fuzzy clustering, pairs of prototypes,
fuzzy rule-based classification

Michal JEZEWSKI[1], Robert CZABANSKI[1], Jacek LESKI[1], Krzysztof HOROBA[2]

A NEW APPROACH FOR THE CLUSTERING USING PAIRS OF PROTOTYPES

In the presented work two variants of the fuzzy clustering approach dedicated for determining the antecedents of the rules of the fuzzy rule-based classifier were presented. The main idea consists in adding additional prototypes ('prototypes in between') to the ones previously obtained using the fuzzy c-means method (ordinary prototypes). The 'prototypes in between' are determined using pairs of the ordinary prototypes, and the algorithm based on distances and densities finding such pairs was proposed. The classification accuracy obtained applying the presented clustering approaches was verified using six benchmark datasets and compared with two reference methods.

1. INTRODUCTION

The goal of clustering [3], [7], [13] is to find groups of similar objects in a given dataset. A one of popular clustering methods is the fuzzy c-means (FCM) method [1], where the clustering results are obtained as locations of centers (prototypes) of groups (clusters), and membership values (within the range $[0, 1]$) of each object to each cluster. In case of the fuzzy rule-based classifiers, the FCM method may be applied to determine the antecedents of the rules [2], [8], the new fuzzy clustering method for such a purpose was proposed in [8]. The goal of the presented work is to propose another fuzzy clustering approach – using the FCM method – for determining the antecedents of the fuzzy if-then rules, and to verify its usefulness by the obtained classification quality.

2. THE PROPOSED CLUSTERING PROCEDURE

The goal of the presented clustering approach is to add additional prototypes – lying between prototypes in two classes of objects ('prototypes in between') – to the prototypes found by the FCM method (ordinary prototypes). In our opinion, the rules of the fuzzy rule-based classifier determined using the 'prototypes in between' may improve the classification quality. The 'prototypes in between' are determined basing on pairs of ordinary prototypes. We proposed two variants of the presented approach: FCMpb (FCM with 'prototypes in between') and FCMpbd (FCM with 'prototypes in between' and with the 'density-based rejection'). In the FCMpbd variant some of the ordinary prototypes with low density values are rejected. We suppose, that such rejection may also improve the quality of the classification.

[1]Silesian University of Technology, Institute of Electronics, Akademicka Str. 16, 44-100 Gliwice, Poland
[2]Institute of Medical Technology and Equipment ITAM, Roosevelt Str. 118, 41-800 Zabrze, Poland

The idea of the classification using 'prototypes in between' determined basing on the pairs of ordinary prototypes we proposed previously (e.g. in [4], [5], [6]). However, in the previous research only the 'prototypes in between' were used, and the pairs of prototypes were determined differently than in the presented work.

The proposed approach concerns two-class classification problem, where two classes: ω_1 and ω_2 of objects $\mathbf{x}_k$ ($\mathbf{x}_k = [x_{k1}, x_{k2}, \cdots, x_{kn}]$) were defined as follows: $\mathbf{x}_k \in \omega_1$ when $\Theta_k = +1$, and $\mathbf{x}_k \in \omega_2$ when $\Theta_k = -1$, where Θ_k denotes the class label (equals to $+1$ or -1) of the object $\mathbf{x}_k$. At the beginning of both algorithm variants, classes of objects are separated and each class is clustered using the FCM method with the given value of the fuzziness degree (m) into c clusters. If there are more prototypes in the same location in a given class, only one is left (first rejection). Next, if there are the same prototypes in both classes, all are removed (second rejection). In case of the FCMpbd variant there is a third rejection of prototypes – with respect to their densities. The densities of prototypes in a given class are calculated and the prototypes with the densities lower than a density threshold are rejected. In case of all rejection stages, the appropriate memberships in the partition matrix are removed. The matrix is not scaled, as a result sum of memberships of a given object to all clusters may not be equal to one. After a given rejection of prototypes, if objects which do not member to any cluster are found, they are removed from further steps (i.e. finding the pairs of prototypes, determining 'prototypes in between' and dispersions of all prototypes).

