Soft Comput (2006) 10: 717–734
DOI 10.1007/s00500-005-0002-1

FOCUS

Rafael Alcalá · Jesús Alcalá-Fdez · Jorge Casillas
Oscar Cordón · Francisco Herrera

Hybrid learning models to get the interpretability–accuracy trade-off in fuzzy modeling

Published online: 12 October 2005
© Springer-Verlag 2005

Abstract One of the problems associated to linguistic fuzzy modeling is its lack of accuracy when modeling some complex systems. To overcome this problem, many different possibilities of improving the accuracy of linguistic fuzzy modeling have been considered in the specialized literature. We will call these approaches as basic refinement approaches. In this work, we present a short study of how these basic approaches can be combined to obtain new hybrid approaches presenting a better trade-off between interpretability and accuracy. As an example of application of these kinds of systems, we analyze seven hybrid approaches to develop accurate and still interpretable fuzzy rule-based systems, which will be tested considering two real-world problems.

Keywords Linguistic fuzzy modeling · Interpretability-accuracy trade-off · Rule selection · Weighted linguistic rules · Tuning of membership functions · Genetic algorithms

1 Introduction

Fuzzy modeling (FM) – 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 with fuzzy predicates. 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, which use fuzzy rules composed of linguistic variables [52] 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 [45].

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 [6]. This drawback sometimes leads linguistic FM to 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 of improving the accuracy of linguistic FM while preserving its intrinsic interpretability have been considered in the specialized literature [10]. A great number of these approaches share the common idea of improving the way in which the linguistic fuzzy model performs the interpolative reasoning by inducing a better cooperation among the rules composing it. We will call these approaches as basic refinement approaches.

Recently, a new trend of research for parameter optimization and rule generation has arisen from this idea. It involves a smart combination of these basic approaches when they present complementary characteristics [1–5, 11, 19, 21, 29], searching for a better trade-off between interpretability and accuracy. The so obtained hybrid approaches usually present a better accuracy than the involved basic approaches and an acceptable interpretability very similar to that obtained by them.

In this paper, we present a short study on how the said basic refinement approaches can be combined to obtain new hybrid approaches presenting a better trade-off between interpretability and accuracy. In this way, as an example of application of these kinds of systems, we analyze seven hybrid approaches to develop accurate and still interpretable FRBSs.

Since these kinds of systems usually act on different parts of the FRBS structure, requiring different representations, the used learning techniques must be able to work with structures of different natures. Genetic algorithms (GAs) [34] can represent any type of fuzzy rules, present flexibility to work with different FRBS architectures and have a good capability to include expert knowledge [15]. The learning approaches analyzed in this work will be based on these kinds of techniques.

This work has been supported by the spanish cicyt project tic2002-04036-c05-01 (keel).

R. Alcalá (✉) · J. Alcalá-Fdez · J. Casillas · O. Cordón · F. Herrera
University of Granada, Department of Computer Science and Artificial Intelligence, 18071 Granada, Spain
E-mail: alcala@decsai.ugr.es; jalcala@decsai.ugr.es; casillas@decsai.ugr.es; ocordon@decsai.ugr.es; herrera@decsai.ugr.es

The paper is arranged as follows. First, a brief summary of different proposals to obtain a good balance between interpretability and accuracy is presented. Then, Sect. 3 briefly introduces the considered basic approaches and analyzes the positive synergy when some of them are combined. Section 4 presents seven different hybrid approaches to improve the balance between interpretability and accuracy. Experimental results are shown in Sect. 5 considering two real-world electrical problems. In Sect. 6, some concluding remarks are pointed out. Finally, the basic refinement approaches considered as starting point for the different hybrid approaches are described in Appendix A.

2 Interpretability–accuracy trade-off

Fuzzy modeling usually comes with two contradictory requirements to the obtained model: the *interpretability*, capability to express the behavior of the real system in an understandable way, and the *accuracy*, capability to faithfully represent the real system.

Of course, the ideal thing would be to satisfy both criteria to a high degree but, since they are contradictory issues, it is not generally possible. In that case, more priority is given to one of them (defined by the problem nature), leaving the other one in the background.

Two FM approaches arise depending on the main objective to be considered:

– *Linguistic FM*, mainly developed by means of linguistic (or Mamdani) FRBSs [33], which is focused on the interpretability.
– *Precise FM*, mainly developed by means of Takagi-Sugeno FRBSs [46], which is focused on the accuracy.

Regardless of the approach, a common scheme is followed to attain the desired balance between interpretability and accuracy (Fig. 1 graphically shows this operation mode):

1. Firstly, the main objective (interpretability or accuracy) is tackled defining a specific model structure to be used, thus setting the FM approach.
2. Then, the modeling components (the model structure and/or the modeling process) are improved by means of different mechanisms to compensate for the initial difference between both requirements. Thus, accuracy improvements are proposed in linguistic FM whilst interpretability improvements are proposed in precise FM.

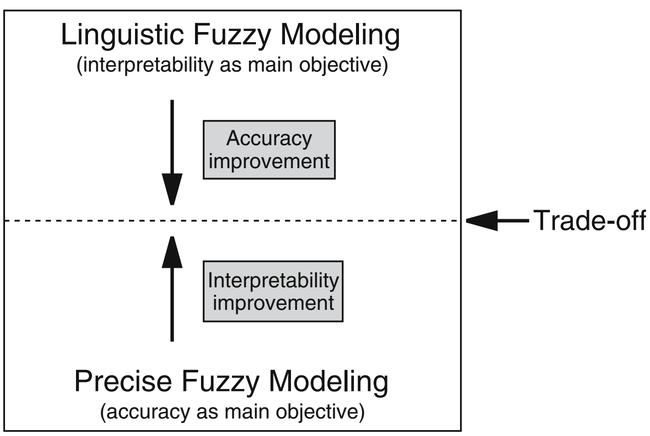


Fig. 1 Improvements of interpretability and accuracy in fuzzy modeling

Some examples found in the existing recent literature are shown as follows:

– *Linguistic FM with improved accuracy* [10] – This approach has been performed by learning/tuning the membership functions by defining their shapes [12,23,29,32, 35], their types (triangular, trapezoidal, etc.) [44], or their context (defining the whole semantic) [40], learning the granularity (number of linguistic terms) of the fuzzy partitions [19], or extending the model structure by using linguistic modifiers [17,22], weights (importance factors for each rule) [18,37], or hierarchical architectures (mixing rules with different granularities) [28], among others. Additionally, although rule base reduction [26,28] and input variable selection [25,31] processes improve the interpretability, they can also be seen as accuracy improvements when redundancy and inconsistency criteria are considered.
– *Precise FM with improved interpretability* [9] – This approach is usually developed by reducing the fuzzy rule set (usually with orthogonal transformations) [49,50], reducing the number of fuzzy sets (usually with similarity measures) with the subsequent merging of rules [41, 43], or exploiting the local description of the rules (basically smoothing the consequent polynomial function of the Takagi–Sugeno rule or isolating the fuzzy rule actions) [7,20,51].

It can be seen how this topic, the interpretability–accuracy trade-off, is a very important branch of research nowadays [9,10]. Our aim in this contribution will be to attain this desired balance by improving the accuracy in linguistic FM. To do so, fuzzy rules are generated by means of hybrid techniques combining several of the different approaches usually considered for linguistic FM with improved accuracy.

3 Hybridization of basic refinement approaches

Basically, two ways of improving the accuracy in linguistic FM can be considered by performing the improvement in:

– the *modeling process*, extending the model design to other components different from the rule base such as the data base or considering more sophisticated derivations of the linguistic fuzzy rules, or in
– the *model structure*, slightly changing the rule structure to make it more flexible.

The following subsections briefly introduce some of the improvements existing in the literature for designing the rule base or the data base with sophisticated methods, and for extending the model structure. Moreover, the basics to combine them will be introduced.

Table 1 Classification of the considered basic refinement approaches

	Extending the model design	**Extending the Rule Structure**
Rule base Design	Rule selection	Weighted rules
	Cooperative rule learning	Double-consequent rules
		Hierarchical knowledge bases
		Linguistic modifiers
Data base design	Tuning membership functions	Hierarchical knowledge bases

3.1 Basic approaches to improve the fuzzy rule-based system accuracy

The basic refinement approaches considered to be combined are presented in this subsection. A wide explanation of these approaches, together with the extended rule structure if it is modified, is included in Appendix A. Table 1 shows two possible classifications, depending on the improvement considered (modeling process or model structure) and depending on the FRBS part affected (data base, rule base or both). Both classifications are interesting to determine the existing synergy among the different approaches. Classifying according to the FRBS part affected, the following basic refinement approaches are considered in this contribution:

- *Approaches acting on the data base:*
  - Membership function tuning [12,23,35]: This approach, usually called data base tuning, involves refining the membership function shapes from a previous definition once the remaining FRBS components have been obtained. Another way to define the membership function shapes is to use more flexible alternative expressions for the membership functions to vary the compatibility degrees to the fuzzy sets [17,32].
- *Approaches acting on the rule base:*
  - Rule selection [12,28,30,48]: It involves obtaining an optimized subset of rules from a previous rule set by selecting some of them.
  - Rule cooperation [8]: This approach follows the primary objective of inducing a better cooperation among the linguistic rules. To do so, the rule base learning is guided by global criteria that jointly consider the action of the different rules. In [8], the cooperative rules (COR) methodology was proposed to induce a better cooperation among the fuzzy rules.
  - Weighted linguistic rule learning [18,37]: This approach considers an additional parameter for each rule that indicates its importance in the inference process, instead of considering all the rules equally important as in the usual case.
  - Double-consequent rule learning [13,36]: This approach allows the fuzzy rule set to present rules where each combination of antecedents may have two consequents associated when it is necessary.
  - Linguistic modifier learning [17,22]: A linguistic modifier is an operator that alters the membership functions of the fuzzy sets associated to the linguistic labels involved on each rule, giving a more or less precise definition as a result depending on the case. In this paper, we will consider an approach that fits the rule surface structure by using linguistic hedges.
- *Approaches acting on the whole knowledge base:*
  - Hierarchical linguistic rule learning [16,28]: This approach is devoted to produce a more general and well defined structure, the hierarchical knowledge base (HKB). In this way, to improve the system accuracy, fuzzy rules consider linguistic terms that are defined in linguistic fuzzy partitions with different granularity levels.

3.2 Positive synergy between the different approaches

As we can see, the previous approaches are not isolated and can be combined among them when they have complementary characteristics, improving even more the performance of the obtained FRBSs. However, it is important to remark that these hybridizations must be carefully made because of two main reasons:

1. Firstly, there is a need that the different approaches to be combined present complementary characteristics, and that this hybridization can be performed in an adequate way since, otherwise, the combination of techniques could worsen the accuracy of the obtained model instead of improving it.
2. Secondly, we must take into account that in linguistic FM the main objective is to improve the linguistic model accuracy *without losing its interpretability to a high degree*. Therefore, prior to the combination of different approaches, the interpretability presented by the final obtained model must be studied to determine if it will be adequate to solve the current problem.

