S. D. Shtovba, Dr. Sc(Eng)., Professor; A. V. Galushchak

COMPARISON OF LEARNING CRITERIA FOR FUZZY CLASSIFIER WITH VOTING RULES

In fuzzy classifiers decision-making is based on linguistic rules <If - then>, antecedents of which contain fuzzy terms "low", "average", "high" etc. To increase the correctness fuzzy classifier is learned by experimental data. We study a fuzzy classifier with voting rules in which by the result of logic inference is class with maximum total supports by all the rules. New criteria of fuzzy classifier learning are suggested, they take into account the difference of memberships of fuzzy inference only to main competitors. In case of correct classification, main competitor of the taken decision is the class with the second membership degree. In case of incorrect classification, decision, taken by mistake is main competitor of the correct class. Computer experiments, dealing with learning of fuzzy classifier for UCI-problem of Italian wines recognition proved significant advantage of new learning criteria.

Key words: *classification, fuzzy knowledge base, learning, voting rules, learning criteria, main competitors.*

Introduction

The problem of classification is in mapping the object by certain features to one of classes. In fuzzy classifiers decision-making is based on linguistic rules <If - then>, antecedents of these rules contain fuzzy terms "low", "average", "high" etc. [1]. Each rule sets the area of feature space, within the boundaries of which objects belong to the same class. Boundaries of those areas are fuzzy. That is why one and the same object may simultaneously belong to several classes, but with different degree.

In fuzzy classifiers aggregation in logic inference by all the rules of knowledge base is realized by two schemes. In accordance with the first scheme with a single winner rule, the consequent of the rule with maximum degree of execution is chosen as a result of logic inference [2]. In accordance with the second scheme with voting rule, the class with maximal total membership by all the rules is selected as a result of logic inference [3]. The advantage of the scheme with a single winner rule is more interpretable algorithm of logic inference, and schemes with voting rules have smoother boundaries of classes devision in feature space [4].

To increase the correctness fuzzy classifier is learned by experimental data. For this its parameters are changed iteratively, to minimize the distance between experimental data and results of fuzzy inference. This distance, we call it learning criterion can be defined in different ways. Accordingly, there appears the interest in choosing such learning criterion, that would provide the best correctness of fuzzy classifier on test set.

In [5-8] the efficiency of three learning criteria for fuzzy classifier with a single winner rule was checked experimentally. Such criteria were studied: 1) misclassification rate; 2) squared distance between two fuzzy sets – desire and real results of classification; 3) squared distance between fuzzy desired and real results of classification with additional penalty for misclassification.

For classifier with voting rules misclassification rate is used as learning criterion [4]. The aim of the paper is testing of five learning criteria of fuzzy classifiers with voting rules – three from [5-8] and two new. New learning criteria of fuzzy classifier take into account membership difference of fuzzy inference only to main competitors. The class with the second membership grade is the main competitor for the decision in case of correct classification. This difference should be maximized to move off from the class, with which the correct decision could be easily confused. In case of incorrect classification decision made by mistake is main competitor of the correct class, that is why, the difference between the degrees of membership to these classes should be reduced in the process of learning.

1. Fuzzy classifier with voting rules

Let us introduce such denotations:

$\mathbf{X} = (x_1, x_2, \ldots, x_n)$ are features of classifying object (vector of input attributes);

$l_1, l_2, \ldots, l_m$ are classes of decisions;

y is result of classification.

Fuzzy classifier is mapping of $\mathbf{X} = (x_1, x_2, \ldots, x_n) \rightarrow y \in \{l_1, l_2, \ldots, l_m\}$ on the base of fuzzy rules. Proceeding from [2] rule base of fuzzy classifier we will write as:

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

where k is number of rules;

$d_j \in \{l_1, l_2, \ldots, l_m\}$ is categorial value of the consequent of the j-th rule;

$w_j \in [0, 1]$ is confidence factor, that describes the trustworthiness of the j-th rule, $j = \overline{1, k}$;

$\tilde{a}_{ij}$ is fuzzy term, that evaluates variable x_i in the j-th rule, $i = \overline{1, n}$, $j = \overline{1, k}$.

Classification of the current object with attributes $\mathbf{X}^* = (x_1^*, x_2^*, \ldots, x_n^*)$ is realized in the following way. First, the degree of j-th rule execution is calculated from the base (1):

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

where $\mu_j(x_i^*)$ is membership degree of x_i^* to fuzzy term $\tilde{a}_{ij}$;

$\wedge$ is t-norm, that is usually realized by operation minimum or product.

The membership degree of input vector $\mathbf{X}^*$ to classes $l_1, l_2, \ldots, l_m$ is calculated in the following way:

$$\mu_{l_s}(y^*) = \frac{\sum\limits_{\forall j: d_j = l_s, \ j = \overline{1, k}} \mu_j(X^*)}{\max\limits_{s = \overline{1, m}} \left(\sum\limits_{\forall j: d_j = l_s, \ j = \overline{1, k}} \mu_j(X^*) \right)} \ , \ s = \overline{1, m}. \tag{3}$$

Fuzzy decision of classification problem will be 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 inference we will choose the core of fuzzy set (4), i.e. the class with maximum sum of membership degrees:

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

The situation is possible, when the core of fuzzy set (4) comprises several elements. In this case, the object simultaneously belongs to several classes with the same degrees, value of which equals $\max\limits_{s = \overline{1, m}} \left(\mu_{l_s}(y^*) \right)$. For selection of one of these competitive classes we apply the scheme, based on single winner rule applying this scheme, among the rules, concerning these competitive classes, we will select the rule with maximum degree of activation.