We used three types of density defined as follows:

$$D1_i^{(j)} = \underset{\substack{1 \le i \le c^{(j)} \\ j \in \{1,2\}}}{\forall} \frac{1}{N_i^{(j)}} \sum_{k \in \mathcal{K}_{i\omega_j}} \left\| \mathbf{v}_i^{(j)} - \mathbf{x}_k \right\|, \tag{1}$$

$$D2_i^{(j)} = \underset{\substack{1 \le i \le c^{(j)} \\ j \in \{1,2\}}}{\forall} \frac{1}{N_i^{(j)}} \sum_{k \in \mathcal{K}_{i\omega_j}} u_{ik}^{(j)}, \tag{2}$$

$$D3_i^{(j)} = \underset{\substack{1 \le i \le c^{(j)} \\ j \in \{1,2\}}}{\forall} \frac{1}{N_i^{(j)}} \sum_{k \in \mathcal{K}_{i\omega_j}} u_{ik}^{(j)} \left\| \mathbf{v}_i^{(j)} - \mathbf{x}_k \right\|. \tag{3}$$

The above equations define densities of the i-th prototype in the j-th class ($\mathbf{v}_i^{(j)}$), $u_{ik}^{(j)}$ denotes membership value (provided by the FCM clustering) of the object $\mathbf{x}_k$ (in the j-th class) to the i-th cluster in the j-th class. We used four types of the density threshold: mean, 1st, 2nd and 3rd quartile of densities of all prototypes in the j-th class. The densities are calculated taking into account only the objects with the membership value greater or equal than assumed membership threshold ($T_i^{(j)}$), the sets $\mathcal{K}_{i\omega_1}$ and $\mathcal{K}_{i\omega_2}$ in the above equations contain indexes of such objects, $N_i^{(1)}$ and $N_i^{(2)}$ are the cardinalities of these sets. The sets $\mathcal{K}_{i\omega_1}$ and $\mathcal{K}_{i\omega_2}$ are defined as follows:

$$\mathcal{K}_{i\omega_1} = \left\{ k \,\middle|\, \mathbf{x}_k \in \omega_1 \cap u_{ik}^{(1)} \ge T_i^{(1)} \right\}, \quad \mathcal{K}_{i\omega_2} = \left\{ k \,\middle|\, \mathbf{x}_k \in \omega_2 \cap u_{ik}^{(2)} \ge T_i^{(2)} \right\}. \tag{4}$$

Five types of the membership threshold were used: 0, mean, 1st, 2nd and 3rd quartile of all membership values to the i-th cluster in the j-th class. The thresholds are always calculated using the up-to-date partition matrix. In case of the membership threshold equals to 0 all current objects in a given class are used.

The 'prototypes in between' are determined basing on the pairs of ordinary prototypes and the algorithm finding such pairs was proposed. Since the are two stages of rejection of prototypes (three in the FCMpbd variant), the final numbers of prototypes in each class ($c^{(1)}$ in the ω_1 class, $c^{(2)}$ in the ω_2 class) may be different. Hence, the maximum number of pairs of prototypes possible to found (N_p) is equal to $min(c^{(1)}, c^{(2)})$. The algorithm for finding pairs of ordinary prototypes may be presented as follows:

1) Fix the 'window' length W and the density type.
2) If $c^{(1)} < c^{(2)}$ then assign ω_1 (ω_2) class as the 1st (2nd) class, if $c^{(1)} > c^{(2)}$ then assign ω_2 (ω_1) class as the 1st (2nd) class. (In the algorithm the 1st (2nd) class denotes the class with the smaller (greater) number of prototypes.)
3) Calculate the densities of the prototypes in the 1st and in the 2nd class.
4) Sort the prototypes in the 1st class in descending order according to their densities.
5) Set $i = 1$, $j = c^{(2)}$.
6) From the sorted prototypes choose the ith one.
7) If $W > j$ then set $W = j$, for the chosen prototype find W nearest prototypes (using the Euclidean distance) in the 2nd class.
8) From the W nearest prototypes choose the one with the highest density, if there are more than one such prototypes choose the first one (i.e. the closest one).
   The prototypes found in the step 6) (in the 1st class) and in the step 8) (in the 2nd class) form the pair.
9) If $i = c^{(1)}$ then stop, else set $i = i + 1$, $j = j - 1$, go to the step 6) and in the step 7) omit the prototypes found previously in the 2nd class.

The proposed algorithm is based on distances and densities, the density imposes the order of finding the pairs. For $W = 1$ the prototypes chosen in the 2nd class are the closest to the prototypes selected – in the order imposed by the densities – in the 1st class. Since three types of the density were assumed, three types of the above algorithm were obtained. If $c^{(1)} = c^{(2)}$ when the above algorithm starts, then it is repeated twice – assigning ω_1 class and next ω_2 class as the 1st class, and the solution leading to the shorter mean length of pairs (using the Euclidean distance) is chosen. In case of equal mean length, the solution obtained assigning ω_1 class as the 1st class is chosen.

