Generating interpretable fuzzy systems for classification problems

Juan A. Contreras Montes[1]

Oscar S. Acuña Camacho[2]

Abstract.

This paper presents a new method to generate interpretable fuzzy systems from training data to deal with classification problems. The antecedent partition uses triangular sets with 0.5 interpolations avoiding the presence of complex overlapping that happens in another method. Singleton consequents are generated form the projection of the modal values of each triangular membership function into the output space. Least square method is used to adjust the consequents. The proposed method gets a higher average classification accuracy rate than the existing methods with a reduced number of rules and parameters and without sacrificing the fuzzy system interpretability. The proposed approach is applied to two classical classification problems: Iris data and the Wisconsin Breast Cancer classification problem.

1   PhD in Technical Sciences from CUJAE, La Habana, Cuba. He graduated as an Electric Engineer in 1987 at the Universidad Tecnológica de Bolívar and as specialist in Industrial Automation in 1998 at the same university. He is working for the Department of Naval Engineer at Navy School in Cartagena, Colombia. epcontrerasj@ieee.org.

2   MSc from UNAB, He graduated as specialist in Industrial Automation from Coruniversitaria, and as Electrical Engineer from UIS. He is working for the Department of Electrial Engineer at Universidad Tecnologica de Bolivar in Cartagena, Colombia. oacuna@unitecnologica.edu.co.

Keywords

Fuzzy systems, Interpretability, classification problems.

Resumen

En este artículo se presenta un nuevo método para generar sistemas difusos interpretables, a partir de datos experimentales de entrada y salida, para resolver problemas de clasificación. En la partición antecedente se emplean conjuntos triangulares con interpolación de 0.5 lo cual evita la presencia de solapamientos complejos que suceden en otros métodos. Los consecuentes, tipo Singleton, son generados por la proyección de los valores modales de cada función de membresía triangular en el espacio de salida y se emplea el método de mínimos cuadrados para el ajuste de los consecuentes. El método propuesto consigue una mayor precisión que la alcanzada con los métodos actuales existentes, empleando un número reducido de reglas y parámetros y sin sacrificar la interpretabilidad del modelo difuso. El enfoque propuesto es aplicado a dos problemas clásicos de clasificación: el Wisconsin Breast Cancer (WBC) y el Iris Data Classification Problem, para mostrar las ventajas del método y comparar los resultados con los alcanzados por otros investigadores.

Palabras claves

Sistemas difusos, interpretabilidad, problemas de clasificación

1. INTRODUCTION

One of the first proposals to automatically design a fuzzy system from data was the table look-up scheme (Wang & Mendel, 1992). After that, was proposed a methodology to identify fuzzy model parameters using singleton consequents, but many rules are required and gives a poor description capacity (Sugeno & Yasukawa, 1993).

Different approaches have been proposed to generate fuzzy models from input-output data. Fuzzy-clustering algorithms perform as the most adequate technique to obtain fuzzy models, being fuzzy C-Means (Bezdek, 1987) and (Gustafson & Kessel, 1979) the most used methods. Modifications to these clustering algorithms have been proposed but they typically seek for a good accuracy while interpretability of the fuzzy model is not their first concern (Nauck & Kruse, 1995; 1999). In fuzzy systems and fuzzy classifiers, it is necessary that the resulting fuzzy models have some transparency, i.e., that their information be interpretable, so as to permit a deeper understanding of the system under study (Paiva & Durado, 2004).

Interpretability is defined for at least five criteria (Paiva & Durado, 2004), (Espinosa & Vandewalle, 2000):

a. Distinguishability. The membership functions should be clearly different and each linguistic label should have semantic meaning.

b. Any element from the universe of discourse should belong to at least one of the fuzzy sets.

c. Due to the fact that each linguistic label has semantic meaning, at least one of the values in the universe of discourse should have a membership degree equal to one. In other words, all the fuzzy sets should be normal.

d. The numbers of functions membership should no exceed the limit of 9 distinct terms.

e. The number of rules should be limited according human cognitive issues.

