the GG and the RNG, two different approaches are defined: the *Gabriel Graph Neighbours* (GGN) and the *Relative Neighbourhood Graph Neighbours* (RNGN) rules, respectively.

Note that the surrounding decision rules (i.e., the k-NCN and the GN rules) can take prototypes which are not sufficiently close to the given sample. Obviously, this constitutes a relative drawback for those rules with respect to the k-NN classifier in some practical problems. Therefore, it could be interesting to look for a certain balance between both categories of non-parametric classifiers, trying to overcome as far as possible some of the mentioned disadvantages of each one.

3 Proposed classification schemes

From the discussion made in Section 2, the neighbourhood-based classifiers considered here can be divided into two groups according to the information used by each one: the *proximity-based* classifiers (the k-NN rule) and the *surrounding* ones (the k-NCN and GN rules). Since our general purpose is to obtain alternative decision rules with the advantages of both groups, it is possible to relate their information in some effective way to derive a single decision, which should presumably improve the corresponding performance. Here, we propose a way to obtain a classification scheme from the cooperating work between the proximity-based classifiers and the surrounding ones. In a few words, it consists of using all the information given by prototypes which belong to a relatively small neighbouring region around a sample.

Let $X = \{x_1, ..., x_n\}$ be a set of n labelled prototypes and let q be a new sample to classify. The set of surrounding neighbours of q (that is, the nearest centroid neighbours or the graph neighbours) can be represented as $SN_q = \{s_1, ..., s_m\}$, $m < n$. Now, let $k = |SN_q|$ in such a way that the set $k\text{-}NN_q = \{p_1, ..., p_k\}$ constitutes the set of its k-proximity-based neighbours (or simply, k-nearest neighbours). Thus, we can define the set of the *nearest surrounding neighbours* of q (NSN_q) as the subset of

surrounding neighbours which are also nearest neighbours. Consequently, this subset can be represented as the intersection of both sets of surrounding and nearest neighbours of q:

$$NSN_q = SN_q \cap k\text{-}NN_q$$

Obviously, the set of the nearest surrounding neighbours, NSN_q, contains the prototypes of X which satisfies both conditions pointed out before: they must be around the sample q (*surrounding neighbourhood*) and, they must be sufficiently close to it (*nearest neighbourhood*). In a first step, the surrounding classifier (i.e., the k-NCN or the GN) takes prototypes around q and in a further step, the k-NN rule constrains that initial set of neighbours (the set of surrounding neighbours SN_q) to only those prototypes which are really close to q. In such a way, we are trying to correct the fact that some surrounding neighbours can be too far from the sample q. There clearly exists a relationship among the size of those sets of neighbours of q:

$$|NSN_q| \leq |k\text{-}NN_q| = |SN_q| < |X|$$

The geometrical meaning of the set NSN_q is as follows: the prototypes of NSN_q constitute a small region (namely, the *nearest surrounding region*) close to the sample q in such a way that all prototypes are distributed around q. It is worth noting that this region does not contain the surrounding neighbours which are too far from q since they are eliminated during the *nearest neighbourhood* test. Formally, the general decision rule can be formulated in the following way:

Step 1: Find the surrounding neighbours of q: $SN_q = \{s_1, ..., s_m\}$, $m < n$.
Step 2: Let $k = SN_q$.
Step 3: Find the k nearest neighbours of q: $k\text{-}NN_q = \{p_1, ..., p_k\}$.
Step 4: Compute the nearest surrounding neighbours of q:

$$NSN_q = SN_q \cap k\text{-}NN_q, \quad j = |NSN_q| \leq k$$

Step 5: Assign to q the class with majority of votes among its j nearest surrounding neighbours in the set NSN_q (resolve ties randomly).

Note that this scheme can lead to three alternative but analogous classification rules, simply varying the surrounding approach used in Step 1 of the algorithm: we can apply either the k-NCN rule or the GN rule and, in the later case, we can choose between the GG or the RNG. A priori, almost nothing can be said about the relative merits of each one of these approaches in terms of classification accuracy. However, it has a practical interest in evaluating other issues related to the surrounding rules: first, according to [11], the averaged computational cost to search for the graph neighbours (i.e., Gabriel neighbours or relative neighbours) of a sample in a d-dimensional feature space is close to $O(dn)$, while k nearest centroid neighbours can be found in $O(kn)$ time [2]. Second, while the GN rules do not require any tuning or external parameter, the k-NCN rule needs the number of neighbours k.

4 Description of the databases

Three different real databases are used to investigate the performance of those classification schemes. For the first and third databases, the *Holdout* method averaged over five different random partitions (half for training and half for testing purposes) of each original set, has been used to obtain the results reported. For the second database [9], only one unbalanced partition of the initial set has been considered to compare the performance achieved here to that obtained in previous related works.

The first database

The aim of the experiments carried out over this database is to distinguish among 11 different textures, each pattern (pixel) being characterised by 40 attributes built by the estimation of fourth order modified moments in four different orientations: 0, 45, 90 and 135 degrees. There are a total number of 5,500 patterns.

