

Interpretation of Artificial Neural Networks by Means of Fuzzy Rules

Juan L. Castro, Carlos J. Mantas, and José M. Benítez, *Member, IEEE*

Abstract—This paper presents an extension of the method presented by Benítez *et al.* for extracting fuzzy rules from an artificial neural network (ANN) that express exactly its behavior. The extraction process provides an interpretation of the ANN in terms of fuzzy rules. The fuzzy rules presented in this paper are in accordance with the domain of the input variables. These rules use a new operator in the antecedent. The properties and the intuitive meaning of this operator are studied. Next, the role of the biases in the fuzzy rule-based systems is analyzed. Several examples are presented to comment on the obtained fuzzy rule-based systems. Finally, the interpretation of ANNs with two or more hidden layers is also studied.

Index Terms—Artificial neural networks (ANNs), extraction, fuzzy rules, interpretation.

I. INTRODUCTION

ARTIFICIAL neural networks (ANNs) [9], [18] are well known massively parallel computing models which have exhibited excellent behavior in the resolution of problems in many areas such as artificial intelligence, engineering, etc. However, they suffer from the shortcoming of being “black boxes,” i.e., determining why an ANN makes a particular decision is a difficult task.

The “principle of incompatibility” of Zadeh [21] established “the complexity of a system and the precision with which it can be analyzed bear a roughly inverse relation to one another.” This principle can be applied to the ANNs. The ANNs are systems with high complexity, which can achieve a good approximation to the solutions of a problem. Against that, it is very difficult to analyze their performance. According to this principle, the methods for understanding the action carried out by a trained ANN can be classified. We have two possibilities for analyzing an ANN.

- 1) To obtain a comprehensible system that approximates the behavior of the ANN (more comprehension \Rightarrow less complexity \Rightarrow less accuracy). In this case, any rule extraction method [1], [6], [14], [15], [17] can be used. A survey of several rule extraction methods can be found in [2].
- 2) To describe the exact action of the ANN as comprehensibly as possible (same complexity \Rightarrow same accuracy \Rightarrow same comprehension but with other words). In this case, the methods presented in [3] and [11] can be used. The method presented in [11] does not generate an exact direct representation of the ANN, but it aims at this philos-

ophy. This method describes the action of the ANN indicating, by means of polyhedra, the locations of the input space which generate a specific output. It has the drawback of lack of conciseness, because it can produce an exponential number of subpolyhedra in each stage of the algorithm.

On the other hand, in [3], it is proved that ANNs with continuous activation function are fuzzy rule-based systems [7], [22]. Fuzzy rules which express exactly the input–output mapping of the ANNs are extracted. In this way, a more comprehensible description of the action of the ANN is achieved.

However, the fuzzy rules presented in [3] have a problem regarding their use for understanding the action of an ANN. The rules are reasonable for understanding the real line domain function which is calculated by the ANN, but sometimes they are not in the domain where the input variables work. To illustrate this problem, let us consider the following fuzzy rule (presented in [3]), extracted from an ANN that solves the iris classification problem [5].

If *sepal-length* is greater than approximately 22.916 *i-or*
sepal-width is not greater than approximately 137.500 *i-or*
petal-length is greater than approximately 14.013 *i-or*
petal-width is greater than approximately 17.886
 then $y = 13.92$.

The input variables of the IRIS classification problem take values on

$sepal-length \in [4.3, 7.9]$
 $sepal-width \in [2.0, 4.4]$
 $petal-length \in [1.0, 6.9]$
 $petal-width \in [0.1, 2.5]$

Even though the fuzzy propositions in the previous rule are comprehensible, they are not in accordance with the domain of the input variables. For example, the fuzzy proposition *sepal-width is not greater than approximately 137.500* is comprehensible, but $sepal-width \in [2.0, 4.4]$, therefore the degree of this proposition is always “almost one” for any correct input value.

To solve the previous problem, an extension of the former method is presented in this paper. This procedure renders rules

Manuscript received March 21, 2000; revised February 7, 2001.

The authors are with the Department of Computer Science and Artificial Intelligence, University of Granada, Granada 18071, Spain (e-mail: castro@decsai.ugr.es; cmantas@decsai.ugr.es; jmbs@decsai.ugr.es).

Publisher Item Identifier S 1045-9227(02)00342-9.

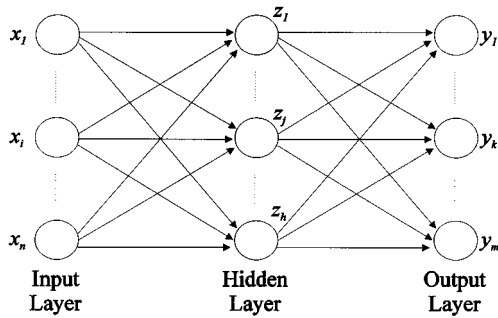


Fig. 1. Multilayer neural network.

whose propositions are in accordance with the domain of the input variables. The logical operator that combines fuzzy propositions changes correspondingly. This new operator has interesting properties with a very intuitive interpretation.

This paper is structured as follows. Section II is a summary of the method for extracting fuzzy rules from an ANN presented in [3]. In Section III, the way of achieving fuzzy rules in accordance with the domains of the input variables is explained. Section IV presents the intuitive meaning of a new operator that appears in the fuzzy rules. Its properties are described in Appendix II. The method for extracting correct fuzzy rules from an ANN is presented in Section V. The role of the biases in the obtained fuzzy rule-based system is analyzed in Section VI. Some examples are presented in Section VII. In Section VIII, the way for interpreting ANNs with two or more hidden layers is exposed. Finally, some conclusions are drawn.

II. ANNs AS FUZZY RULE-BASED SYSTEMS (SUMMARY)

Multilayered feedforward ANNs are the most common model of neural nets, hence they are studied in this work. Let us consider an ANN with input, hidden, and output layers. Let us suppose that the net has n input neurons (x_1, \dots, x_n), h hidden neurons (z_1, \dots, z_h) and m output neurons (y_1, \dots, y_m). Let τ_j the bias for neuron z_j and φ_k for neuron y_k . Let w_{ij} be the weight of the connection from neuron x_i to neuron z_j and β_{jk} the weight of the connection from neuron z_j to neuron y_k . Fig. 1 shows the general layout of these nets. The function the net calculates is

$$F: \mathcal{R}^n \rightarrow \mathcal{R}^m; \quad F(x_1, \dots, x_n) = (y_1, \dots, y_m)$$

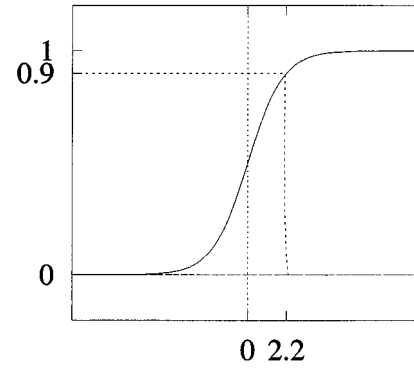
$$y_k = g_A \left(\sum_{j=1}^h (z_j \beta_{jk}) + \varphi_k \right)$$

$$\text{with } z_j = f_A \left(\sum_{i=1}^n (x_i w_{ij}) + \tau_j \right)$$

where g_A and f_A are activation functions. For example, $g_A(x) = x$ and $f_A(x) = 1/(1 + e^{-x})$.

In [3], multilayer feedforward ANNs are seen as additive fuzzy rule-based systems [8]. In these systems, the outputs of each rule are weighted by the activation degree of the rule and then, are added. In the obtained fuzzy system from an ANN, there is a rule R_{jk} per pair of neurons (hidden, output), (z_j, y_k)

$$R_{jk}: \text{if } x_1 \text{ is } A_{jk}^1 \text{ } ior \text{ } x_2 \text{ is } A_{jk}^2 \text{ } ior \dots \text{ } ior \text{ } x_n \text{ is } A_{jk}^n \\ \text{then } y_k = \beta_{jk}$$

Fig. 2. Membership function for fuzzy set A ("greater than approximately 2.2").

where we have the following.

- The system output is the vector whose components are given by $y_k = \sum_{j=1}^h v_{jk} \beta_{jk}$ (additive fuzzy system), where v_{jk} is the firing strength for rule R_{jk} (matching degree between inputs and antecedents). The fuzzy rules R_{jk} can be modified for obtaining a TSK fuzzy system [16] (see Appendix I).
- A_{jk}^i are fuzzy sets obtained from the weights w_{ij} , the biases τ_j and the fuzzy set A defined by the membership function $\mu_A(x) = f_A(x)$ [19], [20] (A may be understood as "greater than approximately 2.2" because $f_A^{-1}(0.9) = 2.2$ [3], see Fig. 2). So

$$\begin{aligned} & x_i w_{ij} + \frac{\tau_j}{n} \text{ is [not] } A \\ \equiv & x_i w_{ij} + \frac{\tau_j}{n} \text{ is [not] greater than approximately 2.2} \\ \equiv & x_i \text{ is [not] greater than approximately } \left(\frac{2.2 - \frac{\tau_j}{n}}{w_{ij}} \right) \\ \equiv & x_i \text{ is [not] } A_{jk}^i \text{ where } \mu_{A_{jk}^i}(x) = f_A \left(x_i w_{ij} + \frac{\tau_j}{n} \right). \end{aligned}$$

- ior is a logic connective defined as

$$ior(a, b) = \frac{ab}{(1-a)(1-b) + ab}, \text{ with } a, b \in (0, 1)$$

where

$$f_A(x_1 w_1 + x_2 w_2) = f_A(x_1 w_1) \text{ } ior \text{ } f_A(x_2 w_2)$$

and

$$f_A(x_i w_i) \equiv \begin{cases} x_i w_i \text{ is } A, & \text{if } w_i > 0.0 \\ x_i |w_i| \text{ is not } A, & \text{if } w_i < 0.0 \end{cases}$$

- There are rules " R_{0k} : If True then $y_k = \varphi_k$," derived from the biases φ_k .

As we have mentioned in the introduction, it may happen that the fuzzy propositions " x_i is [not] A_{jk}^i " are not in accordance with the domains where the variables x_i take values. In the next section, it will be exposed a way of achieving that the fuzzy propositions work within the domain of the variables x_i .

