A coevolutionary genetic algorithm able to learn a FRBS and detect the less informative instances in low quality problems

Luciano Sánchez

[1] Universidad de Oviedo. Departamento de Informática. luciano@uniovi.es

Abstract

The evaluation of the fitness function in genetic fuzzy systems with low quality data is a costly process, and this conditions the applicability of this technique to large datasets. Nevertheless, we have observed that, in many practical cases, a few instances are responsible for the main part of the dispersion of the fitness function.

In this paper we propose a new technique for coevolving a FRBS and a list containing the less informative instances in the dataset. We guide the learning with an approximate fitness, that combines a crisp component with an estimation of the dispersion of the fitness value, obtained from the mentioned subset of instances.

1 Introduction

Obtaining Fuzzy Rule Based Systems (FRBS) from low quality data with Genetic Algorithms (GA) differs from the crisp case [2] on the use of a fuzzy fitness function [16], and on an order structure defined in the set of images of this function, that allows us to compare two FRBSs [9, 11, 18]. Nonstandard GAs are also used; a modified scalar genetic algorithm can find the primal element of this order, or an adapted multicriteria algorithm can retrieve a set of nondominated individuals [8, 17].

At the root of this search, we need to know the error of each candidate FRBS in the population, for any selection of the input data which is compatible with our set of imprecise measurements [16]. If our dataset is interval-valued, we need to obtain the extrema of the error of the model for each instance, when their inputs are box-constrained to the corresponding interval-valued data. The interval arithmetic-based sum of these couples of bounds is regarded as an imprecise perception of the quality of the FRBS at this instance [3]. Else, if our data is fuzzy, these bounds must be produced for each α-cut of the data, and fuzzy arithmetic is used to compute the fuzzy fitness of the FRBS.

There are some difficulties that limit the applicability of this schema. Obtaining the bounds of the error of a model in an interval is a costly computational problem unless these extrema are known to be in the boundary of the region [15]. Since, generally speaking, the error of the FRBS will not be monotonic, it is expected that this stage will consume a large part of the learning time, precluding the application of this kind of algorithms to large datasets. That is the reason why, in this paper, we try to overcome this limitation by exploiting certain characteristics of the fitness function.

Let us introduce our idea with an example. In Figure 1 we have plotted the decision surface of a classifier, whose error we want to evaluate. The squares represent imprecise measurements of objects; that is to say, the actual measurements could be anywhere inside each square. Observe that, in the worst case, the error of this classifier can be as high as 3/15. In the best case it is 0. Observe also that the vagueness of the error of the classifier in the figure depends only on the three instances that have been marked: from a computational point of view, there is no need for obtaining the bounds of the value of the decision function at the 12 intervals that do not intersect the decision surface. We might as well compute the decision function at the centerpoint of each square except at the three "conflicting" instances, where we must launch that costly optimization algorithm we have just mentioned.

Let us introduce now the regression case with a second example. Observe the graph in Figure 2. In this case, all instances (A, B and C) would contribute to the vagueness of the fitness. But we can replace the output of the model in A and C by the output in the midpoints

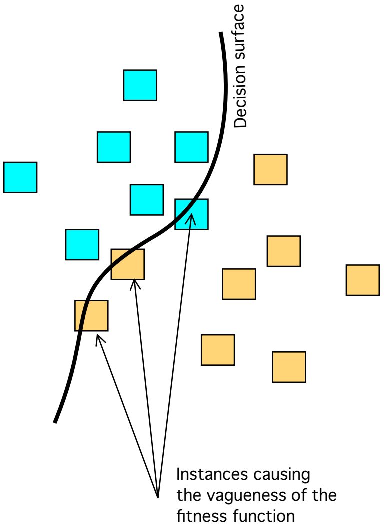


Figure 1: The vagueness in the estimation of the classification error is caused by the three marked elements.

of these intervals, without introducing large errors. In this case, we say that the less informative instance is B.

