Flexible Neuro-Fuzzy Systems

Leszek Rutkowski, *Senior Member, IEEE*, and Krzysztof Cpalka

***Abstract*—In this paper, we derive new neuro-fuzzy structures called flexible neuro-fuzzy inference systems or FLEXNFIS. Based on the input–output data, we learn not only the parameters of the membership functions but also the type of the systems (Mamdani or logical). Moreover, we introduce: 1) softness to fuzzy implication operators, to aggregation of rules and to connectives of antecedents; 2) certainty weights to aggregation of rules and to connectives of antecedents; and 3) parameterized families of T-norms and S-norms to fuzzy implication operators, to aggregation of rules and to connectives of antecedents. Our approach introduces more flexibility to the structure and design of neuro-fuzzy systems. Through computer simulations, we show that Mamdani-type systems are more suitable to approximation problems, whereas logical-type systems may be preferred for classification problems.**

***Index Terms*—Certainty weights, logical approach, Mamdani approach, neuro-fuzzy inference systems (NFIS).**

I. Introduction

In the last decade, various neuro-fuzzy systems have been developed (see, e.g., [4]–[7], [10], [19], [21], [26], [27], [29]–[33], [37]–[39], [41]–[48], [64]–[66]). They combine the natural language description of fuzzy systems and the learning properties of neural-networks. Some of them are known in the literature under short names like ANFIS [20], ANNBFIS [7], DENFIS [23], FALCON [31], GARIC [2], NEFCLASS [38], NEFPROX [37], [38], SANFIS [57] and others. In this paper, we study a wide class of fuzzy systems trained by the back propagation method. Following other authors we call them neuro-fuzzy inference systems (NFIS). To emphasize their main feature-flexibility, we also use name FLEXNFIS.

In the literature to date, two approaches [7], [44], [58], [67] have been proposed to design fuzzy systems.

1) The first approach, called the Mamdani method, uses conjunction for inference and disjunction to aggregate individual rules. In the Mamdani approach, the most widely used operators measuring the truth of the relation between input and output are the following:

$$I(a,b) = \min\{a,b\} \tag{1}$$

and

$$I(a,b) = a \cdot b \tag{2}$$

or more generally

$$I(a,b) = T\{a,b\}. \tag{3}$$

It should be emphasized that formulas (1) and (2) do not satisfy the conditions of fuzzy implication formulated by Fodor [11]. We refer to (1) and (2) as to "engineering implications" (see Mendel [34], [35]) contrary to the fuzzy implications satisfying the axiomatic definition (see Definition 1).

The aggregation is performed by an application of S-norm

$$S\{a_1, a_2, \ldots, a_n\} = S\{\mathbf{a}\} = a_1 \overset{S}{*} a_2 \overset{S}{*} \ldots \overset{S}{*} a_n = S_{i=1}^{n}\{a_i\} \tag{4}$$

e.g.,

$$S\{\mathbf{a}\} = \max_{i=1,\ldots,n}\{a_i\}. \tag{5}$$

It should also be noted that in most cases the aggregation of rules is performed as a part of defuzzification (see, e.g., [35] and [58]).

2) The second paradigm applies fuzzy implications to inference and conjunction to aggregation.

Definition 1. (Fuzzy Implication): A fuzzy implication is a function $I:[0,1]^2 \to [0,1]$ satisfying the following conditions:

(I1) if $a_1 \le a_3$, then $I(a_1,a_2) \ge I(a_3,a_2)$, for all $a_1, a_2, a_3 \in [0,1]$.

(I2) if $a_2 \le a_3$, then $I(a_1,a_2) \le I(a_1,a_3)$, for all $a_1, a_2, a_3 \in [0,1]$.

(I3) $I(0,a_2) = 1$, for all $a_2 \in [0,1]$ (falsity implies anything).

(I4) $I(a_1,1) = 1$, for all $a_1 \in [0,1]$ (anything implies tautology).

(I5) $I(1,0) = 0$ (Booleanity).

Selected fuzzy implications satisfying the above conditions are listed in Table I. In this table, implications 1–4 are examples of an S-implication associated with an S-norm

$$I(a,b) = S\{1-a,b\} \tag{6}$$

e.g.,

$$I(a,b) = \max\{1-a,b\}. \tag{7}$$

For fuzzy systems with a logical implication, the aggregation is realized by a T-norm

$$T\{a_1, a_2, \ldots, a_n\} = T\{\mathbf{a}\} = a_1 \overset{T}{*} a_2 \overset{T}{*} \ldots \overset{T}{*} a_n = T_{i=1}^{n}\{a_i\} \tag{8}$$

e.g.,

$$T\{\mathbf{a}\} = \min_{i=1,\ldots,n}\{a_i\}. \tag{9}$$

Neuro-fuzzy inference systems of a logical-type are described in Section III-B. It should be noted that the aggregation of antecedents in each rule is performed by the same formula (8) for both Mamdani and logical-type systems.

Manuscript received June 20, 2002; revised February 18, 2003.

The authors are with the Department of Computer Engineering, Technical University of Czestochowa, Czestochowa 42 200, Poland.

Digital Object Identifier 10.1109/TNN.2003.811698

TABLE I
FUZZY IMPLICATIONS

No	Name	Implication $I(a,b)$
1	Kleene-Dienes	$\max\{1-a,b\}$
2	Łukasiewicz	$\min\{1,1-a+b\}$
3	Reichenbach	$1-a+a\cdot b$
4	Fodor	$\begin{cases}1 & \text{if } a\le b\\ \max\{1-a,b\} & \text{if } a>b\end{cases}$
5	Sharp	$\begin{cases}1 & \text{if } a\le b\\ 0 & \text{if } a>b\end{cases}$
6	Goguen	$\begin{cases}1 & \text{if } a=0\\ \min\left\{1,\frac{b}{a}\right\} & \text{if } a>0\end{cases}$
7	Gödel	$\begin{cases}1 & \text{if } a\le b\\ b & \text{if } a>b\end{cases}$
8	Yager	$\begin{cases}1 & \text{if } a=0\\ b^a & \text{if } a>0\end{cases}$
9	Zadeh	$\max\{\min\{a,b\},1-a\}$
10	Willmott	$\min\left\{\begin{matrix}\max\{1-a,b\},\\ \max\{a,1-b,\min\{1-a,b\}\}\end{matrix}\right\}$

It was emphasized by Yager [64], [65] that "no formal reason exists for the preponderant use of the Mamdani method in fuzzy logic control as opposed to the logical method other than inertia." Moreover, Yager said [66] that "as a matter of fact the Mamdani approach has some disadvantages: its inability to distinguish more specific information in the face of the rules span the whole input space." This statement was an inspiration for us to determine the type of fuzzy inference (Mamdani or logical) in the process of learning. We decided to study the problem, despite the widely held belief about the inferiority of the logical method (see Remark 2 in Section III).

In this paper, we present a novel approach to fuzzy modeling. The novelty is summarized as follows.

1) We propose a new class of NFIS characterized by automatic determination of a fuzzy inference (Mamdani/logical) in the process of learning. Consequently, the structure of the system is determined in the process of learning. This class is based on the definition of an H-function which becomes a T-norm or S-norm depending on a certain parameter ν which can be found in the process of learning. We refer to this class as OR-type fuzzy systems.
2) We develop AND-type neuro-fuzzy inference systems by making use of the concept of flexible structures studied by Yager and Filev [67]. The AND-type fuzzy inference systems exhibit simultaneously Mamdani and logical type inferences.
3) We introduce
    - softness to fuzzy implication operators, to aggregation of rules and to connectives of antecedents;
    - certainty weights to aggregation of rules and to connectives of antecedents;
    - parameterized families of T-norms and S-norms to fuzzy implication operators, to aggregation of rules and to connectives of antecedents

TABLE II
AND-TYPE SYSTEM

λ	System
0	Mamdani type
1	logical type
(0,1)	compromise (Mamdani **AND** logical)

in both AND-type and OR-type neuro-fuzzy inference systems.
4) Through computer simulations we show that Mamdani-type systems are more suitable to approximation problems, whereas logical-type systems may be preferred for classification problems. Moreover, we observe that the most influential parameters in FLEXNFIS are certainty weights (introduced in this paper in a novel way) in connectives of antecedents and in aggregations of rules. They significantly improve the performance of NFIS in the process of learning.

This paper is organized into eight sections. In the next section, we discuss and propose various flexibility issues in NFIS. In Section III, a formal description of NFIS is presented, which also provides a general architecture [Fig. 2 and formula (44)] of all systems (flexible and nonflexible) studied in this paper. In Section IV, we introduce an H-function and give a framework for the description, unification and development of NFIS. The OR-type and AND-type FLEXNFIS are studied in Sections V and VI, respectively. Section VII shows the simulation results and comparative studies with other neuro-fuzzy systems. Conclusions and discussions are drawn in Section VIII.

II. FLEXIBILITY IN NFIS

A. AND-Type Compromise NFIS

Obviously, the Mamdani and the logical systems lead to different results and, in the literature, there are no formal proofs as to which of them is superior. Therefore, Yager and Filev [67] proposed to combine both methods. The AND-type compromise NFIS is characterized by the simultaneous appearance of Mamdani-type and logical-type systems. In this paper, we study the following combination of "engineering" and fuzzy implications

$$I(a,b) = (1-\lambda)\,T\{a,b\} + \lambda S\{1-a,b\} \tag{10}$$

e.g.,

$$I(a,b) = (1-\lambda)\min\{a,b\} + \lambda\max\{1-a,b\}. \tag{11}$$

In Section VI, we develop compromise NFIS based on formula (10). It should be emphasized that parameter λ can be found in the process of learning subject to the constraint $0 \le \lambda \le 1$. In Section VII, based on the input-output data, we learn a system type starting from $\lambda = 0.5$ as an initial value. The behavior of the AND-type compromise NFIS is depicted in Table II.

B. OR-Type NFIS

OR-type NFIS have recently been proposed by Rutkowski and Cpalka [46], [47]. Depending on a certain parameter ν this class of systems exhibits "more Mamdani" $(0 < \nu < 0.5)$

TABLE III
OR-TYPE SYSTEM

v	System
0	Mamdani type
1	logical type
0.5	undefined
(0,0.5)	"more Mamdani"
(0.5,1)	"more logical"

or "more logical" ($0.5 < \nu < 1$) behavior. At the boundaries the system becomes more of a Mamdani-type ($\nu = 0$) or logical-type ($\nu = 1$). The definition of OR-type systems heavily relies on the concept of an H-function (see Section IV and Rutkowski and Cpalka [46], [47]). The H-function exhibits the behavior of fuzzy norms. More precisely, it is a T-norm for $\nu = 0$ and S-norm for $\nu = 1$. For $0 < \nu < 0.5$ the H-function resembles a T-norm and for $0.5 < \nu < 1$ the H-function resembles an S-norm. In a similar spirit, we construct OR-type implications. The parameter ν can be found in the process of learning subject to the constraint $0 \leq \nu \leq 1$. In Section VII, based on the input-output data, we learn a system type starting from $\nu = 0.5$ as an initial value. The behavior of the OR-type systems is shown in Table III (see Section IV for details). Observe that this system-contrary to the AND-type system-does not exhibit simultaneously Mamdani and logical features. It is strictly an OR-type system. The OR-type NFIS are studied in Section V.

C. Soft NFIS

The soft versions of operators (8) and (4) were proposed by Yager and Filev [67]. They are defined as follows:

$$\tilde{T}\{\mathbf{a};\alpha\} = (1-\alpha)\frac{1}{n}\sum_{i=1}^{n} a_i + \alpha T\{\mathbf{a}\} \tag{12}$$

and

$$\tilde{S}\{\mathbf{a};\alpha\} = (1-\alpha)\frac{1}{n}\sum_{i=1}^{n} a_i + \alpha S\{\mathbf{a}\} \tag{13}$$

where $\alpha \in [0,1]$.

In the same spirit, we define softening of "engineering implication" (3) by

$$\tilde{I}(a,b;\beta) = (1-\beta)\frac{1}{2}(a+b) + \beta T\{a,b\} \tag{14}$$

and logical "fuzzy implication" (6) by

$$\tilde{I}(a,b;\beta) = (1-\beta)\frac{1}{2}(1-a+b) + \beta S\{1-a,b\} \tag{15}$$

where $\beta \in [0,1]$.

The soft compromise NFIS are studied in Sections V-B, V-C, VI-B, and VI-C.

D. NFIS Realized by Parameterized Families of T-Norms and S-Norms

Most fuzzy inference structures studied in the literature employ the triangular norms shown in Table IV. There is only a little knowledge within the engineering community about so-called parameterized families of T-norm and S-norms. They include the Aczél-Alsina, Dombi, Dubois-Prade, Frank, Hamacher, Schweizer-Sklar, Sugeno-Weber, and Yager families [28].

TABLE IV
BASIC TRIANGULAR NORMS

No	Name	T-norm S-norm
1	Mamdani	$T_M\{a_1,a_2\} = \min\{a_1,a_2\}$ $S_M\{a_1,a_2\} = \max\{a_1,a_2\}$
2	Product	$T_P\{a_1,a_2\} = a_1 a_2$ $S_P\{a_1,a_2\} = a_1 + a_2 - a_1 a_2$
3	Łukasiewicz	$T_L\{a_1,a_2\} = \max\{a_1+a_2-1,0\}$ $S_L\{a_1,a_2\} = \min\{a_1+a_2,1\}$
4	Drastic	$T_D\{a_1,a_2\} = \begin{cases} 0 & \text{if } S_M\{a_1,a_2\}<1 \\ T_M\{a_1,a_2\} & \text{if } S_M\{a_1,a_2\}=1 \end{cases}$ $S_D\{a_1,a_2\} = \begin{cases} 1 & \text{if } T_M\{a_1,a_2\}>0 \\ S_M\{a_1,a_2\} & \text{if } T_M\{a_1,a_2\}=0 \end{cases}$

It should be noted that these parameterized families include the triangular norms listed in Table IV. For example, the Dombi family is defined as follows:

1) the Dombi T-norm

$$\overset{\leftrightarrow}{T}_D\{\mathbf{a};p\} = \begin{cases} T_D\{\mathbf{a}\} & \text{if } p=0 \\ T_M\{\mathbf{a}\} & \text{if } p=\infty \\ \dfrac{1}{1+\left(\sum_{i=1}^{n}\left(\frac{1-a_i}{a_i}\right)^p\right)^{1/p}} & \text{if } p\in(0,\infty) \end{cases} \tag{16}$$

where $\overset{\leftrightarrow}{T}_D$ stands for the T-norm of a Dombi family parameterized by p;

2) the Dombi S-norm

$$\overset{\leftrightarrow}{S}_D\{\mathbf{a};p\} = \begin{cases} S_D\{\mathbf{a}\} & \text{if } p=0 \\ S_M\{\mathbf{a}\} & \text{if } p=\infty \\ 1-\dfrac{1}{1+\left(\sum_{i=1}^{n}\left(\frac{a_i}{1-a_i}\right)^p\right)^{1/p}} & \text{if } p\in(0,\infty) \end{cases} \tag{17}$$

where $\overset{\leftrightarrow}{S}_D$ stands for the S-norm of a Dombi family parameterized by p.

The parameter $p \in [0,\infty)$ can be found in the process of learning.

Obviously formula (16) defines the "engineering implication." Combining (6) and (17) we get the fuzzy S-implication generated by the Dombi family

$$\overset{\leftrightarrow}{I}_D(a,b;p) = 1-\frac{1}{1+\left(\left(\frac{1-a}{a}\right)^p+\left(\frac{b}{1-b}\right)^p\right)^{1/p}}. \tag{18}$$

The NFIS realized by parameterized families of T-norms and S-norms are studied in Sections V-B, V-C, VI-B, and VI-C.

E. NFIS Realized by T-Norms and S-Norms With Weighted Arguments

In this paper, we propose the weighted T-norm

$$T^*\{a_1,a_2;w_1,w_2\} = T\{1-w_1(1-a_1),1-w_2(1-a_2)\}. \tag{19}$$

Parameters a_1 and a_2 can be interpreted as antecedents of a rule. The weights w_1 and w_2 are corresponding certainties (credibilities) of both antecedents.

Observe the following.

1) If $w_1 = w_2 = 1$, then the weighted T-norm (19) is reduced to the standard T-norm. In the context of linguistic values we assign the truth to both antecedents a_1 and a_2 of the rule.
2) If $w_1 = 0$, then

$$T^* \{a_1, a_2; 0, w_2\} = T\{1, 1 - w_2(1 - a_2)\} = 1 - w_2(1 - a_2). \quad (20)$$

Therefore, the antecedent a_1 is discarded since its certainty is equal to zero. Similarly, if $w_2 = 0$ then the antecedent a_2 vanishes

$$T^* \{a_1, a_2; w_1, 0\} = T\{1 - w_1(1 - a_1), 1\} = 1 - w_1(1 - a_1). \quad (21)$$

3) If $0 < w_1 < 1$ and $0 < w_2 < 1$ then we assume a partial certainty of antecedents a_1 and a_2.

