Theoretical and Applied Informatics
ISSN 1896–5334
Vol.23 (2011), no. 2
pp. 107–126
DOI: 10.2478/v10179-011-0008-6

# Merging of Fuzzy Models for Neuro-fuzzy Systems

KRZYSZTOF SIMIŃSKI

Institute of Informatics,
Silesian University of Technology,
ul. Akademicka 16, 44-100 Gliwice, Poland
*Krzysztof.Siminski@polsl.pl*

*Received 12 April 2011, Revised 6 June 2011, Accepted 13 July 2011*

**Abstract:** The merging of fuzzy model is widely used for reduction of rule number in fuzzy model. The supernumerosity of rules is mainly caused by grid partition of input domain. In the paper different cause for model merging is described. It is the need for creation of fuzzy model for large data set. In our solution the models are build basing data subset and then the submodels are merged into one. This approach enables quicker elaboration of submodels with relatively good knowledge generalisation ability without waiting for the whole data set to be processed. With passing time, the subsequent submodels are created and merged to create the better model.

**Keywords:** neuro-fuzzy, fuzzy set, rule merging, similarity, annbfis

## 1. Introduction

Merging of fuzzy models may be encountered in three main situations. The first case are the large data sets. Extraction of fuzzy models for large data sets may be executed (for memory reasons) partially for subsets of data and then the elaborated models need merging to obtain one model. The second one is the incremental input of data. Basing of recently acquired data a model is created that is later combined with yet existing one. The third reason is the need for simplification of model (mainly by reducing the number of rules). The objective of the merge-and-learn stategy is not the optimisation of precision, but the production of small rule base with reasonable precision [21].

Most papers on rules merging focus on the problem of supernumeruous fuzzy rules in the model. This is mainly caused by applying grid partition of the input domain while extracting the rules from the presented data.

In [18] the problem of rule base simplification and reduction is discussed. Two problems are: simplification of rules (the reductions of attributes in rule premises) and reduction of rule base (deleting supernumeruous rules). This effect is done with similarity measure (Eq. 10) with two threshold parameters. In [21] the grid partition of the input domain is used to extract rules from learning examples. Then greedy algorithm is applied to merge the premises of the rules (satisfying the precision bounds). In [5] the Gaussian functions are approximated with triangle or trapezoidal functions and then merged. The method is developed for reducing the number of fuzzy rules.

In our approach the reduction of the supernumeruous rules the minor aim. Applying of scatter partition (clustering) [7, 13] or hierarchical partition [17, 19, 20] of the input domain can prevent from creation of unnecessary rules, thus their reduction is no more needed. The major aim of our approach is handling of large data sets. Such data sets cannot be handled in the hitherto existing approaches. The idea of incremental creation of fuzzy rule model is arisen. The models are created basing on the parts of the data set and then the elaborated models are merged into one. This approach is also valid when not all data are available (streaming data from industry measurements) and the model has to be created basing only on a part of data and when the next part of data is available has to be refreshed.

## 2. Measures of rule similarity

The paper [18] describes the procedure of rules merging. This approach is based on «similarity» notion. It is defined as the degree to which the fuzzy sets are equal. It is not, as commonly colloquially understood, having characteristics in common, similar in shape, but in size or position.

For crisp sets the equality is obvious, two sets $A$ and $B$ are equal

$$A = B \Leftrightarrow \forall_x \mu_A(x) = \mu_B(x). \tag{1}$$

In [18] the criteria for rule similarity measure $S(A, B)$ are proposed. These are:

1. Nonoverlapping fuzzy sets should be considered totally non equal

$$S(A, B) = 0 \Leftrightarrow \mu_A(x)\mu_B(x) = 0, \qquad \forall x \in X \tag{2}$$

2. Overlapping fuzzy sets should have a similarity value

$$S(A, B) > 0 \Leftrightarrow \exists x \in X, \qquad \mu_A(x)\mu_B(x) \neq 0 \tag{3}$$

3. Only equal fuzzy sets should have a similarity value

$$S(A, B) = 1 \Leftrightarrow \mu_A(x) = \mu_B(x), \qquad \forall x \in X \tag{4}$$

4. Similarity between two fuzzy sets should not be influenced by scaling or shifting the domain on which they are defined

$$S(A', B') = S(A, B) \tag{5}$$
$$\mu_{A'}(l + kx) = \mu_A(x) \tag{6}$$
$$\mu_{B'}(l + kx) = \mu_B(x) \tag{7}$$

where $k, l \in \mathbb{R}, \qquad k > 0$.

Similarity measures can be divided into [18]:

1. Geometric measures handle the similarity as proximity, not as a measure of equality.
2. Set-theoretical measures are based on union and intersections of sets, not influenced by scaling and ordering of the domain. The similarity of the sets is defined as an inverse of their distance in metric space. The interpretation of similarity as approximate equality can better be represented by a set-theoretic approach.
3. Pattern recognition approach [4] each fuzzy set is represented by a limited number of features, so the distance computation is simplified.

The geometric similarity measures are best suited for measuring (dis)similarity among distinct fuzzy sets, while the set-theoretic measures are the most suitable for capturing similarity among overlapping fuzzy sets [23]. The tuples of data can be interpreted as vectors in space. For high dimensional spaces the metrics used in 3D is no more useful. Other metrics should be used to better fit the features of high dimensional spaces. In spaces with more than 15 dimensions the Euclidean metrics has no meaning [6]. Even in spaces with as few as 5 dimensions some better metrics than Euclidean one should be used [6, 2]. The dimensional (Minkowski) metrics have been proposed for handling high dimensional data sets [1, 6]. Generalisation of the Hausdorff distance to fuzzy sets, Minkowski class of distance functions [10]

$$d_r(x, y) = \left[ \sum_{i=1}^{n} |x_i - y_i|^r \right]^{\frac{1}{r}}, \text{ where } r \geq 1, \tag{8}$$

for various values of $r$ we get various measures:

$r = 1$ city-block model
$r = 2$ Euclidean metric
$r = \infty$ $d_\infty(x, y) = \max_i |x_i - y_i|$

Kacprzyk [11] proposed $(d_2)^2$ measure.

A set-theoretical measure (common in literature) "is consistency index which is the maximum membership degree of the intersection of two fuzzy sets" [18]

$$S_C(A,B) = \sup_{x\in X} \mu_{A\cap B} = \max_{x\in X} [\mu_A(x) \wedge \mu_B(x)] \tag{9}$$

where $\wedge$ is minimum operator. The measure does not fulfil the 3 criterion, because it takes into consideration only one value of the $x$ variable [18].