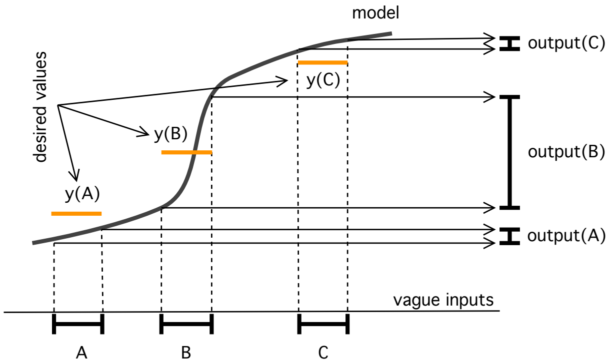


Figure 2: Most of the vagueness in the estimation of the regression error will depend on the instance 'B'.

The rationale of this work is, therefore: when learning FRBSs from low quality data, we can save most of the computing time if we replace each imprecise instance by a crisp value in all but the less informative instances. If we identify these points then we can divide the fitness computation into two parts: fuzzy (or interval) values at these instances, and crisp values at the remaining part of the dataset.

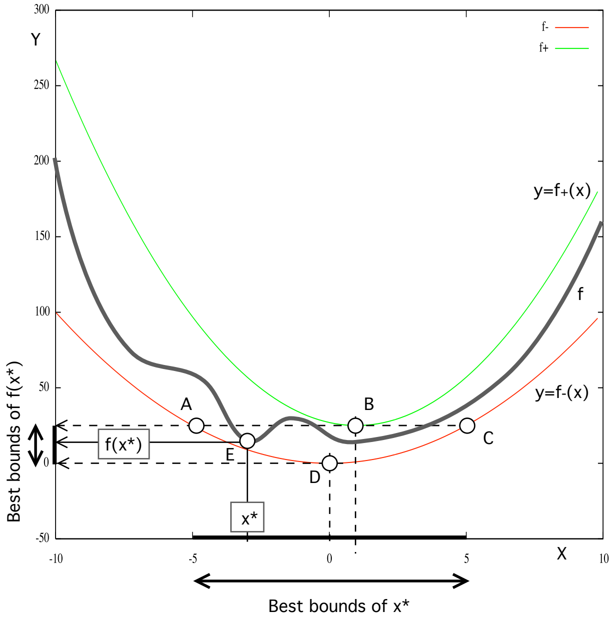


Figure 3: Interval optimization: $f^{-}(x) \leq f(x) \leq f^{+}(x)$. f^{-} and f^{+} are known, but f is not. Hence, the minimum of f cannot be known; at the most we can bound the values of x and $f(x)$ at the minimum.

The remaining part of the paper is devoted to define a GA that coevolves a population of FRBSs and the mentioned set of instances (Section 2), thus solving at the same time two problems: the learning of a FRBS from vague data, and the detection of the instances of data that contribute the most to the variability of the fitness function. In Section 3 we provide compared numerical results. The paper finishes with the concluding remarks in Section 4.

2 A proposal of coevolutionary learning of FRBS from vague data

The objective of this work is to learn a FRBS from a dataset whose input values are imprecise. In the following, we will use an interval valued dataset $\{(A_1, y_1), (A_2, y_2), \ldots, (A_n, y_n)\}$, $A_i \subset \mathbf{R}^n$. In turn, we will assume that a FRBS is completely defined by a vector of real parameters x. We will denote the output of the FRBS, given a crisp input t, by $F_x(t)$, with $F_x : \mathbf{R}^n \to \mathbf{R}$. The error of the FRBS in the mentioned dataset will be denoted by $f(x)$, and this function is unknown, but we are given a couple of functions f^{-} and f^{+} (see Figure 3),

$$f^{-}(x) = \sum_{i=1\ldots N} \min_{t \in A_i} ||F_x(t) - y_i||$$

$$f^{+}(x) = \sum_{i=1\ldots N} \max_{t \in A_i} ||F_x(t) - y_i||$$

where $f^{-}(x) \leq f(x) \leq f^{+}(x)$ for all x.