Figure 1 shows three membership functions where a strong overlapping occurs and, because that, it is very difficult to label these membership functions.
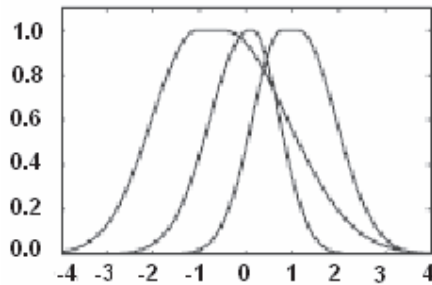
FIGURE 1. No INTERPRETABLE DISTRIBUTION OF MEMBERSHIP FUNCTIONS

This paper presents a new approach for the development of linguistically interpretable fuzzy models from data. The approach has been used in system identification but in this time it will be used as a fuzzy classifier (Contreras et al., 2008).

The methodology used in this paper to get the fuzzy model from input and output data is based on the inference error method from Sala (1998) and is presented in three phases: At the first, the inference error method is used to generate an interpretable fuzzy model and also, to detect possible classes or clusters in data; at the second phase, the consequent parameters are adjusted by means of least square method; at the third phase, the method is applied to two well known classification problems: Wisconsin breast cancer data and Iris data (Chen & Fang, 1995), (Hong & Lee, 1996), (Setiono & Liu, 1997), (Castro et al., 1999), (Hong & Chen, 1999), (Lee et al., 2001).

2. FUZZY IDENTIFICATION APPROACH

2.1 Inference error

a fuzzy rule: "if u is A, then y is B ", where u and y represent two numeric variables, and $A \subset U$ and $B \subset Y$, are two fuzzy input

and output sets respectively, defined at the universes U and Y, is equivalent to the equation:

$$u_A(u) \leq u_B(y) \qquad (1)$$

The inference error ε, is given by

$$\varepsilon = \begin{cases} 0 & \dots & u_A(u) \leq u_B(y \\ u_A(u) - u_B(y) & \dots & u_A(u) > u_B(y) \end{cases} \qquad (2)$$

A fuzzy rule of the kind "If u is A, then y is B" with a null inference error, must fulfill the condition

$$u_A(u) = u_B(y) \qquad (3)$$

If the system has n inputs, it must be represented by rules of the kind "If u_1 is A_1 and u_2 is A_2 and …and u_m is A_m, then y is B", and the generated system must fulfill the condition

$$((u_{A1}(x_k) \wedge (u_{A2}(x_k) \wedge \dots (u_{Am}(x_k)) = u_B(y_k) \qquad (4)$$

where $\wedge$ represents a t-norm, or an aggregation operator, of fuzzy logic

2.1 fuzzy model structure

2.2.1 Membership functions

The universe partitioning of the input variables in the learning process will be done with normalized triangular sets with specific overlapping of 0.5. The triangular membership functions allow the reconstruction of the linguistic value at the same numeric value after a defuzzyfication method has been applied (Pedriycz, 1994); also, the overlapping in 0.5 assures that the supports of the fuzzy sets are different. The fuzzy sets generated by the output variable will be a singleton

2.2.2 Distribution of the memberships Functions

The triangular fuzzy sets of input variables will be distributed symmetrically at each respective universe.

2.2.3 Operators

For combining the antecedents OWA operators will be used (Babuska, 2001).

2.2.4 Inference Methods

$$f(x^{(i)}) = \frac{\sum_{j=1}^{L} \bar{y}^{j}\, m_j(x^{(i)})}{\sum_{j=1}^{L} m_j(x^{(i)})} \qquad (5)$$

where

$$m_j(x^{(i)}) = u_{A1}(x_1^{(j)}),\ u_{A^{j}_2}(x_1^{(i)})..u_{A^{j}_n}\ (x_n^{(i)}) \qquad (6)$$

is the output grade of the j-est rule of a Sugeno fuzzy system, $\bar{y}^{j}$ is the singleton value corresponding to rule j.