2. Learning criteria for fuzzy classifier

Let us denote the training set of M input-output pairs as follows:

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

where $\mathbf{X}_r = \left(x_{r1}, x_{r2}, \ldots, x_{rn}\right)$ are input attributes of r-th object;

$y_r \in \{l_1, l_2, \ldots, l_m\}$ is class of r-th object.

We will introduce such denotations:

$\mathbf{P}$ is vector of membership functions parameters of the terms from knowledge base (1);

$\mathbf{W}$ is vector of confidence factor knowledge base (1);

$F(\mathbf{K}, \mathbf{X}_r) \in \{l_1, l_2, \ldots, l_m\}$ is classification result by knowledge base (1) with parameters $\mathbf{K} = (\mathbf{P}, \mathbf{W})$ for input vector $\mathbf{X}_r$ from r-th row of the training set (5).

Learning of fuzzy classifier consists in finding of such vector $\mathbf{K}$ that minimizes the misclassifications on test set. For tuning $\mathbf{K}$ parameters only training set (5) is used. Learning is considered as optimization problem of searching such controlled variables $\mathbf{K}$ that minimize the distance between the results of logic inference and experimental data from the set (5). This distance that we call learning criterion, may be determined in different ways.

Criterion 1 – misclassification rate [4]:

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

where $\Delta_r(\mathbf{K}) = \begin{cases} 1, & if \quad y_r \neq F(\mathbf{K}, \mathbf{X}_r) \\ 0, & if \quad y_r = F(\mathbf{K}, \mathbf{X}_r) \end{cases}$.

The advantages of the criterion consist in its simplicity and clear content interpretation. But goal function in corresponding optimization problem takes discrete values, that complicates the application of rapid gradient methods of optimization, especially in case of small data samples.

Criterion 2 – squared distance between two fuzzy sets – desired and real results of classification [2]. For its calculation values of output variable y in training set is transformed into following fuzzy set:

$$\left.\begin{array}{ll} \tilde{y} = \left(\dfrac{1}{l_1}, \dfrac{0}{l_2}, \ldots, \dfrac{0}{l_m}\right), & if \quad y = l_1 \\ \tilde{y} = \left(\dfrac{0}{l_1}, \dfrac{1}{l_2}, \ldots, \dfrac{0}{l_m}\right), & if \quad y = l_2 \\ \vdots & \\ \tilde{y} = \left(\dfrac{0}{l_1}, \dfrac{0}{l_2}, \ldots, \dfrac{1}{l_m}\right), & if \quad y = l_m \end{array}\right\}. \tag{7}$$

Learning criterion is based on the distance between logic inference in form of fuzzy set (4) and desired fuzzy value of output variable (7) is written as follows:

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

where $D_r(\mathbf{K}) = \sum\limits_{s=\overline{1, m}} \left(\mu_{l_s}(y_r) - \mu_{l_s}(\mathbf{K}, \mathbf{X}_r)\right)^2$ is distance between the desired and real output fuzzy sets in the process of classification of r-th object from training set (5);

For calculation of $D_r(\mathbf{K})$ Euclidean metric is used:

$$D_r(\mathbf{K}) = \sum_{s=\overline{1, m}} \left(\mu_{l_s}(y_r) - \mu_{l_s}(\mathbf{K}, \mathbf{X}_r)\right)^2\,, \tag{9}$$

where $\mu_{l_s}(y_r)$ is membership degree of r-th object of training set to class l_s according to (7); $\mu_{l_s}(\mathbf{K}, \mathbf{X}_r)$ is calculated by the formula (3) membership degree of the inferred result by fuzzy model with $\mathbf{K}$ parameters to class l_s for input vector $\mathbf{X}_r$.

The advantage of $Crit_2$ consist in taking into account the degree of confidence in decision made, based on the degrees of object membership to different classes. In $Crit_1$ it is considered that the result of object classification is reliable, i.e., it is of no importance how membership degree of the decision is greater, than in other alternatives - by 0.0001 or by 1. Besides, goal function in optimization problem by the criterion (8) has no long plateaus, that is why it is suitable for optimization with gradient methods. But objects, close to the boundaries of classes, make almost the same contribution in learning criterion (8), both at correct and wrong classification, that is why, learning maybe unproductive.

Criterion 3 – squared distance between fuzzy desires and real results of classification with additional penalty for wrong decision [5 - 8]. This criterion inherits the advantages of two previous criteria. The idea consists in increasing the distance D for misclassificated objects:

$$Crit_3 = \sum_{r=1,\,M} \left(\Delta_r(\mathbf{K}) \cdot p + 1\right) \cdot D_r(\mathbf{K}) \ , \tag{10}$$

where $p > 0$ is penalty factor.

Criterion 4 – distance between main competitors with the penalty for wrong decision. This is new learning criterion. According to the inference algorithm class with maximum membership degree is chosen as a decision. We will denote this class-winner as *win* and assign it the first rank.

