An approach of fuzzy modeling towards intelligible modeling

CARLES GARRIGA-BERGA
Electronics Department - Enginyeria i Arquitectura La Salle
Ramon Llull University
Pg. Bonanova, 8 – 08022 Barcelona
SPAIN

Abstract: - In this paper we present a set of heuristic criteria devised to address the problems encountered in designing a fuzzy system to fit a set of input-output data. The objective is to obtain in a simple and fast manner a good starting model to undergo further refinements. The result is a simple algorithm with a similar performance than other techniques but with a low computational cost.

Key-Words: - fuzzy curves, fuzzy modeling, intelligible approach, clustering

1 Introduction

After Zadeh proposition of linguistic modeling in place of the quantitative one [1], fuzzy modeling has become one of the most relevant issues in qualitative analysis benefited from the linguistic capabilities of fuzzy logic. Since, several methods to obtain fuzzy models have appeared, most of them based in clustering techniques. Among the most popular ones we can mention Fuzzy C-means.

Many authors have focused their efforts to achieve models with a high precision in terms of error but have omitted the linguistic capabilities of fuzzy systems. For this reason most use linear equations as consequents of their rules (Takagi-Sugeno's model) instead of fuzzy quantities (Sugeno-Yasukawa's model). In general most methods fulfill the premise to diminish the error but unfortunately they are often excessively complex to be linguistically interpreted, which supposes to fail to take advantage of one of the main virtues of fuzzy systems.

Recall that the initial philosophy of fuzzy logic was to be the bridge between the human understanding and the machine processing. In this challenge, the ability of fuzzy models to express the behavior of real systems in a comprehensible manner acquires great importance. That's why our efforts have been devoted to discuss a new approach to fuzzy modeling able to obtain in a simple manner a good starting model to undergo further refinements based on qualitative analysis.

In this paper we present a set of heuristic criteria devised to address model estimation. In particular, we are able to determine the number of fuzzy sets, place them in the universe of scope and propose a set of linguistic rules. The result is an algorithm with low computational cost but still similar performance than others.

Our goal is not to seek the lowest possible error but to obtain an acceptable precision while keeping the linguistic capabilities of the fuzzy model. In fact with this algorithm we will be able to choose the precision of the model and consequently its degree of interpretability.

2 Outline of the method

2.1 Optimal fuzzy curve

Trying to adjust the fuzzy sets from the original data can be very arduous if many samples must be considered, so it is preferred to work with a simplified relation between the output variable and each possible input.

For this purpose many techniques are considered in the literature. The most popular among them are: linear regression, weighted average and Bézier curves. Among these alternatives we choose the weighted average because we are not interested in an extremely accurate result but in a solution with low computational cost showing the tendency adopted for the output variable when the input varies. So finally the proposed weighted function is the following one where the β parameter can be adjusted to diminish the square error as will be detailed later:

$$\hat{y}_i = \frac{\sum_{k=1}^{N} \phi_{ik} y_k}{\sum_{k=1}^{N} \phi_{ik}} \tag{1}$$

where

$$\phi_{ik} = \exp\left(-\left(\frac{x_k - x_i}{\beta} \right)^2 \right) \tag{2}$$

In fact this function was proposed by Lin *et al.* in [2] and was called fuzzy curve because of its similarities with a fuzzy system with N rules, one per sample, where ϕ_{ik} represents the antecedent set placed at x_k and evaluated when $x=x_i$.

Until now β has usually been adjusted empirically and in fact Lin *et al.* suggest a value equal to the 20% of the variable's range. But it is possible to work with an optimized fuzzy curve because the β of the fuzzy curve can be adjusted in such a way that this curve diminishes the square error defined as the difference between the value of the curve and the real value [3].

At this point it is necessary to clarify the apparently incoherence of diminishing the error when we have bet for simple but not accurate results. The fact is that when working with weighted average functions anyone must answer the following question: how many neighbor samples should I average for each point? Certainly the precision of the resulting function is not its most valuable characteristic but obviously if the final result displays a good performance in terms of error no one would reject it. The method we propose will search a weighted function with a low square error but without assuring it because we will just find a statistic of it and thus this optimized function will have a certain tolerance. This tolerance will be the trade-off between the desired precision and the required computational cost.