The S-norm corresponding to the T-norm (19) is defined as follows:

$$S^* \{a_1, a_2; w_1, w_2\} = S\{w_1 a_1, w_2 a_2\}. \quad (22)$$

In the same spirit we propose the weighted triangular norms

$$S^* \{a_1, a_2; w_1^{\text{agr}}, w_2^{\text{agr}}\} = S\{w_1^{\text{agr}} a_1, w_2^{\text{agr}} a_2\} \quad (23)$$

and

$$T^* \{a_1, a_2; w_1^{\text{agr}}, w_2^{\text{agr}}\} = T\{1 - w_1^{\text{agr}}(1 - a_1), 1 - w_2^{\text{agr}}(1 - a_2)\} \quad (24)$$

to aggregate individual rules in Mamdani-type and logical-type systems, respectively. The weights w_1 and w_2 in (19), as well as w_1^{agr} and w_2^{agr} in (23) or (24), can be found in the process of learning subject to the constraints w_1, w_2, w_1^{agr}, $w_2^{\text{agr}} \in [0, 1]$. In Sections V-C and VI-C we apply the weighted T-norm (19) to a selection of significant inputs, and the weighted S-norm (23) or T-norm (24) to a selection of important rules. The results are depicted in the form of diagrams in Section VII (dark areas correspond to low values of weights and vice versa).

Remark 1: It was pointed out by one of the reviewers that designing of neuro-fuzzy systems should be a compromise between accuracy of the model and its transparency. The measure of accuracy is usually the RMSE-criterion (approximation problems) and percentage of correct or wrong decisions (classification problems). The measure of transparency is the number and form of fuzzy rules obtained. It was mentioned by several authors (see, e.g., [1] and [14]) that the lack of transparency is a major drawback of many neuro-fuzzy systems. Most designers focus their effort on approximation accuracy, while the issue of transparency has received less attention. In this context our method of weighted triangular norms seems to be a promising tool for extracting both transparent and accurate rule-based knowledge from empirical data. More specifically, diagrams (weights representation) presented in Section VII can be used for the analysis and pruning of the fuzzy-rule bases in all the simulation examples. The FLEXNFIS realized by T-norms and S-norms with weighted arguments are studied in Sections V-C and VI-C. Note that our application of weights in NFIS is different from those studied in [16], [36], [55], and [69].

III. Formal Description of the NFIS

In this paper, we consider multi-input–single-output fuzzy NFIS mapping $\mathbf{X} \to \mathbf{Y}$, where $\mathbf{X} \subset \mathbf{R}^n$ and $\mathbf{Y} \subset \mathbf{R}$.

The fuzzifier performs a mapping from the observed crisp input space $\mathbf{X} \subset \mathbf{R}^n$ to the fuzzy sets defined in $\mathbf{Y}$. The most commonly used fuzzifier is the singleton fuzzifier which maps $\bar{\mathbf{x}} = [\bar{x}_1, \ldots, \bar{x}_n] \in \mathbf{X}$ into a fuzzy set $A' \subset \mathbf{X}$ characterized by the membership function

$$\mu_{A'}(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} = \bar{\mathbf{x}} \\ 0 & \text{if } \mathbf{x} \neq \bar{\mathbf{x}} \end{cases}. \quad (25)$$

The fuzzy rule base consists of a collection of N fuzzy IF-THEN rules, aggregated by disjunction or conjunction, in the form

$$R^{(k)} : \begin{cases} \text{IF} & x_1 & \text{is } A_1^k \text{ AND} \\ & x_2 & \text{is } A_2^k \text{ AND} \ldots \\ & x_n & \text{is } A_n^k \\ \text{THEN} & y & \text{is } B^k \end{cases} \quad (26)$$

or

$$R^{(k)} : \text{IF } \mathbf{x} \text{ is } A^k \text{ THEN } y \text{ is } B^k \quad (27)$$

where $\mathbf{x} = [x_1, \ldots, x_n] \in \mathbf{X}$, $y \in \mathbf{Y}$, $A^k = A_1^k \times A_2^k \times \cdots \times A_n^k$, $A_1^k, A_2^k, \ldots, A_n^k$ are fuzzy sets characterized by membership functions $\mu_{A_i^k}(x_i)$, $i = 1, \ldots, n$, $k = 1, \ldots, N$, whereas B^k are fuzzy sets characterized by membership functions $\mu_{B^k}(y)$, respectively, $k = 1, \ldots, N$.

The fuzzy inference determines a mapping from the fuzzy sets in the input space $\mathbf{X}$ to the fuzzy sets in the output space $\mathbf{Y}$. Each of N rules (26) determines a fuzzy set $\overline{B}^k \subset \mathbf{Y}$ given by the compositional rule of inference

$$\overline{B}^k = A' \circ \left(A^k \to B^k\right) \quad (28)$$

where $A^k = A_1^k \times A_2^k \times \cdots \times A_n^k$.

Fuzzy sets $\overline{B}^k$, according to the formula (28), are characterized by membership functions expressed by the *sup-star* composition

$$\mu_{\overline{B}^k}(y) = \sup_{\mathbf{x} \in \mathbf{X}} \left\{\mu_{A'}(\mathbf{x}) * \mu_{A_1^k \times \cdots \times A_n^k \to B^k}(\mathbf{x}, y)\right\} \quad (29)$$

where $*$ can be any operator in the class of T-norms. It is easily seen that for a crisp input $\bar{\mathbf{x}} \in \mathbf{X}$, i.e., a singleton fuzzifier (25), formula (29) becomes

$$\mu_{\bar{B}^k}(y) = \mu_{A_1^k \times \cdots \times A_n^k \to B^k}(\bar{\mathbf{x}}, y) = \mu_{A^k \to B^k}(\bar{\mathbf{x}}, y) = I\left(\mu_{A^k}(\bar{\mathbf{x}}), \mu_{B^k}(y)\right) \quad (30)$$

where $I(\cdot)$ is an "engineering implication" or fuzzy implication.

The aggregation operator, applied in order to obtain the fuzzy set B' based on fuzzy sets $\overline{B}^k$, is the T-norm or S-norm operator, depending on the type of fuzzy implication.

The defuzzifier performs a mapping from a fuzzy set B' to a crisp point $\bar{y}$ in $\mathbf{Y} \subset \mathbf{R}$. The COA (centre of area) method is defined by the following formula:

$$\bar{y} = \frac{\int_{\mathbf{Y}} y \cdot \mu_{B'}(y)\, dy}{\int_{\mathbf{Y}} \mu_{B'}(y)\, dy} \quad (31)$$

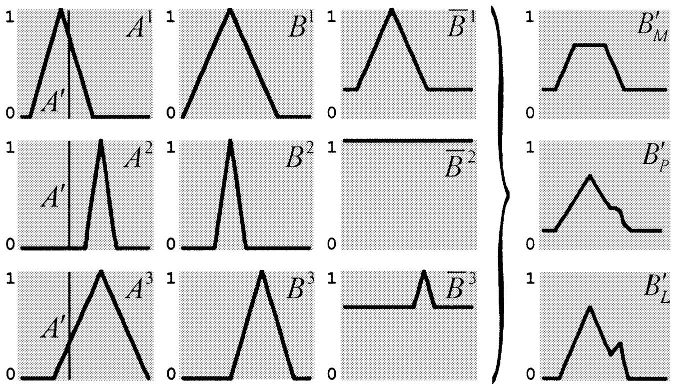

Fig. 1. Illustration of inference based on the binary implication, and Zadeh, product, and Lukasiewicz aggregations.

or by

$$\bar{y} = \frac{\sum_{r=1}^{N} \bar{y}^r \cdot \mu_{B'}(\bar{y}^r)}{\sum_{r=1}^{N} \mu_{B'}(\bar{y}^r)} \tag{32}$$

in the discrete form, where $\bar{y}^r$ denotes centres of the membership functions $\mu_{B^r}(y)$, i.e., for $r = 1, \ldots, N$

$$\mu_{B^r}(\bar{y}^r) = \max_{y \in \mathbf{Y}} \{\mu_{B^r}(y)\}. \tag{33}$$

For other definitions of the defuzzifier, the reader is referred to [7].

Remark 2: Several authors (e.g., Jager [18], Mendel [35]) reported problems with the application of logical implications to NFIS. A major problem is caused by the indeterminant part of the membership function. We illustrate such a situation in Fig. 1, showing the inference for binary implication (7). The aggregation is performed by making use of Zadeh T-norm B'_M, product T-norm B'_P and Lukasiewicz T-norm B'_L as listed in Table IV. Observe that there is no indeterminancy in the case of Lukasiewicz T-norm applied to aggregation.

The indicated problem can be easily resolved by the application of a modified center of gravity defuzzifier

$$\bar{y} = \frac{\int_{\mathbf{Y}_\alpha} y \cdot (\mu_{B'}(y) - \alpha)\, dy}{\int_{\mathbf{Y}_\alpha} (\mu_{B'}(y) - \alpha)\, dy} \tag{34}$$

where

$$\mathbf{Y}_\alpha = \{y \in \mathbf{Y} | \, \mu_{B'}(y) \geq \alpha\}. \tag{35}$$

The value $\alpha \in [0, 1]$ describes the indeterminancy that accompanies the corresponding part of information. It is easily seen that in order to eliminate the indeterminant part of the membership function $\mu_{B\prime}(y)$, the informative part has to be parallely shifted downward by the value of α. Neuro-fuzzy inference systems of a logical-type with defuzzifier (34) have been studied by Czogala and Leski [7].

Depending on implication (30), two types of NFIS can be distinguished.

A. Nonflexible NFIS: Mamdani-Type

In this approach, the implication (30) is a T-norm (e.g., minimum, product, Dombi)

$$I(\mu_{A^k}(\bar{\mathbf{x}}), \mu_{B^k}(\bar{y}^r)) = T\{\mu_{A^k}(\bar{\mathbf{x}}), \mu_{B^k}(\bar{y}^r)\} \tag{36}$$

and the aggregated output fuzzy set $B\prime \subset \mathbf{Y}$ is given by

$$\mu_{B\prime}(\bar{y}^r) = \mathop{S}_{k=1}^{N} \{\mu_{\bar{B}^k}(\bar{y}^r)\} = \mathop{S}_{k=1}^{N} \{T\{\mu_{A^k}(\bar{\mathbf{x}}), \mu_{B^k}(\bar{y}^r)\}\}. \tag{37}$$

Consequently, (32) takes the form

$$\bar{y} = \frac{\sum_{r=1}^{N} \bar{y}^r \cdot \mathop{S}_{k=1}^{N} \left\{ T\left\{ \mathop{T}_{i=1}^{n} \left\{\mu_{A_i^k}(\bar{x}_i)\right\}, \mu_{B^k}(\bar{y}^r)\right\}\right\}}{\sum_{r=1}^{N} \mathop{S}_{k=1}^{N} \left\{ T\left\{ \mathop{T}_{i=1}^{n} \left\{\mu_{A_i^k}(\bar{x}_i)\right\}, \mu_{B^k}(\bar{y}^r)\right\}\right\}}. \tag{38}$$

Obviously, the T-norms used to connect the antecedents in the rule and in the "engineering implication" do not have to be the same. Besides, they can be chosen as differentiable functions like Dombi families.

Remark 3: If

1) the implication is of a Mamdani-type;
2) $\mu_{B^k}(\bar{y}^r) = 0$ for $k \neq r$;

then formula (38) reduces to the well-known fuzzy system studied by Wang [58]

$$\bar{y} = \frac{\sum_{r=1}^{N} \bar{y}^r \cdot \mathop{T}_{i=1}^{n} \left\{\mu_{A_i^r}(\bar{x}_i)\right\}}{\sum_{r=1}^{N} \mathop{T}_{i=1}^{n} \left\{\mu_{A_i^r}(\bar{x}_i)\right\}}. \tag{39}$$

B. Nonflexible NFIS: Logical-Type

In this approach, the fuzzy implication (30) is an S-implication in the form

$$I(\mu_{A^k}(\bar{\mathbf{x}}), \mu_{B^k}(\bar{y}^r)) = S\{N(\mu_{A^k}(\bar{\mathbf{x}})), \mu_{B^k}(\bar{y}^r)\} \tag{40}$$

e.g., binary implication (known as the Kleene–Dienes implication)

$$I(\mu_{A^k}(\bar{\mathbf{x}}), \mu_{B^k}(\bar{y}^r)) = \max\{1 - \mu_{A^k}(\bar{\mathbf{x}}), \mu_{B^k}(\bar{y}^r)\}. \tag{41}$$

The aggregated output fuzzy set $B' \subset \mathbf{Y}$ is given by

$$\begin{aligned} \mu_{B\prime}(\bar{y}^r) &= \mathop{T}_{k=1}^{N} \{\mu_{\bar{B}^k}(\bar{y}^r)\} \\ &= \mathop{T}_{k=1}^{N} \{S\{N(\mu_{A^k}(\bar{\mathbf{x}})), \mu_{B^k}(\bar{y}^r)\}\} \end{aligned} \tag{42}$$

and formula (32) becomes

$$\bar{y} = \frac{\sum_{r=1}^{N} \bar{y}^r \cdot \mathop{T}_{k=1}^{N} \left\{ S\left\{ N\left(\mathop{T}_{i=1}^{n} \left\{\mu_{A_i^k}(\bar{x}_i)\right\}\right), \mu_{B^k}(\bar{y}^r)\right\}\right\}}{\sum_{r=1}^{N} \mathop{T}_{k=1}^{N} \left\{ S\left\{ N\left(\mathop{T}_{i=1}^{n} \left\{\mu_{A_i^k}(\bar{x}_i)\right\}\right), \mu_{B^k}(\bar{y}^r)\right\}\right\}}. \tag{43}$$

Now, we generalize both approaches described in points A and B and propose a general architecture of NFIS. It is easily seen that the systems (38) and (43) can be presented in the form

$$\bar{y} = f(\bar{\mathbf{x}}) = \frac{\sum_{r=1}^{N} \bar{y}^r \cdot \mathrm{agr}_r(\bar{\mathbf{x}}, \bar{y}^r)}{\sum_{r=1}^{N} \mathrm{agr}_r(\bar{\mathbf{x}}, \bar{y}^r)} \tag{44}$$

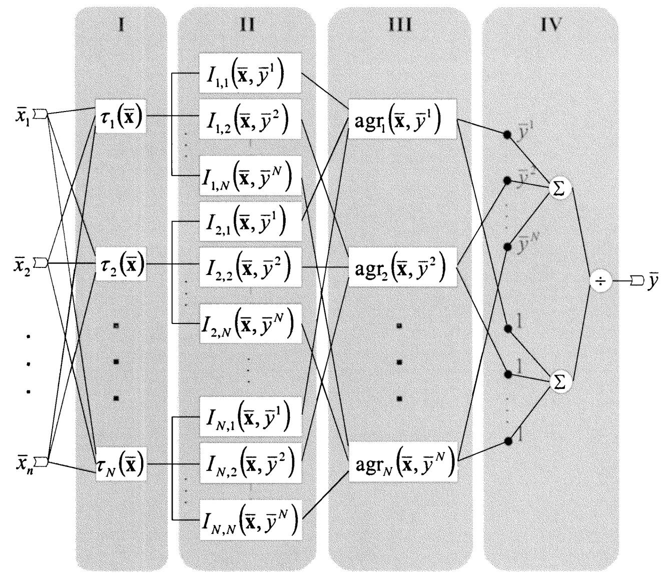

Fig. 2. General architecture of NFIS studied in this paper (flexible and nonflexible).

where

$$\operatorname{agr}_r(\bar{\mathbf{x}},\bar{y}^r)=\begin{cases}\overset{N}{\underset{k=1}{S}}\{I_{k,r}(\bar{\mathbf{x}},\bar{y}^r)\}, & \text{for Mamdani approach}\\ \overset{N}{\underset{k=1}{T}}\{I_{k,r}(\bar{\mathbf{x}},\bar{y}^r)\}, & \text{for logical approach}\end{cases} \tag{45}$$

and

$$I_{k,r}(\bar{\mathbf{x}},\bar{y}^r) =\begin{cases}T\{\tau_k(\bar{\mathbf{x}}),\mu_{B^k}(\bar{y}^r)\}, & \text{for Mamdani approach}\\ S\{N(\tau_k(\bar{\mathbf{x}})),\mu_{B^k}(\bar{y}^r)\}, & \text{for logical approach}\end{cases}. \tag{46}$$

Moreover, the firing strength of rules is given by

$$\tau_k(\bar{\mathbf{x}})=\overset{n}{\underset{i=1}{T}}\left\{\mu_{A_i^k}(\bar{x}_i)\right\}. \tag{47}$$

The general architecture of (44) is depicted in Fig. 2.

