Cooperative Evolutionary Learning of Linguistic Fuzzy Rules and Parametric Aggregation Connectors for Mamdani Fuzzy Systems

Francisco Alfredo Márquez, Antonio Peregrín, and Francisco Herrera

***Abstract*—There are two tasks in the design of linguistic fuzzy models for a concrete application: The derivation of the linguistic rule base and the setup of the inference system and the defuzzification method. Traditionally, the derivation of the linguistic rule base has been considered the most important task, but the use of the appropriate aggregation connectors in the inference system and the defuzzification interface can improve the fuzzy system behavior. In this paper, we take in consideration this idea, we propose an evolutionary learning method to learn a linguistic rule base and the parametric aggregation connectors of the inference and defuzzification in a single step. The aim of this methodology is to make possible a high level of positive synergy between the linguistic rule base and the aggregation connectors, improving the accuracy of the linguistic Mamdani fuzzy systems. Our proposal has shown good results solving three different applications. We introduce a statistical analysis of results for validating the model behavior on the applications used in the experimental study. We must remark that we present an experimental study with a double intention: a) to compare the behavior of the new approach in comparison with those ones that first learn the rule base and then adapt the connectors, and b) to analyze the rule bases obtained with fixed aggregation connectors and with the adaptive ones for showing the changes on the consequent rules, changes on labels that produce a better behavior of the linguistic model than the classic ones.**

***Index Terms*—Adaptive defuzzification, adaptive inference, fuzzy rule-based systems, genetic fuzzy systems, linguistic rule base, parametric aggregation connectors, parametric t-norms.**

I. Introduction

FUZZY modeling, 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 with fuzzy predicates. Mamdani linguistic fuzzy models use a kind of fuzzy rules composed of linguistic variables [46] that take values in a term set with a real-world meaning, in order to describe the behavior of the system being modeled [40].

Manuscript received April 26, 2006; revised March 20, 2007. This work was supported by the Spanish Ministry of Education and Science (MEC) under Grant TIN2005-08386-C05-01.

F. A. Márquez and A. Peregrín are with the Department of Electronic Engineering, Computer Science and Automatics, University of Huelva, 21071 Huelva, Spain (e-mail: alfredo.marquez@diesia.uhu.es; peregrin@diesia.uhu.es).

F. Herrera is with the Department of Computer Science and Artificial Intelligence University of Granada, 18071 Granada, Spain (e-mail: herrera@decsai.ugr.es).

Digital Object Identifier 10.1109/TFUZZ.2007.904121

Interpretability and accuracy are contradictory requirements. While *interpretability* is the capability to express the behavior of the real system in an understandable way, a*ccuracy* is the capability to represent faithfully the real system. In practice, depending on the application details, one of the two properties normally prevails over the other. The higher interpretability with lower accuracy or lower interpretability with higher accuracy. Thus, designers try to find a tradeoff between the two edges, producing an increasing interest [5], [6], recently using evolutionary multi-objective optimization techniques [26], [43].

There are two tasks in the design of a linguistic fuzzy model for a concrete application: The derivation of the linguistic rule base (RB) and the setup of the inference system and the defuzzification method. Nowadays, in the framework of the trade-off between *interpretability* and *accuracy* in fuzzy modeling, the configuration of the inference system and the defuzzification method can reach a major importance. It is possible to choose appropriate connectors, providing the major cooperation with the linguistic RB to get more accuracy, maintaining interpretability.

We know that it is possible to use parametric aggregation operators in the design of the inference system and the defuzzification method, trying to get the most appropriate parameters configuration in any application. The tuning of these components can be considered to get more accurate fuzzy models. We can find different studies in the literature considering this problem. Recent approaches are described in the following.

- In [44], authors look for better performance than traditional minimum or product t-norms for the antecedent connections, and develop a study on the use of parametric connectors, suggesting the use of adaptive t-norms for the antecedent connection.
- In [9] we can find a study on the use of adaptive defuzzification methods.
- In [38], a generic flexible neuro-fuzzy system [36], [37] based on a quasi-triangular norm and a quasi-implication is showed. These operators allow the system to select between a Mamdani Approximate Reasoning (inference with a t-norm and aggregation with a t-conorm) or Formal Logical Reasoning (inference with a S-implication and aggregation with a t-norm) [16] depending on a parameter to be learnt together with the parameters of input–output membership functions.
- In [1] we find a study on the use of parametric t-norms in the inference process that is also analyzed with the tuning of the membership functions.

- In [13], a proposal of a genetic algorithm that simultaneously determines how the inference will be applied within each rule together with the tuning of the membership functions is showed in the framework of fuzzy classifiers systems.

Following these ideas on the advantage of the use of parametric connectors, we propose an evolutionary learning model for getting a rules base and parametric aggregation connectors for Mamdani linguistic fuzzy systems in order to achieve a positive synergy between the RB and the aggregation connectors used by the model.

Why do that? Usually, the existing models present a postprocessing study on the learning/tuning of parameters with a previously established RB. We want to analyze the advantages of the use of parametric aggregation connectors for learning the RB, evaluating the differences between the RB obtained via a fixed set of connectives and the RB obtained with our proposal, and showing the positive synergy reached between both parts of the model. The concept of "cooperative evolutionary learning" is used for representing this idea, the cooperation of both components, RB and parametric aggregation connectors via a positive synergy between both fuzzy system components. As far as we know, this is the first proposal for learning a linguistic RB base and the aggregation connectives for getting a maximum accuracy of the linguistic fuzzy model without tuning of the membership function parameters.

Genetic algorithms (GAs) are search algorithms based on natural genetics that provide robust search capabilities in complex spaces, and thereby offer a valid approach to problems requiring efficient and effective search processes. GAs based on real-number representation, like the ones used in this work, usually called *real-coded* GAs, seem to be adequate when tackling optimization problems of parameters with variables in continuous domains ([14], [15], [21], [31]) due to their ability to avoid becoming trapped at a local optimum, which is of special interest in real-world optimization problems. The use of GAs to design Fuzzy Systems (FSs) allows us to introduce the learning and adaptation capabilities. The result of this hybridization is a Genetic Fuzzy System which is basically a FS augmented by a learning process based on a GA [8], [10].

We use a GA as a tool to evolve the linguistic RB and the connector parameters, learning them with the aim of getting the maximum cooperation. We evaluate this cooperation via the accuracy of the model, using the accuracy measurement as the fitness function of the GA. The three components that are the base of our learning proposal are coded as a single chromosome, defining a specific GA to evolve this structure:

- The linguistic RB learning that is based on the COR methodology [4]. It is an *ad hoc* data-driven approach that does not select the rules looking for the best individual performance as usually do most of the data-driven techniques. It finds a set of cooperative rules searching for the consequents with the best global performance. This methodology manages a set of consequent label sets (one per rule), considering the linguistic RB learning as a combinatorial optimization problem.
- The parameterized connector of the inference system that is the conjunction operator of the antecedents, using a parametric t-norm.
- The parameterized expression of the defuzzification method that uses a weighted average aggregation operator as described in [9] or the parametric SLIDE defuzzification method [45].

Why use GAs? We must point out that, on the one hand, we code the linguistic RB via an integer coding representation and on the other hand, we code the connector parameters via real coding. GAs allow us to evolve this complex structure with different kind of variables defining the adequate operators, and as we have mentioned before, they present the ability to avoid becoming trapped at a local optimum.

We analyze the cooperation between the different components developing an experimental study with three different applications where we compare the accuracy results. To do so, we provide a statistical analysis using some statistical tests (ANOVA, Levene and Tamhane [2]) with the aim to show the significance in the accuracy improvements obtained with the proposed model.

In order to do that, the paper is organized as follows. Section II introduces the parametric aggregation connectors, the adaptive conjunction operators and the adaptive defuzzification methods. Section III is devoted to describing the evolutionary learning proposal. Section IV studies the behavior of the evolutionary fuzzy models with the three considered applications. Finally, Section V presents some concluding remarks. The Appendix is devoted to shortly describing the statistical tests used for our study and to showing the extended results of the statistical study.

II. Parametric Aggregation Connectors in Fuzzy Modeling

In this section, we introduce the notation used in the paper for FRBSs, and we show the parametric aggregation operators used in our learning proposal. In the first subsection we justify the use of the Dubois parametric t-norm as conjunction operator, and in the second we present the two adaptive defuzzification methods used in our study: one uses a weighted average aggregation operator and the other one is based on SLIDE.

A. Adaptive Conjunction

Linguistic FRBSs for system modeling use IF - THEN rules of the following form:

$$R_i : \text{If } X_{i1} \text{ is } A_{i1} \text{ and} \cdots \text{and} X_{im} \text{ is } A_{im} \text{ then } Y \text{ is } B_i$$

with $i = 1$ to N, and with X_{i1} to X_{im} and Y being the input and output variables respectively, and with A_{i1} to A_{im} and B_i being the involved antecedents and consequent labels, respectively.

The expression of the Compositional Rule of Inference in fuzzy models with singleton fuzzification is the following one:

$$\mu_{\mathrm{B}'}(y) = \mathrm{I}(\mathrm{C}(\mu_{\mathrm{A1}}(x_1), \ldots, \mu_{\mathrm{Am}}(x_m)), \mu_{\mathrm{B}}(y))$$

where $\mu_{\mathrm{B}'}(\cdot)$ is the membership function of the inferred consequent, $\mathrm{I}(\cdot)$ is the rule connective, $\mathrm{C}(\cdot)$ is the conjunction operator, $\mu_{\mathrm{Ai}}(\mathrm{x_i})$ are the values of the matching degree of each input with the membership functions of the rule antecedents, and $\mu_{\mathrm{B}}(\cdot)$ is the consequent of the rule.

