Neuro-fuzzy system with weighted attributes

Krzysztof Simiński

Published online: 25 May 2013

© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract The paper presents the neuro-fuzzy system with weighted attributes. Its crucial part is the fuzzy rule base composed of fuzzy rules (implications). In each rule the attributes have their own weights. In our system the weights of the attributes are numbers from the interval $[0, 1]$ and they are not global: each fuzzy rule has its own attributes' weights, thus it exists in its own weighted subspace. The theoretical description is accompanied by results of experiments on real life data sets. They show that the neuro-fuzzy system with weighted attributes can elaborate more precise results than the system that does not apply weights to attributes. Assigning weights to attributes can also discover knowledge about importance of attributes and their relations.

Keywords Weights of attributes · Importance of attributes · Weighted dimension space · Subspace clustering · Neuro-fuzzy system

1 Introduction

Neuro-fuzzy systems proved to be efficient in many fields of data mining. They combine the ability to handle imprecise data and to modify the parameters of elaborated models to better fit the data. The more complicated a model is, the more suitable it is to use fuzzy approach

(Zadeh et al. 1973). The fuzzy approach can provide better models, even for non-fuzzy data, than non-fuzzy systems.

The crucial part of the fuzzy system is the fuzzy rule base composed of fuzzy rules (implications). Creation of the fuzzy rule base is a difficult task. This procedure has enormous influence on the quality of results elaborated by the system. The rules can implement the knowledge of experts or can be created automatically from the presented data. The rules of the fuzzy model split the input domain into regions. This procedure can be reversed in order to obtain the rules from presented data. The domain is split into regions and the regions are transformed into premises of the rules. This approach is commonly used. There are three main ways of domain partition grid split (Jang 1993; Setnes and Babuška 2001), scatter split (clustering) and hierarchical split (Hoffmann and Nelles 2001; Jakubek et al. 2006; Nelles and Isermann 1996; Nelles et al. 2000; Simiński 2008, 2009, 2010). The most common method is scatter split (clustering) (Abonyi et al. 2002; Bauman et al. 1990; Chen et al. 1998; Czogała et al. 2000; Wang et al. 1994). Clustering avoids the curse of dimensionality, which is the main problem of grid partition. The main disadvantage of many clustering algorithms is their inability to discover the number of clusters. In such cases the number of clusters is passed to the algorithm as a parameter.

In high dimensional data sets not always all dimensions (attributes) are relevant. Some of them can be treated as noise and have minor importance. The reduction of dimensionality may be done for a whole data set (global dimensionality reduction) or individually for each cluster. The global feature transformation (e.g. PCA or SVD) causes problems with interpretability of elaborated models. Dimension reduction without feature transformation can be achieved by feature selection. The global approach selects the same subset of attributes for all clusters whereas each

Communicated by W. Pedrycz.

K. Simiński (✉)
Institute of Informatics, Silesian University of Technology ul,
Akademicka 16, 44-100 Gliwice, Poland
e-mail: krzysztof.siminski@polsl.pl

cluster may need its own subspace. This is the idea of subspace clustering (Friedman and Meulman 2004; Gan et al. 2006; Kriegel et al. 2009; Müller et al. 2009; Parsons et al. 2004; Sim et al. 2012) where each cluster may be extracted in its own subspace. There are two kinds of subspace clustering: bottom-up and top-down (Parsons et al. 2004). The former approach splits the clustering space with a grid and analyses the density of data examples in each grid cell extracting the relevant dimensions [e.g. CLIQUE (Agrawal et al. 1998), ENCLUS (Cheng et al. 1999), MAFIA (Goil et al. 1999)]. The latter (top-down) approach starts with full dimensional clusters and tries to throw away the dimensions of minor importance [e.g. PROCLUS (Aggarwal et al. 1999), ORCLUS (Aggarwal et al. 2000), δ -Clusters (Yang et al. 2002), FSC (Gan and Wu 2008; Gan et al. 2006)]. In algorithms mentioned above the attribute is valid or invalid in a certain cluster, the weight of the attribute in each cluster is either 0 or 1. In our solution the clustering algorithm assigns values from the interval [0, 1]. The attributes have partial importance in the subspace. This approach creates fuzzy rules in individual weighted subspaces.

The contribution of the paper is the neuro-fuzzy system with weighted attributes.

In the paper we follow the general rule for symbols: the blackboard bold uppercase characters ($\mathbb{A}$) are used to denote the sets, uppercase italics (A)—the cardinality of sets, uppercase bolds ($\mathbf{A}$)—matrices, lowercase bolds ($\mathbf{a}$)—vectors, lowercase italics (a)—scalars and set elements. Table 1 lists the symbols used in the paper.

The paper is organised as follows: Sect. 2 introduces the new neuro-fuzzy system with parameterized consequences and weighted attributes (architecture—Sect. 2.1, creation of a fuzzy model—Sect. 2.2). Section 3 describes the data sets (Sect. 3.1) and experiments with results (Sect. 3.2). Finally Sect. 4 summarises the paper.

2 Fuzzy inference system with parameterized consequences and attributes' weights

Fuzzy inference system with parameterized consequences and weights attributes is an extension of the neuro-fuzzy system with parameterized consequences ANNBFIS (Czogala et al. 2000; Łeski and Czogala 1999) which is the combination of the Mamdani and Assilan (1975), Takagi and Sugeno (1985) and Sugeno and Kang (1988) approach. The fuzzy sets in consequences are isosceles triangles (as in the Mamdani–Assilan system), but are not fixed—their location is calculated as a linear combination of attribute values as in the Takagi–Sugeno–Kang system. The important feature is the logical interpretation of fuzzy implication (cf. Eq. 11). The idea of the system with

Table 1 Symbols used in the papers

$\mathbb{X}$	Set of tuples, data examples, $\mathbb{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_X\}$
$\mathbf{x}$	Vector of tuple's descriptors, data example, $\mathbf{x} \in \mathbb{X}$
X	Number of tuples, $X = \ \mathbb{X}\ $
$[\mathbf{x}, y]^T$	Data tuple with vector $\mathbf{x}$ of attributes and decision attribute y
x	Descriptor of a tuple, $\mathbf{x} = [x_1, \dots, x_N]^T$
y	Decision attribute of the tuple
y_l	Localisation of the fuzzy set in consequence of l th rule
y_0	Defuzzified output of the system
$\mathbb{N}$	Set of attributes
n	Attribute, $n \in \mathbb{N}$
N	Number of attributes in a tuple, $N = \ \mathbb{N}\ $
$\mathbb{C}$	Set of clusters
C	Number of clusters, $C = \ \mathbb{C}\ $
c	Cluster, $c \in \mathbb{C}$
$\mathbf{U}$	Partition matrix, $\mathbf{U} = \{u_{ij}\}$
u_{ij}	Membership value of the j -th tuple to i -th cluster
d_{ij}	Distance between i -th cluster's centre and j -th tuple
$\mathbf{v}_i$	Cores of fuzzy sets in premise of i -th rule, $\mathbf{v}_i = [v_{i1}, \dots, v_{iN}]$
$\mathbf{s}_i$	Fuzziness of sets in premise of i -th rule, $\mathbf{s}_i = [s_{i1}, \dots, s_{iN}]$
$\mathbf{z}_i$	Weights of attributes in i -th rule, $\mathbf{z}_i = [z_{i1}, \dots, z_{iN}]$
f	Fuzzification parameter
$\mathbb{L}$	Set of rules, rule base
l	Rule, $l \in \mathbb{L}$
L	Number of rules, $L = \ \mathbb{L}\ $
$\mathbf{a}, \mathbf{b}$	Fuzzy linguistic terms for premise and consequence
$\mathbb{A}$	Set representing premise of the fuzzy rule
$\mathbb{B}$	Triangle set in consequence
$\mathbb{B}'$	Fuzzy set of rule's implication
$\rightsquigarrow$	Fuzzy implication
$\star$	T-norm
F	Firing strength

