NIFS 1 (1995) 61-65

Intuitionistic Fuzzy Interpretation of Two-level Classifiers

Ognian K. Asparoukhov

Bulgarian Academy of Sciences, Central Laboratory of Biomedical Engineering,
Acad. G. Bonchev str., bl. 105, 1113 Sofia, Bulgaria

1. Introduction

The central problem of discriminant analysis is that of assigning an object (case, patient, pattern etc.) into one of several possible categories (classes). The result of any discriminant procedure is a classifier. Many different single - stage (one step) discriminant procedures (classifiers) have been described in the literature - statistical (Fukunaga, 1972; Hand, 1981; McLachlan, 1992), fuzzy (Asparoukhov, 1985; Kissiov and Hadjitodorov, 1992; Pedrycz, 1990), mathematical programming (Freed and Glover, 1986; Glover, 1990; Silva and Stam, 1994). Unfortunately conventional single stage classifiers have some disadvantages:

- usually, for a specific application problem, these classifiers differ in accuracy, but maybe none of them is totally perfect, or even not as good as expected for practical application;

- only one of the possible combination of object features (variables) is used in the classification;

-problem arises in case of a large number of variables as it leads to a small sample size and this is frequently met when the variables are mixed (continuous and categorical), because the classifiers appropriate for this situation (logistic discrimination, linear and quadratic local discriminant functions, kernel method - Krzanowski, 1975, 1993; McLachlan, 1992) are not so effective if the number of variables is large.

In order to overcome these disadvantages multi-stage approaches for classifier design were proposed: multiclassifiers (Xu et al., 1992; Wernecke, 1992), multi (two) level classifiers (Jozefczyk, 1986; Kissiov et al., 1990; Kuncheva, 1990, 1993; Ng and Abramson, 1992); tree classifiers (Asparoukhov and Ivanov, 1992; Casey and Nagy, 1984; Kurzynski, 1983; Landeweed et al, 1983; Li and Dubes, 1986; Lin and Fu, 1983; Mui and Fu, 1980; Shlien, 1988, 1990; Sturt, 1980), tree-structured rules (Brieman et al., 1984; Loh and Vanichsetakul, 1988).

In this paper we propose intuitionistic fuzzy interpretation of the two-level classifier design.

2. One -Stage Discriminant Procedures (Classifiers)

Let us consider a classical problem of discriminant analysis: $\omega,\dots,\omega_m$ are m distinct classes, whose objects are described by an n-component vector of variables $\mathbf{x}=(x_1,\dots,x_n)'$. Suppose that we have available a design set of N objects, N_1 of which are known to have come from ω_1, N_2 from $\omega_2,\dots$, N_m from ω_m ($N = N_1+N_2+\dots+N_m$). Design set object j from ω_i is described by vector $\mathbf{x}_{ij}$ consisting of the values of the m variables.

An individual $\mathbf{x}$ is classified into the class with the greatest estimated posterior probability $p(\omega_i/\mathbf{x})$. The posterior probabilities are estimated from the estimated class-conditional distributions and the prior probabilities via Bayes' theorem. If the prior probabilities are equal for the c classes, then $\mathbf{x}$ is allocated to ω_i if $p(x/\omega_i)=\max_j\ p(x/\omega_j)$.

There are two types of error rates, characterizing the effectiveness of every concrete classifier: resubstitution and leave-one-out and they have to be minimized.

This work was supported by National Science Fund of Bulgaria, grant No. L-318/93.

3. Two Level Classifiers

Let us have v statistical classifiers - CL_1, CL_2,..., Cl_v. Generally speaking, the output information that various classification algorithms supply or are able to supply can be divided into three levels (Xu et al., 1992):

1) Abstract level - every classifier only outputs a unique class or a subset of classes.

2) Rank level - every classifier ranks all the classes in a queue with the class at the top being the first choice.

3) Measurement level - every classifier attributes each class a measurement value to address the degree that the object has this class.