Usually, these requirements are easily met when:

- The basic refinement approaches only extend the model design, maintaining the classical rule structure. Therefore, techniques as rule selection or rule cooperation present the ideal framework for hybridization, specially rule selection since this technique allows us to derive simpler FRBSs.
- They are applied to common parts of the FRBS structure. This is due to the following reasons:
  1. They search for the same objective.
  2. They can accomplish a complementary search among them, each of them helping the others to improve even more a part of the FRBS.
  3. If they only act on a part of the FRBS structure, the remaining components are fixed maintaining its original interpretability.

– The model structure extension provides useful, additional information about the system behavior (rules with more importance degree, space zones with more complexity, etc).
– The different approaches can be applied together, within the same learning process. This is due to the dependency among the different components of the FRBS and it is possible when the search space is not too large.

Taking into account the said requirements, we propose the following hybrid approaches to be analyzed:

– *Rule weight learning and rule selection*: (rule weights + rule selection). The FRBS structure is extended to consider the use of weights. This approach only acts on the rule base.
– *The weighted COR methodology*: (rule cooperation + rule weights + rule selection). The FRBS structure is extended to consider the use of weights. Again, this approach only acts on the rule base.
– *Weighted double-consequent rules*: (double-consequents + rule weights + rule selection). The FRBS structure is extended to consider the use of weights and double-consequent rules. This approach only acts on the rule base.
– *Weighted hierarchical rules*: (hierarchical rules + rule weights + rule selection). The FRBS structure is extended to consider the use of weights and hierarchical rules. This approach acts both on the rule base and the data base.
– *Data base tuning and linguistic modifiers*: (membership function tuning + non-linear scaling factors + linguistic hedges). The FRBS structure is extended to consider the use of non-linear scaling factors and linguistic hedges. This approach also acts on the rule base and the data base.
– *Rule selection, data base tuning and linguistic modifiers*: As the previous one but also considering rule selection.
– *COR learning, data base tuning and linguistic modifiers*: As the previous one, but also considering rule cooperation.

The following section describes the proposed hybrid approaches obtained combining the basic ones.

4 Seven different hybrid approaches

In this section, the different hybrid models proposed in the previous section are briefly presented. First, some common aspects are described. Then, four models considering the use of weighted rules are introduced (tuning at rule level). And finally, three models considering parameter optimization (tuning at data base level) and linguistic modifiers are reviewed.

4.1 Preliminaries: common aspects

All the studied methods can be intended as meta-methods over any other linguistic rule generation method, developed to obtain simpler linguistic fuzzy models improving the system accuracy. In this way, the corresponding algorithms are based on the optimization of an initial set of candidate rules obtained from automatic fuzzy rule learning methods or obtained from experts. In this work, we will consider the Wang and Mendel's (WM) method [47] or different extensions of this method as the initial linguistic rule generation method (although any other method could be applied). To derive this set of candidate linguistic rules, we will consider symmetrical fuzzy partitions of triangular-shaped membership functions (see Fig. 2).

On the other hand, to evaluate the different linguistic fuzzy models we will use the well-known mean square error (MSE):

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

with N being the data set size, $F(x^l)$ being the output obtained from the FRBS when the lth example is considered, and y^l being the known desired output.

4.2 Rule weight learning and rule selection

This approach was presented in [5]. As said, the hybridization of the rule weight derivation and the rule selection processes could result in important improvements of the system accuracy, obtaining simpler, and thus easily understandable, linguistic fuzzy models by removing unnecessary rules. In this way, the interpretability is maintained to an acceptable level. This method will be called as WM+WS in the experiments.

4.2.1 Learning scheme

To generate linguistic models combining both approaches, we may follow an operation mode composed of two steps:

1. Firstly, a preliminary fuzzy rule set is derived considering the Wang and Mendel's algorithm [47].
2. Then, after performing the first step, where an initial set of promising rules is generated, the two following tasks must be performed:
   – Genetic selection of a subset of rules presenting good cooperation.
   – Genetic derivation of the weights associated to these rules (see the weighted rule structure in Appendix A.3).

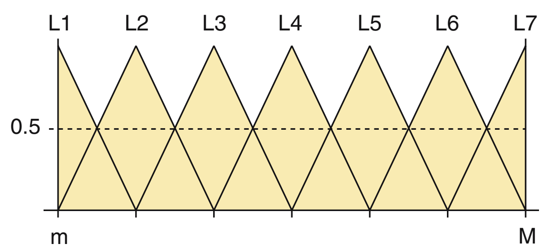


Fig. 2 Graphical representation of a possible fuzzy partition

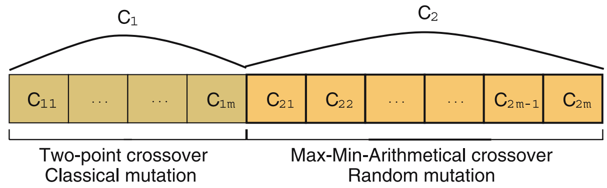


Fig. 3 Genetic representation and operators' application scope

To select the subset of rules with the best cooperation and the weights associated to them (second step), we will consider a GA coding all of them (rules and weights) in a chromosome. This algorithm is presented in the following subsection and will be called as WS.

4.2.2 *Step 2: Genetic weight derivation and rule selection process (WS)*

The proposed algorithm uses the classical *generational* scheme together with the Baker's stochastic universal sampling procedure and an elitist mechanism (that ensures to maintain the best individual of the previous generation) [34]. In this way, the convergence speed is significantly increased.

A double *coding scheme* ($C = C_1 + C_2$) for both rule selection and weight derivation is used:

- For the C_1 part, the coding scheme generates binary-coded strings of length m (number of single fuzzy rules in the previously derived rule set). Depending on whether a rule is selected or not, the alleles '1' or '0' will be respectively assigned to the corresponding gene.
- For the C_2 part, the coding scheme generates real-coded strings of length m. The value of each gene indicates the weight used in the corresponding rule. They may take any value in the interval [0, 1].

The *initial pool* is obtained with the first individual having all genes with value '1' in both parts, and the remaining individuals generated at random. In this way, the first individual represents the initial set of candidate rules.

As said, to evaluate the p-th chromosome we will use the MSE (see Sect. 4.1). In this case, $F(x^l)$ will be computed following the extended fuzzy reasoning method in order to consider the rule weights influence.

Due to the different nature of the chromosomes involved in the rule set definition process, different operators working on each part, C_1 and C_2, are required. Taking into account this aspect, the following operators are considered.

The *crossover* operator will depend on the chromosome part where it is applied: in the C_1 part, the standard two-point crossover is used, whilst in the C_2 part, the max–min-arithmetical crossover [24] is considered. In this case, eight offspring are generated by combining the two ones from the C_1 part (two-point crossover) with the four ones from the C_2 part (max–min-arithmetical crossover [24]). The two best offspring so obtained replace the two corresponding parents in the population.

As regards the *mutation* operator, it flips the gene value in C_1 and takes a value at random within the interval [0, 1] for the corresponding gene in C_2.

Figure 3 shows the application scope of these operators.

4.3 The WCOR methodology

In [3], we present the Weighted COR (WCOR) methodology, which includes the weight learning within the original COR methodology. The proposed method automatically learns the best consequent label, and its associated weight, corresponding to each possible antecedent combination in the problem space. This method that also performs rule selection will be called as WCOR in the experiments.

Since the only change in the classical model structure is the use of weights (see Appendix A.3), being used to improve the rule cooperation, the final obtained model preserves a good interpretability and presents a significantly improved accuracy, thus showing an appropriate trade-off between both requirements.

To learn the subset of rules with the best cooperation and the weights associated to them, different search techniques could be considered [39]. In this contribution, we will consider a GA for this purpose [15].

With this aim, we include the weight derivation within the original cooperative rule learning process. The following subsections present the WCOR methodology to obtain weighted cooperative rules.

4.3.1 *Learning scheme*

Since WCOR involves an extension of the original COR methodology, it consists of the following steps:

1. *Generate a candidate linguistic rule set*. This set will be formed by the rule best covering each example contained in the input–output data set. The structure of each rule, RC_l, is obtained by taking a specific example, e_l, and setting each one of the rule variables to the linguistic label associated to the fuzzy set best covering every example component.
2. *Obtain the antecedents* R_i^{ant} *of the rules composing the* FRBS *and a set of candidate consequents* $C_{R_i^{\text{ant}}}$ *associated to them*. Firstly, the rules are grouped according to their antecedents, determining each group a different antecedent R_i^{ant}. Then, for each group, a set of candidate consequents $C_{R_i^{\text{ant}}}$ is composed from the associated consequents, B_{k_i}, of the grouped rules.
3. *Perform a combinatorial search among the sets* $C_{R_i^{\text{ant}}}$ *looking for the combination of consequents and weights with the best cooperation*. For each rule R_i we have:
$$R_i^{\text{ant}}, \quad C_{R_i^{\text{ant}}}, \quad \text{and} \quad w_i \in [0, 1].$$

Since R_i^{ant} is kept fixed, the problem will involve determining the consequent and the weight associated to each rule. Two vectors of size m (number of rules finally obtained) are defined to represent this information, C_1 and C_2, where,

$$C_1[i] = k_i \mid B_{k_i} \in C_{R_i^{\text{ant}}}, \text{ and} \\ C_2[i] = w_i, \ \forall i \in \{1, \ldots, m\}, \quad (1)$$

and considering rule simplification, in which $B_{k_i} \in C_{R_i^{\text{ant}}} \cup \mathcal{N}$.

In this way, the C_1 part is an integer-valued vector in which each cell represents the index of the consequent used to build the corresponding rule. The C_2 part is a real-valued vector in which each cell represents the weight associated to this rule. Finally, a problem solution is represented as follows:

$C = C_1\ C_2$

Therefore, a search will be performed on the compound solution space to obtain the consequents and weights with the best cooperation. The main objective will be to minimize the MSE (see Sect. 4.1). To do that, the GA presented in the following section will be considered.

