Soft Comput (2007) 11:401–419
DOI 10.1007/s00500-006-0106-2

FOCUS

Rule base reduction and genetic tuning of fuzzy systems based on the linguistic 3-tuples representation

Rafael Alcalá · Jesús Alcalá-Fdez · María José Gacto · Francisco Herrera

Published online: 10 June 2006

Abstract Recently, a new linguistic rule representation model was presented to perform a genetic lateral tuning of membership functions. It is based on the linguistic 2-tuples representation model, that allows the symbolic translation of a label considering an unique parameter. It involves a reduction of the search space that eases the derivation of optimal models. This work presents a new symbolic representation with three values (s, α, β), respectively representing a label, the lateral displacement and the amplitude variation of the support of this label. Based on this new representation we propose a new method for fine tuning of membership functions that is combined with a rule base reduction method in order to extract the most useful tuned rules. This approach makes use of a modified inference system that consider non-covered inputs in order to improve the final fuzzy model generalization ability, specially in highly non-linear problems with noise points. Additionally, we analyze the proposed approach showing its behavior in two real-world applications.

Supported by the Spanish Ministry of Science and Technology under Projects TIC-2002-04036-C05-01 and TIN-2005-08386-C05-01.

R. Alcalá (✉) · J. Alcalá-Fdez · M. J. Gacto · F. Herrera
Department of Computer Science and Artificial Intelligence, University of Granada, 18071 Granada, Spain
e-mail: alcala@decsai.ugr.es

J. Alcalá-Fdez
e-mail: jalcala@decsai.ugr.es

M. J. Gacto
e-mail: mjgacto@ugr.es

F. Herrera
e-mail: herrera@decsai.ugr.es

Keywords Linguistic fuzzy modeling · Interpretability-accuracy trade-off · Evolutionary tuning · Linguistic 3-tuples representation · Rule selection

1 Introduction

Fuzzy modeling (FM)—i.e., system modeling with fuzzy rule-based systems (FRBSs)—may be considered as an approach used to model a system making use of a descriptive language based on fuzzy logic [45,46] with fuzzy predicates [47]. Several types of modeling can be performed by using different types of FRBSs and depending of the desired degree of interpretability and accuracy of the final model. Unfortunately, both requirements are contradictory properties directly depending on the learning process and/or the model structure.

In this framework, one of the most important areas is *linguistic* FM, where the interpretability of the obtained model is the main requirement. This task is usually developed by means of linguistic FRBSs (also called Mamdani FRBSs [37,38]), which use fuzzy rules composed of linguistic variables [47] taking values in a term set with a real-world meaning. Thus, the linguistic fuzzy model consists of a set of linguistic descriptions regarding the behavior of the system being modeled.

One of the problems associated to linguistic FM is its lack of accuracy when modeling some complex systems. It is due to the inflexibility of the concept of linguistic variable, which imposes hard restrictions to the fuzzy rule structure [3]. This drawback leads linguistic FM to sometimes move away from the desired trade-off between interpretability and accuracy, thus losing the usefulness of the model. To overcome this problem, many different possibilities to improve the accuracy of linguistic FM while preserving its intrinsic interpretability have been considered in the specialized literature [4].

Recently, to improve the accuracy of FRBSs (improving the said trade-off), a new linguistic rule representation model was proposed to perform a genetic lateral tuning of membership functions [1]. This new approach was based on the linguistic 2-tuples representation [23], that allows the symbolic translation of a label by considering an unique parameter per label. In this way, two main objectives were achieved:

- To obtain membership functions containing a set of samples with a better covering degree maintaining their original shapes (accuracy improvements), and
- To reduce the search space respect to the classical tuning [2,13,19–21,31,32] (usually considering three parameters in the case of triangular membership functions), in order to easily obtain optimal models.

However, the amplitude of the support of the membership functions is fixed through this tuning process. This amplitude determines the specificity of a label and involves a potential accuracy improvement, since it could determine the best covering region of such label.

On the other hand, in order to improve the interpretability, rule reduction methods directly aggregate multiple rules and/or select a subset of rules from a given fuzzy rule set [30,34]. The combination of tuning approaches with rule selection methods can present a positive synergy, reducing the tuning search space, easing the readability and even improving the system accuracy.

This work presents a new symbolic representation with three values (s, α, β), respectively representing a label, the lateral displacement and the amplitude variation of the support of this label. Based on this new representation, we propose a new method for the Lateral and Amplitude tuning (LA-tuning) of membership functions, that is combined with a rule base (RB) reduction method with the main aim of improving the system accuracy and trying to maintain part of the interpretability as much as possible respect to the lateral tuning. The use of the new parameter β improves the accuracy but loses part of the interpretability, since it has no meaning associated in principle. However, the RB reduction method extracts the most useful tuned rules (the rules with a better information), obtaining a simpler knowledge base (KB) and therefore, enhancing the interpretability. This way to work still gives a reasonable trade-off between accuracy and interpretability.

Additionally, we introduce a new inference system that consider non-covered inputs in order to improve the final fuzzy model generalization ability, specially in highly non-linear problems with noise (which can in fact present new data in the test or the real system).

The paper is arranged as follows. The next section presents the LA-tuning of membership functions and the new inference system for non-covered data. Section 3 describes an evolutionary algorithm to perform the LA-tuning. In Sect. 4, the evolutionary method to perform the LA-tuning together with a rule selection is proposed. Section 5 shows an experimental study of the methods behavior applied to two real-world electrical distribution problems. Finally, Sect. 6 points out some conclusions.

2 A proposal for the LA-tuning of fuzzy rule-based systems

In this section, we will introduce the lateral tuning of membership functions —a part of the LA-tuning. Then, the extension of the lateral tuning to also perform the amplitude tuning will be described, presenting the new rule representation and two different tuning approaches (global approach and local approach).

2.1 Preliminaries: the lateral tuning

In [1], a new model of tuning of FRBSs was proposed considering the linguistic 2-tuples representation scheme introduced in [23], which allows the lateral displacement of the support of a label and maintains a good interpretability associated to the obtained linguistic FRBSs. This proposal also introduces a new model for rule representation based on the concept of symbolic translation (the lateral displacement of a label).

The symbolic translation of a linguistic term is a number within the interval [−0.5, 0.5) that expresses the domain of a label when it is moving between its two lateral labels. Let us consider a set of labels S representing a fuzzy partition. Formally, we have the pair,

$$(s_i, \alpha_i), \quad s_i \in S, \quad \alpha_i \in [-0.5, 0.5).$$

Figure 1 depicts the symbolic translation of a label represented by the pair $(S_2, S - 0.3)$, considering a set S with five linguistic terms represented by their ordinal values ({0, 1, 2, 3, 4}).

In [23], both the linguistic 2-tuples representation model and the needed elements for linguistic information comparison and aggregation are presented and applied to the decision making framework. In the context of the FRBSs, we are going to see its use in the linguistic rule representation. Below we present this approach considering a simple control problem.

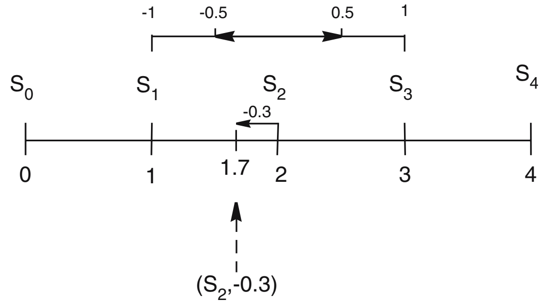


Fig. 1 Symbolic translation of a label

Let us consider a control problem with two input variables, one output variable and a data base (DB) defined from experts determining the membership functions for the following labels:

X1: Error $\rightarrow \{N, Z, P\}$,

X2: $\triangledown$Error $\rightarrow \{N, Z, P\}$,

Y: Power $\rightarrow \{L, M, H\}$.

Based on this DB definition, an example of classical rule and linguistic 2-tuples represented rule is:

Classical rule,
R1: If the **error** is Zero and the $\triangledown$**Error** is Positive then the **Power** is High .

Rule with 2-tuples representation,
R1: If the **error** is (Zero, 0.3) and the $\triangledown$**Error** is (Positive, -0.2) then the **Power** is (High, -0.1).

Analyzed from the rule interpretability point of view, we could interpret the linguistic 2-tuples represented rule in the following way:

If the **Error** is "higher than Zero" and
the **Error Variation** is "a little smaller than Positive"
then the **Power** is "a bit smaller than High".

This proposal decreases the tuning problem complexity, since the three parameters usually considered per label [2,13,19–21,31,32] are reduced to only one symbolic translation parameter.

2.2 The LA-tuning of membership functions

This approach is an extension of the lateral tuning to also perform a tuning of the support amplitude of the membership functions. To adjust the displacements and amplitudes of the membership function supports we propose a new rule representation that considers two parameters, α and β, relatively representing the lateral displacement and the amplitude variation of a label. In this way, each label can be represented by a 3-tuple (s, α, β), where α is a number within the interval $[-0.5, 0.5)$ that expresses the domain of a label when it is moving between its two lateral labels (as in the 2-tuples representation), and β is also a number within the interval $[-0.5, 0.5)$ that allows to increase or reduce the support amplitude of a label until a 50% of its original size. Let us consider a set of labels S representing a fuzzy partition. Formally, we have the triplet,

$$(s_i, \alpha_i, \beta_i), \quad s_i \in S, \quad \{\alpha_i, \beta_i\} \in [-0.5, 0.5)$$

Figure 2 depicts the lateral and amplitude variation of the label M considering triangular and symmetrical equidistant membership functions. The new label "y_2" is located between labels S and M, and has a shorter support than the original label M. Let us represent the new label "y_2" as the 3-tuple (M, α, β). The support of this label, Sup_{y_2}, can be computed in the following way:

$$\mathrm{Sup}_M = c_M - a_M$$
$$\mathrm{Sup}_{y_2} = \mathrm{Sup}_M + \beta * \mathrm{Sup}_M \,,$$

where c_M and a_M are respectively the right and the left extreme of the support of M, and Sup_M is the size of the support of M.