Remark 4: It should be emphasized that (44) and the scheme depicted in Fig. 2 are applicable to all the systems, flexible and nonflexible, studied in this paper with different definitions of $\operatorname{agr}_r(\bar{\mathbf{x}},\bar{y}^r)$ and $I_{k,r}(\bar{\mathbf{x}},\bar{y}^r)$. Nonflexible systems are described by (44), (45), (46) and (47), whereas flexible systems by (44) and $\operatorname{agr}_r(\bar{\mathbf{x}},\bar{y}^r)$, $I_{k,r}(\bar{\mathbf{x}},\bar{y}^r)$, $\tau_k(\bar{\mathbf{x}})$ defined in Sections V and VI. How we define the aggregation operator $\operatorname{agr}_r(\bar{\mathbf{x}},\bar{y}^r)$ and the implication operator $I_{k,r}(\bar{\mathbf{x}},\bar{y}^r)$, depends on the particular class of the system.

Remark 5: It is well known that the basic concept of the backpropagation algorithm, commonly used to train neural networks, can be also applied to any feedforward network. Let $\bar{\mathbf{x}}(t)\in\mathbf{R}^n$ and $d(t)\in\mathbf{R}$ be a sequence of inputs and desirable output signals, respectively. Based on the learning sequence $(\bar{\mathbf{x}}(1),d(1)),(\bar{\mathbf{x}}(2),d(2)),\ldots$ we wish to determine all parameters (including the system's type ν or λ) and weights of NFIS such that

$$e(t)=\frac{1}{2}\left[f(\bar{\mathbf{x}}(t))-d(t)\right]^2 \tag{48}$$

is minimized, were $f(\cdot)$ is given (44). The steepest descent optimization algorithm can be applied to solve this problem. For instance, the parameters $\bar{y}^r$, $r=1,\ldots,N$, are trained by the iterative procedure

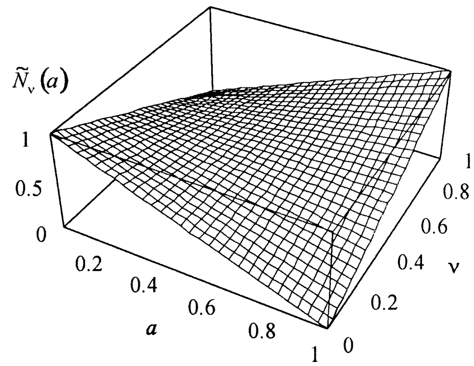

Fig. 3. 3-D plot of function (53).

$$\bar{y}^r(t+1)=\bar{y}^r(t)-\eta\frac{\partial e(t)}{\partial\bar{y}^r(t)}. \tag{49}$$

Directly calculating partial derivatives in recursion like (49) is rather complicated. Therefore, we recall that our NFIS has a layered architecture (Fig. 2) and apply the idea of back propagation to train the system. The exact recursions reflect that idea, however, they are not a copy of the standard backpropagation. For details, the reader is referred to our previous paper [47].

IV. Framework for Description, Unification, and Development of NFIS

In this section the following properties of dual T-norms and dual S-norms will be used

$$T\{\mathbf{a}\}=N(S\{N(a_1),N(a_2),\ldots,N(a_n)\}) \tag{50}$$
$$S\{\mathbf{a}\}=N(T\{N(a_1),N(a_2),\ldots,N(a_n)\}). \tag{51}$$

Our goal is to find a framework for the description, unification and development of all systems studied in this paper. We achieve this goal using two definitions (see Rutkowski and Cpalka [46]).

Definition 2. (Compromise Operator): A function

$$\tilde{N}_\nu:[0,1]\rightarrow[0,1] \tag{52}$$

given by

$$\begin{aligned}\tilde{N}_\nu(a)&=(1-\nu)N(a)+\nu N(N(a))\\&=(1-\nu)N(a)+\nu a\end{aligned} \tag{53}$$

is called a compromise operator where $\nu\in[0,1]$ and $N(a)=\tilde{N}_0(a)=1-a$.

Observe that

$$\tilde{N}_\nu(a)=\begin{cases}N(a), & \text{for } \nu=0\\ \frac{1}{2}, & \text{for } \nu=\frac{1}{2}\\ a, & \text{for } \nu=1\end{cases}. \tag{54}$$

Obviously, function $\tilde{N}_\nu$ is a strong negation (see, e.g., [28]) for $\nu=0$. The 3-D plot of function (53) is depicted in Fig. 3.

Remark 6: The formula (50) can be rewritten with the notation of definition 2

$$T\{\mathbf{a}\}=\tilde{N}_0\left(S\left\{\tilde{N}_0(a_1),\tilde{N}_0(a_2),\ldots,\tilde{N}_0(a_n)\right\}\right) \tag{55}$$

for $\nu = 0$. Apparently

$$S\{\mathbf{a}\} = \tilde{N}_1\left(S\left\{\tilde{N}_1(a_1), \tilde{N}_1(a_2), \ldots, \tilde{N}_1(a_n)\right\}\right) \quad (56)$$

for $\nu = 1$.

The right-hand sides of (55) and (56) can be written as follows:

$$\tilde{N}_\nu\left(\overset{n}{\underset{i=1}{S}}\left\{\tilde{N}_\nu(a_i)\right\}\right) \quad (57)$$

with $\nu = 0$ or $\nu = 1$. One may wish to vary the parameter ν in (57) from 0 to 1. This concept leads to the following definition, allowing us to switch smoothly between S-norm and T-norm:

Definition 3. (H-Function): A function

$$H : [0,1]^n \to [0,1] \quad (58)$$

given by

$$H(\mathbf{a};\nu) = \tilde{N}_\nu\left(\overset{n}{\underset{i=1}{S}}\left\{\tilde{N}_\nu(a_i)\right\}\right) = \tilde{N}_{1-\nu}\left(\overset{n}{\underset{i=1}{T}}\left\{\tilde{N}_{1-\nu}(a_i)\right\}\right) \quad (59)$$

is called an H-function where $\nu \in [0,1]$.

Observe that

$$H(\mathbf{a};\nu) = \begin{cases} T\{\mathbf{a}\}, & \text{for } \nu = 0 \\ \frac{1}{2}, & \text{for } \nu = \frac{1}{2} \\ S\{\mathbf{a}\}, & \text{for } \nu = 1 \end{cases}. \quad (60)$$

It is easily seen that for $0 < \nu < 0.5$ the H-function resembles a T-norm and for $0.5 < \nu < 1$ the H-function resembles an S-norm.

Example 1. (An Example of H-Function): We will show how to switch smoothly from T-norm to S-norm by making use of definition 3. Let $n = 2$ and the standard min-norm and max-conorm are chosen

$$H(a_1, a_2; 0) = T\{a_1, a_2\} = \min\{a_1, a_2\} \quad (61)$$
$$H(a_1, a_2; 1) = S\{a_1, a_2\} = \max\{a_1, a_2\}. \quad (62)$$

The H-function generated by formulas (61) and (62) takes the form

$$\begin{aligned} H(a_1, a_2; \nu) &= \tilde{N}_{1-\nu}\left(\min\left\{\tilde{N}_{1-\nu}(a_1), \tilde{N}_{1-\nu}(a_2)\right\}\right) \\ &= \tilde{N}_\nu\left(\max\left\{\tilde{N}_\nu(a_1), \tilde{N}_\nu(a_2)\right\}\right) \end{aligned} \quad (63)$$

and varies form (61) to (62) as ν goes from zero to one.

In Fig. 4, we illustrate function (63) for $\nu = 0.00$, $\nu = 0.15$, $\nu = 0.50$, $\nu = 0.85$, $\nu = 1.00$.

Example 2. (An Example of H-Implication): In this example, we illustrate how an H-implication based on definition three changes from "engineering implication" (1) to fuzzy implication (7). Let

$$\begin{aligned} I_{\text{eng}}(a,b) &= H(a,b;0) \\ &= T\{a,b\} \\ &= \min\{a,b\} \end{aligned} \quad (64)$$

and

$$\begin{aligned} I_{\text{fuzzy}}(a,b) &= H\left(\tilde{N}_0(a), b; 1\right) \\ &= S\{N(a), b\} \\ &= \max\{N(a), b\}. \end{aligned} \quad (65)$$

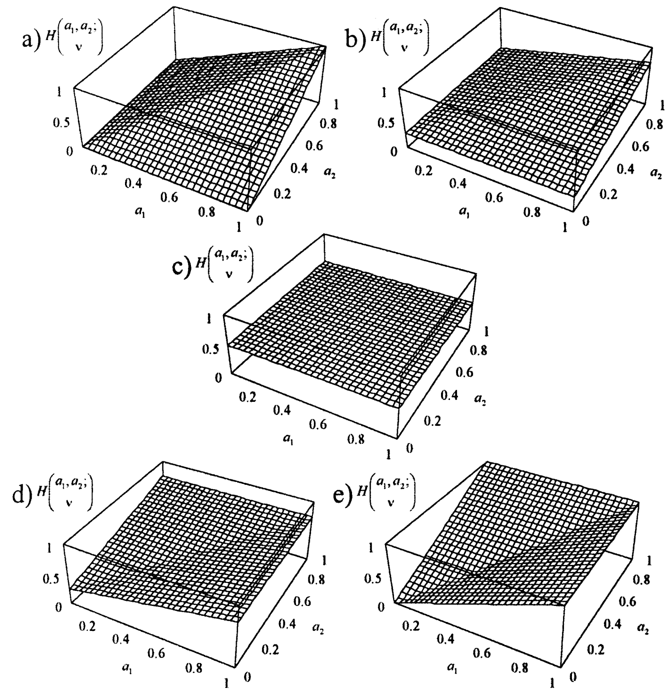

Fig. 4. 3-D plot of function (63) for (a) $\nu = 0.00$. (b) $\nu = 0.15$. (c) $\nu = 0.50$. (d) $\nu = 0.85$. (e) $\nu = 1.00$.

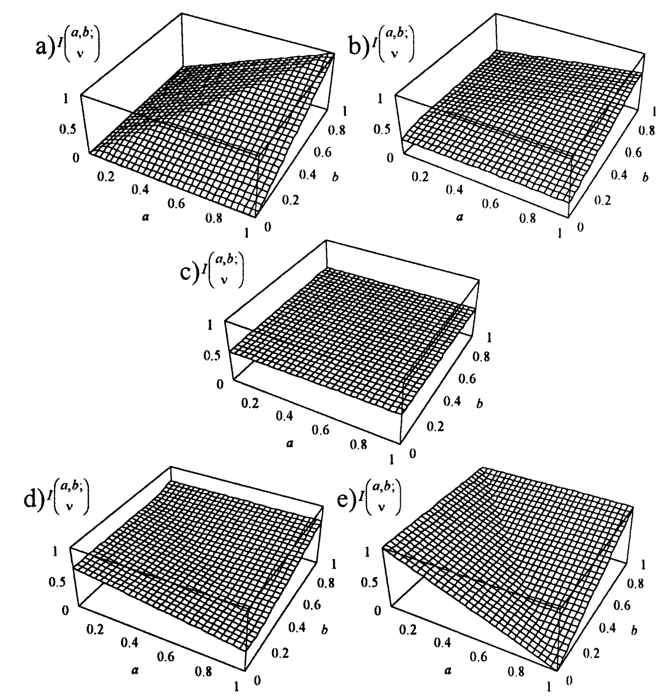

Fig. 5. 3-D plot of function (66) for (a) $\nu = 0.00$. (b) $\nu = 0.15$. (c) $\nu = 0.50$. (d) $\nu = 0.85$. (e) $\nu = 1.00$.

Then

$$I(a,b;\nu) = H\left(\tilde{N}_{1-\nu}(a), b; \nu\right) \quad (66)$$

goes from (64) to (65) as ν varies from zero to one.

In Fig. 5, we illustrate function (66) for $\nu = 0.00$, $\nu = 0.15$, $\nu = 0.50$, $\nu = 0.85$, $\nu = 1.00$.

Remark 7: It is easily seen that the nonflexible NFIS given by formulas (45)–(47) can be alternatively presented by making use of definition 3 with $\nu=0$ or $\nu=1$, shown in (67)–(69) at the bottom of the page.

V. OR-TYPE FLEXNFIS

The OR-type NFIS are based on definition 3 of the H-function. All the systems in this section are described by a general formula (44), see remark 4, with various definitions of $\tau_k(\bar{\mathbf{x}})$, $I_{k,r}(\bar{\mathbf{x}},\bar{y}^r)$ and $\mathrm{agr}_r(\bar{\mathbf{x}},\bar{y}^r)$.

A. Basic NFIS: OR-Type

The basic neuro-fuzzy system of an OR-type is given as follows:

OR I

$$\tau_k(\bar{\mathbf{x}}) = H\begin{pmatrix}\mu_{A_1^k}(\bar{x}_1),\ldots,\mu_{A_n^k}(\bar{x}_n);\\ 0\end{pmatrix} \tag{70}$$

$$I_{k,r}(\bar{\mathbf{x}},\bar{y}^r) = H\begin{pmatrix}\tilde{N}_{1-\nu}(\tau_k(\bar{\mathbf{x}})),\mu_{B^k}(\bar{y}^r);\\ \nu\end{pmatrix} \tag{71}$$

$$\mathrm{agr}_r(\bar{\mathbf{x}},\bar{y}^r) = H\begin{pmatrix}I_{1,r}(\bar{\mathbf{x}},\bar{y}^r),\ldots,I_{N,r}(\bar{\mathbf{x}},\bar{y}^r);\\ 1-\nu\end{pmatrix}. \tag{72}$$

Observe that system (70)–(72) is Mamdani-type for $\nu=0$, like Mamdani-type for $\nu\in(0,0.5)$, undetermined for $\nu=0.5$, like logical-type for $\nu\in(0.5,1)$ and logical-type for $\nu=1$. It is worth noticing that parameter ν can be learned and consequently the type of the system can be determined in the process of learning.

B. Soft NFIS: OR-Type

In this section we propose soft NFIS based on soft fuzzy norms (12) and (13). These systems are characterized by

1) soft strength of firing controlled by parameter α^τ;
2) soft implication controlled by parameter α^I;
3) soft aggregation of rules controlled by parameter α^{agr}.

Moreover, we assume that fuzzy norms (and H-function) in the connection of antecedents, implication and aggregation of rules are parameterized by parameters p^τ, p^I, p^{agr}, respectively. We use notation $\overleftrightarrow{H}(\cdot)$ to indicate parameterized families analogously to (16) and (17).

The soft compromise NFIS of an OR-type are defined as follows:

OR II

$$\tau_k(\bar{\mathbf{x}}) = \left((1-\alpha^\tau)\operatorname{avg}\left(\mu_{A_1^k}(\bar{x}_1),\ldots,\mu_{A_n^k}(\bar{x}_n)\right) + \alpha^\tau\overleftrightarrow{H}\begin{pmatrix}\mu_{A_1^k}(\bar{x}_1),\ldots,\mu_{A_n^k}(\bar{x}_n);\\ p^\tau,0\end{pmatrix}\right) \tag{73}$$

$$I_{k,r}(\bar{\mathbf{x}},\bar{y}^r) = \left(\left(1-\alpha^I\right)\operatorname{avg}\left(\tilde{N}_{1-\nu}(\tau_k(\bar{\mathbf{x}})),\mu_{B^k}(\bar{y}^r)\right) + \alpha^I\overleftrightarrow{H}\begin{pmatrix}\tilde{N}_{1-\nu}(\tau_k(\bar{\mathbf{x}})),\mu_{B^k}(\bar{y}^r);\\ p^I,\nu\end{pmatrix}\right) \tag{74}$$

$$\mathrm{agr}_r(\bar{\mathbf{x}},\bar{y}^r) = \left((1-\alpha^{\mathrm{agr}})\operatorname{avg}\left(I_{1,r}(\bar{\mathbf{x}},\bar{y}^r),\ldots,I_{N,r}(\bar{\mathbf{x}},\bar{y}^r)\right) + \alpha^{\mathrm{agr}}\overleftrightarrow{H}\begin{pmatrix}I_{1,r}(\bar{\mathbf{x}},\bar{y}^r),\ldots,I_{N,r}(\bar{\mathbf{x}},\bar{y}^r);\\ p^{\mathrm{agr}},1-\nu\end{pmatrix}\right). \tag{75}$$

Observe that system (73)–(75) is

1) soft Mamdani-type NFIS for $\nu=0$;
2) soft logical-type NFIS for $\nu=1$;
3) soft like Mamdani-type NFIS for $0<\nu<0.5$;
4) soft like logical-type NFIS for $0.5<\nu<1$;
5) undetermined for $\nu=0.5$.

C. Weighted Soft NFIS: OR-Type

We insert weights to the antecedents and to the aggregation operator of the rules in system ORII:

1) $w_{i,k}^\tau\in[0,1]$, $i=1,\ldots,n$, $k=1,\ldots,N$;
2) $w_k^{\mathrm{agr}}\in[0,1]$, $k=1,\ldots,N$.