First of all it's obvious that we can not define the error from the sum of all the input samples because then the trivial solution for the parameter β would be equal to zero. Consequently we will have to divide the set of values in two groups: the samples that will be used to search the optimal parameter and the samples that will be used to evaluate the square error committed with this adjustment. Therefore we begin dividing the N samples in N_1 samples used to calculate the fuzzy curve and $N_2=N-N_1$ samples used to test the error. In this way the values of the fuzzy curve in each test point will be computed using only the N_1 values with the following equation:

$$\hat{y}_i = \frac{\sum_{k=1}^{N_1} \phi_{ik} y_k}{\sum_{k=1}^{N_1} \phi_{ik}} \qquad (3)$$

Once the fuzzy curve has been computed, the error in each one of the N_2 test points is $\varepsilon_i=y_i-\hat{y}_i$ and the global square error is:

$$\varepsilon = \tfrac{1}{2}\sum_{i=1}^{N_2} \varepsilon_i^2 \qquad (4)$$

As being interested in the value of β which is necessary to diminish this error we come to calculate the equation:

$$\frac{\partial \varepsilon}{\partial \beta} = \sum_{i=1}^{N_2} \left(y_i - \hat{y}_i\right) \frac{\partial \hat{y}_i}{\partial \beta} = 0 \qquad (5)$$

As it is observed the previous equation does not present a trivial solution reason why it must be solved by numerical methods. In order to accelerate the process, it would be interesting to know the limits between which the parameter β can be optimal one.

Therefore, if β is very small then $\hat{y}_i \approx y_k$ for $|x_i-x_k|_{min}$ and only the closest train point to each test point affects the computation of the fuzzy curve. To determine this value we have to analyze from which value the effect of the second train point closest to each test point is insignificant. In this case if d_1 is the distance from a test point to its closest train point and d_2 is the distance from the same test point to the second train point closest to it, the most critical situation will be given for the test point having the two train points with the minimum value for d_2-d_1.

There is also a big value of β from which the error tends to become stabilized because then $\hat{y}_i$ is similar to the mean of all the train points. In this case if d_1 is the distance from a test point to its closest train point and d_∞ is the farthest point to it, the most critical situation will result for the test point having the maximum value for $d_\infty-d_1$.

Therefore and considering an accepted error $\eta \approx 0$, optimal β must be between [3]:

$$\sqrt{-\frac{d_2^2 - d_1^2}{\ln \eta}} < \beta < \sqrt{-\frac{d_\infty^2 - d_1^2}{\ln(1-\eta)}} \qquad (6)$$

Once we have the values between which the optimal β can be found, we can use a numerical method to determine it. But by the fact of dividing the samples in two groups, the optimal value obtained by each division can not be the optimal one for the whole samples, reason why we can not obtain the optimal value but a statistic of it.

So in an iterative way we will create different partitions of the samples from which we will obtain, for each partition, an optimal value of β_i with which we will compute, according to the theorem of the central limit, a confidence interval for this parameter as:

$$\overline{\beta} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}} \qquad (7)$$

We will stop the process after n iterations when the confidence interval allows us to consider the optimal value of β with a lower error than an established level. According to our objectives this level should not be very low because we do not seek very accurate results. Thus we will obtain fuzzy curves with an acceptable error but without increasing our method's computational cost.

2.2 Piecewise linear segmentation

Next step consists in obtaining a fuzzy logic representation of the optimal fuzzy curve with a satisfactory error. The fact of working with a one-dimensional function instead of the original cloud of samples will simplify this step. For this purpose a piecewise linearization of the fuzzy curve is suggested because its mapping to the final fuzzy sets is very simple and any error can be reached based on the number of segments. In general the more segments the less error but the more linguistic complexity. In fact any linear segment can be obtained with a fuzzy system without error so the overall error will be due to the linearization of the curve but not to the fuzzy approximation.

So we suggest the following procedure which is exampled in Figure 1 to clarify it:

- A straight line is plotted to join the first and the last point of the fuzzy curve. In fact these points will fix the boundaries of the universe of the scope and also the extreme fuzzy sets.
- The farthest point of the fuzzy curve to the linear approximation is chosen to define the next linearization and thus the current maximum error is eliminated. So a straight line is plotted to join this point to the next point of the scope used to define the previous linear approximation while another straight line is plotted to join the selected point to the previous point of the scope used in the previous linear approximation. Fuzzy sets are modified to implement this linearization.
- Process is repeated until a certain precision is achieved.