2.3 Fuzzy Identification Algorithm

Given a collection of experimental input and output data $\{x_k,\ y_k\}, k=1,...,N$, where x_k is the n-dimensional input array $x^1_k,\ x^1_k,...\ x^n_k$ and y_k is the one-dimensional output array, the algorithm is defined by the following steps.

Step 1. Organization of the N pair set of input – output data $\{x^i_k,\ y^i\}$ with $i=1,...N; K=1,...p$, where $x_k^{(i)} \in \Re^p$ are input arrays and y^i are output scalars.

$$\begin{bmatrix} x_1^{(1)} \\ x_1^{(2)} \\ \vdots \\ x_1^{(N)} \end{bmatrix} \begin{bmatrix} x_2^{(1)} \\ x_2^{(2)} \\ \vdots \\ x_2^{(N)} \end{bmatrix} \cdots \begin{bmatrix} x_p^{(1)} \\ x_p^{(2)} \\ \vdots \\ x_p^{(N)} \end{bmatrix} \rightarrow \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(N)} \end{bmatrix}$$
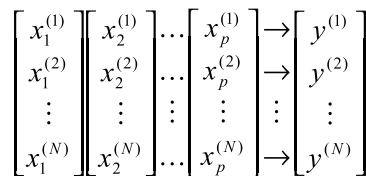
FIGURE 2. DATA ORGANIZATION

Step 2. Determination of universe ranges of each variable, according to maximum and minimum values of associated data $[x_i, x_i^+]$, $[y_i, y_i^+]$.

Step3. Distribution of triangular membership functions over each universe. As a general condition the vortex with ownership value one (modal value) falls at the middle of the region covered by the membership function while the other two vortexes, with membership values equal to zero, fall in the middle of the two neighboring regions
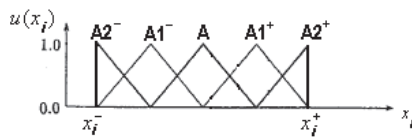

FIGURE 3. TRIANGULAR SUM-1 PARTITION

Step 4. Calculate the position of the modal values from the input variable(s), according to

$$if \quad u_{\underset{k}{(n)}}(x_k^{(i)}) = 1$$
$$ys_k^n = y[i] \qquad\qquad (7)$$
$$end$$

where corresponds to the projection over the output space of data evaluation of the k-th input variable at the n-th set of the corresponding partition. The output value corresponding to this projection is given by the value of the i-th position of output array y.
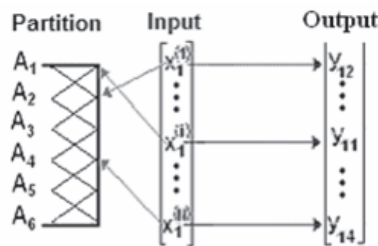

FIGURE 4. GENERATING CONSEQUENTS

Step 5. Rule determination. Initially, the rules number are equal to the number of sets of each input variable multiplied by the number of variables like $n*k$. The membership function associated to a consequent will be the antecedent of this rule. Antecedents of rules with the same consequent are merged by using OWA operator, reducing thus the number of rules.
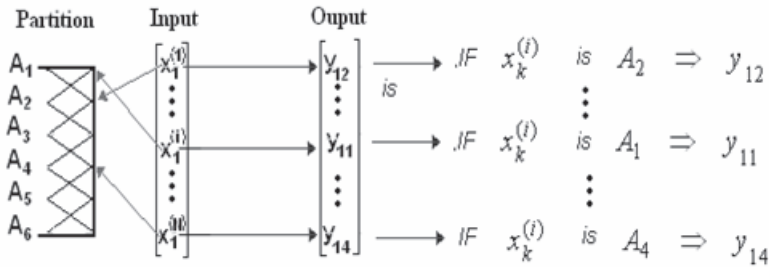

FIGURE 5. GENERATING RULES

Step 6. Model validation using the inference method described by (5).