In case of correct classification main competitor of the decision is *vicewin* – class with the second rank, i.e., class with the second by membership degree (fig. 1a). The greater is the difference between membership degrees *win* and *vicewin*, the greater is the confidence in logic inference, and the object is further from the boundary of classes division.

Let us denote by smax is the operation of finding the element of set with the second value. Then, for r-th object from the set (5) $\mu_{win}(\mathbf{X}_r) = \max\limits_{s=1,\,m}\left(\mu_{l_s}(\mathbf{X}_r)\right)$ and $\mu_{vicewin}(\mathbf{X}_r) = \operatorname{smax}\limits_{s=1,\,m}\left(\mu_{l_s}(\mathbf{X}_r)\right)$. Accordingly, the difference between the competitors equals $\mu_{win}(\mathbf{X}_r) - \mu_{vicewin}(\mathbf{X}_r)$.

In case of incorrect classification, wrong decision will be main competitor of the correct class (fig. 1b). Accordingly, it is desirable to reduce this difference between the degrees of membership to wrong decision and to correct class. The difference between main competitors in this case will be written in the following way: $\mu_{win}(\mathbf{X}_r) - \mu_{y_r}(\mathbf{X}_r)$.

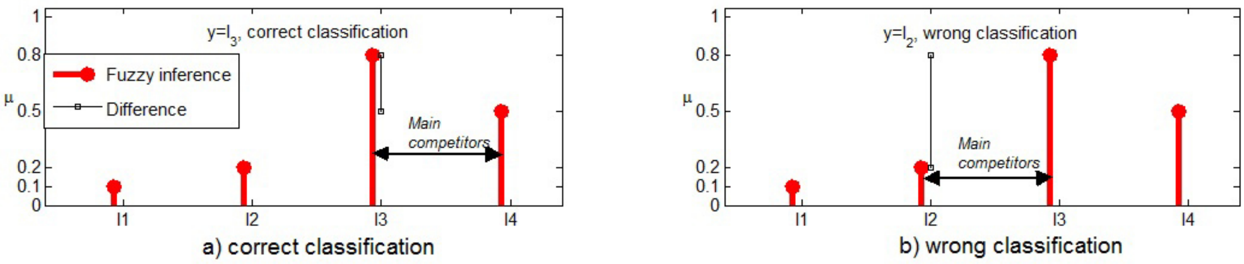

Fig. 1. Main competitors
a) correct classification; б) wrong classification

In learning criterion, we take into account relative indices, having divided the difference into the membership degree of the winner-class. In case of correct classification relative difference equals

$D_r^1 = \dfrac{\mu_{win}(\mathbf{X}_r) - \mu_{vicewin}(\mathbf{X}_r)}{\mu_{win}(\mathbf{X}_r)}$, and in case of wrong classification $D_r^0 = \dfrac{\mu_{win}(\mathbf{X}_r) - \mu_{y_r}(\mathbf{X}_r)}{\mu_{win}(\mathbf{X}_r)}$. Besides, similarly to criterion 3, in case of wrong classification, we multiply the difference by penalty factor. Mathematically, learning criterion will be written in the following way:

$$Crit_4 = p \cdot \sum_{\substack{y_r \neq F(\underline{\mathbf{K}, \mathbf{X}}_r) \\ r=1,M}} D_r^0(\mathbf{K}) - \sum_{\substack{y_r = F(\underline{\mathbf{K}, \mathbf{X}}_r) \\ r=1,M}} D_r^1(\mathbf{K}) \quad , \tag{11}$$

where $p \geq 1$ is penalty factor.

As the example on Fig.1 we calculate the distance (11) by the results of logic inference. In case of correct classification (Fig. 1a) this distance equals: $D_a^1 = \dfrac{0.8 - 0.5}{0.8} = 0.375$. In case of incorrect classification (Fig. 1b) at penalty factor $p = 3$, the distance equals: $D_б^1 = 3 \cdot \dfrac{0.8 - 0.2}{0.8} = 2.25$.

Criterion 5 – squared distance between main competitors with the penalty for wrong decision. This criterion is a modification of the previous one. The difference consists in using of not absolute distances but their squares:

$$Crit_5 = p \cdot \sum_{\substack{y_r \neq F(\underline{\mathbf{K}, \mathbf{X}}_r) \\ r=1,M}} D_r^0(\mathbf{K})^2 - \sum_{\substack{y_r = F(\underline{\mathbf{K}, \mathbf{X}}_r) \\ r=1,M}} D_r^1(\mathbf{K})^2 \quad .$$

Squaring in $Crit_5$ enables to increase the contribution in learning criteria of large differences and decrease the contribution of small differences.