Consequently, we get the weighted soft NFIS of an OR-type

OR III

$$\tau_k(\bar{\mathbf{x}}) = \left((1-\alpha^\tau)\operatorname{avg}\left(\mu_{A_1^k}(\bar{x}_1),\ldots,\mu_{A_n^k}(\bar{x}_n)\right) + \alpha^\tau\overleftrightarrow{H}^*\begin{pmatrix}\mu_{A_1^k}(\bar{x}_1),\ldots,\mu_{A_n^k}(\bar{x}_n);\\ w_{1,k}^\tau,\ldots,w_{n,k}^\tau,p^\tau,0\end{pmatrix}\right) \tag{76}$$

$$I_{k,r}(\bar{\mathbf{x}},\bar{y}^r) = \left(\left(1-\alpha^I\right)\operatorname{avg}\left(\tilde{N}_{1-\nu}(\tau_k(\bar{\mathbf{x}})),\mu_{B^k}(\bar{y}^r)\right) + \alpha^I\overleftrightarrow{H}\begin{pmatrix}\tilde{N}_{1-\nu}(\tau_k(\bar{\mathbf{x}})),\mu_{B^k}(\bar{y}^r);\\ p^I,\nu\end{pmatrix}\right) \tag{77}$$

$$\mathrm{agr}_r(\bar{\mathbf{x}},\bar{y}^r) = \left((1-\alpha^{\mathrm{agr}})\operatorname{avg}\left(I_{1,r}(\bar{\mathbf{x}},\bar{y}^r),\ldots,I_{N,r}(\bar{\mathbf{x}},\bar{y}^r)\right) + \alpha^{\mathrm{agr}}\overleftrightarrow{H}^*\begin{pmatrix}I_{1,r}(\bar{\mathbf{x}},\bar{y}^r),\ldots,I_{N,r}(\bar{\mathbf{x}},\bar{y}^r);\\ w_1^{\mathrm{agr}},\ldots,w_N^{\mathrm{agr}},p^{\mathrm{agr}},1-\nu\end{pmatrix}\right). \tag{78}$$

In the ORIII system we use parameterized families $\overleftrightarrow{H}(\cdot)$ and parameterized families with weights $\overleftrightarrow{H}^*(\cdot)$ analogously to (19).

$$\tau_k(\bar{\mathbf{x}}) = H\left(\mu_{A_1^k}(\bar{x}_1),\ldots,\mu_{A_n^k}(\bar{x}_n);0\right) \tag{67}$$

$$I_{k,r}(\bar{\mathbf{x}},\bar{y}^r) = \begin{cases} H\left(\tilde{N}_1(\tau_k(\bar{\mathbf{x}})),\mu_{B^k}(\bar{y}^r);0\right) & \text{for Mamdani approach}\\ H\left(\tilde{N}_0(\tau_k(\bar{\mathbf{x}})),\mu_{B^k}(\bar{y}^r);1\right) & \text{for logical approach}\end{cases} \tag{68}$$

$$\mathrm{agr}_r(\bar{\mathbf{x}},\bar{y}^r) = \begin{cases} H\left(I_{1,r}(\bar{\mathbf{x}},\bar{y}^r),\ldots,I_{N,r}(\bar{\mathbf{x}},\bar{y}^r);1\right) & \text{for Mamdani approach}\\ H\left(I_{1,r}(\bar{\mathbf{x}},\bar{y}^r),\ldots,I_{N,r}(\bar{\mathbf{x}},\bar{y}^r);0\right) & \text{for logical approach}\end{cases}. \tag{69}$$

More specifically, in (76) and (78), we use the following definition:

$$\overset{\leftrightarrow *}{H}\begin{pmatrix} a_1,\ldots,a_n; \\ w_1,\ldots,w_n,p,\nu \end{pmatrix} = \overset{\leftrightarrow}{H}\begin{pmatrix} \arg_1(a_1,w_1,\nu),\ldots,\arg_n(a_n,w_n,\nu); \\ p,\nu \end{pmatrix} \quad (79)$$

where

$$\arg_i(a_i,w_i,\nu) = (1-\nu)(1-w_i(1-a_i)) + \nu w_i a_i. \quad (80)$$

VI. AND-Type NFIS

In this section, we study AND-type neuro-fuzzy inference systems. They will be presented in two alternative forms: by using T and S-norms or by using an H-function with $\nu = 0$ or $\nu = 1$.

A. Basic NFIS: AND-Type

The basic neuro-fuzzy inference systems of an AND-type employ combinations of "engineering" and fuzzy implication, see, e.g., (10) and (11). The systems are given by the formula:

AND Ia

$$\tau_k(\bar{\mathbf{x}}) = T\left\{\mu_{A_1^k}(\bar{x}_1),\ldots,\mu_{A_n^k}(\bar{x}_n)\right\} \quad (81)$$

$$I_{k,r}(\bar{\mathbf{x}},\bar{y}^r) = \Big((1-\lambda) I_{\mathrm{eng}}\big(\tau_k(\bar{\mathbf{x}}),\mu_{B^k}(\bar{y}^r)\big) + \lambda I_{\mathrm{fuzzy}}\big(\tau_k(\bar{\mathbf{x}}),\mu_{B^k}(\bar{y}^r)\big)\Big) \quad (82)$$

$$\mathrm{agr}_r(\bar{\mathbf{x}},\bar{y}^r) = \Big((1-\lambda) S\big\{I_{1,r}(\bar{\mathbf{x}},\bar{y}^r),\ldots,I_{N,r}(\bar{\mathbf{x}},\bar{y}^r)\big\} + \lambda T\big\{I_{1,r}(\bar{\mathbf{x}},\bar{y}^r),\ldots,I_{N,r}(\bar{\mathbf{x}},\bar{y}^r)\big\}\Big) \quad (83)$$

or

AND Ib

$$\tau_k(\bar{\mathbf{x}}) = H\begin{pmatrix} \mu_{A_1^k}(\bar{x}_1),\ldots,\mu_{A_n^k}(\bar{x}_n); \\ 0 \end{pmatrix} \quad (84)$$

$$I_{k,r}(\bar{\mathbf{x}},\bar{y}^r) = \left((1-\lambda) H\begin{pmatrix} \tilde{N}_1\big(\tau_k(\bar{\mathbf{x}})\big),\mu_{B^k}(\bar{y}^r); \\ 0 \end{pmatrix} + \lambda H\begin{pmatrix} \tilde{N}_0\big(\tau_k(\bar{\mathbf{x}})\big),\mu_{B^k}(\bar{y}^r); \\ 1 \end{pmatrix}\right) \quad (85)$$

$$\mathrm{agr}_r(\bar{\mathbf{x}},\bar{y}^r) = \left((1-\lambda) H\begin{pmatrix} I_{1,r}(\bar{\mathbf{x}},\bar{y}^r),\ldots,I_{N,r}(\bar{\mathbf{x}},\bar{y}^r); \\ 1 \end{pmatrix} + \lambda H\begin{pmatrix} I_{1,r}(\bar{\mathbf{x}},\bar{y}^r),\ldots,I_{N,r}(\bar{\mathbf{x}},\bar{y}^r); \\ 0 \end{pmatrix}\right). \quad (86)$$

It is easily seen that the above system is of a Mamdani-type for $\lambda = 0$ and logical-type for $\lambda = 1$.

B. Soft NFIS: AND-Type

In this section, we propose soft compromise NFIS based on soft fuzzy norms (12) and (13).

The soft compromise NFIS of an AND-type are given by

AND IIa

$$\tau_k(\bar{\mathbf{x}}) = \left((1-\alpha^\tau)\,\mathrm{avg}\big(\mu_{A_1^k}(\bar{x}_1),\ldots,\mu_{A_n^k}(\bar{x}_n)\big) + \alpha^\tau \overset{\leftrightarrow}{T}\begin{Bmatrix} \mu_{A_1^k}(\bar{x}_1),\ldots,\mu_{A_n^k}(\bar{x}_n); \\ p^\tau \end{Bmatrix}\right) \quad (87)$$

$$I_{k,r}(\bar{\mathbf{x}},\bar{y}^r) = \left((1-\alpha^I)\,\mathrm{avg}\big(\tilde{N}_{1-\lambda}(\tau_k(\bar{\mathbf{x}})),\mu_{B^k}(\bar{y}^r)\big) + \alpha^I\left((1-\lambda)\overset{\leftrightarrow}{I}_{\mathrm{eng}}\begin{pmatrix} \tau_k(\bar{\mathbf{x}}),\mu_{B^k}(\bar{y}^r); \\ p^I \end{pmatrix} + \lambda \overset{\leftrightarrow}{I}_{\mathrm{fuzzy}}\begin{pmatrix} \tau_k(\bar{\mathbf{x}}),\mu_{B^k}(\bar{y}^r); \\ p^I \end{pmatrix}\right)\right) \quad (88)$$

$$\mathrm{agr}_r(\bar{\mathbf{x}},\bar{y}^r) = \left((1-\alpha^{\mathrm{agr}})\,\mathrm{avg}\big(I_{1,r}(\bar{\mathbf{x}},\bar{y}^r),\ldots,I_{N,r}(\bar{\mathbf{x}},\bar{y}^r)\big) + \alpha^{\mathrm{agr}}\left((1-\lambda)\overset{\leftrightarrow}{S}\begin{Bmatrix} I_{1,r}(\bar{\mathbf{x}},\bar{y}^r),\ldots,I_{N,r}(\bar{\mathbf{x}},\bar{y}^r); \\ p^{\mathrm{agr}} \end{Bmatrix} + \lambda \overset{\leftrightarrow}{T}\begin{Bmatrix} I_{1,r}(\bar{\mathbf{x}},\bar{y}^r),\ldots,I_{N,r}(\bar{\mathbf{x}},\bar{y}^r); \\ p^{\mathrm{agr}} \end{Bmatrix}\right)\right). \quad (89)$$

Formulas (87)–(89) describe the soft compromise NFIS – AND-type in terms of parameterized families $\overset{\leftrightarrow}{T}$ and $\overset{\leftrightarrow}{S}$. Alternatively, this system can be presented by making use of the H-function definition

AND IIb

$$\tau_k(\bar{\mathbf{x}}) = \left((1-\alpha^\tau)\,\mathrm{avg}\big(\mu_{A_1^k}(\bar{x}_1),\ldots,\mu_{A_n^k}(\bar{x}_n)\big) + \alpha^\tau \overset{\leftrightarrow}{H}\begin{pmatrix} \mu_{A_1^k}(\bar{x}_1),\ldots,\mu_{A_n^k}(\bar{x}_n); \\ p^\tau, 0 \end{pmatrix}\right) \quad (90)$$

$$I_{k,r}(\bar{\mathbf{x}},\bar{y}^r) = \left((1-\alpha^I)\,\mathrm{avg}\big(\tilde{N}_{1-\lambda}(\tau_k(\bar{\mathbf{x}})),\mu_{B^k}(\bar{y}^r)\big) + \alpha^I\left((1-\lambda)\overset{\leftrightarrow}{H}\begin{pmatrix} \tilde{N}_1(\tau_k(\bar{\mathbf{x}})),\mu_{B^k}(\bar{y}^r); \\ p^I, 0 \end{pmatrix} + \lambda \overset{\leftrightarrow}{H}\begin{pmatrix} \tilde{N}_0\{\tau_k(\bar{\mathbf{x}})),\mu_{B^k}(\bar{y}^r); \\ p^I, 1 \end{pmatrix}\right)\right) \quad (91)$$

$$\begin{aligned}&\operatorname{agr}_r\left(\bar{\mathbf{x}}, \bar{y}^r;\right)\\ &=\Bigg((1-\alpha^{\mathrm{agr}}) \operatorname{avg}\left(I_{1,r}\left(\bar{\mathbf{x}}, \bar{y}^r\right), \ldots, I_{N,r}\left(\bar{\mathbf{x}}, \bar{y}^r\right)\right)\\ &\quad+\alpha^{\mathrm{agr}}\Bigg((1-\lambda) \overset{\leftrightarrow}{H}\left(\begin{array}{c} I_{1,r}\left(\bar{\mathbf{x}}, \bar{y}^r\right), \ldots, I_{N,r}\left(\bar{\mathbf{x}}, \bar{y}^r\right); \\ p^{\mathrm{agr}}, 1 \end{array}\right)\\ &\quad+\lambda \overset{\leftrightarrow}{H}\left\{\begin{array}{c} I_{1,r}\left(\bar{\mathbf{x}}, \bar{y}^r;\right), \ldots, I_{N,r}\left(\bar{\mathbf{x}}, \bar{y}^r\right); \\ p^{\mathrm{agr}}, 0 \end{array}\right\}\Bigg)\Bigg). \end{aligned} \tag{92}$$

C. Weighted Soft NFIS: AND-Type

Introducing weights to soft NFIS we get the weighted soft NFIS of an AND-type

AND IIIa

$$\begin{aligned}&\tau_k\left(\bar{\mathbf{x}}\right)\\ &=\Bigg((1-\alpha^\tau) \operatorname{avg}\left(\mu_{A_1^k}\left(\bar{x}_1\right), \ldots, \mu_{A_n^k}\left(\bar{x}_n\right)\right)\\ &\quad+\alpha^\tau \overset{\leftrightarrow *}{T}\left\{\begin{array}{c} \mu_{A_1^k}\left(\bar{x}_1\right), \ldots, \mu_{A_n^k}\left(\bar{x}_n\right); \\ w_{1,k}^\tau, \ldots, w_{n,k}^\tau, p^\tau \end{array}\right\}\Bigg) \end{aligned} \tag{93}$$

$$\begin{aligned}&I_{k,r}\left(\bar{\mathbf{x}}, \bar{y}^r\right)\\ &=\Bigg((1-\alpha^I) \operatorname{avg}\left(\tilde{N}_{1-\lambda}\left\{\tau_k\left(\bar{\mathbf{x}}\right)\right\}, \mu_{B^k}\left(\bar{y}^r\right)\right)\\ &\quad+\alpha^I\Bigg((1-\lambda) \overset{\leftrightarrow}{I}_{\mathrm{eng}}\left(\begin{array}{c} \tau_k\left(\bar{\mathbf{x}}\right), \mu_{B^k}\left(\bar{y}^r\right); \\ p^I \end{array}\right)\\ &\quad+\lambda \overset{\leftrightarrow}{I}_{\mathrm{fuzzy}}\left(\begin{array}{c} \tau_k\left(\bar{\mathbf{x}}\right), \mu_{B^k}\left(\bar{y}^r\right); \\ p^I \end{array}\right)\Bigg)\Bigg) \end{aligned} \tag{94}$$

$$\begin{aligned}&\operatorname{agr}_r\left(\bar{\mathbf{x}}, \bar{\mathbf{y}}^{\mathbf{r}}\right)\\ &=\Bigg((1-\alpha^{\mathrm{agr}}) \operatorname{avg}\left(I_{1,r}\left(\bar{\mathbf{x}}, \bar{\mathbf{y}}^{\mathbf{r}}\right), \ldots, I_{N,r}\left(\bar{\mathbf{x}}, \bar{\mathbf{y}}^{\mathbf{r}}\right)\right)\\ &\quad+\alpha^{\mathrm{agr}}\Bigg((1-\lambda) \overset{\leftrightarrow *}{S}\left(\begin{array}{c} I_{1,r}\left(\bar{\mathbf{x}}, \bar{\mathbf{y}}^{\mathbf{r}}\right), \ldots, I_{N,r}\left(\bar{\mathbf{x}}, \bar{\mathbf{y}}^{\mathbf{r}}\right); \\ w_1^{\mathrm{agr}}, \ldots, w_N^{\mathrm{agr}}, p^{\mathrm{agr}} \end{array}\right)\\ &\quad+\lambda \overset{\leftrightarrow *}{T}\left(\begin{array}{c} I_{1,r}\left(\bar{\mathbf{x}}, \bar{\mathbf{y}}^{\mathbf{r}}\right), \ldots, I_{N,r}\left(\bar{\mathbf{x}}, \bar{\mathbf{y}}^{\mathbf{r}}\right); \\ w_1^{\mathrm{agr}}, \ldots, w_N^{\mathrm{agr}}, p^{\mathrm{agr}} \end{array}\right)\Bigg)\Bigg). \end{aligned} \tag{95}$$