In [1], two different rule representation approaches were proposed for the lateral tuning of membership functions, a global approach and a local approach. We

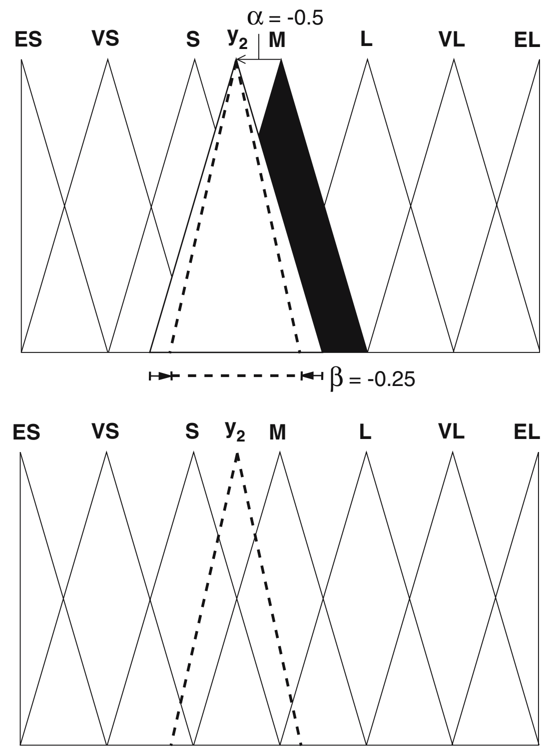


Fig. 2 Lateral displacement and amplitude variation of the linguistic label M considering the set of labels S={ES, VS, S, M, L, VL, EL}

will consider the same two possibilities to perform the LA-tuning, the most interpretable one and the most accurate one:

- Global tuning of the semantics. In this case, the tuning is applied to the level of linguistic partition. In this way, the pair (X_i, label) takes the same tuning values in all the rules where it is considered. For example, X_i is (High, 0.3, 0.1) will present the same values for those rules in which the pair "X_i is High" was initially considered.
- Local tuning of the rules. In this case, the tuning is applied to the level of rule. The pair (X_i, label) is tuned in a different way in each rule:

  Rule k: X_i is (High, 0.3, 0.1)
  Rule j: X_i is (High, -0.2, 0.25)

Notice that, since symmetrical triangular membership functions and a FITA (*First Infer, Then Aggregate*) fuzzy inference [8] will be considered in both, the global and the local approach, a tuning of the amplitude of the consequents has no sense, by which the β parameter will be only applied on the antecedents. In this way, considering the same control problem of the previous subsection, an example of a 3-tuples represented rule is (amplitude variation only applied in the antecedents):

Rule with 3-tuples representation,
R1: If the **error** is (Zero,0.3,0.1) and the $\bigtriangledown$**Error** is (Positive,–0.2,–0.4) then the **Power** is (High,–0.1) .

Analized from the rule interpretability point of view, we could interpret the lateral displacement as said in Sect. 2.1. However, it is not clear a meaning for the amplitude factor β. In this way, if the final membership functions are more or less well distributed and no strong amplitude changes have been performed, an expert perhaps could rename these labels given them a more or less representative meaning.

On the other hand, the use of the β factor (amplitude) is close to the use of non-linear scaling factors [9,36] or linguistic modifiers [9,18]. However there are some differences with these approaches:

a) By using non-linear scaling factors or linguistic modifiers an example that is covered by a label can not be uncovered and *vice versa*, which imposes some restrictions to the search. Determining the amplitude of a membership function is a way to decide which examples are covered or not, better grouping a set of data. Therefore, tuning the amplitude of the membership functions can help,
   - To reduce the number of negative examples (those covered in antecedents but not in the consequents),
   - To increase the number of positive examples (those covered in antecedents and consequents), or
   - To reduce the number of rules if a rule selection method is considered.
b) Contrary to the non-linear scaling factors or linguistic modifiers, the tuning of the support amplitude keeps the shape of the membership functions (triangular and symmetrical). In this way, from the parameters α and β applied to each label, we could obtain the equivalent triangular membership functions, by which the final tuned FRBS could be represented as a classical Mamdani FRBS [37,38].

2.3 Fuzzy inference system

The fuzzy reasoning method is the *minimum* playing the role of the implication and conjunctive operators, and the *center of gravity weighted by the matching* strategy acting as defuzzification operator [8] (FITA scheme). These kinds of inference is applied once the 3-tuples represented model is transformed to (represented by) its equivalent classical Mamdani FRBS.

Since the support of the final membership functions comprising the rules can be reduced and displaced and a rule selection method can be also considered, there could be non-covered zones in the input space. Taking into account that the tuning algorithm is biased by error measures, this fact is not a problem. Non-covered training data usually provokes high errors in the system and finally they would be covered as a result of the optimization process. However, in real problems with strong non-linearity, a few number of training data (usually not totally representative of the modeling surface) and even with noise data, a good behavior of the obtained model is not ensured for the non-covered test data (i.e., the generalization of the final linguistic model could not be good for uncovered inputs). Furthermore, when noise points are present in the training data, the optimization process could try to not cover these points in order to improve the behavior of the truly available data. In this way, to consider non-covered input data for the system output computation, the following mechanism is applied:

1. The nearest rule to the non-covered point is identified (normalized euclidean distance to the vertex of the labels). The non-covered coordinates of the point are set to the value of the vertex of the corresponding label. The euclidean distance is normalized by the maximum distance among the vertex of the

different labels considered for each coordinate in the corresponding RB.
2. The second nearest rule is identified. Then, if the consequent labels of both rules present overlapping to some degree, we only infer with the nearest rule since it will be the most representative in a subspace that does not present strong changes in the output domain.
3. In other case, the final FRBS output should be obtained by interpolation of both rules, since strong changes are detected in this subspace output domain. To do that, the coordinates of the point that are initially covered are displaced towards the second rule, ensuring a minimum covering degree of the nearest rule (nearing these coordinates to the corresponding label extreme at the 10% of the support size). As example, let e_i be a coordinate of the non-covered point e that is initially covered by the corresponding label of the nearest rule $\{a_i^{1st}, b_i^{1st}, c_i^{1st}\}$ (left extreme, vertex and right extreme). And let $\{a_i^{2nd}, b_i^{2nd}, c_i^{2nd}\}$ be the definition points of the corresponding label of the second nearest rule. Then, the new value e_i' is computed as follows:

$$e_i' = \begin{cases} a_i^{1st} + (c_i^{1st} - a_i^{1st}) * 0.1, & \text{If } b_i^{2nd} < b_i^{1st}, \\ c_i^{1st} - (c_i^{1st} - a_i^{1st}) * 0.1, & \text{If } b_i^{2nd} > b_i^{1st}, \\ e_i, & \text{If } b_i^{2nd} = b_i^{1st}. \end{cases} \quad (1)$$

4. Finally, we infer with the new obtained point and considering the complete RB.

2.4 Trade-off between interpretability and accuracy

As said, the support amplitude of the membership functions is kept fixed in the lateral tuning. In this work, we will also tune this amplitude in order to determine the specificity or generality of a label. It involves some advantages (accuracy improvements) and also some drawbacks (interpretability costs) respect to the lateral tuning, where only the membership function position is adjusted:

– Advantages:
  1. There is a potential accuracy improvement respect to the lateral tuning since the 3-tuples representation involves more freedom degrees.
  2. To determine the subset of training points that should be covered is easier, getting more specific and consistent rules.
– Drawbacks:
  1. The models obtained are less interpretable than those considering the 2-tuple representation. However, since they could be transformed to classical Mamdani systems, the models considering the linguistic 3-tuples representation can be considered at least as interpretable as those obtained by classical methods adjusting the three membership function definition points.
  2. The search space increases respect to the lateral tuning, making more difficult the derivation of optimal models. However, this approach still involves a reduction of the search space respect to the classical tuning (one less parameter), which is still well handled by means of an smart use of the search technique.

Both approaches, lateral tuning and LA-tuning, present a good trade-off between interpretability and accuracy. However, the proposed approach is closer to the accuracy than the lateral tuning, being this last closer to the interpretability. The choice between how interpretable and how accurate the model must be, usually depends on the user's needs for a specific problem and it will condition the selection of the type of tuning considered.

From now on, it must be taken into account that when we remark the differences among the classical tuning, the lateral tuning and the LA-tuning, we are separately considering the global and the local approach, i.e., on the one hand the global approaches are compared and on the other hand the local approaches are compared. In this way, the global LA-tuning reduces the search space respect to the classical tuning of the DB and the local LA-tuning reduces the search space respect to the classical tuning of approximate fuzzy rules. From the point of view of the interpretability, they also should be individually compared, since the global approach tries to obtain more interpretable models and the local approach tries to obtain more accurate ones.

The evolutionary LA-tuning method based on the global and local 3-tuples representation models is shown in the next section.

3 Evolutionary LA-tuning algorithm

The automatic definition of fuzzy systems can be considered as an optimization or search process and nowadays, evolutionary algorithms, particularly genetic algorithms (GAs) [16,25], are considered as the more known and used global search technique. Moreover, the genetic coding that they use allow them to include prior knowledge and to use it leading the search up. For this reason, Evolutionary Algorithms have been successfully applied to learn fuzzy systems in the last years,

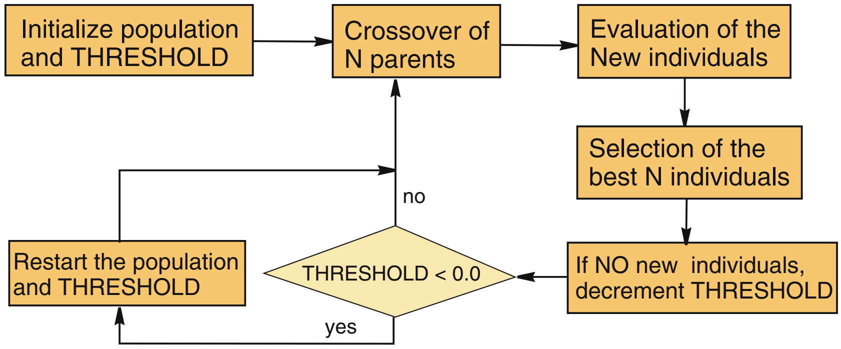


Fig. 3 Scheme of CHC

giving way to the appearance of the so called genetic fuzzy systems (GFSs) [13,14]. Taking into account these interesting characteristics, in this section we propose a new GFS to perform the LA-tuning of membership functions.