4.3.2 Genetic algorithm applied to the WCOR methodology

The said GA performs an approximate search among the candidate consequents with the main aim of selecting the set of consequents with the best cooperation and simultaneously learning the weights associated to the obtained rules. The main characteristics of the said algorithm are presented as follows:

- *Genetic approach*. An elitist generational GA with the Baker's stochastic universal sampling procedure.
- *Initial pool*. The initial pool is obtained by generating a possible combination at random for the C_1 part of each individual in the population. And for the C_2 part, it is obtained with an individual having all the genes with value '1', and the remaining individuals generated at random in [0, 1].
- *Crossover*. The standard two-point crossover in the C_1 part combined with the max–min-arithmetical crossover [24] in the C_2 part. Again, eight offspring are generated by combining the two ones from the C_1 part (two-point crossover) with the four ones from the C_2 part (max–min-arithmetical crossover [24]). The two best offspring so obtained replace the two corresponding parents in the population.
- *Mutation*. The operator considered in the C_1 part randomly selects a specific fuzzy subspace ($i \in \{1, \ldots, m\}$), at least containing two candidate consequents, and changes at random the current consequent k_i by other consequent k_i' such that $B_{k_i'} \in C_{R_i^{\text{ant}}}$ and $k_i' \neq k_i$. On the other hand, the selected gene in the C_2 part takes a random value within the interval [0, 1].

4.4 Weighted double-consequent rule learning

In [4], we propose the use of weighted double-consequent rules to design fuzzy linguistic models by means of a cooperative coevolutionary algorithm coevolving two species, the subset of rules best cooperating and the weights associated to them. In the experiments, this method will be called as DC+WS$_{CC}$.

With this aim, a more flexible linguistic model structure that combines the use of double-consequent and weighted rules was presented, thus having rules with the following structure:

IF X_1 is A_1 AND $\cdots$ AND X_n is A_n
THEN Y is $\{B_1, B_2\}$ $with$ $[w_1, w_2]$,

with w_1 and w_2 being the weights associated to the rules composed using the consequents B_1 and B_2, respectively. Hence, a weighted double-consequent rule can be seen as two weighted single-consequent rules with the same antecedent and different consequents.

4.4.1 Learning scheme

To generate linguistic models with this new structure, we may follow an operation mode based on the ALM methodology [13], but including the weight learning (in a similar way to the learning scheme presented in Sect. 4.2.1):

1. Firstly, two rules, the primary and secondary in importance, are obtained in each fuzzy input subspace considering an extension of the Wang and Mendel's algorithm [13]. When a fuzzy input subspace have two rules associated, both consequents are considered to compose a double-consequent linguistic fuzzy rule.
2. Then, after decomposing each double-consequent rule into two simple ones (obtaining an initial set of numerous promising rules), the two following tasks must be performed
   - Selection of a subset of cooperative rules.
   - Derivation of the weights for these rules.

Since two genes have to be considered in C_1 (rule selection) and C_2 (rule weight learning) for each weighted double-consequent rule, the search space is significantly increased respect to the original methodology, making the choice of the search technique considered crucial. To solve this problem, the search is accomplished considering an advanced optimization technique, the cooperative coevolution.

4.4.2 The cooperative coevolutionary algorithm

A coevolutionary algorithm [38] involves two or more species (populations) that permanently interact among them by a coupled fitness. Thereby, in spite of each species has its own coding scheme and reproduction operators, when an individual must be evaluated, its goodness will be calculated considering some individuals of the other species. This coevolution makes easier to find solutions to complex problems.

As we have seen, the problem that concerns us can be easily decomposed into two subtasks, the rule selection and the weight derivation. Therefore, it can be solved by coevolving two species cooperating to form the complete solution by learning a set of weighted rules. In the following subsections, the main characteristics of the proposed cooperative coevolutionary algorithm are presented.

Interaction scheme between species. The objective will be to minimize the well known MSE (see Sect. 4.1) considering that:

$$\mathrm{MSE}_{ij} = \frac{1}{2 \cdot N} \sum_{l=1}^{N} (F_{ij}(x^l) - y^l)^2 \ ,$$

with N being the number of training data, $F_{ij}(x^l)$ being the output inferred from the model obtained by combining the individuals i and j of the species 1 and 2 when the input x^l is presented, and y^l being the known desired output.

Thus, individuals in the species 1 and 2, are respectively evaluated with the fitness functions f_1 and f_2, defined as follows:

$$f_1(i) = \min_{j \in R_2 \cup P_2} \mathrm{MSE}_{ij} \text{ and } f_2(j) = \min_{i \in R_1 \cup P_1} \mathrm{MSE}_{ij} \ ,$$

with i and j being individuals of species 1 and 2, respectively, R_1 and R_2 being the set of the fittest individuals in the previous generation of the species 1 and 2, respectively, and P_1 and P_2 being individual sets selected at random from the previous generation of the species 1 and 2, respectively. The combined use of these kinds of sets make the algorithm have a trade-off between exploitation ($R_{1|2}$) and exploration ($P_{1|2}$). The cardinalities of the sets $R_{1|2}$ and $P_{1|2}$ are previously defined by the designer.

Species 1: Fuzzy rule selection. For the species 1, we will use the genetic rule selection method considered in [13]. The *coding scheme* generates binary-coded strings of length m (number of single-consequent rules in the previously derived rule set). Depending on whether a rule is selected or not, the alleles "1" or "0" will be respectively assigned to the corresponding gene. Thus, a chromosome C_1^p will be a binary vector representing the subset of rules finally obtained.

The whole *initial pool* is generated at random but one individual, which represents the complete previously obtained rule set. For this species, the standard two-point *crossover* operator is used. As regards the *mutation* operator, it flips the value of the gene.

Species 2: Weight derivation. The *coding scheme* generates real-coded strings of length m. The value of each gene indicates the weight used in the corresponding rule. They may take any value in the interval [0, 1]. Now, a chromosome C_2^p will be a real-valued vector representing the weights associated to the fuzzy rules considered.

The *initial pool* for this species is generated with a chromosome having all the genes with the value "1", and the remaining individuals taking values randomly generated within the interval [0, 1]. The max–min-arithmetical *crossover* operator [24] is considered. As regards the *mutation* operator, it simply involves changing the value of the selected gene by other value obtained at random within the interval [0, 1].

4.5 Weighted hierarchical rule learning

In [2], the hybridization of both hierarchical and weighted linguistic fuzzy rules to derive hierarchical systems of weighted linguistic rules (HSWLRs) is presented. In this work, the structure and inference system of the HSWLR, and a GA-based process to learn these kinds of improved linguistic models, are proposed. This method that also performs rule selection will be called HSWLR in the experiments.

4.5.1 Weighted hierarchical knowledge base

An HKB is composed of a set of layers, and each layer is defined by its components in the following way:

$$\mathrm{layer}(t, n(t)) = \mathrm{DB}(t, n(t)) + \mathrm{RB}(t, n(t)),$$

with $n(t)$ being the number of linguistic terms in the fuzzy partitions of layer t, $\mathrm{DB}(t, n(t))$ being the data base which contains the linguistic partitions with granularity level $n(t)$ of layer t (t-linguistic partitions), and $\mathrm{RB}(t, n(t))$ being the rule base formed by those linguistic rules whose linguistic variables take values in the former partitions (t-linguistic rules). For simplicity in the learning methodology, the following notation equivalences are established: $\mathrm{DB}(t, n(t)) \equiv \mathrm{DB}^t$ and $\mathrm{RB}(t, n(t)) \equiv \mathrm{RB}^t$.

The number of linguistic terms in the partitions of layer t will be defined in the following way (using strong fuzzy partitions):

$$n(t) = (n(1) - 1) \cdot 2^{t-1} + 1$$

The main purpose of developing an hierarchical rule base (HRB) is to model the problem space in a more accurate way. To do so, those linguistic rules from $\mathrm{RB}(t, n(t))$ that model a subspace with bad performance are expanded into a set of more specific linguistic rules, which become their image in $RB(t+1, 2 \cdot n(t) - 1)$. This set of rules models the same subspace that the former one and replaces it. As a consequence of the previous definitions, we could now define the Weighted HKB (WHKB) as the union of every layer t:

$$\mathrm{WHKB} = \cup_t \mathrm{layer}(t, n(t)) + \cup_t W^t,$$

with W^t being the set of weights associated to the rules from layer t. We should notice that these weights are obtained over the whole HRB (and not over the isolated layers) since they must consider the way in which all the rules interact. Therefore, the fuzzy reasoning must be extended as in the case of weighted linguistic rules, considering the matching degree of the rules fired (see Appendix A.3). However, the hybridization of the said approaches to derive two-layer HSWLRs maintains a good interpretability level.

4.5.2 A Two-level HSWLR Learning Algorithm

In this subsection, we present the two-level HSWLR-LM to generate two-layer WHKBs [16]. To do so, we use Wang and Mendel's method —WM— [47] based on the existence of a set of input-output training data $E = \{e^1, ..., e^l, ..., e^q\}$ with $e^l = (ex_1^l, \ldots, ex_m^l, ey^l)$, and a previously defined DB^1. The measure of error used in the algorithm is the MSE(E, RB) (see Sect. 4.1). The algorithm basically consists of the following steps:

1. RB^1 *generation.* Generate the rules from DB^1 by means of the WM method: RB^1 = WM(DB^1, E).
2. RB^2 *generation*. Generate RB^2 from RB^1, DB^1 and DB^2.
   (a) *Calculate the error of* RB^1: MSE $\left(E, \mathrm{RB}^1\right)$.
   (b) *Ditto for each* 1*-linguistic rule:* $\mathrm{MSE}\left(E_i, R_i^{n(1)}\right)$, with E_i being a set of the examples matching the ith rule antecedents to degree $\tau \in (0, 1]$.
   (c) *Select the* 1*-linguistic rules with bad performance which will be expanded:*[1]
   $IF\ MSE(E_i, R_i^{n(1)}) \geq \alpha \cdot MSE(E, RB^1) THEN$
   $R_i^{n(1)} \in RB^1_{bad}$
   $ELSE$
   $R_i^{n(1)} \in RB^1_{good}$.
   (d) *Create*[2] DB^2*:* $DB^2_{x_j}$ and DB^2_y.
   (e) *Select the 2-linguistic partition terms from DB^2 that δ-intersect the ones of the bad performance 1-linguistic rules:* $I(R_i^{n(1)}),\ \forall\ R_i^{n(1)} \in RB^1_{bad}$, where $\delta \in [0, 1]$ is a cross level of *"significant intersection"*.
   (f) *Extract a candidate set of L 2-linguistic rules:*
   $$CLR(R_i^{n(1)}) = WM(I(R_i^{n(1)}), E_i) = \left\{R_{i_1}^{2 \cdot n(1)-1}, ..., R_{i_L}^{2 \cdot n(1)-1}\right\}.$$
3. *Summarization. Obtain a joined set of candidate linguistic rules (JCLR),* performing the union of the group of the new generated *2-linguistic rules* and the former good performance *1-linguistic rules:*
   $$\mathrm{JCLR} = \mathrm{RB}^1_{\mathrm{good}} \cup (\cup_i \mathrm{CLR}(R_i^{n(1)})),\ \ R_i^{n(1)} \in RB^1_{\mathrm{bad}}.$$
   More than one copy of a rule in the same layer can be produced as a consequence of the generation process (Steps 1, 2 and 3). This fact can be interpreted as a *weight* on that rule by using the extended fuzzy reasoning model presented in Appendix A.3 with w_i being the number of times that the ith rule is repeated. In this work, we will consider this approach (which theoretically should obtain equivalent models with the same accuracy level). To do so, repeated rules are excluded of the HKB by obtaining an equivalent HSWLR without them:
   $$\mathrm{WHRB} = (\text{Extract repeated}(\mathrm{JCLR}) + \text{Weights}).$$
   To obtain an equivalent system without repeated rules, we maintain a single instance for each rule, with w_i being the sum of the weights of the corresponding repeated rules. Other equivalent HSWLR could be found with $w_i \in [0, 1]$ by means of a normalization process over the weights.
4. *Genetic weight derivation and rule selection process. Simplify the WHRB by removing the unnecessary rules from it and learning the weights associated to those rules to obtain a* WHRB *with good cooperation.* To do so, the algorithm presented in Sect. 4.2.2 is used considering the inclusion of the initial weights in the first individual and considering the following fitness function (penalizing an excessive number of rules):
   $$F(C^j) = w_1 \cdot \mathrm{MSE}(C^j) + w_2 \cdot N^j_{\mathrm{rules}}$$
   with N^j_{rules} being the number of rules of that WHRB, and with w_1 and w_2 being weighting coefficients defining the relative importance of each objective. In the present experiments, these coefficients are initialized as follows:
   $$w_1 = 1.0; \quad w_2 = 0.1 \cdot \frac{\mathrm{MSE}_{\mathrm{initial}}}{N_{\mathrm{initial\ rules}}}$$