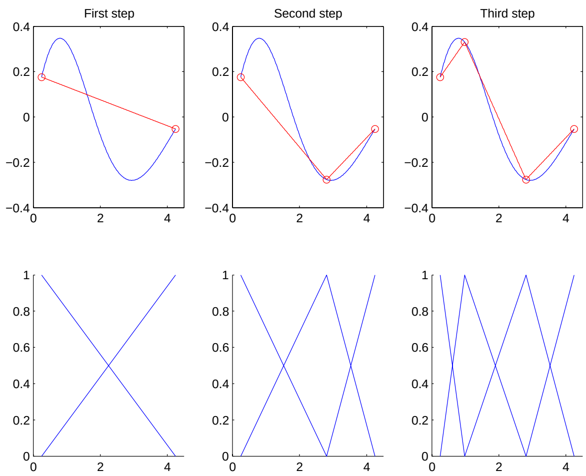

Fig.1 An example of linearization

The error will not always decrease monotonically but it can be easily concluded that if the fuzzy curve is defined with N points, the former method will remove the error in N-1 iterations at the most.

2.3 Wang-Mendel's modified algorithm

After placing all the fuzzy sets and consequently fixing the maximum number of rules as the product of the number of sets of each variable, it's only necessary to define the linguistic rules computing the output set for each possible rule.

This problem is commonly solved searching the value that best fits from the set of samples. This method was proposed by Wang and Mendel [4]. For each rule they chose from a predefined bag of fuzzy sets the one with its core closest to the output value of the sample with the maximum implication level.

We propose a similar version of this algorithm which basically consists on using the implication value of each sample instead of its distance, like Wang and Mendel in [4], but without the predefined bag of output sets. For instance suppose that we must assign the output set of a rule plotted in Figure 2 from the samples given in Table 1.

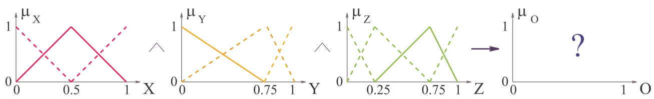
Fig.2 Example of adjusting the output set

Sample	x	y	z	o
1	0.30	0.25	0.50	0.40
2	0.80	0.50	0.75	0.80
3	0.50	0.00	0.00	0.20
4	0.50	0.25	0.75	0.50

Table 1 Example of possible samples

If the product is used to compute the T-norms:

Sample 1 $\rightarrow \mu_x(0.30)\times\mu_y(0.25)\times\mu_z(0.50)=0.20$

Sample 2 $\rightarrow \mu_x(0.80)\times\mu_y(0.50)\times\mu_z(0.75)=0.13$

Sample 3 $\rightarrow \mu_x(0.50)\times\mu_y(0.00)\times\mu_z(0.00)=0.00$

Sample 4 $\rightarrow \mu_x(0.50)\times\mu_y(0.25)\times\mu_z(0.75)=0.66$

Then the output set will be centered at 0.50 for being the output value of the fourth sample which is the one with the highest implication value.

This solution, apart from avoiding the frequent need of using relative distances if the magnitudes of the variables are very different, can find out the rules that can be omitted for not having any sample referred to it. This would happen if the implication value of all the samples was zero. In this way the rule matrix can be simplified and at the end only the rules suited with the set of samples will be held.

For example suppose we have placed the input sets for a certain set of samples plotted in Figure 3 and also the output sets. The input values of the different samples are also plotted over the rule matrix as a shadow. As there are two rules without any sample referred to it, they will be removed from the rule matrix, thus there will be only 40 rules instead of the maximum of 42.

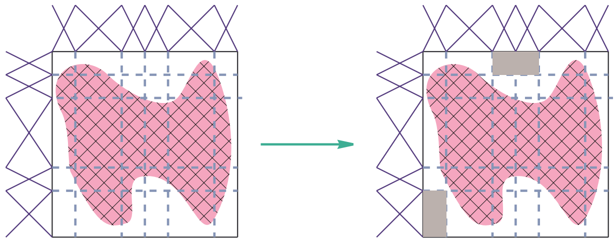

Fig.3 A possible reduction of rules

3 Examples

In this section we deal with three illustrative examples in order to validate the proposed algorithm. All of them have been used previously in many articles and can be considered as a benchmark.

3.1 Nonlinear static function

Here we consider the following double-input single-output function proposed in [5]:

$$y = \left(1 + x_1^{-2} + x_2^{-1.5}\right)^2 \quad 1 \le x_1, x_2 \le 5 \tag{8}$$

plotted in Figure 4 from which 50 samples are given.

