ELSEVIER

European Journal of Operational Research 80 (1995) 365–370

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

Theory and Methodology

Using degree of consensus in two-level fuzzy pattern recognition

Ludmila I. Kuncheva

Central Laboratory of Bioinstrumentation and Automation, Bulgarian Academy of Sciences, Acad. G. Bonchev Street, Block 105, 1113 Sofia, Bulgaria

Received August 1992; revised February 1993

Abstract

A two-level fuzzy pattern recognition scheme is proposed which implements different classification rules at the second level in dependence on the degree of consensus between first-level decision makers. The first-level classification decisions are treated in a fuzzy manner, i.e. each one is supposed to provide class memberships to all classes for each object submitted to classification. The main aim is to increase the classification accuracy of automatic classifiers. The presumption behind the idea is that different rules may appear suitable for the cases of low, medium, or high consensus between first-level decisions and neither of these rules works sufficiently well in all cases. Some implementation details are specified in the text but the real application of the scheme is not confined to them only. A set of experimental results with real data from aviation medicine is presented which show the anticipated increase of the classification accuracy.

Keywords: Fuzzy sets; Pattern recognition; Heuristics; Medicine; Decisions

1. Introduction

Classification is an active research issue due to its intrinsic connection with pattern recognition, decision making, knowledge-based reasoning, neural networks, etc. The use of hybrid methodologies integrating multiple paradigms to cope with the complexity of the classification task is becoming a prominent direction.

Fuzzy pattern recognition is a discipline from the scope of pattern recognition and decision making where fuzzy set theory is applied. It allows for including various heuristic techniques and suggestions in the pattern recognition paradigm. This combination of fuzzy and standard pattern recognition tools in the light of operational research is expected to lead to better results than those in either of the two fields. An especially fruitful application field for fuzzy pattern recognition is medicine since medical data, knowledge, and practice defy deterministic decision models and rules. Statistical tools also fail in handling medical data, usually because of the small sample size and the significant amount of subjective information and missing data.

The simple two-level classification paradigm mimics the group decision making where each first-level classifier corresponds to an 'expert'. The expert is asked to label a given object to one (or more) preliminary defined classes, and, in

SSDI 0377-2217(93)E0220-R

addition, to attach the respective membership grades. The aggregation rule at the second level yields the final decision of the group.

The two-level classifiers based on statistical first-level decisions [3,9] do not express any further similarity with group decision making, e.g. no feedback between 'experts', neither consensus assessment are implied. Including of one or more of these features might lead to an enhanced performance of the classifier.

A two-level fuzzy classification scheme was proposed earlier [6] which aimed at a better classification accuracy than both crisp classifier and one-level fuzzy classifiers. The precise proving of this ability is hardly possible since such a scheme is based mainly on hypotheses and assumptions. An experimental illustration was, however, supplied.

In this paper a fuzzy classification rule is proposed which is supposed to be a further step towards improvement of classification accuracy. It makes use of the degree of consensus between first-level fuzzy classifiers which is taken into account in forming the final decision. Some notions about two-level fuzzy pattern recognition are presented in Section 2. Section 3 contains some considerations on consensus assessment and the description of the new rule. An experimental example is presented in Section 4.

2. Two-level fuzzy pattern recognition

Let $\Omega = \{\omega_1, \dots, \omega_M\}$ be the set of classes (e.g. diseases), and $X = \{X_1, \dots, X_n\}$ be the set of features (e.g. signs, symptoms, laboratory test results, clinical observations, etc.) describing the objects to be classified. Let S^n denote the feature space generated by the set X. The classical pattern recognition task implies that the decision rule D performs the mapping

$$D : S^n \to \Omega.$$

Fuzzy decision rule **FD** is given by

$$\mathbf{FD} : S^n \times \Omega \to [0, 1]. \qquad (1)$$

Note that not only one class may be attached to each point in S^n but a set of classes with different degrees of membership. This degree may denote different characteristics: typicalness, strength of confirmation, intensity, degree of illness, belief of the diagnosis, and many other entities stemming either form uncertainty or from other sources.