[1] The expansion factor α may be adapted in order to have more or less expanded rules.

[2] DB^t is referred to as $\mathrm{DB}^t_{x_j}$ ($j = 1, ..., m$), meaning that it contains the *t-linguistic partition* where the input variable x_j takes values, and as DB^t_y for the output variable y.

4.6 Data base tuning and linguistic modifiers learning

A tuning process based on GAs was introduced in [11] to jointly fitting the membership functions by changing their basic and additional parameters and fitting the rule surface structure using linguistic hedges. The tuning involves starting from a previous knowledge base (rule base + data base) derived by the Wang and Mendel's algorithm [47]. This method will be called as WM+PAL in the experiments.

The following sections introduce the genetic process and its components.

4.6.1 Genetic process

This proposal of FRBS genetic tuning is characterized as follows:

1. The objective (fitness function) will be to minimize the well-known MSE (see Sect. 4.1).
2. A threefold coding scheme ($C_P + C_A + C_L$) is used. C_P will encode the basic membership function parameters, C_A the α membership function parameters (i.e., the non-linear scaling factors), and C_L the linguistic hedges included in the different rules. Therefore, C_P and C_A are used to tune the semantics of the deep structures and C_L to adjust the surface structures. Fig. 4 graphically shows such a scheme.
   – For the C_P *part*, a 3-tuple of real values for each triangular membership function is used, thus being the data base encoded into a real-coded chromosome built by joining the membership functions involved in each variable fuzzy partition. A variation interval is defined for each membership function basic parameter. It will be discussed in next subsection.

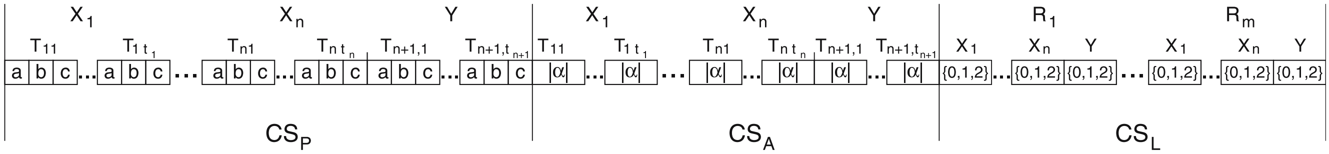


Fig. 4 Coding scheme for tuning Fuzzy rule-based systems (FRBSs) with n being the number of input variables, T_{ij} being the jth linguistic term of the ith variable (with $n+1$ being the output variable), t_i the number of linguistic terms of the ith variable, and m the number of linguistic fuzzy rules

– For the C_A *part*, a real-coded chromosome that encodes the value of the additional parameter α for each membership function is used. Each gene can take any value in the interval $[-1, 1]$ with the following mapping between alleles and actual value:

$$c_{ij}^{A} \in [-1, 0] \longleftrightarrow \alpha \in [0, 1],$$
$$c_{ij}^{A} \in [0, 1] \longleftrightarrow \alpha \in [1, 5],$$

with c_{ij}^{A} being the gene associated to the membership function for the jth linguistic term of the ith variable.

– For the C_L *part*, the coding scheme generates integer-coded strings of length $m \cdot (n+1)$ (with m being the number of rules and n being the number of input variables). Each gene can take any value in the set $\{0, 1, 2\}$ with the following correspondece to the linguistic hedge used:

$$c_{ij}^{L} = 0 \longleftrightarrow \quad \text{the “very” linguistic hedge,}$$
$$c_{ij}^{L} = 1 \longleftrightarrow \quad \text{no linguistic hedge,}$$
$$c_{ij}^{L} = 2 \longleftrightarrow \quad \text{the “more-or-less” linguistic hedge,}$$

with c_{ij}^{L} being the gene associated to the linguistic term used in the jth variable of the ith rule.

4.6.2 Genetic components

The genetic tuning method has the following components:

– When generating the initial population, some of the original information in the initial knowledge base will be mixed up with random values.
To include the original values in the C_P part, the actual values will be directly included.
For the C_A part, the original values will depend on whether these parameters were used in the initial knowledge base or not. If so, the α parameters will be encoded following the said scheme; if not, the allele 0 (which means $\alpha = 1$) will be used.
For the C_L part, the modifiers used in the initial knowledge base are encoded with the said scheme. If no linguistic hedges were previously considered, alleles 1 will be used.
The following four steps are considered to initialize the population:

1. A chromosome that represents the initial data base and rule set is included. Therefore, genes in C_P, C_A, and C_L parts will directly encode the values corresponding to the original knowledge base.
2. A third of the population is generated with the C_P part at random (within the variation interval for each gene) whilst the alleles in C_A and C_L will encode the original values.
3. Another third of the population is generated with original values in the C_P, alleles at random (within the interval $[-1, 1]$) in the C_A, and original values in the C_L part.
4. The remaining chromosomes are generated with the original values of the data base in the C_P and C_A parts, and alleles at random (within the set $\{0, 1, 2\}$) in the C_L part.

The crossover operator will depend on the chromosome part where it is applied:

– In C_P and C_A *parts*, the max-min-arithmetical crossover is considered [24].
– In the C_L *part*, the standard two-point crossover is used.

After recombining each part, the two best chromosomes among the eight *(four* different C_P and C_A parts combined with *two* different C_L parts) descendants obtained will be selected to replace their parents.

The mutation operator will also depend on the chromosome part where it is applied:

– In C_P and C_A *parts*, the Michalewicz's non-uniform mutation operator within the interval allowed for each gene is considered.
– In the C_L *part*, the mutation operator changes the gene to the allele 1 when a gene with alleles 0 or 2 must be mutated, and randomly to 0 or 2 when a gene with allele 1 must be mutated.

A generational GA with the Baker's stochastic universal sampling procedure together with elitism is considered.

4.7 Rule selection, data base tuning and linguistic modifiers

An extension of the previous approach is also presented in [11]. In this case, the previous hybrid approach is combined with a rule selection by allowing an additional binary vector C_S that determines when a rule is selected or not (alleles ‘1’ and ‘0’, respectively). The remaining components are considered as in the previous case. This method will be called as WM+PALS in the experiments.

4.8 COR learning, data base tuning and linguistic modifiers

In [1], three different mechanisms to improve the accuracy of linguistic FM are jointly considered: COR (linguistic rule set learning with COR), PA (learning the membership function parameters and non-linear scaling factors), and L (learning the linguistic hedges used for each linguistic variable in each linguistic rule) improvements. This method will be called as CORPAL in the experiments (and also performs rule selection).

To develop these hybridizations, two main mechanisms are considered. On the one hand, we may distinguish between sequential or simultaneous learning. When several components of the FRBS are designed, we may opt to make a *sequential learning* by dividing it into two or more stages, each of them performing a partial or complete derivation of the linguistic models. Other possibility is to consider a *simultaneous learning* that directly obtain the whole model.

Different combinations are regarded by differentiating between sequential or simultaneous learning and between basic GAs or cooperative coevolution. Among all the possible combinations we will exclusively consider the simultaneous learning with simple GAs. It involves a process to learn both fuzzy rules and membership functions by including the three improvement mechanisms in a single chromosome.

As in the previous subsection, this approach can be considered as an extension of that presented in Sect. 4.6, by allowing an additional integer vector C_C that represents the index of the consequent used to build the corresponding rule (C_1 part of the algorithm presented in Sect. 4.3.1 considering the same operators).

5 Experiments and analysis of results

To evaluate the goodness of the proposed techniques, several experiments have been carried out considering two real-world problems [14]. The first of them presents *strong non-linearities* and the second considers four input variables, and therefore *a large search space*. In order to see the advantages of the combined action of the studied basic approaches, two different studies have been performed: only considering the basic approaches to improve the system accuracy and considering the different hybrid approaches to combine them (the approaches reviewed in this work). Table 2 shows a short description of the methods considered for this study.

From now on, any reference to an application of these methods is represented by the following expression:

method(r[, q])

with r (and q in the case of the hierarchy-based methods) being the granularity level of the linguistic partitions used in the method. At this point we should remark that, the main aim of this paper is to analyze the good cooperation of the basic improved approaches when they are combined, and not to establish a competitive analysis of the said hybrid approaches. Therefore, although the granularity considered for the studied approaches is the same in practically all of them, two of them (WM+WS and DC+WS$_{CC}$) consider a higher granularity to solve the first problem since the obtained improvements are better in that case.

With respect to the fuzzy reasoning method used, we have selected the *minimum t-norm* playing the role of the implication and conjunctive operators, and the *center of gravity weighted by the matching* strategy acting as the defuzzification operator.

Finally, the following values have been considered for the parameters of each method:[3]

- Genetic approaches: 61 individuals, 2,000 generations, 0.6 as crossover probability, 0.2 as mutation probability per chromosome, and 0.35 for the a factor in the max–min-arithmetical crossover.
- Hierarchical Generation: 0.1 as $\delta - (2 \cdot n - 1)$*-linguistic partition* terms selector, 0.5 as τ: used to calculate E_i, and 1.1 as α: used to decide the expansion of a rule.

