ISSN 0146-4116, Automatic Control and Computer Sciences, 2015, Vol. 49, No. 3, pp. 123–132. © Allerton Press, Inc., 2015.
Original Russian Text © S.D. Shtovba, O.D. Pankevich, A.V. Nagorna, 2015, published in Avtomatika i Vychislitel'naya Tekhnika, 2015, No. 3, pp. 5–16.

Analyzing the Criteria for Fuzzy Classifier Learning

S. D. Shtovba, O. D. Pankevich, and A. V. Nagorna

Vinnytsia National Technical University,
Khmel'nitskoe shosse 95, Vinnitsa, 21021 Ukraine
e-mail: shtovba@gmail.com

Received September 4, 2014; in final form, February 2, 2015

Abstract—In a fuzzy classifier, the maping "inputs—output" is described by the linguistic ⟨If—then⟩ rules, the antecedents in which contain the fuzzy terms "low," "average," "high," and so on. To increase the correctness, the fuzzy classifier is learned by using experimental data. The problems with equal and different costs of various classification errors are discussed. A new criterion is offered for problems with undistinguishable types of errors, in addition to the two known criteria. A new one implies that the distance between the desired and real fuzzy results of classification for the cases of a wrong decision is weighted by the penalty factor. The learning criteria are generalized for problems of classification with the cost matrix. The conducted computer experiments on the wine recognition and heart disease diagnostics problems show that the best quality parameters of tuning fuzzy classifiers are achieved by a new learning criterion.

Keywords: learning, learning criteria, fuzzy rules, fuzzy inference, classification, cost matrix

DOI: 10.3103/S0146411615030098

1. INTRODUCTION

The classification problem lies in the assignment an object to one of preassigned classes. It is implemented by analyzing the attributes of the classified object. Various engineering, management, economic, political, medical, sport, and other problems are reduced to classification.

Fuzzy classifiers, namely, those using fuzzy sets during functioning or learning [1], have recently become more and more popular. The application of fuzzy sets for classification problems is presented in [2] for the first time. At present, classifiers based on logic inference in terms of production rules, the antecedents of which contain the fuzzy terms "low," "average," "high," and so on, are most popular. Each rule describes an area of factor space, wherein the objects belong to one class. Since the borders of these areas are fuzzy, one object can belong to several classes but with different degrees.

The main advantages of fuzzy classifiers are caused by the following factors:

• The logic inference over the fuzzy rule base is transparent. It is clear for the customers, among which are doctors, economists, politicians, and other specialists with low cybernetic engineering background.

• The classification models are compact. Only a few linguistic rules are required to describe the complicated dividing surfaces.

• Generating a base of linguistic rules is commonly simple for an expert.

• The logic inference can be implemented not only for numerical but for categorical and fuzzy values of input features as well. In this case, only the fuzzification procedure is modified in the logic inference algorithm [3], while the classification model remains constant.

The aforementioned advantages allow fuzzy decision-making models to be successful rivals for classifiers based on Bayesian rules, the nearest neighbor method, support vector machines, neural networks, and other data induction processing methods.

To increase the correctness, the fuzzy classifier is learned by experimental data. There are two approaches to the learning of the fuzzy classifier. The first one is based on the structural identification of the "inputs—output" relationship with fuzzy rules. It consists in the generation of a base of rules from the candidate-list [4], the selection of linguistic hedges, including "very" and "more or less" for the terms of rule antecedents [5], etc. Here, the learning is reduced to solving the discrete optimization problem. The second approach is based on the parametrical identification of the "inputs—output" relationship with the fuzzy rules. During learning, the rule semantics remains constant, and the membership functions of fuzzy terms and weight factors of rules are modified [1, 6, 7]. The learning is reduced to solving the optimization problem with continuous controllable variables.

The authors of the present paper consider parametrical identification, during which the classifier's parameters are iteratively changed to provide the minimum distance between the experimental data and the fuzzy inference results. There are several methods to define such distance, which is called a learning criterion. The purpose of the article is to reveal the criterion for which learning provides the best correctness of the fuzzy classifier. Cases with equal and different costs of various errors are studied. The last one assumes that the cost matrix is known.