Two-level pattern recognition implements the mapping (1) in two levels. The first one consists of L fuzzy classifiers performing the rules

$$R_k : S^n \times \Omega \to [0, 1], \quad k = 1, \dots, L.$$

They yield competitive classification decisions which are to be aggregated at the second level:

$$\mathbf{FD}(x, \omega_i) = A[R_1(x, \omega_i), \dots, R_L(x, \omega_i)], \quad x \in S^n,$$

where A is an aggregation function and $R_k(x, \omega_i)$ denotes the degree of membership of the object x to the class ω_i as assessed by the k-th first-level classifier, $k = 1, \dots, L$. At this point many aggregation rules from fuzzy decision theory may be applied.

3. Changing the rule in dependence on the degree of consensus

The underlying idea is that the degree of consensus between first-level decision makers could direct the decision rule. One rule may appear suitable in the case of high degree of consensus while another may fit the case of a lower degree. Suppose that in the former case a simple rule is applied since all classifiers agree as to the class label and the strength of membership for the object under consideration. If some doubt arises another rule seems to be most useful. If the decision makers exhibit total dissensus the reasons should be scrutinized.

The principal structure of a two-level classifier in which the final decision rule changes in dependence on the degree of consensus is shown in Fig. 1. Naturally, the presumption that a higher degree of consensus leads to a better decision does not always hold. In real expert decision procedures it may occur that the consensus reached at the first iteration is caused by an authority with

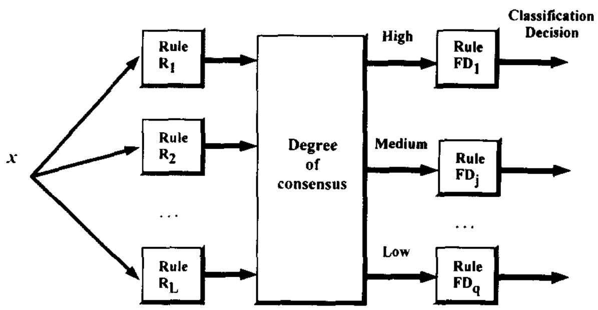

Fig. 1.

wrong opinion. The case considered here is a more favorable one since the mistakes of the decision makers are supposed to be uncorrelated.

Formally, the feature space is partitioned into subspaces in the course of the decision process. It is expected that neither of the decision rules at the second level could perform sufficiently well alone on the whole space S^n.

Bearing in mind the fact, that for each method there exists a data set for which the method yields perfect results while all other methods fail, an example is presented in Fig. 2.

The intention is to illustrate the possibility to derive benefit from the proposed approach. Obviously, the presented scatterplot of the classes is artificial and a lot of akin cases and examples could be designed. Let the k-Nearest Neighbors (k-NN) of the object to be classified serve as first-level decision makers. Each neighbor votes with degree 1 for the class where it belongs and with 0 for the others. Suppose that 3-Nearest Neighbor rule is performed for the whole space.

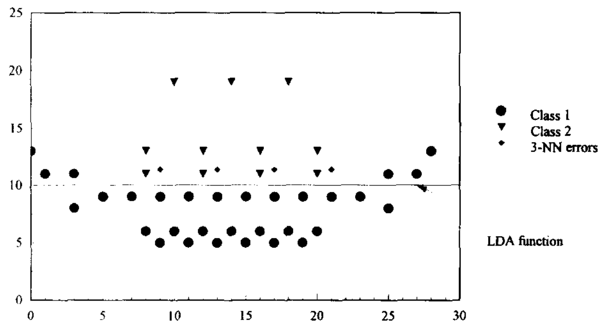

Fig. 2.

The percent of correctly classified objects will be 90.5. If apply discriminant analysis this percent will be 85.7.

Let us now perform the 3-NN rule for the objects with high degree of consensus (those for which all three neighbors vote for the same class) and discriminant rule for the rest. The result will be 100% correctly classified objects.

There are different views about consensus assessment in fuzzy decision making [1,4,10] but they are predominantly focused on preference relations on the set of alternatives. In the case under consideration the consensus is to be evaluated between L values in the interval [0, 1].