To do so, we consider the genetic model of CHC [15]. CHC is a GA that presents a good trade-off between exploration and exploitation, being a good choice in problems with complex search spaces. In the following, the components needed to design the evolutionary tuning process are explained. They are:

- Evolutionary model of CHC.
- Coding scheme and initial gene pool.
- Chromosome evaluation.
- Crossover operator.
- Restarting approach.

3.1 Evolutionary model of CHC

We will consider a population-based selection approach, by using the CHC evolutionary model [15] in order to perform an adequate global search. The genetic model of CHC makes use of a "Population-based Selection" approach. N parents and their corresponding offspring are combined to select the best N individuals to take part of the next population. The CHC approach makes use of an incest prevention mechanism and a restarting process to provoke diversity in the population, instead of the well known mutation operator.

This incest prevention mechanism will be considered in order to apply the crossover operator, i.e., two parents are crossed if their hamming distance divided by 2 is over a predetermined threshold, L_T. Since, we will consider a real coding scheme, we have to transform each gene considering a Gray Code with a fixed number of bits per gene ($BITSGENE$) determined by the system expert. In this way, the threshold value is initialized as:

$$L_T = (\#Genes * BITSGENE)/4.0,$$

where $\#Genes$ is the number of genes in the chromosome. Following the original CHC scheme, L_T is decremented by one when there is no new individuals in the population in one generation. In order to make this procedure independent of $\#Genes$ and $BITSGENE$, in our case, L_T will be decremented by a $\varphi\%$ of its initial value (being φ determined by the user, usually 10%). The algorithm restarts when L_T is below zero.

A scheme of this algorithm is shown in Fig. 3.

3.2 Coding scheme and initial gene pool

Taking into account that two different types of tuning have been proposed (global tuning of the semantics and local tuning of the rules), there are two different kinds of coding schemes. In both cases, a real coding is considered, i.e., the real parameters are the GA representation units (genes). Depending on the type of tuning we want to perform, we will consider one or another of the following coding schemes:

- Global tuning of the semantics: Joint of the parameters of the fuzzy partitions, lateral (C^L) and amplitude (C^A) tuning. Let us consider the following number of labels per variable: $(m^1, \ldots, m^n)$, with n being the number of system variables ($n-1$ input variables and 1 output variable). Then, a chromosome has the following form (where each gene is associated to the tuning value of the corresponding label),

$$C_T = (C^L + C^A),$$

$$C^L = (c^L_{11}, \ldots, c^L_{1m^1}, \ldots, c^L_{n1}, \ldots, c^L_{nm^n}),$$

$$C^A = (c^A_{11}, \ldots, c^A_{1m^1}, \ldots, c^A_{(n-1)1}, \ldots, c^A_{(n-1)m^n}).$$

  See the C_T part of Fig. 5 (in the next section) for an example of coding scheme considering this approach.
- Local tuning of the rules: Joint of the lateral (C^L) and amplitude (C^A) rule parameters. Let us consider that the FRBS has M rules, $(R1, R2, \ldots, RM)$,

with n system variables ($n-1$ input variables and 1 output variable). Then, the chromosome structure is,

$$C_T = (C^L + C^A),$$
$$C^L = (c^L_{11}, \ldots, c^L_{1n}, \ldots, c^L_{m1}, \ldots, c^L_{mn}),$$
$$C^A = (c^A_{11}, \ldots, c^A_{1(n-1)}, \ldots, c^A_{m1}, \ldots, c^A_{m(n-1)}).$$

To make use of the available information, the initial FRBS obtained from automatic fuzzy rule learning methods is included in the population as an initial solution. To do so, the initial pool is obtained with the first individual having all genes with value '0.0' (no displacement or amplitude variation to represent the initial solution), and the remaining individuals generated at random in $[-0.5, 0.5)$.

3.3 Chromosome evaluation

To evaluate a determined chromosome we will use the well-known mean square error (MSE):

$$\mathrm{MSE} = \frac{1}{2 \cdot |E|} \sum_{l=1}^{|E|} (F(x^l) - y^l)^2,$$

with $|E|$ being the size of a data set E, $F(x^l)$ being the output obtained from the FRBS decoded from the said chromosome when the l-th example is considered and y^l being the known desired output.

3.4 Crossover operator

The crossover operator is based on the the concept of environments (the offspring are generated around their parents). These kinds of operators present a good cooperation when they are introduced within evolutionary models forcing the convergence by pressure on the offspring (as the case of CHC). Figure 4 depicts the behavior of these kinds of operators, which allow the offspring genes to be around the genes of one parent, Parent Centric BLX (PCBLX), or around a wide zone determined by both parent genes BLX-α. Particularly, we consider the PCBLX operator that is based on the BLX-α [24].

The PCBLX is described as follows. Let us assume that $X = (x_1 \cdots x_g)$ and $Y = (y_1 \cdots y_g)$, $(x_i, y_i \in [a_i, b_i] \subset \Re, i = 1 \cdots g)$, are two real-coded chromosomes that are going to be crossed.

This crossover operator generates the offspring $Z = (z_1 \cdots z_g)$, where z_i is a randomly (uniformly) chosen number from the interval $[l_i, u_i]$, with $l_i = \max\{a_i, x_i - I\}$, $u_i = \min\{b_i, x_i + I\}$, and $I = \mid x_i - y_i \mid$.

The parents X and Y will be named differently: X will be called *female parent*, and Y will be called *male parent*. In this way, by taking X as female parent (Y as male), and then by taking Y as female parent (X as male) this algorithm generates two offspring.

3.5 Restarting approach

To get away from local optima, this algorithm uses a restart approach [15]. In this case, the best chromosome is maintained and the remaining are generated at random by adding to each gene of the best chromosome a random number generated within the variation interval $[-0.125, 0.125)$. If the resulting value is less (greater) than -0.5 (0.5) it is replaced by the extreme value -0.5 (0.5). It follows the principles of CHC [15], performing the restart procedure when the threshold value L_T is lower than zero.

4 Rule selection and LA-tuning

Sometimes, a large number of fuzzy rules must be used to reach an acceptable accuracy degree. However, an excessive number of rules makes difficult to understand the model behavior. Moreover, we may find different kinds of rules in a large fuzzy rule set: *irrelevant rules*, which do not contain significant information; *redundant rules*, whose actions are covered by other rules; *erroneous rules*, which are wrong defined and distort the FRBS performance; and *conflicting rules*, which perturb the FRBS performance when they coexist with others. These kinds of rules are usually obtained in the following situations:

1. When the final RB is generated by only considering the expert's knowledge, redundant, conflicting and even erroneous rules are usually obtained.
2. When we consider a fuzzy rule set learning process with tendency to generate too many rules (sometimes advisedly), redundant or conflicting rules could be found in the obtained RB. For example, methods as

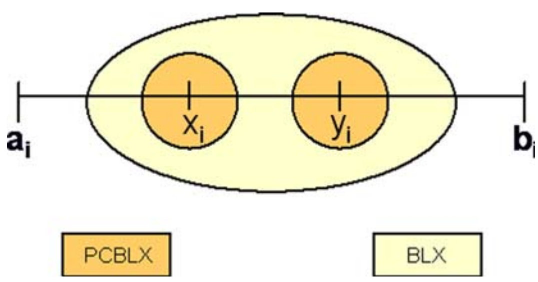


Fig. 4 Scheme of the behavior of the crossover operators based on environments

the Wang and Mendel algorithm [42] or the Input Space oriented Strategy [28] are biased by covering criteria trying to ensure a high covering degree that sometimes is not needed. The Mixed Method (MM) [6] consists of adding rules to the linguistic model obtained by the Wang and Mendel algorithm in the fuzzy input subspaces that having examples do not still have a rule, trying to improve the linguistic model accuracy by adding even more rules. However, these kinds of methods are still useful to obtain a set of promising candidate rules in order to subsequently select those with the best cooperation as a second stage [11,29]. In these cases, two main situations favor the generation of these kinds of rules:

- In complicated multidimensional problems with highly nonlinear input-output relations, in which the cooperation of the obtained rules is more difficult to be obtained and more rules are usually needed.
- In high-dimensional problems, in which the number of rules in the RB grows exponentially as more inputs are added. A large rule set might contain more redundant or even conflicting rules.

To face this problem, a fuzzy rule set reduction process can be developed to achieve the goal of minimizing the number of rules used while maintaining (or even improving) the FRBS performance. To do that, erroneous and conflicting rules that degrade the performance are eliminated, obtaining a more cooperative fuzzy rule set and therefore involving a potential improvement of the system accuracy. Moreover, in many cases the accuracy is not the only requirement of the model but also the interpretability becomes an important aspect. Reducing the model complexity is a way to improve the system readability, i.e., a compact system with few rules requires a minor effort to be interpreted.

Fuzzy rule set reduction is generally applied as a post-processing stage, once an initial fuzzy rule set has been derived. We may distinguish between two different approaches to obtain a *compact* fuzzy rule set:

- *Selecting fuzzy rules*: It involves obtaining an optimal subset of fuzzy rules from a previous fuzzy rule set by selecting some of them. We may find several methods for rule selection, with different search algorithms that look for the most successful combination of fuzzy rules [11,22,27,29,30,34].
  In [35], an interesting heuristic rule selection procedure is proposed where, by means of statistical measures, a relevance factor is computed for each fuzzy rule composing the linguistic FRBSs to subsequently select the most relevant ones. The philosophy of ordering the rules with respect to an importance criterion and selecting a subset of the bests seems similar to the orthogonal transformation-methods considered by Takagi-Sugeno-type FRBSs [43,44].
- *Merging fuzzy rules*: It is an alternative approach that reduces the fuzzy rule set by merging the existing rules. In [33], the authors propose to merge neighboring rules, i.e., fuzzy rules where the linguistic terms used by the same variable in each rule are adjacent. Another proposal is presented in [26], where a special consideration to the merging order is made. On the other hand, in Takagi-Sugeno-type FRBSs, processes that simplify the fuzzy models by merging fuzzy rules have also been proposed [39–41].