The objective of the learning is obtaining the parameter vector x^* that minimizes the error. Observe in

Figure 3 that the most we can know about the value of the objective function in the minimum, $f(x^*)$, is that it is in the segment we have labelled "Best bounds of $f(x^*)$". In turn, x^* is in the area marked "Best bounds of x^*". Observe also that the actual value of x^* is not related to the minimum of f^- neither f^+: we just know that it cannot be out of the mentioned region.

Alternatively, this can be expressed by defining a partial order between fitness values as follows:

$$[a,b] \prec [c,d] \iff b \leq c. \tag{1}$$

If the fitness of the individual x is the interval $[f^-(x), f^+(x)]$ then the best bounds of x^* coincide with the set of all nondominated values of x under that order. This alternative expression is the basis of the use of multicriteria genetic algorithms for learning FRBS with low quality data; those algorithms produce sets of individuals contained in the "Best bounds of x^*" [17].

Let us introduce now the approximation mentioned in the introduction: let $\mathcal{A}$ be the subset of $\{A_1, \ldots, A_n\}$ containing the less informative input intervals. In this paper we want to approximate the fitness function $[f^-, f^+]$ by the pair of functions

$$\begin{aligned} g^-(x) &= \textstyle\sum_{A_i \in \mathcal{A}^-} \min_{t \in A_i} ||F_x(t) - y_i|| \\ &+ \textstyle\sum_{A_i \notin \mathcal{A}^-} ||F_x(\mathrm{mid}(A_i)) - y_i|| \end{aligned} \tag{2}$$

$$\begin{aligned} g^+(x) &= \textstyle\sum_{A_i \in \mathcal{A}^+} \max_{t \in A_i} ||F_x(t) - y_i|| \\ &+ \textstyle\sum_{A_i \notin \mathcal{A}^+} ||F_x(\mathrm{mid}(A_i)) - y_i|| \end{aligned} \tag{3}$$

where $\mathcal{A} = \mathcal{A}^- \cup \mathcal{A}^+$ and $\mathrm{mid}(A)$ is the center of gravity of the interval A.

We propose to obtain the three unknowns ($\mathcal{A}^-$, $\mathcal{A}^+$ and also the best FRBS) at the same time, with a coevolutionay algorithm based on three populations. The first population contains different model candidates (each individual will represent a model, Pitts style [7]), and the other two codify the sets $\mathcal{A}^-$ and $\mathcal{A}^+$, respectively. In these two last populations, each individual represents one point in the sample, and the whole population is the solution (cooperative approach [6]). In the sections that follow we describe this algorithm in detail, detailing the representation of an individual, the fitness function, the coevolutionary scheme and the genetic operators.

2.1 Representation of an individual

For representing the individuals in the first population we will use the same representation that was defined in [18], that we will not repeat here because of space reasons: DNF rules and a single chromosome with binary codification for representing the linguistic terms and real coding of the weights. Observe that will not evolve the membership functions, but assign weights to the rules instead [5].

The second population codify as a whole the set $\mathcal{A}^-$ and also the points of these intervals where the error of the FRBS is minimum. An individual is a pair comprising the index s of an interval, $A_s \in \mathcal{A}^-$, and a point $t_s \in A_s$. We have coded this pair with an integer and a real-coded number (s, δ), where $\delta(t_s) = (t_s - t_s^-)/(t_s^+ - t_s^-)$.

For instance, if we are given a dataset formed by two imprecise tuples $\{(A_1 = [0,3] \times [1,2] \times [3,4], y_1 = 1), (A_2 = [3,4] \times [1,1] \times [3,3], y_2 = 2)\}$, the list $\{1, (0.5, 1, 0.25)\}$ is a valid individual, and it represents a point $(1.5, 2, 3.25) \in [0,3] \times [1,2] \times [3,4]$. The presence of this individual in the second population means that $A_1 \in \mathcal{A}^-$. The codification of the third population is similar.