See page 3 for general rule for symbols

parameterized consequences is presented in Fig. 1. The figure is taken from (Czogala et al. 2000) with modifications.

2.1 Architecture of the system

The system with parameterized consequences is the MISO system. The rule base $\mathbb{L}$ contains fuzzy rules l in form of fuzzy implications

$$l: \mathbf{x} \text{ is } \mathbf{a} \rightsquigarrow y \text{ is } \mathbf{b}, \quad (1)$$

where $\mathbf{x} = [x_1, x_2, \dots, x_N]^T$ and y are linguistic variables, $\mathbf{a}$ and $\mathbf{b}$ are fuzzy linguistic terms (values). Data tuples are represented by vectors $[\mathbf{x}, y]^T$, where $\mathbf{x}$ is a vector of descriptors and y is the decision attribute of the tuples. Both the descriptors and decision are real numbers.

Fig. 1 The scheme of the neuro-fuzzy system with parameterized consequences. The input has two attributes and the rule base is composed of two fuzzy rules. The premises of the rules are responsible for determining the firing strength of the rules. The firing strength is the left operand of the fuzzy implication. The right hand operand is the $\mathbb{B}$ fuzzy triangle set, the location of which

determined by formula 7. The result of the r th fuzzy implication is a fuzzy set $\mathbb{B}'_r$. The fuzzy results of the implications are then aggregated. The non-informative part (the gray rectangular in the picture) is thrown away in aggregation. The informative part (the white mountain-like part of $\mathbb{B}'$ set) is then defuzzified with the centre of gravity method. The defuzzified answer of the system is number y_0

In the following text we will describe the situation only for one rule, but we will omit the index of the rule in the following formulae as not to complicate the notation.

The linguistic variable α (in the rule's premise) is represented in the system as a fuzzy set $\mathbb{A}$ in N -dimensional space. Each fuzzy rule has its own premise and consequence. For each dimension n the set $\mathbb{A}_n$ is described with the Gaussian membership function:

$$u_{\mathbb{A}_n}(x_n) = \exp\left(-\frac{(x_n - v_n)^2}{2s_n^2}\right), \quad (2)$$

where v_n is the core location for n th attribute and s_n is this attribute Gaussian bell deviation (fuzziness).

The membership of a tuple $\mathbf{x}$ to the premise $\mathbb{A}$ of the rule is the T-norm of memberships to all dimensions in the rule's premise. Because each dimension i has its own weight z_i , we use the weighted T-norm (Rutkowski and Cpałka 2003) to determine the membership of the data example to the fuzzy set $\mathbb{A}$ in rule's premise:

$$u_{\mathbb{A}} = T(u_{\mathbb{A}_1}, \dots, u_{\mathbb{A}_N}; z_1, \dots, z_N) = T(1 - z_1(1 - u_{\mathbb{A}_1}), \dots, 1 - z_N(1 - u_{\mathbb{A}_N})). \quad (3)$$

In the system the product T-norm is used so the above Eq. (3) is expressed as:

$$u_{\mathbb{A}} = T(u_{\mathbb{A}_1}, \dots, u_{\mathbb{A}_N}; z_1, \dots, z_N) = \prod_{n=1}^N (1 - z_n(1 - u_{\mathbb{A}_n})). \quad (4)$$

Membership of a data tuple to the fuzzy set in l th rule's premise is the firing strength of the rule for the tuple (from now on we use the rule's index l)

$$F_l(\mathbf{x}) = u_{l\mathbb{A}}(\mathbf{x}) \in [0, 1]. \quad (5)$$

To avoid misunderstandings please keep in mind the meanings of the symbols: $u_{\mathbb{A}_n}$ stands for membership of the n th descriptor to the fuzzy set $\mathbb{A}_n$ in the premise for n th attribute of a certain rule (the index of which we omit here) as in formulae 2, 3, 4, $u_{l\mathbb{A}}$ stands for membership of the whole data tuple to the premise of the l th rule—it is l th rule's firing strength (as in formula 5).

Combining 2 and 4 we get firing strength F of l th rule for data vector (tuple) $\mathbf{x}$:

$$F_l(\mathbf{x}) = \prod_{n=1}^N \left(1 - z_{ln}^f \left\{ 1 - \exp\left[-\frac{(x_n - v_{ln})^2}{2s_{ln}^2}\right] \right\} \right). \quad (6)$$

The term b (in formula (1)) describing the l th rule's consequence is represented by an isosceles triangle fuzzy set $\mathbb{B}_l$ with the base width w_l , the altitude of the triangle

equals 1. The localisation y_l of the core of the triangle fuzzy set is determined by linear combination of input attribute values with attribute weights taken into account:

$$y_l = \mathbf{p}_l^T \cdot \text{diag}([1, \mathbf{z}_l^T]) \cdot [1, \mathbf{x}^T]^T$$

$$= [p_{l0}, p_{l1}, \dots, p_{lN}] \cdot \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & z_{l1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & z_{lN} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_N \end{bmatrix}. \quad (7)$$

The above formula 7 can also be written as

$$y_l = \sum_{n=1}^N p_{ln} z_{ln} x_n + p_{l0} = \sum_{n=0}^N p_{ln} z_{ln} x_n, \quad (8)$$

where $z_{l0} = 1$ and $x_0 = 1$.