Nine measures of consensus for the case of two-level fuzzy classifier have been proposed and examined [7,8]. Here, the following ones are used:

- *Highest discrepancy*:

$$J_2 = J_{\mathrm{hd}}(x, \omega_i) = 1 - \sup_{k1,k2=1,\dots,L} |R_{k1}(x, \omega_i) - R_{k2}(x, \omega_i)| = 1 - \left[\sup_k R_k(x, \omega_i) - \inf_k R_k(x, \omega_i)\right].$$

- *Integral mean coincidence*:

$$J_3 = J_{\mathrm{imc}}(x, \omega_i) = 1 - \frac{1}{L}\sum_{k=1}^{L} |R_k(x, \omega_i) - \bar{R}(x, \omega_i)|,$$

where

$$\bar{R}(x, \omega_i) = \frac{1}{L}\sum_{k=1}^{L} R_k(x, \omega_i).$$

- *Integral pairwise coincidence*:

$$J_4 = J_{\mathrm{ipc}}(x, \omega_i) = 1 - \frac{2}{L(L-1)} \times \sum_{k1=1}^{L-1} \sum_{k2=k1+1}^{L} |R_{k1}(x, \omega_i) - R_{k2}(x, \omega_i)|.$$

- *Integral highest discrepancy*:

$$J_5 = J_{\mathrm{ihd}}(x, \omega_i) = 1 - \sup_k |R_k(x, \omega_i) - \bar{R}(x, \omega_i)|.$$

These definitions show that if perfect agreement is reached the criterion takes value one regardless of the value of the class membership. The measure J_2 is designed to find out if there are two persons with drastically different opinions. The 'integral' measures J_3, J_4 and J_5 are to assess the consensus degree in a more robust form using the average of the opinions of the first-level classifiers, while the former one J_2 is more sensitive. Similarly to the measure of highest discrepancy J_2, J_5 is sensitive to a drastically different opinion from the average one.

The classification rule may be changed once or more times in accordance with respective thresholds on the criterion value. In the real classification process for a given object all classifiers at the first level work in parallel, while only one aggregation rule at the second level turns active.

4. Experimental example

A set of real data from aviation medicine (from Kunchev, 1982) was used containing measurements of systolic blood pressure for 200 healthy male pilots in 7 time moments during examination in a hypobaric chamber (barocamera). For each individual an expert evaluation was made and a degree of hypoxic resistance between 0 and 1 was attached to him. According to these degrees two classes were formed: ω_1: pilots with low hypoxic resistance, and ω_2: pilots