2.2 Fitness function

The fitness of an individual x in the first population is the interval $[g^-(x), g^+(x)]$, defined in eqs. (2) and (3). For regression problems,

$$||F_x(t) - y|| = (F_x(t) - y)^2 \tag{4}$$

and for classification problems

$$||F_x(t) - y|| = \begin{cases} 1 & \text{if } FRBS(t) = y_i \\ 0 & \text{otherwise .} \end{cases} \tag{5}$$

The fitness of an individual in the first population depends on the other two populations, as follows. Let M be the size of populations 2 and 3, let s_j^- be the first part of j-th element of population 2, and let t_j^- be the value encoded in the second part of the same element. Similarly, let s_j^+ and t_j^+ be the first and second part of an element of the third population. Therefore, $\mathcal{A}^- = \{A_{s_j^-} \mid j = 1, \ldots M\}$ and $\mathcal{A}^+ = \{A_{s_j^+} \mid j = 1, \ldots M\}$. Then the fitness of an individual x is the interval $[g^-(x), g^+(x)]$ defined as follows:

$$\begin{aligned} g^-(x) &= \textstyle\sum_{j=1}^{M} ||F_x(t_j^-) - y_{s_j^-}|| \\ &+ \textstyle\sum_{\{i | A_i \notin \mathcal{A}^-\}} ||F_x(\mathrm{mid}(A_i)) - y_i|| \end{aligned} \tag{6}$$

$$\begin{aligned} g^+(x) &= \textstyle\sum_{j=1}^{M} ||F_x(t_j^+) - y_{s_j^+}|| \\ &+ \textstyle\sum_{\{i | A_i \notin \mathcal{A}^+\}} ||F_x(\mathrm{mid}(A_i)) - y_i||. \end{aligned} \tag{7}$$

The fitness value of an individual in the second or the third last two populations is, respectively, the gain or loss in the lower and upper bounds of the error of the elite model, when the point contained in the individual

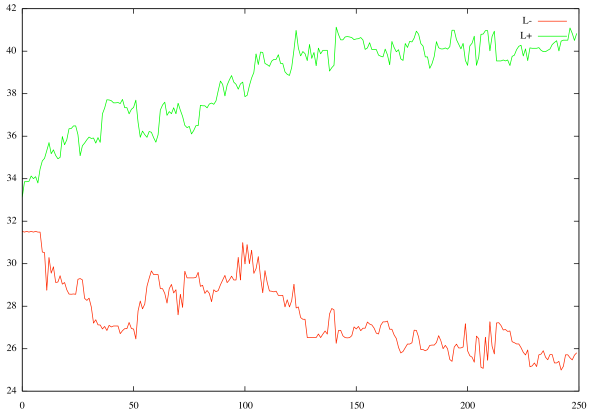


Figure 4: Example run of the GA: Bounds of the error of the best model in the first population when the second and third populations evolve

is replaced by the middlepoint of A_s (where s is the index codified in the individual, as mentioned); that is to say, the fitness of the j-th element of the second and third populations are

$$\begin{aligned} \mathrm{fit}_{j1} &= -\|\mathrm{FRBS}_{\mathrm{elite}}(t_j^-) - y_{s_j^-}\| \\ &+\|\mathrm{FRBS}_{\mathrm{elite}}(\mathrm{mid}(A_{s_j^-})) - y_{s_j^-}\| \end{aligned} \tag{8}$$

and

$$\begin{aligned} \mathrm{fit}_{j2} &= \|\mathrm{FRBS}_{\mathrm{elite}}(t_j^+) - y_{s_j^+}\| \\ &-\|\mathrm{FRBS}_{\mathrm{elite}}(\mathrm{mid}(A_{s_j^+})) - y_{s_j^+}\|. \end{aligned} \tag{9}$$