5.1 Description of the problems considered

As said, two different real-world problems [14] are solved considering the proposed hybrid approaches. Both problems will be introduced in the following.

5.1.1 PROBLEM I: 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 cannot 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.

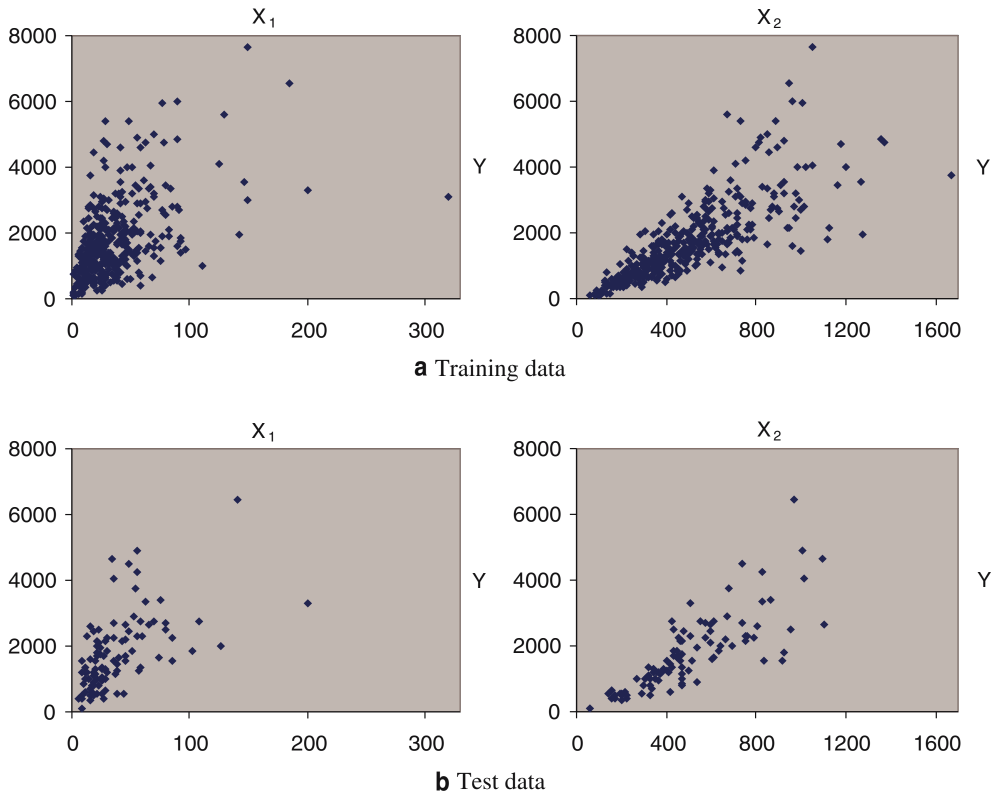
Due to this reason, the length of low voltage lines cannot be directly computed. Therefore, it must be estimated by means of indirect models. The problem involves relating *the length of low voltage line of a certain village* with

[3] 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 generations in order to allow the compared algorithms to achieve an appropriate convergence. No significant changes were achieved by increasing that number of generations.

Table 2 Methods considered for comparison

Method	Ref.	Sect.	Description
			CLASSICAL SIMPLE APPROACH
WM	[47]	—	A well-known ad hoc data-driven method to obtain classical rules
			BASIC REFINEMENT APPROACHES
WM+S	[5]	—	(Rule selection) WM+WS C_1 part
WM+W	[5]	—	(Rule weight learning) WM+WS C_1 part
COR	[8]	—	(Rule cooperation + rule selection) WCOR C_1 part
DC+S, ALM	[13]	—	(double-consequents + rule selection) ALM algorithm
HSLR	[16]	—	(hierarchical rules + rule selection) HSLR algorithm
WM+P	[11]	—	(MF tuning) WM+PAL C_P part
WM+AL	[11]	—	(NL scaling factors + linguistic hedges) WM+PAL C_A and C_L parts
			HYBRID IMPROVED APPROACHES
WM+WS	[5]	4.2	(Rule weights + rule selection)
WCOR	[3]	4.3	(Rule cooperation + rule weights + rule selection)
DC+WS$_{CC}$	[4]	4.4	(Double-consequents + rule weights + rule selection)
HSWLR	[2]	4.5	(Hierarchical rules + rule weights + rule selection)
WM+PAL	[11]	4.6	(MF tuning + NL scaling factors + linguistic hedges)
WM+PALS	[11]	4.7	(MF tuning + NL scaling factors + linguistic hedges + rule selection)
CORPAL	[1]	4.8	(MF tuning + NL scaling factors + linguistic hedges + rule cooperation + rule selection)

MF Membership functions; *NL* Non-linear



a Training data

b Test data

Fig. 5 **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

the following two variables: *the radius of the village* and *the number of users in the village* [14]. 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 farthest clients in a sample of 495 rural nuclei.

In order to evaluate the models obtained from the different methods considered in this paper, this sample has been randomly divided into two subsets, the training set with 396 elements and the test set with 99 elements, 80 and 20%, respectively. The existing dependency of the two input variables with the output variable in the training and test data sets is shown in Fig. 5 (notice that they present strong non-linearities). Both data sets considered are available at `http://decsai.ugr.es/~casillas/fmlib/`.

5.1.2 PROBLEM II: Estimating the maintenance costs of medium voltage lines

Estimating the maintenance costs of the optimal installation of medium voltage electrical network in a town [14] is another interesting electrical problem. Clearly, it is impossible to obtain this value by directly measuring it, since the medium voltage lines existing in a town have been installed incrementally, according to its own electrical needs in each moment. In this case, the consideration of models becomes the only possible solution. Moreover, the model must be able to explain how a specific value is computed for a certain town. These estimations allow electrical companies to justify their expenses. 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.

To develop the different experiments in this contribution, the sample has been randomly divided into two subsets, the training and test ones, with an 80%-20% of the original size respectively. Thus, the training set contains 847 elements, whilst the test one is composed of 212 elements. These data sets used are available at `http://decsai.ugr.es/~casillas/fmlib/`.

5.2 Results and analysis

The results obtained by the methods analyzed are shown in Table 3 for both problems, where $\#R$ stands for the number of rules, and $\mathrm{MSE}_{\mathrm{tra}}$ and $\mathrm{MSE}_{\mathrm{tst}}$ for the error obtained over the training and test data, respectively. The best results are shown in boldface for each problem. These results were obtained for an AMD K7 (Athlon) with clock rate of 1500 MHz and 256 MB of main memory. The run times for the different algorithms in the first problem do not exceeded 20 min and, in the second problem, they never exceed 60 min.

In the following paragraphs we perform an analysis of the results from the accuracy and simplicity point of view. This analysis will be made for each approach respect to its related basic approaches. In this way, Table 4 presents the improvement rates obtained by the studied hybrid methods respect to their related basic ones in the said problems. This information will help us to analyze the behavior of the proposed approaches, showing how they get the desired trade-off between interpretability (simplicity) and accuracy (inexistence of error).

Taking into account the information presented in Tables 3 and 4, it seems that those hybrid approaches based on tuning at rule level (weights combined with rule selection) learns the simplest models in both problems, specially in problem II which presents a higher dimensionality. Furthermore, these approaches show an appropriate balance between approximation and generalization. Methods performing tuning at rule level are:

(a) WM-WS: This method presents interesting improvements in both precision and simplicity respect to the basic WM. It shows a similar precision to WM-W but a similar simplicity to WM-S, inheriting the best characteristics of both approaches (obtaining accurate but simpler models).
(b) WCOR: The results obtained significantly improve those from WM, WM+W and COR, even reducing the number of rules respect to them. It is due to the use of weights that indicates the appropriate interaction level among rules, provoking slight changes in the consequents respect to the original COR methodology and improving the cooperation among the rules so obtained.
(c) DC+WS$_{CC}$: In this case, the results obtained are better for the second problem, overfitting the model obtained for the first problem. However, also in this case (problem I), the model obtained improves those obtained from WM and WM+W, reducing the number of rules in these models. Respect to DC+S, similar results are presented in the first problem (with only one more rule) but significant improvements are obtained in the case of the second problem. Notice that, the obtained FRBS only contains four and five double-consequent weighted rules (problems I and II), maintaining the desired interpretability.
(d) HSWLR: In the first problem, this approach presents the best trade-off between accuracy and interpretability, obtaining only nine rules and presenting the most accurate results. In the second problem, without considering those approaches performing tuning of the data base (WM+PAL, WM+PALS and CORPAL), the best results are obtained by HSWLR, which would present the best trade-off between interpretability (with only 47 rules) and accuracy (about a 30% better than WM-W and similar results to HSLR).

A graphical representation of the decision table of the FRBS obtained by HSWLR is presented in Fig. 6. Each cell of the table represents a fuzzy subspace and contains its associated output consequent(s), i.e., the correspondent label(s) together with its(their) respective rounded rule weight(s). The *absolute importance weight* for each fuzzy rule has been graphically showed by means of the grey color scale, from black (1.0) to white (0.0). In this way, we can easily see the importance of a rule with respect to their neighbor ones which could help the system experts to identify important rules.