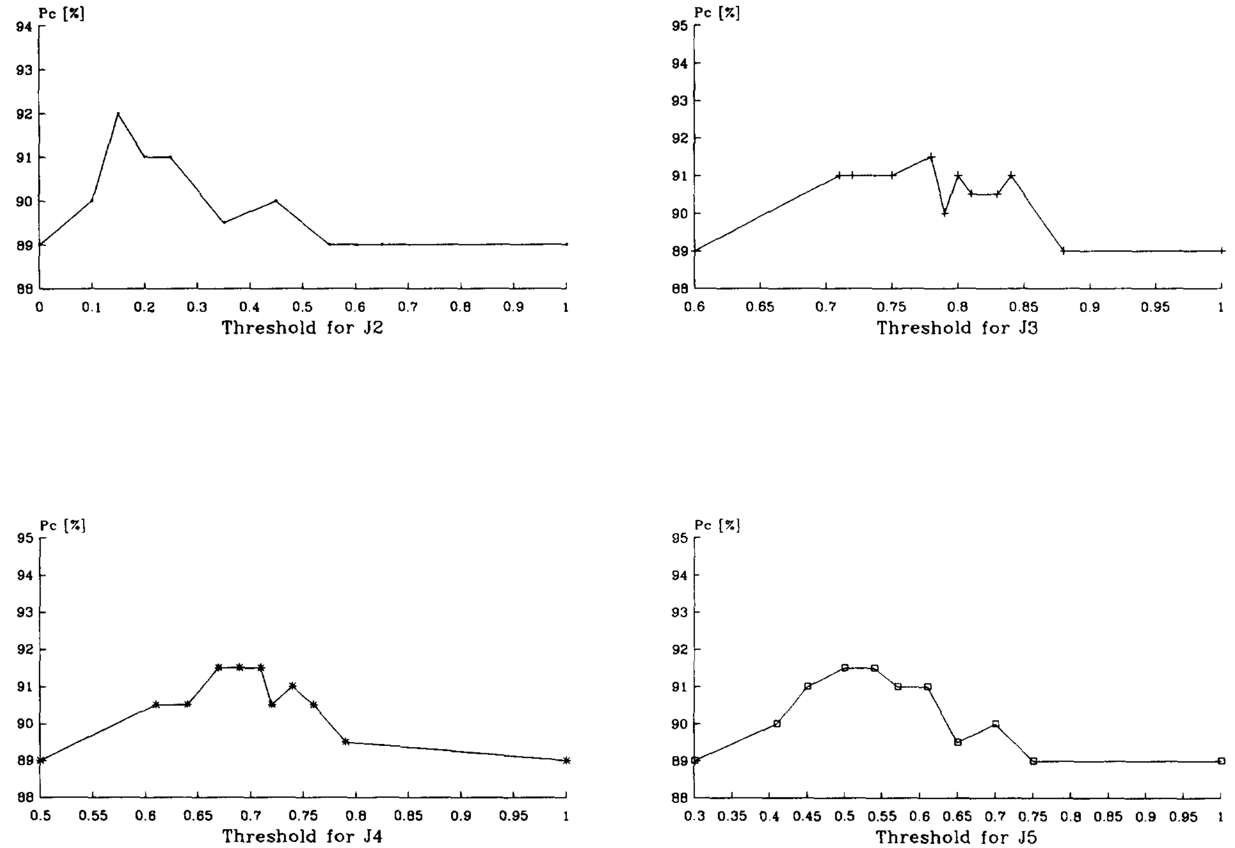

Fig. 3.

with high hypoxic resistance. The degree of hypoxic resistance was directly used as a degree of membership to the class 'low resistance'. Thus, the set $Z = \{Z_1, \dots, Z_N\}$ of preliminary labeled reference patterns was stated and two membership grades $\{\mu^1(Z_j), \mu^2(Z_j)\}$ were attached to each element Z_j where

$$\mu^2(Z_j) = 1 - \mu^1(Z_j).$$

In this example fuzzy k-Nearest Neighbors rule is used. Let $x \in S^n$ be an unknown pattern to be classified which is presented via its feature values. Let $Z^1(x), \dots, Z^k(x)$ be the k nearest neighbors of x in the set Z. For the current problem $L = k = 9$ were chosen.

For each class $\omega_i \in \Omega$ a voting is performed. In the classical fuzzy k-NN classifier [2] the score for the respective class serves as a degree of membership and is calculated by

$$\mathbf{FD}(x, \omega_i) = \mu^i(x) = \frac{1}{k} \sum_{t=1}^{k} \mu^i[Z^t(x)].$$

In the two-level scheme based on degree of consensus the rules take different forms for the respective intervals. The following variants are implemented:

$$\mathbf{FD}_j(x, \omega_i) = \mu^i(x) = \frac{1}{\| w_j \|} \sum_{t=1}^{k} w_j^t \mu^i[Z^t(x)], \tag{2}$$

where $w_j = [w_j^1, \dots, w_j^k]$ is a binary weight vector determining which neighbors are used in the formula and j varies from one to the number of consensus intervals. The difference consists in the weight vectors which can be used, i.e. in the subsets of nearest neighbors. Following the analogy with group decision making, w corresponds to the set of experts which expresses optimal performance at a given level of agreement. It may appear that a smaller group, even excluding the most competent experts, could surpass the initial group in solving nontrivial cases.

For each measure J_2–J_5 the following experiments were carried out. Nine nearest neighbors, which were meant to play the role of 'experts' of the first level, were found for each object. Their votes for the two classes were calculated through (1). The consensus between them was assessed by the respective measure. Stating a threshold for the consensus value, the sample was divided into two parts. The optimal subset of neighbors (vector w) was selected via exhaustive search on the set of subsets. The criterion for this search was an estimate of the probability of correct classification P_c using leave-one-out method. The calculation of P_c was performed in the following way.

For each vector w, subject to evaluation, membership grades $\mu^2(Z_j)$ for the respective part of the sample Z were calculated via (2). Then an optimal value (cutpoint) was selected which served as a boundary between the two classes and for which the percent (P_c) of correctly classified objects was maximal. Then the consensus threshold was changed, causing a new partition on Z, until it passed the whole range of the respective measure. The dependencies of the classification accuracy on the threshold value for J_2 to J_5 are presented in Fig. 3.

It can be seen that there exist optimal threshold values for all measures of consensus for which the classification accuracy reaches maximal values.