With this definition, the sum of the fitness values of all the individuals in the population equals the difference between the error on the sample comprising the midpoints of the interval-valued training set and the error commited when $\mathcal{A}^-$ and $\mathcal{A}^+$ are codified by populations 2 and 3. The genetic evolution produces, therefore, a cooperative behavior and the populations evolve to encode sets of values with respectively lower and higher errors (see Figure 4 for an actual plot of the bounds of the error of the best model in the first population when the second and third populations evolve).

It is remarked that, in case that an index appears more than once in the same population, the fitness values of all the individuals but the best one must be set to zero, or else the sum of the fitness values is no longer the mentioned difference.

2.3 Coevolutionary scheme

The coevolutionary scheme is as follows:

1. The three populations are initialized with random values.

2. Repeat steps 3 to 9, G_1 times:

3. The fitness of individuals in the first population are computed

4. The first population is ranked by means of a precedence operator between intervals [9]. The elite is copied apart. Tournament selection, crossover and mutation are performed in this population, and the offspring is inserted in place of the worst individuals in the tournament.

5. Repeat steps 6 to 9, G_2 times:

6. Evaluate the fitness of population 2
7. Evaluate the fitness of population 3
8. Assign fitness 0 to duplicates in pop. 2 and 3
9. Crossover and mutation are performed in these last two populations, and the offspring is inserted back in place (steady state).

2.4 Genetic operators

All algorithms are steady state and based in a tournament selection. The offspring of the winners of the tournament replace the last two elements of the tournament, whose length is used to control the selective pressure.

Standard two-point crossover and mutation are used in the first population. The application of the crossover operator is followed by a Lamarckian local search (Nelder and Mead's algorithm) with certain probability.

The other two populations need custom operators; two individuals (s_1, δ_1) and (s_2, δ_2) are crossed as follows:

- If $s_1 = s_2$, we do an arithmetic crossover between δ_1 and δ_2 [10].
- If $s_1 \neq s_2$, we insert a copy of the best individual and randomly generate the other.

3 Numerical results

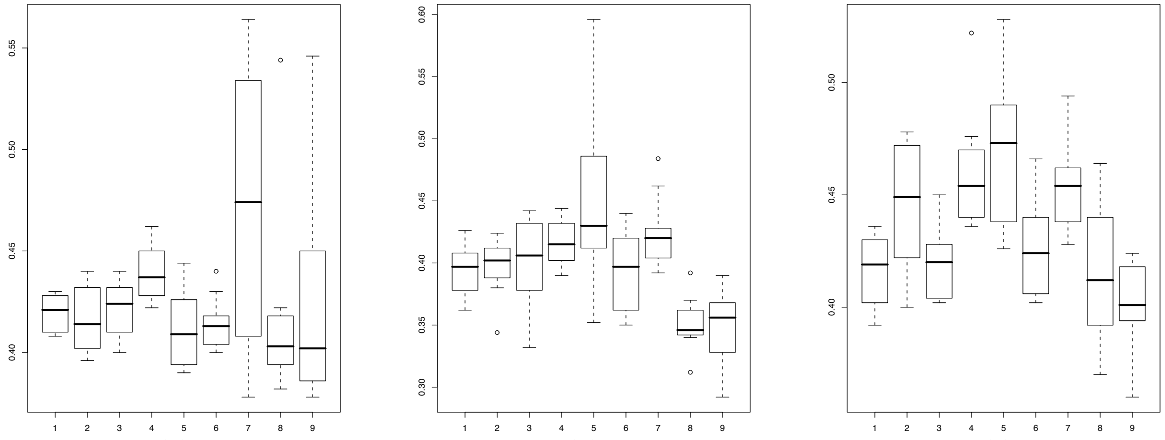
This section contains the initial results of the algorithm, when applied to synthetic problems. There are two categories of datasets that should be suitable for a method like this:

1. Data for which the classification rules can be expressed with a compact rulebase: low to moderate number of features, not too complex decision surface.
2. Low quality data: censoring, interval valued and missing features in the data.