These kinds of techniques for rule reduction could be easily combined with other post-processing techniques to obtain more compact and accurate FRBSs. In this way, some works have considered the selection of rules together with the tuning of membership functions by coding all of them (rules and parameters) in the same chromosome [5,17].

4.1 Positive synergy between both approaches

There are several reasons explaining the positive synergy between the rule selection and the tuning of membership functions. Some of them are:

- The tuning process is affected when erroneous or conflictive rules are included in the initial RB. When the RB of a model being tuned contains bad rules (greatly increasing the system error), the tuning process tries to reduce the effect of these kinds of rules, adapting them and the remaining ones to avoid the bad behavior of such rules. This way to work imposes hard restrictions reducing the process ability to obtain precise linguistic models. Furthermore, in some cases this also affect to the model interpretability, since the membership functions comprising bad rules does not have the shape and location best representing the information being modeled.
  This problem grows as the problem complexity grows (i.e., problems with a large number of variables and/or rules) and when the rule generation method does not ensure the generation of rules with good quality (e.g., when the initial RB is obtained from experts). In these cases, the tuning process is very complicated since the search ability is dedicated to reduce the bad behavior of some rules instead of improving the behavior of the remaining ones.

In these cases, the rule selection could help the tuning mechanism by removing the rules that really degrade the model accuracy.
- Sometimes redundant rules can not be removed by only using a rule selection method, since these kinds of rules could reinforce the action of poor rules improving the model accuracy. The tuning of membership functions can change the behavior of these rules making unnecessary the reinforcement action, and therefore, helping the rule selection technique to remove redundant rules.

Therefore, combining rule selection and tuning approaches could result in important improvements of the system accuracy, maintaining the interpretability to an acceptable level [5,17]. However, in some cases, the search space considered when both techniques are combined is too large, which could provoke the derivation of sub-optimal models [5].

In this section, we propose the selection of a cooperative set of rules from a candidate fuzzy rule set together with the learning of the α and β parameters. This pursues the following aims:

- To improve the linguistic model accuracy selecting the set of rules best cooperating while the LA-tuning is performed to improve the location and amplitude of the membership functions.
- To obtain simpler, and thus easily understandable, linguistic models by removing unnecessary rules.
- To favor the combined action of the tuning and selection strategies (which involves a larger search space) by considering the simpler search space of the LA-tuning (only two parameters per label).

4.2 Evolutionary algorithm

To select the subset of rules best cooperating and to obtain the α and β parameters, we consider a GA coding all of them (rules and parameters) in a chromosome. This algorithm is based on the one proposed in Sect. 3, also considering the genetic model of CHC [15].

To do so, we must take into account the existence of binary genes (rule selection) and real values (lateral displacements and amplitude variation) within the same chromosome. Therefore, the algorithm proposed in Sect. 3 is changed in order to consider a double coding scheme and to apply the appropriate genetic operators for each chromosome part. The following changes are considered in order to integrate the reduction process within the proposed tuning method:

- *Coding scheme*: A double *coding scheme* for both, LA-tuning and rule selection, is considered:

  $C = C_T + C_S.$

  In this case, the previous approach (part C_T) is combined with the rule selection by allowing an additional binary vector C_S that determines when a rule is selected or not (alleles '1' and '0' respectively). Considering the M rules contained in the preliminary/candidate rule set, the chromosome part $C_S = (c_1, \ldots, c_M)$ represents a subset of rules composing the final RB, such that:

  $IF\ c_i = 1\ THEN\ (Ri \in RB)\ ELSE\ (Ri \notin RB),$

  with Ri being the corresponding i-th rule in the candidate rule set and RB being the final RB. Figure 5 graphically depicts an example of correspondence between a chromosome and its associated KB when the global LA-tuning and the rule selection are considered.
- *Initial gene pool*: The initial pool is obtained with an individual having all genes with value '0.0' in the C_T part and '1' in the C_S part, and the remaining individuals generated at random in $[-0.5, 0.5)$ and $\{0, 1\}$ respectively.
- *Crossover*: The environment-based crossover operator presented in Sect. 3 for the C_T part combined with the HUX crossover [15] in the C_S part. The HUX crossover exactly interchanges the mid of the alleles that are different in the parents (the genes to be crossed are randomly selected among those that are different in the parents). This operator ensures the maximum distance of the offspring to their parents (exploration).

  In this case, the incest prevention is applied considering a different threshold for each part, L_T applied to the C_T part (as in Sect. 3) and L_S applied to the C_S part (the HUX crossover is only applied if the hamming distance in C_S between both parents divided by 2 is over L_S). L_S is initialized as:

  $L_S = \#GenesC_S/4.0,$

  where $\#GenesC_S$ is the number of genes in C_S. Finally, the offspring are generated in the following way:
  1. If any crossover is applied, no offspring are generated.
  2. If only the PBLX crossover is applied, two offspring are generated by respectively copying the C_S part of their parents.
  3. If only the HUX crossover is applied, two offspring are generated by respectively copying the C_T part of their parents.

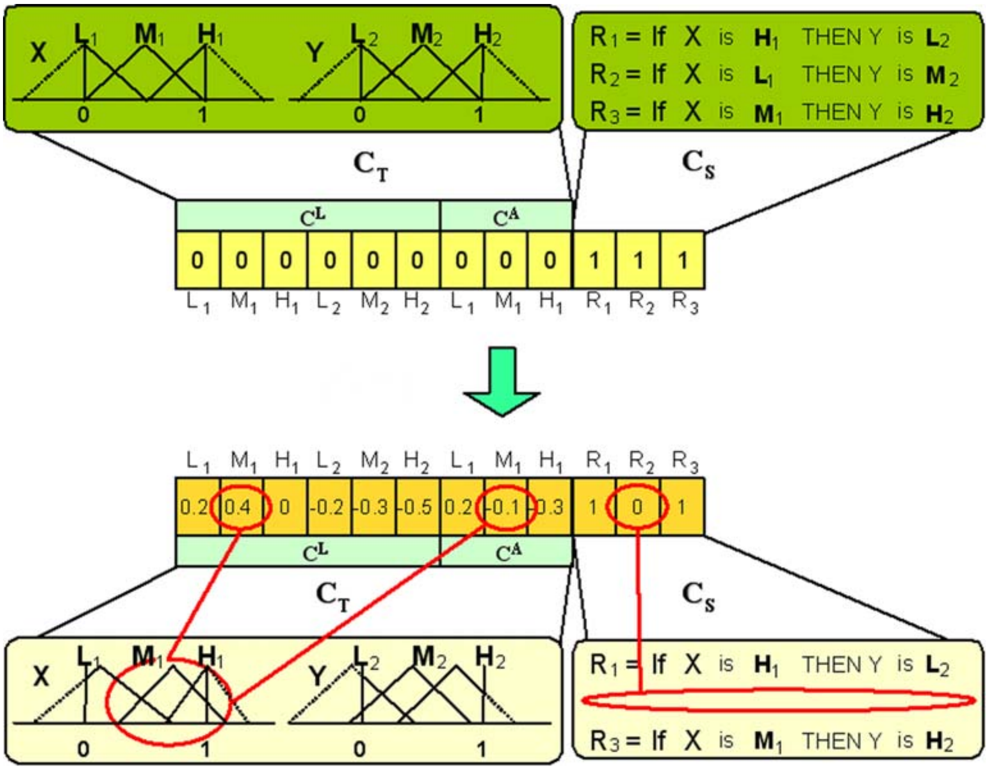


Fig. 5 Example of coding scheme considering the global LA-tuning and rule selection

4. If both PBLX and HUX are applied, two offspring are generated by joining the two ones from the C_T part with the two ones from the C_S part.

If no offspring are included in the new population by the application of PBLX, L_T is decreased as said in Sect. 3. If no offspring are included in the new population by the application of HUX, L_S is decreased by one.

– *Restarting approach*: To get away from local optima, the restarting operator is applied when both threshold values L_T and L_S are lower than zero. In this case, all the chromosomes set up their C_S parts to that of the best chromosome. Once the C_S part is fixed (first restarting), the L_S threshold is no more considered since it has no sense. The parameters of the C_T parts are generated at random by adding to each gene of the best chromosome a random number generated within the variation interval $[-0.125, 0.125)$.

5 Experimental study

To evaluate the goodness of the four proposed approaches (local and global tuning with and without rule selection), two real-world electrical energy distribution problems [10] with different complexities (different number of variables and rules) are considered to be solved:

– Estimating the length of low voltage lines in rural nuclei. This problem presents *noise and strong nonlinearities* which makes the modeling surface very complicated. On the other hand, a short number of variables have to be considered involving a short search space (small complexity).
– Estimating the maintenance costs of medium voltage lines in a town. This problem presents four input variables and a considerable number of rules, and therefore involves *a larger search space* (high complexity).

The initial set of fuzzy rules will be obtained from automatic learning methods. In both cases, the well-known ad hoc data-driven learning algorithm of Wang and Mendel [42] is applied to generate an initial set of candidate linguistic rules. To do so, we will consider symmetrical fuzzy partitions of triangular-shaped membership functions. Once the initial RB is generated, the proposed post-processing algorithms will be applied. In the following subsections these problems are introduced and solved to analyze the behavior of the proposed methods.

Table 1 Methods considered for comparison

References	Method	Description
[42]	**WM**	Learning method
[34]	**S**	Rule selection Method
[7]	**T**	Classical genetic tuning
[5]	**PAL**	Tuning of parameters, domains and local linguistic modifiers
[1]	**GL**	Global lateral tuning
[1]	**LL**	Local lateral tuning
–	**GLA**	Global lateral amplitude tuning
–	**LLA**	Local lateral amplitude tuning

5.1 Experimental set-up

The methods considered for the experiments are briefly described in Table 1. The use of two of them connected by '+' indicates that they are applied as a combination. E.g., GLA+S indicates a global LA-tuning together with rule selection.