3. Computer experiments

The aim of the experiments is to determine criterion, learning by which provides the best correctness. Test problem Wine Dataset from UCI Machine Learning Repository is considered. It consists in determining the sort of grapes (y), the wine is made of. Database contains the results of laboratory analyses, made by 13 indices of 178 specimen of Italian wine, produced in the same region. For each specimen one of 3 grape sorts used for wine producing is indicated. Traning set will be formed of the lines of data set with extreme values of each of 13 attributes. Additionally, all the odd lines of data base are included in training set. All often data will be written in test sample. As a result, we will obtain training set, containing 100 lines and test sample, containing 78 lines. We will design fuzzy classifier of wines with three features: x_7 - flavanoids, x_{10} - color intensity and x_{13} - proline. After the visualization of experimental data (Fig. 2) we will form fuzzy knowledge base (Table 1) with 5 rules. Fuzzy terms will be described by Gaussian curve:

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

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

Parameters of membership functions of initial fuzzy classifier are given in Table 2.

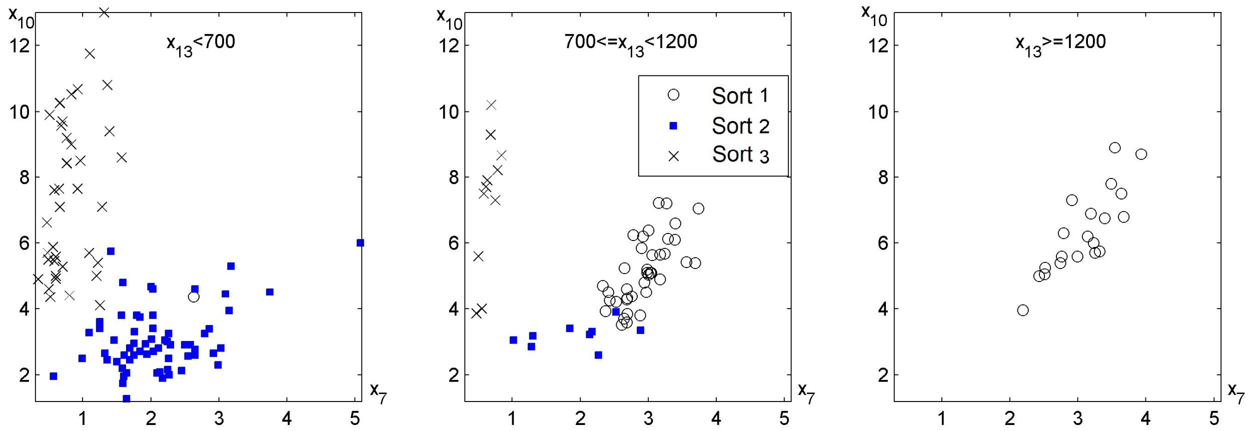

Fig. 2. Data distribution in Wine Database

Table 1

Fuzzy knowledge base

№	x_7	x_{10}	x_{13}	y
1	–	–	High	Sort 1
2	High	High	Average	Sort 1
3	–	Low	Low	Sort 2
4	Low	Low	Average	Sort 2
5	Low	High	–	Sort 3

Table 2

Parameters of membership functions of the terms of fuzzy classifier

Feature	Term	Parameters	
		b	c
x_7	Low	2	0.34
	High	2	5.08
x_{10}	Low	6	1.28
	High	6	13
x_{13}	Low	3	2.78
	Average	3	10
	High	3	16.8

For each criterion we will carry out 1000 experiments of fuzzy knowledge base learning based on quasi-Newton algorithm. After learning each classifier will be checked on test set according to mistakes frequency (criterion $Crit_1$). During the learning we tune confidence factor of the first four rules. Trustworthiness of the fifth rule does not excite doubts, that is why, according to [9] its confidence factor is not tuned. Let us also tune the coefficients of concentration (c) of membership function of each fuzzy term. In order to keep interpretability of knowledge base according to [10] we will tune coordinates of maxima (b) of membership functions only of non-extreme terms. In knowledge base there is only one intermediate term – "Average", we will change its coordinate of maximum. Thus, total number of tuned parameters is $4+7+1=12$. Initial points for learning will be chosen randomly – for confidence factor of the rules from the range [0, 1] and for parameters of membership functions within the limits of $\pm 30\%$ from the values in Table 2.

We will carry out two series of experiments. The first series – for fuzzy classifier with the realization of t-norm by operation minimum (min), and the second – with the realization of t-norm by operation product (prod). In learning experiments by criteria $Crit_3$, $Crit_4$ and $Crit_5$ at first we will define the acceptable level of penalty factor. For this purpose we will carry out 200 experiments for $p = 1, 3, ..., 9$. The results of all experiments (fig. 3 and 4, table 3) showed that learning is much better, if for criterion $Crit_3$ $p = 1$. At criteria $Crit_3$ and $Crit_5$ quality of learning is not so sensitive to penalty factor. When t-norm is realized by operation minimum the best correctness is achieved, if $p = 3$ for $Crit_4$ and $p = 5$ for $Crit_5$. When t-norm is realized by the product the best correctness is achieved, if $p = 5$ for $Crit_4$ and $p = 3$ for $Crit_5$. The rest 800 experiments are carried at these values of penalty factor.