The number of 'prototypes in between' ($c^{(b)}$) was set as equal to $\lfloor (c/2) \rfloor$. If the number of found pairs of prototypes (N_p) is greater than $c^{(b)}$, then $c^{(b)}$ shortest pairs is chosen, otherwise all N_p pairs are used. The remained (after the rejections) ordinary prototypes $\mathbf{v}_i^{(1)}$ ($i = 1, 2, \cdots, c^{(1)}$) and $\mathbf{v}_i^{(2)}$ ($i = 1, 2, \cdots, c^{(2)}$) are the result of the FCM clustering. The 'prototypes in between' are determined basing on the obtained pairs of ordinary prototypes using the following formula:

$$\underset{1 \le i \le c^{(b)}}{\forall} \quad \mathbf{v}_i^{(b)} = \frac{\sum\limits_{k \in \mathcal{K}_{r\omega_1}} \left(u_{rk}^{(1)}\right)^m \mathbf{x}_k + \sum\limits_{k \in \mathcal{K}_{t\omega_2}} \left(u_{tk}^{(2)}\right)^m \mathbf{x}_k}{\sum\limits_{k \in \mathcal{K}_{r\omega_1}} \left(u_{rk}^{(1)}\right)^m + \sum\limits_{k \in \mathcal{K}_{t\omega_2}} \left(u_{tk}^{(2)}\right)^m}, \tag{5}$$

where: r and t denote the indexes of the ordinary prototypes making pairs, i.e. $\mathbf{v}_r^{(1)}$ and $\mathbf{v}_t^{(2)}$, the sets $\mathcal{K}_{r\omega_1}$ and $\mathcal{K}_{t\omega_2}$ contain indexes of the objects with the membership value greater or equal than the assumed membership threshold:

$$\mathcal{K}_{r\omega_1} = \left\{ k \,\middle|\, \mathbf{x}_k \in \omega_1 \cap u_{rk}^{(1)} \ge T_r^{(1)} \right\}, \quad \mathcal{K}_{t\omega_2} = \left\{ k \,\middle|\, \mathbf{x}_k \in \omega_2 \cap u_{tk}^{(2)} \ge T_t^{(2)} \right\}. \tag{6}$$

The $c^{(b)}$ 'prototypes in between' along with $\left(c^{(1)} + c^{(2)}\right)$ ordinary prototypes form $K = c^{(1)} + c^{(2)} + c^{(b)}$ prototypes ($\mathbf{v}$), which components are the centers of the Gaussian membership functions in the antecedents of the K fuzzy if-then rules of the classifier presented in the next section. The dispersions ($\mathbf{s}$) of the Gaussian membership functions are calculated using the formulas (7) – for the ordinary prototypes, and (8) – for the 'prototypes in between':

$$\underset{\substack{1 \le i \le c^{(j)} \\ j \in \{1,2\}}}{\forall} \quad \left[\mathbf{s}_i^{(j)}\right]^{(\bullet 2)} = \frac{\sum\limits_{k \in \mathcal{K}_{i\omega_j}} u_{ik}^{(j)} \left[\mathbf{x}_k - \mathbf{v}_i^{(j)}\right]^{(\bullet 2)}}{\sum\limits_{k \in \mathcal{K}_{i\omega_j}} u_{ik}^{(j)}}, \tag{7}$$

$$\underset{1 \le i \le c^{(b)}}{\forall} \quad \left[\mathbf{s}_i^{(b)}\right]^{(\bullet 2)} = \frac{\sum\limits_{\mathcal{K}_{r\omega_1}} u_{rk}^{(1)} \left[\mathbf{x}_k - \mathbf{v}_r^{(1)}\right]^{(\bullet 2)} + \sum\limits_{\mathcal{K}_{t\omega_2}} u_{tk}^{(2)} \left[\mathbf{x}_k - \mathbf{v}_t^{(2)}\right]^{(\bullet 2)}}{\sum\limits_{\mathcal{K}_{r\omega_1}} u_{rk}^{(1)} + \sum\limits_{\mathcal{K}_{t\omega_2}} u_{tk}^{(2)}}. \tag{8}$$

The symbol $^{(\bullet 2)}$ denotes component-by-component squaring $[x_{k1}, x_{k2}, \cdots, x_{kn}]^{(\bullet 2)} = [x_{k1}^2, x_{k2}^2, \cdots, x_{kn}^2]$. Each FCM clustering providing the ordinary prototypes was started from the initial prototypes selected from the boundary of the convex hull [8] of the class being clustered. The clustering was performed as long as the change between successive values of the FCM criterion function was greater or equal than 10^{-4}. The maximum number of iterations was established at 500, distances between objects and prototypes less than 10^{-10} were treated as equal to 0, and then a special update of the partition matrix was used.