Alternatively, it can be expressed by

AND IIIb

$$\begin{aligned}&\tau_k\left(\bar{\mathbf{x}}\right)\\ &=\Bigg((1-\alpha^\tau) \operatorname{avg}\left(\mu_{A_1^k}\left(\bar{x}_1\right), \ldots, \mu_{A_n^k}\left(\bar{x}_n\right)\right)\\ &\quad+\alpha^\tau \overset{\leftrightarrow *}{H}\left(\begin{array}{c} \mu_{A_1^k}\left(\bar{x}_1\right), \ldots, \mu_{A_n^k}\left(\bar{x}_n\right); \\ w_{1,k}^\tau, \ldots, w_{n,k}^\tau, p^\tau, 0 \end{array}\right)\Bigg) \end{aligned} \tag{96}$$

$$\begin{aligned}&I_{k,r}\left(\bar{\mathbf{x}}, \bar{y}^r\right)\\ &=\Bigg((1-\alpha^I) \operatorname{avg}\left(\tilde{N}_{1-\lambda}\left(\tau_k\left(\bar{\mathbf{x}}\right)\right), \mu_{B^k}\left(\bar{y}^r\right)\right)\\ &\quad+\alpha^I\Bigg((1-\lambda) \overset{\leftrightarrow}{H}\left(\begin{array}{c} \tilde{N}_1\left(\tau_k\left(\bar{\mathbf{x}}\right)\right), \mu_{B^k}\left(\bar{y}^r\right); \\ p^I, 0 \end{array}\right)\\ &\quad+\lambda \overset{\leftrightarrow}{H}\left(\begin{array}{c} \tilde{N}_1\left(\tau_k\left(\bar{\mathbf{x}}\right)\right), \mu_{B^k}\left(\bar{y}^r\right); \\ p^I, 1 \end{array}\right)\Bigg)\Bigg) \end{aligned} \tag{97}$$

TABLE V
EXPERIMENTAL RESULTS

Experiment number	Name of flexibility parameter	Initial values	Final values after learning	Mistakes [%] (learning sequence)	Mistakes [%] (testing sequence)
i	ν	0.5	1.0000	6.66%	7.81%
ii	λ	0.5	1.0000	7.33%	7.81%
iii	ν	0	-	10.00%	7.81%
iv	ν	0.5	1.0000	3.33%	6.25%
	p^τ	10	9.9953		
	p^I	10	9.9998		
	p^{agr}	10	9.9999		
	α^τ	1	0.9576		
	α^I	1	0.9931		
	α^{agr}	1	0.8482		
v	ν	0.5	1.0000	2.00%	6.25%
	p^τ	10	9.6501		
	p^I	10	9.9997		
	p^{agr}	10	9.9836		
	α^τ	1	0.9213		
	α^I	1	0.9939		
	α^{agr}	1	0.8456		
	$\mathbf{w}^\tau$	1	Fig. 6		
	$\mathbf{w}^{agr}$	1	Fig. 6		

TABLE VI
COMPARISON TABLE

No	Name	T-norm S-norm
1	Zadeh	$T_M\{a_1,a_2\}=\min\{a_1,a_2\}$ $S_M\{a_1,a_2\}=\max\{a_1,a_2\}$
2	Product	$T_P\{a_1,a_2\}=a_1 a_2$ $S_P\{a_1,a_2\}=a_1+a_2-a_1 a_2$
3	Łukasiewicz	$T_L\{a_1,a_2\}=\max\{a_1+a_2-1,0\}$ $S_L\{a_1,a_2\}=\min\{a_1+a_2,1\}$
4	Drastic	$T_D\{a_1,a_2\}=\begin{cases}0 & \text{if } S_M\{a_1,a_2\}<1\\ T_M\{a_1,a_2\} & \text{if } S_M\{a_1,a_2\}=1\end{cases}$ $S_D\{a_1,a_2\}=\begin{cases}1 & \text{if } T_M\{a_1,a_2\}>0\\ S_M\{a_1,a_2\} & \text{if } T_M\{a_1,a_2\}=0\end{cases}$

$$\begin{aligned}&\operatorname{agr}_r\left(\bar{\mathbf{x}}, \bar{y}^r\right)\\ &=\Bigg((1-\alpha^{\mathrm{agr}}) \operatorname{avg}\left(I_{1,r}\left(\bar{\mathbf{x}}, \bar{\mathbf{y}}^{\mathbf{r}}\right), \ldots, I_{N,r}\left(\bar{\mathbf{x}}, \bar{\mathbf{y}}^{\mathbf{r}}\right)\right)\\ &\quad+\alpha^{\mathrm{agr}}\Bigg((1-\lambda) \overset{\leftrightarrow *}{H}\left\{\begin{array}{c} I_{1,r}\left(\bar{\mathbf{x}}, \bar{y}^r\right), \ldots, I_{N,r}\left(\bar{\mathbf{x}}, \bar{y}^r\right); \\ w_1^{\mathrm{agr}}, \ldots, w_N^{\mathrm{agr}}, p^{\mathrm{agr}}, 1 \end{array}\right\}\\ &\quad+\lambda \overset{\leftrightarrow *}{H}\left(\begin{array}{c} I_{1,r}\left(\bar{\mathbf{x}}, \bar{\mathbf{y}}^{\mathbf{r}}\right), \ldots, I_{N,r}\left(\bar{\mathbf{x}}, \bar{\mathbf{y}}^{\mathbf{r}}\right); \\ w_1^{\mathrm{agr}}, \ldots, w_N^{\mathrm{agr}}, p^{\mathrm{agr}}, 0 \end{array}\right)\Bigg)\Bigg). \end{aligned} \tag{98}$$

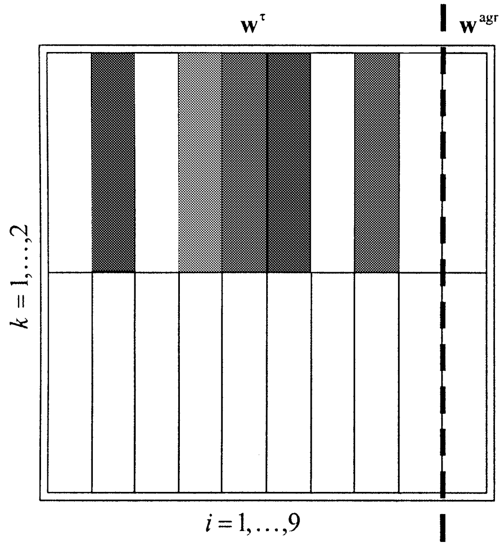

Fig. 6. Weights representation in the Glass Identification problem $w^{\tau}_{i,k} \in [0,1]$, $w^{agr}_{k} \in [0,1]$, $i = 1,\ldots,9$, $k = 1,\ldots,2$ (dark areas correspond to low values and vice versa).

TABLE VII
EXPERIMENTAL RESULTS

Experiment number	Name of flexibility parameter	Initial values	Final values after learning	Mistakes [%] (learning sequence)	Mistakes [%] (testing sequence)
i	ν	0.5	1.0000	4.87%	9.52%
ii	λ	0.5	1.0000	4.87%	9.52%
iii	ν	0	-	35.77%	36.19%
iv	ν p^{τ} p^{I} p^{agr} α^{τ} α^{I} α^{agr}	0.5 10 10 10 1 1 1	1.0000 9.9185 9.9999 9.9999 0.0000 0.9620 0.9922	2.43%	9.52%
v	ν p^{τ} p^{I} p^{agr} α^{τ} α^{I} α^{agr} $\mathbf{w}^{\tau}$ $\mathbf{w}^{agr}$	0.5 10 10 10 1 1 1 **1** **1**	1.0000 9.9052 9.9997 9.9954 0.0000 0.9598 0.9883 **≈1** **≈1**	2.43%	7.61%

VII. Performance Evaluation

We have conducted very extensive simulations in order to test our new structures, find an appropriate system type and compare the results with other authors. The results are summarized and discussed in Section VIII. We present 11 simulations by making

TABLE VIII
COMPARISON TABLE

Method	Testing Acc.
González and Pérez (CN2) [15]	83.96
González and Pérez (C4.5) [15]	86.50
González and Pérez (LVQ) [15]	87.90
González and Pérez (SLAVE[2]) [15]	82.40
González and Pérez (SLAVE[2] with OR and AND) [15]	85.40
González and Pérez (SLAVE[2] with OR, AND and Generalization) [15]	90.90
González and Pérez (SLAVE[2] with all operators) [15]	91.80
our result	**92.38**

TABLE IX
EXPERIMENTAL RESULTS

Experiment number	Name of flexibility parameter	Initial values	Final values after learning	Mistakes [%] (learning sequence)	Mistakes [%] (testing sequence)
i	ν	0.5	1.0000	1.90%	6.66%
ii	λ	0.5	1.0000	1.90%	6.66%
iii	ν	0	-	4.76%	11.11%
iv	ν p^{τ} p^{I} p^{agr} α^{τ} α^{I} α^{agr}	0.5 10 10 10 1 1 1	1.0000 10.0041 9.9879 9.9834 0.9003 0.9914 0.9517	1.90%	4.44%
v	ν p^{τ} p^{I} p^{agr} α^{τ} α^{I} α^{agr} $\mathbf{w}^{\tau}$ $\mathbf{w}^{agr}$	0.5 10 10 10 1 1 1 **1** **1**	1.0000 10.0033 9.9950 9.9836 0.9080 0.9974 0.9582 Fig. 7 Fig. 7	0.00%	2.22%

use of commonly known benchmarks. Each of the simulations is designed in the same fashion:

1) In the first experiment, based on the input-output data, we learn the parameters of the membership functions and a system type $\nu \in [0,1]$ of the OR I neuro-fuzzy inference system. It will be seen that the optimal values of ν, determined by a gradient procedure, are either zero or one.
2) In the second experiment, we learn the parameters of the membership functions and a system type $\lambda \in [0,1]$ of the AND I neuro-fuzzy inference system. It will be seen that optimal values of λ, determined by a gradient procedure, are either zero or one.

TABLE X
COMPARISON TABLE

Method	Testing Acc.
González and Pérez (CN2) [15]	94.16
González and Pérez (C4.5) [15]	92.70
González and Pérez (LVQ) [15]	95.70
González and Pérez (SLAVE[2]) [15]	95.70
González and Pérez (SLAVE[2] with OR and AND) [15]	95.70
González and Pérez (SLAVE[2] with OR, AND and Generalization) [15]	95.70
González and Pérez (SLAVE[2] with all operators) [15]	95.70
our result	**97.78**

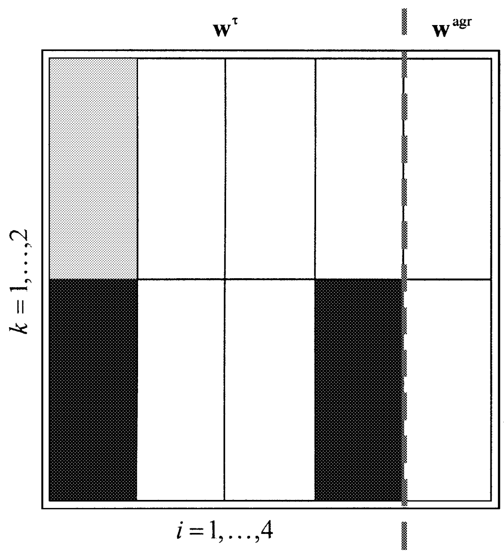

Fig. 7. Weights representation in the Iris problem $w_{i,k}^{\tau} \in [0,1]$, $w_k^{\mathrm{agr}} \in [0,1]$, $i = 1,\ldots,4$, $k = 1,\ldots,2$ (dark areas correspond to low values and vice versa).

3) In the third experiment, we learn the parameters of the membership functions of the OR I/AND I neuro-fuzzy inference systems choosing values of ν and λ as opposite of those obtained in 1) and 2). Obviously, we expect a worse performance of both systems. Note that OR-type are equivalent to AND-type systems if $\nu = \lambda = 0$ or $\nu = \lambda = 1$.
4) In the fourth experiment, we learn the parameters of the membership functions, system type $\nu \in [0,1]$ of the OR II neuro-fuzzy inference system and soft parameters $\alpha^{\tau} \in [0,1]$, $\alpha^{I} \in [0,1]$, $\alpha^{\mathrm{agr}} \in [0,1]$. Moreover, we learn parameters $p^{\tau} \in [0,\infty)$, $p^{I} \in [0,\infty)$, $p^{\mathrm{agr}} \in [0,\infty)$ of the Dombi norm used for the connection of antecedents, implication and aggregation of rules, respectively.
5) In the fifth experiment, we learn the same parameters as in the fourth experiment and, moreover, the weights $w_{i,k}^{\tau} \in [0,1]$, $i = 1,\ldots,n$, $k = 1,\ldots,N$, in the antecedents of rules and weights $w_k^{\mathrm{agr}} \in [0,1]$, $k = 1,\ldots,N$, of the aggregation operator of the rules. In all diagrams (weights representation) we separate $w_{i,k}^{\tau} \in [0,1]$, $i = 1,\ldots,n$, $k = 1,\ldots,N$, from $w_k^{\mathrm{agr}} \in [0,1]$, $k = 1,\ldots,N$, by a vertical dashed line.

TABLE XI
EXPERIMENTAL RESULTS

Experiment number	Name of flexibility parameter	Initial values	Final values after learning	RMSE (learning sequence)
i	ν	0.5	0.0000	0.1021
ii	λ	0.5	0.0000	0.0915
iii	ν	1	-	0.1265
iv	ν	0.5	0.0000	0.0857
	p^{τ}	10	9.8804	
	p^{I}	10	9.9752	
	p^{agr}	10	9.5912	
	α^{τ}	1	0.0331	
	α^{I}	1	0.9741	
	α^{agr}	1	0.7611	
v	ν	0.5	0.0000	0.0739
	p^{τ}	10	9.9513	
	p^{I}	10	9.9699	
	p^{agr}	10	9.4197	
	α^{τ}	1	0.1250	
	α^{I}	1	0.9612	
	α^{agr}	1	0.8456	
	$\mathbf{w}^{\tau}$	**1**	Fig. 8	
	$\mathbf{w}^{\mathrm{agr}}$	**1**	Fig. 8	

TABLE XII
COMPARISON TABLE

Method	No of rules	Training RMSE
Kim et al (SI) [26]	3	0.0935
Sugeno and Yasukawa [52]	6	0.2810
Kim et al (SI with FMAIUP) [27]	2	0.1403
our result	**4**	**0.0739**

The parameters learned in experiments 1)-5) can be determined by standard recursive gradient procedures with the constraints listed above. In order to avoid arduous gradient calculations, we have developed [47] a universal network trainer that can tune the parameters and weights of FLEXNFIS based on their architectures. The idea of the trainer comes from the backpropagation method. It should be noted that Gaussian membership functions are used in all the experiments.

A. Glass Identification

The Glass Identification problem [56] contains 214 instances and each instance is described by nine attributes (RI: refractive index, Na: sodium, Mg: magnesium, Al: aluminum, Si: silicon, K: potassium, Ca: calcium, Ba: barium, Fe: iron). All attributes are continuous. There are two classes: window glass and non-window glass. In our experiments, all sets are divided into a learning sequence (150 sets) and testing sequence (64 sets). The study of classification of types of glass was motivated by criminological investigation. At the scene of the crime, the glass left

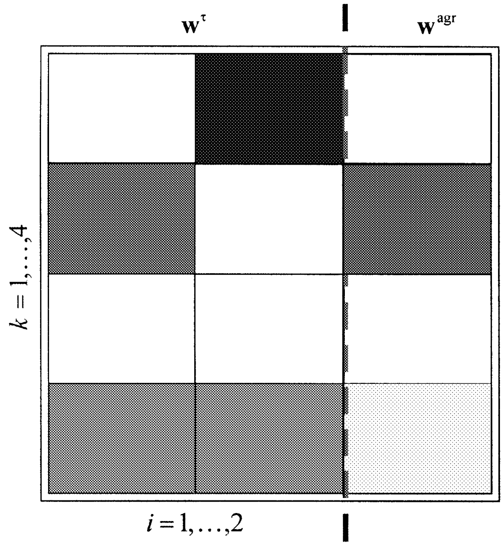

Fig. 8. Weights representation in the HANG problem $w_{i,k}^{\tau} \in [0,1]$, $w_k^{\mathrm{agr}} \in [0,1]$, $i = 1,\ldots,2$, $k = 1,\ldots,4$ (dark areas correspond to low values and vice versa).

TABLE XIII
EXPERIMENTAL RESULTS

Experiment number	Name of flexibility parameter	Initial values	Final values after learning	RMSE (learning sequence)
i	ν	0.5	0.0000	0.5079
ii	λ	0.5	0.0000	0.4917
iii	ν	1	-	0.6420
iv	ν	0.5	0.0000	0.4232
	p^{τ}	10	9.9774	
	p^{I}	10	9.8540	
	p^{agr}	10	9.2408	
	α^{τ}	1	0.0215	
	α^{I}	1	0.9433	
	α^{agr}	1	0.9686	
v	ν	0.5	0.0000	0.2416
	p^{τ}	10	9.9581	
	p^{I}	10	9.8592	
	p^{agr}	10	9.1067	
	α^{τ}	1	0.0590	
	α^{I}	1	0.9713	
	α^{agr}	1	0.9655	
	$\mathbf{w}^{\tau}$	**1**	Fig. 9	
	$\mathbf{w}^{agr}$	**1**	Fig. 9	

can be used as evidence if it is correctly identified. The experimental results are depicted in Table V, Table VI, and Fig. 6.