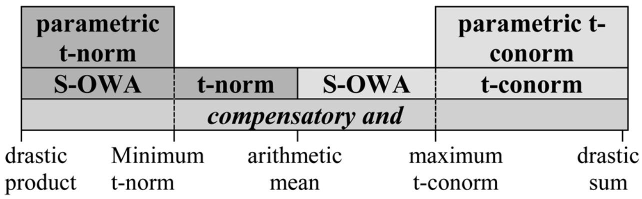

Fig. 1. Ranges covered by connectives operators.

Therefore, the inference system performs the two following tasks.

1) First, it computes $\mathrm{C}(\mu_{\mathrm{A1}}(x_1), \ldots, \mu_{\mathrm{An}}(x_m))$, that is the *matching degree* of each rule, $\mathrm{h_i}$. The conjunction operator $\mathrm{C}(\cdot)$ is usually modeled with a t-norm.
2) Second, it infers using the rule connective $\mathrm{I}\ (\cdot)$, the matching degree and the consequent of the rule. Rule connectives can be classified into different families, being implication functions [41] and t-norms [20] the most well known. T-norms are the most used in practical fuzzy modeling.

Hence, the inference system uses two components: the conjunction $\mathrm{C}(\cdot)$, and the rule connective $\mathrm{I}(\cdot)$.

The aforementioned two components, conjunction and rule connective are suitable to be parameterized in order to adapt the IS. Our previous studies in [1], [22] show that the model based on the adaptive conjunction operator is a more valuable option than the one based on the adaptive rule connective, in the framework of improving the accuracy of linguistic FSs. Consequently, we have selected the use of the adaptive conjunction in this study, in order to parameterize the IS.

In [44], looking for better performance than traditional minimum or product t-norms for the antecedents connections, authors develop a study on the use of parametric connectors that are extended from t-norms and t-conorms in order to cover the range between them, including *compensatory and* and *S-OWA* operators and many others. Fig. 1 shows the ranges covered by parametric t-norms, parametric t-conorms, *compensatory and*, and S-OWA operators.

In our previous studies, we obtained good performance with t-norms, and we will use them in this study. Table I exemplifies three classical parametric T-norms [32] that can be used to model the adaptive conjunction operator $\mathrm{C}(\cdot)$. The parameter for the adaptive conjunction will be α, therefore the adaptive component is $\mathrm{C}(\alpha, \cdot)$.

Table II shows the relation between the five classical t-norms and the values of the parameter of the adaptive t-norms.

The use of adaptive conjunction connectives in Table I, allows to adapt the influence of the matching degree in a non linear way. The effect of the parameter in the adaptive conjunction is sometimes equivalent to one of the well-known mechanisms to modify the linguistic meaning of the rule structure, the use of linguistic modifiers [30]. The goal of linguistic rule modifiers is also to improve the accuracy of the model, slightly relaxing the rule structure by changing the meaning of the involved labels. The parameter plays a similar role by changing the shape of the membership function associated with the linguistic label antecedents of the rule, as shown in Fig. 2, where h is the matching for the trapezoidal fuzzy set when the input value is e, h' and $''$ are the values computed for $\alpha = 0.2$ and $\alpha = 0.1$ respectively. We must point out that the effect of the adaptive t-norm playing the role of conjunction operator does not modify the shape of the inferred fuzzy set.

TABLE I
ADAPTIVE T-NORMS

T-norm		
Name	**Expression**	**Domain**
Dubois	$T_{\mathrm{Dubois}}(x,y,\alpha) = \dfrac{x \cdot y}{\mathrm{Max}(x,y,\alpha)}$	$(0 \le \alpha \le 1)$
Dombi	$T_{\mathrm{Dombi}}(x,y,\alpha) = \dfrac{1}{1+\sqrt[\alpha]{\left(\frac{1-x}{x}\right)^{\alpha}+\left(\frac{1-y}{y}\right)^{\alpha}}}$	$(\alpha > 0)$
Frank	$T_{\mathrm{Frank}}(x,y,\alpha) = \log_{\alpha}\left[1+\dfrac{(\alpha^{x}-1)(\alpha^{y}-1)}{\alpha-1}\right]$	$(\alpha > 0)$, $(\alpha \ne 1)$

TABLE II
RELATION BETWEEN CLASSICAL AND PARAMETRIZED T-NORMS DEPENDING ON THE A PARAMETER

	$\mathbf{T_{Minimum}}$	$\mathbf{T_{Hamacher}}$	$\mathbf{T_{Algebraic}}$	$\mathbf{T_{Einstein.}}$	$\mathbf{T_{Bounded}}$	$\mathbf{T_{Drastic}}$
$\mathbf{T_{Dubois}}$	0		1			
$\mathbf{T_{Dombi}}$	∞	1				$\to 0$
$\mathbf{T_{Frank}}$	$\to 0$		$\to 1$		∞	

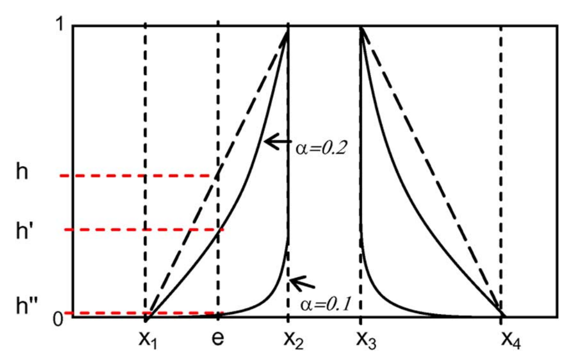

Fig. 2. Graphical representation of the antecedent linguistic modification produced by different values of Dombi t-norm.

Two models of Adaptive Inference System can be considered depending on the amount of parameters they use: a single parameter α to tune globally the behavior of the connector, or individual parameters for every rule, α_{i}, having a local tuning mechanism of the behavior of the inference system for every rule.

- The single parameter model lets us to adapt the behavior of the conjunction operator globally between the classical t-norms. However, the benefits of this model will not yield remarkable improvements in accuracy. The reason is the low importance in choice of the conjunction operator in the design of linguistic fuzzy systems [11] with a similar behavior with the use of different t-norms considering the same operator for all rules.

- On the contrary, the model that uses individual parameters for each rule, has got a mechanism to alter the behavior of the inference system for every rule. This model shows the highest accuracy in [1] because of its high degree of freedom.

Therefore, in this study, we learn the conjunctive connector for every rule separately.

Taking into account the studies in [1], [22], we have selected the Dubois t-norm with a separate connector for every rule. It showed the highest accuracy in the studies performed before, compared with Frank and Dombi t-norms and it is more efficiently computed. The use of an adaptive t-norm for the antecedent connection reminds the suggestion of [44] in order to look for better performance than traditional minimum or product t-norms.

Dubois t-norm is between minimum and algebraic product, when $\alpha = 0$ and $\alpha = 1$ it achieves like a minimum or product respectively. When $0 < \alpha < 1$, it continues performing like minimum excepting when every match with antecedents are below α, that takes values between minimum and product, being similar to a concentration effect. Thus, Dubois t-norm connects with minimum in those cases when the matches with antecedents are more significant, while the rest are connected with a value between minimum and product.

B. Adaptive Defuzzification Methods

The most used technique in practice, due to its fine performance, efficiency and easier implementation, is to apply the defuzzification function to every rule inferred fuzzy set (getting a characteristic value) and to compute then by a weighted average operator. This way of working is named Mode B [11] or FITA (First Infer, Then Aggregate) [3]. Its formula is

$$y_0 = \frac{\sum_i^N h_i \cdot V_i}{\sum_i^N h_i}$$

where $\mathrm{h_i}$ is the matching degree between the input variables and the rule antecedent fuzzy sets, and $\mathrm{V_i}$ represents a characteristic value of the fuzzy set inferred from rule $\mathrm{R_i}$, the Maximum Value or the Gravity Center (GC).

The general formula that generates some parametric defuzzification methods is

$$y_0 = \frac{\sum_i^N \mathrm{f}(h_i) \cdot V_i}{\sum_i^N \mathrm{f}(h_i)}$$

where $\mathrm{f(h_i)}$ is a functional of the matching degree [1].

The functional term can be defined with a single parameter, β, or with a set of parameters β_i, corresponding to one parameter for each rule $R_i, i = 1$, to N, in the RB. Moreover, the functional term could be defined as a product or as a power among other possible functions. Combining both functional operators and the aforementioned single or several parameters, the functional term could take any of these four forms

$$\mathrm{f}(h_i) = h_i \cdot \beta, \quad \mathrm{f}(h_i) = h_i^\beta$$
$$\mathrm{f}(h_i) = h_i \cdot \beta_i, \quad \mathrm{f}(h_i) = h_i^{\beta_i}.$$

However, it doesn't make sense to consider the form $\mathrm{f}(h_i) = h_i \cdot \beta$ as the effect of β is cancelled in the final expression. Thus, combining the three aforementioned possibilities with the two characteristic values, Maximum Value or GC, six different defuzzification methods may be obtained (expressions can be seen in [9]). Some of them have been used by several authors, like the functional term $\mathrm{f}(h_i) = h_i \cdot \beta_\mathrm{i}$ in [35], or Accurate Center of Gravity with $\mathrm{f}(h_i) = h_i^{\beta_i}$ and GC [29].