An attempt was made to implement the scheme with two thresholds, i.e. three decision rules of type (2) were designed for each measure by the exhaustive search described above. The resulting classification accuracy is shown in Fig. 4. Three other values are presented for comparison in Fig. 4. The application of stepwise linear discriminant analysis from BMDP software package to the

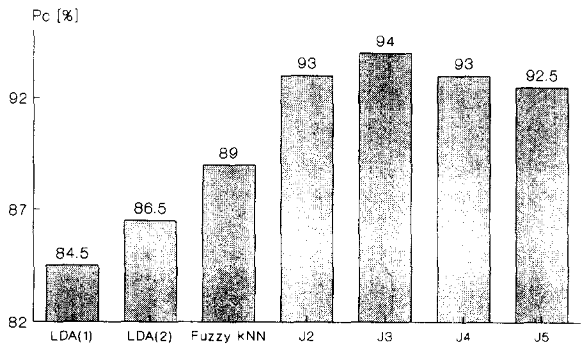

Fig. 4.

considered problem resulted in 84.5% correct classification (designation LDA(1) in Fig. 4). Then a voting two-level classification scheme was tried where classifiers were built on single features and their outputs were treated in a fuzzy way (LDA(2) in Fig. 4). The posterior probabilities for correct classification were used as membership degrees. The optimal combination of votes was detected by exhaustive search. The result for P_c was 86.5%. The third additional value in Fig. 4 expresses the maximal criterion value derived from the classical fuzzy k-NN recognition method. In the experiment k varied from 1 to 9. The best result ($P_c = 89\%$) was found for $k = 1$.

In calculation of the results with measure J_3 the rule $\mathbf{FD}_1$ (the case of lowest consensus) was replaced with the two-level linear discriminant rule (LDA(2)). For this case it appeared better than using (2) three times.

The experimental results exhibit the expected improvement of the classification accuracy. Although the difference is not great it exists in all experiments. It should be taken into account that a drastic jump in classification accuracy is hardly possible for this type of problems.

5. Conclusions

The two-level scheme based on consensus which is proposed in this paper is a pattern recognition paradigm resembling to some degree multi-tour expertise. This inclusion of heuristic elements into classical methods is supposed to lead to more accurate decisions. The idea may be viewed as a step towards integration between pattern recognition and artificial intelligence in the context of classification.

The positive experimental results with real data draw a perspective for further examination and possible applications of the scheme. Naturally, the classification rules at the first and the second levels are not confined to fuzzy k-NN nor to linear discriminant analysis. Any set of rules may be used taking into account the presumption that better results may be expected if the first level classifiers yield uncorrelated decisions.

References

[1] Fedrizzi, M., and Kacprzyk, J., "On the concept of fuzzy majority in consensus formation", Report of the System Research Institute - Polish Academy of Sciences, 1990.

[2] Jozwik, A., "A learning scheme for a fuzzy k-NN rule", *Pattern Recognition Letters* 1 (1983) 287–289.

[3] Jozefczyk, J., "Determination of optimal recognition algorithm in the two-level system", *Pattern Recognition Letters* 4 (1986) 413–420.

[4] Kacprzyk, J., and Fedrizzi, M., "A 'soft' measure of consensus in the setting of partial (fuzzy) preferences", *European Journal of Operational Research* 34 (1988) 316–325.

[5] Kunchev, K.H., "Investigations on the resistance of the flying staff in respect to a moderate degree of hypoxia", Thesis, Sofia, 1981 (in Bulgarian).

[6] Kuncheva, L.I., "Fuzzy multilevel classifier for medical applications", *Computers in Biology and Medicine* 20 (1990) 421–431.

[7] Kuncheva, L.I., "Additional criteria in fuzzy two-level pattern recognition", *Fuzzy Sets and Systems* 47 (1992) 281–291.

[8] Kuncheva, L., "Five measures of consensus in group decision making using fuzzy sets", in: *Proc. IFSA'92*, Brussels, Belgium, 1991, 141–144.

[9] Rastrigin, L., and Erenstein, R., "Method of group recognition", Moscow Energoizdat, 1981 (in Russian).

[10] Spillman, B., Spillman, R., and Bezdek, J., "A fuzzy analysis of consensus in small groups", in: P.P. Wang and S.K. Chang (eds.), *Fuzzy Sets. Theory and Application to Policy Analysis and Information Systems*, Plenum Press, New York, 1980, 291–308.