To comply with our first requirement, we have built a FRBS comprising 9 rules in a problem with two

								Crisp	Interval
	Linear	Quadratic	Neural	KNN	WM	ISH	PM	FRBS	FRBS
censored - 100	0.492	0.478	0.460	0.448	0.448	0.488	0.478	0.478	**0.424**
censored - 1000	0.421	0.414	0.424	0.437	0.409	0.413	0.474	0.403	**0.402**
interval - 100	0.554	0.478	0.490	0.506	0.460	0.478	0.458	0.442	**0.432**
interval - 1000	0.394	0.397	0.402	0.416	0.450	0.393	0.424	0.351	**0.346**
missing - 100	0.408	0.372	0.426	0.376	0.364	**0.328**	0.518	0.330	0.372
missing - 1000	0.416	0.445	0.412	0.461	0.470	0.426	0.456	0.415	**0.401**

Table 1: Numerical results: Crisp algorithms (LDA and QDA discriminant analysis [4], multilayer perceptron, KNN classifier, Chi [1], Ishibuchi [5], Pal-Mandal [13] and FRBS [12] were compared to Interval-FRBS. The best test results are boldfaced.



(a) (b) (c)

Figure 5: Boxplots showing the dispersion of the results in Figure 1. Censored data (a), interval valued data (c) and missing data (c). The algorithms being compared are in the same order as they appear in Figure 1.

inputs between 0 and 1, and two classes. Since we know the distribution of the population, we have generated datasets whose Bayes error is also known, and for which there exists a FRBS which is the optimal solution. Two datasets of sizes 100 and 1000 were generated.

The second requirement has been fulfilled by adding imprecision to these datasets. We have considered three different categories of imprecision:

1. Censoring: in the 50% of cases, the training data x_s is replaced by the interval $[0, x_s]$. The other cases were replaced by the interval $[x_s, 1]$.
2. Interval valued data: each training data is replaced by the interval $[x_s, x_s + 0.4]$. or $[x_s, 1]$ if $x_s + 0.4 > 1$
3. Missing values: 40% of the points in the training set had one of their features replaced by the interval $[0, 1]$.

These three additions were performed for both datasets, giving the six problems we will use in this section. Other details of the experimental setup are: each experiment has been repeated 10 times, with a 5x2cv experimental design. The size of the first genetic population is 50. Second and third populations have sizes 100 or 1000, depending on the dataset. The number of generations G_1 is 50 and G_2 is 5 (see Section 2.3). The probabilities of crossover and mutation in the first population are 0.7 and 0.1, and the probability of crossover in the second and third populations are equal to 0.9. The probability of the local search is 0.25. The tournament size is 5.

For crisp algorithms (LDA and QDA discriminant analysis [4], multilayer perceptron, KNN classifier, Chi [1], Ishibuchi [5], Pal-Mandal [13] and FRBS [12]), we replaced each interval by its midpoint. We expect that our approach performs the best in all the cases we selected, and also that the final populations 2 and 3 contain the less informative points for the classifier (i.e., those points that, if removed, reduce the most the width of the interval of likelihoods of the model).

The mean value of the test results are shown in Table 1, and the boxplots depicting the relevance of the differences are displayed in Figure 5. We have obtained the expected results in all cases but one (40% of missing data, datasets of size 100), where the crisp version of the same algorithm improved the results. At the sight of these preliminary results, we think that this algorithm is a promising new technique for exploiting interval data in rule-based classification problems.