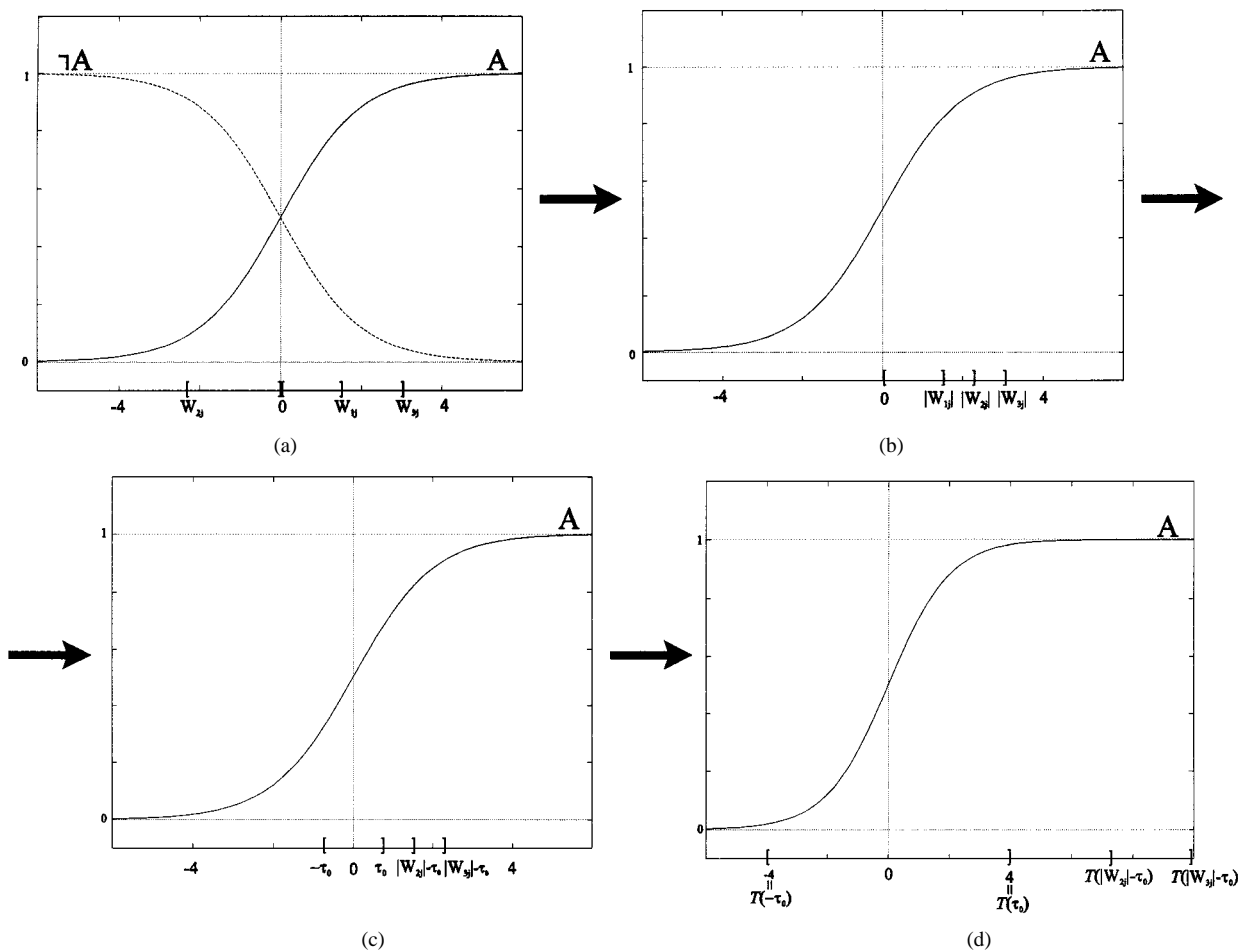


Fig. 3. (a) Domains where (x_1w_{1j}) , (x_2w_{2j}) and (x_3w_{3j}) work together with the fuzzy sets A and $\neg A$. (b) Domains of $(x_i|w_{ij}|)$, $i = 1, 2, 3$. (c) Domains where $(x_i|w_{ij}| - \tau_0)$, $i = 1, 2, 3$ work. (d) Domains where $T(x_i|w_{ij}| - \tau_0)$, $i = 1, 2, 3$ work.

III. COHERENT FUZZY PROPOSITIONS

The idea for attaining fuzzy propositions in accordance with the domains of the input variables consists of transforming these domains with the aim that they become to contain to the intervals where the fuzzy propositions are coherent. To better explain the idea, let us consider a simple example.

Let us suppose three input variables of an ANN, x_i , $i = 1, 2, 3$. These variables will be normalized ($x_i \in [0, 1]$, $i = 1, 2, 3$). The output of the j hidden neuron will be

$$z_j = f_A \left(\sum_{i=1}^3 (x_i w_{ij}) + \tau_j \right)$$

that is equivalent to the degree to which the following proposition activates (coherent in the real line domain):

$$\sum_{i=1}^3 (x_i w_{ij}) + \tau_j \text{ is } A$$

where $\mu_A(x) = f_A(x)$ ($A \equiv$ "greater than approximately 2.2").

Let us suppose that $w_{2j} < 0$, $w_{1j} > 0$, $w_{3j} > 0$ and if $i < k$ then $|w_{ij}| < |w_{kj}|$. Since x_1, x_2 and x_3 take values in $[0, 1]$, the domains for variables (x_1w_{1j}) , (x_2w_{2j}) and (x_3w_{3j}) are $[0, w_{1j}]$, $[w_{2j}, 0]$ and $[0, w_{3j}]$, respectively. Fig. 3(a) illustrates these domains along with the fuzzy sets A and $\neg A$. We

can study the domains of $(x_i|w_{ij}|)$, $i = 1, 2, 3$ [Fig. 3(b)] because if $w_{ij} < 0$, the following lemma can be applied.

Lemma 1: If $\mu_A(x) = f_A(x)$ then $\forall x \in \mathfrak{R}$, $x \text{ is } A \Leftrightarrow -x \text{ is not } A$.

Let $\tau_0 = (\text{minimum}\{|w_{ij}|, i = 1, 2, 3\}/2) = |w_{1j}|/2$. τ_0 can be used for centering the most narrow domain of $(x_i|w_{ij}|)$ $i = 1, 2, 3$ on the origin. The domains of the remaining variables $(x_i|w_{ij}|)$ are correspondingly displaced. Fig. 3(c) illustrates this fact (domains where $(x_i|w_{ij}| - \tau_0)$, $i = 1, 2, 3$ work).

This change can be carried out without difficulty because

$$\begin{aligned} z_j &= f_A \left(\sum_{i=1}^3 (x_i w_{ij}) + \tau_j \right) \\ &= f_A ((x_1|w_{1j}| - \tau_0) - (x_2|w_{2j}| - \tau_0) \\ &\quad + (x_3|w_{3j}| - \tau_0) + (\tau_j + \tau_0 - \tau_0 + \tau_0)) \\ &= f_A ((x_1w_{1j} - \tau_0) + (x_2w_{2j} + \tau_0) \\ &\quad + (x_3w_{3j} - \tau_0) + (\tau_j + \tau_0)) \\ &= f_A(x_1w_{1j} - \tau_0) \text{ ior } f_A(x_2w_{2j} + \tau_0) \\ &\quad \text{ior } f_A(x_3w_{3j} - \tau_0) \text{ ior } f_A(\tau_j + \tau_0) \\ &\equiv (x_1w_{1j} - \tau_0) \text{ is } A \text{ ior } (x_2w_{2j} + \tau_0) \\ &\quad \text{is } A \text{ ior } (x_3w_{3j} - \tau_0) \text{ is } A \text{ ior } (\tau_j + \tau_0) \text{ is } A \\ &\equiv (x_1w_{1j} - \tau_0) \text{ is } A \text{ ior } -(x_2w_{2j} + \tau_0) \\ &\quad \text{is not } A \text{ ior } (x_3w_{3j} - \tau_0) \text{ is } A \text{ ior } (\tau_j + \tau_0) \text{ is } A. \end{aligned}$$

Next, the domains are transformed by the function

$$T(x) = \frac{4}{\tau_0} \cdot x = \frac{8}{\text{minimum}\{|w_{ij}|, i = 1, 2, 3\}} \cdot x = u \cdot x.$$

Since the most interesting area of the neurons activation function (and correspondingly fuzzy set membership function) is the interval $[-4.0, 4.0]$, the transformation T maps the most narrow domain of $(x_i|w_{ij} - \tau_0)$ into $[-4.0, 4.0]$. The fuzzy proposition “ x is [not] A ” is reasonable in this interval.

Hence, the least domain of $(x_i|w_{ij} - \tau_0)$ $i = 1, 2, 3$, is transformed into the interval $[-4.0, 4.0]$, where the fuzzy proposition “ x is [not] A ” is reasonable. The domains of the remaining variables $(T(x_i|w_{ij} - \tau_0))$ contain to the interval $[-4.0, 4.0]$. Fig. 3(d) illustrates this transformation. u will be greater or equal than 1.0, because if u were less than 1.0 the domains would contain into the interval $[-4.0, 4.0]$ and therefore, the transformation T would not be necessary.

The last step consists of finding a logical connective \otimes verifying

$$f_A(x_1+x_2) = f_A(T(x_1)) \otimes f_A(T(x_2)) = f_A(u \cdot x_1) \otimes f_A(u \cdot x_2).$$

With this logical connective we can get

$$\begin{aligned} z_j &= f_A \left(\sum_{i=1}^3 (x_i w_{ij}) + \tau_j \right) \\ &= f_A((x_1|w_{1j} - \tau_0) - (x_2|w_{2j} - \tau_0) \\ &\quad + (x_3|w_{3j} - \tau_0) + (\tau_j + \tau_0 - \tau_0 + \tau_0)) \\ &= f_A(u \cdot (x_1|w_{1j} - \tau_0)) \otimes f_A((-1.0) \cdot u \cdot (x_2|w_{2j} - \tau_0)) \\ &\quad \otimes f_A(u \cdot (x_3|w_{3j} - \tau_0)) \otimes f_A(u \cdot (\tau_j + \tau_0)) \\ &\equiv u \cdot (x_1 w_{1j} - \tau_0) \text{ is } A \otimes (-1.0) \cdot u \cdot (x_2|w_{2j} - \tau_0) \text{ is } A \\ &\quad \otimes (u \cdot (x_3 w_{3j} - \tau_0)) \text{ is } A \otimes u \cdot (\tau_j + \tau_0) \text{ is } A \\ &\equiv u \cdot (x_1 w_{1j} - \tau_0) \text{ is } A \otimes (-1.0) \cdot u \cdot (x_2 w_{2j} + \tau_0) \text{ is not } A \\ &\quad \otimes (u \cdot (x_3 w_{3j} - \tau_0)) \text{ is } A \otimes u \cdot (\tau_j + \tau_0) \text{ is } A. \end{aligned}$$

These fuzzy propositions are in accordance with the domain of the variables $x_i \in [0, 1]$ because of the following.

1) For positive weights, i.e., $w_{ij} > 0.0$

$$\begin{aligned} &u \cdot (x_i w_{ij} - \tau_0) \text{ is } A \\ &\equiv u \cdot (x_i w_{ij} - \tau_0) \text{ is greater than approximately } 2.2 \\ &\equiv x_i \text{ is greater than approximately } \frac{2.2 + u \cdot \tau_0}{u \cdot w_{ij}} \\ &\text{and } \frac{2.2 + u \cdot \tau_0}{u \cdot w_{ij}} \in (0, 1) \text{ (see the following Lemma 2).} \end{aligned}$$

Lemma 2: If $w_{ij} > 0.0$, $\tau_0 = (\text{minimum}\{|w_{ij}|, i = 1, \dots, n\}/2)$ and $u = 4/\tau_0$ then $((2.2 + u \cdot \tau_0)/(u \cdot w_{ij})) \in (0, 1)$.

2) For negative weights, $w_{ij} < 0.0$

$$\begin{aligned} &(-1.0) \cdot u \cdot (x_i w_{ij} + \tau_0) \text{ is not } A \\ &\equiv (-1.0) \cdot u \cdot (x_i w_{ij} + \tau_0) \\ &\quad \text{is not greater than approximately } 2.2 \\ &\equiv x_i \text{ is not greater than approximately } \frac{2.2 + u \cdot \tau_0}{-u \cdot w_{ij}} \\ &\text{and } \frac{2.2 + u \cdot \tau_0}{-u \cdot w_{ij}} \in (0, 1) \text{ (see the next Lemma 3).} \end{aligned}$$

Lemma 3: If $w_{ij} < 0.0$, $\tau_0 = (\text{minimum}\{|w_{ij}|, i = 1, \dots, n\}/2)$ and $u = 4/\tau_0$ then $((2.2 + u \cdot \tau_0)/(-u \cdot w_{ij})) \in (0, 1)$.

Now, the wanted operator \otimes can be defined

Definition 1: Let $n > 0$, $u \geq 1.0$ and $a_i \in (0, 1)$, $i = 1, \dots, n$. The operator $\otimes_n^u : (0, 1)^n \rightarrow (0, 1)$ is defined as

$$\begin{aligned} &\otimes_n^u(a_1, \dots, a_n) \\ &= \frac{a_1^{u-1} \dots a_n^{u-1}}{a_1^{u-1} \dots a_n^{u-1} + (1 - a_1)^{u-1} \dots (1 - a_n)^{u-1}}. \end{aligned}$$

The following lemma confirms that, actually, this is the operator we were looking for in order to state the equality between the newly formulated rules and the network neurons.

Lemma 4: The operator \otimes_n^u verifies

$$f_A(x_1 + \dots + x_n) = \otimes_n^u(f_A(T(x_1)), \dots, f_A(T(x_n)))$$

where $T(x) = u \cdot x$, $n > 0$, $u \geq 1.0$, $f_A(x) = 1/(1 + e^{-x})$ and $x_i \in \mathfrak{R}$.

It is immediate to check the resemblance between this new operator and the *i-or* operator [3]. This relationship extends through a number of properties that both operators share. The most interesting properties of the new operator are described in Appendix II. One aspect of the utmost importance for its intended role in the interpretation of ANNs is that the new operator also has a natural, intuitive meaning.