3. Fuzzy rule-based classifier

To perform the classification we applied the fuzzy rule-based classifier [8] that uses the fuzzy if-then rules with the Gaussian membership functions in the antecedents and singletons in the consequents. The final output of such classifier for the ℓth object is given with the formula:

$$y_{0\ell} = \mathcal{F}(\mathbf{x}_\ell) = \frac{\sum_{k=1}^{K} \mu_{\mathbf{A}_k}(\mathbf{x}_\ell)\, y_k}{\sum_{k=1}^{K} \mu_{\mathbf{A}_k}(\mathbf{x}_\ell)}, \tag{9}$$

where: y_k stands for the crisp value provided by the k-th rule, and

$$\underset{1 \le k \le K}{\forall} \mu_{\mathbf{A}_k}(\mathbf{x}_\ell) = \exp\left[-\frac{1}{2}\sum_{j=1}^{n}\left(\frac{x_{\ell j} - v_{kj}}{\delta s_{kj}}\right)^2\right] \tag{10}$$

denotes membership function of the antecedent of the k-th rule. The parameter δ is used for scaling the dispersions. The equation (9) may be written in the form:

$$y_{0\ell} = \mathcal{F}(\mathbf{x}_\ell) = \sum_{k=1}^{K} \overline{\mu_{\mathbf{A}_k}}(\mathbf{x}_\ell)\, y_k = \mathbf{g}(\mathbf{x}_\ell)^{\mathsf{T}} \mathbf{y}, \tag{11}$$

where:

$$\overline{\mu_{\mathbf{A}_k}}(\mathbf{x}_\ell) = \frac{\mu_{\mathbf{A}_k}(\mathbf{x}_\ell)}{\sum_{k=1}^{K} \mu_{\mathbf{A}_k}(\mathbf{x}_\ell)} \tag{12}$$

denotes the normalized activation of the rule, $\mathbf{g}(\mathbf{x}_\ell)^{\mathsf{T}} = [\overline{\mu_{\mathbf{A}_1}}(\mathbf{x}_\ell), \overline{\mu_{\mathbf{A}_2}}(\mathbf{x}_\ell), \ldots, \overline{\mu_{\mathbf{A}_K}}(\mathbf{x}_\ell)]$, and $\mathbf{y}^{\mathsf{T}} = [y_1, y_2, \ldots, y_K]$. Defining $\mathbf{G}^{\mathsf{T}} = [\Theta_1 \mathbf{g}(\mathbf{x}_1), \Theta_2 \mathbf{g}(\mathbf{x}_2), \cdots, \Theta_N \mathbf{g}(\mathbf{x}_N)]$ as the $N \times K$ matrix (N denotes the number of objects in the training set), the vector $\mathbf{y}$ such that:

$$\mathcal{F}(\mathbf{x}_\ell) = \mathbf{g}(\mathbf{x}_\ell)^{\mathsf{T}} \mathbf{y} \begin{cases} > 0, & \mathbf{x}_\ell \in \omega_1, \\ \le 0, & \mathbf{x}_\ell \in \omega_2. \end{cases} \tag{13}$$

is found according to [8] using the 'ASQR' loss function [8]. If the sum $\sum_{k=1}^{K} \mu_{\mathbf{A}_k}(\mathbf{x}_\ell)$ was equal to 0, then $\mu_{\mathbf{A}_k}(\mathbf{x}_\ell) = 10^{-6}$ for $k = 1, 2, \cdots, K$ was established.

4. Results and discussion

To verify the classification quality obtained by the fuzzy rule-based classifier with the proposed clustering approach we applied six benchmark datasets. Five of them – Breast-cancer (BRE), Diabetis (DIA), Heart (HEA), Thyroid (THY) and Titanic (TIT) – are described in [10], [11], and were obtained from http://ida.first.fraunhofer.de/projects/bench with the divisions into 100 training and testing sets. The sixth dataset is the synthetic dataset (SYN) generated by Ripley [12], with the division into 100 training and testing sets the same as in [8]. The obtained classification quality was compared with the results from [8] (for the 'ASQR' loss function) and with the Lagrangian SVM (LSVM) method [9] (obtained from http://www.cs.wisc.edu/dmi/lsvm) with the Gaussian kernel $K(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\chi \|\mathbf{x} - \mathbf{x}_i\|^2\right)$; $\chi \in \mathbb{R}_+$. Both LSVM parameters (ν and χ) values providing the best classification quality were searched within the set $\{0.00001, 0.00004, 0.00007, 0.0001, 0.0004, \cdots, 70000, 100000\}$ using 10 first pairs of training and testing sets, the rest of the LSVM parameters were set to default values [9]. For the LSVM the processed data were normalized to the range $[-1, +1]$.