In paper [18] the set-theoretic criterion from [8] is used:

$$S(A,B) = \frac{|A\cap B|}{|A\cup B|} = \frac{|A\cap B|}{|A| + |B| - |A\cap B|}, \tag{10}$$

where $|\cdot|$ denotes the cardinality of the set, and $\cap$ and $\cup$ operators denote intersection and union respectively. For discrete universe $X = \{x_j : j = 1, 2, \ldots, m\}$ we get

$$S(A,B) = \frac{\sum_{j=1}^{m} [\mu_A(x_j) \wedge \mu_B(x_j)]}{\sum_{j=1}^{m} [\mu_A(x_j) \vee \mu_B(x_j)]}, \tag{11}$$

where $\wedge$ stands for minimum and $\vee$ stands for maximum. Measure of set similarity introduced in [23]

$$S(A,B) = \frac{f(A\cap B)}{f(A\cap B) + \alpha f(A\backslash B) + \beta f(B\backslash A)} \in [0,1], \tag{12}$$

where $\alpha, \beta \geqslant 0$. Researchers propose various values of parameters, as $\alpha = \beta = 1$, $\alpha = \beta = 1/2$, $\alpha = 1 \wedge \beta = 0$. Typically the $f$ function is taken to be the cardinality function [23].

Pattern recognition based approach is proposed by Bonissone [4], who extract for each set four-dimensional vector. The dimensions are:

1. power (cardinality of the fuzzy set)

$$|A| = \int_{-\infty}^{+\infty} \mu_A(x)dx, \tag{13}$$

2. entropy

$$E(A) = \int_{-\infty}^{+\infty} S\left(\mu_A(x)\right) dx, \tag{14}$$

where

$$S(y) = -y \ln y - (1-y)\ln(1-y), \tag{15}$$

3. first moment (centre of gravity of the membership function)

$$F(A) = \frac{\int_{-\infty}^{+\infty} x\mu_A(x)dx}{\int_{-\infty}^{+\infty} \mu_A(x)dx} \tag{16}$$

4. skewness of the membership function

$$K(A) = \int_{-\infty}^{+\infty} (x - F(A))^3 \mu_A(x)dx. \tag{17}$$

The Euclidean weighted distance between two concepts is used to determine the similarity of the sets.

Wenstøp [22] assigns two dimensional feature vector to each set: location (centre of gravity) and imprecision (fuzzy scalar cardinality) of the membership function. The distance is measured with Euclidean metrics. Similar geometrical approach is proposed by Eshragh and Mamdami [9] and by Kacprzyk [11].

Some further similarity measures are also proposed. Bhattacharya [3] distance is defined as

$$R(A,B) = \sqrt{1 - \int_{-\infty}^{+\infty} \sqrt{\mu_A^*(x) \cdot \mu_A^*(x)}dx}, \tag{18}$$

where $\mu_A^*(x)$ is normalised membership function:

$$\mu_A^*(x) = \frac{\mu_A(x)}{|A|}. \tag{19}$$

Correlation Index proposed in [16] is defined as

$$\mathrm{CORR}(A,B) = 1 - \left(\frac{4}{X_A + X_B}\right)(d_2)^2, \tag{20}$$

where $X_A$ is defined as

$$X_A = \int_{-\infty}^{+\infty} (2\mu_A(x) - 1)^2 \, dx \tag{21}$$

In [10] the fuzzy measure of similarity of fuzzy sets is presented.

## 3. Fuzzy models

The neuro-fuzzy system used in this paper is ANNBFIS system proposed in [7]. For the brevity of the paper the ANNBFIS system will not be described in details. For further information see [7, 14]. The ANNBFIS system is system with parametrised consequences and logical interpretation of fuzzy rules. The rule base is composed of fuzzy rules (fuzzy implications)

$$R_i : \underline{\mathbf{x}} \text{ is } \underline{\mathbf{A}}_i \Rightarrow y_i \text{ is } B_i \left( \underline{p}_i \right), \tag{22}$$

where $\underline{\mathbf{x}} = [x_1, x_2, \ldots, x_N]^{\mathrm{T}}$ and $y_i$ are linguistic variables, $N$ – number of attributes. $\underline{\mathbf{A}}$ and $B$ are fuzzy linguistic terms and $\underline{\mathbf{p}}$ is the consequence parameter vector. The variable $\underline{\mathbf{A}}$ represents the region in input domain. The linguistic variable $A_{ij}$ ($\underline{\mathbf{A}}_i$ for $j$th attribute of $i$th rule) is described with the Gaussian membership function:

$$\mu_{A_{ij}}(x_j) = \exp\left( -\frac{(x_j - c_{ij})^2}{2s_{ij}^2} \right), \tag{23}$$

where $c_{ij}$ is the core location for $j$th attribute of $i$th rule and $s_{ij}$ is this attribute Gaussian bell deviation.

The firing strength $F_i$ of the $i$th rule is defined as T-norm (here product T-norm is used) of membership function values of all attributes.

$$F = T(\mu_1, \mu_2, \ldots, \mu_N). \tag{24}$$

The term $B$ in Eq. 22 is represented by an isosceles triangle with the base width $w$, the altitude equal to one. The localisation of the core of the triangle fuzzy set is determined by linear combination of input attribute values:

$$y_i = \underline{\mathbf{p}}_i^{\mathrm{T}} \cdot \left[1, \underline{\mathbf{x}}^{\mathrm{T}}\right]^{\mathrm{T}} = [p_{i0}, p_{i1}, \ldots, p_{iN}] \cdot [1, x_1, \ldots, x_N]^{\mathrm{T}}. \tag{25}$$

The fuzzy output of the system can be written as

$$\mu_{B'}(y, \underline{\mathbf{x}}) = \bigoplus_{i=1}^{I} \Psi\left( \mu_{\underline{\mathbf{A}}_i}(\underline{\mathbf{x}}), \mu_{B_i}(y_i, \underline{\mathbf{x}}) \right), \tag{26}$$

where $\bigoplus$ denotes the aggregation, $\Psi$ – the fuzzy implication and $I$ – the number of rules.