Step 7. Parameters adjust, relocating the output singletons using the least square method (Contreras, 2006). Equation (5) can be written like:

$$f(x^{(i)}) = \sum_{j=1}^{L} \bar{y}^{j} w_j(x^{(i)})$$ (8)

Where

$$w_j(x^{(i)}) = \frac{m_j(x^{(i)})}{\sum_{j=1}^{L} m_j(x^{(i)})}$$ (9)

Output values can be represented as $Y=W\theta+E$ like equation 10.

$$
\underbrace{\begin{bmatrix} y^1 \\ y^2 \\ \vdots \\ y^L \end{bmatrix}}_{Y} = \underbrace{\begin{bmatrix} w_1^1 & w_2^1 & \ldots & w_L^1 \\ w_1^2 & w_2^2 & \ldots & w_L^1 \\ \vdots & \vdots & \ddots & \vdots \\ w_1^n & w_2^n & \ldots & w_1^1 \end{bmatrix}}_{W} \underbrace{\begin{bmatrix} \bar{y}^1 \\ \bar{y}^2 \\ \vdots \\ \bar{y}^L \end{bmatrix}}_{\theta} + \underbrace{\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}}_{E} \tag{10}
$$

where E is the approximation error to minimize. Using the quadratic error norm, we have:

$$
\theta = \frac{W^T Y}{W^T W} = (W^T W)^{-1} W^T Y \tag{11}
$$

If $W^T W$ is no singular, the least square estimation can be done recursively.

Step 8. Finally if either the square error measure MES is not greater than a measure previously established or the number of membership functions is more than nine. In any other case increment by 1 the number n of sets in the input variable (the number of partition member) and turn back to step 3.

3. RESULTS AND DISCUSSION

3.1 Iris Data Classification Problem

The Iris data is a common benchmark in the classification and the pattern recognition studies, consisting of feature measurements for the speciation of iris flowers. There are three classes corresponding to three species of iris: setosa, versicolor, and virginica, where each flower can be identified according to four continuous attributes measured in centimeters: (1) sepal length (SL); (2) sepal width (SW), (3) petal length (PL), and (4) petal width (PW). The 150 database entries include 50 sample cases for each of the three species. The goal is to classify irises into one of the three classes in accordance with the four inputs.

First the algorithm has to organize input and output variables and define the ranges of each input variable, thus:

Sepal length SL:	[4.3, 7.9]
Sepal width SW:	[2.0, 4.4]
Petal length PL:	[1.0, 6.9]
Petal width PW:	[0.1, 2.5]

After training, each input variable has three triangular memberships with overlap in 0.5 and fully distinguishable: Small (S), Medium (M) and Big (B). These sets are distributed uniformly on the universe of each input variable. The consequents are placed, before the adjust of least square method, in the following position: 1, 3 and 5, which represents "setosa" class, "versicolor" class and "virginica" class, respectively.

Figure 6 shows the relation of the membership function for each one of the input variables, X1-X4, to the setosa class, versicolor class and virginica class.

	X1	X2	X3	X4
S	1	3	1	1
M	3	1	3	3
B	5	1	5	5

FIGURE 6. RELATION OF THE MEMBERSHIP FUNCTION FOR EACH ONE OF THE INPUT VARIABLES TO THE SETOSA CLASS, VERSICOLOR CLASS AND VIRGINICA CLASS

From the Figure 6 we can extract 12 rules.

Rule1: IF X1 is Small, THEN "Setosa"

Rule2: IF X2 is Small, THEN "Versicolor"

Rule3: IF X3 is Small, THEN "Setosa"

Rule10: IF X2 is Big, THEN "Setosa"

Rule11: IF X3 is Big, THEN "Virginica"

Rule12: IF X4 is Big, THEN "Virginica"

The rules with common consequents can to merge in exclusively one rule unchanged the results subject to the antecedents were combined by an operator average type. In this manner the rule base is reduced only to three rules.