The output of the l th rule is the fuzzy value of the fuzzy implication:

$$u_{l\mathbb{B}'}(\mathbf{x}) = u_{l\mathbb{A}}(\mathbf{x}) \rightsquigarrow u_{l\mathbb{B}}(\mathbf{x}), \quad (9)$$

where squiggle arrow ($\rightsquigarrow$) stands for fuzzy implication. The shape of the fuzzy set $\mathbb{B}'$ depends on the used fuzzy implication (Czogala et al. 2000). In our system we use Reichenbach implication (Reichenbach et al. 1935)

$$p \rightsquigarrow q = 1 - p + pq. \quad (10)$$

The answers $u_{l\mathbb{B}'}$ of all L rules are then aggregated into one fuzzy answer of the system:

$$u_{\mathbb{B}'}(\mathbf{x}) = \bigoplus_{l=1}^L u_{l\mathbb{B}'}(\mathbf{x}), \quad (11)$$

where $\bigoplus$ stands for the aggregation operator. In order to get the non-fuzzy answer y_0 the fuzzy set $\mathbb{B}'$ is defuzzified with MICO method (Czogala et al. 2000). This approach removes the non-informative parts of the aggregated fuzzy sets and takes into account only the informative parts (cf. description of Fig. 1). The aggregation and defuzzification may be quite expensive, but it has been proved (Czogala et al. 2000) that the defuzzified system output can be expressed as:

$$y_0 = \frac{\sum_{l=1}^L g_l(\mathbf{x}) y_l(\mathbf{x})}{\sum_{l=1}^L g_l(\mathbf{x})}. \quad (12)$$

The function g depends on the fuzzy implication, in the system the Reichenbach one is used, so for the l th rule function g is

$$g_l(\mathbf{x}) = \frac{w_l}{2} F_l(\mathbf{x}). \quad (13)$$

The forms of g function for various implications can be found in the original work introducing the ANNBFIS system (Czogala et al. 2000). Some inaccuracies are discussed in Nowicki (2006) and Łęski (2008).

2.2 Creation of the fuzzy model

Creation of the fuzzy model (fuzzy rule base) is done in three steps: partition of the input domain (Sect. 2.2.1), extraction of rules' premises (Sect. 2.2.2) and tuning of the rules (this step is also responsible for creation of rules consequences)—Sect. 2.2.3.

2.2.1 Partition of the input domain

For domain partition we use modification (Simiński 2012) of the FCM clustering algorithm (Dunn 1973) where the weights are the values from the interval $[0, 1]$. Thus each cluster is fuzzy in two ways:

1. Data tuples have fuzzy membership to clusters. The sum of membership of one data tuple to all clusters is 1 (cf. Eq. 16). This is common in the fuzzy clustering paradigm.
2. The cluster itself has fuzzy possession of attributes. This means that the cluster spreads in a fuzzy way upon the dimensions. The sum of dimensions weight of one cluster is 1 (cf. Eq. 16).

Our clustering method is based on minimising the criterion function J

$$J = \sum_{c=1}^C \sum_{i=1}^X u_{ci}^m \sum_{n=1}^N z_{cn}^f (x_{in} - v_{cn})^2. \quad (14)$$

where m and $f \neq 1$ (the case of $f = 1$ is discussed on Eq. 20) are parameters, u_{ci} stands for membership of i th data example ($\mathbf{x}_i$) to c th cluster, z_{cn} stands for weight of n th attribute (descriptor) in c th cluster, x_{in} is n th descriptor of i th data tuple, v_{cn} is n th attribute of centre of c th cluster.

The centre of c th cluster is defined as

$$\mathbf{v}_c = \frac{\sum_{i=1}^X u_{ci} \mathbf{x}_i}{\sum_{i=1}^X u_{ci}}. \quad (15)$$

Two constraints are put on dimension weights and partition matrix:

1. The sum of membership values to all clusters for each data tuple is one:

$$\forall i \in [1, X] : \sum_{c=1}^C u_{ci} = 1. \quad (16)$$

2. The sum of dimension weights z for all dimensions N in each cluster c equals one:

$$\forall c \in [1, C] : \sum_{n=1}^N z_{cn} = 1. \quad (17)$$

Applying of Lagrange multipliers leads to the following formulae:

$$u_{ci} = \frac{\left(\sum_{n=1}^N z_{cn}^f (x_{in} - v_{cn})^2\right)^{\frac{1}{1-m}}}{\sum_{j=1}^C \left(\sum_{n=1}^N z_{jn}^f (x_{in} - v_{jn})^2\right)^{\frac{1}{1-m}}}, \quad (18)$$

$$z_{ci} = \frac{\left(\sum_{k=1}^X u_{ck}^m (x_{ki} - v_{ci})^2\right)^{\frac{1}{1-f}}}{\sum_{n=1}^N \left(\sum_{k=1}^X u_{ck}^m (x_{kn} - v_{cn})^2\right)^{\frac{1}{1-f}}}. \quad (19)$$

The data are clustered by alternating application of formulae 15, 18 and 19.

The procedure described above cannot be used if $f = 1$. The objective function (14) becomes

$$J = \sum_{l=1}^C \sum_{i=1}^X u_{li}^m \sum_{n=1}^N z_{ln} (x_{in} - v_{ln})^2. \quad (20)$$

In such a case the attribute n of the l th rule for which the sum

$$\sum_{i=1}^X u_{li}^m z_{ln} (x_{in} - v_{ln})^2 \quad (21)$$

is minimal gets the weight $z_{ln} = 1$ and other attributes of this rule get zero weights (because of the constraint expressed by formula 17).

2.2.2 Extraction of rules

The clustering procedure elaborates memberships and weights gathered in matrices $\mathbf{U} = \{u_{ij}\}$ and $\mathbf{Z} = \{z_{ij}\}$ respectively which are then converted into premises' parameters v , s and z . The number of rules is equal to the number of clusters: $L = C$.

The cores v of rules' premises are calculated with formula 15. The fuzzification parameter s is calculated with formula (Czogala et al. 2000)

$$s_i = \sqrt{\frac{\sum_{j=1}^X u_{ij}^m (v_i - x_j)^2}{\sum_{j=1}^X u_{ij}^m}}. \quad (22)$$

The extraction of the weights of attributes is slightly more complicated. The constraint expressed by formula 17 makes the sum of all weights in a rule equal one. If two attributes have weights greater than zero, their values have to be lower than one. If all N attributes have the same weights, their weight is $z = 1/N$ (cf. Eq. 17) and if firing strengths of all attributes are the same and equal F_n , the firing strength of the whole rule is (cf. Eq. 6)

$$F(N, F_n) = \prod_{n=1}^N [1 - z^f (1 - F_n)] = \left[1 - \frac{1}{N^f} (1 - F_n)\right]^N. \quad (23)$$

Fig. 2 Firing strength F for the whole rule (Eq. 6) when all attributes have the equal firing strength F_n in function of number of attributes (N) and attribute's weight exponent $f = 2$ without augmentation. If the weights of the attributes are not augmented the firing strength of the whole rule tends to one independently whether the attributes are fired or not. The figure comprises 11 draws for values from $F_n = 0.0$ to 1.0 with 0.1 step. The gray lines are only to join the firing strengths for the same F_n values. They have no physical meaning, because the number of attributes N has only discrete values

If all attributes are minimally fired (zero firing strengths) the total firing strength of the whole rule tends to one (with increase in the number of attributes), so there is no difference if the attributes are fired or not. This is highly unsatisfactory. The Fig. 2 presents this phenomenon.