2. FUZZY CLASSIFIER

Let us denote by $\mathbf{X} = (x_1, x_2, \ldots, x_n)$ the vector of informative features (attributes) of the classification object and by $l_1, l_2, \ldots, l_m$ the decision classes. Then, the representation $\mathbf{X} = (x_1, x_2, \ldots, x_n) \to y \in \{l_1, l_2, \ldots, l_m\}$, implemented by fuzzy rules, is said to be the fuzzy classifier. Based on [1, 4, 6, 7], the fuzzy rule base of this representation can be written as follows:

$$\text{If } \left(x_1 = \tilde{a}_{1j} \text{ and } x_2 = \tilde{a}_{2j} \text{ and } \ldots \text{ and } x_n = \tilde{a}_{nj} \text{ with the weight } w_j\right), \text{ then } y = d_j, \quad j = \overline{1, k}, \tag{1}$$

where k is the number of rules;

$d_j \in \{l_1, l_2, \ldots, l_m\}$ is the value of consequent of the jth rule;

$w_j \in [0, 1]$ is the weight factor specified the reliability of the jth rule, $j = \overline{1, k}$;

$\tilde{a}_{ij}$ is a fuzzy term that is the evaluating attribute x_i in the jth rule, $i = \overline{1, n}$, $j = \overline{1, k}$.

The classification of the current object given by the attribute vector $\mathbf{X}^* = (x_1^*, x_2^*, \ldots, x_n^*)$ is implemented as follows. At first, the degree of fulfillment of the jth rule from base (1) is calculated:

$$\mu_j(\mathbf{X}^*) = w_j \cdot \left(\mu_j(x_1^*) \wedge \mu_j(x_2^*) \wedge \ldots \wedge \mu_j(x_n^*) \right), \quad j = \overline{1, k}, \tag{2}$$

where $\mu_j(x_i^*)$ is the membership degree of x_i^* to the fuzzy term $\tilde{a}_{ij}$; $\wedge$ is the t-norm, which is generally realized by minimum operation or product.

The membership degree of the input vector $\mathbf{X}^*$ to classes $l_1, l_2, \ldots, l_m$ is estimated as follows:

$$\mu_{l_s}(y^*) = \underset{\forall j: d_j = l_s}{agg} \left(\mu_j(X^*)\right), \quad s = \overline{1, m}, \tag{3}$$

where *agg* is the aggregation of the results of fuzzy inference by the rules with the same consequents. The aggregation is realized by the maximum operation over the membership degrees that corresponds to the logic inference scheme with a *single winner rule* [8].

The fuzzy solution of the classification problem is the fuzzy set

$$\tilde{y}^* = \left(\frac{\mu_{l_1}(y^*)}{l_1}, \frac{\mu_{l_2}(y^*)}{l_2}, \ldots, \frac{\mu_{l_m}(y^*)}{l_m} \right). \tag{4}$$

The result of fuzzy inference is selected to be the core of the fuzzy set (4), namely, the class with the maximum membership degree:

$$y^* = \underset{\{l_1, l_2, \ldots, l_m\}}{\arg} \max_{s = \overline{1, m}} \left(\mu_{l_s}(y^*)\right).$$

The core of the fuzzy set (4) can include several elements. The object then concurrently belongs to several classes with equal degrees, the value of which is $\max_{s=\overline{1,m}}\left(\mu_{l_s}(y^*)\right)$. Let us apply the-voting-based scheme [8]. According to this scheme, the sum of degrees (2) of fulfillment of the corresponding rules is calculated for each class. The class with the maximum sum is selected as the decision.

3. LEARNING CRITERIA FOR THE FUZZY CLASSIFIER WITHOUT REGARD TO THE COST MATRIX

It is assumed that the learning set from M pairs "inputs–output" is known:

$$\left(\mathbf{X}_r, y_r\right), \quad r = \overline{1, M}, \tag{5}$$

where $y_r \in \{l_1, l_2, \ldots, l_m\}$.

Let us make the following denotations:

P is the vector of parameters of membership functions of the fuzzy terms from the rule base (1);

W is the vector of weight factors of rules from the base (1);

$F(\mathbf{K}, \mathbf{X}_r) \in \{l_1, l_2, \ldots, l_m\}$ is the classification result over the fuzzy rule base (1) with the parameters $\mathbf{K} = (\mathbf{P}, \mathbf{W})$ for the input vector $\mathbf{X}_r$ from the rth row of set (5).

The idea of fuzzy classifier learning is to find the vector **K**, which minimizes the distance between the results of the logic inference and experimental data from set (5). Below are three techniques for calculating this distance, which is said to be the learning criteria.

Criterion 1. The distance between the desired and real behaviors of the model can be defined by the rate of the misclassification on the learning set:

$$Crit_1 = \frac{1}{M} \sum_{r=\overline{1,M}} \Delta_r(\mathbf{K}), \tag{6}$$

where $\Delta_r(\mathbf{K}) = \begin{cases} 1, & \text{if } y_r \neq F(\mathbf{K}, \mathbf{X}_r) \\ 0, & \text{if } y_r = F(\mathbf{K}, \mathbf{X}_r) \end{cases}$ is the misclassification of the object $\mathbf{X}_r$.