The role of the individual parameter β is interpreted as a modulation of the matching influence, which can be improved or attenuated. We should note that this modulation is only linear for the product case.

The interpretation is quite different when one parameter for each rule is used. Instead of a global modulation of the matching influence, the local action of each rule defuzzified with a product or a power functional is changed. The difference between the meanings of each of these functional terms is discussed as follows.

The product functional term with a different parameter for each rule has the effect of weighted rules [7], [35]. The β_i value associated with rule R_i gets the meaning of how significant or important that rule is for the inference process. An improved accuracy is the system modeling goal when using this kind of rule. The following is an example of a set of weighted rules, where the weights are β_i:

$$\mathrm{R}_1 : \text{ If } X_{11} \text{ is } \mathrm{A}_{11} \text{ and} \ldots \text{ and} X_{1m} \text{ is} A_{1m} \text{ then } Y \text{ is} B_1 \text{with } \beta_1$$
$$\mathrm{R}_2 : \text{If } X_{21} \text{is } A_{21} \text{and} \ldots \text{and} X_{2m} \text{is } A_{1m} \text{then } Y \text{is } B_2 \text{ with } \beta_2$$
$$\ldots$$
$$\mathrm{R}_n : \text{If } X_{n1} \text{is } A_{n1} \text{and} \ldots \text{and} X_{nm} \text{is } A_{nm} \text{then } Y \text{is } B_n \text{with } \beta_n$$

The rule weight adaptation process will produce a rule subset with better cooperation among the rules composing it [9]. This fact has shown to be of special interest when the rule set has been generated using a quick data-driven fuzzy rule generation method. These methods usually look for the best individual rule performance, and generate a linguistic RB with a low cooperation degree. Using the product functional and a parameter learning process will be equivalent to look for a subset of rules with the best global cooperation.

Overall, the influence of rule weights on the interpretability of fuzzy systems is usually discussed. Some authors consider they can be equivalently replaced by modifications in the membership functions in order to avoid negative effects on the interpretability [33], while other authors claim the importance of weights as a certainty grade and its importance in some problems [24], [27].

As regards the power functional case, the effect on defuzzification is equivalent to one of the well known mechanisms to modify the linguistic meaning of the rule structure, the use of linguistic modifiers [30]. The defuzzifier parameter plays the role of a linguistic modifier changing the shape of the membership function associated with the linguistic label antecedents of the rule, as shown in Fig. 3, where h is the matching for the trapezoidal fuzzy set when the input value is e. We must point out that this effect does not modify the shape of the inferred fuzzy set because the matching is only modified by defuzzification effects.

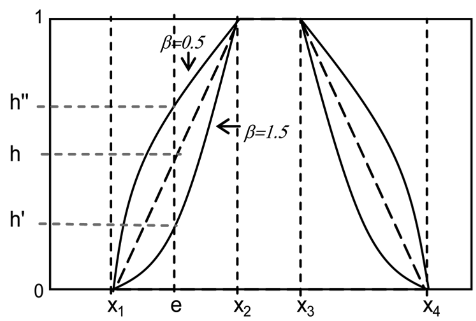

Fig. 3. Graphical representation of the effect produced by the power based linguistic modifier on the defuzzification process.

- When the fuzzy set is modified by power values greater than one, the membership function is concentrated. The modified matching will be h' in Fig. 3. Examples of these kind of linguistic modifiers are *absolutely, very, much more, more and plus* [23]
- On the contrary, when the fuzzy set is modified with values below one, the membership function is expanded or dilated. Observing Fig. 3, the modified matching will now be h''. Sometimes, these linguistic modifiers are named as minus, more or less and slightly [23].

  In this work, we have selected the two following parametric defuzzification methods:
- One obtained with the functional product with one parameter for each fule (that is weighting rule) and GC. The resulting expression is a well known weighted average aggregation operator. It has been selected due to its good results shown in [9]).

$$\mathrm{D}_W \frac{\sum_i^N h_i \cdot \beta_i \cdot GC_i}{\sum_i^N h_i \cdot \beta_i}$$

  where GC is computed with $(\int_Y y \cdot \mu_{B'}(y)dy)/(\int_Y \mu_{B'}(y)dy)$, (so called standard WCOA, for $\beta_\mathrm{i} = 1$, $\mathrm{i} = 1, \ldots, \mathrm{N}$).
- The second adaptive defuzzification method is the well known SLIDE [45], which formula is

$$D_{\mathrm{SLIDE}} = \frac{\sum_{i=1}^n \mu_{B'}(y_i) \cdot J_i \cdot y_i}{\sum_{i=1}^n \mu_{B'}(y_i) \cdot J_i}$$

$$\text{where } J_i = \begin{cases} 1-\delta, & \mu_{B'}(y_i) < \sigma \\ 1, & \mu_{B'}(y_i) \geq \sigma \end{cases}.$$

The parameters for this defuzzification method, σ and δ, are defined in the following intervals: $\sigma \in [0,1]$ and $\delta \in [0,1]$. We adapted this defuzzification method to Mode B—FITA.

According to Nauck and Kruse [33], the addition of parameters in fuzzy systems can deteriorate its interpretability. Even though, in the trade-off between interpretability and accuracy where this work is positioned, the sacrifice of a part of the system interpretability in order to get more accuracy is accepted. The objective is to get the best accuracy with the lowest loss of interpretability.

III. Cooperative Evolutionary Learning of Fuzzy Rules and Parametric Aggregation Connectors

In this section, we describe the evolutionary model proposed to learn the linguistic RB and the parametric aggregation connectors at the same time with the aim of obtaining fuzzy models with cooperation between fuzzy rules and fuzzy connectors. As we have mentioned, we use an *ad hoc data-driven method called COR*. The evolutionary algorithm used in the cooperative model is a well known GA that is called CHC [17].

In the following three subsections we first introduce the linguistic RB learning, we present the cooperative evolutionary learning model, and finally we describe CHC algorithm.

A. Rule Base Learning

As we have mentioned, COR, *cooperative rules* [4], is an *ad hoc* data-driven methodology that gets particularly accurate sets of rules because it does not select the rules looking for the best individual performance as usually do most of data-driven techniques. COR methodology finds a set of cooperative rules searching the consequents with the best global performance. This methodology manages a set of consequent label sets (one per rule), designing the linguistic RB learning as a combinatorial optimization problem.

We also decided to use the well known data-driven method proposed by Wang and Mendel (WM-method) in [42] because of its simplicity, clarity and quickness. Furthermore, it is used by COR method to generate the antecedent parts of the rules. The WM-method is a learning method we use to compare the cooperative evolutionary proposal with the learning of the connectors parameter for a learning method. Any other reference method might be considered as well. The algorithms descriptions are given below.

WM Method: The WM algorithm is an *ad hoc* data-driven linguistic rule learning method. It considers a previous definition of the linguistic term sets composed of the input and output primary fuzzy partitions that may be obtained from expert information (if it is available) or by a normalization process.

The generation of the linguistic rules is guided by covering criteria of the data in the example set (hence the name data-driven). The learning mechanism is specifically developed for this purpose, and consist of giving an importance degree to each linguistic RB on its covering and at last selecting the rule with the highest importance degree for each group, that is, for each antecedents combination.

A description of the WM rule generation process is shown in the following steps:

1) *Consider a fuzzy partition of the variable spaces.*
2) *Generate a candidate linguistic rule set*—This set will be formed by the rule best covering each example container in the input-output data set.

   The structure of each rule, CR^l, is obtained by taking a specific example, e_l, and setting each of the rule variables to the linguistic label associated with the fuzzy set best covering every example component, $(A_1^l, \ldots, A_m^l, B^l)$, with $A_i^l \in A_i$ and $B^l \in B$.
3) *Give an importance degree to each rule*—Let $\mathrm{CR}^l =$ IFX_1 is A_1^l and ... and X_m is A_n^l THEN Y is B^l be the linguistic rule generated from the example e_l.

The importance degree associated with it will be obtained by computing the covering value of the rule over the corresponding example as follows:

$$\mathrm{CV}_{\Pi}(\mathrm{CR}^l, e_l) = \mu_{A_1^l(x_1^l)} \ldots \mu_{A_m^l(x_m^l)} \cdot \mu_{B_1^l(x_1^l)}.$$

4) *Obtain a final linguistic RB from the candidate linguistic rule set*—Group the candidate linguistic rules according to their antecedents and select the rule with the highest importance degree in each group.

COR Method: An *ad hoc* data-driven method usually looks for the fuzzy rules with the best individual performance (e.g., the aforementioned [42]) and therefore the global interaction among the rules of the linguistic RB is not considered.

With the aim of addressing these drawbacks keeping the interesting advantages of *ad hoc* data-driven methods, the COR methodology was proposed in [4]. It is based on a *combinatorial search of cooperative rules* performed on the set of candidate rules to find the best cooperating rule set. Instead of selecting the consequent with the highest performance in each subspace as usual, the COR methodology considers the possibility of using another consequent, different from the best one, when it allows the FRBS to be more accurate thanks to having a RB with best cooperation. For this purpose, COR performs a combinatorial search among the candidate rules looking for the set of consequents which globally achieves the best accuracy.