This can be easily avoided by augmenting of the weights of the attributes in a rule. The attribute weights for one rule are divided by the maximal values of them. This maximal values is always greater than zero. In this procedure all weights in this rule are scaled and the maximum weights become one:

$$\forall l \in \mathbb{L} : z_{ln} \leftarrow \frac{z_{ln}}{\max_{i \in [1..N]} z_{li}}. \quad (24)$$

2.2.3 Tuning of rule parameters

In neuro-fuzzy systems the parameters of the model are tuned to better fit the data. In this system the parameters of the premises (v and s in Eq. 2, z in Eq. 4) and the values of the supports w of the sets in consequences are tuned with the gradient method. The linear coefficients p (Eq. 7) for the calculation of the localisation of the consequence sets are calculated with the pseudoinverse matrix. For tuning parameters of the model the square error is used

$$E = \frac{(y - y_0)^2}{2}, \quad (25)$$

where y is the original value and y_0 is the value elaborated by the system (cf. Eq. 12). For q parameter in j th rule the differential has the following form:

$$\frac{\partial E}{\partial q_j} = \frac{\partial E}{\partial y_0} \cdot \frac{\partial y}{\partial g} \cdot \frac{\partial g}{\partial F} \cdot \frac{\partial F}{\partial q_j}. \quad (26)$$

Formula (26) is valid for v , s and z parameters. For width w of the isosceles triangle in the rule's consequence the following formula is used:

$$\frac{\partial E}{\partial w_i} = \frac{\partial E}{\partial y} \cdot \frac{\partial y}{\partial g} \cdot \frac{\partial g}{\partial w_i}. \quad (27)$$

The differentials in Eq. 26 are:

$$\frac{\partial E}{\partial y_0} = -(y - y_0) \quad (28)$$

and

$$\frac{\partial y_0}{\partial g} = \frac{y_j - y_0}{\sum_{i=1}^L g(F_i(\mathbf{x}), w_i)}. \quad (29)$$

The differentials $\frac{\partial g}{\partial F}$ and $\frac{\partial g}{\partial w}$ depend on the used implication (cf. Eq. 13). For Reichenbach implication we have:

$$\frac{\partial g}{\partial F} = \frac{w}{2} \quad (30)$$

and

$$\frac{\partial g}{\partial w} = \frac{F}{2}. \quad (31)$$

For q_j being v_{jm} parameter (the core of the m th attribute in j th rule) we get (cf. Eq. 6)

$$\begin{aligned} \frac{\partial F}{\partial v_{jm}} &= \prod_{n=1, n \neq m}^N \left(1 - z_{in}^f \left\{ 1 - \exp \left[-\frac{(x_n - v_{in})^2}{2s_{in}^2} \right] \right\} \right) \cdots \\ &\cdots z_{jm}^f \exp \left[-\frac{(x_n - v_{jm})^2}{2s_{jm}^2} \right] \cdot \left[2 \frac{(x_n - v_{jm})}{2s_{jm}^2} \right]. \end{aligned} \quad (32)$$

For q_j being s_{jm} parameter (the fuzzification of the m th attribute in j th rule) in Eq. 26 we get:

$$\begin{aligned} \frac{\partial F}{\partial s_{jm}} &= \prod_{n=1, n \neq j}^N \left(1 - z_{in}^f \left\{ 1 - \exp \left[-\frac{(x_n - v_{in})^2}{2s_{in}^2} \right] \right\} \right) \cdots \\ &\cdots z_{jm}^f \exp \left[-\frac{(x_n - v_{jm})^2}{2s_{jm}^2} \right] \cdot \left[2 \frac{(x_n - v_{jm})}{2s_{jm}^2} \right]. \end{aligned} \quad (33)$$

And finally for q_j being z_{jm} (the weight of the m th attribute in j th rule) parameter in Eq. 26 we get:

$$\begin{aligned} \frac{\partial F}{\partial z_{jm}} &= \prod_{n=1, n \neq j}^N \left(1 - z_{in}^f \left\{ 1 - \exp \left[-\frac{(x_n - v_{in})^2}{2s_{in}^2} \right] \right\} \right) \cdots \\ &\cdots \left\{ -1 + \exp \left[-\frac{f(x_n - v_{jm})^2}{2s_{jm}^2} \right] \right\}. \end{aligned} \quad (34)$$

The linear parameters for localisation of the cores of triangle fuzzy sets in consequences are calculated as a solution to the linear equation expressed by Eq. 7. To avoid numerical problems the pseudoreverse matrix is calculated. In the calculation the weights are also taken into account.

For $f = 0$ (which switches off the attributes' weights) the proposed system is identical with ANNBFIS system described in (Czogala et al. 2000).

3 Experiments

The experiments were conducted on real-life data sets depicting methane concentration, death rate, breast cancer recurrence time, concrete compressive strength and ozone concentration. All real life data sets are normalised (to mean 0 and standard deviation 1). Some parameters of data sets are gathered in Table 2.

3.1 Data set description

The 'Methane' data set contains the real life measurements of air parameters in a coal mine in Upper Silesia (Poland). The parameters (measured in 10 s intervals) are: AN31—the flow of air in the shaft, AN32—the flow of air in the adjacent shaft, MM32—concentration of methane (CH_4), production of coal, the day of week. The 10-min sums of measurements of AN31, AN32, MM32 are added to the tuples as dynamic attributes (Sikora et al. 2005). The task is to predict the concentration of the methane in 10 min. The data is divided into a train set (499 tuples) and test set (523 tuples).

The 'Death' data represent the tuples containing information on various factors, the task is to estimate the death rate (Späth 1992). The first attribute (the index) is excluded from the dataset. The precise description of the attributes is available with the data set, the names of the attributes are listed in Table 7, so the description is omitted here. The data can be downloaded from a public repository.¹

Table 2 Number of tuples and attributes in the real life data sets

Data set	Number of tuples		Number of attributes
	Train	Test	
'Concrete'	515	515	8
'Methane'	499	523	7
'Death'	30	30	15
'Breast cancer'	97	97	32
'Ozone'	924	923	72

¹ <http://orion.math.iastate.edu/burkardt/data/regression/x28.txt>.