IV. INTUITIVE INTERPRETATION OF THE OPERATOR \otimes_n^u

The properties of the operator \otimes_n^u (presented in Appendix II) are used for providing it an intuitive interpretation. It can be best exposed through an example.

Let us consider a person who wants to buy a car. He/she is studying a particular model. Initially, he/she does not know whether it is advisable to buy this model (buying advisability ≈ 1.0) or it is unsuitable (buying advisability ≈ 0.0), i.e., buying advisability = $1/2$. With the aim of modifying the *buying advisability*, he/she asks about n characteristics of the car model (motor, starter, speeding up, comfort, etc). For example, what about the motor? (good ≈ 1.0 or bad ≈ 0.0). With these n answers (n values in $(0, 1)$), he/she obtains a new value for the *buying advisability* ($\in (0, 1)$). This last process of aggregation may be modeled using the operator \otimes_n^u .

Let us have a closer look at several points.

- 1) The final decision of buying the car model and the *buying advisability* are different things, although dependent.
- 2) The answers and the *buying advisability* are also different. Hence, it is not reasonable to aggregate a value of *buying advisability* with an answer, i.e., it is reasonable that the operator \otimes_n^u is not associative (Property 2 of Appendix II).
- 3) If a characteristic is not good or bad (answer = $1/2$), it should not influence in the modification of the *buying advisability* (Property 3 of Appendix II). On the other hand, if all the characteristics are neuter (answers = $1/2$), the *buying advisability* goes on with the initial value (= $1/2$) (Property 4 of Appendix II).

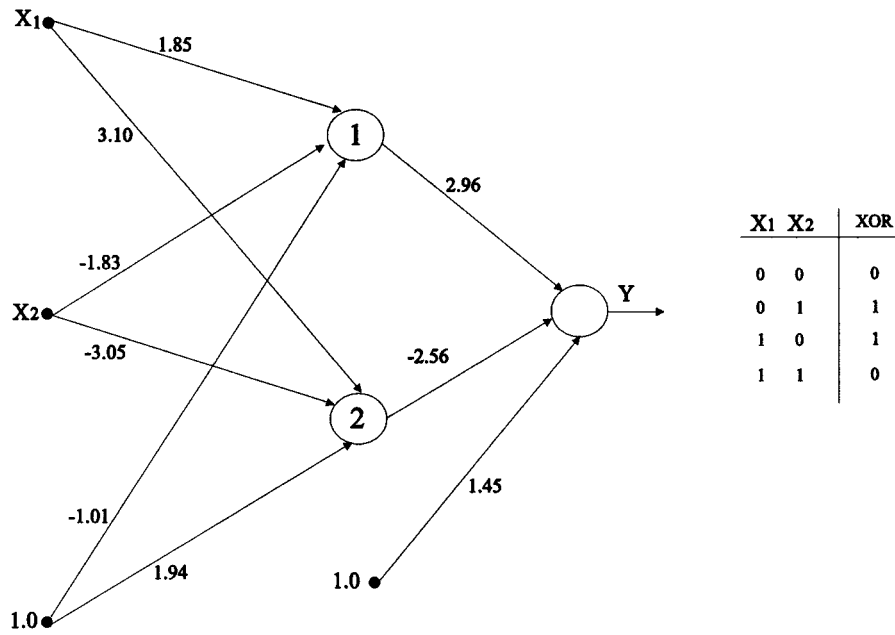


Fig. 4. ANN that solves the XOR problem.

- 4) If the qualities of two characteristics are opposed but of equal strength, they do not influence in the modification of the *buying advisability* (Property 5 of Appendix II).
- 5) If all the characteristics are positive (answers > 1/2), the *buying advisability* increases from 1/2 to 1.0 (Property 6 of Appendix II). On the other hand, if all the characteristics are negative (answers < 1/2), the *buying advisability* decreases from 1/2 to 0.0 (Property 7 of Appendix II).
- 6) If $\Sigma(\text{positive_answers} - 1/2)$ is greater than $\Sigma(1/2 - \text{negative_answers})$, the *buying advisability* increases from 1/2 (Property 8 of Appendix II). On the other hand, if $\Sigma(\text{positive_answers} - 1/2)$ is less than $\Sigma(1/2 - \text{negative_answers})$, the *buying advisability* decreases from 1/2 (Property 9 of Appendix II).
- 7) If a characteristic of the car model is totally perfect, the *buying advisability* will be 1.0 (Property 10 of Appendix II). On the other hand, if a characteristic is totally dreadful, the *buying advisability* will be 0.0 (Property 11 of Appendix II).
- 8) A *buying advisability* of a car model will be less than the *buying advisability* of other car model with better characteristics (Property 12 of Appendix II).
- 9) The credibility of the answers about the characteristics is determined in terms of the parameter u . The greater is u , the lesser is the importance of the answers (influence on the result) and vice versa (Property 13–16 of Appendix II).

We can find \otimes_n^u applicability in many other situations. For example, an editor can use \otimes_n^u to determine whether it is advisable to publish a scientific paper or not, starting from the features of the paper included in the reviewers' reports.

The operator \otimes_n^u determines the advisability of carrying out an action through the aggregation of several evaluations or opinions about independent facts.

V. ANNs AS FUZZY RULE-BASED SYSTEMS

At this point, we have all the tools necessary to express a given ANN (of the kind under study) in terms of fuzzy rules coherent with their input domains. In this section, we present a procedure to produce such transformation.

Let us consider the multilayer ANN with one hidden layer illustrated in Fig. 1. Let us suppose that some w_{ij} weights are positive and others are negative. Without loss of generality, let us consider that $w_{ij} < 0.0$ for $1 \leq i \leq p$ and $w_{ij} > 0.0$ for $p < i \leq n$. The procedure for transforming this ANN into a fuzzy rule-based system is composed of the following steps.

- 1) Transform the domains of the input variables. This can be carried out for each hidden neuron independently, so the obtained system will have fuzzy rules with different operators \otimes_n^u . To obtain fuzzy rules with the same operator \otimes_n^u , a common value u must be used for all the transformations. Hence, this step consists of calculating

$$\tau_0 = \frac{\text{minimum}\{|w_{ij}|, 1 \leq i \leq n, 1 \leq j \leq h\}}{2} \text{ and } u = \frac{4}{\tau_0}.$$

Now, the function $T(x) = u \cdot x$ and the operator \otimes_n^u are common for all the fuzzy rules.

- 2) For each output neuron y_k , a fuzzy rule “ R_{0k} : If True then $y_k = \varphi_k$ ” is added to the rule base.
- 3) For each pair of neurons (hidden, output) (z_j, y_k) , the following rule is added:

$$R_{jk}: \text{ If } -T(x_1 w_{1j} + \tau_0) \text{ is not greater than approximately } 2.2 \quad \otimes_{n+1}^u$$

.....

$-T(x_p w_{pj} + \tau_0)$ is not greater than approximately 2.2 \otimes_{n+1}^u
 $T(x_{(p+1)} w_{(p+1)j} - \tau_0)$ is greater than approximately 2.2 \otimes_{n+1}^u

 $T(x_n w_{nj} - \tau_0)$ is greater than approximately 2.2 \otimes_{n+1}^u
 γ_j
 then $y_k = \beta_{jk}$,

where

- $-T(x_i w_{ij} + \tau_0)$ is not greater than approximately 2.2 $\equiv x_i$ is not greater than approximately $(2.2 + u \cdot \tau_0) / (-u \cdot w_{ij}) = 6.2 / (-u \cdot w_{ij}) = \lambda_{ij}$ and μ is not greater than approximately $\lambda_{ij}(x) = f_A((6.2/\lambda_{ij}) \cdot x - 4)$.
- $T(x_i w_{ij} - \tau_0)$ is greater than approximately 2.2 $\equiv x_i$ is greater than approximately $(2.2 + u \cdot \tau_0) / (u \cdot w_{ij}) = (6.2/u \cdot w_{ij}) = \lambda_{ij}$ and μ is not greater than approximately $\lambda_{ij}(x) = f_A((6.2/\lambda_{ij}) \cdot x - 4)$.
- $\gamma_j = f_A(T(\tau_j - p \cdot \tau_0 + (n - p) \cdot \tau_0))$
Therefore

R_{jk} : If x_1 is not greater than approximately λ_{1j} \otimes_{n+1}^u

 x_p is not greater than approximately λ_{pj} \otimes_{n+1}^u
 $x_{(p+1)}$ is greater than approximately $\lambda_{(p+1)j}$ \otimes_{n+1}^u

 x_n is greater than approximately λ_{nj} \otimes_{n+1}^u
 γ_j
 then $y_k = \beta_{jk}$.

This procedure is an extension of the one presented in the proof for the theorem of equality [3]. It shares its main characteristics specially its constructive nature and its efficiency, running in a time proportional to the size of the net. Moreover, even though it produces propositions coherent with the input domains, this task does not increase its algorithmic complexity.

VI. INTERPRETATION OF THE BIASES

So far we have analyzed the role of the network weights. Now we turn our attention to the other free parameters of the ANN, namely, the biases.

Two kinds of biases appear in the multilayer ANN of Fig. 1: biases φ_k ($k = 1, \dots, m$) of the output neurons and biases τ_j ($j = 1, \dots, h$) of the hidden neurons. Each kind of bias has a different role in the fuzzy rule-based system.

- The biases φ_k generate the rules " R_{0k} : If True then $y_k = \varphi_k$ " in the fuzzy rule-based system. These rules provide a default value for each output y_k . If the remaining rules are fired, they only modify this default output value. This is in straight connection with the human reasoning process

where a default value is modified as new information is considered [10].

- The biases τ_j generate the constants $\gamma_j = f_A(T(\tau_j - p \cdot \tau_0 + (n - p) \cdot \tau_0))$ that appear in the antecedents of the fuzzy rules R_{jk} . In order to understand the interpretation of the constants γ_j , the following lemma explains the role of the γ_j parameters.

Lemma 5: Let $b, a_1, \dots, a_n \in (0, 1)$, $u \geq 1$ and $\otimes_1^u(b) = k \in (0, 1)$. Then we have the following.

- If $(\sum_{i=1}^n a_i/n) > 1/2 \Rightarrow \otimes_{n+1}^u(b, a_1, \dots, a_n) > k$.
- If $(\sum_{i=1}^n a_i/n) < 1/2 \Rightarrow \otimes_{n+1}^u(b, a_1, \dots, a_n) < k$.

According to the previous lemma, the constants γ_j provide a default value, $\otimes_1^u(\gamma_j)$, for the firing strength of the antecedents. The remaining fuzzy propositions in the antecedents only modify this default value. This is a kind of rule more flexible than usual ones where unmatched rules always activate at zero level, so this rule representation power is greater.

VII. EXAMPLES

In this section, two problems are used to illustrate the interpretation of ANNs as fuzzy rule-based systems: the XOR problem [12] and the breast cancer problem [4]. These binary classification problems are solved by means of multilayer feedforward ANNs with one hidden layer composed of two neurons. An output value of the ANN greater than 0.5 corresponds to a classification value equal to 1.0 and an output value less than 0.5 corresponds with a classification value equal to 0.0. These networks are trained with the Backpropagation algorithm [13]. Then, the procedure presented in Section V is applied to extract fuzzy rules from the ANNs.

A. XOR Problem

This problem has two binary inputs. Fig. 4 displays an ANN which solves this problem. Now, we detail the steps for the extraction of fuzzy rules.

Step 1)

$$\begin{aligned} \tau_0 &= \frac{\text{minimum}\{1.85, |3.10|, |-1.83|, |-3.05|\}}{2} \\ &= \frac{1.83}{2} = 0.915 \\ u &= \frac{4}{\tau_0} = \frac{4}{0.915} = 4.38. \end{aligned}$$

Step 2) From the bias $\varphi_1 = 1.45$, the following rule is obtained.

Rule 1. If True then $Y = 1.45$.

Step 3) For the pair of neurons (z_1, y_1)

$$\begin{aligned} \lambda_{11} &= \frac{6.2}{4.38 \cdot 1.85} = 0.76. \\ \lambda_{21} &= -\frac{6.2}{4.38 \cdot (-1.83)} = 0.77. \\ \gamma_1 &= f_A(4.38 \cdot ((-1.01) + \tau_0 - \tau_0)) = 0.012. \end{aligned}$$

Therefore, the following rule is added.

