Proc. 12 International Conference on Pattern Recognition, Jerusalem, Israel, 1994, 467-469.

A Two-Level Classification Scheme Trained by a Fuzzy Neural Network

Ludmila I. Kuncheva[1], Sushmita Mitra[2]

[1]CLBA, Bulgarian Academy of Sciences, Acad. G. Bonchev Street, Block 105
1113 Sofia, BULGARIA, e-mail: lucy@inf.nbu.bg
[2]Machine Intelligence Unit, Indian Statistical Institute, 203 B.T. Road,
Calcutta 700035, INDIA, e-mail: sushmita@isical.ernet.in

Abstract

A two-level pattern classification scheme is considered. At its first level the scheme labels the input object as "doubtful" or "certain" and at the second one applies the respective classification rule. Complicating the classifier in such a way we aim at a more accurate result than that obtained through either of the classification rules itself. A fuzzy neural network with linguistically interpretable inputs has been applied to detect the boundaries of the "doubtful" region(s) in the feature space. A fuzzy k-Nearest Neighbors rule with k=1 and k=5 has been used at the second level for the "doubtful" and "certain" regions, respectively. The idea has been demonstrated on a generated data set (two separable classes with uniform distribution). The results show the tendency of improvement of the classification accuracy.

1: Introduction

Two-level classification schemes are heuristic pattern recognition tools that are supposed to yield better classification accuracy than single-level ones at the expense of a certain complication of the classification structure [1-4].

There are different strategies to build a two-level classification scheme. The most widely used one is to include several classifiers in parallel at the first level and to aggregate their decisions at the second level. Another approach, called 'change-glasses', consists in switching to another classification rule if the current one leads to a decision with low certainty for the object being classified [2]. This strategy corresponds to the well known medical practice to summon a consultation if there is some ambiguity which the current staff has no competence to resolve. A different hypothesis underlies the two-level scheme in [4]. The feature space is preliminarily partitioned into subspaces and only one rule is authorized to make the decision for each subspace. Hopefully, each such rule is the most 'competent' one for the respective region. Here we use a combination of this last approach and the 'change-glasses' one.

The problem of partitioning the feature space into appropriate regions is considered in the literature (usually implicitly) in a different context. In fact, every rule-based classifier performs a partitioning through antecedent clauses and assigns a classification rule to each region through the implication. In the fuzzy classification rule described in [5,6] the partitioning is uniform, i.e. the regions continue to be split until a sufficiently high certainty of the rule, generated by each region, is achieved. In this way, the decision boundary is approximated as precisely, as necessary. The problem is how to guarantee that the generalization capability of the classifier is sufficiently high if the regions contain only few objects. Trying to prevent this case, we suggest to use only two regions designated as 'doubtful' and 'certain', respectively. Note that each of these may not necessarily be compact and may consist of more than one disjoint subregions. These regions are then treated by different classification rules.

For the purposes of partitioning we used a previously developed technology based on a neural network with fuzzy inputs [7,8]. The details are summarized in Section 2. Some considerations about the choice of the classification rules for both regions are presented in Section 3. Section 4 contains an experimental illustration.

2: A fuzzy neural network for partitioning the feature space

Let $\Omega = \{\omega_1 ,..., \omega_M\}$ be the set of classes, $\mathbf{X} = \{X_1 ,..., X_n\}$ be the set of features and $x = [x_1 ,...,x_n]^T$ be the vector representing an object in the feature space. The fuzzy multi-layer perceptron (MLP) partitions the feature space in terms of the linguistic properties low, medium and high, represented as π-functions. The centres and radii of these three π-functions along each feature axis are determined automatically from the distribution of the training patterns [8]. An n-dimensional input feature space is therefore partitioned into 3^n overlapping regions. In this work we evaluate, in turn, the classification accuracy of the network for pattern points belonging to each of these 3^n partitions. The output class labels of the training set are used to determine the particular region (expressed in terms of center c and radius r of the π-function) whose pattern points yield poor classificatory performance. This is designated to be the 'doubtful' region. Note that a threshold could be established to generate one or more such 'doubtful' regions. Then the boundaries of the doubtful region are computed as $[c - r/2,\ c + r/2]$ along each feature axis. The remaining portion of the feature space constitutes the 'certain' region.

3: Choice of the classification rules for the two regions

The choice of the classification rules for the two regions is the next question which has to be investigated. Due to its robustness, feasibility of implementation, and good practical results, we decided to apply variants of the k-Nearest Neighbors (k-NN) rule for both regions but with different parameters.

Some heuristic considerations can be used to guide the choice. From statistical pattern recognition theory it is known that for the asymptotic case with $k \to \propto$ and $N \to \propto$ the probability of error tends to the optimal (Bayesian) one. For a finite sample, however, there should be some reasonable ratio k/N. Here we rely on the hypothesis that the class structures for the certain regions are more or less **compact**, so that the network has generated rules with high certainty. Therefore, we could achieve better results if we use k-NN with k>1 as compared to the 1-NN rule. On the contrary, for the doubtful region, we suppose that the decision boundary is eventually more complex and the interconnection between the objects in the same class is not high. Therefore the 1-NN rule would be more appropriate. Furthermore, it is reported in different papers that distance-based k-NN rules [9] can outperform the crisp k-NN for a finite data set.

Then the two-level classifier operates as follows. The object is analyzed in order to detect whether it belongs to the doubtful region or not. The second level turns on the one of the rules depending on the first-level decision.