The crisp output of the system is calculated using the MICOG method:

$$y = \frac{\sum_{i=1}^{I} g\left(F_i(\underline{\mathbf{x}}), w_i\right) y_i(\underline{\mathbf{x}})}{\sum_{i=1}^{I} g\left(F_i(\underline{\mathbf{x}}), w_i\right)}, \tag{27}$$

where $y_i(\underline{\mathbf{x}})$ stands for the location of the core of the consequent fuzzy set, $F_i$ – the firing strength of the $i$th rule, $w_i$ – the width of the base of the isosceles triangle consequence function of the $i$th rule. The function $g$ depends on the fuzzy implication, in the system the Reichenbach one is used, so for the $i$th rule function $g$ is [7]

$$g_i(\underline{\mathbf{x}}) = \frac{w_i}{2} F_i(\underline{\mathbf{x}}). \tag{28}$$

The scatter partition of input domain (clustering) is used for extraction of fuzzy rules basing on presented train examples. The clustering procedure elaborates the premises of the rules. The consequences are then determined in tuning process. The tuning of the rules is done with two methods: the premises of the rules are tuned by gradient method and consequences (linear coefficients) are calculated with LMSE iterative algorithm [12].

## 4. Fuzzy model merging

Each rule is characterised by parameters $\langle \underline{\mathbf{c}}, \underline{\mathbf{s}}, \underline{\mathbf{p}}, w \rangle$, where $\underline{\mathbf{c}}_i$ and $\underline{\mathbf{s}}_i$ are the the parametres of the fuzzy set for $i$-th attribute (cf. Eq. 23); $\underline{\mathbf{p}}$ is a vector of linear consequences in the rule's consequence (cf. Eq. 25) and finally $w$ is the length of the support of the fuzzy set in the consequence of the rule. For merging only the parameters of the rule's premise are used. These parameters describe the Gaussian membership function (cf. Eq. 23). For two Gaussian functions describing the same parameter in two rules the set theoretic criterion (Eq. 10) proposed in [8] are applied. The measure fulfils all four criteria for rule similarity mentioned in Sec. 2.

### 4.1. Similarity of rules

For two Gaussian sets $A$ and $B$ the value of $|A \cap B|$ has to be calculated. Each set is defined with equation

$$g(c, s; x) = \exp\left(-\frac{(x-c)^2}{2s^2}\right), \tag{29}$$

so

$$|A \cap B| = \int_{-\infty}^{\infty} \min\left[g(c_a, s_a; x), g(c_b, s_b; x)\right] dx. \tag{30}$$

To avoid numerical integration the cumulative distribution function $\Phi$ of standard normal distribution can be used. Then the integral of function $g$ (Eq. 29) may be expressed as

$$G(x) = \int_{-\infty}^{x} g(c, s; x) dx = s\sqrt{2\pi}\, \Phi\left(\frac{x-c}{s}\right). \tag{31}$$

Four cases should be discussed:

1. $c_a = c_b \wedge s_a = s_b$,
2. $c_a = c_b \wedge s_a \neq s_b$,
3. $c_a \neq c_b \wedge s_a = s_b$,
4. $c_a \neq c_b \wedge s_a \neq s_b$.

The first case is trivial, the similarity $S$ of the fuzzy sets $A$ and $B$ is maximal:

$$S(A, B) = 1. \tag{32}$$

The second case, $c_a = c_b \wedge s_a \neq s_b$,

$$|A \cap B| = G(c, \min(s_a, s_b); \infty) = \min(s_a, s_b)\sqrt{2\pi} \tag{33}$$

$$S(A, B) = \frac{\min(s_a, s_b)\sqrt{2\pi}}{s_a\sqrt{2\pi} + s_b\sqrt{2\pi} - \min(s_a, s_b)\sqrt{2\pi}}, \tag{34}$$

what leads to

$$S(A, B) = \frac{\min(s_a, s_b)}{\max(s_a, s_b)}. \tag{35}$$

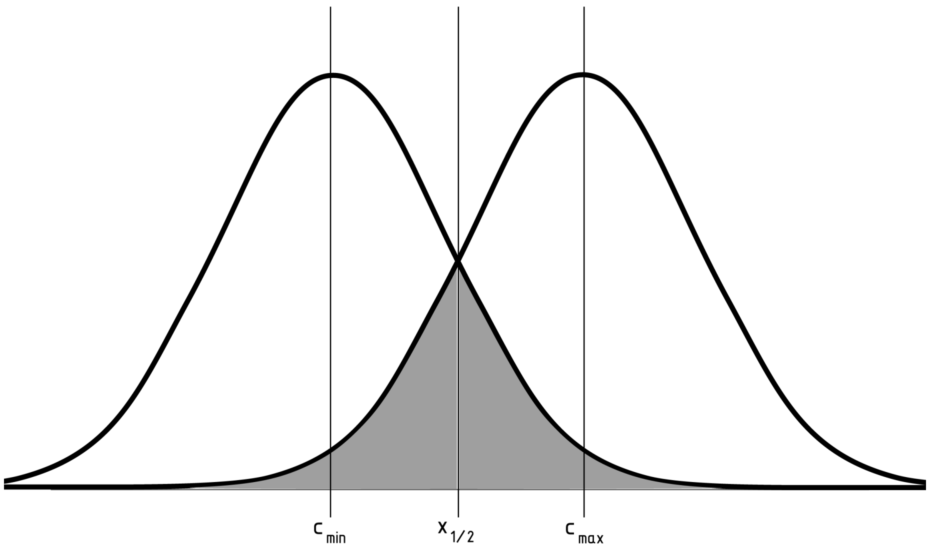


Fig. 1. Intersection of Gaussian functions with unequal core locations ($c_a \neq c_b$) and the same fuzzyfication ($s_a = s_b$). The gray area denotes $|A \cap B|$

The third case (cf. Fig. 1), $c_a \neq c_b \wedge s_a = s_b = s$, is more complicated. Let'a denote $c_{\max} = \max(c_a, c_b)$ and $c_{\min} = \min(c_a, c_b)$. Then

$$|A \cap B| = G(c_{\max}, s; x_{1/2}) + 1 - G(c_{\min}, s; x_{1/2}), \tag{36}$$

where $x_{1/2}$ is solution to the equation $g(c_a, s; x) = g(c_b, s; x)$ and is equal

$$x_{1/2} = (c_a + c_b)/2. \tag{37}$$

The area $|A \cap B|$ is symmetric and can be expressed as

$$|A \cap B| = 2G(c_{\max}, s; x_{1/2}) = 2\sqrt{2\pi}\Phi\left(\frac{x_{1/2} - c_{\max}}{s}\right). \tag{38}$$