As said, the WM method is considered to obtain the initial RB to be tuned. The initial linguistic partitions to obtain the initial RB are comprised by *five linguistic terms*. The tuning methods are applied once this initial RB has been obtained. T is a classical membership function parameter tuning algorithm. The PAL method has been compared with tuning methods of the parameters, domain, linguistic modifiers and with any combination of any two of them obtaining the best results [5]. For this reason, we only consider the PAL method (parameters, domains and linguistic edges) in this study. And finally, the GL and LL methods respectively performs a global and a local lateral tuning of the membership functions.

To develop the different experiments we consider a *5-folder cross-validation model*, i.e., five random partitions of data[1] with a 20%, and the combination of four of them (80%) as training and the remaining one as test. For each one of the five data partitions, the tuning methods have been run six times, showing for each problem the averaged results of a total of 30 runs. Moreover, a *t-test* (with 95 percent confidence) was applied in order to ascertain if differences in the performance of the proposed approaches are significant when compared against the one for the other algorithms in the respective table.

The following values have been considered for the parameters of each method:[2] 51 individuals, 50,000 evaluations, 30 bits per gene for the Gray codification and $\varphi = 10$ (0.2 and 0.6 as mutation and crossover probability per chromosome and 0.35 for the factor *a* in the max-min-arithmetical crossover for T and PAL).

5.2 Estimating the length of low voltage lines

For an electric company, it may be of interest to measure the maintenance costs of its own electricity lines. These estimations could be useful to allow them to justify their expenses. However, in some cases these costs can not be directly calculated. The problem comes when trying to compute the maintenance costs of low voltage lines and it is due to the following reasons. Although maintenance costs depend on the total length of the electrical line, the length of low voltage lines would be very difficult and expensive to be measured since they are contained in little villages and rural nuclei. The installation of these kinds of lines is often very intricate and, in some cases, one company can serve to more than 10,000 rural nuclei.

Due to this reason, the length of low voltage lines can not be directly computed. Therefore, it must be estimated by means of indirect models. The problem involves relating *the length of low voltage line of a certain village* with the following two variables: *the radius of the village* and *the number of users in the village* [10]. We were provided with the measured line length, the number of inhabitants and the mean distance from the center of the town to the three furthest clients in a sample of 495 rural nuclei. Five partitions[1] considering an 80% (396) in training and a 20% (99) in test are considered for the experiments. The existing dependency of the two input variables with the output variable in the training and test data sets of one of the five partitions is shown in Fig. 6 (notice that they present strong non-linearities).

The results obtained in this problem by the analyzed methods are shown in Table 2, where $\#R$ stands for the number of rules, MSE_{tra} and MSE_{tst} respectively for the averaged error obtained over the training and test data, σ for the standard deviation and *t-test* represents the following information:

- ⋆ represents the best averaged result.
- \+ means that the best result has better behavior than the one in the corresponding row.
- = denotes that the results are statistically equal according to the *t-test*.

[1] These data sets are available at: *http://www.decsai.ugr.es/∼casillas/fmlib/*

[2] With these values we have tried to ease the comparisons selecting standard common parameters that work well in most cases instead of searching very specific values for each method. Moreover, we have set a large number of evaluations in order to allow the compared algorithms to achieve an appropriate convergence. No significant changes were achieved by increasing that number of evaluations.

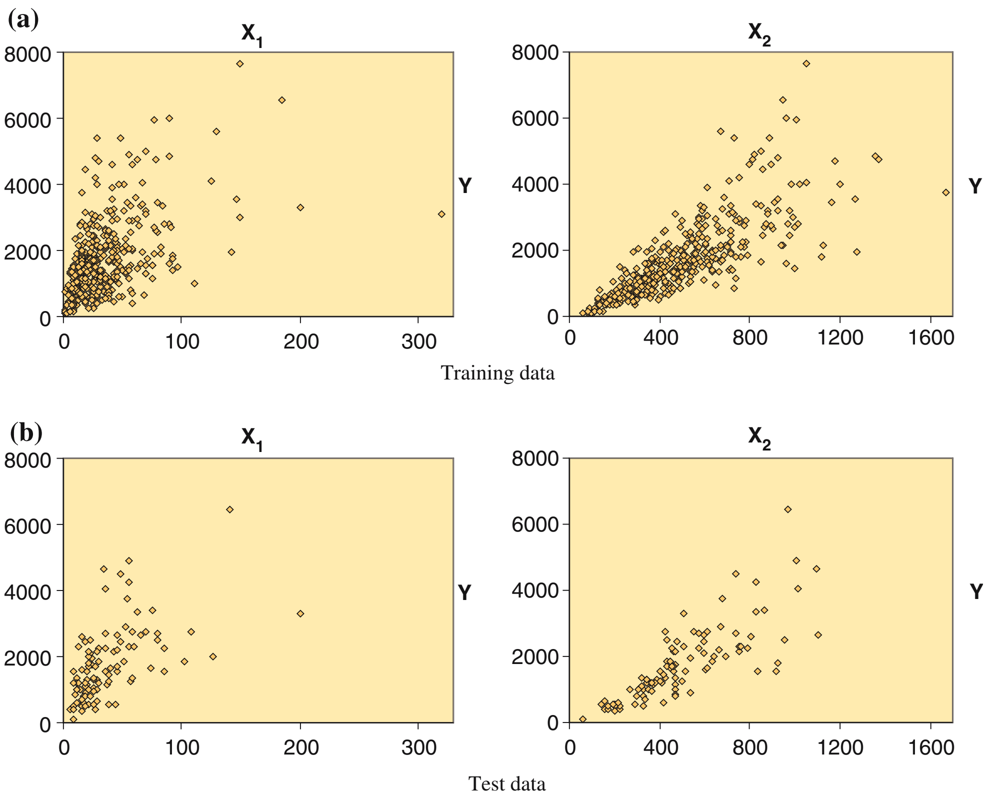


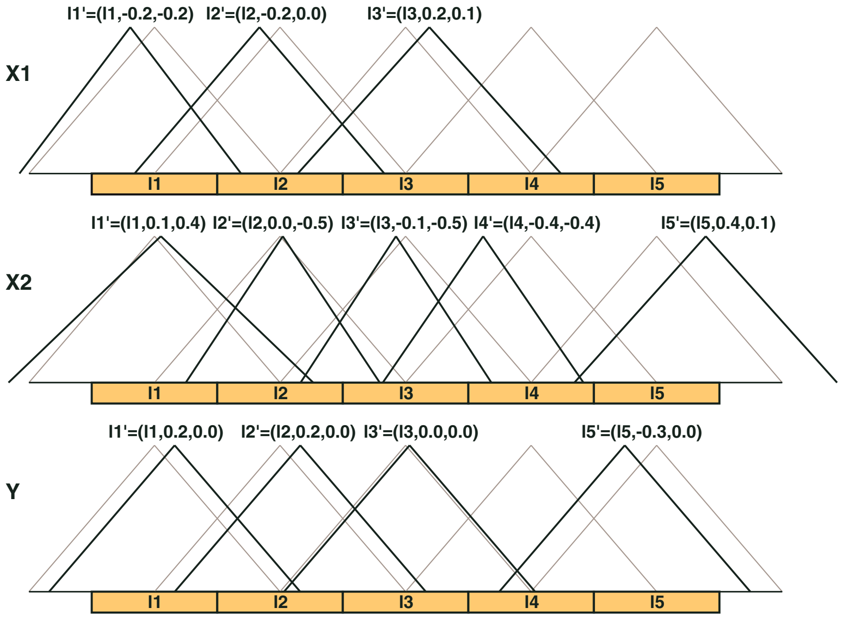
Fig. 6 **a** (X_1, Y) and (X_2, Y) dependency in the training data; **b** (X_1, Y) and (X_2, Y) dependency in the test data

Table 2 Results obtained in the line length estimation problem

Method	**#R**	**MSE**$_{tra}$	σ_{tra}	$t-test$	**MSE**$_{tst}$	σ_{tst}	$t-test$
WM	12.4	234712	32073	+	242147	24473	+
S	10.0	226135	19875	+	241883	19410	+
T	12.4	158662	6495	+	221613	29986	+
PAL	12.4	141638	4340	+	189279	19523	=
T+S	8.9	156313	2967	+	193477	49912	=
PAL+S	10.6	145712	5444	+	191922	16987	=
GL	12.4	166674	11480	+	189216	14743	=
LL	12.4	139189	3155	+	191604	18243	=
GL+S	9.0	160081	7316	+	189844	22448	=
LL+S	10.5	141446	3444	+	186746	15762	=
GLA	12.4	157604	9158	+	**185810**	18812	⋆
LLA	12.4	**133076**	4330	⋆	191945	16456	=
GLA+S	10.2	155404	9264	+	189472	20393	=
LLA+S	10.4	134541	5752	=	189057	20106	=

Analyzing the results presented in Table 2 we can point out the following conclusions:

– Considering these techniques, the RB is obtained by means of a generation method to learn few rules (12.4 from the 25 possible rules). It allows us to obtain a compact and accurate tuned model.
– In the test partitions, most of the methods show a similar error, although the GLA method obtains the best averaged results. This behavior is due to the noise and the strong non-linearities that this problem presents, which makes very difficult to avoid the over-learning, specially when we are trying to reduce the training error more and more.
– Regardless, in the training sets, the local LA-tuning obtains better results than the remaining approaches. The real achievement of the local LA-tuning is that the test error is not increased respect to



#R: 8 *MSE-tra: 163975.969* *MSE-tst: 188394.868*

X1	X2	Y
l1'	l1'	l1'
l1'	l4'	l2'
l1'	l5'	l3'
l2'	l1'	l1'
l2'	l2'	l2'
l2'	l4'	l3'
l3'	l2'	l2'
l3'	l3'	l5'

Fig. 7 Initial DB, tuned DB and RB of a model obtained with GLA+S in the low voltage line length estimation problem (the unused labels were removed from this figure)

the remaining approaches, despite of the very good results that this method obtains in training.