Table 3 Root mean square error elaborated by our system

f	'Death'		'Methane'		'Concrete'		'Breast cancer'	
	DA	KG	DA	KG	DA	KG	DA	KG
0.0	1.4590	2.6367	0.4210	0.4021	1.0464	1.2609	2.4243	5.5836
0.3	1.7372	2.0286	0.3884	0.3204	0.4404	1.1930	1.3124	3.0173
0.5	1.0276	1.5364	0.3959	0.3051	0.4623	0.8925	2.0098	3.7444
0.8	0.2840	1.1046	0.3999	0.2993	0.4988	0.9203	0.7836	1.2458
1.2	0.5746	0.9959	0.4012	0.3225	0.8744	0.9804	0.9460	0.9996
1.3	0.6983	4.2052	0.3852	0.3147	0.6579	0.7481	0.7990	1.0337
1.4	0.5698	0.8276	0.4381	0.3633	0.5603	0.7725	0.6221	1.6346
1.5	0.2365	1.5162	0.3874	0.3169	0.5327	0.7539	0.5953	1.4051
1.6	0.4799	2.0888	0.3871	0.3528	0.7585	0.7949	0.5359	1.7692
1.7	0.0029	1.9332	0.3923	0.3169	0.5845	0.7240	0.6330	1.2517
1.8	0.0031	1.7067	0.3915	0.3081	0.6582	0.7191	0.5396	1.4417
1.9	0.0034	1.5482	0.3902	0.2963	0.6606	0.7125	0.7560	1.2005
2.0	0.0034	3.2544	0.4168	0.3220	0.6586	0.7289	0.7527	1.1810
2.5	0.0018	3.3457	0.5989	0.7574	0.6625	0.7575	0.6927	1.2029
3.0	0.0014	3.2089	0.3969	0.2998	0.9120	1.0742	0.6879	2.4426
3.5	0.0025	2.7339	0.3757	0.8383	0.9081	1.0679	0.4396	2.8398
4.0	0.0074	2.4294	0.3757	0.8211	0.9045	1.0623	0.3597	3.0749
4.5	0.0122	1.9510	0.3879	0.7762	0.6557	0.7181	0.5227	2.4006
5.0	0.0102	1.9008	0.3918	0.2997	0.6615	0.7530	0.5197	2.5246
10.0	0.0113	1.8498	0.4597	0.3769	0.6683	0.7097	0.7905	2.6919
20.0	0.0127	1.9301	0.4001	0.2978	0.6660	0.7550	0.9721	3.0642

Number of rules $L = 5$. The values for $f = 0$ are elaborated by the ANNBFIS system (where no weights are assigned to attributes)

The 'Breast cancer' data set represents the data for the breast cancer case (Asuncion and Newman 2007). Each data tuple contains 32 continuous attributes and one predictive attribute (the time to recur). Here again we will omit the description of attributes, their names are listed in Table 6. The symbol 'se' in the attribute's name stands for 'standard error' and the adjective 'worst' means the 'largest'. The data can be downloaded² from a public repository (Frank et al. 2010).

The 'Concrete' set is a real life data set describing the parameters of the concrete sample and its strength (Yeh 1998). The attributes are: cement ratio, amount of blast furnace slag, fly ash, water, superplasticizer, coarse aggregate, fine aggregate, age; the decision attribute is the concrete compressive strength. The original data set can be downloaded³ from public repository (Frank et al. 2010).

The 'Ozone' set—a real life data set—describes the level of ozone in the air (Zhang and Fan 2008). The data set includes 2536 tuples with 73 attributes. The original data set can be downloaded⁴ from a public repository

(Frank et al. 2010). The data set has 687 tuples with missing values, these were deleted from the data set and 1847 full tuples were left. The first attribute (date) was deleted from the tuples. The tuples were numbered starting with 1. The tuples with odd numbers are used as a train set (924 tuples), the even numbered tuples constitute the test data set (923 tuples). All attributes are real numbers. The task is to predict the level of ozone (high 1 or low 0).

3.2 Results of experiments

The fuzzy models were created with train sets. The number of rules is always the same as the number of clusters and was assumed a priori as $L = 5$ (for the 'Ozone' dataset $L = 3$). Finding the optimal number of clusters in clustering is a difficult task. Our aim here is not to discuss this problem, but to compare the precision of our system with the one already existing. This is why we assume the a priori number of rules.

The experiments were conducted in two paradigms. In the first one—data approximation (DA)—the models are created and tested with the same train data sets. In the other—knowledge generalisation (KG)—the models are created with train data sets and tested with unseen tuples of test data sets.

² [http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)).

³ <http://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength>.

⁴ <http://archive.ics.uci.edu/ml/datasets/Ozone+Level+Detection>.

Fig. 3 Root mean square errors for the real life data sets. The results for data approximation (DA) are denoted with *times signs*, for knowledge generalisation (KG)—*black squares*. The symbols are accompanied by auxiliary lines for higher readability

Root mean square error (RMSE) measure is used to evaluate the elaborated results:

$$RMSE = \sqrt{\frac{1}{X} \sum_{i=0}^X (y(\mathbf{x}_i) - y_0(\mathbf{x}_i))^2}, \quad (35)$$

where $y(\mathbf{x}_i)$ stands for original (expected) value for i th tuple and $y_0(\mathbf{x}_i)$ is the value elaborated by the system; X is the number of tuples.

Fig. 4 The values elaborated for the 'Death' data set (KG). The original values are marked with the black squares, the values elaborated by ANNBFIS by the *gray line*, elaborated by NFS with weighted attributes by the *black line*. Number of rules $L = 5$, $f = 2$

Two main features were tested: (1) the precision of created models and (2) the weights assigned to dimensions (attributes).

3.3 Precision of models

Table 3 presents the RMSE results elaborated for various values of f parameter. For $f = 0$ the system elaborates the same results as ANNBFIS system. The results gathered in Table 3 are also presented as graphs in Fig. 3a–d.

The experiments reveal that for $f \in [1.5, 2]$ the RMSE, elaborated by the systems for various data sets, achieves its most advantageous values both for DA and KG. For $f > 2$ the KG error starts to grow, whereas the DA error is kept on more or less the same level. The optimal interval of f parameter seems independent from the data sets.

The Figs. 4, 5, 6 and 7 present the comparison of results elaborated by ANNBFIS (the gray lines) and our system (the black lines) with $f = 2$. The squares denote the expected values. The Figs. 8 and 9 present in a more detailed way the result for tuples 50–150 and 250–400 respectively. Similarly, the Fig. 10 presents the details of the Fig. 5 for tuples 400–500.

The figures show that applying attribute weights in the fuzzy rule base results in a more precise prediction. The better rule prediction can be observed in Fig. 8, where the expected values are better elaborated by our systems (the black line) than by the original ANNBFIS system that does not use attribute weights in rules (the gray line).

3.3.1 Weights of attributes

Another feature tested in experiments are the weights assigned to attributes (dimensions). The Tables 4, 5, 6, 7 and Figs. 11, 12, 13 and 14 present the weights of attributes in models elaborated for real life data sets.