The advantages of criterion (6) are simplicity and a clear, informative interpretation. The percent of errors is frequently used as the criterion of learning different pattern recognition systems [9]. In (6), the goal function of the corresponding optimization problem takes the discrete values, which hinders the use of rapid gradient optimization methods, especially for small learning sets.

Criterion 2. The learning quality can be related to the distance between the logic inference, resulting in the form of fuzzy set (4) and the values of output variable in the learning set. To do this, the value of output variable in the learning set (5) is transformed into such a fuzzy set [7]:

$$\left.\begin{aligned} \tilde{y} &= \left(\frac{1}{l_1}, \frac{0}{l_2}, \ldots, \frac{0}{l_m}\right), \text{ if } y = l_1 \\ \tilde{y} &= \left(\frac{0}{l_1}, \frac{1}{l_2}, \ldots, \frac{0}{l_m}\right), \text{ if } y = l_2 \\ &\vdots \\ \tilde{y} &= \left(\frac{0}{l_1}, \frac{0}{l_2}, \ldots, \frac{1}{l_m}\right), \text{ if } y = l_m \end{aligned}\right\}. \tag{7}$$

The learning criterion based on the distance between (4) and (7) is written as follows:

$$Crit_2 = \sqrt{\frac{1}{M} \sum_{r=\overline{1,M}} D_r(\mathbf{K})}, \tag{8}$$

where $D_r(\mathbf{K}) = \sum_{s=\overline{1,m}} \left(\mu_{l_s}(y_r) - \mu_{l_s}(\mathbf{K}, \mathbf{X}_r)\right)^2$ is the distance between the desired and real output fuzzy sets at classification of the rth object from the learning set (5); $\mu_{l_s}(y_r)$ is the degree of membership of the rth object from the learning set to the class l_s according to (5); $\mu_{l_s}(\mathbf{K}, \mathbf{X}_r)$ is the degree of membership to the class l_s of the output of the fuzzy model with parameters K in case of the input vector $\mathbf{X}_r$ according to (3).

The advantage of the criterion $Crit_2$ is that the extent of confidence in a solution based on the degrees of object membership to various classes is taken into account. In $Crit_1$, this information is neglected, i.e., the extent to which the membership degree of the solution is greater in comparison with other alternatives: by 0.0001 or 1 is unimportant. This means that the object classification result is taken to be absolutely reliable in case of $Crit_1$. Moreover, the goal function for the optimization problem by the criterion $Crit_2$ does not contain the extended plateau, and the fuzzy classifier can be learned by the rapid gradient methods. However, the fuzzy model optimal by (8) does not provide the minimal correctness of classification (6) in some cases [6, 10]. This is because the objects close to the borders of the classes make practically the same contribution D into the learning criterion (8) in both correct and wrong classifications.

Criterion 3. Below is a new learning criterion which has all of the mentioned advantages. The idea is to increase the distance D for the incorrectly classified objects:

$$Crit_3 = \sqrt{\frac{1}{M}\sum_{r=\overline{1,M}} \left(\Delta_r(\mathbf{K})\cdot penalty + 1\right)\cdot D_r(\mathbf{K})}, \tag{9}$$

where $penalty > 0$ is the penalty factor.

At $penalty \to 0$, the criteria (8) and (9) remain equivalent. At $penalty \to \infty$, the reliefs of goal functions of the optimization problems based on the criteria (6) and (9) are similar. During learning in terms of $Crit_3$, the selection of the direction of motion to the optimum is mostly defined by misclassified objects. Such behavior simulates the adaptive optimization method [11], wherein the incorrectly recognized objects are frequently presented for the repeated learning. The results of experiments from [11] prove that the learning at such approach is rapid.

4. LEARNING CRITERIA WITH THE COST MATRIX

The cost matrix if the following square matrix:

$$\mathbf{C} = \begin{bmatrix} 0 & c(l_1,l_2) & \dots & c(l_1,l_m) \\ c(l_2,l_1) & 0 & \dots & c(l_2,l_m) \\ \vdots & & & \\ c(l_m,l_1) & c(l_m,l_2) & \dots & 0 \end{bmatrix}, \tag{10}$$

where $c(l_i,l_j)$ is the cost of error of type $l_i \to l_j$, when the decision l_j, $i = \overline{1,m}$, $j = \overline{1,m}$, $i \neq j$ is wrongly selected as a result of classification instead of correct l_i. Zeros on the main diagonal of matrix (10) indicate the absence of the cost for correct classification.