Among the three levels, the measurement level contains the highest amount of information and if we reduce this information we could rank all classes (rank level) or we can assign a unique class (abstract level). That is why below we will talk only about measurement level classifiers. The best developed measurement level classifiers are the statistical classifiers ones (Fukunaga, 1972; Hand, 1981; McLachlan, 1992). The output of every statistical classifier for a particular object $\mathbf{x}_j$ contains a probability for each of m classes (see table 1).

Table 1. Legend: $\mathbf{P_k(i,j)}$ - the estimated by classifier **k** probability that an object $\mathbf{x}_j$ belongs to class **i**. The sum of probabilities for every row is equal to 1.

Classifier No	**CLASSES**				
	ω_1	...	ω_i	...	ω_m
Classifier 1	$P_1(1,j)$	...	$P_1(i,j)$	...	$P_1(m,j)$
...	...	...	...	...	...
Classifier k	$P_k(1,j)$	...	$P_k(i,j)$	...	$P_k(m,j)$
...	...	...	...	...	...
Classifier v	$P_v(1,j)$	...	$P_v(i,j)$	...	$P_v(m,j)$

The main idea of two level and multi- classifiers is the aggregation at the second level of the received at the first level object probabilities for every class and the assignment of the object to the class with greatest aggregated estimate. The main problem is what this aggregation should beat the second level? There are a lot of types of aggregation (Xu et al., 1992), but all of them have one disadvantage - they use for every classifier only one value (probability) for a class.

Below, an intuitionistic fuzzy interpretation of the two-level classifier design, that used more first level information than conventional approaches, is presented.

4. Intuitionistic Fuzzy Interpretation of Two-level Classifiers Design

The output information for object $\mathbf{x}_j$ of the statistical classifiers CL_k for all classes contains the probabilities $P_k(1,j)$,...,$P_k(i,j)$,...,$P_k(m,j)$ and the conventional approaches use in object aggregation for one class only one value (i.e. probability) from all m values of this classifier.

We propose that in one class object aggregation two values (function of probabilities) per classifier be involved. For this purpose we define the following functions - degree of membership, degree of non membership and degree of indeterminacy:

Definition 1: The degree of membership of an object $\mathbf{x}_j$ to a class ω_i according to k-th classifier $\mu_{ik}(\mathbf{x}_j)$ is equal to the probability $P_k(i,j)$.

Definition 2: The degree of non membership of an object $\mathbf{x}_j$ to a class ω_i according to k-th classifier is equal to:

$$\nu_{ik}(x_j) = P_{(\max)} \quad for\ i \neq r$$
$$\nu_{ik}(x_j) = P_{(\max-1)} \quad for\ i = r\ ,$$

where $P_{(\max)} = \max\limits_{i=1,m} \{P_k(i,j)\} = P_k(r,j)$ and $P_{(\max\text{-}1)} = \max\limits_{i=1,m;i\neq r} \{P_k(i,j)\}$.

Both the degree of membership $\mu_{ik}(\mathbf{x}_j)$ and the degree of non membership $\nu_{ik}(\mathbf{x}_j)$ belong to the set $\{P_k(i,j),\ i = 1, m;\ \sum_{i=1}^{m} P_k(i,j) = 1\}$, so they are non negative and $\mu_{ik}(\mathbf{x}_j) + \nu_{ik}(\mathbf{x}_j) \leq 1$.

Definition 3: The degree of indeterminacy of an object $\mathbf{x}_j$ to a class ω_i according to k-th classifier $\pi_{ik}(\mathbf{x}_j)$ is equal to:

$\pi_{ik}(\mathbf{x}_j) = 1 - (\mu_{ik}(\mathbf{x}_j) + \nu_{ik}(\mathbf{x}_j))$.

How these functions could be interpreted?

The class degree of membership is equal to the class probability and this is a probability measure that this class is true for the object. The class degree of non membership is equal to the maximum probability from the other competitive classes, and it is measure that this class is false for the object. The difference between the measures of truth and falsity is the degree of indeterminacy and indeed this is measure (again probability one) for uncertainty (doubtfulness).

The functions involved by the definitions 1, 2 and 3 are typical functions for the Intuitionistic Fuzzy Set Theory (Atanassov, 1986) and its concepts, mathematical actions and relations could be used for their aggregation and decision making. Here we will turn our attention to the decision making problem after aggregation.