- Respect to the global approach, only small improvements have been obtained in training respect to the global lateral tuning. The short number of variables considered in this problem makes the search space not too complex for the global approach and therefore, there are not significant differences in the tuning methods.
- With regard to the combination of the tuning and the selection methods, GLA+S and LLA+S respectively show more or less the same accuracy than GLA and LLA considering a reduced number of rules and therefore reaching a better trade-off between interpretability and accuracy. Notice that, the combination of several techniques (specially when they act on different parts of the FRBS) increases the search space complexity. However, in the case of the global lateral tuning, the accuracy even improves when it is combined with the rule selection approach, which could indicate a better behavior of the global approach when the search space grows.

#R: 9 *MSE-tra: 139943.845* *MSE-tst: 169602.558*

X1	X2	Y
l1,-0.06,-0.09	**l1**,-0.26, 0.03	**l1**,-0.11
l1,-0.26,-0.30	**l3**,-0.04,-0.36	**l2**,-0.43
l1,-0.06,-0.34	**l5**,-0.49, 0.35	**l3**,-0.50
l2,-0.19,-0.09	**l1**, 0.10, 0.44	**l1**,-0.09
l2,-0.01, 0.11	**l2**,-0.11,-0.08	**l2**, 0.39
l2,-0.21, 0.07	**l4**,-0.43, 0.14	**l3**, 0.07
l3, 0.19, 0.47	**l2**, 0.31, 0.46	**l2**,-0.35
l3, 0.13,-0.02	**l3**,-0.05,-0.26	**l5**, 0.50
l5,-0.36,-0.15	**l3**, 0.04, 0.32	**l3**,-0.41

Fig. 8 RB and lateral displacements of a model obtained with LLA+S in the low voltage line length estimation problem

Figure 7 presents the evolved fuzzy linguistic partitions and the decision table obtained by GLA+S from one of the 30 runs performed in the first problem. On the other hand, Fig. 8 depicts one of the 30 RBs obtained with LLA+S in this problem, where we can see how the local tuning evolves each label of the different rules in a different way. The initial membership functions considered by this method are the same shown in Fig. 7 with grey color.

Table 3 Results obtained in the maintenance costs estimation problem

Method	#R	$\mathbf{MSE}_{tra}$	σ_{tra}	$t-test$	$\mathbf{MSE}_{tst}$	σ_{tst}	$t-test$
WM	65	57605	2841	+	57934	4733	+
S	40.8	41086	1322	+	59942	4931	+
T	65	18602	1211	+	22666	3386	+
PAL	65	10545	279	+	13973	1688	+
T+S	41.9	14987	391	+	18973	3772	+
PAL+S	57.4	12851	362	+	16854	1463	+
GL	65	23064	1479	+	25654	2611	+
LL	65	3664	390	+	5858	1798	+
GL+S	49.1	18801	2669	+	22586	3550	+
LL+S	58.0	3821	385	+	6339	2164	+
GLA	65	17950	1889	+	21212	2686	+
LLA	65	**2747**	282	⋆	**4540**	788	⋆
GLA+S	49.4	17538	2391	+	21491	4168	+
LLA+S	47.5	3404	433	+	5633	1452	+

To easy the graphical representation, in all these figures, the labels are named from ‘l1’ to ‘lL^i (with L^i being the number of labels of the i-th variable). Nevertheless, such labels would have associated a linguistic meaning determined by an expert. In this way, if the ‘l1’ label of the ‘X1’ variable represents ‘LOW’, ‘l1+0.11’ could be interpreted as ‘a little smaller than LOW’ (based on the expert opinion) or, as in the case of the classical tuning approach, it could be interpreted maintaining the original meaning of such label. It is the case of Fig. 7, where practically all the new labels could maintain their initial meanings or be easily renamed.

These figures show how small variations in the membership functions lead to important improvements in the behavior of the obtained FRBSs. Furthermore, the difficult trade-off between accuracy and complexity can be observed taking into account both, the global RB and the local RB (Fig. 7, 8). The accuracy can be improved considering a local approach but always at the expense of losing some interpretability.

5.3 Estimating the maintenance costs of medium voltage lines

Estimating the maintenance costs of the medium voltage electrical network in a town [10] is a complex but interesting problem. Since a direct measure is very difficult to obtain, the consideration of models becomes useful. These estimations allow electrical companies to justify their expenses. Moreover, the model must be able to explain how a specific value is computed for a certain town. Our objective will be to relate the *maintenance costs of medium voltage line* with the following four variables: *sum of the lengths of all streets in the town*, *total area of the town*, *area that is occupied by buildings*, and *energy supply to the town*. We will deal with estimations of minimum maintenance costs based on a model of the optimal electrical network for a town in a sample of 1,059 towns. Five partitions[1] considering an 80% (847) in training and a 20% (212) in test are considered for the experiments.

The results obtained in this problem by the analyzed methods are shown in Table 3 (these kinds of table was described in the previous subsection). Analyzing the results presented in Table 3 we can stress the following facts:

- The initial RBs are also comprised by few rules, 65 from the 625 possible rules, which is a reasonable number for this problem.
- The LA-tuning methods show an important reduction of the MSE respect to the classical methods (specially the LLA method) and reasonable improvements respect to the lateral tuning.
- The best results are obtained by the local approach, presenting a good relationship between the search space complexity and the results obtained, and getting a good trade-off between accuracy and local interpretability. Furthermore, since the lateral and amplitude variations are related to the original global labels, a global interpretation could be done in these terms.
- GLA+S and LLA+S respectively shows more or less the same accuracy than GLA and LLA considering a reduced number of rules. Notice that the LLAT+S method removes 10 more rules than LL+S.

Figures 9 and 10 depict one of the 30 KBs respectively obtained with GLA and GLA-S. Analyzing both linguistic models, we can see that in most of the cases,

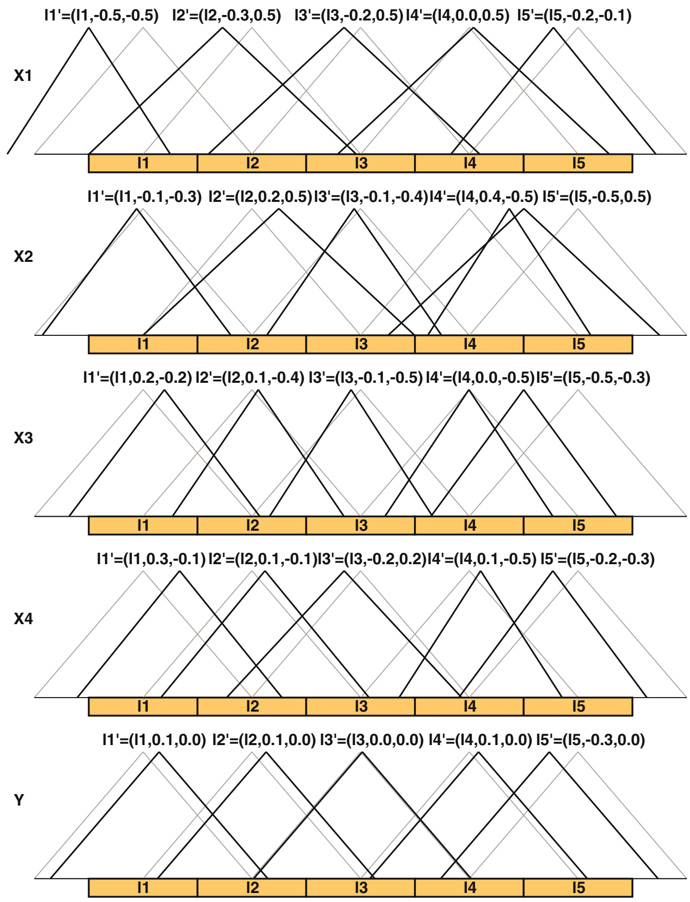


Fig. 9 Initial DB, tuned DB and RB of a model obtained with GLA in the maintenance costs estimation problem

#R: 65 *MSE-tra: 19636.934*
MSE-tst: 22378.074

X1	X2	X3	X4	Y
l1'	l1'	l1'	l1'	l1'
l1'	l1'	l1'	l2'	l2'
l1'	l2'	l1'	l1'	l1'
l1'	l2'	l1'	l2'	l2'
l1'	l2'	l2'	l1'	l1'
l1'	l2'	l2'	l2'	l2'
l2'	l1'	l1'	l1'	l1'
l2'	l1'	l1'	l2'	l2'
l2'	l1'	l2'	l1'	l1'
l2'	l1'	l2'	l2'	l2'

X1	X2	X3	X4	Y
l2'	l2'	l2'	l2'	l2'
l2'	l3'	l2'	l1'	l2'
l2'	l3'	l2'	l2'	l2'
l2'	l3'	l3'	l1'	l3'
l2'	l3'	l3'	l2'	l3'
l3'	l2'	l1'	l1'	l1'
l3'	l2'	l1'	l2'	l2'
l3'	l2'	l1'	l3'	l2'
l3'	l2'	l2'	l1'	l2'
l3'	l2'	l2'	l2'	l2'