The second database

This database was already considered in [9] in order to study colour segmentation to locate oranges in outdoor scenes under daylight conditions. It consists of values from RGB colour images with a resolution of 256 x 256 pixels. There are 19,164 patterns (6,386 samples for training and 12,778 for testing) with two attributes (coordinates φ and θ of the colour vectors in the RGB space) and three classes.

The third database

This database was used in [1] about the development of a real time analytical system for French and Spanish speech recognition. There are two different classes: the nasal vowels and the oral vowels. It contains samples coming from 1,809 different isolated syllables. Five different attributes were chosen to characterise each vowel: the amplitudes of the five first harmonics, normalised by the total energy. There are 5,404 patterns: 3,818 for nasal vowels and 1,586 for oral ones.

5 Experimental results

Some experiments by using four single neighbourhood-based classifiers (k-NN, k-NCN, GGN, and RNGN) and the three schemes proposed here (k-NN with k-NCN, k-NN with GGN, and k-NN with RNGN) have been carried out. For the k-NN and k-NCN rules, typical values for the parameter k (ranging from 1 to 11) have been tried in each experiment.

Table 1 shows the accuracy levels obtained by the four single classifiers. Note that the results for the first and third databases correspond to the averaged classification rates and the standard deviations over the five random partitions of the original data

sets. These results should be taken as illustrative in the sense that they will be used to compare the single classifiers to the schemes proposed here.

	Database 1	Database 2	Database 3
1-NN (1-NCN)	98.38 (± 0.04)	97.56	88.41 (± 0.21)
3-NN	98.03 (± 0.18)	97.61	87.41 (± 0.10)
5-NN	97.75 (± 0.14)	97.61	86.33 (± 0.12)
7-NN	97.65 (± 0.18)	97.72	86.12 (± 0.12)
9-NN	97.44 (± 0.16)	97.60	85.87 (± 0.06)
11-NN	97.33 (± 0.17)	97.46	85.13 (± 0.19)
3-NCN	98.14 (± 0.05)	97.64	88.11 (± 0.09)
5-NCN	98.20 (± 0.11)	97.75	86.97 (± 0.13)
7-NCN	98.48 (± 0.08)	97.71	87.02 (± 0.09)
9-NCN	98.58 (± 0.04)	97.60	87.00 (± 0.18)
11-NCN	98.73 (± 0.07)	97.68	86.34 (± 0.15)
GGN	80.01 (± 0.08)	97.68	84.85 (± 0.20)
RNGN	97.90 (± 0.14)	98.08	88.03 (± 0.12)

Table 1. Accuracy levels by using single neighbourhood-based classifiers

It is also important to point out the large difference in performance between the GGN rule versus the other ones for the first and third databases. This is in contrast to the RNGN (the other GN classifier), which obtains rates comparable to the k-NN and k-NCN rules, specially in the case of the last two databases. The "surprising" differences between the GGN and the RNGN rules can be interpreted as follows: the number of Gabriel neighbours for a given sample is much bigger than the number of its relative neighbours (i.e., the neighbours in the corresponding RNG), so the GGN is inspecting into a too large region while in practice, the neighbourhood of a sample should be relatively small. This can be assessed in Table 2 where the averaged number of graph neighbours as well as the percentage of prototypes in the test set with respect to the neighbourhood size for each database are reported. Note that the cases for which the GGN obtains a sufficiently high performance correspond to the second and third databases, where the percentage is smaller than 1.

	Database 1		Database 2		Database 3	
	Size	%	Size	%	Size	%
GGN	33.30	1.21	77.72	0.60	10.21	0.37
RNGN	3.35	0.12	50.23	0.39	2.87	0.11

Table 2. Neighbourhood size for the GGN and RNGN rules

The classification accuracy and the standard deviations achieved with the schemes proposed are summarised in Table 3. As expected, the approaches introduced here generally present a certain improvement with respect to the single classifiers. This confirms that the use of the information given by both groups of neighbourhood-based classifiers really manages to bound better the more reliable neighbourhood of a sample (i.e., the decision boundaries among classes).

	Database 1	Database 2	Database 3
1-NCN/1-NN	98.38 (± 0.04)	97.56	88.41 (± 0.21)
3-NCN/3-NN	98.34 (± 0.03)	97.54	88.15 (± 0.27)
5-NCN/5-NN	98.25 (± 0.08)	97.44	88.29 (± 0.30)
7-NCN/7-NN	98.02 (± 0.10)	97.39	87.82 (± 0.29)
9-NCN/9-NN	98.00 (± 0.16)	97.58	87.65 (± 0.13)
11-NCN/11-NN	98.03 (± 0.23)	97.57	87.44 (± 0.21)
GGN/*k*-NN	92.36 (± 0.06)	97.80	87.21 (± 0.28)
RNGN/*k*-NN	98.57 (± 0.06)	98.09	88.70 (± 0.21)

Table 3. Percentage of test cases correctly classified by the proposed schemes

Note that the GGN/*k*-NN rule achieves the highest improvement rates on the first database, in fact for the one in which the single application of the GGN classifier obtains the worst results. In this case, it seems that the *k*-NN rule adjusts as far as possible the disproportionate neighbourhood given by the GGN classifier. On the other hand, it is clear that the cooperation between the GGN and the *k*-NN rules achieves the highest increase in performance in all experiments.