Let the results of aggregation be denoted by $\mu_i(\mathbf{x}_j)$, $\nu_i(\mathbf{x}_j)$ and $\pi_i(\mathbf{x}_j)$.

There are three one-function basic criteria:

Criterion 1: Maximum true.

This means that the degree of membership $\mu_i(\mathbf{x}_j)$ must be maximum. The functional of aggregation is:

$$\mu_i(\mathbf{x}_j) \to \max_i .$$

Criterion 2: Minimum false.

This means that the degree of non membership $\nu_i(\mathbf{x}_j)$ must be minimum. The functional of aggregation is:

$$\nu_i(\mathbf{x}_j) \to \min_i .$$

Criterion 3: Minimum indeterminacy (doubtfulness).

This means that the degree of indeterminacy $\pi_i(\mathbf{x}_j)$ must be minimum. It seems that this criterion is fulfilled if the first two are fulfilled, but it is not difficult to construct an example where the final decision for object class, derived by the criterion 3 is different from the one derived by both criteria 1 and 2. The functional of aggregation for the criterion 3 is:

$$\pi_i(\mathbf{x}_j) \to \min_i .$$

We accept that the product of two or three criteria leads to their simultaneous performance.

There three two criteria product combinations:

Combination 12 of the criteria 1 and 2: The functional of aggregation is:

$$\mu_i(\mathbf{x}_j) - \nu_i(\mathbf{x}_j). \to \max_i ,\ \text{or}$$

$\mu_i(x_j)/\nu_i(x_j) \to \max_i$, if the denominator is always different from zero.

Combination 13 of the criteria 1 and 3: The functional of aggregation is:

$$\mu_i(x_j) - \pi_i(x_j) = 2\mu_i(x_j) + \nu_i(x_j) - 1 \to \max_i, \text{ or } 2\mu_i x_j) + \nu_i(x_j) \to \max_i.$$

Combination 23 of the criteria 2 and 3: The functional of aggregation for this combination is:

$$\nu_i(x_j) + \pi_i(x_j) = 1 - \mu_i(x_j) \to \min_i.$$

There is only one three criteria combination - combination 123 of the criteria 1, 2 and 3. The aggregation functional for this combination could have a different form:

$$\frac{\mu_i(x_j)}{\nu_i(x_j) + \pi_i(x_j)} = \frac{\mu_i(x_j)}{1 - \mu_i(x_j)} \to \max_i, \text{ if } \mu_i(x_j) \neq 1, \text{ or}$$

$$\{\mu_i(x_j) - \nu_i(x_j)\}\{1 - \pi_i(x_j)\} = \{\mu_i(x_j) - \nu_i(x_j)\}\{\mu_i(x_j) + \nu_i(x_j)\} \to \max_i, \text{ or}$$

$$\frac{\mu_i(x_j) - \nu_i(x_j)}{\pi_i(x_j)} = \to \max_i, \text{ if } \mu_i(x_j) + \nu_i(x_j) \neq 1.$$

Of course there are other possibilities for formulation of the functional of the combination 123, which can not be described here due to the lack of space.

Here we only give a brief description of the idea of an intuitionistic fuzzy interpretation of two-level classifiers design. This idea is very appropriate for medical diagnostics and we now investigate it for preoperative prediction of postoperative pulmonary embolism.

References

Asparoukhov, O. K. (1985). Microprocessor system for investigation of thromboembolic complications. Unpublished Ph.D. dissertation, Technical University, Sofia.

Asparoukhov, O. and Ivanov, T. (1992). Tree classifier differential diagnostics of laryngeal diseases. Proceedings of the XVI colloquium MEDICOMP'92, 11-15.

Atanassov, K. (1986). Intuitionistic fuzzy sets. Fuzzy sets and Systems. Vol. 20, **1**, 87-96.

Breiman, L., Friedman, J., H., Olshen, R. A. and Stone, C. J. (1984). Classification and regression trees. Belmont, California: Wadsworth.