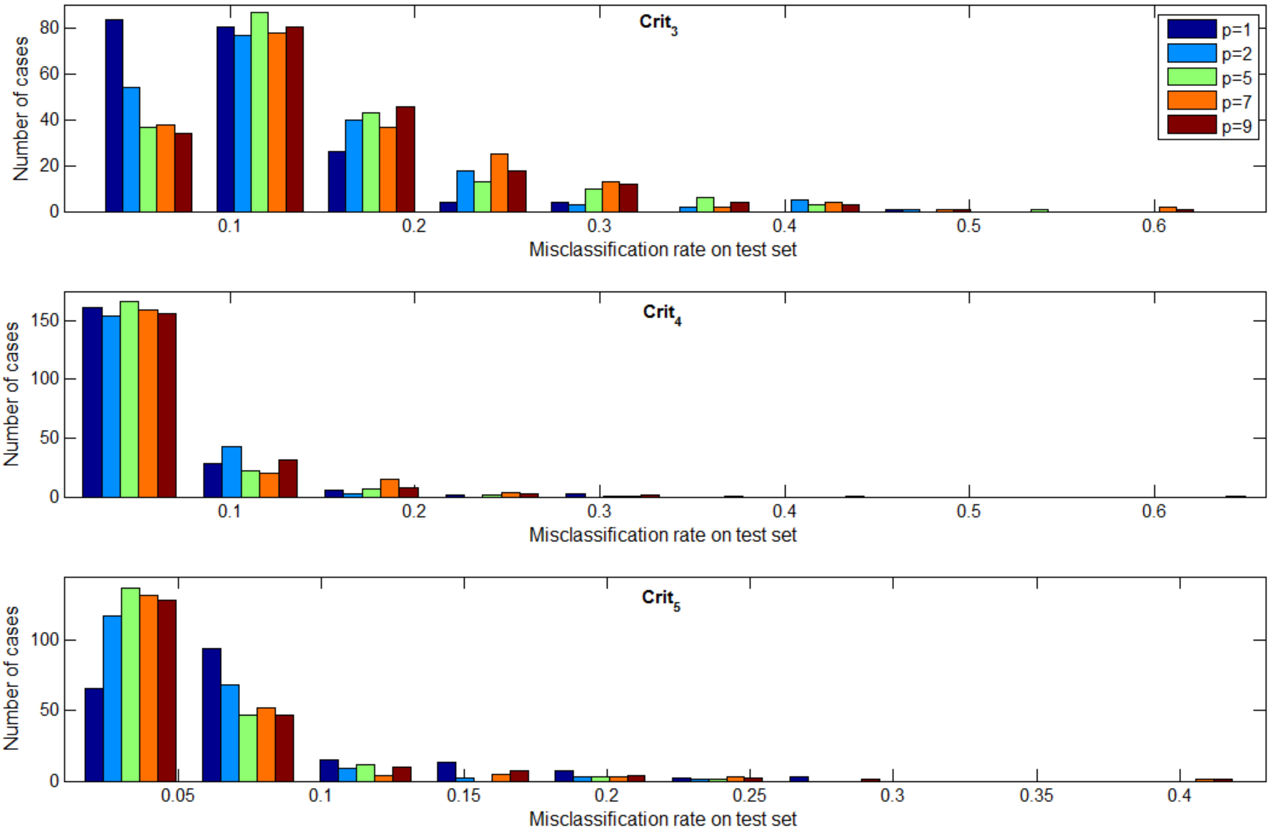

Fig. 3. Influence of penalty factor on correctness of the classifier, in which t-norm is realized by minimum operation