When the cost matrix (10) is known, criterion 1 transforms to the following form:

$$Crit_{1C} = \frac{1}{M}\sum_{r=\overline{1,M}} c(u,v), \tag{11}$$

where $u = y_r$ and $v = F(\mathbf{K},\mathbf{X}_r)$.

Criterion 2 is modified so that the summands $D_r(\mathbf{K})$ are weighted by the costs of corresponding errors. As a result, formula (8) takes the form:

$$Crit_{2C} = \sqrt{\frac{1}{M}\sum_{r=\overline{1,M}} D_r(\mathbf{K},\mathbf{C})}, \tag{12}$$

where $D_r(\mathbf{K},\mathbf{C}) = \sum_{s=\overline{1,m}} \left(1 + c(y_r,l_s)\right)\cdot\left(\mu_{l_s}(y_r) - \mu_{l_s}(\mathbf{K},\mathbf{X}_r)\right)^2$ is the weighted distance between the desired and real output sets at classification of the rth object from the learning set (5).

Let us put the weighted distance $D_r(\mathbf{K},\mathbf{C})$ form (12) instead of $D_r(\mathbf{K})$ into criterion 3:

$$Crit_{3C} = \sqrt{\frac{1}{M}\sum_{r=\overline{1,M}} \left(\Delta_r(\mathbf{K})\cdot penalty + 1\right)\cdot D_r(\mathbf{K},\mathbf{C})}. \tag{13}$$

5. COMPUTER EXPERIMENTS

The purpose of experiments is to reveal the criterion by which learning provides the best correctness. Two test problems from UCI *Machine Learning Repository* [12] are considered. In the first problem on grape type recognition, all of the classification errors are equal, and the learning is implemented without the cost matrix. In the second problem on heart disease diagnosis, the cost of target pass is five times higher than that of the false alarm, and the corresponding cost matrix is used in learning.

Table 1. A base of fuzzy rules of wine classifiers

No.	x_1	x_7	x_{13}	y
1	–	–	Low	Type 1
2	Low	–	–	Type 2
3	–	Low	–	Type 3

Table 2. Parameters of the membership functions of terms of fuzzy wine classifiers

Attribute	Term	Initial		Best_min		Best_prod	
		b	c	b	c	b	c
x_1	Low	11	1.65	11	1.05	11	1.11
x_7	Low	0.34	2	0.34	0.91	0.34	0.764
x_{13}	Low	2.78	6	2.78	10	2.78	10

5.1. Problem of Grape Type Recognition

The problem of recognition the grape type (y) from which a wine is made is considered. The database *Wine Dataset* contains the results of chemical analysis over 13 factors of 178 samples of Italian wine made in the same region. One of three types of grape is pointed for each sample.

The learning set is formed from rows of a database with the boundary values from 13 attributes. Let us include all the odd rows of database into the learning set. The rest data are put to the test set. As a result, a learning set from 100 rows and a test set from 78 rows are obtained. Let us project the fuzzy classifier of wine with the following inputs: x_1 – alcohol; x_7 – flavanoids, and x_{13} – proline. Let us use the rule base (Table 1) of the fuzzy classifier of wine from [13]. The fuzzy terms are specified by the Gaussian membership function with two parameters:

$$\mu(x) = \exp\left(-\frac{(x-b)^2}{2c^2}\right), \tag{14}$$

where b is the coordinate of the maximum and $c > 0$ is the concentration factor.

The parameters of the membership functions of the initial fuzzy classifier is given in Table 2.

For each criterion, 500 experiments in learning the fuzzy model on the basis of quasi-Newton algorithm are carried out. Then, each classifier is verified on the test set by the criterion $Crit_1$. The experiments are conducted for two fuzzy models, the t-norms in which are implemented by the operation of minimum and the product, respectively.

During learning, the weight factors of each of three rules of knowledge base and the concentration factors (c) of the membership function of each fuzzy terms are tuned. Since all of the fuzzy terms in the knowledge base are extreme, the coordinates of the maximums of membership function (coefficients b) are not adjusted according to [14]. They are set equal to the minimum value of the corresponding attribute. Thus, the total number of the tuning parameters is 3 + 3 = 6. The initial points for learning were selected at random for the weight factors of rules from the range [0, 1] and for the membership function in the range ±20% from Table 2.