Rule 2. If x_1 is greater than approximately 0.76 $\otimes_3^{4.38}$
 x_2 is not greater than approximately 0.77 $\otimes_3^{4.38}$
 0.012 (where $\otimes_1^{4.38}(0.012) = 0.267$)
 then $Y = 2.96$.

For the pair of neurons (z_2, y_1)

$$\lambda_{12} = \frac{6.2}{4.38 \cdot 3.10} = 0.456.$$

$$\lambda_{22} = -\frac{6.2}{4.38 \cdot (-3.05)} = 0.464.$$

$$\gamma_2 = f_A(4.38 \cdot ((1.94) + \tau_0 - \tau_0)) = 0.999.$$

So, the following rule is added.

Rule 3. If x_1 is greater than approximately 0.456 $\otimes_3^{4.38}$
 x_2 is not greater than approximately 0.464 $\otimes_3^{4.38}$
 0.999 (where $\otimes_1^{4.38}(0.999) = 0.874$)
 then $Y = -2.56$.

Please note the following remarks.

- The fuzzy rule 1 establishes the default output value ($Y = 1.45$). Rules 2 and 3 modify this value in terms of their output values and the firing strength of their antecedents.
- The values $\otimes_1^{4.38}(0.012)$ and $\otimes_1^{4.38}(0.999)$ provide the default firing strength of the antecedents of the fuzzy rules 2 and 3.
- The value $u = 4.38$ determines the influence of the inputs on the modification of the default firing strength in the antecedents of the rules 2 and 3.
- All the fuzzy propositions of the antecedents are in accordance with the domains of the input variables ($x_i \in [0, 1]$, $i = 1, 2$). This property is not fulfilled by the fuzzy rules extracted with the method presented in [3]. These rules are as follows.

- 1) If True then $Y = 1.45$.
- 2) If x_1 is greater than approximately 1.46 *ior*
 x_2 is not greater than approximately 1.48
 then $Y = 2.96$.
- 3) If x_1 is greater than approximately 0.39 *ior*
 x_2 is not greater than approximately 0.40
 then $Y = -2.56$.

B. Breast Cancer Problem

This problem is composed of nine continuous variables ($x_i \in [1, 10]$, normalized to $x_i \in [0, 1]$, $i = 1, \dots, 9$) and one binary output ($Y = 0 \Rightarrow$ benign, $Y = 1 \Rightarrow$ malignant). The input variables are

- x_1 : Clump thickness.
- x_2 : Uniformity of cell size.
- x_3 : Uniformity of cell shape.
- x_4 : Marginal adhesion.
- x_5 : Single epithelial cell size.
- x_6 : Bare nuclei.
- x_7 : Bland chromatin.
- x_8 : Normal nucleoli.
- x_9 : Mitoses.

Next, the fuzzy rules extracted from an ANN that achieves 98.468% of successes on 457 training examples without missing values, are presented.

- 1) If True then $Y = 0.99$.
- 2) If x_1 is not greater than approximately 0.03 $\otimes_{10}^{7.59}$
 x_2 is not greater than approximately 0.24 $\otimes_{10}^{7.59}$
 x_3 is greater than approximately 0.09 $\otimes_{10}^{7.59}$
 x_4 is not greater than approximately 0.07 $\otimes_{10}^{7.59}$
 x_5 is not greater than approximately 0.27 $\otimes_{10}^{7.59}$
 x_6 is not greater than approximately 0.07 $\otimes_{10}^{7.59}$
 x_7 is not greater than approximately 0.03 $\otimes_{10}^{7.59}$
 x_8 is greater than approximately 0.26 $\otimes_{10}^{7.59}$
 x_9 is not greater than approximately 0.03 $\otimes_{10}^{7.59}$
 $1 - 10^{-25} (\approx 1.0)$ (where $\otimes_1^{7.59} (1 - 10^{-25}) = 1 - 10^{-12} (\approx 1.0)$)
 then $Y = -0.97$.
- 3) If x_1 is not greater than approximately 0.34 $\otimes_{10}^{7.59}$
 x_2 is not greater than approximately 0.72 $\otimes_{10}^{7.59}$
 x_3 is greater than approximately 0.20 $\otimes_{10}^{7.59}$
 x_4 is not greater than approximately 0.21 $\otimes_{10}^{7.59}$
 x_5 is not greater than approximately 0.76 $\otimes_{10}^{7.59}$
 x_6 is not greater than approximately 0.27 $\otimes_{10}^{7.59}$

x_7 is not greater than approximately 0.14 $\otimes_{10}^{7.59}$
 x_8 is greater than approximately 0.07 $\otimes_{10}^{7.59}$
 x_9 is not greater than approximately 0.21 $\otimes_{10}^{7.59}$
 10^{-20} (≈ 0.0) (where $\otimes_1^{7.59}(10^{-20}) = 83 \cdot 10^{-7}$ (≈ 0.0))
 then $Y = 3.84$.

It can be noted that the default firing strength of the antecedent of the fuzzy rule 2 is almost 1.0 and the one of the fuzzy rule 3 is almost 0.0. This fact is reasonable thanks to the output value of each rule

$$\begin{aligned}
 & Y_{\text{rule}_1} + 1.0 \cdot Y_{\text{rule}_2} + 0.0 \cdot Y_{\text{rule}_3} \\
 & = 0.99 + 1.0 \cdot (-0.97) + 0.0 \cdot 3.84 = 0.02.
 \end{aligned}$$

Therefore, the classification of the ANN will be benign except that either the input values fire very little the antecedent of the rule 2, or the input values fire very much the antecedent of the rule 3. This last analysis would have to do it a specialist doctor.

VIII. INTERPRETATION OF ANNS WITH TWO OR MORE HIDDEN LAYERS

Previously, the interpretation of ANNs with a single hidden layer has been explained. However, the presented method can not be used for interpreting ANNs with two or more hidden layers, because it is hard to understand the obtained rules. For example, fuzzy rules with the following format can be attained.

If $(x_1 \text{ is } A_{1j}^1 \otimes_{n+1}^u \quad x_2 \text{ is } A_{1j}^2 \otimes_{n+1}^u$
 $\dots \otimes_{n+1}^u \quad x_n \text{ is } A_{1j}^n \otimes_{n+1}^u \gamma_1^A) \text{ is } B_{jk}^1 \otimes_{q+1}^u$
 $(x_1 \text{ is } A_{2j}^1 \otimes_{n+1}^u \quad x_2 \text{ is } A_{2j}^2 \otimes_{n+1}^u$
 $\dots \otimes_{n+1}^u \quad x_n \text{ is } A_{2j}^n \otimes_{n+1}^u \gamma_2^A) \text{ is } B_{jk}^2 \otimes_{q+1}^u$
 \dots
 $(x_1 \text{ is } A_{qj}^1 \otimes_{n+1}^u \quad x_2 \text{ is } A_{qj}^2 \otimes_{n+1}^u$
 $\dots \otimes_{n+1}^u \quad x_n \text{ is } A_{qj}^n \otimes_{n+1}^u \gamma_q^A) \text{ is } B_{jk}^q \otimes_{q+1}^u$
 γ_j^B
 then $y_k = \beta_{jk}$.

The idea for solving this problem consists of transforming the ANN with N ($N \geq 2$) hidden layers into N chained ANNs with a single hidden layer each. Next, every one of these ANNs is interpreted as a comprehensible fuzzy system. This way, chained fuzzy rule-based systems are obtained that express the action of the ANN with N hidden layers.

In order to explain this procedure, let us consider the ANN with two hidden layers illustrated in Fig. 5. This ANN calculates

$$y_k = g_A \left(\sum_{j=1}^h (z_j \beta_{jk}) + \varphi_k \right), \quad k = 1, \dots, m$$

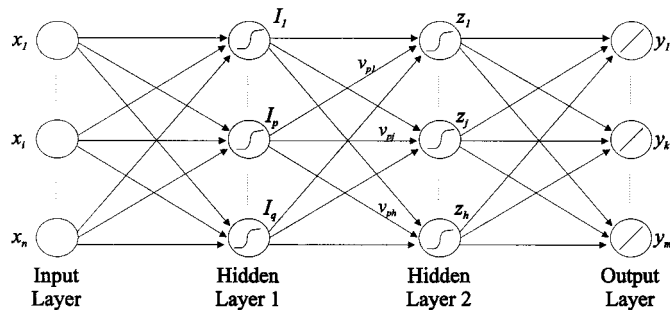


Fig. 5. ANN with two hidden layers.

$$\begin{aligned}
 z_j &= f_A \left(\sum_{p=1}^q (I_p v_{pj}) + \theta_j \right), \quad j = 1, \dots, h \\
 I_p &= f_A \left(\sum_{i=1}^n (x_i w_{ip}) + \tau_p \right), \quad p = 1, \dots, q.
 \end{aligned}$$

We can insert a new hidden layer between the original hidden layers (Fig. 6). This new layer has q neurons, $\bar{I}_{p'}$, $p' = 1, \dots, q$, with linear activation function. The output weights of these neurons $\bar{I}_{p'}$ are the same than the ones of the neurons I_p in the original ANN, that is, $v_{p'j} = v_{pj}$ if $p = p'$, $p = 1, \dots, q$; $p' = 1, \dots, q$ and $j = 1, \dots, h$. The input weights of the neurons $\bar{I}_{p'}$ are

$$\delta_{pp'} = \begin{cases} 1 & \text{if } p = p' \\ 0 & \text{otherwise} \end{cases}, \quad p = 1, \dots, q \text{ and } p' = 1, \dots, q.$$

Besides, the neurons $\bar{I}_{p'}$ do not have biases, i.e., $\bar{I}_{p'} = g_A(\sum_{p=1}^q (I_p \cdot \delta_{pp'}))$.

This new layer does not modify the output of the original ANN with two hidden layers (Fig. 5) because

$$\begin{aligned}
 \bar{I}_{p'} &= g_A \left(\sum_{p=1}^q I_p \cdot \delta_{pp'} \right) = g_A(I_{p'}) = I_{p'} \\
 p' &= 1, \dots, q, \text{ therefore} \\
 z_j &= f_A \left(\sum_{p'=1}^q \bar{I}_{p'} \cdot v_{p'j} \right) = f_A \left(\sum_{p'=1}^q I_{p'} \cdot v_{p'j} \right) \\
 &\equiv f_A \left(\sum_{p=1}^q I_p \cdot v_{pj} \right).
 \end{aligned}$$

The action of the new ANN with three hidden layers (Fig. 6) is equivalent to the chained action of two ANNs with a single hidden layer (Fig. 7).

According to previous sections, an additive fuzzy rule-based system can be extracted from each ANN with a single hidden layer of Fig. 7. Therefore, starting from an ANN with two hidden layers, we can extract two chained fuzzy rule-based systems. Similar reasoning can be used for extracting N chained fuzzy systems from an ANN with N hidden layers.

In this manner, starting from an ANN with two or more hidden layers, we can obtain chained fuzzy systems that are more comprehensible than a single fuzzy system obtained from the ANN.

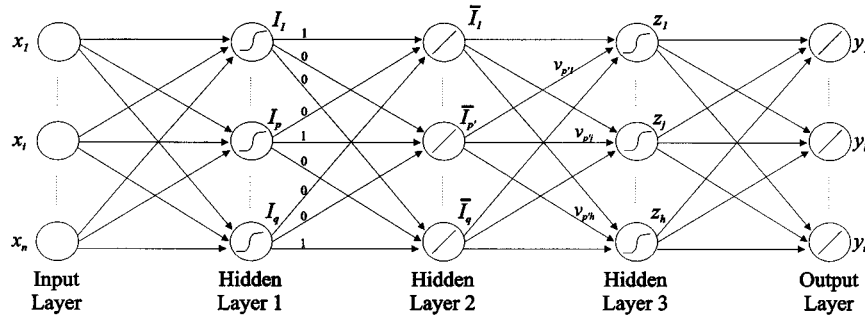


Fig. 6. ANN of Fig. 5 with a new hidden layer.