Casey, R. G. and Nagy, G. (1984). Decision tree design using a probabilistic model. *IEEE Transactions on information theory* IT-30, **1**, 93-99.

Freed, N. and Glover, F. (1986). A linear programming approach to the discriminant problem. *Decision Sciences*, 12, 68-74.

Fukunaga, K. (1972). Introduction to Statistical Pattern Recognition. Academic Press, London.

Glover, F. (1990). Improved linear programming models for discriminant analysis. *Decision Sciences*, 21, 771-785.

Hand, D. J. (1981). Discrimination and Classification. New York, Wiley.

Jozefczyk, J. (1986). Determination of optimal recognition algorithms in the two-level system. *Pattern Recognition Letters* 4, 413-420.

Kissiov, V. and Hadjitodorov, S. (1992). A fuzzy version of K-NN method. *Fuzzy Sets and Sistems* **49**, 323-329.

Kissiov, V., Hadjitodorov, S. and Kuncheva, L. (1990). Using key features in pattern classification. *Pattern Recognition Letters* **11**, 1-5.

Krzanowski, W.J. (1975). Discrimination and Classification Using Both Binary and Continuous Variables. *J. Amer. Statist. Assoc.*, vol. 70, 352.

Krzanowski, W.J. (1993). The location model for mixtures of categorical and continuous variables. *Journal of Classification* **10**, 25-49.

Kuncheva, L. (1993). Two-level classification schemes in medical diagnostics. *Int. J. Biomed. Comput.* **32**, 197-210.

Kuncheva, L. (1990). Fuzzy multi-level classifier for medical applications. *Comput. Biology Medicine* **20**, 421-431.

Landeweed, G. H., Timmers, T. and Gelsema, E. S. (1983). Binary tree versus single level tree classification of white blood cells. *Pattern Recognition*, Vol. 16, **6**, 571-577.

Li , X. and Dubes, R. C. (1986). Tree classifier design with a permutation statistic. *Pattern Recognition*, Vol. 19, **3**, 229-235.

Lin, Y. K. and Fu, K. S. (1983). Automatic classification of cervical cells using binary tree classifier. *Pattern Recognition*, Vol. 16, **1**, 69-80.

Loh, W.-Y. and Vanichsetakul, N. (1988). Tree-structured classification via generalized discriminant analysis (with discussion). *J. Amer. Statist. Assoc.* **83**, 715-728.

Kurzynski, M. W. (1983). The optimal strategy of a tree classifier. *Pattern Recognition*, Vol. 16, **1**, 81-87.

McLachlan, G. J. (1992). Discriminant Analysis and Statistical Pattern Recognition. New York: Wiley.

Mui, J. K. and Fu, K. S. (1980). Automated classification of nucleated blood cells using a binary tree classifier. *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-2, **5**, 429-443.

Ng, K. -C. and Abramson, B. (1992). Consensus diagnosis: A simulation study. *IEEE Trans. Syst., Man and Cybern.* **22**, 916-928.

Pedrycz, W. (1990). Fuzzy sets in pattern recognition: methodology and methods. *Pattern Recognition* 23, 121-146.

Schuermann, J. and Doster, W. (1984). A decision theoretic approach to hierarchical classifier design, *Pattern Recognition*, Vol. 17, **3**, 359-369.

Shlien, S. (1990). Multiple binary decision tree classifiers. *Pattern Recognition*, Vol. 23, **7**, 757-763.

Shlien, S. (1988). Multifont character recognition for typeset documents. *Pattern Recognition*, Vol. 2, **4**, 603-620.

Silva, P. D. and Stam, A. (1994). Second order mathematical programming formulation for discriminant analysis. *European Journal of Operational Research* 72, 4-22.

Sturt, E. (1980). Algorithm AS 165. An algorithm to construct a discriminant function in Fortran for categorical data. *Appl. Statist.* **30**, 313-325.

Xu, L., Krzyzak, A. and Suen, C. (1992). Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transaction on Systems, Man, and Cybernetics*, Vol. 22, **3**, 418-435.

Wernecke, K.-D. (1992). A coupling procedure for the discrimination of mixed data. *Biometrics* 48, 497-506.