On the other hand, those approaches based on the data base tuning get the fittest models in terms of approximation but, in some cases, the obtained models are overfitted (as in the case of the first problem) obtaining poor results. However, even in these cases, they still improve the behavior of their related basic approaches and present the best results in the second problem. Methods performing data base tuning are:

(a) WM+PAL: This method gets very good results respect to WM and its related basic approaches WM+P and WM+AL

Table 3 Results obtained in the electrical estimation problems

Problem I: low voltage lines					Problem II: medium voltage lines				
Method	**#R**	← (S+DC)	**MSE$_{tra}$**	**MSE$_{tst}$**	**Method**	**#R**	← (S+DC)	**MSE$_{tra}$**	**MSE$_{tst}$**
WM(5)	13		298450	282029	WM(5)	66		71294	80934
WM(7)	24		222654	239962					
WM+S(7)	17		214177	265179	WM+S(5)	**43**		57025	59942
WM+W(5)	13		242680	252483	WM+W(5)	66		33639	33319
WM+W(7)	24		191577	221583					
COR(5)	13		221569	196808	COR(5)	66		67237	69457
DC+S(7), ALM	17	(14+ 3)	155898	178534	DC+S(5)	47	(44+ 3)	51714	58806
HSLR(3,5)	11	(10+ 1)	180057	168211	HSLR(3,5)	131	(64+1+66R)	23525	22328
WM+P(5)	13		179126	203429	WM+P(5)	66		23440	31988
WM+AL(5)	13		230794	251068	WM+AL(5)	66		19665	23425
WM+WS(7)	20		191565	219370	WM+WS(5)	**43**		32476	32638
WCOR(5)	12		161414	161511	WCOR(5)	**43**		29417	31019
DC+WS$_{CC}$(7)	18	(14+ 4)	144290	176057	DC+WS$_{CC}$(5)	54	(49+ 5)	24961	28225
HSWLR(3,5)	**9**		163406	**156434**	HSWLR(3,5)	47	(45+ 2)	20425	22873
WM+PAL(5)	13		157264	171825	WM+PAL(5)	66		11464	20724
WM+PALS(5)	10		156943	197999	WM+PALS(5)	64		9896	13617
CORPAL(5)	16		**138935**	188797	CORPAL(5)	50		**4751**	**8422**

S Simple rules; *DC* Double Consequent rules; *R* Repeated rules

Table 4 Improvement percentages respect to the related basic approaches in Problems I and II

Methods	Problem	WM	WM+S	WM+W	COR	DC+S	HSLR	WM+P	WM+AL
WM+WS	P I	17,14,9	−18,11,17	17,0,1	—	—	—	—	—
	P II	35,54,60	0,43,46	35,3,2	—	—	—	—	—
WCOR	P I	8,46,43	—	8,33,36	8,27,18	—	—	—	—
	P II	35,59,62	—	35,13,7	35,56,55	—	—	—	—
DC+WS$_{CC}$	P I	25,35,27	—	25,25,21	—	−5,7,1	—	—	—
	P II	18,65,65	—	18,26,15	—	−14,52,52	—	—	—
HSWLR	P I	**31,45,45**	—	31,33,38	—	—	18,9,7	—	—
	P II	29,71,72	—	29,39,31	—	—	64,13,-2	—	—
WM+PAL	P I	0,47,39	—	—	—	—	—	0,12,16	0,32,32
	P II	0,84,74	—	—	—	—	—	0,51,35	0,42,12
WM+PALS	P I	23,47,30	—	—	—	—	—	23,12,3	23,32,21
	P II	3,86,83	—	—	—	—	—	3,58,57	3,50,42
CORPAL	P I	−23,53,33	—	—	−23,37,4	—	—	−23,22,7	−23,40,25
	P II	**24,93,90**	—	—	24,93,88	—	—	24,80,74	24,76,64

Each cell in the table represents the improvement percentages (with format: % #R, % MSE$_{tra}$, % MSE$_{tst}$) of the proposed hybrid approaches respect to their related basic ones in problems I and II. These percentages are rounded due to space restrictions (see the first column–row where values are: around a 17% of reduction in the number of rules and around a 14 and 9% of improvement in training and test respect to the values obtained by WM in problem I).

in both problems. However, it overfits the models obtained and maintains the same number of rules of the classical approach (13 and 66 rules in problems I and II, respectively). In any case, the generalization error of the model obtained for the second problem is very good related to those approaches only considering tuning at rule level.

Figure 7 graphically depicts the tuned data base and rule base obtained by the WM+PAL for a specific data set partition of the electrical problem. In this figure we can see how the combined use of basic and additional membership functions parameters does not disturb the legibility of the data base finally obtained.

(b) WM+PALS: This approach shows a similar behavior to WM+PAL with two main differences. By allowing rule selection the models obtained are simpler than those obtained by WM+PAL which, in the case of the second problem (medium complexity) avoids the overfitting and produces significant improvements respect to its related basic models.

(c) CORPAL: Again, this method presents better results than those obtained by WM, COR, WM+P and WM+AL. In the case of the first problem, the model obtained is overfitted, presenting even more rules that in the classical approach. However, CORPAL presents the best results of problem II, removing ten rules respect to the classical approach and, therefore, presenting the best trade-off between accuracy and interpretability. In this problem, this approach obtains improvements over the 85% in training and test respect to the original COR methodology

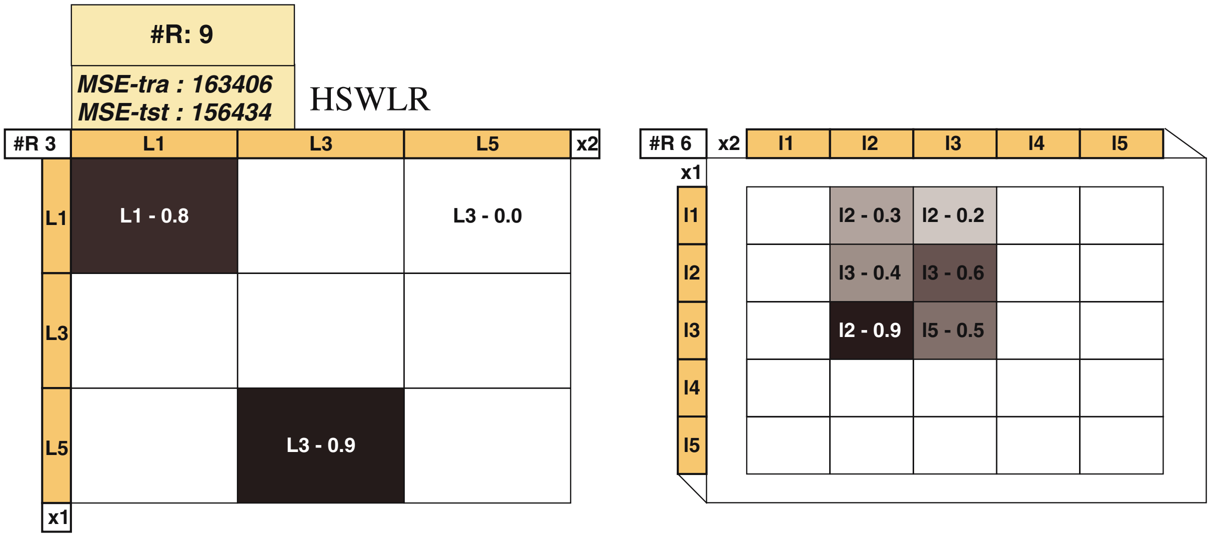


Fig. 6 Decision table of the FRBS obtained from HSWLR(3,5)

and over the 64% in training and test respect to WM+P and WM+AL.

To summarize, we can say that all the hybrid approaches show more accurate results respect to its related basic approaches, obtaining FRBS with an acceptable interpretability and, in several cases, even simpler than the ones obtained by the basic approaches. Moreover, notice that:

- Those approaches only performing tuning at rule level (approaches considering weights in combination with rule selection) seems to present a better behavior in these kinds of problems presenting strong non-linearities (as in the case of the problem I). Moreover, they usually get the simpler models in terms of number of rules.
- Those approaches based on the data base tuning have presented the better results in the second problem (problem II), showing the good behavior of these kinds of techniques in problems of medium complexity (having not so strong non-linearities) with large search spaces. In any case, notice that the consideration of techniques acting on the RB, such as the rule selection (PALS) and the rule cooperation methodology (CORPAL), shows a better behavior, specially in generalization, than the one only considering tuning of the parameters, scaling factors and linguistic edges (PAL).

6 Concluding remarks

In this work, we present a short study of how some basic refinement approaches can be combined to obtain new hybrid improved approaches presenting a better trade-off between interpretability and accuracy. We propose the use of seven hybrid approaches to develop accurate FRBSs, which have been tested considering two real-world problems. All the considered hybrid approaches have shown more accurate results respect to its related basic approaches, obtaining FRBS with an acceptable interpretability and, in many cases, even simpler than the ones obtained by the basic approaches.

In view of the obtained results, the proposed approaches seems to inherit the accuracy and interpretability characteristics of their involved basic approaches, obtaining simple but powerful fuzzy linguistic models. This is due to the following reasons:

- A smart hybridization has been performed selecting different basic approaches which present complementary characteristics. In this way, we have made use of the ability of these approaches to induce a better cooperation among the different rules in an adequate way and, therefore, the accuracy of the obtained FRBSs has been improved.
- Before combining the different approaches, there is a need to take into account the interpretability presented by the final obtained model. Following this premise, the interpretability has been maintained to a good level.

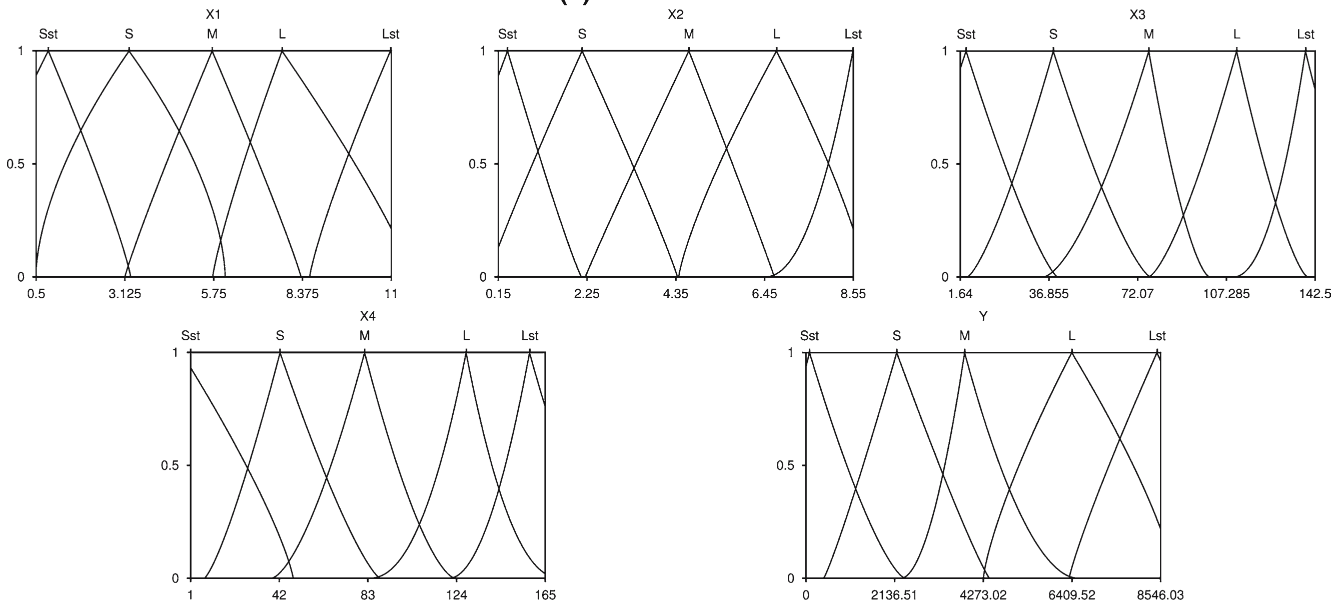
A Basic refinement approaches

A.1 Rule selection

Rule selection involves obtaining an optimized subset of rules from a previous rule set by selecting some of them. We may find several methods to do so with different search algorithms in the specialized literature [12, 13, 28].