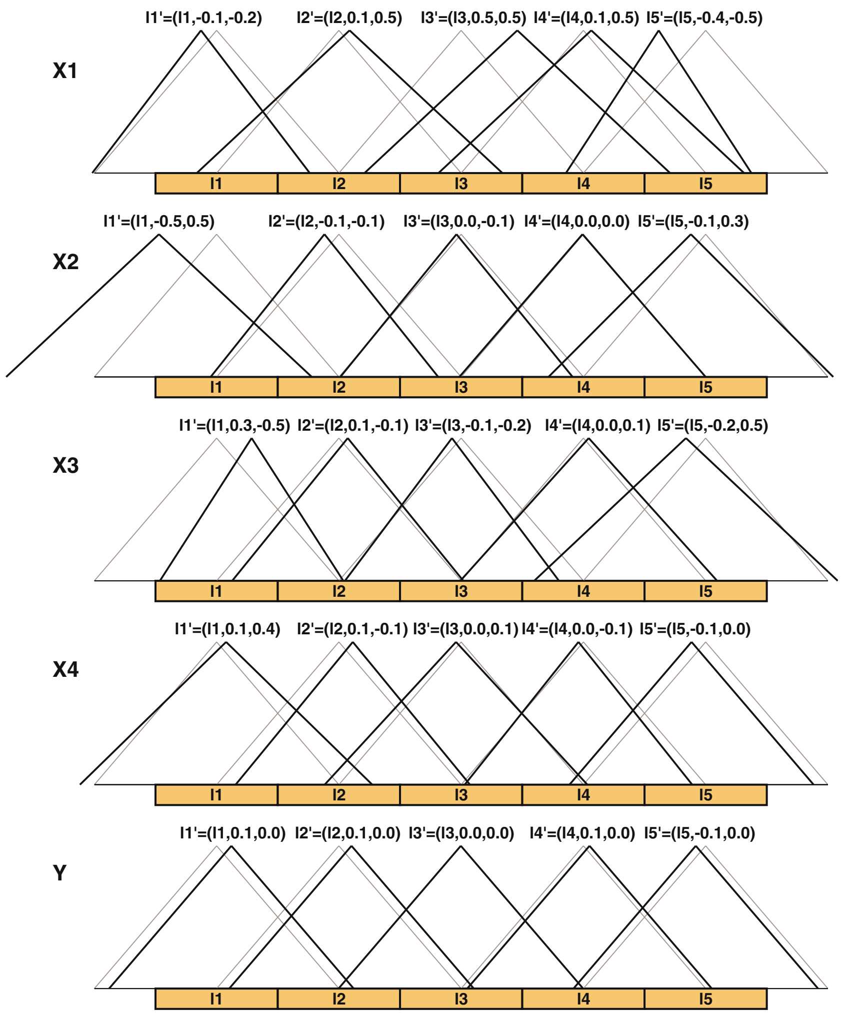
X1	X2	X3	X4	Y
l3'	l3'	l2'	l3'	l3'
l3'	l3'	l3'	l2'	l3'
l3'	l3'	l3'	l3'	l3'
l3'	l4'	l3'	l2'	l3'
l3'	l4'	l3'	l3'	l3'
l3'	l4'	l4'	l2'	l3'
l3'	l4'	l4'	l3'	l4'
l4'	l2'	l2'	l1'	l2'
l4'	l2'	l2'	l2'	l2'
l4'	l2'	l2'	l3'	l3'

X1	X2	X3	X4	Y
l4'	l3'	l2'	l3'	l3'
l4'	l3'	l2'	l4'	l3'
l4'	l3'	l3'	l2'	l3'
l4'	l3'	l3'	l3'	l4'
l4'	l3'	l3'	l4'	l4'
l4'	l4'	l3'	l1'	l3'
l4'	l4'	l3'	l2'	l3'
l4'	l4'	l3'	l3'	l4'
l4'	l4'	l3'	l4'	l4'
l4'	l4'	l4'	l2'	l4'
l5'	l4'	l3'	l5'	l5'
l4'	l5'	l4'	l2'	l3'

X1	X2	X3	X4	Y
l4'	l5'	l4'	l3'	l4'
l4'	l5'	l4'	l4'	l5'
l4'	l5'	l5'	l2'	l5'
l4'	l5'	l5'	l3'	l5'
l5'	l2'	l2'	l2'	l2'
l5'	l2'	l2'	l4'	l3'
l5'	l2'	l2'	l5'	l4'
l5'	l2'	l3'	l2'	l3'
l5'	l2'	l3'	l4'	l3'
l5'	l2'	l3'	l5'	l4'
l5'	l4'	l3'	l2'	l3'
l5'	l4'	l3'	l4'	l4'

X1	X2	X3	X4	Y
l2'	l2'	l1'	l1'	l1'
l2'	l2'	l1'	l2'	l2'
l2'	l2'	l2'	l1'	l2'
l3'	l2'	l2'	l3'	l3'
l3'	l3'	l2'	l1'	l2'
l3'	l3'	l2'	l2'	l3'
l4'	l2'	l2'	l4'	l3'
l4'	l3'	l2'	l1'	l2'
l4'	l3'	l2'	l2'	l3'
l4'	l4'	l4'	l3'	l4'
l4'	l4'	l4'	l4'	l5'

Fig. 10 Initial DB, tuned DB and RB of a model obtained with GLA+S in the maintenance costs estimation problem



#R: 50 *MSE-tra: 21657.293*
MSE-tst: 26021.017

X1	X2	X3	X4	Y	X1	X2	X3	X4	Y	X1	X2	X3	X4	Y
l1'	l1'	l1'	l1'	l1'	l2'	l3'	l3'	l2'	l3'	l4'	l2'	l2'	l2'	l2'
l1'	l1'	l1'	l2'	l2'	l3'	l2'	l1'	l2'	l2'	l4'	l2'	l2'	l3'	l3'
l1'	l2'	l1'	l1'	l1'	l3'	l2'	l1'	l3'	l2'	l4'	l2'	l2'	l4'	l3'
l1'	l2'	l2'	l2'	l2'	l3'	l2'	l2'	l1'	l2'	l4'	l3'	l2'	l1'	l2'
l2'	l1'	l1'	l1'	l1'	l3'	l2'	l2'	l2'	l2'	l4'	l3'	l2'	l2'	l3'
l2'	l1'	l1'	l2'	l2'	l3'	l2'	l2'	l3'	l3'	l4'	l3'	l2'	l3'	l3'
l2'	l1'	l2'	l1'	l1'	l3'	l3'	l2'	l1'	l2'	l4'	l3'	l2'	l4'	l3'
l2'	l1'	l2'	l2'	l2'	l3'	l3'	l2'	l2'	l2'	l4'	l3'	l3'	l2'	l3'
l2'	l2'	l1'	l1'	l1'	l3'	l3'	l2'	l3'	l3'	l4'	l3'	l3'	l3'	l4'
l2'	l2'	l1'	l2'	l2'	l3'	l3'	l3'	l2'	l3'	l4'	l4'	l3'	l2'	l3'
l2'	l2'	l2'	l1'	l2'	l3'	l4'	l3'	l2'	l3'	l4'	l4'	l3'	l4'	l4'
l2'	l2'	l2'	l2'	l2'	l3'	l4'	l3'	l3'	l3'	l4'	l4'	l4'	l2'	l4'
l2'	l3'	l3'	l1'	l3'	l3'	l4'	l4'	l3'	l4'	l4'	l4'	l4'	l3'	l4'
l4'	l4'	l4'	l4'	l5'	l4'	l5'	l5'	l2'	l5'	l5'	l2'	l3'	l4'	l3'
l4'	l5'	l4'	l2'	l3'	l4'	l5'	l5'	l3'	l5'	l5'	l2'	l3'	l5'	l4'
l4'	l5'	l4'	l3'	l4'	l5'	l2'	l2'	l5'	l4'	l5'	l4'	l3'	l2'	l3'
l5'	l4'	l3'	l4'	l4'	l5'	l4'	l3'	l5'	l5'					

the final tuned membership functions are very similar to the original ones, probably preserving their original meanings from an expert point of view.

In Fig. 9, only the second variable presents two labels whose meaning could change. This fact, does not occur in the case of GLA-S, Fig. 10, where only the first variable presents some membership functions with significant variations respect to the original ones. However, the corresponding labels could probably maintain their original meanings. It can be due to the existence of bad rules in the RB, which can provoke strong variations in the membership functions in order improve the behavior of such rules. Removing these kinds of rules can solve this problem.

5.4 Analysis on the use of the new inference system

In this section, we present a comparative study of the proposed methods with and without considering the new inference system to work with non-covered data. This study is only performed on the first problem presented in this section (estimation of the length of low voltage lines). This problem presents strong non-linearities, two noise points and a few number of training examples, which forces to the tuned linguistic models to sometimes not cover any test datum (a new case is given for the test in one of the five data partitions) or even to not cover any training datum (a noise point is given for training in one of the five data partitions). The two example data that are not covered are the two noise points presented in Table 4.

In the second problem, the example data are enough, well distributed and no noise points are presented. Therefore, in this problem, all the example data (in training and test) are finally covered by the obtained models. For this reason, an analysis on this problem has no sense.

Table 4 Non covered data

X_1	X_2	Y
320	880	3090
29	1200	3985

The results obtained are presented in Table 5. Classically, when an input pattern is not covered by any rule, a default output is given by the system. The two most used approaches are: to consider zero as default output (typical in fuzzy control problems) or to consider the mid of the output domain as default output. Both approaches are considered in our study to be shown in the table. In our problem, this mid value of the output domain is exactly 3797.5, that is very near to the desired output of the non-covered points presented in Table 4. This fact favors a better generalization ability of the methods considering this second approach. However, we though that this would not be a fair way to present our results, by which we developed the new inference system presented in Sect. 2.3 to better provide for non-covered data. In this way, the proposed methods could also work in other problems in which the mid value of the output domain does not coincide with the desired output of the non-covered data.

As can be seen in the results presented in Table 5, the default output considered has a great influence in the generalization ability of the obtained models. The new proposed inference seems to be an effective way to determine a good output value based on the information summarized in the final tuned rules. At this point, we must point out that all the methods considered for comparisons in Sects. 5.2 and 5.3 consider the mid of the output domain as default output (helping them to obtain better test errors) since practically all of them were proposed following this approach.