B. Ionosphere

This radar data was collected by a system in Goose Bay, Labrador [56]. This system consists of a phased array of 16 high-frequency antennas with a total transmitted power in the order of 6.4 kW. The targets were free electrons in the ionosphere. The database is composed of 34 continuous attributes plus the class variable, using 351 examples. In our experiments, all sets are divided into a learning sequence (246 sets) and testing sequence (105 sets), $n = 33$, $N = 2$ and after learning all weights are equal to one. The experimental results are depicted in Tables VII and VIII.

TABLE XIV
COMPARISON TABLE

Method	No of inputs/ rules	Training RMSE
Tong [54]	2/19	0.6848
Pedrycz [40]	2/81	0.5656
Xu and Lu [63]	2/25	0.5727
Box and Jenkins [3]	6/-	0.4494
Sugeno and Yasukawa [51]	3/6	0.4358
Wang and Langari [59]	6/2	0.2569
Sugeno and Tanaka [51]	6/2	0.2607
Lin and Cunningham [33]	5/4	0.2664
Kim et al [26]	6/2	0.2345
Kim et al [27]	6/2	0.2190
Delgado et al [8]	2/4	0.4100
Yoshinari [70]	2/6	0.5460
our result	**6/4**	**0.2416**

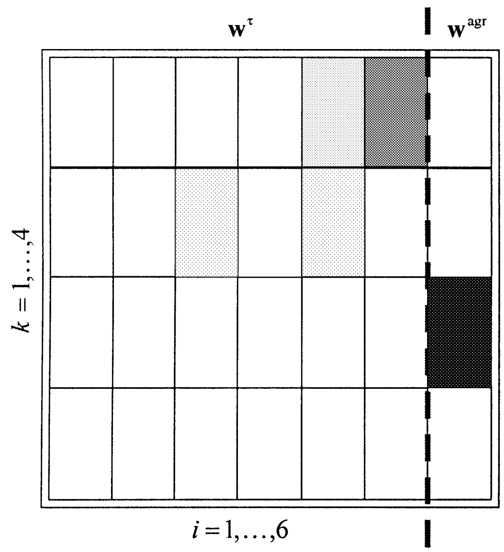

Fig. 9. Weights representation in the Modeling of Box and Jenkins Gas Furnace problem $w_{i,k}^{\tau} \in [0,1]$, $w_k^{\mathrm{agr}} \in [0,1]$, $i = 1,\ldots,6$, $k = 1,\ldots,4$ (dark areas correspond to low values and vice versa).

C. Iris

The Iris data [56] is a common benchmark in classification and pattern recognition studies. It contains 50 measurements of four features (sepal length in cm, sepal width in cm, petal length in cm, petal width in cm) from each of three species: iris setosa, iris versicolor, and iris virginica. In our experiments, all sets are divided into a learning sequence (105 sets) and testing sequence (45 sets). The experimental results are depicted in Table IX, Table X, and Fig. 7.

TABLE XV
EXPERIMENTAL RESULTS

Experiment number	Name of flexibility parameter	Initial values	Final values after learning	RMSE (learning sequence)	RMSE (testing sequence)
i	ν	0.5	0.0000	0.0375	0.0270
ii	λ	0.5	0.0000	0.0379	0.0232
iii	ν	1	-	0.0381	0.0280
iv	ν	0.5	0.0000	0.0364	0.0241
	p^{τ}	10	10.0129		
	p^{l}	10	9.9789		
	p^{agr}	10	10.0819		
	α^{τ}	1	0.5893		
	α^{l}	1	0.9996		
	α^{agr}	1	0.9998		
v	ν	0.5	0.0000	0.0328	0.0211
	p^{τ}	10	10.0492		
	p^{l}	10	9.9483		
	p^{agr}	10	10.0453		
	α^{τ}	1	0.5286		
	α^{l}	1	0.9947		
	α^{agr}	1	0.9980		
	$\mathbf{w}^{\tau}$	**1**	Fig. 10		
	$\mathbf{w}^{agr}$	**1**	Fig. 10		

TABLE XVI
COMPARISON TABLE

Method	No of rules	Training RMSE	Testing RMSE
Wang and Yen [62]	40	0.0182	0.0263
Wang and Yen [62]	28	0.0182	0.0245
Wang and Yen [61]	36	0.0053	0.0714
Wang and Yen [61]	23	0.0057	0.0436
Wang and Yen [61]	36	0.0014	0.0539
Wang and Yen [61]	24	0.0014	0.0253
Yen and Wang [68]	25	0.0152	0.0202
Yen and Wang [68]	20	0.0261	0.0155
Setnes and Roubos [49]	7	0.1265	0.0346
Setnes and Roubos [49]	7	0.0548	0.0221
Setnes and Roubos [49]	5	0.0762	0.0500
Setnes and Roubos [49]	5	0.0274	0.0187
Setnes and Roubos [49]	4	0.0346	0.0217
Roubos and Setnes [43]	5	0.0700	0.0539
Roubos and Setnes [43]	5	0.0374	0.0243
Roubos and Setnes [43]	5	0.0288	0.0187
our result	**5**	**0.0328**	**0.0211**

D. Modeling of a Static Nonlinear Function (HANG)

In this example, a double-input and single output static function is chosen to be a target system for the new fuzzy modeling strategy. This function is represented as

$$z = \left(1 + x^{-2} + y^{-1.5}\right)^2. \qquad (99)$$

From the evenly distributed grid point of the input range $x, y \in [1, 5]$ of the preceding equation, 50 training data pairs were obtained. The experimental results are depicted in Table XI, Table XII, and Fig. 8.

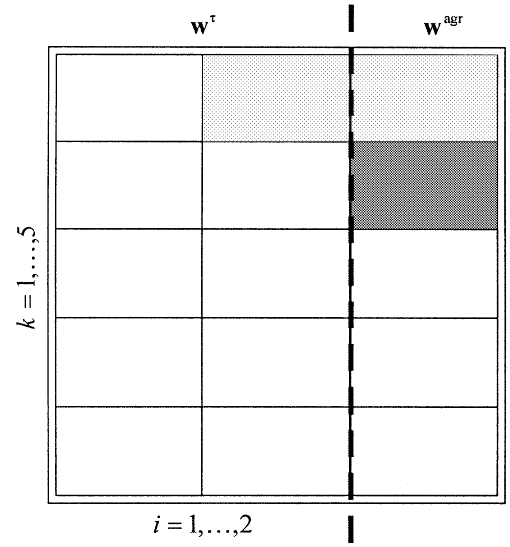

Fig. 10. Weights representation in the nonlinear dynamic plant problem $w^{\tau}_{i,k} \in [0,1]$, $w^{agr}_{k} \in [0,1]$, $i = 1,\dots,2$, $k = 1,\dots,5$ (dark areas correspond to low values and vice versa).

TABLE XVII
EXPERIMENTAL RESULTS

Experiment number	Name of flexibility parameter	Initial values	Final values after learning	Mistakes [%] (learning sequence)	Mistakes [%] (testing sequence)
i	ν	0.5	1.0000	21.88%	22.92%
ii	λ	0.5	1.0000	23.09%	23.96%
iii	ν	0	-	28.47%	29.17%
iv	ν	0.5	1.0000	21.70%	22.40%
	p^{τ}	10	10.2563		
	p^{l}	10	9.9973		
	p^{agr}	10	9.4531		
	α^{τ}	1	0.9349		
	α^{l}	1	0.9572		
	α^{agr}	1	0.0591		
v	ν	0.5	1.0000	20.83%	21.35%
	p^{τ}	10	10.2070		
	p^{l}	10	9.9890		
	p^{agr}	10	9.5711		
	α^{τ}	1	0.9259		
	α^{l}	1	0.9463		
	α^{agr}	1	0.0211		
	$\mathbf{w}^{\tau}$	**1**	Fig. 11		
	$\mathbf{w}^{agr}$	**1**	Fig. 11		

E. Modeling of Box and Jenkins Gas Furnace

The Box and Jenkins Gas Furnace data consists of 296 measurements of a gas furnace system: the input measurement $u(k)$

TABLE XIX
EXPERIMENTAL RESULTS

Experiment number	Name of flexibility parameter	Initial values	Final values after learning	RMSE (learning sequence)	RMSE (testing sequence)
i	ν	0.5	0.0000	0.0169	0.0491
ii	λ	0.5	0.0000	0.0173	0.0504
iii	ν	1	-	0.0175	0.0536
iv	ν	0.5	0.0000	0.0167	0.0482
	p^{τ}	10	9.9079		
	p^{I}	10	9.9955		
	p^{agr}	10	10.0081		
	α^{τ}	1	0.5980		
	α^{I}	1	0.9846		
	α^{agr}	1	0.9445		
v	ν	0.5	0.0000	0.0155	0.0427
	p^{τ}	10	9.9120		
	p^{I}	10	9.9968		
	p^{agr}	10	10.0259		
	α^{τ}	1	0.6782		
	α^{I}	1	0.9888		
	α^{agr}	1	0.9401		
	$\mathbf{w}^{\tau}$	**1**	Fig. 12		
	$\mathbf{w}^{agr}$	**1**	Fig. 12		

TABLE XX
COMPARISON TABLE

Method	Training RMSE	Testing RMSE
Ishibuchi et al (Heuristic. K=4) [17]	0.0800	0.0624
Ishibuchi et al (Heuristic. K=5) [17]	0.0414	0.0764
Ishibuchi et al (Learning. K=3) [17]	0.0339	0.0518
Ishibuchi et al (Learning. K=5) [17]	0.0205	0.1887
Ishibuchi et al (Hybrid. K=2) [17]	0.0434	0.0563
Ishibuchi et al (Hybrid. K=5) [17]	0.0154	0.0812
Ishibuchi et al (NN 5 20 1) [17]	0.0330	0.0547
Ishibuchi et al (NN 5 30 1) [17]	0.0333	0.0542
Nozaki et al (Heuristic. α=50. K=2) [39]	0.0752	0.0872
Nozaki et al (Heuristic. α=10. K=3) [39]	0.0542	0.0698
Nozaki et al (Heuristic. α=5. K=4) [39]	0.0429	0.0600
Nozaki et al (Heuristic. α=10. K=4) [39]	0.0423	0.0606
Nozaki et al (Heuristic (α=5. K=5) [39]	0.0372	0.0762
Nozaki et al (Learning. e=200. K=2) [39]	0.0420	0.0560
Nozaki et al (Learning. e=500. K=2) [39]	0.0400	0.0611
Nozaki et al (Learning. e=100. K=3) [39]	0.0339	0.0516
Nozaki et al (Learning. e=500. K=3) [39]	0.0283	0.0658
Nozaki et al (Learning. e=200. K=4) [39]	0.0248	0.1183
Nozaki et al (Learning. e=500. K=4) [39]	0.0179	0.1200
Nozaki et al (Learning. e=200. K=5) [39]	0.0163	0.1883
Nozaki et al (Learning. e=400. K=5) [39]	0.0126	0.1887
our result	**0.0155**	**0.0427**

is the gas flow rate into the furnace and the output measurement is the CO_2 concentration in outlet gas. The sampling interval is 9 s. The experimental results are depicted in Table XIII, Table XIV, and Fig. 9.

TABLE XVIII
COMPARISON TABLE

Method	Testing Acc.
Smith et al [56]	76.0
Ster and Dobnikar (Logdisc) [50]	77.7
Ster and Dobnikar (IncNet) [50]	77.6
Ster and Dobnikar (DIPOL92) [50]	77.6
Ster and Dobnikar (LDA) [50]	77.5
Ster and Dobnikar (SMART) [50]	76.8
Ster and Dobnikar (ASI) [50]	76.6
Ster and Dobnikar (FDA) [50]	76.5
Ster and Dobnikar (BP) [50]	76.4
Ster and Dobnikar (LVQ) [50]	75.8
Ster and Dobnikar (RBF) [50]	75.7
Ster and Dobnikar (LFC) [50]	75.8
Ster and Dobnikar (NB) [50]	75.3
Ster and Dobnikar (SNB) [50]	75.4
Ster and Dobnikar (DB-CART) [50]	74.4
Ster and Dobnikar (ASR) [50]	74.3
Ster and Dobnikar (CART: 11nodes) [50]	73.7
Ster and Dobnikar (C4.5) [50]	73.0
Ster and Dobnikar (CART) [50]	72.8
Ster and Dobnikar (Kohonen SOM) [50]	72.2
Ster and Dobnikar (kNN) [50]	71.9
our result	**78.6**

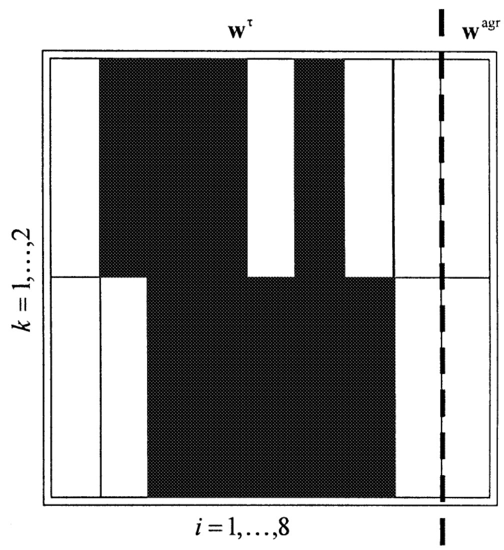

Fig. 11. Weights representation in the Pima Indians Diabetes problem $w^{\tau}_{i,k} \in [0,1]$, $w^{agr}_{k} \in [0,1]$, $i = 1,\ldots,8$, $k = 1,\ldots,2$ (dark areas correspond to low values and vice versa).

F. Nonlinear Dynamic Plant

We consider the second-order nonlinear plant studied by Wang and Yen [61]

$$y(k) = g(y(k-1), y(k-2)) + u(k) \tag{100}$$

with

$$g(y(k-1), y(k-2)) = \frac{y(k-1)y(k-2)(y(k-1)-0.5)}{1+y^2(k-1)+y^2(k-2)}. \tag{101}$$

The goal is to approximate the nonlinear component $g(y(k-1), y(k-2))$ of the plant with a fuzzy model. In [61], 400 simulated data were generated from the plant model (101). Starting from the equilibrium state (0, 0), 200 samples of identification data were obtained with a random

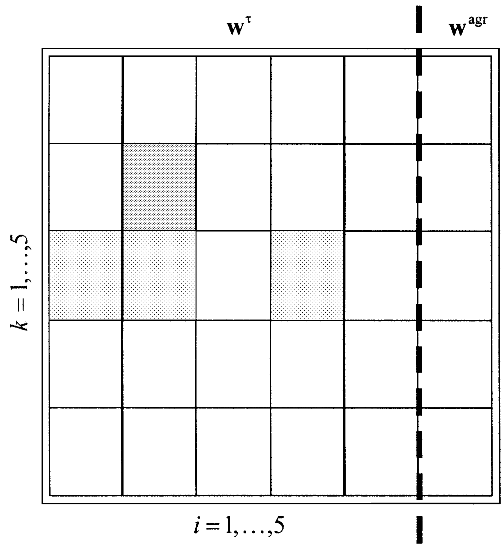

Fig. 12. Weights representation in the Rice Taste problem $w_{i,k}^{\tau} \in [0, 1]$, $w_k^{\text{agr}} \in [0, 1]$, $i = 1, \ldots, 5$, $k = 1, \ldots, 5$ (dark areas correspond to low values and vice versa).

TABLE XXI
EXPERIMENTAL RESULTS

Experiment number	Name of flexibility parameter	Initial values	Final values after learning	Mistakes [%] (learning sequence)	Mistakes [%] (testing sequence)
i	ν	0.5	0.0000	8.87%	13.89%
ii	λ	0.5	0.0000	8.87%	13.89%
iii	ν	1	-	11.29%	19.68%
iv	ν	0.5	0.0000	6.45%	8.33%
	p^{τ}	10	9.9894		
	p^{I}	10	9.9997		
	p^{agr}	10	9.8545		
	α^{τ}	1	0.9876		
	α^{I}	1	0.9799		
	α^{agr}	1	1.0000		
v	ν	0.5	0.0000	0.00%	0.00%
	p^{τ}	10	9.9950		
	p^{I}	10	9.9990		
	p^{agr}	10	8.2958		
	α^{τ}	1	0.9805		
	α^{I}	1	0.9852		
	α^{agr}	1	1.0000		
	$\mathbf{w}^{\tau}$	**1**	Fig. 13		
	$\mathbf{w}^{agr}$	**1**	Fig. 13		

input signal $u(k)$ uniformly distributed in [−1.5, 1.5], followed by 200 samples of evaluation data obtained using a sinusoidal input signal $u(k) = \sin(2\pi k/25)$. The experimental results are depicted in Table XV, Table XVI, and Fig. 10.