The parameters of the fuzzy rule-based classifier with the proposed clustering (FCMpb or FCMpbd) were searched within the following ranges:

- the fuzziness degree in the FCM clustering (m, searched within the set $\{1.1, 1.5, 2\}$),
- the initial number of clusters per class (c, changed from 2 to 16 with the step of 1),
- the density type (1 of 3: $D1$, $D2$, $D3$),
- the type of the membership threshold (1 of 5: M_0 (0), M_m (mean), M_{q1} (1st quartile), M_{q2} (2nd quartile), M_{q3} (3rd quartile)),
- the type of the density threshold (only in the FCMpbd variant, 1 of 4: D_m (mean), D_{q1} (1st quartile), D_{q2} (2nd quartile), D_{q3} (3rd quartile)),
- the W in the algorithm finding pairs (changed from 1 to 5 with the step of 1),
- the parameter δ scaling the dispersions (changed from 0.2 to 1.6 with the step of 0.1).

As the maximum number of the initial number of clusters per class (c) was established at 16, the maximum number of all obtained prototypes (K) is equal to $16 \times 2 + \lfloor (16/2) \rfloor = 40$. If after the clustering with a given combination of the values of the above parameters (excluding δ) the numbers of prototypes were to small (i.e. $c^{(1)} < 2$ or $c^{(2)} < 2$) then pairs of prototypes were not being found and such clustering was excluded from further analysis.

The 10 first pairs of training and testing sets were merged into single dataset which was being clustered with a given combination of the values of the above parameters (excluding δ). This way the antecedents of the rules (centers and dispersions of the Gaussian membership functions) were obtained, and they were common for all 100 training and testing sets. The consequents (vector $\mathbf{y}$) were being determined separately for each of the 10 training sets. The combination of the values of all above parameters providing the highest classification quality for the first 10 testing sets was chosen to calculate the final result for 100 testing sets (with the consequents being determined separately for each of all 100 training sets). The 100 training and testing sets were not modified, i.e. they contained all objects. In the classification process (in case of both the fuzzy rule-based classifier and the LSVM), if the classifier output value was > 0 (≤ 0), then the ω_1 (ω_2) class was assigned.

Tables 1 and 2 present the obtained classification qualities for both variants of the proposed approach – without (FCMpb) and with (FCMpbd) the 'density-based rejection', and for the reference procedures: the $(c + p)$-means [8] and the LSVM. Each cell contains mean value and standard deviation of the misclassification error for all 100 testing sets (top) and values of parameters providing them (c_f denotes the final number of prototypes in total). The best result for each dataset is in boldface, separately for both variants. In case of BRE, DIA and especially HEA, the FCMpbd variant provided higher misclassification error for all types of the density. For SYN, THY and TIT the FCMpbd variant resulted in higher or lower misclassification error, depending on the density type, but provided the lowest misclassification error for these datasets. However, it should be emphasized, that the FCMpbd variant is characterized by higher computing time than FCMpb variant since it has additional parameter – the density threshold. Omitting this variant worsens the best results for SYN, THY and TIT, however the differences seem to be small (0.31% for SYN, 0.02% for THY and 0.03% for TIT). Taking into account the density type and the classification quality, it is difficult to indicate the density type leading to the highest classification quality – it depends on the dataset and the applied variant. Both $(c+p)$-means and the presented approaches provided lower misclassification error than LSVM. We could not get the better results than the $(c+p)$-means only in case of the HEA and SYN datasets. For the remaining datasets both FCMpb and FCMpbd based on the third type of the density provided better classification accuracy. When considering the classification efficacy defined as mean classification error calculated for all databases (Table 3), the best results were obtained for the FCMpb variant and the density type 3.

Table 1. The classification error rates in the FCMpb variant.