The attributes' weights for the 'Methane' data set (prediction of methane concentration in a coal mine shaft) gathered in Table 4 and presented in Fig. 11 show a very

Fig. 5 The values elaborated for the ‘Methane’ data set (KG). The original values are marked with the black squares, the values elaborated by ANNBFIS by the gray line, elaborated by NFS with weighted attributes by the black line. Number of rules $L = 5$, $f = 2$

Fig. 6 The values elaborated for the ‘Breast cancer’ data set (KG). The original values are marked with the black squares, the values elaborated by ANNBFIS by the gray line, elaborated by NFS with weighted attributes by the black line. Number of rules $L = 5$, $f = 2$

Fig. 7 The values elaborated for the ‘Concrete’ data set (KG). The original values are marked with the black squares, the values elaborated by ANNBFIS by the gray line, elaborated by NFS with weighted attributes by the black line. Number of rules $L = 5$, $f = 2$

interesting fact: the actual concentration of methane (the third attribute) turned out to be of minor importance in all the rules, although the task was the 10-min prediction of the concentration of the methane in the shaft. The most important attributes are the flow of air in the mine shaft (the first attribute) and the production of coal (the fourth attribute). It can be explained by the fact that excavation of coal causes tensions and splits in the rock that may release the methane gas. In two rules the most important attribute is the first one, the flow of the air in the shaft in question. In the fifth rule an interesting phenomenon can be observed. The most important attribute is the fifth one, the 10-min

Fig. 8 The values elaborated for the ‘Concrete’ data set (KG). The original values are marked with the black squares, the values elaborated by ANNBFIS by the gray line, elaborated by NFS with weighted attributes by the black line. Number of rules $L = 5$, $f = 2$. The figure presents in a more detailed way the part of Fig. 7

Fig. 9 The values elaborated for the ‘Concrete’ data set (KG). The original values are marked with the black squares, the values elaborated by ANNBFIS by the gray line, elaborated by NFS with weighted attributes by the black line. Number of rules $L = 5$, $f = 2$. The figure presents in a more detailed way the part of Fig. 7

Fig. 10 The values elaborated for the ‘Methane’ data set (KG). The original values are marked with the black squares, the values elaborated by ANNBFIS by the gray line, elaborated by NFS with weighted attributes by the black line. Number of rules $L = 5$, $f = 2$. The figure presents in a more detailed way the part of Fig. 5

sum of the first attribute (flow of the air), whereas the first attribute itself has lower weight. The similar situation occurs in the case of the second attribute (flow of the air in the adjacent shaft), where the sum of air flow measurements in the adjacent shaft (the sixth attribute) is more important than the summed air flow itself (the second attribute).

The weights of attributes elaborated for the ‘Concrete’ data set are presented in Table 5 and Fig. 12. The most

Table 4 Weights of attributes elaborated for the ‘Methane’ data set (cf. Fig. 11)

Attribute	Attributes’ weights in rules				
	I	II	III	IV	V
AN31: flow of air in the shaft	1.000	0.000	0.132	1.000	0.361
AN32: flow of air in the adjacent shaft	0.009	0.000	0.087	0.002	0.102
MM32: concentration of methane	0.004	0.000	0.065	0.002	0.058
Production of coal	0.028	1.000	1.000	0.011	0.885
Sum of AN31	0.020	0.000	0.156	0.005	1.000
Sum of AN32	0.009	0.000	0.085	0.004	0.291
Sum of MM32	0.005	0.000	0.093	0.002	0.059

Table 5 Weights of attributes elaborated for the ‘Concrete’ data set (cf. Fig. 12)

Attribute	Attributes’ weights in rules				
	I	II	III	IV	V
Cement ratio	0.000	0.065	0.006	0.065	0.005
Blast furnace slag	0.000	0.050	0.019	0.535	1.000
Fly ash	1.000	0.008	0.121	0.332	0.007
Water	0.000	0.013	0.190	0.086	0.003
Superplasticizer	0.000	0.057	0.057	0.112	0.004
Coarse aggregate	0.000	0.028	0.006	0.090	0.003
Fine aggregate	0.000	1.000	1.000	0.063	0.002
Age	0.000	0.027	0.031	1.000	0.004

important attributes (all others have low weights) are: blast furnace slag (the second attribute), ratio of fly ash (the third attribute), fine aggregate (the seventh attribute) and age of concrete (the eighth attribute). In one rule the weights are more varied: the most important attribute is age, but concentration of blast furnace slag and fly ash have also quite high weights.

The weights of attributes elaborated for the ‘Breast cancer’ data set are presented in Table 6 and Fig. 6. In rule I the most important attribute is the first one (lymph nodes), which is in concordance with medical diagnose procedures. In all the rules the importance of three attributes: radius mean (the second attribute), perimeter mean (the fourth attribute) and area mean (the fifth attribute) are correlated. In rule III there are two triples of attributes of higher importance. The triple of high importance comprises: area worst (the 25th attribute), perimeter worst (the 24th attribute) and radius worst (the 22nd attribute). This triple is accompanied by the triple of slightly lower importance: area mean (the fifth attribute), perimeter mean (the fourth attribute) and radius mean (the second attribute). In one rule the weights are more varied. The important attributes are fractal dimension worst (the 31st attribute), fractal

Table 6 Weights of attributes elaborated for the ‘Breast cancer’ data set (cf. Fig. 13)

Attribute	Attributes’ weights in rules				
	I	II	III	IV	V
Lymph_node	1.000	0.008	0.030	0.061	0.044
Radius_mean	0.001	0.009	0.277	0.208	0.953
Texture_mean	0.001	0.015	0.037	0.147	0.039
Perimeter_mean	0.001	0.009	0.293	0.194	1.000
Area_mean	0.001	0.008	0.470	0.163	0.910
Smoothness_mean	0.000	0.035	0.051	0.198	0.047
Compactness_mean	0.000	0.016	0.041	0.241	0.039
Concavity_mean	0.000	0.021	0.061	0.108	0.059
Concave_points_mean	0.001	0.012	0.090	0.160	0.054
Symmetry_mean	0.000	0.037	0.041	0.123	0.033
Fractal_dimension_mean	0.001	0.022	0.037	0.431	0.046
Radius_se	0.001	0.028	0.022	0.223	0.059
Texture_se	0.000	0.052	0.028	0.431	0.034
Perimeter_se	0.000	0.029	0.026	0.252	0.052
Area_se	0.000	0.019	0.064	0.300	0.076
Smoothness_se	0.001	0.094	0.076	0.543	0.025
Compactness_se	0.000	0.040	0.069	0.356	0.064
Concavity_se	0.000	0.049	0.052	0.190	0.048
Concave_points_se	0.001	0.085	0.035	0.123	0.032
Symmetry_se	0.000	1.000	0.051	0.175	0.066
Fractal_dimension_se	0.000	0.030	0.044	0.608	0.032
Radius_worst	0.001	0.009	0.544	0.196	0.379
Texture_worst	0.000	0.023	0.038	0.165	0.077
Perimeter_worst	0.001	0.010	0.336	0.204	0.197
Area_worst	0.001	0.009	1.000	0.156	0.332
Smoothness_worst	0.000	0.020	0.033	0.231	0.079
Compactness_worst	0.000	0.016	0.029	0.482	0.079
Concavity_worst	0.001	0.018	0.021	0.174	0.104
Concave_points_worst	0.001	0.016	0.036	0.145	0.052
Symmetry_worst	0.000	0.057	0.028	0.185	0.050
Fractal_dimension_worst	0.001	0.014	0.021	1.000	0.106
Tumor_size	0.003	0.016	0.023	0.051	0.044

dimension standard deviation (the 21st attribute) and fractal dimension mean (the 11th attribute), smoothness standard deviation (the 16th attribute) and compactness worst (the 27th attribute).