In [30], 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 them seems similar to the orthogonal transformation-methods used for TSK-type FRBSs [49, 50]. Another heuristic rule selection procedure is proposed in [48].

(a) Tuned data base



(b) Tuned rule base

Rule	*Street*	*Town*	*Building*	*Energy*	*Costs*	*Rule*	*Street*	*Town*	*Building*	*Energy*	*Costs*
R_{1}	*m-l* Sst	Sst	*very* Sst	*m-l* Sst	Sst	R_{34}	*m-l* L	*m-l* S	*very* S	*very* Sst	S
R_{2}	Sst	Sst	Sst	S	S	R_{35}	*very* L	S	*very* S	*m-l* S	S
R_{3}	Sst	*m-l* S	*very* Sst	*m-l* Sst	*very* Sst	R_{36}	*m-l* L	*very* S	S	*very* M	M
R_{4}	Sst	*m-l* S	Sst	*m-l* S	*m-l* S	R_{37}	L	S	*m-l* S	*very* L	M
R_{5}	*very* Sst	*m-l* S	S	*very* Sst	*very* Sst	R_{38}	L	M	*m-l* S	*m-l* Sst	*very* S
R_{6}	Sst	S	*very* S	*m-l* S	*very* S	R_{39}	*m-l* L	*very* M	*very* S	S	*very* M
R_{7}	S	*m-l* Sst	*very* Sst	*m-l* Sst	Sst	R_{40}	L	*m-l* M	*m-l* S	*m-l* M	M
R_{8}	*m-l* S	Sst	*m-l* Sst	*very* S	*m-l* S	R_{41}	*m-l* L	*m-l* M	*m-l* S	*very* L	*very* M
R_{9}	S	*very* Sst	*very* S	*very* Sst	Sst	R_{42}	*m-l* L	*m-l* M	*m-l* M	S	*very* M
R_{10}	*m-l* S	*m-l* Sst	*m-l* S	*m-l* S	S	R_{43}	*m-l* L	M	M	*m-l* M	L
R_{11}	*very* S	S	*m-l* Sst	*very* Sst	*m-l* Sst	R_{44}	L	*m-l* M	*very* M	*m-l* L	L
R_{12}	*m-l* S	S	*m-l* Sst	S	*m-l* S	R_{45}	*m-l* L	*very* L	*very* M	*very* Sst	M
R_{13}	*m-l* S	S	S	Sst	*m-l* S	R_{46}	L	L	*very* M	*m-l* S	M
R_{14}	*very* S	*very* S	S	S	S	R_{47}	L	*very* L	M	*very* M	*m-l* L
R_{15}	S	*very* M	*very* S	Sst	S	R_{48}	L	*very* L	*very* M	*m-l* L	*very* L
R_{16}	S	*very* M	*very* S	*very* S	*m-l* S	R_{49}	L	*very* L	L	*very* S	*m-l* L
R_{17}	S	*m-l* M	M	*m-l* Sst	*m-l* M	R_{50}	*m-l* L	*m-l* L	*very* L	M	L
R_{18}	*m-l* S	*m-l* M	*m-l* M	S	*very* M	R_{51}	*m-l* L	*very* L	L	L	*m-l* Lst
R_{19}	*m-l* M	*m-l* S	Sst	*m-l* Sst	Sst	R_{52}	*m-l* L	*very* Lst	*m-l* L	S	M
R_{20}	*m-l* M	*m-l* S	*m-l* Sst	S	*very* S	R_{53}	*m-l* L	*very* Lst	*m-l* L	*m-l* M	*very* L
R_{21}	*m-l* M	S	*m-l* Sst	*m-l* M	S	R_{54}	*m-l* L	Lst	*very* L	*very* L	Lst
R_{22}	*very* M	*very* S	*very* S	*very* Sst	S	R_{55}	L	*m-l* Lst	*m-l* Lst	S	Lst
R_{23}	M	*very* S	*m-l* S	S	S	R_{56}	L	Lst	Lst	*m-l* M	*m-l* Lst
R_{24}	M	S	S	*m-l* M	M	R_{57}	L	*very* Lst	*very* Lst	*very* L	*very* Sst
R_{25}	M	*very* M	*m-l* S	Sst	*m-l* S	R_{58}	Lst	*m-l* S	*very* S	S	*m-l* S
R_{26}	*m-l* M	M	*very* S	*m-l* S	*m-l* S	R_{59}	Lst	*very* S	*m-l* S	*very* L	*m-l* M
R_{27}	*m-l* M	M	*m-l* S	*m-l* M	M	R_{60}	Lst	S	S	*m-l* Lst	L
R_{28}	M	M	M	S	*m-l* M	R_{61}	*very* Lst	*m-l* S	M	S	M
R_{29}	*very* M	*m-l* M	M	*m-l* M	*m-l* M	R_{62}	*very* Lst	S	*very* M	*very* L	M
R_{30}	*very* M	*m-l* L	*very* M	S	*very* M	R_{63}	Lst	*m-l* S	*m-l* M	*m-l* Lst	L
R_{31}	*m-l* M	*very* L	*very* M	*m-l* M	*very* M	R_{64}	*very* Lst	*m-l* L	M	*very* S	*very* M
R_{32}	*m-l* M	L	*m-l* L	*very* S	M	R_{65}	*very* Lst	*m-l* L	M	*very* L	*m-l* L
R_{33}	*very* M	*very* L	*m-l* L	*m-l* M	*m-l* L	R_{66}	*very* Lst	*very* L	*very* M	*very* Lst	*m-l* Lst

Fig. 7 Knowledge base generated by the WM+PAL method for the electrical problem. *Sst* Smallest; *S* Small; *M* Medium; *L* Large; *Lst* Largest; *m-l* for more-or-less

A.2 Sophisticated rule base learning methods: COR

In this case, the improvements arise as an effort to exploit the accuracy ability of linguistic FRBSs by exclusively focusing on the rule base design. In this case, the data base and the model structure keep invariable, thus resulting in the highest interpretability. Usually, all these improvements have the final goal of enhancing the *interpolative reasoning* the FRBS develops. This is one of the most interesting features of FRBSs and plays a key role in their high performance, being a consequence of the cooperative action of the linguistic fuzzy rules.

The original COR method proposed in [8] follows the primary objective of inducing a better cooperation among the linguistic rules. To do that, the rule base design is made using global criteria that consider the action of the different rules jointly. It is attained by means of a strong, smart reduction of the search space. The main advantages of the COR methodology are its capability to include heuristic information, its flexibility to be used with different metaheuristics, and its easy integration within other derivation processes.

A.3 Weighted rules

This approach involves using an additional parameter for each rule that indicates its importance degree in the inference process, instead of considering all rules equally important as in the usual case. Thus, the FRBS presents more flexibility to improve the interpolative reasoning and, therefore, the model performance [18,37]. The rule structure will be the following one:

$$\textbf{IF } X_1 \text{ is } A_1 \text{ AND } \cdots \text{ AND } X_n \text{ is } A_n$$
$$\textbf{THEN } Y \text{ is } B \textit{ with } [w]\ ,$$

with w being the real-valued rule weight. Following this approach, some changes must be made to the classical inference system must be made to consider the weighted action of each rule.

The operator with, which attaches a weight to a rule, may be defined in different ways. One of the most usual options is to multiply the matching degree of the antecedent by the corresponding weight before applying the implication operator, which relates antecedent and consequent. Another possibility is to change the conclusion derived from the implication operator according to the corresponding weight (e.g., changing the support of the obtained fuzzy set).

These weights are usually considered to handle inconsistencies [18]. Moreover, some proposals make use of them to improve the model accuracy with an automatic learning of weights using different techniques such as heuristic methods [27,42], gradient descent processes [37], or evolutionary algorithms [4].

A.4 Double-consequent rules

This approach involves allowing the rule base to present rules where each combination of antecedents may have two consequents associated when it is necessary to improve the model accuracy [13,19,36]. It is clear that this will improve the capability of the model to perform the interpolative reasoning and, thus, its performance. The rule structure obtained will be as follows:

$$\textbf{IF } X_1 \text{ is } A_1 \text{ AND } \cdots \text{ AND } X_n \text{ is } A_n$$
$$\textbf{THEN } Y \text{ is } \{B_1, B_2\}.$$

Since each double-consequent fuzzy rule can be decomposed into two different rules with a single consequent, the usual plain fuzzy inference system can be applied. The only restriction imposed is that the defuzzification method must consider the matching degree of the rules fired. For example, the *center of gravity weighted by the matching degree* defuzzification strategy may be used.

When using two consequents per rule, the interpretation of the action performed by every rule may be confusing to some extent. However, we should note this fact does not constitute an inconsistency from the linguistic FM point of view but only a shift of the main labels making the final output of the rule lie in an intermediate zone between both consequents. Indeed, let us suppose that a specific combination of antecedents, "X_1 is A_1 AND $\cdots$ AND X_n is A_n," has two different consequents associated, B_1 and B_2. The resulting double-consequent rule may be interpreted as follows [13]:

$$\textbf{IF } X_1 \text{ is } A_1 \text{ AND} \ldots \text{AND } X_n \text{ is } A_n$$
$$\textbf{THEN } Y \text{ is } \textit{between } B_1 \text{ and } B_2\ .$$

A.5 Hierarchical knowledge bases

A deeper change in the model structure involves considering HKBs. In this case, the HKB is composed of a set of layers where each one contains linguistic partitions with different granularity levels (a layer of the hierarchical data base) and linguistic rules whose linguistic variables take values in these partitions (a layer of the HRB) [16]. Different learning methods have been proposed to design this extended model structure.

The method proposed in [28] obtains a HKB by creating several hierarchical linguistic partitions with different granularity levels, generating the complete set of linguistic rules in each of these partitions, taking the union of all of these sets, and finally performing a genetic rule selection process on the whole rule set.

The method introduced in [16] uses an inductive linguistic rule generation method to progressively refine the controversial regions (those covered by linguistic fuzzy rules with a bad performance) by defining new rules in a deeper layer. The obtained HRB is compacted by a subsequent selection process. Therefore, this latter method follows a *descending* approach refining the regions by increasing the granularity.

A.6 Learning of membership functions

Basic linguistic FM methods are exclusively focused on determining the set of fuzzy rules composing the rule base of the model. In these cases, the membership functions are usually obtained from expert information (if available) or by a normalization process and it remains fixed during the rule set derivation process.

However, the automatic design of the membership functions has shown to be a very suitable mechanism to increase the approximation capability of the linguistic models. Generally speaking, the procedure involves either defining the most appropriate shapes for the membership functions that give meaning to the fuzzy sets associated to the considered linguistic terms or determining the optimum number of linguistic terms used in the variable fuzzy partitions, i.e., the granularity.