The Figure 7 shows the approximation of the output of the fuzzy model to the real dates. The system or fuzzy model rules based try to approximate the discontinuous function represented by the assigned values to the three classes: 1, 3 and 5.
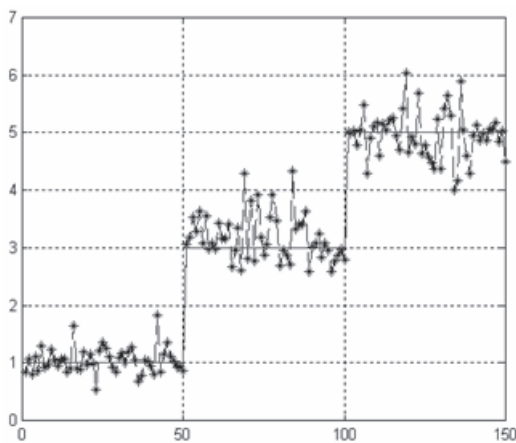

FIGURE 7. REAL SYSTEM OUTPUT (CONTINUOUS) VS FUZZY MODEL OUTPUT (*--)

Model output was obtained using equation (10); in this case error vector was not considered.

$$
\underbrace{\begin{bmatrix} y^1 \\ y^2 \\ \vdots \\ y^{150} \end{bmatrix}}_{Y} = \underbrace{\begin{bmatrix} u_s(X1)^1 & u_M(X1)^1 & \ldots & u_B(X4)^1 \\ u_s(X1)^2 & u_M(X1)^2 & \ldots & u_B(X4)^2 \\ \vdots & \vdots & \ddots & \vdots \\ u_s(X1)^{150} & u_M(X1)^{150} & \ldots & u_B(X1)^{150} \end{bmatrix}}_{W} \underbrace{\begin{bmatrix} \bar{y}^1 \\ \bar{y}^2 \\ \vdots \\ \bar{y}^{12} \end{bmatrix}}_{\theta}
$$

where Y is an array with the outputs of the fuzzy model; W is an array containing

$u_s(X1), u_M(X1), u_B(X1), u_s(X2), u_M(X2), u_B(X2), u_s(X3), u_M(X3), u_B(X3), u_s$ $(X4), u_M(X4)$, and u_B $(X4)$, and θ is the consequents vector (12 singleton consequents) after being adjusted by using LQM.

Next, a single algorithm is included for classification proposes.

If 0<y<2

Then class 2

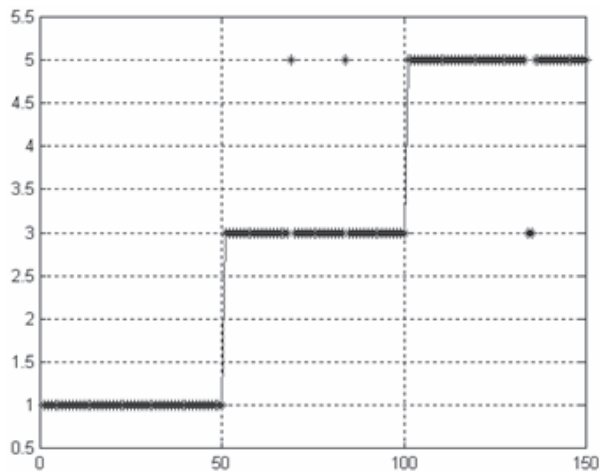Elseif 2<=y<4

Then class 3

Else

Then class 5

Result is shown in Figure 8.

FIGURE 8. REAL SYSTEM OUTPUT (CONTINUOUS) VS FUZZY MODEL OUTPUT (*)

In the Figure8 only 4 of 150 cases were not well classified. These results have a good performance, on the order of 97.33%. In this example were utilized 150 instances for the training. In the table 1 there is a comparison with the results obtained for other researcher by means of different values for training and validation instances.