Taking into consideration Eq. 37 we get

$$|A \cap B| = 2\sqrt{2\pi}\Phi\left(\frac{c_{\min} - c_{\max}}{2s}\right). \tag{39}$$

Finally the similarity is expressed as

$$S(A, B) = \frac{\Phi\left(\frac{c_{\min} - c_{\max}}{2s}\right)}{s - \Phi\left(\frac{c_{\min} - c_{\max}}{2s}\right)} \tag{40}$$

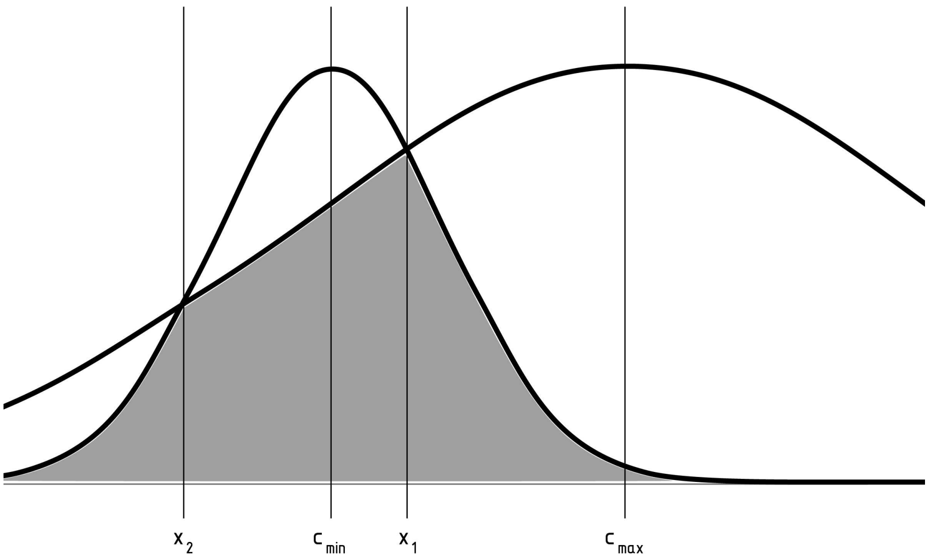


Fig. 2. Intersection of Gaussian functions with unequal core locations ($c_a \neq c_b$) and fuzzyfication ($s_a \neq s_b$). The gray area denotes $|A \cap B|$

The fourth case, $c_a \neq c_b \wedge s_a \neq s_b$, is the most complicated one. The value of $|A \cap B|$ can be calculated as (cf. Fig. 2) sum of integrals in intervals $[-\infty, x_2]$, $[x_2, x_1]$ and $[x_1, \infty]$. The values $x_1$ and $x_2$ are the intersection arguments of both Gaussian function. There are two solutions to the equation

$$g(c_a, s_a; x) = g(c_b, s_b; x) \tag{41}$$

$$x_1 = \frac{s_b c_a + s_a c_b}{s_b + s_a} \tag{42}$$

and

$$x_2 = \frac{s_b c_a - s_a c_b}{s_b - s_a} \tag{43}$$

The $x_1$ value is the weighted mean of $c_a$ and $c_b$, so $\min(c_a, c_b) < x_1 < \max(c_a, c_b)$.

Let's denote $c_l = \min(c_a, c_b)$, $c_h = \max(c_a, c_b)$ and $s_l = s_i$, where $i = \arg\min(c_a, c_b)$, $s_h = s_j$, where $j = \arg\max(c_a, c_b)$. It means that $c_l$ i $s_l$ apply to the same function. Of course it is not always valid that $s_h = \max(s_a, s_b)$ and $s_l = \min(s_a, s_b)$. Thus two subcases should be discussed:

1. If $s_l < s_h$, then

$$\begin{aligned}
|A \cap B| &= \int_{-\infty}^{x_1} g(c_h, s_h; x)dx - \int_{-\infty}^{x_2} g(c_h, s_h; x)dx + \\
&+ \int_{-\infty}^{x_2} g(c_l, s_l; x)dx + \int_{-\infty}^{\infty} g(c_l, s_l; x)dx + \\
&- \int_{-\infty}^{x_1} g(c_l, s_l; x)dx = \qquad (44) \\
&= s_h\sqrt{2\pi}\left[\Phi\left(\frac{x_1 - c_h}{s_h}\right) - \Phi\left(\frac{x_2 - c_h}{s_h}\right)\right] + \\
&+ s_l\sqrt{2\pi}\left[\Phi\left(\frac{x_2 - c_l}{s_l}\right) + 1 - \Phi\left(\frac{x_1 - c_l}{s_l}\right)\right] = \\
&= \sqrt{2\pi}\left\{s_h\left[\Phi\left(\frac{x_1 - c_h}{s_h}\right) - \Phi\left(\frac{x_2 - c_h}{s_h}\right)\right] + \right. \\
&+ \left. s_l\left[\Phi\left(\frac{x_2 - c_l}{s_l}\right) + 1 - \Phi\left(\frac{x_1 - c_l}{s_l}\right)\right]\right\} = \\
&= \sqrt{2\pi}\left(\Psi + s_l\right) \qquad (45)
\end{aligned}$$

where $\Psi$ denotes

$$\begin{aligned}
\Psi &= s_h\left[\Phi\left(\frac{x_1 - c_h}{s_h}\right) - \Phi\left(\frac{x_2 - c_h}{s_h}\right)\right] + \\
&+ s_l\left[\Phi\left(\frac{x_2 - c_l}{s_l}\right) - \Phi\left(\frac{x_1 - c_l}{s_l}\right)\right]. \qquad (46)
\end{aligned}$$

Similarity measure:

$$S(A, B) = \frac{\Psi + s_l}{s_l + s_h - \Psi - s_l} = \frac{\Psi + s_l}{s_h - \Psi} \tag{47}$$