COR consists of two stages

1) *Search space construction*—It obtains a set of candidate consequents for each rule.
2) *Selection of the most cooperative fuzzy rule set*—It performs a combinatorial search among these sets looking for the combination of consequents with the best global accuracy.

In order to perform this combinatorial search, an *explicit enumeration* or an *approximate search technique* can be considered.

1) The *explicit enumeration* accomplishes a full search through the set of possible combinations. Although this technique ensures that the optimal solution is obtained, it may take a long time, or simply be unapproachable in terms of run time, when there is a great number of combinations. Therefore, this technique is only used in confined spaces.
2) On the other hand, when the use of an explicit enumeration is not possible, an approximate search technique is needed. Any search technique can be used. However, since one of the main advantages of *ad hoc* data-driven methods is their ability to find good fuzzy models quickly, the search technique should be both effective and quick.

A description of the COR-based rule generation process is shown in the following steps.

Inputs:

- *An input-output data set*—$\mathrm{E} = \{e_1, \ldots\ldots, e_l, \ldots \ldots, e_M\}$, with $e_l = (x_1^l, \ldots\ldots,\ x_m^l, y_{1,\ldots\ldots,}^l y_p^l)$, $l \in \{1, \ldots\ldots, N\}, N$ being the data set size, and $n(m)$ being the number of input (output) variables—*representing the behavior of the problem being solved.*
- *A fuzzy partition of the variable spaces*. In our case, uniformly distributed fuzzy sets are regarded. Let A_i be the set of linguistic terms of the ith input variable, with $i \in \{1, \ldots\ldots, m\}$, and B_j be the set of linguistic terms of the jth output variable, with $j \in \{1, \ldots\ldots, p\}$, with $|A_i|(|B_j|)$ being the number of labels of the ith (jth) input (output) variable.

Algorithm:

1) *Search space construction:*

1.1 *Define the fuzzy input subspaces containing positive examples:* To do so, we should define the positive example set ($\mathrm{E}^{\,+}(S_s)$) for each fuzzy input subspace $S_s = (A_1^s, \ldots\ldots, A_i^s, \ldots\ldots, A_m^s)$, with $A_i^s \in A_i$ being a label, $s \in \{1, \ldots\ldots, N_s\}$, and $N_s = \prod_{i=1}^{n} |A_i|$ being the number of fuzzy input subspaces. In this paper, we use the following:

$$\mathrm{E}^+(S_s) = \{\mathrm{e_l} \in \mathrm{E} \,|\, \forall \mathrm{i} \in \{1, \ldots\ldots, \mathrm{m}\}, \\ \forall \mathrm{A}'_i \in \mathrm{A_i}, \mu_{\mathrm{A}_i^S(x_i^l)} \geq \mu_{A'_i(x_i^l)}\}$$

with $\mu_{\mathrm{A}_i^S(\cdot)}$ being the membership function associated with the label A_i^s.

Among all the N_s possible fuzzy input subspaces, consider only those containing at least one positive example. To do so, the set of subspaces with positive examples is defined as $S^+ = \{S_h | E^+(S_h) \neq \emptyset\}$.

1.2. *Generate the set of candidate rules in each subspace with positive examples:* First, the candidate consequent set associated with each subspace containing at least an example, $S_h \in S^+$, is defined. In this paper, we use the following:

$$\mathrm{C}(S_h) = \Big\{ \left(\mathrm{B}_1^{\mathrm{k}}\mathrm{h}, \ldots\ldots, \mathrm{B}_p^k h\right) \in B_1 \mathrm{x} \ldots \mathrm{xB_p} \,|\, \exists \mathrm{e_l} \\ \in \mathrm{E}^+(\mathrm{S_h}) \text{ where } \forall \mathrm{j} \in \{1, \ldots\ldots, \mathrm{p}\}, \forall \mathrm{B'_j} \in \mathrm{B_j}, \\ \mu_{B_j^{k_h}(y_j^l)} \geq \mu_{B'_j(y_j^l)} \Big\}$$

Then, the candidate rule set for each subspace is defined as $\mathrm{CR}(S_h) = \{R_{k_h} =$ [IF X_1 is A_1^h and ... and X_m is A_m^h THEN Y_1 is $B_1^{k_h}$ and ... and Y_p is $B_p^{k_h}$] such that $B^{k_h} = (B_1^{k_h}, \ldots\ldots, B_p^{k_h}) \in \mathrm{C}(S_h)\}$.

To allow COR to reduce the initial number of fuzzy rules, the special element $\mathrm{R}_\emptyset$ (which means "don't care") is added to each candidate rule set, i.e., $\mathrm{CR}(S_h) = \mathrm{CR}(S_h) \cup \mathrm{R}_\emptyset$. If it is selected, no rules are used in the corresponding fuzzy input subspace.

2) *Selection of the most cooperative fuzzy rule set*—This stage is performed by running a combinatorial search algorithm to look for the combination $\mathrm{RB} = \{R_1 \in \mathrm{CR}(S_1), \ldots\ldots, R_h \in \mathrm{CR}(S_h), \ldots\ldots, R_{|S+1|} \in \mathrm{CR}(S_{|S+1|})\}$ with the best accuracy. Since the tackled search space is usually large, approximate search techniques should be used.

An index f(RB) measuring the global quality of the encoded rule set is considered to evaluate the quality of each solution. In order to obtain solutions with a high interpretability, the original function is modified to penalize excessive number of rules:

$$\mathrm{f}'(\mathrm{RB}) = \mathrm{f}(\mathrm{RB}) + \gamma \cdot \mathrm{f}(\mathrm{RB}_O) - \#\mathbf{RB}/|S^+|$$

with $\gamma \in [0, 1]$ being a parameter defined by the designer to regulate the importance of the number of rules, #RB being

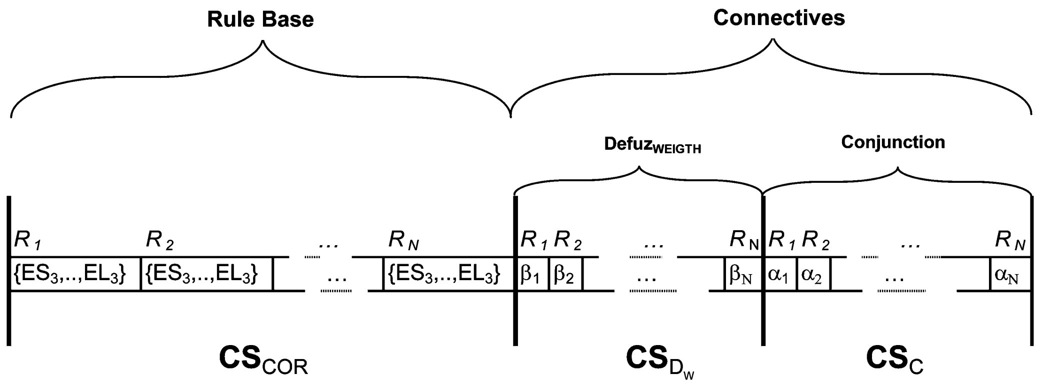

Fig. 4. Coding scheme for the evolutionary algorithm with N rules and weighted based defuzzification method.

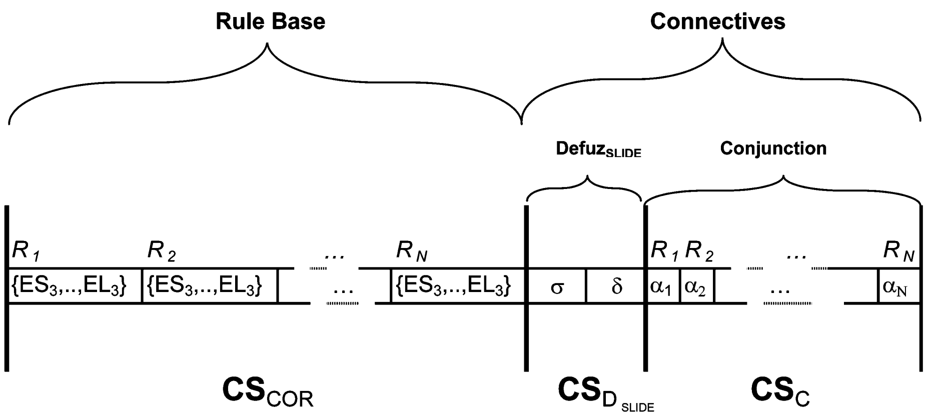

Fig. 5. Coding scheme for the evolutionary algorithm with N rules and SLIDE defuzzification method.

the number of rules used in the evaluated solution (i.e., $|S^+| - |\{R_h \in \mathbf{RB}\ \textit{such that}\ R_h = R_0\}|$), and RB_0 being the initial linguistic RB considered by the search algorithm.

B. Proposal: Cooperative Evolutionary Learning Model

With the aim of improving the accuracy of the fuzzy model, we propose a cooperative fuzzy model based on the cooperation between the linguistic RB and the fuzzy connectors.

The evolutionary model was based on the CHC Genetic Algorithm [17] whose chromosome has got a threefold coding scheme $(\mathrm{CS}_{\mathrm{COR}} + \mathrm{CS}_{\mathrm{C}} + \mathrm{CD}_{\mathrm{DW}})$ as represented in Fig. 4, where $\mathrm{CS}_{\mathrm{COR}}$ encodes the consequents of COR methodology, CS_{C} the α_i parameters of the conjunction connective, and $\mathrm{CS}_{\mathrm{DW}}$ the β_{i} parameters of the defuzzification.