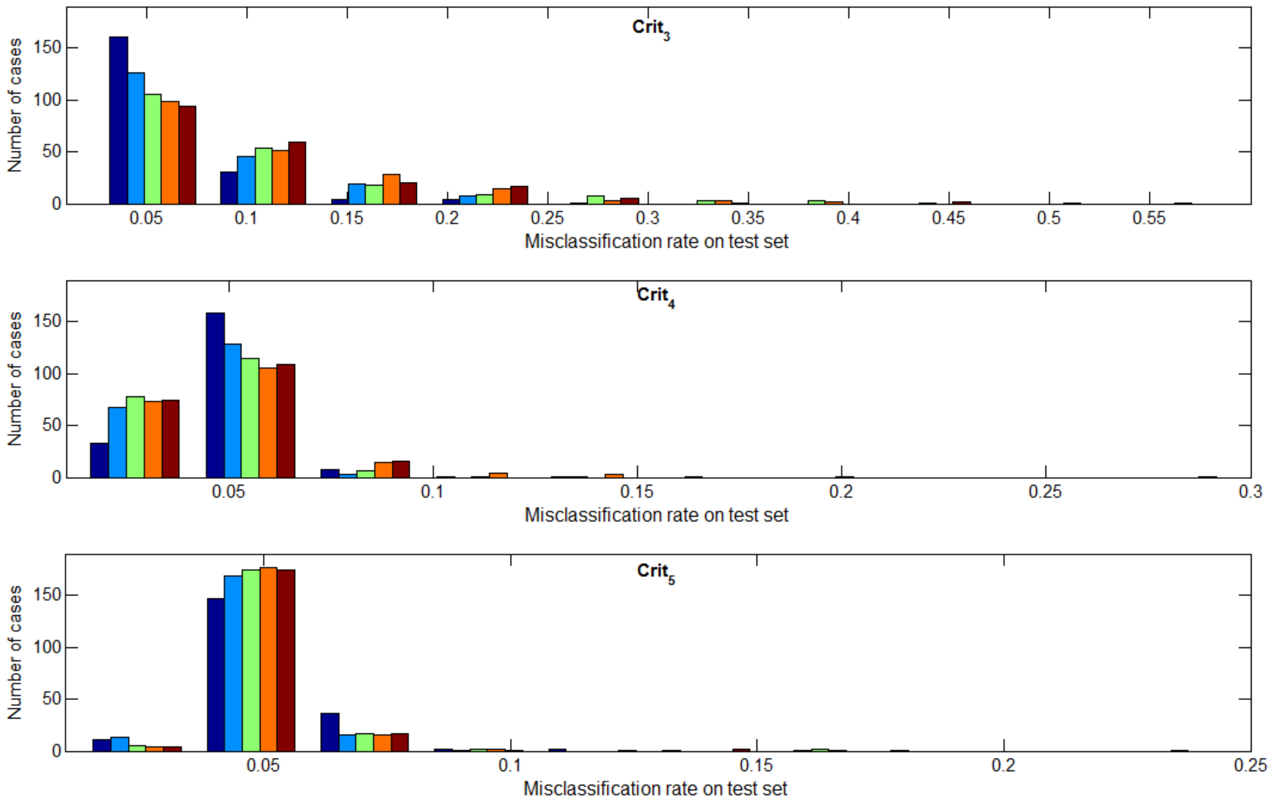

Fig. 4. Influence of penalty factor on correctness of the classifier, in which t-norm is realized by the product operation.

Table 3

Influence of penalty factor in learning criterion on correctness of the classifier on test set (the best results are in bold font

t-norm	Criterion	Mean correctness				
		$p=1$	$p=3$	$p=5$	$p=7$	$p=9$
min	$Crit_3$	0.1050	0.1356	0.1469	0.1560	0.1553
	$Crit_4$	0.0712	0.0645	0.0696	0.0757	0.0721
	$Crit_5$	0.0812	0.0595	0.0572	0.0608	0.0631
prod	$Crit_3$	0.0715	0.0879	0.1091	0.1061	0.1133
	$Crit_4$	0.0546	0.0506	0.0496	0.0528	0.0517
	$Crit_5$	0.0510	0.0446	0.0448	0.0461	0.0460

The results of the experiments prove the correlation of $Crit_1$ – $Crit_5$ criteria values on training set with misclassification rate on test set (fig. 5). Accordingly, these criteria can be applied for learning fuzzy classifier with voting rules. Regarding the quality of learning (Table 4 and Fig. 6, 7), it is significantly better, if new criteria $Crit_4$ and $Crit_5$ are used. New criteria provide better correctness both on average (Table 4) and by the number of the best learning cases. Among new criteria $Crit_5$ has minor advantage.