In this contribution, we will focus on learning the membership functions by defining their parameters and using non-linear scaling factors to vary their shapes (these shapes will have a high influence in the FRBS performance):

– *Learning/tuning the membership function parameters* The most common way to derive the membership functions is to change their definition parameters [12,23]. For example, if the following triangular-shape membership function is considered:

$$\mu(x) = \begin{cases} \frac{x-a}{b-a}, & \text{if } a \le x < b, \\ \frac{c-x}{c-b}, & \text{if } b \le x \le c, \\ 0, & \text{otherwise.} \end{cases}$$

changing the basic parameters (a, b, and c) will vary the shape of the fuzzy set associated to the membership function, thus influencing the FRBS performance. The same yields for other shapes of membership functions (trapezoidal, gaussian, sigmoid, etc.).

– *Using non-linear scaling factors* — Another way to define the membership function shapes is to use more flexible alternative expressions for the membership functions to vary the compatibility degrees to the fuzzy sets [17, 32]. For example, a new membership function can be obtained raising the membership value to the power of α, i.e.,

$\mu'(x) = \mu(x)^{\alpha}, \quad 0 < \alpha.$

By changing the α value we may define different membership function shapes.

A.7 Fuzzy rules with linguistic hedges

A third possibility to increase the accuracy in linguistic FM is to relax the rule structure by including certain operators that slightly change the meaning of the linguistic labels involved in the system when necessary [17,22]. A way to do so without losing the description to a high degree is to use linguistic hedges.

A linguistic hedge is an operator that alters the membership functions for the fuzzy sets associated to the linguistic labels, giving a more or less precise definition as a result depending on the case. For example, the linguistic hedges ‘very’ and ‘more-or-less’ performs as follows: $\mu^{\text{very}}(x) = \mu(x)^2$ and $\mu^{\text{more-or-less}}(x) = \sqrt{\mu(x)}$. An example of a rule with this structure is the following:

IF X_1 is very high and X_2 is low
THEN Y is more-or-less large.

Actually, the consideration of linguistic modifiers does not define a new meaning to the so-called *primary terms* — *high*, *low*, and *large* in our example — but they are used as generators whose meaning is defined in the context. Certainly, the fact of using fuzzy rules with linguistic hedges will have a significative influence in the behavior of the linguistic FRBS because the matching degree of the rule antecedent as well as the output fuzzy set obtained when applied the implication operator in the inference process are changed.

References

1. Alcalá R, Casillas J, Cordón O, Herrera F (2002) Hybridization of components to improve the accuracy in linguistic fuzzy modeling. In: Proceedings of the 9th international conference on information processing and management of uncertainty in knowledge-based systems (IPMU 2002), pp 1969–1977
2. Alcalá R, Cano JR, Cordón O, Herrera F, Villar P, Zwir I (2003) Linguistic modeling with hierarchical systems of weighted linguistic rules. Int J Approx Reason 32(2–3):187–215
3. Alcalá R, Casillas J, Cordón O, Herrera F (2003) Applying rule weight derivation to obtain cooperative rules. In: Benítez JM, Cordón O, Hoffmann F, Roy R (eds) Advances in soft computing – engineering, design and manufacturing, Springer, Berlin Heidelberg New York, pp 139–147
4. Alcalá R, Casillas J, Cordón O, Herrera F (2003) Linguistic modeling with weighted double-consequent fuzzy rules based on cooperative coevolutionary learning. Int J Int Comput Aided Eng (to appear)
5. Alcalá R, Cordón O, Herrera F (2003) Combining rule weight learning and rule selection to obtain simpler and more accurate linguistic fuzzy models. In: Lawry J (ed) Modelling with words, Springer, Berlin Heidelberg New York (in press)
6. Bastian A (1994) How to handle the flexibility of linguistic variables with applications. Int J Uncertainty Fuzziness Knowledge-Based Syst 2(4):463–484
7. Bikdash M (1999) A highly interpretable form of Sugeno inference systems. IEEE Trans Fuzzy Syst 7(6):686–696
8. Casillas J, Cordón O, Herrera F (2002) COR: A methodology to improve ad hoc data-driven linguistic rule learning methods by inducing cooperation among rules. IEEE Trans Syst Man Cybern 32(4):526–537
9. Casillas J, Cordón O, Herrera F, Magdalena L (eds) (2003) Interpretability issues in fuzzy modeling. Studies in fuzziness and soft computing, 128, Springer, Berlin Heidelberg New York
10. Casillas J, Cordón O, Herrera F, Magdalena L (eds) (2003) Accuracy improvements in linguistic fuzzy modeling. Studies in

fuzziness and soft computing, 129, Springer, Berlin Heidelberg New York
11. Casillas J, Cordón O, Del Jesus MJ, Herrera F (2001) Genetic tuning of fuzzy rule deep structures for linguistic modeling. Technical Report #DECSAI-01-01-02, Department of Computer Science and Artificial Intelligence, University of Granada
12. 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
13. Cordón O, Herrera F (2000) A proposal for improving the accuracy of linguistic modeling. IEEE Trans Fuzzy Syst 8(3):335–344
14. Cordón O, Herrera F, Sánchez L (1999) Solving electrical distribution problems using hybrid evolutionary data analysis techniques. Appl Intell 10:5–24
15. Cordón O, Herrera F, Hoffmann F, Magdalena L (2001) Genetic fuzzy systems: evolutionary tuning and learning of fuzzy knowledge bases, World Scientific, Singapore
16. Cordón O, Herrera F, Zwir I (2002) Linguistic modeling by hierarchical systems of linguistic rules. IEEE Trans Fuzzy Syst 10(1):2–20
17. 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
18. Cho JS, Park DJ (2000) Novel fuzzy logic control based on weighting of partially inconsistent rules using neural network. J Intell Fuzzy Syst 8:99–110
19. Espinosa J, Vandewalle J (2000) Constructing fuzzy models with linguistic integrity from numerical data-AFRELI algorithm. IEEE Trans Fuzzy Syst 8(5):591–600
20. Fiordaliso A (2001) A constrained Takagi-Sugeno fuzzy system that allows for better interpretation and analysis. Fuzzy Sets Syst 118(2):307–318
21. Gómez-Skarmeta A, Jiménez F (1999) Fuzzy modeling with hybrid systems. Fuzzy Sets Syst 104:199–208
22. González A, Pérez R (1999) A study about the inclusion of linguistic hedges in a fuzzy rule learning algorithm. Int J Uncertainty Fuzziness Knowledge-Based Systems 7(3):257–266
23. Gürocak HB (1999) A genetic-algorithm-based method for tuning fuzzy-logic controllers. Fuzzy Sets Syst 108(1):39–47
24. Herrera F, Lozano M, Verdegay JL (1997) Fuzzy connectives based crossover operators to model genetic algorithms population diversity. Fuzzy Sets Syst 92(1):21–30
25. Hong TP, Chen JB (1999) Finding relevant attributes and membership functions. Fuzzy Sets Syst 103(3):389–404
26. Hong TP, Lee CY (1999) Effect of merging order on performance of fuzzy induction. Int Data Anal 3(2):139–151
27. Ishibuchi H, Nakashima T (2001) Effect of rule weights in fuzzy rule-based classification systems. IEEE Trans Fuzzy Syst 9(4):506–515
28. 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
29. Jin Y, von Seelen W, Sendhoff B (1999) On generating FC^3 fuzzy rule systems from data using evolution strategies. IEEE Trans Syst Man Cybern 29(4):829–845
30. Krone A, Taeger H (2001) Data-based fuzzy rule test for fuzzy modelling. Fuzzy Sets Syst 123(3):343–358
31. Lee HM, Chen CM, Chen JM, Jou YL (2001) An efficient fuzzy classifier with feature selection based on fuzzy entropy. IEEE Trans Syst Man Cybern 31(3):426–432
32. 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 31(1):32–53
33. Mamdani EH (1974) Applications of fuzzy algorithms for control a simple dynamic plant. In: Proceedings of the IEEE 121(12):1585–1588
34. Michalewicz Z (1996) Genetic algorithms + data structures = evolution programs, Springer, Berlin Heidelberg New York
35. Nauck D, Kruse R (1999) Neuro-fuzzy systems for function approximaton. Fuzzy Sets Syst 101(2):261–271
36. 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
37. Pal NR, Pal K (1999) Handling of inconsistent rules with an extended model of fuzzy reasoning. J Intell Fuzzy Syst 7:55–73
38. Paredis J (1995) Coevolutionary computation. Artifi Life 2:355–375
39. Pardalos PM, Resende MGC (2002) Handbook of applied optimization, Oxford University Press, NY
40. Pedrycz W, Gudwin RR, Gomide FAC (1997) Nonlinear context adaptation in the calibration of fuzzy sets. Fuzzy Sets Syst 88(1):91–97
41. Roubos H, Setnes M (2001) Compact and transparent fuzzy models and classifiers through iterative complexity reduction, *IEEE Transactions on Fuzzy Systems*, **9:4** (2001) 516–524
42. Sánchez L, Casillas J, Cordón O, del Jesus MJ (2003) Some relationships between fuzzy and random set-based classifiers and models. Int J Approx Reason 29(2):175–213
43. Setnes M, Babuška R, Kaymak U, van Nauta Lemke HR (1998) Similarity measures in fuzzy rule base simplification. IEEE Trans Syst Man Cybern 28(3):376–386
44. Shi Y, Eberhart R, Chen Y (1999) Implementation of evolutionary fuzzy systems. IEEE Trans Fuzzy Syst 7(2):109–119
45. Sugeno M, Yasukawa T (1993) A fuzzy-logic-based approach to qualitative modeling. IEEE Trans Fuzzy Syst 1(1):7–31
46. Takagi T, Sugeno M (1985) Fuzzy identification of systems and its application to modeling and control. IEEE Trans Syst Man Cybern 15:116–132
47. Wang LX, Mendel JM (1992) Generating fuzzy rules by learning from examples. IEEE Trans Syst Man Cybern 22:1414–1427
48. Wang X, Hong J (1999) Learning optimization in simplifying fuzzy rules. Fuzzy Sets Syst 106(3):349–356
49. Yam Y, Baranyi P, Yang CT (1999) Reduction of fuzzy rule base via singular value decomposition. IEEE Trans Fuzzy Syst 7(2):120–132
50. Yen J, Wang L (1999) Simplifying fuzzy rule-based models using orthogonal transformation methods. IEEE Trans Syst Man Cybern 29(1):13–24
51. Yen J, Wang L, Gillespie CW (1998) Improving the interpretability of TSK fuzzy models by combining global learning and local learning. IEEE Trans Fuzzy Syst 6(4):530–537
52. Zadeh LA (1975) The concept of a linguistic variable and its application to approximate reasoning, Parts I, II and III. Inform Sci 8, 8, 9,:199–249, 301–357, 43–80