Fig. 4 shows the chromosome where $\mathrm{CS}_{\mathrm{COR}}$ belongs to COR methodology. This part has got N genes, each one representing a candidate label of the consequent rules, being the possibilities between ES_3 and EL_3, represented in the implementation with integer values.

The proposed algorithm performs an approximate search among the candidate rules with the main aim of selecting the set of consequents with the best cooperation and simultaneously learning the rest of the chromosome.

The connective parameters are coded in the right side of the chromosome of Fig. 4. They are composed of two parts:

- Conjunction part, CS_{C}, with N parameters $\boldsymbol{\alpha}_{\mathrm{i}}$ (genes) for each R_{i} rule of the linguistic RB. Each gene can take any value in the interval [0, 1], that is, among minimum and algebraic product.
- Defuzzification part, with two possibilities:
  - $\mathrm{CS}_{\mathrm{D_W}}$: when using defuzzification method based on weighting rules, with N parameters β_{i} for each R_{i} rule of the linguistic RB. Each gene can take any value in the interval [0, 10]. This interval has been selected according with the study developed in [9]. It allows attenuation as well as enhancement of the matching degree.
  - $\mathrm{CS}_{\mathrm{D_{SLIDE}}}$: when using SLIDE defuzzification method (see Fig. 5), with two parameters σ and δ for the whole expression. The gene that represents σ or δ can take any value in the interval [0, 1].

C. Questions Related to the Evolutionary CHC Algorithm

The evolutionary algorithm used is the CHC [17]. It is considered as an evolutionary model with a good trade-off between diversity and convergence in high-dimensional search spaces in different applications.

During each generation, the CHC algorithm [17] uses a parent population of size M to generate an intermediate population of M individuals, which are randomly paired and used to generate two M' potential offspring (the value of M' depends on the crossover operator selected). Then, a survival competition is held, where the best M chromosomes from the parent and offspring populations are selected to form the next generation.

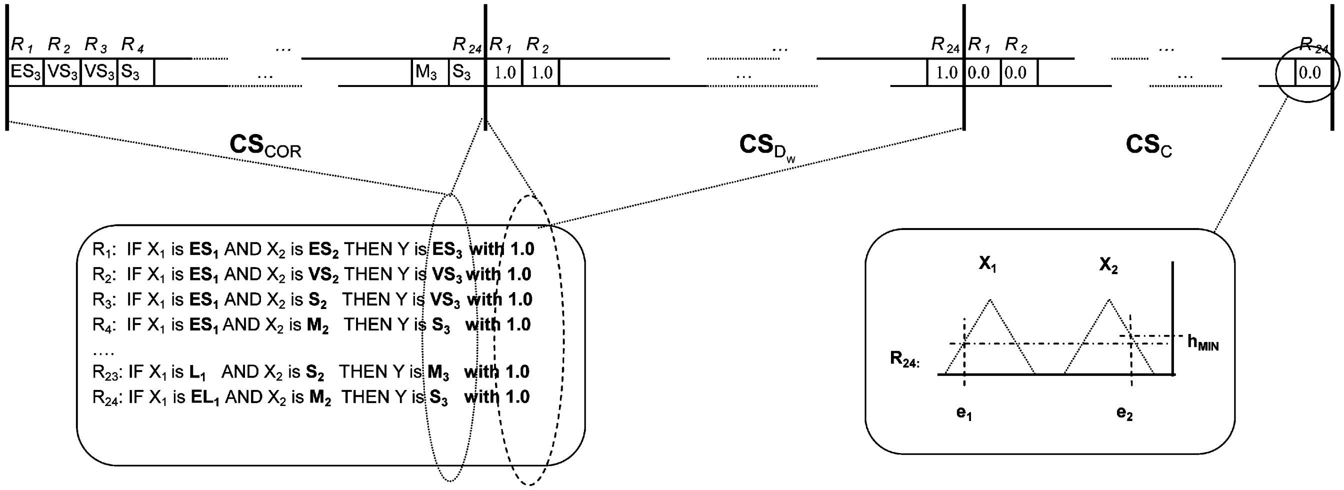

Fig. 6. Example of initial state for the chromosome whose has got the WM-method linguistic RB, the adaptive connector set to minimum t-norm, and the defuzzification set to WCOA.

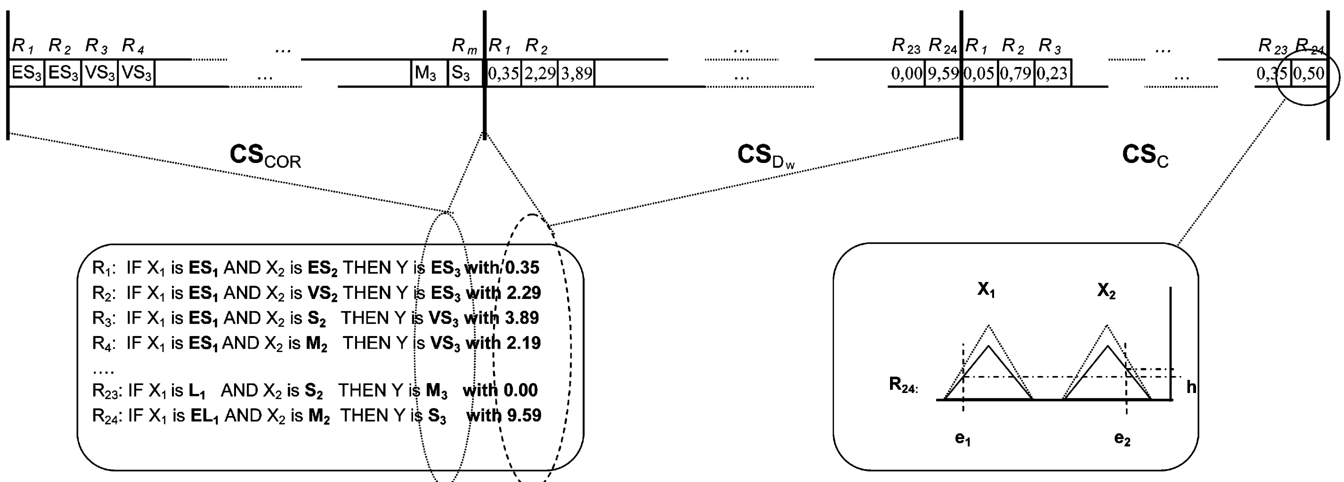

Fig. 7. Example of adapted chromosome after the evolutionary process.

No mutation is applied during the recombination phase. Instead, when the population converges or the search stops making progress (i.e., the difference threshold has dropped to zero and none of the new generated offspring are better than any member of the parent population), the population is reinitialized. The restarted population completely consists of random individuals except for one of them which must be the best individual found so far [18].

Although CHC was conceived for binary-coded problems, there are real-coded versions, like the one we use in this work to tune the parameters of the fuzzy operators. In these cases, the BLX-α crossover $(\alpha = 0.5)$ is used in order to recombine the parent's genes. It produces two descendents for each pair of parents, thus, the offspring generated by this crossover operator is of the same size than the initial population. The Hamming distance is computed by translating the real-coded genes into strings and by taking into account whether each character is different or not. Only those string pairs which differ from each other by some number of bits (mating threshold) are mated. The initial threshold is set to $\mathrm{L}/4$ where L is the length of the string. When no offspring is inserted into the new population, the threshold is reduced by 1.

The population size was 50, randomly initialized with the exception of a single chromosome with the following setup:

- Linguistic RB part, $\mathrm{CS_{COR}}$, with the N rules obtained by the WM-method.
- Connectors part:
  - Conjunction, $\mathrm{CS_C}$, with the N genes initiated to 0, in order to make Dubois t-norm equivalent to Minimum t-norm initially.
  - Defuzzification, depending on the defuzzifier considered:
    - $\mathrm{D_W}, \mathrm{CS_{D_W}}$, with the N genes initiated to 1, with the objective to make it act like the model without weights, equivalent to the well known WCOA defuzzification method.
    - $\mathrm{D_{SLIDE}}, \mathrm{CS_{D_{SLIDE}}}$, with the 2 genes initiated to 0, equivalent to WCOA defuzzification method.

Fig. 6 illustrates the initial values mentioned with weighted defuzzification method, while Fig. 7 shows the best fitness chromosome of the population, after the evolutionary process. It shows changes in the consequents of some rules, their associ-

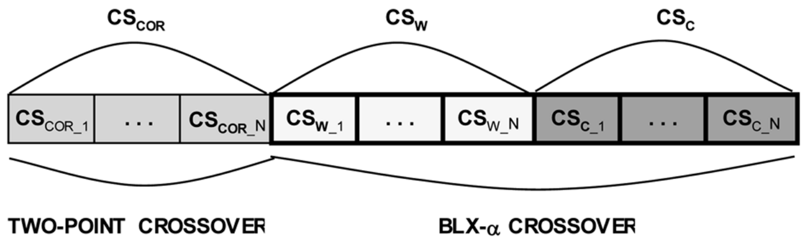

Fig. 8. Genetic representation and crossover scheme for the weighted defuzzification type model.