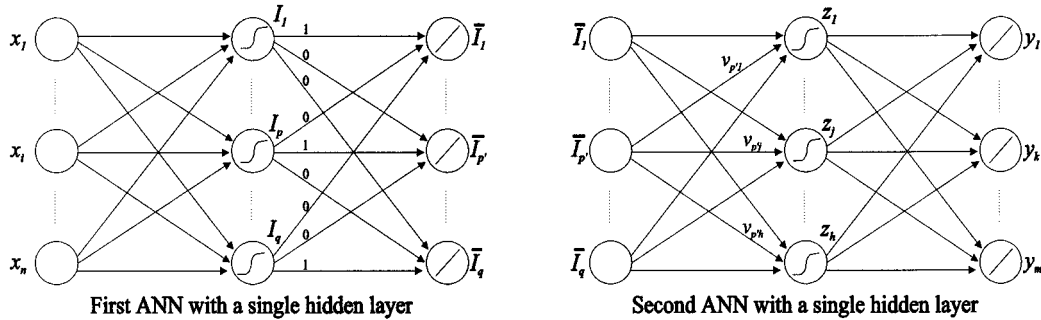


Fig. 7. Two chained ANNs with a single hidden layer that have the same behavior than the ANN of Fig. 6.

IX. CONCLUSION

In this paper, we have presented a procedure to represent the action of an ANN in terms of fuzzy rules. This method extends another one which we had proposed previously. The main achievement of the new method is that the fuzzy rules obtained are in agreement with the domain of the input variables.

In order to keep the equality relationship between the ANN and a corresponding fuzzy rule-based system, a new logical operator has been presented. This operator has a higher representational power than the *i-or* operator. Nevertheless, both operators are very similar, they share many properties. In particular, both of them have a natural interpretation.

These new fuzzy rules along with the new operator render a more coherent and understandable interpretation of multilayered ANNs.

We have also studied the role of biases and therefore, gained further knowledge in the understanding of ANNs behavior.

Finally, the range of neural models to which the proposed interpretation may be applied has been extended by considering ANNs with two or more hidden layers. These networks can be expressed as chained fuzzy rule-based systems.

APPENDIX I

ADDITIVE FUZZY SYSTEM AS TSK FUZZY SYSTEM

Let us consider the following fuzzy rules extracted from a trained ANN:

$$\begin{aligned}
 R_{jk} : & \text{ If } x_1 \text{ is } A_{jk}^1 \text{ ior } x_2 \text{ is } A_{jk}^2 \text{ ior } \dots \text{ ior } x_n \text{ is } A_{jk}^n \\
 & \text{ then } y_k = \beta_{jk} \\
 R_{0k} : & \text{ If True then } y_k = \varphi
 \end{aligned} \tag{a}$$

$j = 1, \dots, h; k = 1, \dots, m.$

Without loss of generality, we can make the assumption $k=1$. In this case, the rules are

$$\begin{aligned}
 R_j : & \text{ If } x_1 \text{ is } A_j^1 \text{ ior } x_2 \text{ is } A_j^2 \text{ ior } \dots \text{ ior } x_n \text{ is } A_j^n \\
 & \text{ then } y = \beta_j \\
 R_0 : & \text{ If True then } y = \varphi
 \end{aligned} \tag{b}$$

$j = 1, \dots, h.$

For a particular input $(x_1^0, x_2^0, \dots, x_n^0)$, the output provided by the additive fuzzy rule-based system [8] is equal to

$$y = \left(\sum_{j=1}^h \beta_j \cdot (A_j^1(x_1^0) \text{ ior } A_j^2(x_2^0) \text{ ior } \dots \text{ ior } A_j^n(x_n^0)) \right) + \varphi \cdot 1. \tag{a.1}$$

Our goal is to modify the fuzzy rules for obtaining the output y using a weighted average (TSK fuzzy rule-based system [16]). If we consider the previous rules as component of a TSK-type fuzzy rule-based system, y would be computed as shown in (a.2) at the bottom of the next page.

In order to achieve this, we can modify the expression (a.1) so we have (a.3) shown at the bottom of the next page, where

$$\beta_j = \frac{b_j - c_j}{h + 1}, \quad \varphi = \sum_{j=1}^h \frac{c_j + \gamma}{h + 1}.$$

Expression (a.3) allow us to rewrite the rules of a fuzzy additive system in (b) as the following rules for a TSK system:

$$\begin{aligned}
 R_j^1 : & \text{ If } x_1 \text{ is } A_j^1 \text{ ior } x_2 \text{ is } A_j^2 \text{ ior } \dots \text{ ior } x_n \text{ is } A_j^n \\
 & \text{ then } y = b_j
 \end{aligned}$$

R_j^2 : If $\neg(x_1 \text{ is } A_j^1 \text{ ior } x_2 \text{ is } A_j^2 \text{ ior } \dots \text{ ior } x_n \text{ is } A_j^n)$
 then $y = c_j$
 R_0 : If True then $y = \gamma$
 $j = 1, \dots, h.$

In addition, the following lemma give us a way to rewrite the previous R_j^2 rule.

A. Lemma

$\neg(x_1 \text{ is } A^1 \text{ ior } x_2 \text{ is } A^2 \text{ ior } \dots \text{ ior } x_n \text{ is } A^n)$
 $= x_1 \text{ is not } A^1 \text{ ior } x_2 \text{ is not } A^2 \text{ ior } \dots \text{ ior } x_n \text{ is not } A^n.$

Proof (Lemma): See (1) shown at the bottom of the page. \lfloor
 Now the rule R_j^2 can be expressed as

R_j^2 : If $(x_1 \text{ is not } A_j^1 \text{ ior } x_2 \text{ is not } A_j^2 \text{ ior } \dots \text{ ior } x_n \text{ is } A_j^n)$
 then $y = c_j$
 $j = 1, \dots, h.$

The expressions and results presented in this section provide a straightforward procedure to transform an additive fuzzy rule-based system into a TSK fuzzy rule-based system. In particular,

ANNs of the kind considered in this paper can be translated into TSK systems.

APPENDIX II

PROPERTIES OF THE OPERATOR \otimes_n^u

Let $a, a_1, \dots, a_n \in (0, 1)$. The operator \otimes_n^u verifies the following properties:

- 1) \otimes_n^u is commutative.
- 2) \otimes_n^u is not associative.
- 3) $\otimes_n^u(1/2, a_1, \dots, a_{n-1}) = \otimes_{n-1}^u(a_1, \dots, a_{n-1})$.
- 4) $\otimes_n^u(1/2, 1/2, \dots, 1/2) = \otimes_0^u() = 1/2$.
- 5)

$$\begin{aligned} & \otimes_n^u(a_1, \dots, a_j, (1 - a_j), \dots, a_n) \\ &= \otimes_{n-2}^u(a_1, \dots, a_{j-1}, a_{j+2}, \dots, a_n). \end{aligned}$$

- 6) If $a_i > 1/2, \forall i = 1, \dots, n \Rightarrow \lim_{n \rightarrow \infty} \otimes_n^u(a_1, \dots, a_n) = 1.0$.
- 7) In particular, if $a_i = a > 1/2 \Rightarrow \lim_{n \rightarrow \infty} \otimes_n^u(a, \dots, a) = 1.0$.
- 8) If $a_i < 1/2, \forall i = 1, \dots, n \Rightarrow \lim_{n \rightarrow \infty} \otimes_n^u(a_1, \dots, a_n) = 0.0$.
- 9) In particular, if $a_i = a < 1/2 \Rightarrow \lim_{n \rightarrow \infty} \otimes_n^u(a, \dots, a) = 0.0$.
- 10) If $((\sum_{i=1}^n a_i)/n) > 1/2 \Rightarrow \otimes_n^u(a_1, \dots, a_n) > 1/2$.

$$y = \frac{\left(\sum_{j=1}^h \beta_j \cdot (A_j^1(x_1^0) \text{ ior } A_j^2(x_2^0) \text{ ior } \dots \text{ ior } A_j^n(x_n^0)) \right) + \varphi \cdot 1}{\left(\sum_{j=1}^h (A_j^1(x_1^0) \text{ ior } A_j^2(x_2^0) \text{ ior } \dots \text{ ior } A_j^n(x_n^0)) \right) + 1}. \quad (\text{a.2})$$

$$\begin{aligned} y &= \left(\sum_{j=1}^h \beta_j \cdot (A_j^1(x_1^0) \text{ ior } A_j^2(x_2^0) \text{ ior } \dots \text{ ior } A_j^n(x_n^0)) \right) + \varphi \cdot 1 \\ &= \frac{\left(\sum_{j=1}^h b_j \cdot (A_j^1(x_1^0) \text{ ior } A_j^2(x_2^0) \text{ ior } \dots \text{ ior } A_j^n(x_n^0)) \right) + \left(\sum_{j=1}^h c_j \cdot (1 - (A_j^1(x_1^0) \text{ ior } A_j^2(x_2^0) \text{ ior } \dots \text{ ior } A_j^n(x_n^0))) \right) + \gamma \cdot 1}{\left(\sum_{j=1}^h (A_j^1(x_1^0) \text{ ior } A_j^2(x_2^0) \text{ ior } \dots \text{ ior } A_j^n(x_n^0)) \right) + \left(\sum_{j=1}^h (1 - (A_j^1(x_1^0) \text{ ior } A_j^2(x_2^0) \text{ ior } \dots \text{ ior } A_j^n(x_n^0))) \right) + 1} \end{aligned} \quad (\text{a.3})$$

$$\begin{aligned} & \neg(x_1 \text{ is } A^1 \text{ ior } x_2 \text{ is } A^2 \text{ ior } \dots \text{ ior } x_n \text{ is } A^n) \\ &= 1 - (x_1 \text{ is } A^1 \text{ ior } x_2 \text{ is } A^2 \text{ ior } \dots \text{ ior } x_n \text{ is } A^n) \\ &= 1 - \frac{A^1(x_1) \cdot A^2(x_2) \cdot \dots \cdot A^n(x_n)}{A^1(x_1) \cdot A^2(x_2) \cdot \dots \cdot A^n(x_n) + (1 - A^1(x_1)) \cdot (1 - A^2(x_2)) \cdot \dots \cdot (1 - A^n(x_n))} \\ &= \frac{(1 - A^1(x_1)) \cdot (1 - A^2(x_2)) \cdot \dots \cdot (1 - A^n(x_n))}{A^1(x_1) \cdot A^2(x_2) \cdot \dots \cdot A^n(x_n) + (1 - A^1(x_1)) \cdot (1 - A^2(x_2)) \cdot \dots \cdot (1 - A^n(x_n))} \\ &= (1 - x_1 \text{ is } A^1) \text{ ior } (x_2 \text{ is } A^2) \text{ ior } \dots \text{ ior } (1 - x_n \text{ is } A^n) \\ &= \neg x_1 \text{ is } A^1 \text{ ior } \neg x_2 \text{ is } A^2 \text{ ior } \dots \text{ ior } \neg x_n \text{ is } A^n \\ &= x_1 \text{ is not } A^1 \text{ ior } x_2 \text{ is not } A^2 \text{ ior } \dots \text{ ior } x_n \text{ is not } A^n. \end{aligned} \quad (\text{1})$$

- 11) If $((\sum_{i=1}^n a_i)/n) < 1/2 \Rightarrow \otimes_n^u(a_1, \dots, a_n) < 1/2$.
- 12) $\lim_{a_j \rightarrow 1} \otimes_n^u(a_1, \dots, a_j, \dots, a_n) = 1.0$.
- 13) $\lim_{a_j \rightarrow 0} \otimes_n^u(a_1, \dots, a_j, \dots, a_n) = 0.0$.
- 14) $a_j^1 < a_j^2 \Rightarrow \otimes_n^u(a_1, \dots, a_j^1, \dots, a_n) > \otimes_n^u(a_1, \dots, a_j^2, \dots, a_n)$.
- 15) $\otimes_n^1(a_1, \dots, a_n) = ior(a_1, \dots, a_n)$.
- 16) If $u_1 < u_2$ and $((\sum_{i=1}^n a_i)/n) > 1/2 \Rightarrow \otimes_n^{u_1}(a_1, \dots, a_n) > \otimes_n^{u_2}(a_1, \dots, a_n) > 1/2$.
- 17) If $u_1 < u_2$ and $((\sum_{i=1}^n a_i)/n) < 1/2 \Rightarrow \otimes_n^{u_1}(a_1, \dots, a_n) < \otimes_n^{u_2}(a_1, \dots, a_n) > 1/2$.
- 18) $\lim_{n \rightarrow \infty} \otimes_n^u(a_1, \dots, a_n) = 1/2$.