At first, the appropriate values of the penalty factor in $Crit_3$ are determined. To do this, 500 experiments, in which the penalty factor was selected at random from the range (0, 10] for each fuzzy model, were conducted. Then, this interval is divided into five equal sections, and the frequency of successful experiments (α) is calculated for each one. The experiment, the result of which falls in the top 25% in terms of correctness, is taken to be successful. During testing, it was found (Fig. 1) that the maximum number of successful launches occurs at $penalty \in (0,2]$. Such values of the penalty factor will be used in further experiments.

After a series of 3000 experiments, a set of ten classifiers, each of which provides the minimum misclassification rate (6) on the level 0.039 on the test set, was obtained. Let us select two classifiers, Best_min

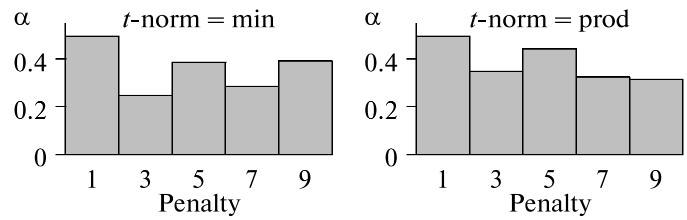

Fig. 1. Relationship between the success of learning the fuzzy wine classifier and the penalty factor.

and Best_prod, with minimum values of this criterion on the test set, namely, $Crit_1$(Best_min) = 0.1 and $Crit_1$(Best_prod) = 0.1. Hereafter, Best_min denotes the best classifier with the t-norm implemented by the operation of minimum and Best_prod identifies the best classifier with the t-norm implemented by the product. The weight factors of the rules in Best_min are $\omega_1 = 0.26$, $\omega_2 = 1$, $\omega_3 = 0.49$, while Best_prod has $\omega_1 = 0.29$, $\omega_2 = 1$, $\omega_3 = 0.72$. The parameters of the membership functions of these classifiers are shown in Table 2.

The experimental results (Table 3 and Fig. 2) show that the best learning quality is observed on average with $Crit_3$. The widest and narrowest dispersions of the learning results are at optimization according to $Crit_1$ and $Crit_2$, respectively.

Let us develop a model for defining the grape type with an alternative method, which will help in selecting the decision tree. Based on the splitting rules by the Gini index, entropy, and towing [15], three decision trees were synthesized. After pruning, the tree with entropy splitting proved to be the best. The tree contains seven rules with the length of antecedents from two to four logic conditions. The misclassification rate of this tree on the test set is 0.103. Therefore, the fuzzy classifier for the problem of grape recognision turned out to be the better than the decision tree in terms of correctness and compactness criteria.

5.2. Problem of Diagnosis of Heart Diseases

The problem of heart disease diagnosis on the basis of data from *Statlog Heart Data Set* is considered. Each of the 270 rows of this base contains a description of 13 features of the patient's state. They are used to make a decision (y) about the presence or absence of the heart disease. The following cost matrix is known: $\mathbf{C} = \begin{bmatrix} 0 & 1 \\ 5 & 0 \end{bmatrix}$.

The learning set is generated from the rows of the database, which involve the boundary values of each of 13 attributes. All the odd rows are included in the learning set. The remaining data are placed in the test set. As a result, we obtain the learning set from 145 rows and the test set from 125 rows.

Table 3. Statistics of learning the fuzzy wine classifiers

t-norm	Learning criterion	Value of criterion $Crit_1$ on test set			
		minimum	average	maximum	MSD
Minimum	$Crit_1$	0.103	0.582	0.962	0.143
	$Crit_2$	0.051	0.229	0.564	0.058
	$Crit_3$	0.039	0.198	0.513	0.114
Product	$Crit_1$	0.115	0.599	0.974	0.142
	$Crit_2$	0.09	0.223	0.449	0.049
	$Crit_3$	0.039	0.2	0.68	0.124