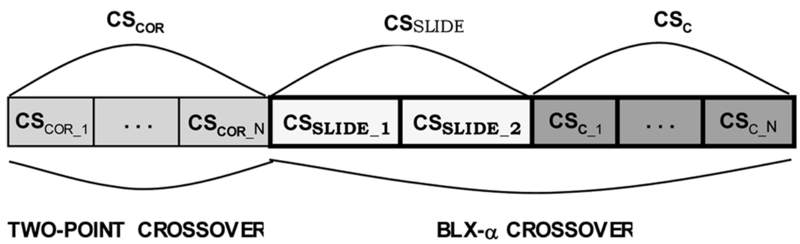

Fig. 9. Genetic representation and crossover scheme for SLIDE defuzzification type model.

ated weights are different and the matching degree of every rule has been modified too.

The problem representation and crossover operators selected are the following, depending on the chromosome part:

- Linguistic RB part, $(\mathrm{CS_{COR}})$: It is an integer-valued vector. A standard two-point crossover has been used. It is graphically illustrated on the left zones of Figs. 8 and 9.
- Connectives part $(\mathrm{CS_C}$ and $\mathrm{CS_{DW}}$ or $\mathrm{CS_{D_{SLIDE}}})$: It is a real-valued vector graphically illustrated on the middle and right zones of Figs. 8 and 9.

Two different thresholds have been used: one for the real-valued vector, and another one for the integer-valued. Both are initially set to $\mathrm{L}/4$, where L is the length of the vector in the integer valued, or the string in the real-valued. When no offspring is inserted into the new population, the threshold is reduced by 1, independently.

The fitness function used was the classical Mean Square Error, MSE $(\cdot)$ for a fuzzy model

$$\mathrm{MSE}(S[i])_B = \frac{\frac{1}{2}\sum_{k=1}^{P}(y_k - S[i](x_k))^2}{\mathrm{p}}$$

where $S[i]$ denotes the fuzzy model whose inference system uses the Dubois t-norm as conjunction operator, rule connective Minimum t-norm, and defuzzification method $\mathrm{D_i}$ (with $\mathrm{i} = 1$ for $\mathrm{D_w}$ method and $\mathrm{i} = 2$ for $\mathrm{D_{SLIDE}}$). This measure uses a set of system evaluation data formed by P pairs of numerical data $\mathrm{Z_k} = (\mathrm{x_k}, \mathrm{y_k}), \mathrm{k} = 1, \ldots, \mathrm{P}$, with x_k being the values of the input variables, and with y_k being the corresponding values of the associated output variables.

IV. Experimental Study

We analyze the cooperation among the different elements developing an experimental study with three different applications where we compare the accuracy results. To do so, we provide a statistical analysis using some statistical tests (ANOVA, Levene and Tamhane [2]) with the aim of showing the significance in the accuracy improvements obtained with the proposed model.

The following four subsections introduce the problems depiction, describe the experimental methodology, show the results and analysis, and exemplify the analysis of the resulting RBs respectively.

A. *Description of the Problems*

Two electrical distribution problems described in [12] and a classical application of a rice taste evaluation problem [25], [34], have been selected to analyze the performance of the cooperative model in fuzzy modeling. The first application, $E_{1,}$ is the estimation of the low voltage network real length in rural villages, the second application, E_2, is the estimation of the electrical medium voltage network maintenance cost in a town, and the third application, rice taste data.

E_1 *Application:* The data set has two inputs and a single output from 495 villages. The input variables are the *number of clients in the consigned population* which domain is [1, 320] and *the radius of that population in the sample* which domain is [60, 1673]. The output variable is *the estimation of the real length in a particular village* that takes values in the interval [80, 7675]. The input and output variable domains have been partitioned with seven labels $\{\mathrm{ES_i}, \mathrm{VS_i}, \mathrm{S_i}, \mathrm{M_i}, \mathrm{L_i}, \mathrm{VL_i}, \mathrm{EL_i}\}$, (with $\mathrm{i} = 1, 2$ for the two antecedents, and $\mathrm{i} = 3$ for the consequent), as shown in Fig. 10, with the following meaning:

ES is extremely small,
VS is very small,
S is small, M is medium,
L is large,
VL is very large, and
EL is extremely large.

Three kinds of linguistic RBs have been obtained: they are composed of 20 to 24 linguistic rules depending on the partition, obtained with the Wang and Mendel method [42], with the

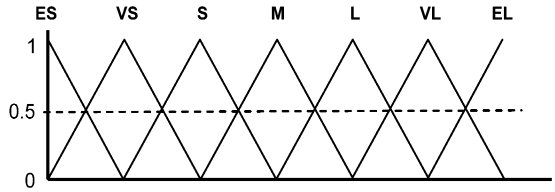

Fig. 10. Fuzzy partition considered for the input and output variables of E_1.

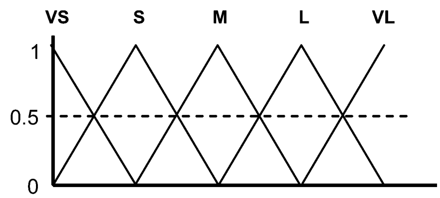

Fig. 11. Fuzzy partition considered for the input and output variables of E_2.

COR method [4], and with the presented cooperative model. All of them have been obtained from data training sets of 80% of the original available data, that is, 396 villages taken randomly. We have considered 5-fold cross validation, therefore we get 5 linguistic RBs associated with the 5 training sets. The evaluation of the different fuzzy models composed have been carried out with the remaining 20% of the initial data set, that is, with data from 99 villages.

E_2 *Application:* The second electrical distribution problem, E_2, has got a data set of 1059 cities with four input variables and a single output. The input variables are *the sum of the lengths of all streets in the town*, which domain is [0.5,11], *the total area of the town*, which domain is [0.15,8.55], *the area that is occupied by buildings*, which domain is [1.64, 142.5], and *the energy supply to the town* which domain is [1, 165]. The output variable is *the maintenance costs of medium voltage line* which domain is [0, 8546.03]. The fuzzy partition used for inputs and output has 5 labels $\{\mathrm{VS}, \mathrm{S}, \mathrm{M}, \mathrm{L}, \mathrm{VL}\}$, (with $\mathrm{i} = 1$ to 4 for the antecedents, and $\mathrm{i} = 5$ for the consequent), (see Fig. 11), where:

VS is very small,
L is large,
S is small,
VL is very large, and
M is middle.

We have three kinds of linguistic RBs composed of 65 linguistic rules achieved with the Wang and Mendel method [42], COR method [4] and the presented cooperative model They have also been obtained from training data sets of 80% of the original available data, that is, 847 cities taken randomly. Evaluation of the fuzzy models has been carried out with the remaining 20% of the initial data set, that is, with data from 212 cities. In the same way, we have considered a 5-fold cross validation.

The Rice Taste Evaluation Problem: Subjective qualification of food taste is a very important but difficult problem. In the case of the rice taste qualification, it is usually put into effect using a subjective evaluation called the *sensory test*. In this test, a group of experts, usually composed of 24 individuals, evaluate the rice according to a set of characteristics associated with it. These factors are *flavor, appearance, taste, stickiness*, and *toughness* [25].

Because of the large quantity of relevant variables, the problem of rice taste analysis becomes very complex, thus requiring the design of a model representing the existing nonlinear relationships. We used the data set presented in [25], [34]. This set is composed of 105 data arrays collecting subjective evaluations of the six variables in question (the five mentioned and the overall evaluation of the rice kind), made up by experts on the number of kinds of rice grown in Japan (for example, Sasanishiki, Akita-Komachi, etc.).

The six variables are normalized, thus taking values in the real interval. Because of the small number of examples used, there is a high risk of biasing the learning process. Thus, we have randomly obtained several partitions of the mentioned set (71% for training and 29% for test). In this way, 10 partitions of training and test sets with 75 and 30 pieces, respectively, are considered. This is the same experimental procedure developed by the authors in the paper where the example data set is presented [25], [34].

Two labels are considered for every linguistic variable domain.

B. Comparison Methodology

We built several fuzzy models combining the WM-method linguistic RB with the parameterized connectors and, on the other hand, we used the cooperative model proposed in combination with different parameterized connectors, in order to compare their accuracy solving the three different fuzzy model applications.

The whole set of fuzzy models are illustrated in Table III. First, we added the two initial non parameterized operators adaptive models based on WM-method and COR linguistic RBs. Next, we have the evolutionary connectives models, which are based on the linguistic RB learned with WM-method, and later altering their parameterized connectors using several combinations. Finally, the evolutionary cooperative models proposed, which learn the linguistic RB and the connectors at the same time, also using several combinations. Note that the evolutionary cooperative models are not marked in Table III as COR linguistic RB, because they are not using a COR previously learned linguistic RB.

We achieved 30 trials for every evolutionary process, running them with six different seeds for the random number generator and five different data sets, five-fold cross-validation approach for the two electrical problems E_1 and E_2, and with three different seeds and 10 different data sets for rice taste evaluation problem.

In order to compare the different fuzzy models obtained, we consider an usual fuzzy model performance measure, the MSE whose expression has been aforesaid. The considered real MSE was computed as the arithmetic mean of the 30 results.

The evolutionary models have been run for different amount of evaluations depending on the particular fuzzy model to be learned. Table IV shows these values.