4: Data sets, statement of the experiments, and results

Two data sets have been generated to illustrate the two-level scheme. Two classes, ω_1 and ω_2, are considered, both from the same uniform distribution. Two numerical features are used so that the data is in the region [0,1]x[0,1] and can be easily visualized. The decision boundary is

$$f = - 0.25 \sin (7\pi x_1^3) + x_2 - 0.5$$

Note that, since the boundary is fixed, the theoretical probability for correct classification for the asymptotic case is 1.

The first set consists of 200 cases (Fig. 1) and is used as the training set, while the second one contains 1000 more cases and is meant to be the test set. The second set has been generated in order to avoid an optimistic bias in assessing the classification accuracy.

Crisp and fuzzy k-NN rules are performed for the training set, and the classification accuracy is assessed both on the training and the test sets. The leave-one-out method was used for the training set. The results for k =1,3,5, and 7are presented in Table 1.

For a comparison, the linear discriminant analysis has been performed on the training set and its classification accuracy assessed both on the training and the test sets. For the training set, the resubstitution method for assessing the classification accuracy yields 81.1 % for ω_1, 85.1 % for ω_2, and 83.0 % for both. For the test sample, the results are 88.0 %, 81.6 %, and 85.4 %, respectively.

Then the neural network described in Section 2 is used to separate the doubtful and certain regions. The partition high, medium (along the first and second feature axes respectively) is found to constitute the doubtful region. The centres and radii along the two axes are determined to be $c_1 = 0.754$, $r_1 = 0.471$ and $c_2 = 0.525$, $r_2 = 0.461$ respectively, from the training set. The detected doubtful region is depicted in Fig 1. as a dashed-line rectangle.

Further experiments have been performed separately on the two sets, subject to first testing whether each object falls into the doubtful region or not. We apply 1-NN for the doubtful region, and 5-NN for the undoubtful one. The result for the doubtful region is **80.0** % correctly classified objects (196 cases out of 245) and **97.9** % (739 cases out of 755) for the certain region. The overall classification accuracy is **93.5 %**.

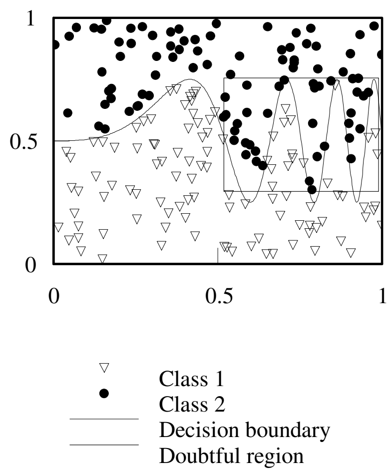

Fig. 1. The training set (200 cases)
and the detected "doubtful" region.

Table 1. Classification accuracy with the crisp and fuzzy k-NN [%] on the training and the test sets

k	Crisp k-NN		Fuzzy k-NN	
	Training set	Test set	Training set	Test set
1	93.5	92.8	93.5	92.8
3	90.5	93.0	92.5	92.9
5	91.0	92.5	93.0	92.9
7	92.0	90.6	93.0	92.4

5: Discussion and conclusions

Considering the whole sample, it appeared that the results from discriminant analysis were worse than those obtained through k-NN. This fact has been expected because the classes are neither Gaussian, nor are they linearly separable. The k-NN rule (in its pure version, or with distance-based modifications), being a robust technique, is recommended in the two-level classification scheme under consideration.

It can be seen from Fig. 1. that the region, detected as 'doubtful' through the fuzzy neural network is really the region with the most complex classification boundary.

Furthermore, according to the results from the last experiments, it appeared appropriate to apply different rules to the objects from the different regions. The accuracy, as assessed on the test set, was **93.5 %** versus **92.8 %** using the plain 1-NN. It should be mentioned that a choice k = 3, or k = 7 is also successful because the overall accuracy is 93.0 % and 93.2 %, respectively. Although the improvement in this example is not high, the tendency can be observed.

6: Acknowledgments

This work was partially supported by the award No. is 572/813-014-3 under German Academic Exchange Service (DAAD) Fellowship, and partially by the contract No ═╚-╠Ë-2-╚═/93 with the Young Scientists Foundation.

7: References

1. L. Xu, A. Krzyzak and C. Y. Suen, Methods of combining multiple classifiers and their application to handwriting recognition, *IEEE Transactions on Systems, Man, and Cybernetics* **22** (1992) 418-435.
2. L. I. Kuncheva, 'Change-glasses' approach in pattern recognition, *Pattern Recognition Letters* **14** (1993) 619-623.
3. Yu. L. Barabash, *Collective Statistical Decisions in Recognition* (Moscow, Radio i Sviaz, 1983) (In Russian).
4. L.A. Rastrigin, R.H. Erenstein, *Method of Collective Recognition* (Moscow, Energoizdat, 1981) (In Russian).
5. H. Ishibuchi H., K. Nozaki, H. Tanaka, Distributed representation of fuzzy rules and its application to pattern classification, *Fuzzy Sets and Systems* **52** (1992) 21-32.
6. H. Ishibuchi, K. Nozaki, H. Tanaka, Efficient fuzzy partition of pattern space for classification problems, *Proc. 2nd International Conference on Fuzzy Logic & Neural Networks, Iizuka, Japan* (1992) 671-674.
7. S. Mitra, S. K. Pal, Fuzzy multi-layer perceptron, inferencing and rule generation, *IEEE Transactions on Neural Networks* (in press).
8. S. Mitra, Fuzzy MLP based expert system for medical diagnosis, *Fuzzy Sets and Systems* (accepted).
9. G. Parthasarathy, B.N. Chatterji, A class of k-NN methods for low sample problems, *IEEE Transactions on Systems, Man, and Cybernetics* **20** (1990) 715-718.