4 Concluding remarks

In this paper we have tried a different approach for obtaining linguistically understandable classifiers from interval-valued data. We have defined a coevolutionary scheme able to search in parallel for the best set of rules, and for the two selections of the training set where the lowest and highest errors are reached. These two bounds are used to find a model which is not dominated by other models, and that results in a robust estimation under vague input data. We have checked, with an ad-hoc dataset, that this approach is able to obtain better models than some statistical and fuzzy classifiers, if the conditions are appropriate.

Acknowledgements

This work was supported by the Spanish Ministry of Science and Innovation, under grant TIN2008-06681-C06-04.

References

[1] Chi, Z., Yan, H., Pham, T. *Fuzzy Algorithms: With Applications to Image Processing and Pattern Recognition.* World Scientific. (1996).

[2] Cordón O, Herrera F, Hoffmann F, Magdalena L, Genetic fuzzy systems. Evolutionary tuning and learning of fuzzy knowledge bases. World Scientific, Singapore (2001)

[3] Ferson S, Kreainovich V, Hajagos J, Oberkampf W, Ginzburg L, Experimental uncertainty estimation and statistics for data having interval uncertainly. Technical Report SAND2007-0939, Sandia National Laboratories (2007)

[4] Hand, D. J. *Discrimination and Classification.* Wiley. (1981)

[5] Ishibuchi, H., Nakashima, T., Murata, T., A fuzzy classifier system that generates fuzzy if-then rules for pattern classification problems. In Proc. of 2nd IEEE CEC, 759–764. (1995)

[6] Ishibuchi H, Nakashima T, Murata T, Performance evaluation of fuzzy classifier systems for multidimensional pattern classification problems. IEEE Transactions on Systms, Man and Cybernetics. Part B-Cybernetics 29(5):601-618. (1999).

[7] De Jong KA, Spears WM, Gordon D.F. Using genetic algorithms for concept learning. Machine Learning 13:161-188. (1993).

[8] Koeppen, M., Franke, K., and Nickolay, B., Fuzzy-Pareto-Dominance driven multi-objective genetic algorithm. in Proc. 10th International Fuzzy Systems Assotiation World Congress (IFSA), Istanbul, Turkey, 2003: 450–453. (2003).

[9] Limbourg, P., Multi-objective optimization of problems with epistemic uncertainty. in EMO 2005: 413–427. (2005).

[10] Michalewicz, Z. Genetic algorithms + Data Structures = Evolution Programs. Springer-verlag, (1992)

[11] Montes, I., Martinetti, D., Diaz, S., Montes, S. Statistical Preference as a comparison method of two imprecise fitness values. ESTYLF 2010. To appear. (2010)

[12] Otero J., Sanchez, L. (2006) Induction of descriptive fuzzy classifiers with the Logitboost algorithm Soft Computing 10 (9): 825–835

[13] Pal, S. K., Mandal, D. P. “Linguistic recognition system based in approximate reasoning”. *Information Sciences* **61**, pp. 135-161. (1992).

[14] Potter, M., De Jong, K. A cooperative coevolutionary approach to function optimization. Lecture Notes in Computer Science 866. 249-257. 2006.

[15] Ratschek, H. Some recent aspects of interval algorithms for global optimization. In: Moore RE (ed) Reliability in Computing: The Role of Interval Methods in Scientific Computing. Acad. Press, New York, pp 325? 339. (1988).

[16] Sánchez L., Couso I. Advocating the use of imprecisely observed data in genetic fuzzy systems IEEE Transactions on Fuzzy Systems 15 (4): 551–562. (2007)

[17] Sánchez, L., Couso, I., Casillas, J. Modeling vague data with genetic fuzzy systems under a combination of crisp and imprecise criteria. First IEEE Symposium on Computational Intelligence in Multi-Criteria Decision-Making (MCDM 2007). Honolulu, Hawaii, USA, (2007).

[18] Sánchez, L., Couso, I., Casillas, J. Genetic learning of fuzzy rules based on low quality data. Fuzzy Sets and Systems. 160 (17) 2524-2552. (2009)