Data	LSVM	$(c+p)$-means	FCMpb$D1$	FCMpb$D2$	FCMpb$D3$
BRE	24.27 (3.95) $\nu = 0.04$ $\chi = 0.4$	21.42 (4.00) $m = 1.1, c = 8$ $c_f = 16, \delta = 1.2$	17.35 (4.04) $m = 1.1, c = 16, c_f = 40$ $M_m, W = 4, \delta = 0.8$	16.29 (3.56) $m = 1.1, c = 16, c_f = 40$ $M_{q3}, W = 3, \delta = 0.9$	**15.69** (3.41) $m = 1.1, c = 16, c_f = 40$ $M_{q1}, W = 4, \delta = 0.8$
DIA	23.00 (1.78) $\nu = 1000$ $\chi = 0.004$	21.71 (1.79) $m = 1.1, c = 19$ $c_f = 38, \delta = 1.6$	**19.71** (1.86) $m = 1.1, c = 13, c_f = 32$ $M_{q1}, W = 5, \delta = 1.1$	20.13 (1.88) $m = 1.1, c = 13, c_f = 32$ $M_0, W = 5, \delta = 1.2$	20.13 (1.92) $m = 1.1, c = 13, c_f = 32$ $M_0, W = 1, \delta = 1.3$
HEA	16.33 (2.67) $\nu = 10000$ $\chi = 0.0001$	**6.16** (2.29) $m = 1.1, c = 16$ $c_f = 32, \delta = 0.2$	6.65 (2.22) $m = 1.1, c = 16, c_f = 40$ $M_m, W = 4, \delta = 0.2$	7.68 (2.88) $m = 1.1, c = 16, c_f = 40$ $M_m, W = 3, \delta = 0.2$	6.28 (2.44) $m = 1.1, c = 16, c_f = 40$ $M_m, W = 5, \delta = 0.2$
SYN	9.54 (0.60) $\nu = 0.7$ $\chi = 7.0$	**8.55** (0.47) $m = 1.1, c = 9$ $c_f = 18, \delta = 0.2$	9.14 (0.95) $m = 1.1, c = 7, c_f = 17$ $M_m, W = 1, \delta = 0.4$	9.33 (0.76) $m = 2, c = 5, c_f = 12$ $M_0, W = 2, \delta = 0.3$	8.92 (0.79) $m = 1.1, c = 5, c_f = 12$ $M_0, W = 4, \delta = 0.3$
THY	4.21 (2.11) $\nu = 10.0$ $\chi = 4.0$	1.51 (1.64) $m = 1.1, c = 7$ $c_f = 14, \delta = 0.9$	1.63 (1.72) $m = 1.1, c = 13, c_f = 32$ $M_{q2}, W = 3, \delta = 1.2$	2.09 (2.30) $m = 1.1, c = 13, c_f = 32$ $M_0, W = 3, \delta = 1.4$	**1.47** (1.71) $m = 1.1, c = 16, c_f = 40$ $M_{q3}, W = 3, \delta = 1.5$
TIT	22.90 (1.33) $\nu = 0.01$ $\chi = 4000$	22.42 (1.21) $m = 1.1, c = 3$ $c_f = 6, \delta = 0.5$	21.95 (1.05) $m = 1.5, c = 13, c_f = 29$ $M_0, W = 1, \delta = 0.7$	**21.94** (0.71) $m = 1.5, c = 4, c_f = 10$ $M_m, W = 2, \delta = 1.1$	22.17 (1.18) $m = 2, c = 10, c_f = 25$ $M_m, W = 2, \delta = 0.7$

The fuzziness degree (m) in the FCM method is usually set to 2. In our research we changed the value of m, including $m = 1.1$ as it was proposed in [8]. Analyzing the values of m in the Tables 1 and 2 it is worth to notice, that for BRE, DIA, HEA (both variants) and THY (FCMpb variant) the best results were always obtained for $m = 1.1$. Table 4 presents the classification quality using the FCMpb variant obtained for $m = 2$ (for the rest of parameters the best values were selected). The results for the mentioned datasets were worse, especially for HEA and THY, where about two times higher misclassification error was observed. Taking the above into account we suppose, that the value of the m has a significant influence on the obtained classification quality.

Figure 1 presents objects of the first testing set of the two-dimensional SYN dataset, and the discrimination curve obtained for the first training set using the values of parameters providing the best result for SYN, i.e. FCMpbd variant and 2nd density type. The initial number of clusters

Table 2. The classification error rates in the FCMpbd variant.