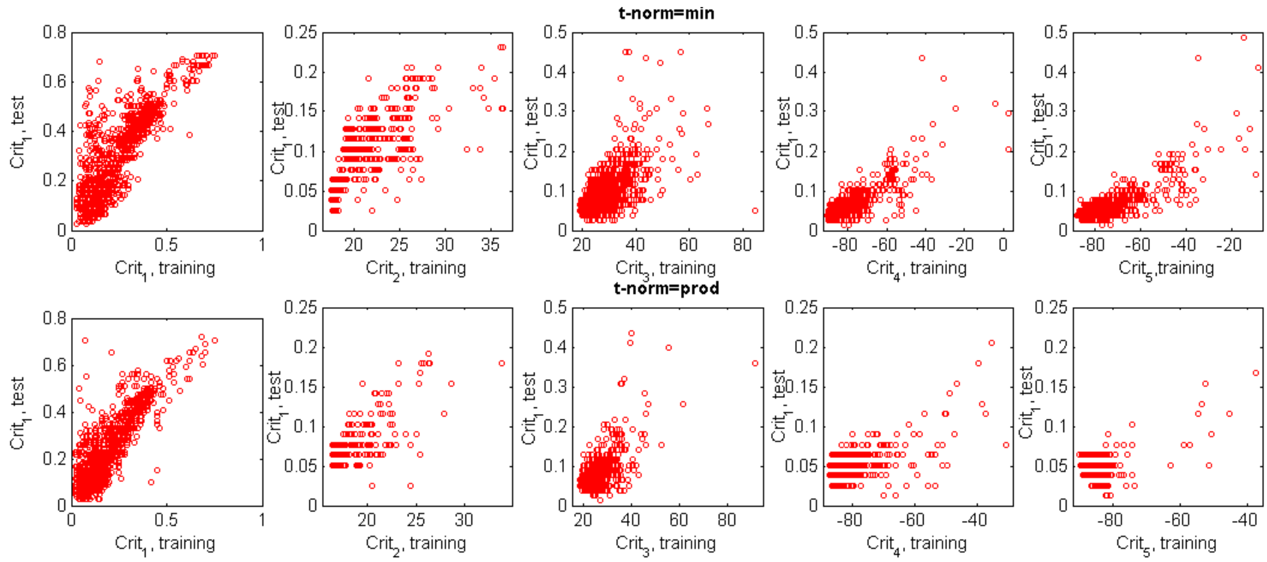

Fig. 5. Distribution of the results of fuzzy classifier learning

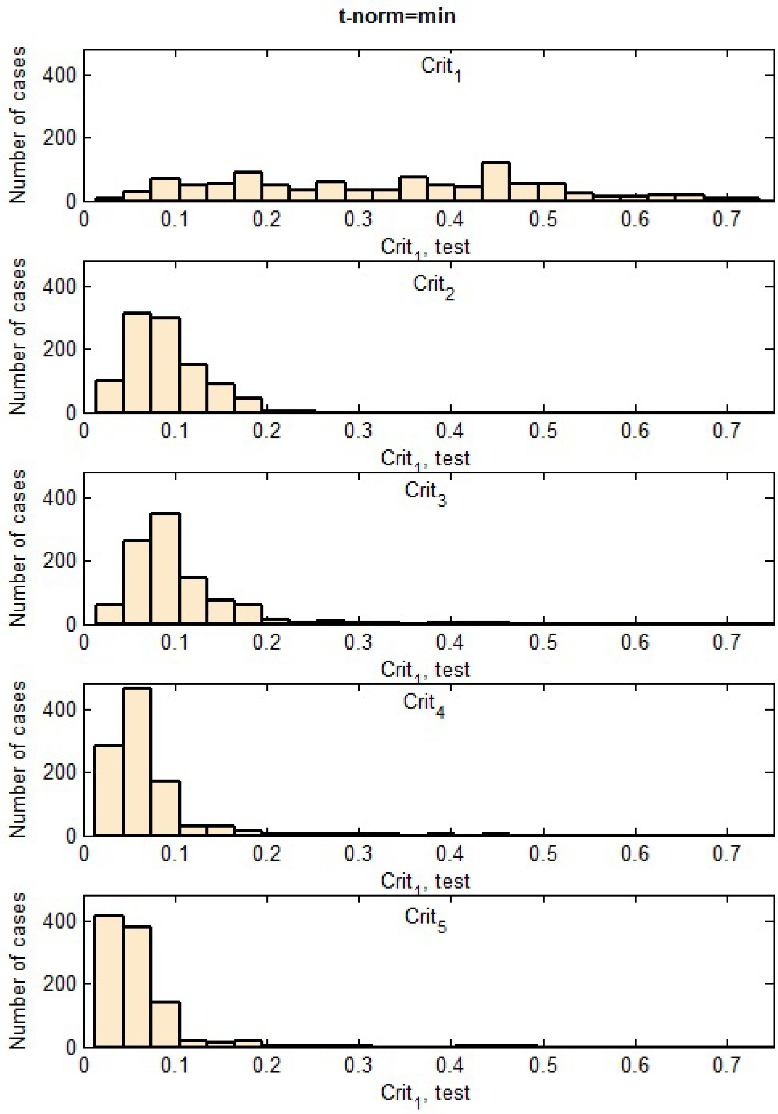

Fig. 6. Distribution of learning quality for classifier with t-norm=min

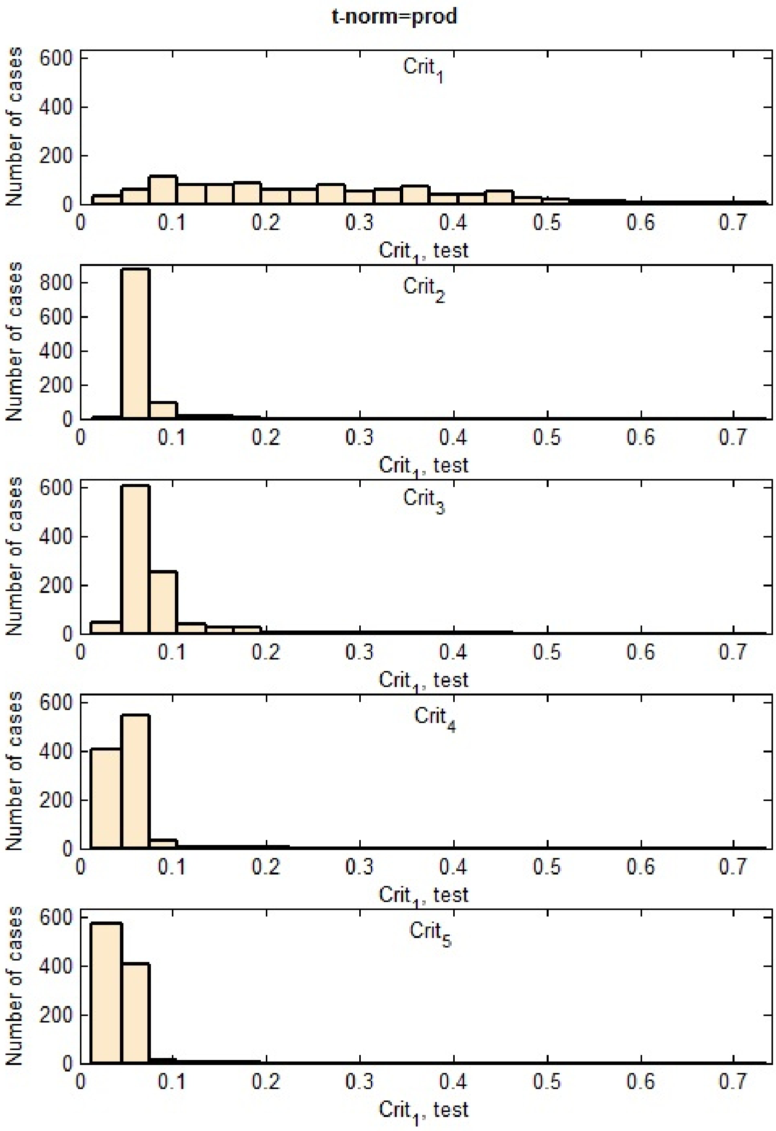

Fig. 7. Distribution of learning quality for classifier with t-norm=prod

Table 4

Statistics of fuzzy classifier learning (the best results are in bold font)