TABLE XXII
COMPARISON TABLE

Method	Training Acc.	Testing Acc.
Dong and Kothari (IG) [9]	100.00	86.75
Dong and Kothari (IG+LA) [9]	100.00	100.00
Dong and Kothari (GR) [9]	100.00	84.72
Dong and Kothari (GR+LA) [9]	100.00	100.00
our result	**100.00**	**100.00**

Monk1

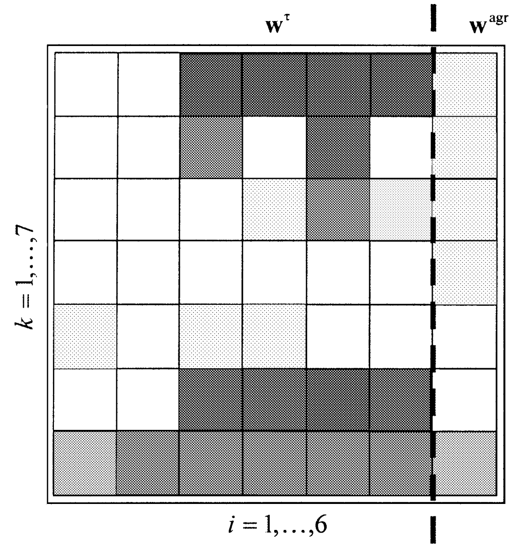

Fig. 13. Weights representation in the Monk 1 problem $w_{i,k}^{\tau} \in [0, 1]$, $w_k^{\text{agr}} \in [0, 1]$, $i = 1, \ldots, 6$, $k = 1, \ldots, 7$ (dark areas correspond to low values and vice versa).

G. Pima Indians Diabetes

The Pima Indians Diabetes data [56] contains two classes, eight attributes (number of times pregnant, plasma glucose concentration in an oral glucose tolerance test, diastolic blood pressure (mm Hg), triceps skin fold thickness (mm), 2-h serum insulin (mu U/ml), body mass index [weight in kg/(height in m)^2), diabetes pedigree function, age (years)]. We consider 768 instances, 500 (65.1%) healthy and 268 (34.9%) diabetes cases. All patients were females at least 21 years old, of Pima Indian heritage. In our experiments, all sets are divided into a learning sequence (576 sets) and testing sequence (192 sets). The experimental results are depicted in Table XVII, Table XVIII, and Fig. 11.

H. Rice Taste

The Rice Taste data contains 105 instances and each instance is described by five attributes: flavor, appearance, taste, stickiness, toughness, and overall evaluation. In simulations the input-output pairs of the rice taste data were normalized in the interval [0, 1]. The experimental results are depicted in Table XIX, Table XX, and Fig. 12.

I. The Three Monk's Problems

The Three Monk's Problems [56] are artificial, small problems designed to test machine learning algorithms. Each of the

TABLE XXIII
EXPERIMENTAL RESULTS

Experiment number	Name of flexibility parameter	Initial values	Final values after learning	Mistakes [%] (learning sequence)	Mistakes [%] (testing sequence)
i	ν	0.5	0.0000	22.49%	27.08%
ii	λ	0.5	0.0000	22.49%	27.08%
iii	ν	1	-	30.18%	35.65%
	ν	0.5	0.0000		
	p^{τ}	10	9.9812		
	p^{I}	10	10.0124		
iv	p^{agr}	10	8.8633	17.16%	31.25%
	α^{τ}	1	0.0000		
	α^{I}	1	0.9887		
	α^{agr}	1	0.9925		
	ν	0.5	0.0000		
	p^{τ}	10	9.9629		
	p^{I}	10	10.514		
	p^{agr}	10	8.9431		
v	α^{τ}	1	0.0016	4.14%	11.81%
	α^{I}	1	0.9958		
	α^{agr}	1	0.9755		
	$\mathbf{w}^{\tau}$	**1**	Fig. 14		
	$\mathbf{w}^{agr}$	**1**	Fig. 14		

TABLE XXIV
COMPARISON TABLE

Method	Training Acc.	Testing Acc.
Dong and Kothari (IG) [9]	80.47	63.42
Dong and Kothari (IG+LA) [9]	82.25	63.65
Dong and Kothari (GR) [9]	81.06	64.81
Dong and Kothari (GR+LA) [9]	81.66	66.51
our result	**95.86**	**88.19**

three monks problem requires determining whether an object described by six features (head shape, body shape, is smiling, holding, jacket color, has tie) is a monk or not.

There are 432 combinations of the six symbolic attributes. In the first problem (**Monk1**), 124 cases were randomly selected for the training set, in the second problem (**Monk2**) 169 cases, and in the third problem (**Monk3**) 122 cases, of which 5% were misclassifications introducing some noise in the data. The experimental results are depicted in Table XXI, Table XXII, and Fig. 13 (**Monk1**), Table XXIII, Table XXIV, and Fig. 14 (**Monk2**), Table XXV, Table XXVI, and Fig. 15 (**Monk3**).

Monk2

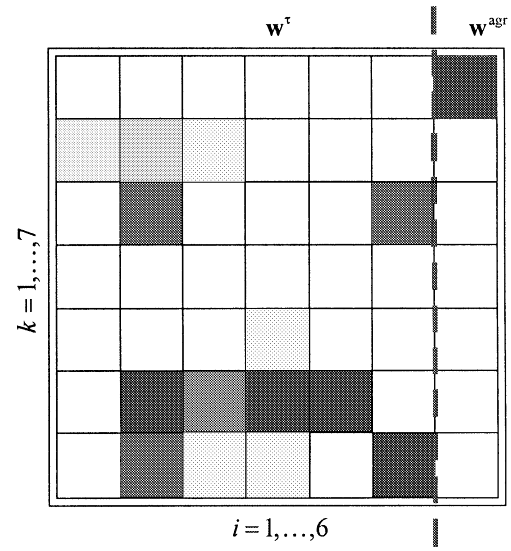

Fig. 14. Weights representation in the Monk 2 problem $w^{\tau}_{i,k} \in [0,1]$, $w^{agr}_{k} \in [0,1]$, $i = 1,\ldots,6$, $k = 1,\ldots,7$ (dark areas correspond to low values and vice versa).

TABLE XXV
EXPERIMENTAL RESULTS

Experiment number	Name of flexibility parameter	Initial values	Final values after learning	Mistakes [%] (learning sequence)	Mistakes [%] (testing sequence)
i	ν	0.5	0.0000	6.55%	3.00%
ii	λ	0.5	0.0000	6.55%	3.70%
iii	ν	1	-	4.91%	6.94%
	ν	0.5	0.0000		
	p^{τ}	10	9.9400		
	p^{I}	10	1.0000		
iv	p^{agr}	10	9.9216	6.55%	2.77%
	α^{τ}	1	1.0000		
	α^{I}	1	0.9815		
	α^{agr}	1	0.7589		
	ν	0.5	0.0000		
	p^{τ}	10	9.9156		
	p^{I}	10	1.0000		
	p^{agr}	10	9.9512		
v	α^{τ}	1	1.0000	0.00%	0.00%
	α^{I}	1	0.9342		
	α^{agr}	1	0.7612		
	$\mathbf{w}^{\tau}$	**1**	Fig. 15		
	$\mathbf{w}^{agr}$	**1**	Fig. 15		

J. Wine Recognition

The Wine data [56] contains the chemical analysis of 178 wines grown in the same region of Italy but derived from three different vineyards. The 13 continuous attributes available for classification are: alcohol, malic acid, ash, alcalinity of ash, magnesium, total phenols, flavanoids, nonflavanoid phenols, proanthocyanins, color intensity, hue, OD280/OD315 of diluted wines and proline. In our experiments all sets are divided into a learning sequence (125 sets) and testing sequence (53 sets). The experimental results are depicted in Table XXVII, Table XXIII, and Fig. 16.

TABLE XXVI
COMPARISON TABLE

Method	Training Acc.	Testing Acc.
Dong and Kothari (IG) [9]	80.47	63.42
Dong and Kothari (IG+LA) [9]	82.25	63.65
Dong and Kothari (GR) [9]	81.06	64.81
Dong and Kothari (GR+LA) [9]	81.66	66.51
our result	**95.86**	**88.19**

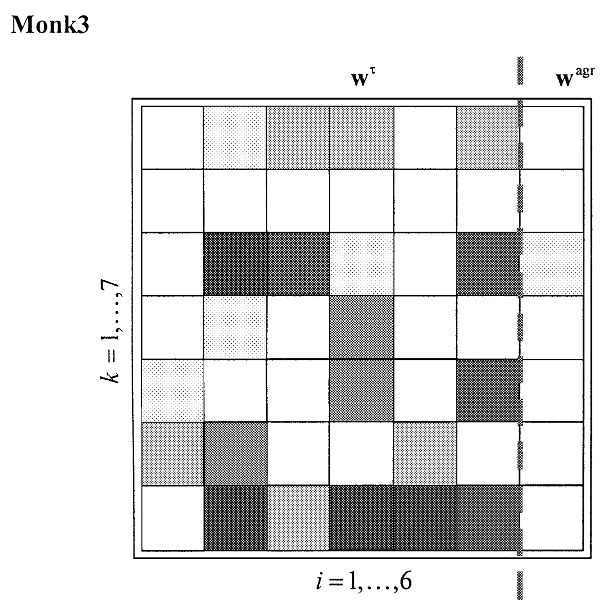

Fig. 15. Weights representation in the Monk 3 problem $w_{i,k}^{\tau} \in [0,1]$, $w_k^{\mathrm{agr}} \in [0,1]$, $i=1,\ldots,6$, $k=1,\ldots,7$ (dark areas correspond to low values and vice versa).

K. Wisconsin Breast Cancer Data

The Wisconsin Breast Cancer data [56] contains 699 instances (of which 16 instances have a single missing attribute) and each instance is described by nine attributes (clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli, mitoses). We removed those 16 instances and used the remaining 683 instances. In our experiments, all sets are divided into a learning sequence (478 sets) and testing sequence (205 sets). The experimental results are depicted in Table XXIX, Table XXX, and Fig. 17.

VIII. FINAL REMARKS

In this paper, we have presented new neuro-fuzzy structures. They are characterized as follows

1) The AND-type system is a combination, controlled by parameter $\lambda \in [0,1]$, of Mamdani-type and logical-type systems. In the process of learning only one type of system is established ($\lambda = 0$ or $\lambda = 1$).
2) The OR-type system is "more Mamdani" ($0 < \nu < 0.5$) or "more logical" ($0.5 < \nu < 1$). In the process of learning one gets $\nu = 0$ or $\nu = 1$.
3) The OR-type is equivalent to the AND-type system if $\nu = \lambda = 0$ (Mamdani-type) or $\nu = \lambda = 1$ (logical-type).

TABLE XXVII
EXPERIMENTAL RESULTS

Experiment number	Name of flexibility parameter	Initial values	Final values after learning	Mistakes [%] (learning sequence)	Mistakes [%] (testing sequence)
i	ν	0.5	1.0000	4.00%	7.54%
ii	λ	0.5	1.0000	4.00%	7.54%
iii	ν	0	-	7.20%	9.43%
iv	ν	0.5	1.0000	0.80%	3.77%
	p^{τ}	10	10.5256		
	p^{I}	10	9.9573		
	p^{agr}	10	9.8363		
	α^{τ}	1	0.3257		
	α^{I}	1	0.9455		
	α^{agr}	1	9.9958		
v	ν	0.5	1.0000	0.00%	0.00%
	p^{τ}	10	10.2556		
	p^{I}	10	9.9999		
	p^{agr}	10	9.9881		
	α^{τ}	1	0.3067		
	α^{I}	1	0.9592		
	α^{agr}	1	1.0000		
	$\mathbf{w}^{\tau}$	**1**	Fig. 16		
	$\mathbf{w}^{agr}$	**1**	Fig. 16		

TABLE XXVIII
COMPARISON TABLE

Method	Testing Acc.
Corcoran and Sen [6]	100.0
Ishibuchi et al [17]	99.4
González and Pérez (SLAVE[2]) [15]	89.8
González and Pérez (SLAVE[2] with OR and AND) [15]	90.4
González and Pérez (SLAVE[2] with OR, AND and Generalization) [15]	90.4
González and Pérez (SLAVE[2] with all operators) [15]	93.8
our result	**100.0**

4) The OR-type system is less complicated than the AND-type system from a computational point of view. Both systems produce the same type of inference (Mamdani or logical) in the process of learning.
5) The most influential parameters are certainty weights $w_{i,k}^{\tau} \in [0,1]$, $i = 1,\ldots,n$, $k = 1,\ldots N$ and $w_k^{\mathrm{agr}} \in [0,1]$, $k = 1,\ldots N$. They significantly improve the performance of the system in the process of learning.
6) The influence of soft parameters $\alpha^{\tau} \in [0,1]$, $\alpha^{I} \in [0,1]$, $\alpha^{\mathrm{agr}} \in [0,1]$ on the performance of the system varies depending on the problem.

The main advantage of our approach is the possibility of learning a system type. The results of simulations are given in Table XXXI.

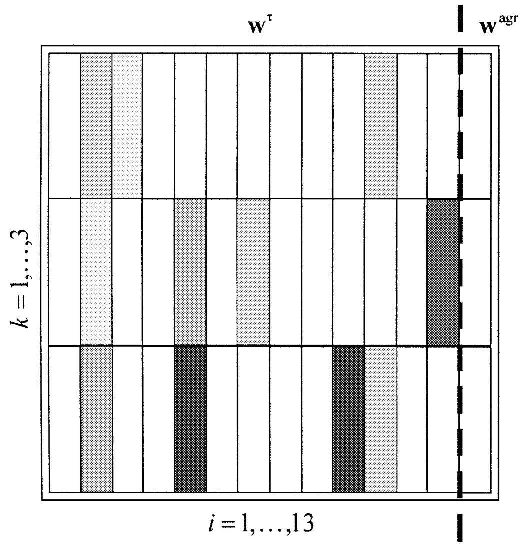

Fig. 16. Weights representation in the Wine Recognition problem $w_{i,k}^{\tau} \in [0,1]$, $w_k^{\mathrm{agr}} \in [0,1]$, $i = 1,\ldots,13$, $k = 1,\ldots,3$ (dark areas correspond to low values and vice versa).

TABLE XXIX
EXPERIMENTAL RESULTS

Experiment number	Name of flexibility parameter	Initial values	Final values after learning	Mistakes [%] (learning sequence)	Mistakes [%] (testing sequence)
i	ν	0.5	1.0000	2.30%	3.90%
ii	λ	0.5	1.0000	2.30%	3.90%
iii	ν	0	-	2.72%	6.34%
iv	ν	0.5	1.0000	2.09%	3.90%
	p^{τ}	10	9.9136		
	p^{I}	10	9.9969		
	p^{agr}	10	10.0631		
	α^{τ}	1	0.9527		
	α^{I}	1	0.9469		
	α^{agr}	1	0.9802		
v	ν	0.5	1.0000	1.67%	3.41%
	p^{τ}	10	8.9705		
	p^{I}	10	9.9999		
	p^{agr}	10	9.9999		
	α^{τ}	1	0.9927		
	α^{I}	1	0.9996		
	α^{agr}	1	0.9798		
	$\mathbf{w}^{\tau}$	**1**	Fig. 17		
	$\mathbf{w}^{agr}$	**1**	Fig. 17		

We conclude that Mamdani-type systems are more suitable to approximation problems, whereas logical-type systems may be preferred for classification problems. It should be emphasized that the results in simulations A, B, C, D, G, H, I, and J outperform the best results known in the literature, although it was not the main goal of our paper.

TABLE XXX
COMPARISON TABLE

Method	Testing Acc.
Dong and Kothari (IG) [9]	94.3
Dong and Kothari (IG+LA) [9]	94.1
Dong and Kothari (GR) [9]	94.7
Dong and Kothari (GR+LA) [9]	94.8
Ster and Dobnikar (Fisher LDA) [50]	96.8
Ster and Dobnikar (MLP+BP) [50]	96.7
Ster and Dobnikar (LVQ) [50]	96.6
Ster and Dobnikar (Bayes) [50]	96.6
Ster and Dobnikar (Naive Bayes) [50]	96.4
Ster and Dobnikar (LDA) [50]	96.0
Ster and Dobnikar (LFC, ASI, ASR) [50]	94.4-95.6
Ster and Dobnikar (Quadratic DA) [50]	34.5
our result	**96.6**

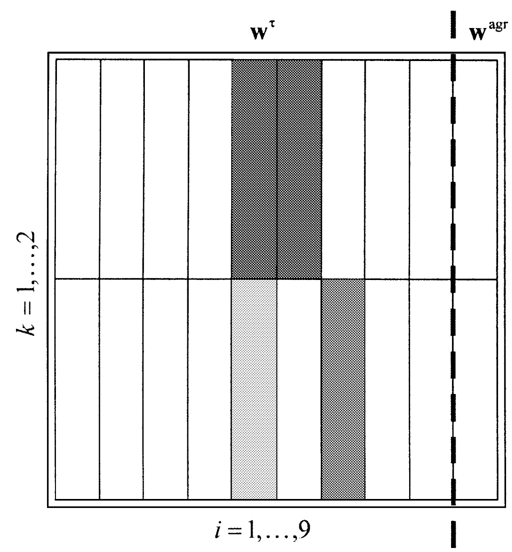

Fig. 17. Weights representation in the Wisconsin Breast Cancer problem $w_{i,k}^{\tau} \in [0,1]$, $w_k^{\mathrm{agr}} \in [0,1]$, $i = 1,\ldots,9$, $k = 1,\ldots,2$ (dark areas correspond to low values and vice versa).