TABLE III
FUZZY MODELS BUILT FOR THE EXPERIMENTAL STUDY

Abbreviation	WM RB	COR RB	Adaptive Conjunc.	Adaptive Defuzz. w. Weights	Adaptive Defuzz. SLIDE
Reference Models					
WM	•				
COR		•			
Evolutionary Connectives Models					
WM+D_W	•			•	
WM+D_{SLIDE}	•				•
WM+C	•		•		
WM+C+D_W	•		•	•	
WM+C+D_{SLIDE}	•		•		•
Evolutionary Cooperative Models: Evolutionary RB and Connectives					
D_W – COR				•	
D_{SLIDE} - COR					•
C - COR			•		
C – D_W - COR			•	•	
C – D_{SLIDE} - COR			•		•

TABLE IV
EVALUATIONS PERFORMED BY THE EVOLUTIONARY CHC MODEL

Fuzzy Model	Application		
	E_1	E_2	Rice
WM+D_W	40000	40000	10000
WM+D_{SLIDE}	5000	5000	2000
WM+C	40000	40000	10000
WM+C+D_W	60000	60000	15000
WM+C+D_{SLIDE}	40000	40000	10000
D_W - COR	200000	200000	40000
D_{SLIDE} - COR	150000	150000	30000
C – COR	200000	200000	40000
C – D_W - COR	300000	300000	60000
C – D_{SLIDE} - COR	210000	210000	45000

TABLE V
MSE FOR THE FUZZY MODEL OF ELECTRICAL APPLICATION E_1.
MEAN # RULES: 22

	Electrical Problem E_1	
Model	MSE_{tra}	MSE_{tst}
WM	211776.69	227583.28
COR	177152.20	208572.24
WM+D_W	190899.21	220496.80
WM+D_{SLIDE}	199221.41	226288.24
WM+C	190526.06	229123.41
WM+C+D_W	186121.31	224284.03
WM+C+D_{SLIDE}	190692.99	224848.88
D_W – COR	147837.55	**184008.67**
D_{SLIDE} – COR	174812.43	221010.64
C – COR	165470.80	202325.15
C – D_W – COR	**139507.39**	186695.08
C – D_{SLIDE} – COR	166352.38	211718.01

C. Results and Analysis

We have organized this section in three subsections: First devoted to MSE, second to the statistical analysis performed, and third to study an example of the obtained solutions.

1) General Results and Analysis: The MSE is shown in Tables V, VI and VII for application E_1, E_2 and rice taste evaluation problem respectively. Tables show in two columns the values for training and test.

We can point out some conclusions analyzing them:

TABLE VI
MSE FOR THE FUZZY MODEL OF ELECTRICAL APPLICATION E_2.
MEAN # RULES: 65

	Electrical Problem E_2	
Model	MSE_{tra}	MSE_{tst}
WM	56135.75	56359.42
COR	50710.82	54584.76
WM+D_W	31443.27	34879.16
WM+D_{SLIDE}	46788.34	49422.80
WM+C	34371.65	36845.51
WM+C+D_W	23291.91	25016.47
WM+C+D_{SLIDE}	30996.38	31567.10
D_W – COR	27077.00	29640.10
D_{SLIDE} – COR	45453.94	48944.55
C – COR	27239.07	29605.33
C – D_W – COR	**17832.85**	**19855.88**
C – D_{SLIDE} – COR	27333.90	29389.17

TABLE VII
MSE FOR THE FUZZY MODEL OF RICE TASTE EVALUATION PROBLEM.
MEAN # RULES: 15

	Rice taste evaluation problem	
Model	MSE_{tra}	MSE_{tst}
WM	0.013284	0.013119
COR	0.007979	0.008244
WM+D_W	0.003180	0.003161
WM+D_{SLIDE}	0.005656	0.006270
WM+C	0.002324	0.001978
WM+C+D_W	0.001902	0.001882
WM+C+D_{SLIDE}	0.002776	0.002288
D_W – COR	0.002458	0.002934
D_{SLIDE} – COR	0.003159	0.003628
C – COR	0.001420	**0.001724**
C – D_W – COR	**0.001139**	0.001802
C – D_{SLIDE} – COR	0.001427	0.001847

a) *Parameterized connectives:*
   - The parameterized connectives show that they are a good tool in order to improve the accuracy of the FS, as it was also showed in [1], [9].
   - The results obtained with parameterized conjunction and defuzzification together improve the accuracy gained with the parameterized conjunction or defuzzification alone. Thus, parameters in the conjunction and defuzzification cooperate and get better precision.
   - In two applications (E_2 and rice), the evolutionary learning of the inference and defuzzification together obtain higher accuracy than COR method without adaptive connectors. Therefore, evolutionary connectors are a good tool for linguistic fuzzy model designers.

b) *Cooperative evolutionary learning model:*
   - Noticeably, cooperative models show the best results of the practical study.
   - Cooperation between fuzzy rules, and connectors is clearly noticeable when comparing the MSE obtained by COR method and the one obtained with complete cooperative methods.

c) *Globally:*
   - The best results are shown by the one with more degrees of freedom: the cooperative model with parametric t-norm with the weighted defuzzification method.

TABLE VIII
ANOVA SUMMARY TABLE, FOR PROBLEM E_1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	75828095896.302	11	6893463263.300	11.786	0.000
Within Groups	188920474593.440	323	584893110.196		
Total	264748570489.742	334			

TABLE IX
ANOVA SUMMARY TABLE, FOR PROBLEM E_2

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	38682467685.980	11	3516587971.453	267.190	0.000
Within Groups	4251117396.729	323	13161354.169		
Total	42933585082.709	334			

TABLE X
ANOVA SUMMARY TABLE, FOR RICE TASTE EVALUATION PROBLEM

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	0.002	11	0.000	123.414	0.000
Within Groups	0.001	328	0.000		
Total	0.003	339			

TABLE XI
SUMMARY TABLE FROM TAMHANE TEST, FOR E_1 PROBLEM

	WM	COR	WM+D_W	WM+D_{SLIDE}	WM+C	WM+C+D_W	WM+C+D_{SLIDE}	D_W – COR	D_{SLIDE} - COR	C - COR	C – D_W - COR	C – D_{SLIDE} – COR
WM		≈	≈	≈	≈	≈	≈	≈	≈	≈	≈	≈
COR	≈		≈	≈	≈	≈	≈	-	≈	≈	-	≈
WM+D_W	≈	≈		≈	≈	≈	≈	-	≈	-	-	≈
WM+D_{SLIDE}	≈	≈	≈		≈	≈	≈	-	≈	-	-	≈
WM+C	≈	≈	≈	≈		≈	≈	-	≈	-	-	≈
WM+C+D_W	≈	≈	≈	≈	≈		≈	-	≈	-	-	≈
WM+C+D_{SLIDE}	≈	≈	≈	≈	≈	≈		-	≈	-	-	≈
D_W – COR	≈	+	+	+	+	+	+		+	+	≈	+
D_{SLIDE} – COR	≈	≈	≈	≈	≈	≈	≈	-		≈	≈	≈
C – COR	≈	≈	+	+	+	+	+	-	≈		-	≈
C – D_W – COR	≈	+	+	+	+	+	+	≈	≈	+		+
C – D_{SLIDE} – COR	≈	≈	≈	≈	≈	≈	≈	-	≈	≈	-	

TABLE XII
SUMMARY TABLE FROM TAMHANE TEST, FOR E_2 PROBLEM

	WM	COR	WM+D_W	WM+D_{SLIDE}	WM+C	WM+C+D_W	WM+C+D_{SLIDE}	D_W – COR	D_{SLIDE} - COR	C - COR	C – D_W - COR	C – D_{SLIDE} - COR
WM		≈	-	≈	≈	-	-	-	≈	-	-	-
COR	≈		-	-	-	-	-	-	-	-	-	-
WM+D_W	+	+		+	≈	-	-	-	+	-	-	-
WM+D_{SLIDE}	≈	+	-		-	-	-	-	≈	-	-	-
WM+C	≈	+	≈	+		-	-	-	+	-	-	-
WM+C+D_W	+	+	+	+	+		+	+	+	+	-	+
WM+C+D_{SLIDE}	+	+	+	+	+	-		≈	+	≈	-	≈
D_W – COR	+	+	+	+	+	-	≈		+	≈	-	≈
D_{SLIDE} – COR	≈	+	-	≈	-	-	-	-		-	-	-
C – COR	+	+	+	+	+	-	≈	≈	+		-	≈
C – D_W – COR	+	+	+	+	+	+	+	+	+	+		+
C – D_{SLIDE} – COR	+	+	+	+	+	-	≈	≈	+	≈	-	

TABLE XIII
SUMMARY TABLE FROM TAMHANE TEST, FOR RICE TASTE PROBLEM

	WM	COR	WM+D_W	WM+D_{SLIDE}	WM+C	WM+C+D_W	WM+C+D_{SLIDE}	D_W – COR	D_{SLIDE} - COR	C - COR	C – D_W - COR	C – D_{SLIDE} - COR
WM		≈	-	-	-	-	-	-	-	-	-	-
COR	≈		-	≈	-	-	-	-	-	-	-	-
WM+D_W	+	+		+	-	-	-	≈	≈	-	-	-
WM+D_{SLIDE}	+	≈	-		-	-	-	-	-	-	-	-
WM+C	+	+	+	+		≈	≈	+	+	≈	≈	≈
WM+C+D_W	+	+	+	+	≈		≈	+	+	≈	≈	≈
WM+C+D_{SLIDE}	+	+	+	+	≈	≈		+	+	-	-	≈
D_W – COR	+	+	≈	+	-	-	-		≈	-	-	-
D_{SLIDE} – COR	+	+	≈	+	-	-	-	≈		-	-	-
C – COR	+	+	+	+	≈	≈	+	+	+		≈	≈
C – D_W – COR	+	+	+	+	≈	≈	+	+	+	≈		≈
C – D_{SLIDE} - COR	+	+	+	+	≈	≈	≈	+	+	≈	≈	

2) Statistical Study: To compare the results provided by the different models, we develop a statistical analysis. First, we compute some tables of descriptive statistics (see Tables XVII, XIX and XXI in Appendix), where the mean values, standard deviation, and so on are showed. Later we use the ANOVA analysis of one factor [2] for each model to be used for that purpose; the factor being the models used on the test data sets. See Tables VIII, IX and X, for E_1, E_2 and rice taste evaluation problem respectively. Given that significant differences were found for all models with respect to the mean result values associated with the different models analyzed, we performed a Tamhane means rank test [2] (see the Statistical Study Developed in Appendix), with a confidence coefficient of 95% due to the case of hypothesis of equality of variances of the results was rejected in all of the analysis performed for each method (Levene test, see Tables XVIII, XX and XXII in Appendix).