Because $s_l < s_h$

$$S(A,B) = \frac{\Psi + s_{\min}}{s_{\max} - \Psi} \tag{48}$$

2. If $s_l > s_h$ then

$$\begin{aligned} |A \cap B| &= \int_{-\infty}^{x_2} g(c_l, s_l; x)dx - \int_{-\infty}^{x_1} g(c_l, s_l; x)dx + \\ &+ \int_{-\infty}^{x_1} g(c_h, s_h; x)dx + \int_{-\infty}^{\infty} g(c_h, s_h; x)dx + \\ &- \int_{-\infty}^{x_2} g(c_h, s_h; x)dx = \qquad (49) \\ &= \sqrt{2\pi} \left\{ s_l \left[ \Phi\left(\frac{x_2 - c_l}{s_l}\right) - \Phi\left(\frac{x_1 - c_l}{s_l}\right) \right] + \right. \\ &+ \left. s_h \left[ \Phi\left(\frac{x_1 - c_h}{s_h}\right) + 1 - \Phi\left(\frac{x_2 - c_h}{s_h}\right) \right] \right\} = \qquad (50) \\ &= \sqrt{2\pi} \left(\Psi + s_h\right), \qquad (51) \end{aligned}$$

where $\Psi$ is defined by Eq. 46. Similarity measure:

$$S(A,B) = \frac{\Psi + s_h}{s_l - \Psi}. \tag{52}$$

Because $s_l > s_h$, so

$$S(A,B) = \frac{\Psi + s_{\min}}{s_{\max} - \Psi}. \tag{53}$$

The Eq. 48 and 53 are the same and in both subcases the similarity is calculated with the same formula.

The considerations above concern only one attribute and the premises of rule are usually composed of descriptors for more attributes. The firing strength of the rule is calculated as a T-norm of firing strengths of all attributes of the rule (Eq. 24). Analogously the similarity of the rules $R_a$ and $R_b$ is calculated as a T-norm of the similarities of adequate attributes.

$$S(R_a, R_b) = T\left[S(A_1, B_1), S(A_2, B_2), \ldots, S(A_N, B_N)\right], \tag{54}$$

where $A_i$ stands for the fuzzy set representing the $i$th attribute of the $R_a$ rule and respectively $B_i$ denotes the set representing the $i$th attribute of the $R_b$ rule.

### 4.2. Merging of rules

For each of $n(n+1)/2$ pairs of $n$ rules the similarity is calculated in the way presented in the Sec. 4.1. Then the pairs are sorted in descending order of the similarity value. Starting from the most similar pair, its similarity value is compared with the merging threshold $\zeta$. If the similarity is greater the pair is merged and the rules being merged are removed from the rule base. Then the next rule pair is checked. If the rule $R$ has been merged, all next pairs containing this rule are not further analysed.

During the rule merging the core location and the fuzzyfication of the new set constituting the merged rule are elaborated. The core location $\underline{\mathbf{c}}$ is calculated as a weighted mean of all data examples presented in this step of merging. For each data example the firing strength $F_1$ and $F_2$ of both parent rules (that are to be merged) are calculated (Eq. 24). The weight of the $k$th example is elaborated with the formula 55:

$$\eta_k = \max\left(m_1 F_1(k), m_2 F_2(k)\right) \tag{55}$$

where $F_i$ stands for the firing strength if the $i$th rule and $m_i$ is the number of data examples that participated in the creation of the $i$th rule. If the rule is created in the merging process, its $m$ value is sum of the $m$ values of the parent rules. Otherwise when the rule has no parent rules (ie. is not a result of merging) the $m$ value is the number of learning examples in the data set used in extraction of the rule.

Having calculated the data examples weights the core location of the fuzzy set in the new rule is elaborated with the formula:

$$\underline{\mathbf{c}}_m = \frac{\sum_{k=1}^{K} \eta_k \underline{\mathbf{x}}_k}{\sum_{k=1}^{K} \eta_k}. \tag{56}$$

The fuzzification of the fuzzy set in question is calculated with formula:

$$\underline{\mathbf{s}}_m = \sqrt{\frac{1}{\sum_{k=1}^{K} \eta_k} \sum_{k=1}^{K} \left(\underline{\mathbf{x}}_k - \underline{\mathbf{c}}_m\right)^2 \eta_k} \tag{57}$$

The new rule base (with new rules and no merged rules) is then tuned in order to better fit the parameters of rules to the presented data and to elaborate the linear coefficients $\underline{\mathbf{p}}$ in the consequences of rules (cf. Eq. 25).

One more question needs discussing. The important parameter in rule merging is the merging parameter $\zeta$ (it is the threshold value for merging). The value of this parameter is determined in dynamic way. The initial value $\zeta_0$ is determined in the following procedure. Having created two models for the first two data sets these models are merged with several values of $\zeta = [1.00 \cdot 10^{-5}, 3.59 \cdot 10^{-5}, 1.29 \cdot 10^{-4}, 4.64 \cdot 10^{-4}, 1.67 \cdot 10^{-3}, 5.99 \cdot 10^{-3}, 2.15 \cdot 10^{-2}, 7.74 \cdot 10^{-2}, 2.78 \cdot 10^{-1}, 1.00]$. These values form

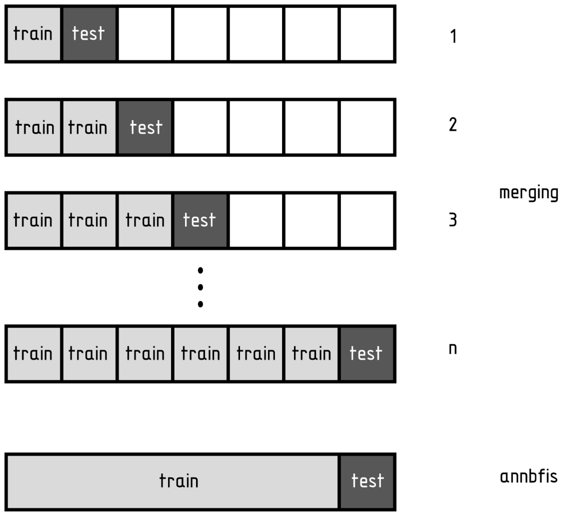


Fig. 3. The scheme of experiments. The data set is divided for merging. First the model is elaborated for the first subset, the second subset is used as test data. Then for the second data set the second model is created. Both models are then merged and the third set is used for evaluation of the model. In general in $i$th step $i$ models were created and merged and $(i+1)$st data set is used for evaluation. For comparison the model is created by annbfis basing on the train data being the sum of all subsets except for the ultimate set that is used for test and evaluation of the created model

the logarithmic sequence. The higher values of $\zeta$ means that it is more difficult to merge rules and consequently the model contains more rules. The initial value $\zeta_0$ of the threshold parameter is chosen the one at which the number of rules in the model increases for the second time. In this way the initial value $\zeta_0$ of the threshold parameter is elaborated. The experiments show that further merging with $\zeta_0 = \mathrm{const}$ leads to faster of slower multiplication of rules when the new model is merged with the previous one. This makes the model highly illegible what is in the conflict with the idea of fuzzy systems. This disadvantageous feature can be avoided in the following way. After the initial value $\zeta_0$ is determined, the temporary value of the threshold parameter $\zeta_t$ is calculated as

$$\zeta_t = \zeta_0 \bigg/ \left(\frac{r_t}{r_0}\right), \tag{58}$$

where $\zeta_0$ is the initial value of the threshold, $r_0$ is the number of rules in the first model and finally $r_t$ is the number of rules in the model merged in the previous merging step. Such approach protects against uncontrollable incrementation of number of rules in the model's rule base and keeps the model intelligible.