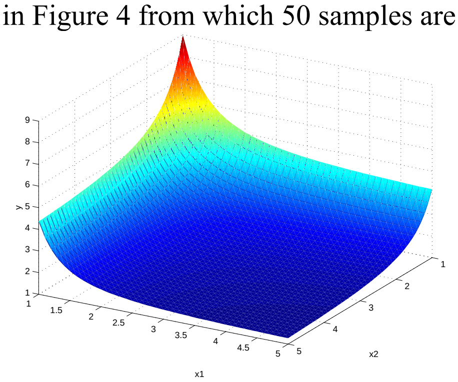

Fig.4 Nonlinear static function

The optimal β parameters for the fuzzy curves that we have computed from the previous data are 0.3477 for x_1 and 0.3779 for x_2. In fact we do not assure the optimal value but a probabilistic distribution of it. For both variables this statistic has been computed with a level of confidence of 95% and a relative error of 10%. The statistic for x_1 is plotted in Figure 5.

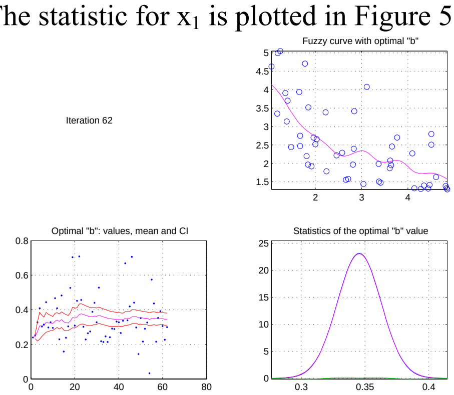

Fig.5 Statistics for the β parameter

Once β is computed and thus the fuzzy curve is defined, we linearize it until a certain error is accomplished. This process for the variable x_1 is shown in Figure 6 where a low error of only 1% has been demanded.

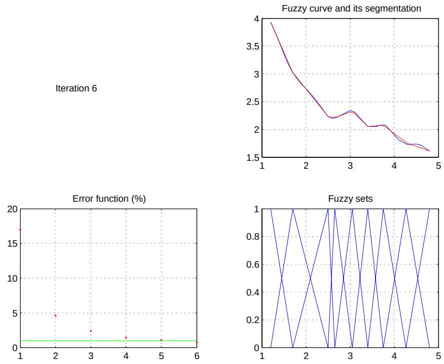

Fig.6 Linearization of a fuzzy curve

The previous low error is obviously desired but then a high number of segments are necessary and thus a high number of sets. Considering for example a precision of p=15% and the values for the three variables rounded to 0.1 we would obtain only the sets plotted in Figure 7.

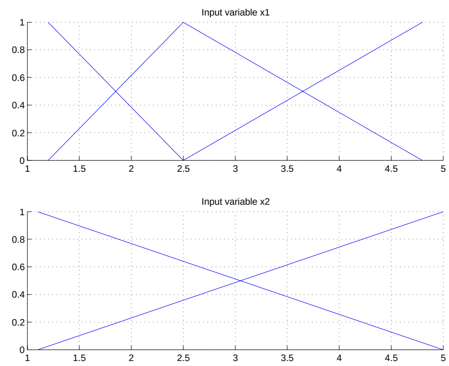

Fig.7 Sets with a precision=15%

Finally the output set for each possible combination of inputs is computed with the proposed modified Wang-Mendel's algorithm. Considering the former sets, the final rules with our method are:

if x_1 is $X_{1.2}$ and x_2 is $X_{1.1}$ then y is $Y_{5.0}$
if x_1 is $X_{1.2}$ and x_2 is $X_{5.0}$ then y is $Y_{3.4}$
if x_1 is $X_{2.5}$ and x_2 is $X_{1.1}$ then y is $Y_{3.4}$
if x_1 is $X_{2.5}$ and x_2 is $X_{5.0}$ then y is $Y_{1.6}$
if x_1 is $X_{4.8}$ and x_2 is $X_{1.1}$ then y is $Y_{2.8}$
if x_1 is $X_{4.8}$ and x_2 is $X_{5.0}$ then y is $Y_{1.3}$

where the subscripts of the fuzzy sets show their core.

In this case the root mean square error (RMSE) obtained is RMSE=0.644.

Obviously a better result can be achieved in terms of error if the precision parameter is reduced but with a higher number of sets and consequently with a poorer linguistic interpretability. The errors obtained in different simulations can be compared with other methods in Table 2 where, apart from the error, the number of sets, the number of rules and the type of the fuzzy model (Takagi-Sugeno or Sugeno-Yasukawa) are given to compare their linguistic capabilities.

