Proc FUBEST, 1994, Sofia, Bulgaria, pp 29-32

CHANGE-GLASSES PATTERN CLASSIFICATION WITH A FUZZY NEURAL NETWORK

S. Mitra
Machine Intelligence Unit, Indian Statistical Institute
203 B.T. Road, Calcutta 700035, INDIA
e-mail: sushmita@isical.ernet.in

L. Kuncheva
CLBA, Bulgarian Academy of Sciences
Acad. G. Bonchev Street, Block 105 1113 Sofia, BULGARIA
☎: +359 (2) 713 36 03; Fax: + 359 (2) 72 37 87; e-mail: lucy@inf.nbu.bg

Keywords: Fuzzy pattern recognition (classification), fuzzy neural networks, "change-glasses" approach, partitioning of the feature space

SUMMARY

"Change-glasses" approach in pattern recognition consists in switching between different classification rules depending on their certainty in the current classification decision. The paper presents the results from an experimental study with a set of generated data. The classification rules are based on a fuzzy neural network with different "resolution" levels.

INTRODUCTION

"Change-glasses" approach in pattern recognition [1] relies on the assumption that there are subspaces of the initial feature space where the chosen classification rule can be replaced by another, more competent, one (see, e.g. [2]). This will hopefully lead to a better classification accuracy in comparison with that obtained through one rule only. This classification strategy bears an analogy with the process of working up a diagnosis in medicine. If the physician does not feel competent to resolve a special case, he summons a consultation team of professionals in that particular field.

The main question that arises here concerns the way of partitioning the feature space. There are different approaches underlying some heuristic classification paradigms. In fact, every rule-based classifier performs a partitioning through antecedent clauses and assigns a classification rule to each region through the implication. A partition may be based on the geometric properties of the classes detected by a preliminary clustering [3] or by sequential groping about for the class boundaries [4]. 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 'undoubtful', 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. Section 3 contains an experimental illustration.

THE FUZZY NEURAL NETWORK

Let $\Omega = \{\omega_1,..., \omega_M\}$ be the set of classes, $\boldsymbol{X} = \{X_1,..., X_n\}$ be the set of features and $x = [x_1,...,x_n]^T$ be the vector representing an object in the feature space. Let $\mathbf{Z} = \{Z_1,..., Z_N\}$ be the available set of labeled objects.

Classical multi-layered perceptron (MLP) feed-forward structure is considered. The output of a neuron in any layer other than the input one is given as

$$y_j^{h+1} = \frac{1}{1+\exp(-\sum_i y_i^h w_{ji}^h)}$$

where y_i^h is the state of the i-th neuron in the preceding h-th layer and w_{ji}^h is the weight of the connection from the i-th neuron in layer h to the j-th neuron in layer $h+1$. Each input node corresponds to a component of the input feature vector. The outputs of these nodes directly yield the respective value of the input. Backpropagation algorithm is used for training.

The fuzzy multi-layer perceptron partitions the feature space in terms of the linguistic properties low, medium and high, represented as π-functions. An n-dimensional pattern $x = [x_1,...,x_n]^T$ is represented as a $3n$-dimensional vector

$$F = [\mu_{1,low}(x_1), \mu_{1,medium}(x_1), \mu_{1,high}(x_1), ..., \mu_{n,low}(x_n), \mu_{n,medium}(x_n), \mu_{n,high}(x_n)]^T$$

where the μ values stand for membership functions of the corresponding linguistic π-functions along each feature axis. The centers and radii of these three π-functions along each feature axis are determined automatically from the distribution of the training patterns [7].

In order to apply the "change-glasses" approach, we evaluate, in turn, the classification accuracy of the network for pattern points belonging to each of these 3^n regions. 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 where the classification rule is to be changed. 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 'undoubtful' region.

For the doubtful region, further partitioning is performed. An effective partitioning of the feature space refers to generating neither too many nor too few partitions along the different feature axes. The distribution of the patterns in the input space is likely to play an important role guiding the choice of the regions to be further processed.

Here we confine our change-glasses approach to two alternatives only although it may include as many stages, as necessary. It should be kept in mind, however, that each new partitioning requires a significant amount of training and test data, and a balance is to be observed.

The main change-glasses scheme that we use in this study is shown in Fig. 1.