Data	LSVM	$(c+p)$-means	FCMpbd$D1$	FCMpbd$D2$	FCMpbd$D3$
BRE	24.27 (3.95) $\nu = 0.04$ $\chi = 0.4$	21.42 (4.00) $m = 1.1, c = 8$ $c_f = 16, \delta = 1.2$	**17.61** (3.83) $m = 1.1, c = 10, c_f = 21$ $M_0, D_{q1}, W = 5, \delta = 0.3$	18.34 (3.93) $m = 1.1, c = 16, c_f = 32$ $M_0, D_{q1}, W = 2, \delta = 0.8$	19.08 (3.90) $m = 1.1, c = 16, c_f = 32$ $M_{q1}, D_{q1}, W = 2, \delta = 1.0$
DIA	23.00 (1.78) $\nu = 1000$ $\chi = 0.004$	21.71 (1.79) $m = 1.1, c = 19$ $c_f = 38, \delta = 1.6$	21.27 (2.14) $m = 1.1, c = 16, c_f = 32$ $M_0, D_{q1}, W = 4, \delta = 0.9$	**20.99** (1.61) $m = 1.1, c = 13, c_f = 26$ $M_0, D_{q1}, W = 2, \delta = 1.2$	21.06 (1.98) $m = 1.1, c = 16, c_f = 32$ $M_m, D_{q1}, W = 1, \delta = 1.2$
HEA	16.33 (2.67) $\nu = 10000$ $\chi = 0.0001$	**6.16** (2.29) $m = 1.1, c = 16$ $c_f = 32, \delta = 0.2$	11.73 (3.01) $m = 1.1, c = 11, c_f = 21$ $M_m, D_{q1}, W = 2, \delta = 0.8$	10.40 (2.94) $m = 1.1, c = 10, c_f = 21$ $M_{q1}, D_{q1}, W = 4, \delta = 0.8$	9.98 (2.67) $m = 1.1, c = 12, c_f = 24$ $M_0, D_{q1}, W = 4, \delta = 0.4$
SYN	9.54 (0.60) $\nu = 0.7$ $\chi = 7.0$	**8.55** (0.47) $m = 1.1, c = 9$ $c_f = 18, \delta = 0.2$	8.93 (0.76) $m = 1.5, c = 6, c_f = 13$ $M_{q3}, D_{q1}, W = 3, \delta = 0.2$	8.61 (0.68) $m = 2, c = 6, c_f = 10$ $M_m, D_m, W = 4, \delta = 0.2$	8.94 (0.88) $m = 2, c = 8, c_f = 16$ $M_m, D_{q1}, W = 4, \delta = 0.2$
THY	4.21 (2.11) $\nu = 10.0$ $\chi = 4.0$	1.51 (1.64) $m = 1.1, c = 7$ $c_f = 14, \delta = 0.9$	1.77 (1.33) $m = 1.5, c = 5, c_f = 8$ $M_{q2}, D_{q2}, W = 2, \delta = 0.9$	2.89 (2.21) $m = 1.5, c = 14, c_f = 12$ $M_{q3}, D_{q3}, W = 2, \delta = 1.6$	**1.45** (1.24) $m = 1.5, c = 11, c_f = 21$ $M_{q3}, D_{q1}, W = 1, \delta = 0.3$
TIT	22.90 (1.33) $\nu = 0.01$ $\chi = 4000$	22.42 (1.21) $m = 1.1, c = 3$ $c_f = 6, \delta = 0.5$	22.15 (1.17) $m = 2, c = 10, c_f = 21$ $M_m, D_{q1}, W = 3, \delta = 0.9$	**21.91** (0.87) $m = 1.5, c = 7, c_f = 13$ $M_{q3}, D_{q1}, W = 1, \delta = 0.3$	22.27 (1.26) $m = 1.5, c = 12, c_f = 16$ $M_m, D_{q2}, W = 2, \delta = 0.5$

Table 3. The mean classification error rates.

LSVM	$(c+p)$-means	FCMpb$D1$	FCMpb$D2$	FCMpb$D3$	FCMpbd$D1$	FCMpbd$D2$	FCMpbd$D3$
16.71	13.63	12.74	12.91	**12.44**	13.91	13.86	13.80

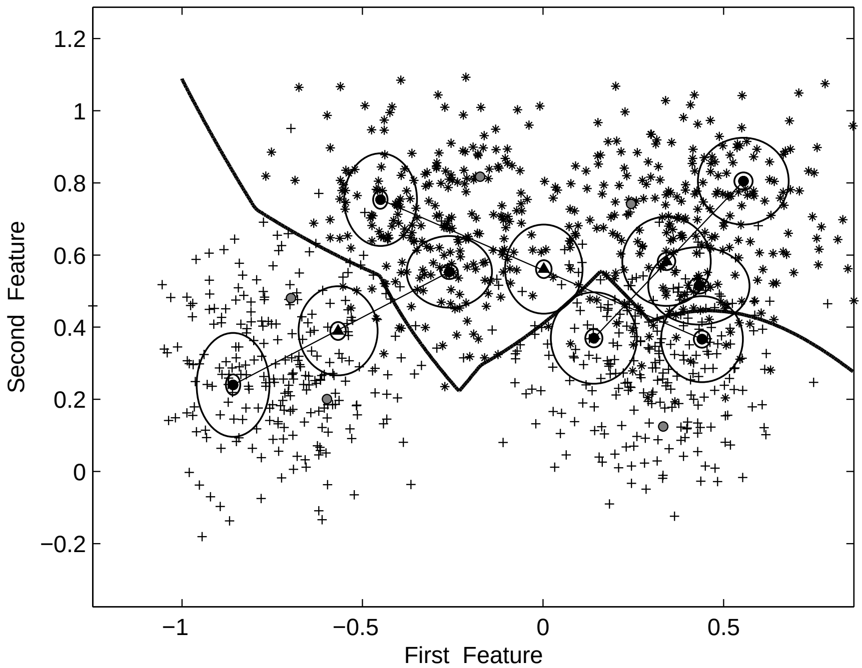

Fig. 1. The discrimination curve obtained for the two-dimensional synthetic SYN dataset.

per class was equal to 6, however 3 (2) prototypes in the ω_1 (ω_2) class were rejected due to the low density. The rejected prototypes are marked by gray circles, the remained ones by black

Table 4. The classification error rates in the FCMpb variant for $m = 2$.