TABLE XXXI
RESULTS OF SIMULATIONS

No	Name of simulation	Type of inference model
A	Glass Identification	logical $\nu = \lambda = 1$
B	Ionosphere	logical $\nu = \lambda = 1$
C	Iris	logical $\nu = \lambda = 1$
D	Static Nonlinear Function (HANG)	Mamdani $\nu = \lambda = 0$
E	Box and Jenkins Gas Furnace	Mamdani $\nu = \lambda = 0$
F	Nonlinear Dynamic Plant	Mamdani $\nu = \lambda = 0$
G	Prima Indians Diabetes	logical $\nu = \lambda = 1$
H	Rice Taste	Mamdani $\nu = \lambda = 0$
I	The Three Monk Problem	Mamdani $\nu = \lambda = 0$
J	Wine Recognition	logical $\nu = \lambda = 1$
K	Wisconsin Breast Cancer Data	logical $\nu = \lambda = 1$

REFERENCES

[1] R. Babuška and M. Setnes, "Data-driven construction of transparent fuzzy models," in *Fuzzy Algorithms for Control*, H. B. Verbruggen, H.-J. Zimmermann, and R. Babuška, Eds. Boston, MA: Kluwer, 1999.

[2] H. R. Berenji and P. Khedkar, "Learning and tuning fuzzy logic controllers through reinforcements," *IEEE Trans. Neural Networks*, vol. 3, pp. 724–740, Sept. 1992.
[3] G. E. P. Box and G. M. Jenkins, *Time Series Analysis, Forecasting and Control*. San Francisco, CA: Holden Day, 1970.
[4] M. Y. Chen and D. A. Linkens, "A systematic neuro-fuzzy modeling framework with application to material property prediction," *IEEE Trans. Fuzzy Syst.*, vol. 31, pp. 781–790, Oct. 2001.
[5] C. C. Chuang, S. F. Su, and S. S. Chen, "Robust TSK fuzzy modeling for function approximation with outliers," *IEEE Trans. Fuzzy Syst.*, vol. 9, pp. 810–821, Dec. 2001.
[6] A. L. Corcoran and S. Sen, "Using real-valued genetic algorithms to evolve rule sets for classification," *Proc. 1st IEEE Conf. Evolutionary Computation*, pp. 120–124, June 1994.
[7] E. Czogala and J. Leski, *Fuzzy and Neuro-Fuzzy Intelligent Systems*. Heidelberg, Germany: Physica-Verlag, 2000.
[8] M. Delgado, A. F. Gómez-Skarmeta, and F. Martin, "Fuzzy clustering-based rapid prototyping for fuzzy rule-based modeling," *IEEE Trans. Fuzzy Syst.*, vol. 5, pp. 223–233, May 1997.
[9] M. Dong and R. Kothari, "Look-ahead based fuzzy decision tree induction," *IEEE Trans. Fuzzy Syst.*, vol. 9, pp. 461–468, June 2001.
[10] J. C. Duan and F. L. Chung, "Cascaded fuzzy neural network based on syllogistic fuzzy reasoning," *IEEE Trans. Fuzzy Syst.*, vol. 9, pp. 293–306, Apr. 2001.
[11] J. C. Fodor, "On fuzzy implication operators," *Fuzzy Sets Syst.*, vol. 42, pp. 293–300, 1991.
[12] A. E. Gaweda and J. M. Zurada, "Data-driven design of fuzzy system with relational input partition," in *Proc. Int. Conf. Fuzzy Systems*, FUZZ-IEEE'2001, Melbourne, Australia, Dec. 2–5, 2001.
[13] ——, "Fuzzy neural network with relational fuzzy rules," in *Proc. Int. Joint Conf. Neural Networks*, vol. 5, Como, Italy, July 23–27, 2000, pp. 3–8.
[14] M. B. Gorzalczany, *Computational Intelligence Systems and Applications. Neuro-Fuzzy and Fuzzy Neural Synergisms*. New York: Springer-Verlag, 2002.
[15] A. González and R. Pérez, "SLAVE: A genetic learning system based on an iterative approach," *IEEE Trans. Fuzzy Syst.*, vol. 7, pp. 176–191, Apr. 1999.
[16] H. Ishibuchi and T. Nakashima, "Effect of rule weights in fuzzy rule-based classification systems," *IEEE Trans. Fuzzy Syst.*, vol. 9, pp. 506–515, Aug. 2001.
[17] H. Ishibuchi, T. Nakashima, and T. Murata, "Performance evaluation of fuzzy classifier systems for multidimensional pattern classification problems," *IEEE Trans. Syst., Man, Cybern. B*, vol. 29, pp. 601–618, Oct. 1999.
[18] R. Jager, "Fuzzy Logic in Control," Ph.D. dissertation, Tech. Univ. Delft, Delft, The Netherlands, 1995.
[19] J. S. R. Jang and C. T. Sun, "Neuro-fuzzy modeling and control," *Proc. IEEE*, vol. 83, pp. 378–406, Mar. 1995.
[20] J. S. R. Jang, "ANFIS: Adaptive-network-based fuzzy inference system," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, pp. 665–685, June 1993.
[21] J. S. Jang, C. T. Sun, and E. Mizutani, *Neuro-Fuzzy and Soft Computing*. Englewood Cliffs, NJ: Prentice-Hall, 1997.
[22] C.-F. Juang and C.-T. Lin, ""An on-line self-constructing neural fuzzy inference network and its applications," *IEEE Trans. Fuzzy Syst.*, vol. 6, pp. 12–32, Feb. 1998.
[23] N. Kasabov, "DENFIS: Dynamic evolving neural-fuzzy inference system and its application for time-series prediction," *IEEE Trans. Fuzzy Syst.*, vol. 10, pp. 144–154, Apr. 2002.
[24] ——, *Foundations of Neural Networks, Fuzzy Systems and Knowledge Engineering*. Cambridge, MA: MIT Press, 1996, p. 550.
[25] ——, "Learning fuzzy rules and approximate reasoning in fuzzy neural networks and hybrid systems," *Fuzzy Sets Syst.*, vol. 82, pp. 135–149, 1996.
[26] E. Kim, M. Park, S. Ji, and M. Park, "A new approach to fuzzy modeling," *IEEE Trans. Fuzzy Syst.*, vol. 5, pp. 328–337, Aug. 1997.
[27] E. Kim, M. Park, S. Kimand, and M. Park, "A transformed input-domain approach to fuzzy modeling," *IEEE Trans. Fuzzy Syst.*, vol. 6, pp. 596–604, Nov. 1998.
[28] E. P. Klement, R. Mesiar, and E. Pap, *Triangular Norms*. Amsterdam, The Netherlands: Kluwer, 2000.
[29] K.-M. Lee and D.-H. Kwang, "A fuzzy neural network model for fuzzy inference and rule tuning," *Int. J. Uncertainty, Fuzziness, Knowledge-Based Syst.*, vol. 2, no. 3, pp. 265–277, 1994.
[30] C. T. Lin and G. C. S. Lee, *Neural Fuzzy Systems-A Neuro-Fuzzy Synergism to Intelligent Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1997.
[31] C. T. Lin and C. S. G. Lee, "Neural-network-based fuzzy logic control and decision system," *IEEE Trans. Comput.*, vol. 40, pp. 1320–1336, Dec. 1991.
[32] C. T. Lin and Y. C. Lu, "A neural fuzzy systems with linguistic teaching signals," *IEEE Trans. Fuzzy Syst.*, vol. 3, pp. 169–189, May 1995.
[33] Y. Lin and G. A. Cunningham III, "A new approach to fuzzy-neural system modeling," *IEEE Trans. Fuzzy Syst.*, vol. 3, pp. 190–198, May 1995.
[34] J. M. Mendel, "Fuzzy logic systems for engineering: a tutorial," *Proc. IEEE*, vol. 83, no. 3, pp. 345–377, Mar. 1995.
[35] ——, *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Upper Saddle River, NJ: Prentice-Hall, 2001.
[36] D. Nauck and R. Kruse, "How the learning of rule weights affects the interpretability of fuzzy systems," in *Proc. World Conf. Computational Intelligence*, Anchorage, AK, 1998, pp. 1235–1240.
[37] ——, "Neuro-fuzzy systems for function approximation," *Fuzzy Sets Syst.*, vol. 101, pp. 261–271, 1999.
[38] D. Nauck, F. Klawon, and R. Kruse, *Foundations of Neuro-Fuzzy Systems*. Chchester, U.K.: Wiley, 1997.
[39] K. Nozaki, H. Ishibuchi, and K. Tanaka, "A simple but powerful heuristic method for generating fuzzy rules from numerical data," *Fuzzy Sets Syst.*, vol. 86, pp. 251–270, 1995.
[40] W. Pedrycz, "A identification algorithm in fuzzy relational systems," *Fuzzy Sets Syst.*, vol. 13, pp. 153–167, 1984.
[41] ——, "Fuzzy neural networks with reference neurons as pattern classifiers," *IEEE Trans. Neural Networks*, vol. 3, pp. 770–775, Sept. 1992.
[42] W. Pedrycz and F. Gomide, *An Introduction to Fuzzy Sets*. Cambridge, MA: MIT Press, 1998.
[43] H. Roubos and M. Setnes, "Compact and transparent fuzzy models and classifiers through iterative complexity reduction," *IEEE Trans. Fuzzy Syst.*, vol. 9, pp. 516–524, Aug. 2001.
[44] D. Rutkowska, *Neuro-Fuzzy Architectures and Hybrid Learning*. New York: Springer-Verlag, 2002.
[45] L. Rutkowski, *New Soft Computing Techniques For System Modeling, Pattern Classification and Image Processing*. New York: Springer-Verlag, to be published.
[46] L. Rutkowski and K. Cpalka, "A general approach to neuro-fuzzy systems," in Proc. 10th IEEE Int. Conf. Fuzzy Syst., Melbourne, Australia, 2001.
[47] ——, Designing and Learning of Adjustable Quasi-Triangular Norms With Applications to Neuro-Fuzzy Systems: Tech. Univ. Czestochowa, Dept. Comput. Eng., 2002.
[48] ——, *Flexible Structures of Neuro-Fuzzy Systems, Quo Vadis Computational Intelligence, Studies in Fuzziness and Soft Computing*: Springer-Verlag, 2000, vol. 54, pp. 479–484.
[49] M. Setnes and H. Roubos, "GA-fuzzy modeling and classification complexity and performance," *IEEE Trans. Fuzzy Syst.*, vol. 8, pp. 509–521, Oct. 2000.
[50] B. Ster and A. Dobnikar, "Neural networks in medical diagnosis: Comparison with other methods," in *Proc. Int. Conf. EANN'96*, A. Bulsari, Ed., 1996, pp. 427–430.
[51] M. Sugeno and K. Tanaka, "Successive identification on a fuzzy model and its applications to prediction of a complex system," *Fuzzy Sets Syst.*, vol. 42, pp. 315–334, 1991.
[52] M. Sugeno and T. Yasukawa, "A fuzzy-logic based approach to qualitative modeling," *IEEE Trans. Fuzzy Syst.*, vol. 1, pp. 7–31, Feb. 1993.
[53] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its application to modeling and control," *IEEE Trans. Syst., Man, Cybern.*, vol. 15, pp. 116–132, 1985.
[54] R. M. Tong, "The evaluation of fuzzy models derived from experimental data," *Fuzzy Sets Syst.*, vol. 4, pp. 1–12, 1980.
[55] E. C. C. Tsang, John. W. T. Lee, and D. S. Yeung, "Tuning certainty factor and local weight of fuzzy production rules using fuzzy neural network," *IEEE Trans. Syst., Man, Cybern. B*, vol. 32, pp. 91–98, Feb. 2002.
[56] UCI Respository of Machine Learning Databases. [Online]. Available: http://www.ics.uci.edu/pub/machine-learning-databases
[57] J. S. Wang and C. S. G. Lee, "Self-adaptive neuro-fuzzy inference systems for classification applications," *IEEE Trans. Fuzzy Syst.*, vol. 10, pp. 790–802, Dec. 2002.

[58] L. X. Wang, *Adaptive Fuzzy Systems and Control*. Englewood Cliffs: Prentice-Hall, 1994.
[59] L. X. Wang and R. Langari, "Building Sugeno-type models using fuzzy discretization and orthogonal parameter estimation techniques," *IEEE Trans. Fuzzy Syst.*, vol. 3, pp. 454–458, Nov. 1995.
[60] L. X. Wang and C. Wei, "Approximation accuracy of some neuro-fuzzy approaches," *IEEE Trans. Fuzzy Syst.*, vol. 8, pp. 470–478, Aug. 2000.
[61] L. Wang and J. Yen, "Application of statistical information criteria for optimal fuzzy model construction," *IEEE Trans. Fuzzy Syst.*, vol. 6, pp. 362–371, Aug. 1998.
[62] ——, "Extracting fuzzy rules for system modeling using a hybrid of genetic algorithms and Kalman filter," *Fuzzy Sets Syst.*, vol. 101, pp. 353–362, 1999.
[63] W. Xu and Y. Z. Lu, "Fuzzy model identification and self-learning for dynamic systems," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-17, pp. 683–689, July 1987.
[64] R. R. Yager, "A general approach to rule aggregation in fuzzy logic control," *Appl. Intell.*, vol. 2, pp. 333–351, 1992.
[65] ——, "Fuzzy logic controller structures," in *Proc. SPIE Symp. Laser Sci Optics Appl.*, 1990, pp. 368–378.
[66] ——, *Fuzzy Sets, Neural Networks and Soft Computing*. New York: Van Nostrand Reinhold, 1997.
[67] R. R. Yager and D. P. Filev, *Essentials of Fuzzy Modeling and Control*. New York: Wiley, 1994.
[68] J. Yen and L. Wang, "Simplifying fuzzy rule based models using orthogonal transformation methods," *IEEE Tran. Syst., Man, Cybern.*, vol. 29, pp. 13–24, Feb. 1999.
[69] D. S. Yeung and C. C. Tsang, "A multi-level weighted fuzzy reasoning algorithm for expert systems," *IEEE Trans. Syst., Man, Cybern. A*, vol. 28, pp. 149–158, Mar. 1998.
[70] Y. Yoshinari, W. Pedrycz, and K. Hirota, "Construction of fuzzy models through clustering techniques," *Fuzzy Sets Syst.*, vol. 54, pp. 157–165, 1993.

Leszek Rutkowski (M'92–SM'94) was born in Wroclaw, Poland, in 1952. He received the M.Sc., Ph. D., and D.Sc. degrees in 1977, 1980, 1986, respectively, all from the Technical University of Wroclaw, Wroclaw, Poland.

Since 1980, he has been with the Technical University of Czestochowa, Czestochowa, Poland, where he is currently a Professor and Chairman of the Computer Engineering Department. From 1987 to 1990, he held a visiting position in the School of Electrical and Computer Engineering, Oklahoma State University, Stillwater. His research interests include neural networks, fuzzy systems, computational intelligence, pattern recognition, and systems identification. He has published more than 100 technical papers including 14 in various series IEEE TRANSACTIONS. He is the author of the books, *Adaptive Filters and Adaptive Signal Processing* and *New Soft Computing Techniques For System Modeling, Pattern Classification and Image Processing* to be published by Springer, and coauthor of two others, *Neural Networks, Genetic Algorithms and Fuzzy Systems* and *Neural Networks for Image Compression*. He is also President and Founder of the Polish Neural Networks Society. He organized and served as General Chair of the Polish Neural Networks Conferences held in 1996, 1997, 1999, 2000, and 2002.

Dr. Rutkowski is an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS.

Krzysztof Cpalka was born in Czestochowa, Poland, in 1972. He received the M.Sc. degree in electrical engineeringand the Ph.D. degree (with honors) in computer engineering, both from the Technical University of Czestochowa, Poland, in 1997and 2002, respectively.

Since 2002, he has been an Assistant Professor in the Department of Computer Engineering, Technical University of Czestochowa, Czestochowa, Poland. His research interests include fuzzy systems, neural networks, evolutionary algorithms, and artificial intelligence. He published over 20 technical papers in journals and conference proceedings.