and

$$\otimes_2^2(0.4, \otimes_2^2(0.3, 0.7)) = \otimes_2^2(0.4, 0.5) = 0.44949.$$

Therefore, \otimes_n^u is not associative. |

- 3) See (2) shown at the bottom of the page. |
- 4) See (3) shown at the bottom of the page. On the other hand, applying Property 3, we have $\otimes_n^u(1/2, 1/2, \dots, 1/2) = \otimes_0^u()$. Therefore, $\otimes_n^u(1/2, 1/2, \dots, 1/2) = \otimes_0^u() = 1/2$. |
- 5) See (4) shown at the bottom of the page. |
- 6)

A. Proof of Properties

- 1) It is trivial because the product and the addition are commutative. |
- 2) Let $u = 2, n = 3, a_1 = 0.4, a_2 = 0.3$ and $a_3 = 0.7$. Then

$$\otimes_2^2(\otimes_2^2(0.4, 0.3), 0.7) = \otimes_2^2(0.348331, 0.7) = 0.527586$$

$$\begin{aligned} & \lim_{n \rightarrow \infty} \otimes_n^u(a_1, \dots, a_n) \\ &= \lim_{n \rightarrow \infty} \left(\frac{a_1^{u-1} \dots a_n^{u-1}}{a_1^{u-1} \dots a_n^{u-1} + (1-a_1)^{u-1} \dots (1-a_n)^{u-1}} \right) \\ &= \lim_{n \rightarrow \infty} \left(\frac{1}{1 + \frac{(1-a_1)^{u-1}}{a_1^{u-1}} \dots \frac{(1-a_n)^{u-1}}{a_n^{u-1}}} \right) \end{aligned}$$

$$\begin{aligned} \otimes_n^u \left(\frac{1}{2}, a_1, \dots, a_{n-1} \right) &= \frac{\left(\frac{1}{2}\right)^{u-1} \cdot a_1^{u-1} \dots a_{n-1}^{u-1}}{\left(\frac{1}{2}\right)^{u-1} \cdot a_1^{u-1} \dots a_{n-1}^{u-1} + \left(1 - \frac{1}{2}\right)^{u-1} \cdot (1-a_1)^{u-1} \dots (1-a_{n-1})^{u-1}} \\ &= \frac{\left(\frac{1}{2}\right)^{u-1} \cdot a_1^{u-1} \dots a_{n-1}^{u-1}}{\left(\frac{1}{2}\right)^{u-1} \cdot \left[a_1^{u-1} \dots a_{n-1}^{u-1} + (1-a_1)^{u-1} \dots (1-a_{n-1})^{u-1} \right]} \\ &= \frac{a_1^{u-1} \dots a_{n-1}^{u-1}}{a_1^{u-1} \dots a_{n-1}^{u-1} + (1-a_1)^{u-1} \dots (1-a_{n-1})^{u-1}} \\ &= \otimes_{n-1}^u(a_1, \dots, a_{n-1}). \end{aligned} \tag{2}$$

$$\begin{aligned} \otimes_n^u \left(\frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2} \right) &= \frac{\left(\frac{1}{2}\right)^{u-1} \cdot \left(\frac{1}{2}\right)^{u-1} \dots \left(\frac{1}{2}\right)^{u-1}}{\left(\frac{1}{2}\right)^{u-1} \cdot \left(\frac{1}{2}\right)^{u-1} \dots \left(\frac{1}{2}\right)^{u-1} + \left(1 - \frac{1}{2}\right)^{u-1} \cdot \left(1 - \frac{1}{2}\right)^{u-1} \dots \left(1 - \frac{1}{2}\right)^{u-1}} \\ &= \frac{\left(\frac{1}{2}\right)^{u-1} \cdot \left(\frac{1}{2}\right)^{u-1} \dots \left(\frac{1}{2}\right)^{u-1}}{\left(\frac{1}{2}\right)^{u-1} \cdot \left(\frac{1}{2}\right)^{u-1} \dots \left(\frac{1}{2}\right)^{u-1} \cdot [1 + 1]} \\ &= \frac{1}{2} = \frac{1}{2}. \end{aligned} \tag{3}$$

$$\begin{aligned} & \otimes_n^u(a_1, \dots, a_j, (1-a_j), \dots, a_n) \\ &= \frac{a_1^{u-1} \dots a_j^{u-1} \cdot (1-a_j)^{u-1} \dots a_n^{u-1}}{a_1^{u-1} \dots a_j^{u-1} \cdot (1-a_j)^{u-1} \dots a_n^{u-1} + (1-a_1)^{u-1} \dots (1-a_j)^{u-1} \cdot (1-1+a_j)^{u-1} \dots (1-a_n)^{u-1}} \\ &= \frac{a_j^{u-1} \cdot (1-a_j)^{u-1} \cdot \left[a_1^{u-1} \dots a_{j-1}^{u-1} \cdot a_{j+2}^{u-1} \dots a_n^{u-1} \right]}{a_j^{u-1} \cdot (1-a_j)^{u-1} \cdot \left[a_1^{u-1} \dots a_{j-1}^{u-1} \cdot a_{j+2}^{u-1} \dots a_n^{u-1} + (1-a_1)^{u-1} \dots (1-a_{j-1})^{u-1} \cdot (1-a_{j+2})^{u-1} \dots (1-a_n)^{u-1} \right]} \\ &= \otimes_{n-2}^u(a_1, \dots, a_{j-1}, a_{j+2}, \dots, a_n). \end{aligned} \tag{4}$$

$$= \lim_{n \rightarrow \infty} \left(\frac{1}{1 + \left(\frac{(1-a_1)}{a_1} \dots \frac{(1-a_n)}{a_n} \right)^{u^{-1}}} \right).$$

As $a_i > 1/2, \forall i = 1, \dots, n$, we have $((1-a_1)/a_i) < 1.0, \forall i$.

Therefore

$$\lim_{n \rightarrow \infty} \left(\frac{1}{1 + \left(\frac{(1-a_1)}{a_1} \dots \frac{(1-a_n)}{a_n} \right)^{u^{-1}}} \right) = \frac{1}{1 + (0.0)^{u^{-1}}} = 1.0.$$

7)

$$\lim_{n \rightarrow \infty} \otimes_n^u (a_1, \dots, a_n)$$

$$= \lim_{n \rightarrow \infty} \left(\frac{a_1^{u^{-1}} \dots a_n^{u^{-1}}}{a_1^{u^{-1}} \dots a_n^{u^{-1}} + (1-a_1)^{u^{-1}} \dots (1-a_n)^{u^{-1}}} \right)$$

$$= \lim_{n \rightarrow \infty} \left(\frac{1}{1 + \frac{(1-a_1)^{u^{-1}} \dots (1-a_n)^{u^{-1}}}{a_1^{u^{-1}} \dots a_n^{u^{-1}}}} \right)$$

$$= \lim_{n \rightarrow \infty} \left(\frac{1}{1 + \left(\frac{(1-a_1)}{a_1} \dots \frac{(1-a_n)}{a_n} \right)^{u^{-1}}} \right).$$

As $a_i < 1/2, \forall i = 1, \dots, n$, we have $((1-a_1)/a_i) > 1.0, \forall i$.

Therefore

$$\lim_{n \rightarrow \infty} \left(\frac{1}{1 + \left(\frac{(1-a_1)}{a_1} \dots \frac{(1-a_n)}{a_n} \right)^{u^{-1}}} \right) = \frac{1}{1 + \infty^{u^{-1}}} = 0.0.$$

8) We proceed by induction.

Let $n = 1$.

As $a > 1/2$, we have $((1-a)/a) < 1.0 \Rightarrow (((1-a)/a)^{u^{-1}}) > 1.0$.

Therefore

$$\begin{aligned} \otimes_1^u(a) &= \frac{a^{u^{-1}}}{a^{u^{-1}} + (1-a)^{u^{-1}}} \\ &= \frac{1}{1 + \left(\frac{1-a}{a} \right)^{u^{-1}}} > \frac{1}{2}. \end{aligned}$$

Let us suppose that Property 8 is true for n ($n \geq 1$), is Property 8 true for $(n+1)$?

Let us consider two cases.

a) If $a_{n+1} > 1/2$ then

a.1) If $a_i > 1/2 \forall i = 1, \dots, (n+1)$, we have $((1-a_i)/a_i) < 1.0, \forall i$. Therefore

$$\begin{aligned} \otimes_{n+1}^u (a_1, \dots, a_n, a_{n+1}) &= \frac{1}{1 + \left(\frac{(1-a_1)}{a_1} \dots \frac{(1-a_{n+1})}{a_{n+1}} \right)^{u^{-1}}} \\ &> \frac{1}{1+1} = \frac{1}{2}. \end{aligned}$$

a.2) If $\exists a_k < 1/2$ with $1 \leq k < (n+1)$ then we can choose the least k and we can do $a_k = a_{n+1}$ and $a_{n+1} = a_k$. Next, case b) can be applied.

b) If $a_{n+1} < 1/2$ then, as $((\sum_{i=1}^{n+1} a_i)/(n+1)) > 1/2$ we have

$$\begin{aligned} \sum_{i=1}^n a_i &> \frac{1}{2}(n+1) - a_{n+1} \\ &> \frac{1}{2}n + \left(\frac{1}{2} - a_{n+1} \right) > \frac{1}{2}n, \text{ therefore} \\ \frac{\sum_{i=1}^n a_i}{n} &> \frac{1}{2}. \end{aligned}$$

On the other hand

$$\begin{aligned} \frac{\sum_{i=1}^{n+1} a_i}{n+1} &> \frac{1}{2} \Rightarrow \\ a_{n+1} &> \frac{1}{2}(n+1) - \sum_{i=1}^n a_i \end{aligned}$$

By induction, we have

$$\begin{aligned} \frac{\sum_{i=1}^n a_i}{n} &> \frac{1}{2} \\ \Rightarrow \otimes_n^u (a_1, \dots, a_n) &= \frac{1}{1 + \left(\frac{(1-a_1) \dots (1-a_n)}{a_1 \dots a_n} \right)^{u^{-1}}} \\ &> \frac{1}{2} \\ \Rightarrow \left(\frac{(1-a_1) \dots (1-a_n)}{a_1 \dots a_n} \right)^{u^{-1}} &< 1.0 \end{aligned}$$

Hence, together with $((\sum_{i=1}^n a_i)/n) > 1/2$, we have

$$\begin{aligned} \left(\frac{(1-a_1) \dots (1-a_n)}{a_1 \dots a_n} \right)^{u^{-1}} \cdot \left(\frac{(1-a_{n+1})}{a_{n+1}} \right)^{u^{-1}} \\ < 1.0 \cdot \left(\frac{(1-a_{n+1})}{a_{n+1}} \right)^{u^{-1}} \\ > \left(\frac{1}{\frac{1}{2}(n+1) - \sum_{i=1}^n a_i} - 1 \right)^{u^{-1}} \\ < \left(\frac{4}{(n+1)n} - 1 \right)^{u^{-1}} \\ \leq \left(\frac{4}{2} - 1 \right)^{u^{-1}} = 1^{u^{-1}} = 1. \end{aligned}$$

Therefore

$$\begin{aligned} \otimes_{n+1}^u (a_1, \dots, a_n, a_{n+1}) &= \frac{1}{1 + \left(\frac{(1-a_1) \dots (1-a_{n+1})}{a_1 \dots a_{n+1}} \right)^{u^{-1}}} \\ &> \frac{1}{1+1} = \frac{1}{2}. \end{aligned}$$

9) Once again, a proof can be obtained by induction.

Let $n = 1$.

As $a < 1/2$, we have $((1-a)/a) > 1.0 \Rightarrow (((1-a)/a)^{u^{-1}}) > 1.0$.