## 5. Experiments

### 5.1. Data sets

The experiments were conducted for the datasets generated with Mackey-Glass differential equation [15]

$$\frac{\partial x(t)}{\partial t} = \frac{ax(t-\tau)}{1+[x(t-\tau)]^{10}} - bx(t), \tag{59}$$

where $x(t)$ is the density of leukocytes, and $a, b, \tau$ are constants. The initial conditions are $a = 0.2$, $b = 0.1$, $\tau = 17$. The equation is solved with Runge-Kutta method (order 4) with sampling step $0.1$ and initial condition $x(0) = 0.1$. The generated data enable preparing data tuples with template

$$\underline{\mathbf{x}}(k) = [x(k-18), x(k-12), x(k-6), x(k), x(k+6)] \, . \tag{60}$$

The last value $x(k+6)$ is the prediction attribute value.

Two series of data have been prepared. The former contains the data sets with 1200, 2000, 5000, 10 000, 20 000, 100 000, $10^6$ and $10^7$ tuples. The data sets where divided into subsets with 500 tuples (the only exception is 1200-tuple data set, which has split into 200-tuple sets). The last subsets may contain less than 500 tuples. The latter series is prepared basing on the 100 000-tuple data set, but each data set is divided into different number of data examples in the subsets. This series contains the same number of data in each data set but divided into various number of subsets: 1000 (100 subsets), 2000 (50 subsets), 5000 (20 subsets), 10 000 (10 subsets) and 20 000 (5 subsets).

For elaboration of fuzzy model the ANNBFIS algorithm [7], briefly described in Sec. 3, has been used. The experiments have been conducted in the following way. The model for the first data subset is tested with the second data subset. Then the model for the second subsets is elaborated, merged with first model and the merged model is tested with the third data subset. In each step the $(i+1)$st subset is used to test the model created in merging of $i$ previous models. Then the comparatory experiment without merging of models where conducted. In this part the last $n$th subset was used to test the model elaborated for the set containing all data from $n-1$ subsets. Fig. 3 presents the experiment paradigm in graphical way. This paradigm serves for testing knowledge generalisation (KG) ability. Also the data approximation (DA) ability was tested. In this case the train set are the same as in KG and the train set is simultaneously used as the test set.

### 5.2. Results

The results of experiments are gathered in tables.

| no. of data in subset | time [s] M $t_m$ | A $t_a$ | $t_m/t_a$ | $\zeta$ M | RMSE $(\cdot 10^{-3})$ DA M | DA A | KG M | KG A | no. of rules |
|---|---|---|---|---|---|---|---|---|---|
| 200 | 3124 | 2309 | 1.35 | 0.00828 | 1.68 | 1.95 | 33.65 | 1.86 | 13 |
| 500 | 3467 | 2440 | 1.42 | 0.02977 | 3.07 | 1.95 | 4.00 | 1.90 | 13 |
| 1000 | 5634 | 4699 | 1.20 | 0.06323 | 1.33 | 1.14 | 1.35 | 1.20 | 20 |
| 2000 | 3933 | 2701 | 1.46 | 0.08695 | 1.43 | 1.95 | 3.03 | 2.00 | 15 |
| 5000 | 2979 | 2417 | 1.23 | 0.09937 | 1.38 | 1.88 | 1.40 | 1.83 | 14 |
| 10000 | 2348 | 1849 | 1.27 | 0.09937 | 2.14 | 2.31 | 2.20 | 2.25 | 12 |
| 20000 | 1665 | 1150 | 1.45 | 0.27825 | 2.53 | 3.48 | 2.60 | 3.44 | 10 |

Table 1: The same data (100 000 tuples in each set) divided into subsets with various data numbers. Abbreviations: M – merging procedure, A – annbfis (simple) procedure, DA – data approximation, KG – knowledge generalisation, $t_m$ – time consumed by merging procedure, $t_a$ – time consumed by simple (annbfis) procedure.

Table 1 presents the results of knowledge generalisation (KG) and data approximation (DA) for merging (M) and annbfis (A). In this data series there are 100 000 tuples split into various numbers of tuples in data subsets (from 200 to 20 000). In all cases here the merging approach takes more time to handle the whole data set (with all subsets), the ratio of time consumed by merging and simple approach is from 1.20 to 1.46. The number of rules is stabilised on 10 – 20 rules. The knowledge generalisation and data approximation ability seems better in merging approach with larger subsets.

Table 2 presents the results elaborated for second data series. In this series the number of data tuples in data subset is 200, 500 and 10 000 whereas the number of all tuples in the data set is various. In this data series the time needed by merging approach is longer then by simple one, but the difference drops with growth of the data set. Also the results of KG and DA are similar in both approaches with one exception: for the data set with 2 million tuples the simple approach failed to create the model, whereas the merging approach managed to elaborate fuzzy model.

The merging approach can create models with reasonable accuracy using only some subsets with no need of handling the whole data set (cf. Fig. 4).