	FCMpb$D1$	FCMpb$D2$	FCMpb$D3$		FCMpb$D1$	FCMpb$D2$	FCMpb$D3$
BRE	22.03 (3.93)	22.03 (3.93)	22.03 (3.93)	SYN	9.19 (0.68)	9.33 (0.76)	9.29 (0.74)
DIA	22.32 (1.65)	22.32 (1.65)	22.45 (1.65)	THY	2.81 (2.28)	3.09 (2.34)	3.09 (2.34)
HEA	14.30 (3.22)	13.89 (3.16)	14.39 (3.53)	TIT	22.14 (1.17)	22.11 (1.14)	22.17 (1.18)

circles. The obtained pairs of prototypes are denoted by lines, the 'prototypes in between' are marked by triangles. The ellipses visualize dispersions: the ones with the greater radius denote the dispersions calculated basing on the obtained prototypes, i.e. using equations (7) and (8), the ones with lower radius represent the dispersions scaled by the parameter $\delta = 0.2$, which was found as providing the best classification quality for SYN in the FCMpbd$D2$ variant.

5. Conclusions

In the paper we presented the clustering approach using the FCM method and dedicated for determining the antecedents of the rules of the fuzzy rule-based classifier. For all six applied benchmark datasets the obtained classification quality was higher than provided by the Lagrangian SVM method. For four datasets the results were also better comparing to the another clustering-based method. Two variants of the presented approach with three types of the prototypes density were examined. It is not possible to indicate the single approach leading to the best results for all applied datasets. However, taking into account mean classification quality for all datasets, the variant without rejection of prototypes with respect to their densities that combine membership values and distances may be regarded as the best. A rather large number of parameters which values should be found may be regarded as a disadvantage.

Acknowledgement

This work was partially supported by the Ministry of Science and Higher Education funding for statutory activities of young researchers (decision no. 8686/E-367/M/2015 of 12 March 2015).

Bibliography

[1] Bezdek J. C. Pattern recognition with fuzzy objective function algorithms. 1981. Plenum.

[2] Czabanski R. Deterministic annealing integrated with ε-insensitive learning in neuro-fuzzy systems. Proceedings of 8th International Conference on Artificial Intelligence and Soft Computing ICAISC 2006, Lecture Notes in Artificial Intelligence 4029, 2006. Springer-Verlag, pp. 220–229.

[3] Doring C., Lesot M.-J., Kruse R. Data analysis with fuzzy clustering methods. Computational Statistics & Data Analysis, 2006, Vol. 51. pp. 192–214.

[4] Jezewski M., Leski J. M. Clustering algorithm for classification methods. Journal of Medical Informatics and Technologies, 2012, Vol. 20. pp. 11–18.

[5] Jezewski M., Leski J. M. Nonlinear extension of the IRLS classifier using clustering with pairs of prototypes. Proc. of the 8th Int. Conf. on Comp. Recog. Sys. CORES 2013, Advances in Intelligent Systems and Computing 226, 2013. Springer Int. Pub. Switzerland, pp. 121–130.

[6] Jezewski M., Leski J. M. Application of the conditional fuzzy clustering with prototypes pairs to classification. Man-Machine Interactions 3, Advances in Intelligent Systems and Computing 242, 2014. Springer Int. Pub. Switzerland, pp. 397–405.

[7] Kruse R., Doring C., Lesot M.-J. Fundamentals of fuzzy clustering. Advances in fuzzy clustering and its applications, 2007. John Wiley & Sons, pp. 3–30.

[8] Leski J. M. Fuzzy $(c+p)$-means clustering and its application to a fuzzy rule-based classifier: toward good generalization and good interpretability. IEEE Transactions on Fuzzy Systems, 2015, Vol. 23 (4). pp. 802–812.

[9] MANGASARIAN O. L., MUSICANT D. R. Lagrangian support vector machines. Journal of Machine Learning Research, 2001, Vol. 1. pp. 161–167.

[10] MIKA S., RATSCH G., WESTON J., SCHOLKOPF B., MULLER K.-R. Fisher discriminant analysis with kernels. Neural Networks for Signal Processing IX, 1999. pp. 41–48.

[11] RATSCH G., ONODA T., MULLER K.-R. Soft margins for adaboost. Machine Learning, 2001, Vol. 42. pp. 287–320.

[12] RIPLEY B. D. Pattern recognition and neural networks. 1996. Cambridge University Press.

[13] XU R., WUNSCH, II, D. C. Clustering. 2009. John Wiley & Sons.