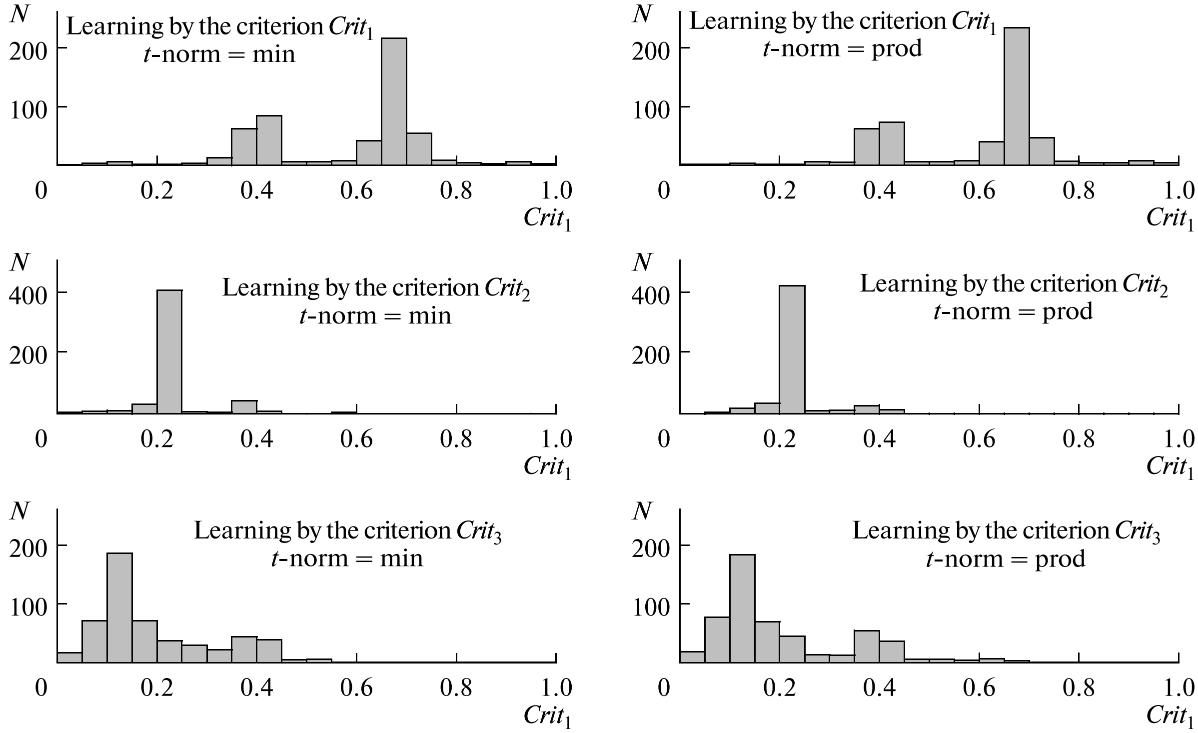

Fig. 2. Distribution of the results of learning the fuzzy wine classifiers (N is the number of cases).

Let us design the fuzzy classifier with three inputs:

x_1 is the age; x_{10} is the old peak; x_{12} is the number of major vessels colored by fluoroscopy.

The fuzzy knowledge base is formed by the distribution of data (Table 4). The fuzzy terms are specified by the Gaussian membership function (14) with parameters from Table 5.

Table 4. A base of fuzzy rules for the diagnosis of heart diseases

No.	x_1	x_{10}	x_{12}	y
1	–	Low	Few	Healthy
2	Young	High	Few	Healthy
3	–	–	Much	Sick
4	Old	High	Few	Sick

Table 5. Parameters of the membership functions of terms of fuzzy classifiers for the diagnosis of heart diseases

Attribute	Term	Initial		Best_min		Best_prod	
		b	c	b	c	b	c
x_1	Low	29	20.4	29	10.4	29	8.75
	High	77	20.4	77	8.69	77	8.38
x_{10}	Low	0	2.63	0	0.83	0	0.93
	High	6.2	2.63	6.2	3	6.2	1.11
x_{12}	Low	0	1.27	0	0.57	0	0.45
	High	3	1.27	3	2	3	2

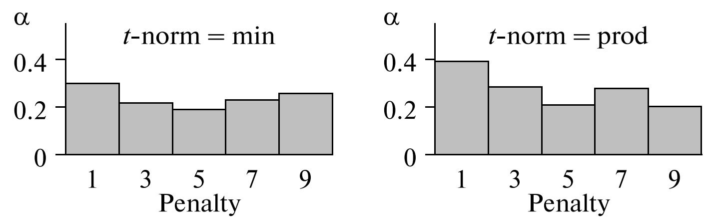

Fig. 3. Relationship between the success of learning the fuzzy classifier for the diagnosis of the heart diseases and the penalty factor.

During learning, the weight coefficients of each from four rules of the knowledge base and the concentration factors (c) of the membership function of each fuzzy term are tuned. Since all of the fuzzy terms in the knowledge base are extreme, the coordinates of the maximums of membership functions are set equal to the borders of the interval of the attribute change. Thus, the general number of the tuning parameters is 4 + 6 = 10. The initial points for learning were selected at random for the weight factors of rules from [0, 1] and for the parameters of the membership functions within $\pm 20\%$ from the values in Table 5.