Method	Type	Sets	Rules	RMSE
Sugeno [5]	S-Y	12 in + 6 out	6	0.564
Delgado [6]	S-Y	10 in + 5 out	25	0.493
Our method p=20%	S-Y	4 in + 4 out	4	1.024
Our method p=15%	S-Y	5 in + 5 out	6	0.644
Our method p=5%	S-Y	9 in + 14 out	18	0.427
Our method p=2%	S-Y	12 in + 20 out	36	0.253
Kim [7]	T-S	6 in + 3 eqs.	3	0.281

Table 2 Comparisons between alternatives

3.2 Fuzzy system model

With this case we will study the reproduction of the proposed method like in [5]. So we will consider 100 samples from the fuzzy system given in Figure 8 which was already used in [5] and whose transfer function is plotted in Figure 9.

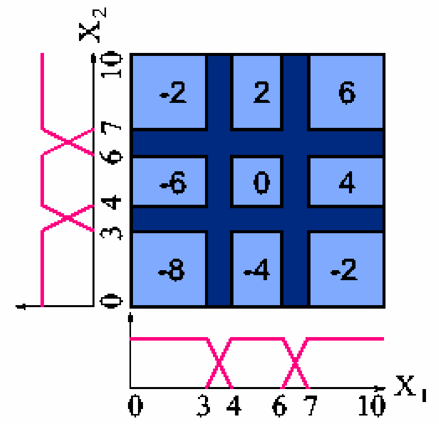

Fig.8 Original fuzzy system

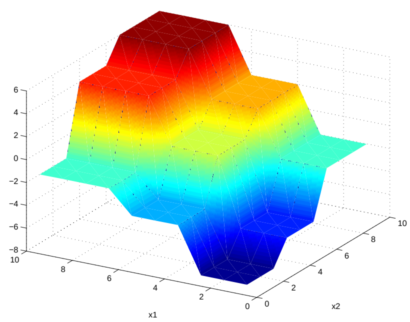

Fig.9 Original transfer function

If we consider 10 equidistant samples per variable from 1 to 10, the resulting optimal β parameters are 0.8529 for x_1 and 0.9903 for x_2.

Like in the previous example, we can consider different values to round the numbers and also the desired precision. In most cases the resulting model is very similar to the original one and in many cases is exactly the same if, for example, the round values are equal to one and the precision is less than 10%.

The only difference of our model in comparison with the original one is the fact that we only work with triangular sets and consequently we need six sets per variable placed at 1, 3, 4, 6, 7 and 10 instead of only three trapezoidal sets and thus the resulting rule matrix is the one given in Table 3. Obviously both fuzzy systems give exactly the same output values.

$x_1 \setminus x_2$	1	3	4	6	7	10
1	-8	-8	-6	-6	-2	-2
3	-8	-8	-6	-6	-2	-2
4	-4	-4	0	0	2	2
6	-4	-4	0	0	2	2
7	-2	-2	4	4	6	6
10	-2	-2	4	4	6	6

Table 3 Resulting model's fuzzy rules

In fact it is very easy to work with trapezoidal sets with our method. Once the rule matrix with triangular sets is obtained we search, for each input set, if the output set assigned to its neighbor rule with the same input sets is the same, apart from the one belonging to the current variable. If this situation is repeated for all the rules with the same input values then both sets can be grouped into a trapezoidal one. If this algorithm is applied to the previous rule matrix then we obtain exactly the original fuzzy system.

If 100 random samples are considered we obtain a good performance in most runs as can be showed in Figure 10.

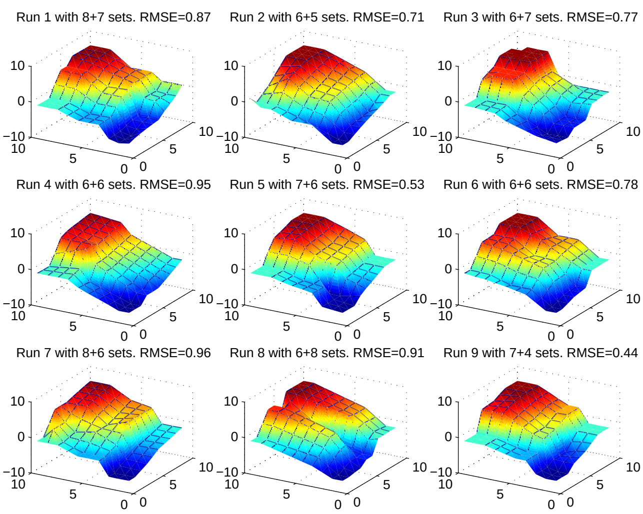

Fig.10 Random samples. Precision=5%.