Table 5 Results obtained in the line length estimation problem

Method	**#R**	**MSE**$_{tra}$	σ_{tra}	**t-test**	**MSE**$_{tst}$	σ_{tst}	**t-test**
Zero as Default Output (0)							
GLA	12.4	158611	9091	+	217953	38780	+
LLA	12.4	**132815**	5715	⋆	201400	25034	+
GLA+S	11	157068	8253	+	228922	44799	+
LLA+S	11.1	136493	5133	=	212655	31089	+
The mid of the Output Domain as Default Output (3797.5)							
GLA	12.4	158284	9540	+	187879	18408	=
LLA	12.4	134003	3815	=	187568	15828	=
GLA+S	9.9	155927	6328	+	187820	20074	=
LLA+S	9.8	139047	6288	+	190678	20892	=
With the New Inference System							
GLA	12.4	157604	9158	+	**185810**	18812	⋆
LLA	12.4	133076	4330	=	191945	16456	=
GLA+S	10.2	155404	9264	+	189472	20393	=
LLA+S	10.4	134541	5752	=	189057	20106	=

6 Conclusion

In this work, we extend the lateral tuning of membership functions proposing a new post-processing method for the lateral and amplitude tuning of membership functions. This approach proposes a new symbolic representation with two parameters (α, β), respectively representing the lateral displacement and the amplitude variation of the support of a label.

In the following, we present our conclusions to subsequently present some further considerations:

- The linguistic 3-tuples based rule representation together with the proposed evolutionary tuning algorithm, provides a good mechanism to obtain accurate models, although it involves a slight lost of interpretability.
- In most of the cases, only small variations have been performed on the original membership functions, which maintains the interpretability to a reasonable level.
- The use of rule selection methods to reduce the number of rules together with the LA-tuning is a good approach to obtain more compact models with a similar accuracy. This combination increases the search space (tuning of the parameters + selection of the rules), which is easily handled thanks to the new rule representation.

Finally, we want point out that the use of different FRBS learning schemes considering the linguistic 3-tuples rule representation model could be a good approach to obtain more compact and precise models. In this way, our main interest for the future is to obtain whole KBs by means of the evolutionary learning of the DB a priori [12], using the linguistic 3-tuples rule representation and a basic rule generation method.

References

1. Alcalá R, Herrera F (2004) Genetic tuning on fuzzy systems based on the linguistic 2-tuples representation. In: Proceedings of the 13th IEEE international conference on fuzzy systems, vol 1, pp 233–238
2. Babuška R, Oosterhoff J, Oudshoorn A, Bruijn PM (2002) Fuzzy selftuning PI control of pH in fermentation, Eng Appl Artif Intell 15(1): 3–15
3. Bastian A (1994) How to handle the flexibility of linguistic variables with applications. Int J Uncertain, Fuzziness Knowl-Based Syst 3(4): 463-484
4. Casillas J, Cordón O, Herrera F, Magdalena L (2005) Accuracy improvements in linguistic fuzzy modeling. Stud Fuzziness Soft Comput 129: 385
5. Casillas J, Cordón O, del Jesus MJ, Herrera F (2005) Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction, IEEE Trans Fuzzy Syst 13(1): 13–29
6. Carmona P, Castro J, Zurita J (2004) Strategies to identify fuzzy rules directly from certainty degrees: a comparison and a proposal, IEEE Trans Fuzzy Sys 12(5): 631–640
7. Cordón O, Herrera F (1997) A three-stage evolutionary process for learning descriptive and approximate fuzzy logic controller knowledge bases from examples. Int J Approx Reason 17(4): 369–407
8. Cordón O, Herrera F, Peregrín A (1997) Applicability of the fuzzy operators in the design of fuzzy logic controllers. Fuzzy Sets Syst 86(1): 15–41
9. Cordón O, del Jesus MJ, Herrera F (1998) Genetic learning of fuzzy rule-based classification systems cooperating with fuzzy reasoning methods. Int J Intell Syst 13: 1025–1053
10. Cordón O, Herrera F, Sánchez L (1999) Solving electrical distribution problems using hybrid evolutionary data analysis techniques. Appl Intell 10: 5–24
11. Cordón O, Herrera F (2000) A proposal for improving the accuracy of linguistic modeling. IEEE Trans Fuzzy Syst 8(3): 335–344
12. Cordón O, Herrera F, Villar P (2001) Generating the knowledge base of a fuzzy rule-based system by the genetic learning of the data base. IEEE Trans Fuzzy Syst 9(4): 667–674
13. Cordón O, Herrera F, Hoffmann F, Magdalena L (2001) Genetic fuzzy systems: Evolutionary tuning and learning of fuzzy knowledge bases World Scientific, Singapore, 488
14. Cordón O, Gomide F, Herrera F, Hoffmann F, Magdalena L (2004) Ten years of genetic fuzzy systems: current framework and new trends. Fuzzy Sets Systems 41(1): 5–31
15. Eshelman LJ (1991) The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination. In: Foundations of genetic Algorithms 1, Rawlin G (Ed.) Morgan Kaufman, 265–283
16. Goldberg DE (1989) Genetic algorithms in search, optimization, and machine learning. Addison-Wesley, New York, P 372
17. Gómez-Skarmeta AF, Jiménez F (1999) Fuzzy modeling with hybrid systems, Fuzzy Sets Syste 104: 1991–208
18. González A, Pérez R (1999) A study about the inclusion of linguistic hedges in a fuzzy rule learning algorithm. Int J Uncertain Fuzziness Knowl-Based Syst 7(3): 257–266
19. Gürocak HB (1999) A genetic-algorithm-based method for tuning fuzzy logic controllers. Fuzzy Sets Syst 108(1): 39–47
20. Gürocak HB (2000) Tuning fuzzy logic controllers using response envelope method. Fuzzy Sets Syst 115(2): 287–304
21. Herrera F, Lozano M, Verdegay JL (1995) Tuning fuzzy logic controllers by genetic algorithms. Int J Approx Reason 12: 299–315
22. Herrera F, Lozano M, Verdegay JL (1998) A learning process for fuzzy control rules using genetic algorithms. Fuzzy Sets Syst 100: 143–158
23. Herrera F, Martínez L (2000) A 2-tuple fuzzy linguistic representation model for computing with words. IEEE Trans Fuzzy Syst 8(6): 746–752
24. Herrera F, Lozano M, Sánchez AM (2003) A taxonomy for the crossover operator for real-coded genetic algorithms: an experimental study. Int J Intell Syst 18: 309–338
25. Holland JH (1975) Adaptation in natural and artificial systems. (The University of Michigan Press, London, p 228
26. Hong TP, Lee CY (1999) Effect of merging order on performance of fuzzy induction. Intell Data Anal 3(2): 139–151
27. Ishibuchi H, Nozaki K, Yamamoto N, Tanaka H (1995) Selecting fuzzy if-then rules for classification problems using genetic algorithms. IEEE Trans Fuzzy Syst (3)3: 260–270

28. Nozaki K, Ishibuchi H, Tanaka H (1997) A simple but powerful heuristic method for generating fuzzy rules from numerical data. Fuzzy Sets Syst 86(3): 251–270
29. Ishibuchi H, Murata T, Türksen IB (1997) Single-objective and two objective genetic algorithms for selecting linguistic rules for pattern classification problems, Fuzzy Sets Syst 89(2): 135–150.
30. Ishibuchi H, Yamamoto T (2003) Trade-off between the number of fuzzy rules and their classification performance. In: Accuracy improvements in linguistic fuzzy modeling, Casillas J, Cordón O, Herrera F, Magdalena L, (eds) Springer Heidelberg, Germany, pp 72–99
31. Jang JSR (1993) ANFIS: adaptive network based fuzzy inference system, IEEE Trans Syst. Man Cyber 23(3): p 665–684.
32. Karr C (1991) Genetic algorithms for fuzzy controllers, AI Expert 6(2): 26–33
33. Klose A, Nurnberger A, Nauck D (1998) Some approaches to improve the interpretability of neuro-fuzzy classifiers. In: Proceeding of the 6th European congress on intelligent techniques and soft computing Vol 1, pp 629–633
34. Krone A, Krause H, Slawinski T (2001) A new rule reduction method for finding interpretable and small rule bases in high dimensional search spaces. In: Proceeding of the 9th IEEE international conference on fuzzy syst, Vol 2, pp 693–699
35. Krone A, Taeger H (2001) Data-based fuzzy rule test for fuzzy modelling. Fuzzy Sets Syst 123(3): 343–358
36. Liu BD, Chen CY, Tsao JY (2001) Design of adaptive fuzzy logic controller based on linguistic-hedge concepts and genetic algorithms. IEEE Trans Syst, Man Cybern. Part B: Cybern 31(1): 32–53
37. Mamdani EH (1974) Application of fuzzy algorithms for control of simple dynamic plant. Proc. IEEE 121(12): 1585–1588
38. Mamdani EH, Assilian S (1975) An experiment in linguistic synthesis with a fuzzy logic controller. Int J Man-Mach Stud 7: 1–13
39. Roubos JA, Setnes M (2001) Compact and transparent fuzzy models and classifiers through iterative complexity reduction. IEEE Trans on Fuzzy Syst 9(4): 516–524
40. Setnes M, Babuška R, Kaymak U, van Nauta Lemke HR (1998) Similarity measures in fuzzy rule base simplification. IEEE Trans Syst, Man Cyber - Part B: Cybern 28(3): 376–386
41. Setnes M, Roubos JA (2000) GA-fuzzy modeling and classification: complexity and performance. IEEE Trans Fuzzy Syst 8(5): 509–522
42. Wang LX, Mendel JM (1992) Generating fuzzy rules by learning from examples. IEEE Trans Syst, Man Cybern **22:6,** (1992) 1414–1427.
43. Yam Y, Baranyi P, Yang CT (1999) Reduction of fuzzy rule base via singular value decomposition. IEEE Trans Fuzzy Syst 7:120–132
44. Yen J, Wang L (1999) Simplifying fuzzy rule-based models using orthogonal transformation methods, IEEE Transa Syst, Man Cybern Part B: Cybern 29: 13–24
45. Zadeh LA (1965) Fuzzy sets. Info Control 8: 338–353
46. Zadeh LA (1973) Outline of a new approach to the analysis of complex systems and decision processes, IEEE Trans Syst, Man Cybern 3:28–44
47. Zadeh LA (1975) The concept of a linguistic variable and its applications to approximate reasoning, parts I, II and III. Inf Sci 8:199–249, 301–357, 9:43–80