Therefore

$$\begin{aligned} \otimes_1^u(a) &= \frac{a^{u-1}}{a^{u-1} + (1-a)^{u-1}} \\ &= \frac{1}{1 + \left(\frac{1-a}{a}\right)^{u-1}} < \frac{1}{2}. \end{aligned}$$

Let us suppose that Property 9 is true for n ($n \geq 1$), is Property 9 true for $(n+1)$?:

Let us consider two cases.

a) If $a_{n+1} < 1/2$ then

- a.1) If $a_i < 1/2 \forall i = 1, \dots, (n+1)$, we have $((1-a_i)/a_i) > 1.0, \forall i$. Therefore

$$\begin{aligned} \otimes_{n+1}^u(a_1, \dots, a_n, a_{n+1}) &= \frac{1}{1 + \left(\frac{(1-a_1)}{a_1} \dots \frac{(1-a_{n+1})}{a_{n+1}}\right)^{u-1}} \\ &< \frac{1}{1+1} = \frac{1}{2}. \end{aligned}$$

- a.2) If $\exists a_k > 1/2$ with $1 \leq k < (n+1)$ then we can choose the least k and we can do $a_k = a_{n+1}$ and $a_{n+1} = a_k$. Next, case b) can be applied.

b) If $a_{n+1} > 1/2$ then, as $((\sum_{i=1}^{n+1} a_i)/(n+1)) > 1/2$ we have

$$\begin{aligned} \sum_{i=1}^n a_i &< \frac{1}{2}(n+1) - a_{n+1} \\ &> \frac{1}{2}n + \left(\frac{1}{2} - a_{n+1}\right) < \frac{1}{2}n, \text{ therefore} \\ \frac{\sum_{i=1}^n a_i}{n} &< \frac{1}{2}. \end{aligned}$$

On the other hand

$$\begin{aligned} \frac{\sum_{i=1}^{n+1} a_i}{n+1} &< \frac{1}{2} \\ \Rightarrow a_{n+1} &> \frac{1}{2}(n+1) \sum_{i=1}^n a_i. \end{aligned}$$

By induction, we have

$$\begin{aligned} \frac{\sum_{i=1}^n a_i}{n} < \frac{1}{2} &\Rightarrow \otimes_n^u(a_1, \dots, a_n) \\ &= \frac{1}{1 + \left(\frac{(1-a_1) \dots (1-a_n)}{a_1 \dots a_n}\right)^{u-1}} < \frac{1}{2} \\ &\Rightarrow \left(\frac{(1-a_1) \dots (1-a_n)}{a_1 \dots a_n}\right)^{u-1} > 1.0. \end{aligned}$$

Hence, together with $((\sum_{i=1}^n a_i)/n) < 1/2$, we have

$$\begin{aligned} &\left(\frac{(1-a_1) \dots (1-a_n)}{a_1 \dots a_n}\right)^{u-1} \cdot \left(\frac{(1-a_{n+1})}{a_{n+1}}\right)^{u-1} \\ &> 1.0 \cdot \left(\frac{(1-a_{n+1})}{a_{n+1}}\right)^{u-1} \\ &> \left(\frac{1}{\frac{1}{2}(n+1) \sum_{i=1}^n a_i} - 1\right)^{u-1} \\ &> \left(\frac{4}{(n+1)n} - 1\right)^{u-1} \\ &\geq \left(\frac{4}{2} - 1\right)^{u-1} = 1^{u-1} = 1. \end{aligned}$$

Therefore

$$\begin{aligned} \otimes_{n+1}^u(a_1, \dots, a_n, a_{n+1}) &= \frac{1}{1 + \left(\frac{(1-a_1) \dots (1-a_{n+1})}{a_1 \dots a_{n+1}}\right)^{u-1}} \\ &< \frac{1}{1+1} = \frac{1}{2}. \end{aligned}$$

- 10) See (5) shown at the bottom of the page.]
 11) See (6) shown at the bottom of the page.]
 12) If $a_j^1 < a_j^2$ then $((1-a_j^1)/a_j^1) > ((1-a_j^2)/a_j^2)$. Therefore we have (7) shown at the bottom of the next page.]

$$\begin{aligned} \lim_{a_j \rightarrow 1} \otimes_n^u(a_1, \dots, a_j, \dots, a_n) &= \lim_{a_j \rightarrow 1} \frac{a_1^{u-1} \dots a_j^{u-1} \dots a_n^{u-1}}{a_1^{u-1} \dots a_j^{u-1} \dots a_n^{u-1} + (1-a_1)^{u-1} \dots (1-a_j)^{u-1} \dots (1-a_n)^{u-1}} \\ &= \frac{a_1^{u-1} \dots a_{j-1}^{u-1} \cdot a_{j+1}^{u-1} \dots a_n^{u-1}}{a_1^{u-1} \dots a_{j-1}^{u-1} \cdot a_{j+1}^{u-1} \dots a_n^{u-1} + 0.0} = 1.0. \end{aligned} \tag{5}$$

$$\begin{aligned} &\lim_{a_j \rightarrow 0} \otimes_n^u(a_1, \dots, a_j, \dots, a_n) \\ &= \lim_{a_j \rightarrow 0} \frac{a_1^{u-1} \dots a_j^{u-1} \dots a_n^{u-1}}{a_1^{u-1} \dots a_j^{u-1} \dots a_n^{u-1} + (1-a_1)^{u-1} \dots (1-a_j)^{u-1} \dots (1-a_n)^{u-1}} \\ &= \frac{0.0}{0.0 + (1-a_1)^{u-1} \dots (1-a_{j-1})^{u-1} \cdot (1-a_{j+1})^{u-1} \dots (1-a_n)^{u-1}} = 0.0. \end{aligned} \tag{6}$$

13)

$$\begin{aligned} \otimes_n^1(a_1, \dots, a_n) &= \frac{a_1^{1-1} \cdots a_n^{1-1}}{a_1^{1-1} \cdots a_n^{1-1} + (1-a_1)^{1-1} \cdots (1-a_n)^{1-1}} \\ &= \frac{a_1 \cdots a_n}{a_1 \cdots a_n + (1-a_1) \cdots (1-a_n)} \\ &= \text{ior}(a_1, \dots, a_n). \end{aligned}$$

14) As $(\sum_{i=1}^n a_i/n) > 1/2$ then $\otimes_n^{u_1}(a_1, \dots, a_n) > 1/2$
and $\otimes_n^{u_2}(a_1, \dots, a_n) > 1/2$.

So

$$\begin{aligned} \left(\frac{(1-a_1) \cdots (1-a_n)}{a_1 \cdots a_n}\right)^{u_1^{-1}} &< 1.0 \text{ and} \\ \left(\frac{(1-a_1) \cdots (1-a_n)}{a_1 \cdots a_n}\right)^{u_2^{-1}} &> 1.0 \\ \Rightarrow \ln \left(\frac{(1-a_1) \cdots (1-a_n)}{a_1 \cdots a_n}\right)^{u_1^{-1}} &< 0.0 \text{ and} \\ \ln \left(\frac{(1-a_1) \cdots (1-a_n)}{a_1 \cdots a_n}\right)^{u_2^{-1}} &> 0.0. \end{aligned}$$

Hence and as $1 \leq u_1 < u_2$, we have

$$\begin{aligned} \ln \left(\frac{(1-a_1) \cdots (1-a_n)}{a_1 \cdots a_n}\right)^{u_1^{-1}} & \\ = u_1^{-1} \ln \left(\frac{(1-a_1) \cdots (1-a_n)}{a_1 \cdots a_n}\right) & \\ < u_2^{-1} \ln \left(\frac{(1-a_1) \cdots (1-a_n)}{a_1 \cdots a_n}\right) & \\ = \ln \left(\frac{(1-a_1) \cdots (1-a_n)}{a_1 \cdots a_n}\right)^{u_2^{-1}} & \\ \Rightarrow \left(\frac{(1-a_1) \cdots (1-a_n)}{a_1 \cdots a_n}\right)^{u_1^{-1}} & \\ > \left(\frac{(1-a_1) \cdots (1-a_n)}{a_1 \cdots a_n}\right)^{u_2^{-1}}. & \end{aligned}$$

Therefore

$$\otimes_n^{u_1}(a_1, \dots, a_n) = \frac{1}{1 + \left(\frac{(1-a_1) \cdots (1-a_n)}{a_1 \cdots a_n}\right)^{u_1^{-1}}}$$

$$\begin{aligned} &> \frac{1}{1 + \left(\frac{(1-a_1) \cdots (1-a_n)}{a_1 \cdots a_n}\right)^{u_2^{-1}}} \\ &= \otimes_n^{u_2}(a_1, \dots, a_n). \end{aligned}$$

15) As $(\sum_{i=1}^n a_i/n) < 1/2$ then $\otimes_n^{u_1}(a_1, \dots, a_n) < 1/2$
and $\otimes_n^{u_2}(a_1, \dots, a_n) < 1/2$.

So

$$\begin{aligned} \left(\frac{(1-a_1) \cdots (1-a_n)}{a_1 \cdots a_n}\right)^{u_1^{-1}} &> 1.0 \text{ and} \\ \left(\frac{(1-a_1) \cdots (1-a_n)}{a_1 \cdots a_n}\right)^{u_2^{-1}} &> 1.0 \\ \Rightarrow \ln \left(\frac{(1-a_1) \cdots (1-a_n)}{a_1 \cdots a_n}\right)^{u_1^{-1}} &> 0.0 \text{ and} \\ \ln \left(\frac{(1-a_1) \cdots (1-a_n)}{a_1 \cdots a_n}\right)^{u_2^{-1}} &> 0.0. \end{aligned}$$

Hence and as $1 \leq u_1 < u_2$, we have

$$\begin{aligned} \ln \left(\frac{(1-a_1) \cdots (1-a_n)}{a_1 \cdots a_n}\right)^{u_1^{-1}} & \\ = u_1^{-1} \ln \left(\frac{(1-a_1) \cdots (1-a_n)}{a_1 \cdots a_n}\right) & \\ > u_2^{-1} \ln \left(\frac{(1-a_1) \cdots (1-a_n)}{a_1 \cdots a_n}\right) & \\ = \ln \left(\frac{(1-a_1) \cdots (1-a_n)}{a_1 \cdots a_n}\right)^{u_2^{-1}} & \\ \Rightarrow \left(\frac{(1-a_1) \cdots (1-a_n)}{a_1 \cdots a_n}\right)^{u_1^{-1}} & \\ > \left(\frac{(1-a_1) \cdots (1-a_n)}{a_1 \cdots a_n}\right)^{u_2^{-1}}. & \end{aligned}$$

Therefore

$$\begin{aligned} \otimes_n^{u_1}(a_1, \dots, a_n) &= \frac{1}{1 + \left(\frac{(1-a_1) \cdots (1-a_n)}{a_1 \cdots a_n}\right)^{u_1^{-1}}} \\ &< \frac{1}{1 + \left(\frac{(1-a_1) \cdots (1-a_n)}{a_1 \cdots a_n}\right)^{u_2^{-1}}} \\ &= \otimes_n^{u_2}(a_1, \dots, a_n). \end{aligned}$$

$$\begin{aligned} \otimes_n^u(a_1, \dots, a_j^1, \dots, a_n) &= \frac{a_1^{u-1} \cdots (a_j^1)^{u-1} \cdots a_n^{u-1}}{a_1^{u-1} \cdots (a_j^1)^{u-1} \cdots a_n^{u-1} + (1-a_1)^{u-1} \cdots (1-a_j^1)^{u-1} \cdots (1-a_n)^{u-1}} \\ &= \frac{1}{1 + \left(\frac{(1-a_1) \cdots (1-a_{j-1}) \cdot (1-a_{j+1}) \cdots (1-a_n)}{a_1 \cdots a_{j-1} \cdot a_{j+1} \cdots a_n}\right)^{u-1} \left(\frac{(1-a_j^1)}{a_j^1}\right)^{u-1}} \\ &< \frac{1}{1 + \left(\frac{(1-a_1) \cdots (1-a_{j-1}) \cdot (1-a_{j+1}) \cdots (1-a_n)}{a_1 \cdots a_{j-1} \cdot a_{j+1} \cdots a_n}\right)^{u-1} \left(\frac{(1-a_j^2)}{a_j^2}\right)^{u-1}} \\ &= \otimes_n^u(a_1, \dots, a_j^2, \dots, a_n). \end{aligned} \tag{7}$$