The weights of attributes elaborated for the ‘Death’ data set are presented in Table 7 and Fig. 14. In rules I, III and V the most important attributes are the hydrocarbon pollution index (the 12th attribute) and the nitric oxide pollution index (the 13th attribute). It is interesting that the pollution index for sulphur dioxide has low weight in all rules. In the second rule the most important attribute describes scholarisation of persons over 22 (the sixth attribute).

Table 7 Weights of attributes elaborated for the ‘Death’ data set (cf. Fig. 14)

Attribute	Attributes’ weights in rules				
	I	II	III	IV	V
Average annual precipitation	0.002	0.000	0.002	0.090	0.001
Average January temperature	0.007	0.000	0.001	0.328	0.001
Average July temperature	0.003	0.000	0.001	0.022	0.001
Size of the population older than 65	0.020	0.000	0.000	0.006	0.002
Number of members per household	0.009	0.000	0.001	0.002	0.002
Years of schooling for persons over 22	0.008	1.000	0.001	0.189	0.001
Households with fully equipped kitchens	0.001	0.000	0.001	0.012	0.001
Population per square mile	0.002	0.055	0.001	0.005	0.001
Size of the nonwhite population	0.002	0.000	0.001	1.000	0.001
Number of office workers	0.066	0.013	0.001	0.054	0.001
Families with an income <3000	0.002	0.001	0.001	0.011	0.001
Hydrocarbon pollution index	1.000	0.000	1.000	0.056	1.000
Nitric oxide pollution index	0.269	0.000	0.440	0.023	0.110
Sulphur dioxide pollution index	0.003	0.000	0.021	0.001	0.004
Degree of atmospheric moisture	0.015	0.000	0.002	0.103	0.000

Fig. 11 Weights of attributes elaborated for the ‘Methane’ data set (cf. Table 4) **Fig. 12** Weights of attributes elaborated for the ‘Concrete’ data set (cf. Fig. 12)

The experiments were also executed on the ‘Ozone’ data set. This real life data set comprises 72 attributes describing the meteorological measurements (the original data set has 73 attributes, but the first one—the date—has been deleted as mentioned in the data set’s description above). The attributes are not listed here, their short description is available at the data repository, from which the data set can be downloaded. The tuples are labelled 0 or 1. The task is to classify the unseen data examples. Our system was trained with 0 or 1 labels, but it elaborates the real value answer. The answers lower than 0.5 were labelled with zero, otherwise with one. The experiments were conducted with ANNBFIS and our subspace neuro-fuzzy systems. The ANNBFIS system assigned the major class to all the answers. The subspace approach elaborated more precise results (precision: 0.926). The weights of attributes are presented in Fig. 15. In the next experiment only the attributes with weights higher than 0.7 in at least two rules were selected. This led to the selection of attributes 27 to 53. All these attributes describe the results of temperature measurements. The results were a bit poorer (precision: 0.921) than in the case when all attributes were used.

4 Summary

The paper describes the novel neuro-fuzzy system with weighted attributes. In this approach the attributes in a fuzzy

Fig. 13 Weights of attributes elaborated for the 'Breast cancer' data set (cf. Table 6)

Fig. 14 Weights of attributes elaborated for the 'Death' data set (cf. Table 7)

rule have weights. The weights of attributes are numbers from the interval $[0, 1]$. The weights of the attributes are not assigned globally, but each fuzzy rule has its own weights of

Fig. 15 Weights of attributes elaborated for the 'Ozone' data set

attributes. Each rule exists in its own subspace. An attribute can be important in a certain rule, but unimportant in another. This approach is inspired by subspace clustering, but in our system the attribute can have partial weight, which is uncommon in subspace clustering where attributes have full (1) or none (0) weights in a subspace.

There are two main advantages of the approach proposed in the paper:

1. The experiments show that fuzzy models with weighted attributes can elaborate more precise results, both for data approximation and knowledge generalisation for real life data sets in comparison with a neuro-fuzzy system that does not assign weights to attributes.
2. Assigning weights to attributes discovers knowledge on importance of attributes in a problem. Individual weights of attributes in each rule discover the relation between attributes. This may explain why the weights of the same attribute are low in one rule and high in another one.

The experiments show that assigned weights of the attributes are in concordance with experts' knowledge on the physical or medical mechanisms described by the data sets.

Acknowledgments The author is grateful to the anonymous reviewers for their constructive comments that have helped to improve the paper.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Abonyi J, Babuška R, Szeifert F (2002) Modified Gath–Geva fuzzy clustering for identification of Takagi–Sugeno fuzzy models. *IEEE Trans Syst Man Cybern B* 32(5):612–621
- Aggarwal CC, Wolf JL, Yu PS, Procopiuc C, Park JS (1999) Fast algorithms for projected clustering. *SIGMOD Rec* 28(2):61–72. doi:10.1145/304181.304188