The Fig. 5 presents the membership functions for the first attribute before and after the last merge. The A plot shows these functions for the model merged of $N-1$ submodels, the B one – the last submodel. The C plot depicts the membership functions of the final model (after last merge). Some membership function are similar because the whole rule with all attributes was taken into account in merging process. And similarity of one attribute may not lead to merging of whole rules.

| no. of data in | | time [s] | | | RMSE $(\cdot 10^{-3})$ | | | | no. of |
|---|---|---|---|---|---|---|---|---|---|
| | | M | A | | DA | | KG | | rules |
| set | subset | $t_m$ | $t_a$ | $t_m/t_a$ | M | A | M | A | |
| 1200 | 200 | 22 | 12 | 1.87 | 0.336 | 4.478 | 5.039 | 4.102 | 9 |
| 2k | 500 | 30 | 12 | 2.72 | 5.460 | 4.301 | 8.340 | 4.410 | 6 |
| 5k | 500 | 102 | 60 | 1.69 | 4.829 | 3.462 | 5.434 | 3.579 | 11 |
| 10k | 500 | 241 | 172 | 1.40 | 1.305 | 1.994 | 4.467 | 1.808 | 13 |
| 20k | 500 | 561 | 408 | 1.37 | 2.071 | 1.935 | 4.663 | 1.832 | 14 |
| 100k | 10k | 2142 | 1744 | 1.23 | 2.147 | 2.316 | 2.208 | 2.254 | 12 |
| 200k | 10k | 5196 | 4180 | 1.24 | 1.689 | 1.912 | 1.696 | 1.838 | 13 |
| 500k | 10k | 14922 | 12758 | 1.17 | 1.913 | 1.872 | 1.911 | 1.855 | 13 |
| 1M | 10k | 37595 | 37123 | 1.01 | 1.653 | 1.548 | 1.653 | 1.531 | 15 |
| 2M | 10k | 92164 | – | – | 1.645 | – | 1.653 | – | 20 |

Table 2: The comparison of results elaborated for the second data series. The number of tuples in data subset is constant and the number of data tuples in data set growth. The abbreviations used are the same as in Table 1.

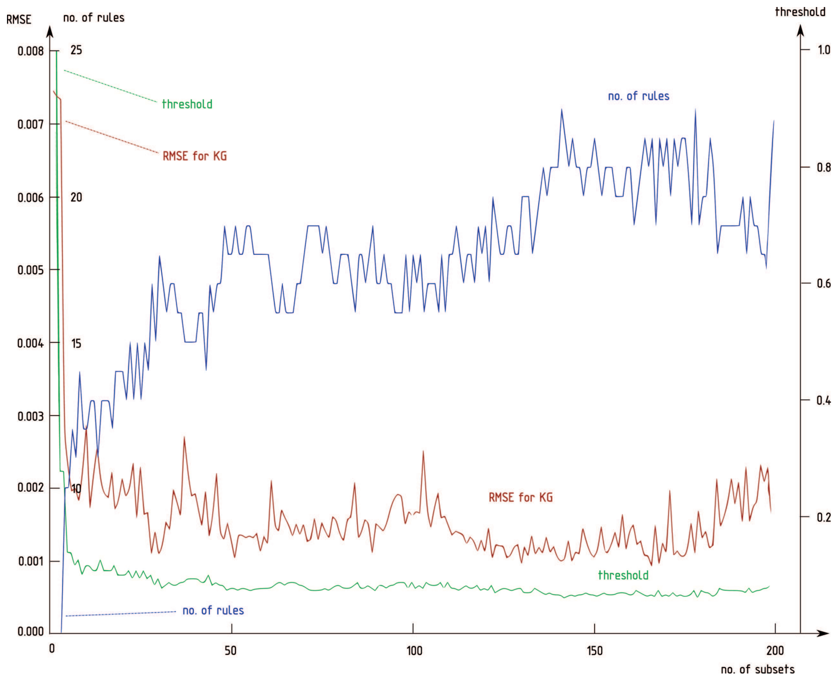


Fig. 4. The results elaborated for 2 mil. tuples split into 200 subsets (each containing 10 000 tuples). The figure presents the root mean square error (RMSE) for knowledge generalisation (KG), number of rules and merging threshold $\zeta$ in function of merged subsets

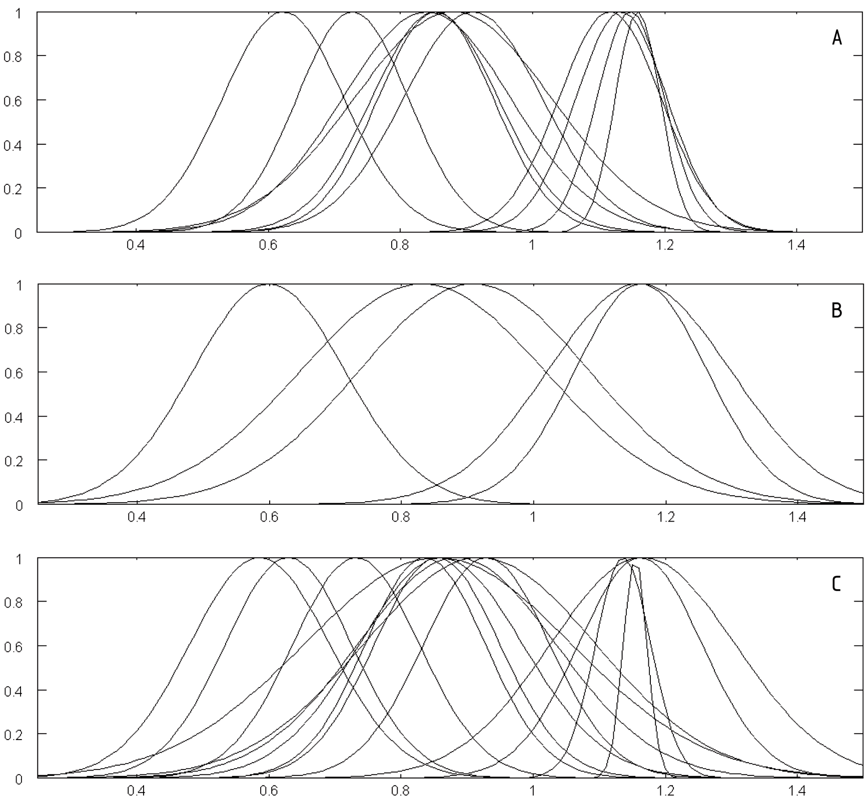


Fig. 5. The membership functions for the first attribute in the last merging. The A figure presents the functions in the models before last merge, the B – the last submodel and finally C – the functions of the merged final model. The figure drawn for the data set split into 10 subsets each with 10000 tuples

## 6. Summary

Merging of fuzzy model may be helpful in some situations. One of them is elaborating of fuzzy models for large data sets. The solution proposed in the paper is suitable for data set neuro-fuzzy systems with parametrised consequences. This method enables creating model for big data sets in reasonable time. Although the elaboration of fuzzy model for whole data set takes more time, but quite good model can be achieved after merging of models created for a few data subsets. The precision of the model elaborated in merging approach is similar to that achieved in simple approach.