3.3 Mackey-Glass chaotic time series

Here we will study the capabilities of the proposed method to model the Mackey-Glass chaotic time series defined with the following delay differential equation 9 under the premise that $\tau > 17$.

$$\frac{dx(t)}{dt} = \frac{0.2x(t-\tau)}{1+x^{10}(t-\tau)} - 0.1x(t) \qquad (9)$$

Higher values of τ yield higher dimensional chaos. In our simulation we will consider $\tau = 30$ like in [4], where Wang and Mendel evaluated their method for adjusting fuzzy rules. To study our algorithm we have used 991 samples with a sampling period of 1.1 seconds.

If we choose two input variables y(k-1) and y(k-2) and asking for a precision=0.5% we have obtained a model with 17 sets for y(k-1), 9 sets for y(k-2), 43 sets for y(k) and 57 rules whose RMSE=0.0333. Results are plotted in Figure 11.

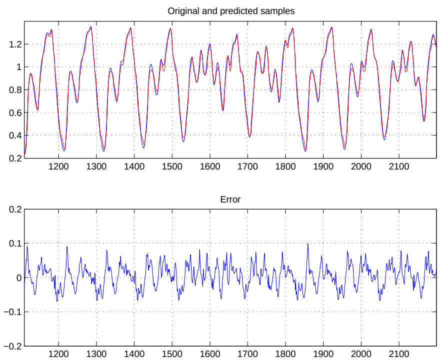

Fig.11 Prediction of chaotic series

We could use again this model to predict samples different from those used when constructing the model, achieving in this case a RMSE=0.0327. This situation is plotted in Figure 12.

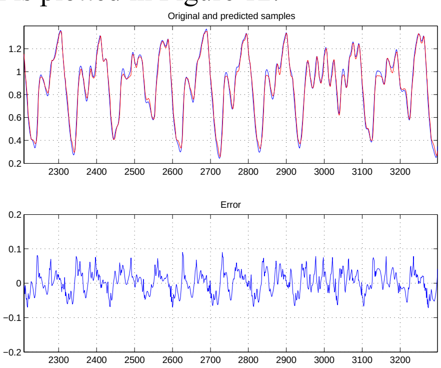

Fig.12 Prediction of chaotic series – Test samples

4 Conclusions

An algorithm that is able to give a first-approach of the fuzzy systems and that relates input-output pairs has been presented and discussed. Its main feature is that it is quite simple and extracts relevant linguistic information in a fast manner.

In spite of the surprising results that this and other identification algorithms can offer, the most important thing one must bear in mind when trying to identify a system, is the necessity of a good set of samples of all the variables because they will be at the end in charge of the result.

We are currently working on applications where linguistic interpretation is required such as econometrics, social analysis and scientific studies. Other applications related with on-line control processes are being considered.

Acknowledgements:

Author would like to thank Dr. Xavier Vilasís for his support on this work.

References:

[1] L.A. Zadeh, Outline of a new approach to the analysis of complex systems and decision processes, *IEEE Trans. on system, man and cybernetics*, vol. 3, 1973, pp. 28-44.

[2] Y. Lin and G.A. Cunningham III, A new approach to fuzzy-neural system modeling, *IEEE Trans. on fuzzy systems*, vol. 3, num. 2, 1995, pp. 190-197.

[3] C. Garriga-Berga, Min-square error optimization of fuzzy curves, *International Fuzzy Systems Association World Congress*, Istanbul, 2003

[4] L.X. Wang and J.M. Mendel, Generating fuzzy rules by learning from examples, *IEEE Trans. on system, man and cybernetics*, vol. 22, num. 6, 1992, pp. 1414-1427.

[5] M. Sugeno and T. Yasukawa, A fuzzy-logic-based approach to qualitative modeling, *IEEE Trans. on fuzzy systems*, vol. 1, num. 1, 1993, pp. 7-31

[6] M. Delgado, F. Gómez-Skarmeta and F. Martín. A fuzzy clustering-based rapid prototyping for fuzzy rule-based modeling. *IEEE Trans. On fuzzy systems*, vol. 5, num. 2, 1997, pp. 223-232.

[7] E. Kim, M. Park and S. Ji and M. Park. A new approach to fuzzy modeling. *IEEE Trans. on fuzzy systems*, vol. 5, num. 3, 1997, pp. 328-337.