- Aggarwal CC, Yu PS (2000) Finding generalized projected clusters in high dimensional spaces. In: SIGMOD '00: Proceedings of the 2000 ACM SIGMOD international conference on management of data. ACM, New York, pp 70–81. doi:[10.1145/342009.335383](https://doi.org/10.1145/342009.335383)
- Agrawal R, Gehrke J, Gunopulos D, Raghavan P (1998) Automatic subspace clustering of high dimensional data for data mining applications. SIGMOD Rec 27(2):94–105. doi:[10.1145/276305.276314](https://doi.org/10.1145/276305.276314)
- Asuncion A, Newman DJ (2007) UCI machine learning repository. University of California, School of Information and Computer Sciences, Irvine, CA
- Bauman E, Dorofeyuk A (1990) Fuzzy identification of nonlinear dynamical systems. In: Proceedings of the international conference on fuzzy logic and neural nets, pp 895–898
- Chen JQ, Xi YG, Zhang ZJ (1998) A clustering algorithm for fuzzy model identification. Fuzzy Sets Syst 98(3):319–329. doi:[10.1016/S0165-0114\(96\)00384-3](https://doi.org/10.1016/S0165-0114(96)00384-3)
- Cheng CH, Fu AW, Zhang Y (1999) Entropy-based subspace clustering for mining numerical data. In: KDD '99: Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, New York, pp 84–93. doi:[10.1145/312129.312199](https://doi.org/10.1145/312129.312199)
- Czogala E, Leski J (2000) Fuzzy and neuro-fuzzy intelligent systems. Series in fuzziness and soft computing. Physica-Verlag, Heidelberg
- Dunn JC (1973) A fuzzy relative of the ISODATA process and its use in detecting compact, well separated clusters. J Cybern 3(3): 32–57
- Frank A, Asuncion A (2010) UCI machine learning repository
- Friedman JH, Meulman JJ (2004) Clustering objects on subsets of attributes. J R Statist Soc B 66:815–849
- Gan G, Wu J (2008) A convergence theorem for the fuzzy subspace clustering (FSC) algorithm. Pattern Recogn 41(6):1939–1947. doi:[10.1016/j.patcog.2007.11.011](https://doi.org/10.1016/j.patcog.2007.11.011)
- Gan G, Wu J, Yang Z (2006) A fuzzy subspace algorithm for clustering high dimensional data. In: Advanced data mining and applications, second international conference, ADMA 2006, Xi'an, China, August 14–16, 2006, Proceedings. Lecture notes in computer science, vol 4093. Springer, Berlin, pp 271–278
- Goil S, Goil S, Nagesh H, Nagesh H, Choudhary A, Choudhary A (1999) Mafia: efficient and scalable subspace clustering for very large data sets. Tech rep (1999)
- Hoffmann F, Nelles O (2001) Genetic programming for model selection of tsf-fuzzy systems. Inf Sci 136:7–28
- Jakubek S, Keuth N (2006) A local neuro-fuzzy network for high-dimensional models and optimization. In: Engineering applications of artificial intelligence, pp 705–717
- Jang JSR (1993) ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybern 23(3):665–684
- Kriegel HP, Kroger P, Zimek A (2009) Clustering high-dimensional data: a survey on subspace clustering, pattern-based clustering, and correlation clustering. ACM Trans Knowl Discov Data (TKDD) 3(1):1–58. doi:[10.1145/1497577.1497578](https://doi.org/10.1145/1497577.1497578)
- Łęski J (2008) Systemy neuronowo-rozmyte [Neuro-fuzzy systems]. Wydawnictwa Naukowo-Techniczne, Warszawa. ISBN 978-83-204-3229-9
- Łęski J, Czogala E (1999) A new artificial neural network based fuzzy inference system with moving consequents in if-then rules and selected applications. Fuzzy Sets Syst 108(3):289–297. doi:[10.1016/S0165-0114\(97\)00314-X](https://doi.org/10.1016/S0165-0114(97)00314-X)
- Mamdani EH, Assilian S (1975) An experiment in linguistic synthesis with a fuzzy logic controller. Int J Man-Mach Stud 7(1):1–13
- Müller E, Günemann S, Assent I, Seidl T (2009) Evaluating clustering in subspace projections of high dimensional data. Proc VLDB Endow 2(1):1270–1281
- Nelles O, Fink A, Babuška R, Setnes M (2000) Comparison of two construction algorithms for Takagi–Sugeno fuzzy models. Int J Appl Math Comput Sci 10(4):835–855
- Nelles O, Isermann R (1996) Basis function networks for interpolation of local linear models. In: Proceedings of the 35th IEEE conference on decision and control, vol 1, pp 470–475
- Nowicki R (2006) Rough-neuro-fuzzy system with MICOG defuzzification. In: 2006 IEEE international conference on fuzzy systems. Vancouver, Canada, pp 1958–1965. doi:[10.1109/FUZZY.2006.1681972](https://doi.org/10.1109/FUZZY.2006.1681972)
- Parsons L, Haque E, Liu H (2004) Subspace clustering for high dimensional data: a review. SIGKDD Explor Newsl 6(1):90–105. doi:[10.1145/1007730.1007731](https://doi.org/10.1145/1007730.1007731)
- Reichenbach H (1935) Wahrscheinlichkeitslogik. Erkenntnis 5: 37–43. doi:[10.1007/BF00172280](https://doi.org/10.1007/BF00172280)
- Rutkowski L, Cpałka K (2003) Flexible neuro-fuzzy systems. IEEE Trans Neural Netw 14(3):554–574. doi:[10.1109/TNN.2003.811698](https://doi.org/10.1109/TNN.2003.811698)
- Setnes M, Babuška R (2001) Rule base reduction: some comments on the use of orthogonal transforms. IEEE Trans Syst Man Cybern C Appl Rev 31(2):199–206. doi:[10.1109/5326.941843](https://doi.org/10.1109/5326.941843)
- Sikora M, Krzykowski D (2005) Application of data exploration methods in analysis of carbon dioxide emission in hard-coal mines dewater pump stations. Mech Autom Mining 413(6)
- Sim K, Gopalkrishnan V, Zimek A, Cong G (2012) A survey on enhanced subspace clustering. In: Data mining and knowledge discovery, pp 1–66. doi:[10.1007/s10618-012-0258-x](https://doi.org/10.1007/s10618-012-0258-x)
- Simiński K (2008) Neuro-fuzzy system with hierarchical partition of input domain. Studia Inf 29(4A (80, 43–53):43–53
- Simiński K (2009) Patchwork neuro-fuzzy system with hierarchical domain partition. In: Kurzyński M, Woźniak M (eds.) Computer Recognition Systems 3. Advances in intelligent and soft computing, vol 57. Springer, Berlin, pp 11–18. doi:[10.1007/978-3-540-93905-4_2](https://doi.org/10.1007/978-3-540-93905-4_2)
- Simiński K (2010) Rule weights in neuro-fuzzy system with hierarchical domain partition. Int J Appl Math Comput Sci 20(2):337–347. doi:[10.2478/v10006-010-0025-3](https://doi.org/10.2478/v10006-010-0025-3)
- Simiński K (2012) Clustering in fuzzy subspaces. Theoret Appl Inf 24(4):313–326. doi:[10.2478/v10179-012-0019-y](https://doi.org/10.2478/v10179-012-0019-y)
- Späth H (1992) Mathematical algorithms for linear regression. Academic Press Professional, Inc., San Diego
- Sugeno M, Kang GT (1988) Structure identification of fuzzy model. Fuzzy Sets Syst 28(1):15–33 9
- Takagi T, Sugeno M (1985) Fuzzy identification of systems and its application to modeling and control. IEEE Trans Syst Man Cybern 15(1):116–132
- Wang L, Langari R (1994) Building Sugeno-type models using fuzzy discretization and orthogonal parameter estimation techniques. NAFIPS/IFIS/NASA '94. In: Proceedings of the first international joint conference of the north american fuzzy information processing society biannual conference. The industrial fuzzy control and intelligent systems conference, and the NASA Joint Technolo, pp 201–206. doi:[10.1109/IJCF.1994.375098](https://doi.org/10.1109/IJCF.1994.375098)
- Yang J, Wang W, Wang H, Yu P (2002) δ -clusters: capturing subspace correlation in a large data set. In: Proceedings 18th international conference on data engineering, 2002, pp 517–528
- Yeh IC (1998) Modeling of strength of high-performance concrete using artificial neural networks. Cement Concrete Res 28(12): 1797–1808. doi:[10.1016/S0008-8846\(98\)00165-3](https://doi.org/10.1016/S0008-8846(98)00165-3)
- Zadeh LA (1973) Outline of a new approach to the analysis of complex systems and decision processes. IEEE Trans Syst Man Cybern SMC-3, pp 28–44
- Zhang K, Fan W (2008) Forecasting skewed biased stochastic ozone days: analyses, solutions and beyond. Knowl Inf Syst 14(3): 299–326