As we did previously, let us first define the appropriate values of the penalty factors in the criterion $Crit_{3C}$. From 500 experiments, it was found (Fig. 3) that the greatest number of successful launches is observed at $penalty \in (0,2]$. Such values of the penalty factor will be used in further experiments.

After a series of 3000 experiments, a set of ten classifiers, each of which provides the minimum value of $Crit_{1C}$ on the level 0.424 on the test set, was obtained. Let us select two classifiers, Best_min and Best_prod, with minimum values of this criterion on the test set, namely, $Crit_{1C}$(Best_min) = 0.683 and $Crit_{1C}$(Best_prod) = 0.683. The weight factors of the rules are $\omega_1 = 0.75$, $\omega_2 = 0.24$, $\omega_3 = 0.96$, and $\omega_4 = 1$ in Best_min and $\omega_1 = 0.67$, $\omega_2 = 0.63$, $\omega_3 = 1$, and $\omega_4 = 0.27$ in Best_prod. The parameters of the membership functions of these classifiers are shown in Table 5.

The experimental results (Table 6 and Fig. 4) proves that the best learning quality is observed at $Crit_{3C}$. The widest and narrowest dispersions are at optimization by $Crit_1$ and $Crit_2$, respectively.

Let us compare the fuzzy classifier with alternative models in form of the decision tree. With three models synthesized with the splitting rules based on Gini index, entropy, and towing, the entropy-based tree proved to be the best after pruning. The tree contains 12 rules, each of which involves from three to four logic conditions. The risk (11) of this tree on the test set is 0.424, which is identical to the fuzzy classifier. However, the fuzzy model turned out to be significantly more compact that the decision tree.

Table 6. Statistics of learning the fuzzy classifiers for diagnosis of heart diseases

t-norm	Learning criterion	Value of criterion $Crit_{1C}$ on test set			
		minimum	average	maximum	MSD
Minimum	$Crit_{1C}$	0.424	1.185	1.408	0.309
	$Crit_{2C}$	0.96	1.081	1.112	0.015
	$Crit_{3C}$	0.424	0.592	1.032	0.094
Product	$Crit_{1C}$	0.424	1.051	1.408	0.399
	$Crit_{2C}$	0.832	0.923	0.952	0.019
	$Crit_{3C}$	0.424	0.59	1.288	0.103

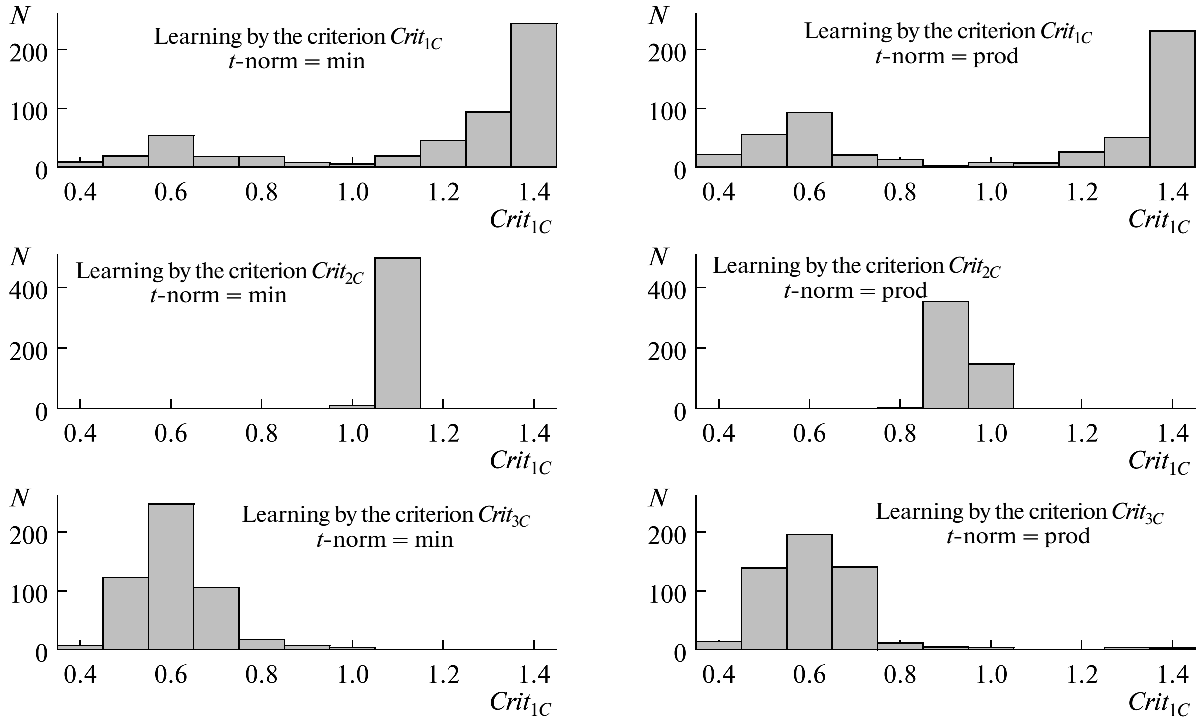

Fig. 4. Distribution of the results of learning the fuzzy classifiers for the diagnosis of heart diseases (N is the number of cases).