TABLE 1. COMPARISON BETWEEN DIFFERENT METHODS OF CLASSIFICATION

Method	Precision (%)
Hong and Lee's method (training set: 75 instances; validation set: 75 instances)	95.57
Chen and Fang's method (training set: 75 instances; validation set: 75 instances)	96.28
Proposed method (training set: 75 instances; validation set: 75 instances)	96.67
Castro, Castro-Sanchez and Zurita's method (training set: 120 instances; validation set: 30 instances)	96.60
Chen and Fang method (training set: 120 instances; validation set: 30 instances)	96.72
Proposed method (training set: 120 instances; validation set: 30 instances)	96.67
Hong and Chen method (training set: 150 instances; validation set: 150 instances)	96.67
Chen and Fang method (training set: 150 instances; validation set: 150 instances)	97.33
Proposed method (training set: 150 instances; validation set: 150 instances)	97.33

3.1 The Wisconsin Breast Cancer (WBC)

The Wisconsin Breast Cancer Diagnostic Data set contains 699 instances and 16 of them are omitted because these are incomplete. The aim of the classification is to distinguish between benign and malignant cancers based on the 9 variables associated with each instance: X1 clump thickness, X2 uniformity of cell size, X3 uniformity of cell shape, X4 marginal adhesion., X5 single epithelial cell size., X6 bare nuclei, X7 bland chromatin, X8 normal nuclei, and X9 mitosis. After removing samples containing missing values, 683 instances are considered: 444 are in the benign class and 239 are in the malignant one.

The Fuzzy model obtained has three triangular membership functions for each input variable: Small (S), Medium (M) and Big (B). The singleton consequents are located, before the adjustment via least square method, on the values 2, that represents benign class, and 4, that represents malignant class. Figure 9 shows the relation between the membership functions of each one of the 9 input variables and the benign (2) and malignant (4) classes.

	X1	X2	X3	X4	X5	X6	X7	X8	X9
S	2	2	2	2	2	2	2	2	2
M	2	4	4	2	4	4	4	4	2
B	4	4	4	4	4	2	4	4	4

FIGURE 9. RELATION BETWEEN THE MEMBERSHIP FUNCTIONS OF EACH ONE OF THE 9 INPUT VARIABLES AND THE BENIGN (2) AND MALIGNANT (4) CLASSES

From the Figure 6 we can extract 27 rules.

Rule 1: IF X1 is Small (S), THEN benign class (2)
Rule 2: IF X1 is Medium (M), THEN benign class (2)
Rule 3: IF X1 is Big (B), THEN malign class (4)
Rule 4: IF X2 is Small (S), THEN benign class (2)

.
.
.

Rule 26: IF X9 is Medium (M), THEN benign class (2)
Rule 27: IF X9 is Big (B), THEN malign class (2)

However, rules with the same consequent can be reduced to one rule combining the antecedents of each rule using OWA operator without affect the precision of the system. In this case, we obtain a model with 2 rules.

Rule1: IF X1 is Small, X1 is Medium, X2 is Small X3 is Small,…, X8 is Small, X9 is Small and X9 is Medium, THEN benign class

Rule2: IF X1 is Big, X2 is Medium, X2 is Big X3 is Medium,…, X8 is Medium, X8 is Big and X9 is Big, THEN malign class

Table 2 shows a comparison with the results obtained by other authors.

TABLE 2. WBC COMPARISON WITH OTHER METHODS

Method	Accuracy %
(Setiono & Liu, 1997)	93.9
(Lee et al 2001)	94.7
(Wang and Lee , 2002)	96.3
Our approach	96.78

4. CONCLUSIONS

A new approach for the development of linguistically interpretable fuzzy models from data was developed in this paper. The fuzzy identification algorithm proposed uses triangular membership function with 0.5 interpolations for antecedent partition avoiding the presence of complex overlapping that happens in other methods. This approach does not require other techniques (neural network, genetic algorithm, etc.) for learning process.