16)

$$\begin{aligned} & \lim_{u \rightarrow \infty} \otimes_n^u (a_1, \dots, a_n) \\ &= \lim_{n \rightarrow \infty} \left(\frac{(a_1 \cdots a_n)^{u^{-1}}}{(a_1 \cdots a_n)^{u^{-1}} + ((1 - a_1) \cdots (1 - a_n))^{u^{-1}}} \right) \\ &= \frac{(a_1 \cdots a_n)^0}{(a_1 \cdots a_n)^0 + ((1 - a_1) \cdots (1 - a_n))^0} = \frac{1}{1 + 1} = \frac{1}{2}. \end{aligned}$$

finally, from $w_{ij} > 0.0$ we have

$$0.0 < \frac{6.2 \text{ minimum}\{|w_{ij}|, i = 1, \dots, n\}}{8 w_{ij}} \leq \frac{6.2}{8} < 1.0.$$

C. Proof (Lemma 3)

From $u = 4/\tau_0$, we have

$$\begin{aligned} \frac{2.2 + u \cdot \tau_0}{-u \cdot w_{ij}} &= \frac{(2.2 + 4) \cdot \tau_0}{-4 \cdot w_{ij}} \\ &= \frac{6.2 \cdot \tau_0}{-4 \cdot w_{ij}} \end{aligned}$$

from $\tau_0 = (\text{minimum}\{|w_{ij}|, i = 1, \dots, n\}/2)$, we have

$$\frac{6.2 \cdot \tau_0}{-4 \cdot w_{ij}} = \frac{6.2 \cdot \text{minimum}\{|w_{ij}|, i = 1, \dots, n\}}{-8 \cdot w_{ij}}.$$

finally, from $w_{ij} < 0.0$ we have

$$\begin{aligned} & \frac{6.2 \text{ minimum}\{|w_{ij}|, i = 1, \dots, n\}}{-8 w_{ij}} \\ &= \frac{6.2 \text{ minimum}\{|w_{ij}|, i = 1, \dots, n\}}{8 |w_{ij}|} \end{aligned}$$

therefore

$$0.0 < \frac{6.2 \text{ minimum}\{|w_{ij}|, i = 1, \dots, n\}}{8 w_{ij}} \leq \frac{6.2}{8} < 1.0.$$

APPENDIX III PROOF OF RESULTS

A. Proof (Lemma 1)

From [3], we have $f_A(-x) = 1 - f_A(x)$. Therefore

$$\begin{aligned} & \forall x \in \mathfrak{R}, x \text{ is } A \\ & \Leftrightarrow \mu_A(x) = f_A(x) = 1 - (1 - f_A(x)) \\ & = 1 - f_A(-x) = 1 - \mu_A(-x) = \mu_{\neg A}(-x) \\ & \Leftrightarrow -x \text{ is } \neg A \\ & \Leftrightarrow -x \text{ is not } A. \end{aligned}$$

B. Proof (Lemma 2)

From $u = 4/\tau_0$, we have

$$\begin{aligned} \frac{2.2 + u \cdot \tau_0}{u \cdot w_{ij}} &= \frac{(2.2 + 4) \cdot \tau_0}{4 \cdot w_{ij}} \\ &= \frac{6.2 \cdot \tau_0}{4 \cdot w_{ij}} \end{aligned}$$

from $\tau_0 = (\text{minimum}\{|w_{ij}|, i = 1, \dots, n\}/2)$, we have

$$\frac{6.2 \cdot \tau_0}{4 \cdot w_{ij}} = \frac{6.2 \cdot \text{minimum}\{|w_{ij}|, i = 1, \dots, n\}}{8 \cdot w_{ij}}$$

D. Proof (Lemma 4)

By Definition 1, we have (8) shown at the bottom of the page.

$$\otimes_n^u (f_A(T(x_1)), \dots, f_A(T(x_n)))$$

$$\begin{aligned} &= \frac{(f_A(T(x_1)))^{u^{-1}} \cdots (f_A(T(x_n)))^{u^{-1}}}{(f_A(T(x_1)))^{u^{-1}} \cdots (f_A(T(x_n)))^{u^{-1}} + (1 - f_A(T(x_1)))^{u^{-1}} \cdots (1 - f_A(T(x_n)))^{u^{-1}}} \\ &= \frac{1}{1 + \left(\frac{1 - f_A(T(x_1))}{f_A(T(x_1))} \cdots \frac{1 - f_A(T(x_n))}{f_A(T(x_n))} \right)^{u^{-1}}} \\ &= \frac{1}{1 + \left(\left(\frac{1}{f_A(T(x_1))} - 1 \right) \cdots \left(\frac{1}{f_A(T(x_n))} - 1 \right) \right)^{u^{-1}}} \\ &= \frac{1}{1 + ((1 + e^{-T(x_1)} - 1) \cdots (1 + e^{-T(x_n)} - 1))^{u^{-1}}} \\ &= \frac{1}{1 + (e^{-(T(x_1) + \cdots + T(x_n))})^{u^{-1}}} \\ &= \frac{1}{1 + (e^{-u(x_1 + \cdots + x_n)})^{u^{-1}}} \\ &= \frac{1}{1 + e^{-(x_1 + \cdots + x_n)}} = f_A(x_1 + \cdots + x_n). \end{aligned} \tag{8}$$

E. Proof (Lemma 5)

1)

$$\begin{aligned}\otimes_1^u(b) &= \frac{b^{u-1}}{b^{u-1} + (1-b)^{u-1}} \\ &= \frac{1}{1 + \left(\frac{1-b}{b}\right)^{u-1}} \\ &= k \Rightarrow \left(\frac{1-b}{b}\right)^{u-1} = \frac{1-k}{k}\end{aligned}$$

Hence and as
 $\left(\frac{(1-a_1)\cdots(1-a_n)}{a_1\cdots a_n}\right)^{u-1} < 1.0$ because
 $\left(\frac{\sum_{i=1}^n a_i}{n}\right) > 1/2$, we have

$$\begin{aligned}\otimes_n^u(b, a_1, \dots, a_n) &= \frac{1}{1 + \left(\frac{(1-a_1)\cdots(1-a_n)}{a_1\cdots a_n}\right)^{u-1} \left(\frac{(1-b)}{b}\right)^{u-1}} \\ &= \frac{1}{1 + \left(\frac{(1-a_1)\cdots(1-a_n)}{a_1\cdots a_n}\right)^{u-1} \left(\frac{(1-k)}{k}\right)} \\ &> \frac{1}{1 + \left(\frac{(1-k)}{k}\right)} = \frac{k}{k+1-k} = k.\end{aligned}$$

2) As $\left(\frac{(1-b)}{b}\right)^{u-1} = (1-k)/k$ because $\otimes_1^u(b) = k$
 and as $\left(\frac{(1-a_1)\cdots(1-a_n)}{a_1\cdots a_n}\right)^{u-1} > 1.0$
 because $\left(\frac{\sum_{i=1}^n a_i}{n}\right) < 1/2$, we have

$$\begin{aligned}\otimes_n^u(b, a_1, \dots, a_n) &= \frac{1}{1 + \left(\frac{(1-a_1)\cdots(1-a_n)}{a_1\cdots a_n}\right)^{u-1} \left(\frac{(1-b)}{b}\right)^{u-1}} \\ &= \frac{1}{1 + \left(\frac{(1-a_1)\cdots(1-a_n)}{a_1\cdots a_n}\right)^{u-1} \left(\frac{(1-k)}{k}\right)} \\ &< \frac{1}{1 + \left(\frac{(1-k)}{k}\right)} = \frac{k}{k+1-k} = k\end{aligned}$$

REFERENCES

- [1] J. A. Alexander and M. C. Mozer, "Template-based procedures for neural network interpretation," *Neural Networks*, vol. 12, pp. 479–498, 1999.
- [2] R. Andrews, J. Diederich, and A. B. Tickle, "Survey and critique of techniques for extracting rules from trained artificial neural networks," *Knowledge-Based Syst.*, vol. 8, no. 6, pp. 373–389, 1995.
- [3] J. M. Benítez, J. L. Castro, and I. Requena, "Are artificial neural networks black boxes?," *IEEE Trans. Neural Networks*, vol. 8, pp. 1156–1164, Sept. 1997.
- [4] K. P. Bennet and O. L. Mangasarian, "Robust linear programming discrimination of two linearly inseparable problems," in *Optimization Methods and Software*. New York: Gordon and Breach, 1992, vol. 1, pp. 23–34.
- [5] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, pp. 179–188, 1936.
- [6] L. Fu, "Rule generation from neural networks," *IEEE Trans. Syst., Man, Cybern.*, vol. 24, pp. 1114–1124, 1994.
- [7] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
- [8] B. Kosko, "Fuzzy systems are universal approximators," *IEEE Trans. Comput.*, vol. 43, pp. 1324–1333, Nov. 1994.
- [9] R. P. Lippmann, "An introduction to computing with neural nets," *IEEE ASSP Mag.*, pp. 4–22, Apr. 1987.

- [10] W. Lukasiewicz, "Non-monotonic reasoning: Formalization of commonsense reasoning," in *Ellis Horwood Series in Artificial Intelligence*, 1990.
- [11] F. Maire, "Rule-extraction by backpropagation of polyhedra," *Neural Networks*, vol. 12, pp. 717–725, 1999.
- [12] M. Minsky and S. Papert, *Perceptrons*. Cambridge, MA: MIT Press, 1969.
- [13] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing: Explanations in the Microstructure of Cognition*, D. Rumelhart and J. McClell, Eds. Cambridge, MA: MIT Press, 1986, vol. 1, Foundation, pp. 318–362.
- [14] R. Setiono, "Extracting M-of-N rules from trained neural networks," *IEEE Trans. Neural Networks*, vol. 11, pp. 512–519, Mar. 2000.
- [15] I. Taha and J. Ghosh, "Symbolic interpretation of artificial neural networks," *IEEE Trans. Knowledge Data Eng.*, vol. 11, pp. 448–463, 1999.
- [16] Y. 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.
- [17] H. Tsukimoto, "Extracting rules from trained neural networks," *IEEE Trans. Neural Networks*, vol. 11, pp. 377–389, Mar. 2000.
- [18] P. D. Wasserman, *Advanced Methods in Neural Computing*. New York: Van Nostrand Reinhold, 1993.
- [19] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 3, no. 8, pp. 338–353, 1965.
- [20] —, "Fuzzy logic and approximate reasoning," *Synthese*, vol. 30, pp. 407–428, 1975.
- [21] —, "The concept of a linguistic variable and its application to approximate reasoning," *Inform. Sci.*, pt. Part 1, vol. 8, pp. 199–249, 1975.
- [22] H. J. Zimmermann, *Fuzzy Set Theory and its Applications*, 2nd ed. Boston, MA: Kluwer, 1991.



Juan L. Castro received the M.S. degree in 1988 and the Ph.D. degree in 1991, both in mathematics, from the University of Granada, Granada, Spain. His doctoral dissertation was on logical models for artificial intelligence.

He is a member of Research Group in Intelligent Systems in the Department of Computer Science and Artificial Intelligence (DECSAI) at the University of Granada. His research interests include fuzzy logic, nonclassical logics, approximate reasoning, knowledge-based systems, neural networks, and related applications.

Dr. Castro serves as a reviewer for some international journals and conferences.



Carlos J. Mantas received the M.S. degree in 1996 and Ph.D. degree in 1999, both in computer science from University of Granada, Granada, Spain. His doctoral dissertation was on knowledge representation and constructive neural networks.

He is currently a Professor in Department of Computer Science and Artificial Intelligence (DECSAI) of the University of Granada. He is a member of Research Group in Intelligent Systems. His research interests include neural networks, data mining, and fuzzy systems.



José M. Benítez (M'98) received the M.S. degree in 1994 and the Ph.D. degree in 1998, both in computer science from University of Granada, Granada, Spain.

Currently, he is a tenured Professor at the Department of Computer Science and Artificial Intelligence (DECSAI) of the University of Granada. He is a member of Research Group in Intelligent Systems. He is a coordinator of the EUSFLAT Working Group on Neuro-Fuzzy Systems. His research interests include neural networks, fuzzy rule-based systems, neuro-fuzzy systems, cluster computing,

and e-commerce.

Dr. Benítez is a member of the IEEE Computer Society, ACM, and EUSFLAT.