6. CONCLUSIONS

The fuzzy rule-based classifiers were considered for problems with the equal and different costs of various classification errors. A new criterion was offered for problems with undistinguished types of errors, in addition to two learning criteria of the fuzzy classifier on the basis of the error frequency and the distance between the fuzzy sets. The distance between the desired and real fuzzy classification results in this criterion is weighted by the penalty factor in the case of an incorrect decision. The learning criteria are generalized for problems wherein the costs of classification errors are specified by the cost matrix. The conducted computer experiments in the wine recognition and heart disease diagnosis problems show that the best quality of tuning the fuzzy classifiers are achieved in case of using the suggested learning criterion.

REFERENCES

1. Kuncheva, L.I., Fuzzy classifier design, in *Studies in Fuzziness and Soft Computing. vol. 49,* Berlin: Springer-Verlag, 2000.
2. Bellman, R., Kalaba, R., and Zadeh, L., Abstraction and pattern classification, *J. Mathem. Anal. Appl.,* 1966, vol. 13, no. 1, pp. 1–7.
3. Rotshtein, A. and Shtovba, S., Identification of a nonlinear dependence by a fuzzy knowledge base in the case of a fuzzy training set, *Cybern. Syst. Anal.*, 2006, vol. 42, pp. 176–182.
4. Ishibuchi, H., Nakashima, T., and Nii, M., *Classification and Modeling with Linguistic Information Granules: Advanced Approaches to Linguistic Data Mining,* Berlin: Springer–Verlag, 2005.
5. Cordon, O., del Jesus, M.J., and Herrera, F., Genetic learning of fuzzy rule–based classification systems cooperating with fuzzy reasoning methods, *Int. J. Intell. Syst.,* 1998, vol. 13, pp. 1025–1053.
6. Shtovba, S.D., *Proektirovanie nechetkikh sistem sredstvami MATLAB* (Fuzzy System Design by MATLAB Means), Moscow: Goryachaya liniya, Telekom, 2007.
7. Rotshtein, A., Design and tuning of fuzzy rule–based system for medical diagnosis, in *Fuzzy and Neuro-Fuzzy Systems in Medicine,* Teodorescu, N.H., Kandel, A., and Jain, L.C., Eds., Boca-Raton: CRC-Press, 1998, pp. 243–289.
8. Ishibuchi, H., Nakashima, T., and Morisawa, T., Voting in fuzzy rule–based systems for pattern classification problems, *Fuzzy Sets and Systems*, 1999, vol. 103, pp. 223–238.

9. Duda, R.O. and Hart, P.E., *Pattern Classification and Scene Analysis,* New York: Wiley, 1973; Moscow: Mir, 1976.
10. Shtovba, S., Pankevich, O., and Dounias, G., Tuning the fuzzy classification models with various learning criteria: The case of credit data classification, *Proc. Int. Conf. on Fuzzy Sets and Soft Computing in Economics and Finance, St. Petersburg, 2004.*
11. Rastrigin, L.A., *Adaptatsiya slozhnykh sistem. Metody i prilozheniya* (Adaptation of Complex Systems. Methods and Applications), Riga: Zinatne, 1981.
12. Bache, K. and Lichman, M., *UCI Machine Learning Repository,* http://archive.ics.uci.edu/ml]. Irvine, CA: Univ. Calif., School Inform. Comp. Sci., 2013.
13. Ishibuchi, H. and Yamamoto, T., Fuzzy rule selection by multi–objective genetic local search algorithms and rule evaluation measures in data mining, *Fuzzy Sets and Systems*, 2004, vol. 141, pp. 59–88.
14. Shtovba, S., Ensuring accuracy and transparency of Mamdani fuzzy model in learning by experimental data, *J. Autom. Inform. Sci.,* 2007, vol. 39, no. 8, pp. 39–52.
15. Breiman, L., Technical note: Some properties of splitting criteria, *Machine Learning,* 1996, vol. 24, no. 1, pp. 41–47.

Translated by A. Evseeva