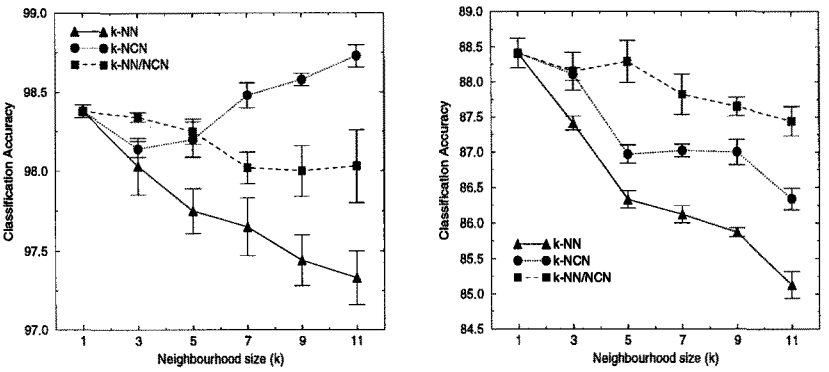

Fig. 1. Classification accuracy for the first and third databases

Fig. 1 shows the performance of the *k*-NN and *k*-NCN classifiers, as well as the one achieved by the cooperation between them on the first and third databases. As can be seen, the combined application of those schemes achieves more significant differences when the single *k*-NCN rule performs worse (that is, from $k = 5$). In fact, these results are also consistent with the discussion made for the GG-based classifier.

6 Concluding remarks

A way of cooperation between different neighbourhood-based classification schemes has been introduced in order to increase the performance achieved by using single classifiers. The general idea of this approach consists of using two complementary neighbourhood definitions: the *nearest* neighbourhood and the *surrounding*

neighbourhood. The former is represented by the well-known k-NN decision rule [3] and, the latter by means of the recently introduced k-NCN [10] and graph-based classifiers [10]. The aim of using jointly the information obtained from different neighbourhood-based classifiers is to benefit from the advantages of both groups of neighbourhood defined in this paper. On the one hand, the k-NN rule gives a suitable measure of distance-based proximity with respect to a sample. On the other hand, the surrounding approaches conveniently define the geometrical distribution of prototypes around the given sample.

From the experimental results, it is possible to draw some conclusions. First, the schemes proposed really achieve a relative improvement in terms of classification accuracy with respect to the single neighbourhood-based rules. Second, it seems that the combination of the k-NN and the GGN classifiers is not appropriate since it systematically achieves the lowest performance in all experiments. Finally, with regard to the jointly use of the information obtained from the k-NN and the k-NCN and RNG-based rules, differences between them are not significant enough and, it is not possible to make a final statement about the suitability of either of them. In this case, one should weigh other related issues such as the computing time.

References

1. Alinat, P. (1993). Periodic Progress Report 4: ROARS Project ESPRIT II 5516.
2. Chaudhuri, B.B. (1996). A new definition of neighbourhood of a point in multi-dimensional space, *Pattern Recognition Letters* **17,** 11-17.
3. Duda, R. and P.E. Hart (1973). *Pattern classification and scene analysis.* John Wiley & Sons, New York.
4. Franke, J. and E. Mandler (1992). A comparison of two approaches for combining the votes of cooperating classifiers, *Proc. of 11th. International Conference on Pattern Recognition*, 611-614.
5. Ho, T.K., J.J. Hull and S.N. Srihari (1994). Decision combination in multiple classifier systems, *IEEE Trans. on Pattern Analysis and Machine Intelligence* **16**, 66-75.
6. Jaromczyk, J.W. and G.T. Toussaint (1992). Relative neighbourhood graphs and their relatives, *Proc. of IEEE* **80**, 1502-1517.
7. Kittler, J., M. Hatef and R.P.W. Duin (1996). Combining classifiers, *Proc. of 13th. International Conference on Pattern Recognition*, 897-901.
8. Lam, L. and C.Y. Suen (1995). Optimal combinations of pattern classifiers, *Pattern Recognition Letters* **16**, 945-954.
9. Pla, F., F. Juste, F.J. Ferri and M. Vicens (1993). Colour segmentation based on a light reflection model to locate citrus fruits for robotics harvesting, *Computers and Electronics in Agriculture* **9**, 53-70.
10. Sánchez, J.S., F. Pla and F.J. Ferri (1997). On the use of neighbourhood-based non-parametric classifiers, *Proc. of Pattern Recognition in Practice V* (In press).
11. Toussaint, G.T., B.K. Bhattacharya and R.S. Poulsen (1985). The application of Voronoi diagrams to nonparametric decision rules, *Computer Science and Statistics*, 97-108.