The proposed approach is applied to Iris Data and the Wisconsin Breast Cancer classification problem, obtaining a high accuracy without sacrificing the fuzzy system interpretability. Results shown reveal that the proposed approach is an effective and promising classification tool.

REFERENCES

Babuska, R. (2001). Fuzzy and Neural Control. Disc Course Lecture Notes. Delft University of Technology. Delft, the Netherlands. 2001.

Bezdek, J. C. (1987). Pattern recognition with Fuzzy Objective Function Algorithms. Ed. Plenum Press.

Castro, J. L., Sánchez, J.J. & Zurita, J.M.(1999). "Learning maximal structure rules in fuzzy logic for knowledge acquisition in expert systems". Fuzzy Sets and Systems. 101(3), pp. 331-342.

Contreras, J., Misa, R. & Murillo, L. (2008). "Obtención de Modelos Borrosos Interpretables de Procesos Dinámicos". RIAI: Revista Iberoamericana de Automática e Informática Industrial, vol. 5, No. 3, pp. 70-77.

Contreras, J., (2006). "Introducción al Control Automático". Cartagena, Colombia. Editorial Escuela Naval Almirante Padilla.

Chen, Sh-M. & Fang, Y-D (2005). "A New Approach for Handling the Iris Data Classification Problem". International Journal of Applied Science and Engineering. 3(1), 37-49.

Espinosa, J & Vandewalle, J. (2000). Constructing Fuzzy Models with Linguistic Integrity from Numerical Data-Afreli Algorithm, IEEE Trans. Fuzzy Systems, vol. 8, pp. 591 – 600.

Guztafson, E. E. & Kessel, W. C. (1979). Fuzzy Clustering with a Fuzzy Covariance Matrix. IEEE CDC, San Diego, California, pp. 503 – 516.

Hong, T.P. & Chen, J.B. (1999). "Finding relevant attributes and membership functions". Fuzzy Sets and Systems. 103, 389-404.

Hong, T. P. & Lee, C.Y. (1996). Induction of fuzzy rules and membership functions from training examples. Fuzzy Sets and Systems, 84: 33-47.

Lee, H.M., Chen, C.M, Chen, J.M. & Jou, Y.L. (2001). "An efficient fuzzy classifier with feature selection based on fuzzy entropy", IEEE Trans. Syst. Man and Cybern, vol. 31, pp.426-432.

Nauck, D. & Kruse,R. (1995). «Nefclass - a neuro-fuzzy approach for the classification of data», In Proceedings of the Symposium on Applied Computing.

Nauck, D. &, Kruse, R. (1999). "Neuro-fuzzy systems for function approximation". Fuzzy Sets and System. 101(2), pp. 261-271.

Paiva, R. P. & Dourado, A. (2004). Interpretability and Learning in Neuro-Fuzzy Systems, Fuzzy Sets and System. 147, pp. 17-38. 2004.

Pedriycz, W. (1994). "Why Triangular Membership Functions?", IEEE Trans. Fuzzy Sets and System, vol. 64, pp.21-30.

Sala, A. (1998). Validación y Aproximación Funcional en Sistemas de Control Basados en Lógica Borrosa. Universidad Politécnica de Valencia. Tesis Doctoral.

Setiono, R. & Liu, H. (1997). "Neural Network feature selector", IEEE Trans. Neural Network, vol. 88, pp.654-662.

Sugeno, M. & Yasukawa, T. (1993). "A fuzzy logic based approach to qualitative modeling". Transactions on Fuzzy Systems, vol. 1, No. 1, pp. 7-31.

Wang, J. S. & Lee, C. S. G. (2002). "Self adaptive neuro-fuzzy inference system for classifications application", IEEE Trans. Fuzzy Systems, vol. 10, pp.790-802.

Wang, L.X. & Mendel, J.M. (1992). "Generating fuzzy rules by learning form examples", IEEE Transactions System, Man and Cybernetics, vol. 22, pp. 1414-1427,