t-norm	Learning criterion	Misclassifications ($Crit_1$) on test sample			
		minimal	mean	median	maximum
min	$Crit_1$	0.0256	0.3222	0.3333	0.7051
	$Crit_2$	0.0256	0.0901	0.0897	0.2308
	$Crit_3$	0.0256	0.1003	0.0897	0.4487
	$Crit_4$	0.0128	0.0638	0.0513	0.4359
	$Crit_5$	0.0128	0.0598	0.0513	0.4872
prod	$Crit_1$	0.0256	0.2518	0.2308	0.7179
	$Crit_2$	0.0256	0.0601	0.0513	0.1923
	$Crit_3$	0.0128	0.0747	0.0641	0.4359
	$Crit_4$	0.0128	0.0496	0.0513	0.2051
	$Crit_5$	0.0128	0.0451	0.0385	0.1667

Conclusions

For the first time learning of fuzzy classifier with voting rules is realized not only by misclassification rate but by other learning criteria. These learning criterion are squared distance between two fuzzy sets – desired and real results of classification; squared distance between desired and real results of classification with additional penalty for wrong decision; distance between main competitors with the penalty for wrong decision. Criteria with distance between main competitors are new and the rest criteria were applied for learning of fuzzy classifier with single winner rule.

Carried computer experiments on tuning fuzzy classifier for UCI-problem about Italian wine recognition showed certain advantage of new learning criteria. Among new learning criteria the criterion in the form of squared distance between main competitors with the penalty for wrong decision has minor advantage.

The given publication is based on the research provided by the grant support of the State Fund For Fundamental Research (project №F62 / 201-2015.).

REFERENCES

1. Kuncheva L. I. Fuzzy classifier design, Studies in Fuzziness and Soft Computing / L. I. Kuncheva . – Vol. 49. – Berlin – Heidelberg: Springer-Verlag, 2000. – 314 p.

2. Rotshtein A. N. Design and tuning of fuzzy rule-based system for medical diagnosis / A. N. Rotshtein, N. H. Teodorescu, , A. Kandel, L. C. Jain // Fuzzy and Neuro-Fuzzy Systems in Medicine. – Boca–Raton : CRC–Press, 1998. - P. 243 – 289.

3. Ishibuchi H. Voting in fuzzy rule-based systems for pattern classification problems / H. Ishibuchi, T. Nakashima, T. Morisawa // Fuzzy Sets and Systems. – 1999. – Vol. 103, №2. – P. 223 – 238.

4. Ishibuchi H. Classification and modeling with linguistic information granules: advanced approaches advanced approaches to linguistic data mining / H. Ishibuchi, T. Nakashima, M. Nii. – Berlin – Heidelberg: Springer-Verlag, 2005. – 307 p.

5. Shtovba S. Tuning the fuzzy classification models with various learning criteria: the case of credit data classification / S. Shtovba, O. Pankevich, G. Dounias // Proc. of Inter. Conference on Fuzzy Sets and Soft Computing in Economics and Finance. St. Petersburg (Russia). St. Petersburg: Russian Fuzzy Systems Association. – 2004. – Vol. 1. – P. 103 – 110.

6. Штовба С. Д. Проектирование нечетких систем средствами MATLAB / С. Д. Штовба. – М. : Горячая линия – Телеком, 2007. – 288 с.

7. Штовба С. Д. Порівняння критеріїв навчання нечіткого класифікатора / С. Д. Штовба // Вісник Вінницького політехнічного інституту. – 2007. – № 6. – С. 84 – 91.

8. Штовба С. Д. Анализ критериев обучения нечеткого классификатора / С. Д. Штовба, О. Д. Панкевич, А. В. Нагорна // Автоматика и вычислительная техника. - 2015. - № 3. - С. 5 - 16.

9. Панкевич О. Д. Діагностування тріщин будівельних конструкцій за допомогою нечітких баз знань. Монографія / О. Д. Панкевич, С. Д. Штовба. – Вінниця: УНІВЕРСУМ – Вінниця, 2005. – 108 с.

10. Штовба С. Д. Обеспечение точности и прозрачности нечеткой модели Мамдани при обучении по экспериментальным данным / С. Д. Штовба // Проблемы управления и информатики. – 2007. – № 4. – С. 102 – 114.

***Shtovba Serhiy* –** Doctor of Sc(Eng.)., Professor, Department of Computer Control Systems, Email: shtovba@ksu.vntu.edu.ua.

***Galushchak Anastasiia* –** Assistant, Department of Computer Control Systems, Email: nastya.nahorna@gmail.com.

Vinnytsia National Technical University.