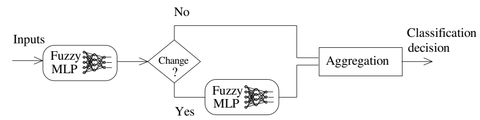

Figure 1. Change-glasses paradigm with FNN

The aggregation block serves to form the final classification decision. It can be omitted and replaced only by a switch. It can, however, be used to combine the decisions obtained from different perspectives. This corresponds to the practice that, after the consultation team has suggested a diagnosis, the physician makes the final decision by himself.

DATA SETS

Two data sets have been generated for illustration. 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. 2.) 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 any 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. 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. The results are presented in Table 1.

Table 1 contains also the results from FNN applied on the whole feature space with six different configurations: 3 layers, with 11, 12 and 13 nodes in the hidden layer, and 4 layers with 10, 11, and 12 nodes in the hidden layers. The same configurations have been applied in the change-glasses approach as described above. The partition *high, medium* (along the first and second feature axes respectively) is found to constitute the doubtful region. The centers and radii along the two

Proc FUBEST, 1994, Sofia, Bulgaria, pp 29-32

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. Therefore, the boundary of the doubtful region is evaluated as [0.5185, 0.9895] and [0.2945, 0.7555] along the two feature axes. The detected doubtful region is depicted in Fig 2. as a dashed-line rectangle.

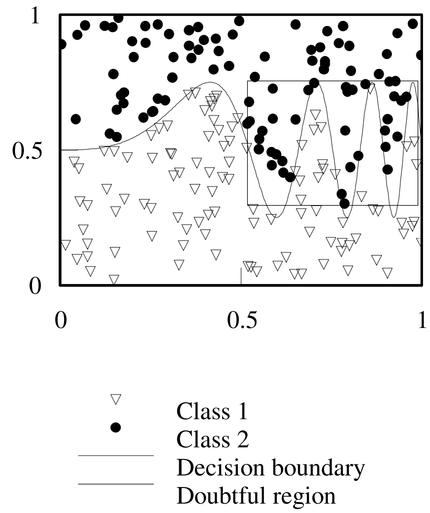

Fig. 2. The training set (200 cases) and the detected "doubtful" region

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. 2. 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 higher with the proposed approach.

Table 1. Classification accuracy [%] on the training and the test sets

Classification technique	Training set	Test set
Crisp 1-NN	93.5	92.8
Crisp 3-NN	90.5	93.0
Fuzzy 1-NN	93.5	92.8
Fuzzy 3-NN	92.5	92.9
LDA	84.0	85.4
FNN **3** layers		
11 nodes	99.0	92.1
12 nodes	93.5	92.1
13 nodes	99.5	92.8
FNN **4** layers		
10 nodes	99.0	92.0
11 nodes	93.0	92.3
12 nodes	93.0	91.5
Change-glasses approach		
FNN **3** layers		
11 nodes	99.5	92.1
12 nodes	99.5	92.3
13 nodes	98.5	**93.2**
FNN **4** layers		
10 nodes	98.0	93.0
11 nodes	99.5	92.5
12 nodes	95.0	92.2

It is worth to be mentioned here that the classification rules that can be used in this approach are not confined to the proposed FNN. It may appear that another combination of classification rules is more suitable. For example, we could use FNN to determine the doubtful region, and to apply a classical technique (e.g. k-NN). In this way we can avoid training of the classifier on small data sets (eventually contained in the doubtful region).

ACKNOWLEDGMENTS

This work was partially supported by the contract No ═╚-╠Ë-2-╚═/93 with the Young Scientists Foundation.

REFERENCES

1. L. I. Kuncheva, 'Change-glasses' approach in pattern recognition, *Pattern Recognition Letters* **14** (1993) 619-623.
2. L. A. Rastrigin, R.H. Erenstein, *Method of Collective Recognition* (Moscow, Energoizdat, 1981) In Russian.

3. K. Hirota, W. Pedrycz, Geometric-logical pattern classification, *Proc. 2nd Int. Conference on Fuzzy Logic & Neural Networks, Iizuka, Japan* (1992) 675-678.
4. D.P. Mandal, C.A. Murthy, S. K. Pal, Formulation of a multivalued recognition system, *IEEE Trans. on Systems, Man, and Cybernetics* **22** (1992) 607-620.
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, Fuzzy MLP based expert system for medical diagnosis, *Fuzzy Sets and Systems* (accepted).
8. S. Mitra, L. I. Kuncheva, Improving classification efficiency of fuzzy MLP using two-level selective partitioning of the feature space, *Fuzzy Sets and Systems* (submitted).