In short:

1. The merging approach to elaborating the model needs more time to create model for whole data set than creating model with all data tuples at once. But the larger

the dataset, the less is the difference between both approaches. When the dataset is too large, only the merging paradigm can afford the task.

2. Although creation of model for whole data set in merging approach takes more time, but quite reasonable model can be elaborated basing only on a few data subset, so the first model is achieved in shorter time (Fig. 4).
3. The results of data approximation (DA) and knowledge generalisation (KG) are similar. Two phenomena can be noticed:
   (a) The bigger the data set the smaller the advantage of the simple paradigm (Tab. 2).
   (b) The bigger the data subset the bigger the advantage of the merging paradigm (Tab. 1).

**Acknowledgements**

This work was supported by the European Union from the European Social Fund.

**References**

1. C. C. Aggarwal, A. Hinneburg, D. A. Keim: *On the surprising behavior of distance metrics in high dimensional spaces*. In ICDT '01: Proceedings of the 8th International Conference on Database Theory, pages 420–434, London, UK, 2001. Springer-Verlag.
2. K. S. Beyer, J. Goldstein, R. Ramakrishnan, U. Shaft: *When is "nearest neighbor" meaningful?* In ICDT '99: Proceeding of the 7th International Conference on Database Theory, pages 217–235, London, UK, 1999. Springer- Verlag.
3. B. Bhattacharya, D. P. Solomatine: *Neural networks and m5 model trees in modelling water level-discharge relationship*. Neurocomputing, 63:381–396, 2005.
4. P. P. Bonissone: *A pattern recognition approach to the problem of linguistic approximation in system analysis*. In Proceedings of the IEEE International COnference on Cybernetics and Society, pages 73–798, 1979.
5. M.-Y. Chen, D. A. Linkens: *Rule-base self-generation and simplification for data-driven fuzzy models*. Fuzzy Sets and Systems, 142(2):243–265, 2004.
6. A. Chmielewski, S. T. Wierzcho´n: *On the distance norms for detecting anomalies in multidimensional datasets*. Zeszyty Naukowe Politechniki Białostockiej, 2:39–49, 2007.
7. E. Czogała, J. Łęski: *Fuzzy and Neuro-Fuzzy Intelligent Systems*. Series in Fuzziness and Soft Computing. Physica-Verlag, A Springer-Verlag Company, Heidelberg, New York, 2000.
8. D. Dubois, H. Prade: *Fuzzy Sets and Systems: Theory and Applications*. Academic Press, Inc., 1980.

9. F. Eshragh, E.H. Mamdani: *A general approach to linguistic approximation*. International Journal of Man-Machine Studies, 11(4):501–519, 1979.

10. Vladimír Janiš and Susana Montes: *Distance between fuzzy sets as a fuzzy quantity*. Acta Universitatis Matthiae Belii, Series Mathematics, 14:41–49, 2007.

11. J. Kacprzyk: *Fuzzy set-theoretic approach to the optimal assignment of work places*. Large-Scale Systems Theory and Application, pages 123–131, 1976.

12. P. Larminat, Y. Thomas: *Automatyka – układy liniowe*. WNT, 1983.

13. J. Łęski. *Systemy neuronowo-rozmyte*. Wydawnictwa Naukowo-Techniczne, Warszawa, 2008.

14. J. Łęski, Ernest Czogała: *A new artificial neural network based fuzzy inference system with moving consequents in if-then rules and selected applications*. Fuzzy Sets and Systems, 108(3):289–297, 1999.

15. M. C. Mackey, L. Glass: *Oscillation and chaos in physiological control systems*. Science, 197(4300):287–289, July 1977.

16. C. A. Murthy, S. K. Pal, D. Dutta Majumder: *Correlation between two fuzzy membership functions*. Fuzzy Sets Syst., 17(1):23–38, 1985.

17. O. Nelles, R. Isermann: *Basis function networks for interpolation of local linear models*. Proceedings of the 35th IEEE Conference on Decision and Control, 1:470–475, 1996.

18. M. Setnes, R. Babuška, U. Kazmak,. H. R. van Nauta Lemke: *Similarity measures in fuzzy rule base simplification*. IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics, 28(3):376–386, June 1998.

19. K. Simiński: *Neuro-fuzzy system with hierarchical partition of input domain*. Studia Informatica, 29(4A (80)):43–53, 2008.

20. K. Simiński. *Patchwork neuro-fuzzy system with hierarchical domain partition*. In Marek Kurzyński and Michał Woźniak, editors, Computer Recognition Systems 3, volume 57 of Advances in Intelligent and Soft Computing, pages 11–18. Springer-Verlag, Berlin, Heidelberg, 2009.

21. T. Sudkamp, A. Knapp, J. Knapp: *A greedy approach to rule reduction in fuzzy models*. In IEEE International Conference on Systems, Man and Cybernetics (IEEE SMC'00), pages 3716–3721, Nashville, Tennessee, USA, 2000.

22. F. Wenstøp: *Quantitative analysis with linguistic values*. Fuzzy Sets and Systems, 4:99–115, 1980.

23. R. Zwick, E. Carlstein, D. V. Budescu: *Measures of similarity among fuzzy concepts: a comparative analysis*. Int. J. Approx. Reas., 1:221–242, 1987.

**Scalanie modeli rozmytych w systemach neuronowo-rozmytych**

Streszczenie

Artykuł opisuje scalanie modeli rozmytych w systemach neuronowo-rozmytych wykorzystywane przy tworzeniu modeli dla dużych zbiorów danych. Nieraz zbiory danych są tak duże, że nie jest możliwe wypracowanie modelu od razu dla całego zbioru. Tworzy się zatem modele dla podzbiorów zbioru danych. Uzyskane w ten sposób modele są następnie scalane, by wypracować jeden model. Podejście to jest także korzystne, gdy wszystkie dane nie są dostępne, ale są dostarczane partiami. Wtedy wstępny model jest wypracowany zanim wszystkie dane zostaną dostarczone do systemu. Artykuł przedstawia sposób wyznaczania podobieństwa reguł w modelu rozmytym oraz opisuje system neuronowo-rozmyty budujący i scalający modele wypracowane dla podzbiorów.