With the aim of summarizing the results of Tamhane tests for multiple comparisons, we built Tables XI, XII and XIII, they show a summary of Tamhane test for every application. Sign $(+)$ means that the selected row (fuzzy model) improves the selected column (fuzzy model), while sign $(-)$ means the contrary. Sign $(\approx)$ means they are similar. Tables must be read beginning with the file and after the column, i.e., *the model D_W-COR (row 8) improves the model COR (column 2) in Table XI.*

Considering these tables, we can point out that:

- The whole cooperative model proposed with adaptive conjunction together with adaptive defuzzification presents the best results, in particular, when adaptive defuzzification is based on weights (D_W) we find the best results.
- The MSE improvements may depend on the application, because applications E_2 and rice show better improvements than E_1.

TABLE XIV
CHROMOSOME INITIALIZED WITH THE WM-METHOD RB, PARAMETERIZEDT-NORM AS MINIMUM AND PARAMETERIZED DEFUZZIFICATIONAS GC WEIGHTED BY THE MATCHING. $\mathrm{MSE}_{\mathrm{TRA}} = 202698.35, \mathrm{MSE}_{\mathrm{TST}} = 210365.15$, FOR APPLICATION E_1

# R 24 / X_1 \ X_2	ES_2	VS_2	S_2	M_2	L_2	VL_2	EL_2
ES_1	ES_3 0.00 1.00	VS_3 0.00 1.00	VS_3 0.00 1.00	S_3 0.00 1.00		VS_3 0.00 1.00	M_3 0.00 1.00
VS_1	ES_3 0.00 1.00	VS_3 0.00 1.00	VS_3 0.00 1.00	L_3 0.00 1.00	S_3 0.00 1.00	L_3 0.00 1.00	
S_1		VS_3 0.00 1.00	M_3 0.00 1.00	S_3 0.00 1.00	S_3 0.00 1.00	L_3 0.00 1.00	
M_1		VS_3 0.00 1.00	S_3 0.00 1.00	VL_3 0.00 1.00	EL_3 0.00 1.00	M_3 0.00 1.00	
L_1			M_3 0.00 1.00				
VL_1							
EL_1				S_3 0.00 1.00			

TABLE XV
BEST ADAPTED CHROMOSOME AFTER THE EVOLUTIONARY PROCESS. $\mathrm{MSE}_{\mathrm{TRA}} = 138891.40, \mathrm{MSE}_{\mathrm{TST}} = 174889.92$, FOR APPLICATION E_1 USING C-$\mathrm{D_W}$-COR

# R 24 / X_1 \ X_2	ES_2	VS_2	S_2	M_2	L_2	VL_2	EL_2
ES_1	ES_3 0.05 0.35	ES_3 0.79 2.29	VS_3 0.23 3.89	VS_3 0.04 2.19		VS_3 1.00 3.03	M_3 0.65 6.18
VS_1	ES_3 0.93 7.37	VS_3 1.00 9.99	S_3 0.70 6.34	M_3 0.91 5.36	S_3 1.00 0.52	L_3 0.55 2.28	
S_1		VS_3 1.00 0.03	M_3 0.06 0.01	VS_3 1.00 0.13	S_3 0.47 0.01	L_3 0.68 0.01	
M_1		VS_3 0.75 0.00	S_3 0.58 0.01	VL_3 0.57 0.27	EL_3 1.00 0.49	M_3 0.16 9.99	
L_1			M_3 0.35 0.01				
VL_1							
EL_1				S_3 0.50 9.59			

TABLE XVI
RB OBTAINED WITH THE COR METHOD. $\mathrm{MSE}_{\mathrm{TRA}} = 181910.54, \mathrm{MSE}_{\mathrm{TST}} = 179266.96$, FOR APPLICATION E_1

# R 24 / X_1 \ X_2	ES_2	VS_2	S_2	M_2	L_2	VL_2	EL_2
ES_1	ES_3	ES_3	VS_3	VS_3		VS_3	M_3
VS_1	**VS_3**	VS_3	S_3	M_3	**M_3**	L_3	
S_1		VS_3	M_3	**S_3**	S_3	L_3	
M_1		VS_3	S_3	VL_3	EL_3	M_3	
L_1			M_3				
VL_1							
EL_1				S_3			

D. Analysis of the Rule Bases: A Study on the Application E_1 With Methods C-D_W-COR, COR and WM

Now we analyze an example of a complete chromosome for application E_1. We have represented the chromosome initialized with the WM-method linguistic RB, the Dubois t-norm as the minimum ($\alpha_{\mathrm{i}} = 0$) and the defuzzification method as the GC weighted by the matching with weights ($\beta_{\mathrm{i}} = 1$), in Table XIV.

Table XV shows the best adapted chromosome after the evolutionary process, that is, for every rule, first the consequent, below the parameter of the t-norm, and in the lower the parameter of defuzzification.

For example, observing Table XIV we can see the rule

If X_1 is ES_1 andX_2 is VS_2 then Y is VS_3

$$\text{with } \alpha_i = 0 \text{ and } \beta_i = 1,$$

while viewing Table XV, we can see the result of learning process experimented by the same rule

If X_1 is ES_1 and X_2 is VS_2 then Y is ES_3

$$\text{with } \alpha_i = 0.79 \text{ and } \beta_i = 2.29$$

so the consequent has changed and the values of aggregation and defuzzification method have been tuned.

Table XVI has been added to compare the consequents obtained with COR and the ones obtained with C-$\mathrm{D_W}$-COR in Table XV. The three tables have been obtained using the same data set.

Analyzing them we can point out the following.

- Comparing Tables XIV and XV, the initial linguistic RB, obtained with WM-method, has got the consequents with the highest performance in each subspace. Whereas the adapted linguistic RB, built with C-$\mathrm{D_W}$-COR, has got five different consequents, that is, it uses different consequents from the best one when they allow, together with the parameters, the linguistic fuzzy model to be more accurate.
- Comparing the linguistic RB obtained with COR-method in Table XVI, with the linguistic RB obtained with C-$\mathrm{D_W}$ -COR in Table XV, three differences can be observed, that corroborate the cooperation between the connectors

TABLE XVII
DESCRIPTIVE STATISTICS, FOR APPLICATION E_1

Algorithms	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimun	Maximun
					Lower Bound	Upper Bound		
WM	5	227583.28	22253.75120	9952.18009	199951.60	255214.96	196974.7	252227.9
COR	30	208572.23	30114.52577	5498.13502	197327.28	219817.18	175267.0	254778.4
WM+D_W	30	220496.79	18152.29187	3314.13991	213718.62	227274.97	189257.5	248690.9
WM+D_{SLIDE}	30	226288.23	6616.60429	1208.02114	223817.55	228758.91	216578.6	242156.2
WM+C	30	229123.41	19968.52668	3645.73750	221667.04	236579.78	203396.0	263376.8
WM+C+D_W	30	224284.03	25776.78864	4706.17620	214658.81	233909.24	186469.4	284759.1
WM+C+D_{SLIDE}	30	224848.88	4484.32273	818.72157	223174.40	226523.35	215356.6	231355.7
D_W - COR	30	184008.67	9304.19895	1698.70655	180534.42	187482.91	165690.5	206325.0
D_{SLIDE} - COR	30	221010.64	50847.17256	9283.38113	202023.99	239997.28	168979.4	377295.0
C - COR	30	202325.15	19594.35522	3577.42345	195008.50	209641.80	172862.1	231386.3
C – D_W - COR	30	186695.07	11545.87033	2107.97788	182383.78	191006.37	166814.6	210310.5
C – D_{SLIDE} - COR	30	211718.01	29891.11583	5457.34614	200556.48	222879.54	170910.9	351128.3
Total	335	212892.69	28154.22509	1538.22964	209866.85	215918.52	165690.5	377295.0

and the linguistic RB. The consequents have been selected taking into account the best global accuracy together with the adaptive operators.

We have also analyzed this question on the problem E_2 and we find five different consequents between COR and WM, and we also find six different consequents between C-D_W-COR and COR, with the following errors:

	$C - D_W - COR$	COR	WM
MSE_{tra}	19265.306641	51236.710938	56081.953125
MSE_{tst}	15046.996094	48863.687500	54982.914062

- Studying the parameters of the t-norm (the upper ones in figures), the higher values of α mean that these rules are connected with product mainly: there are six rules with $\alpha = 1$, and two more rules with values near to 1. Although, the lower values mean they are using minimum predominantly, there are four rules connecting with minimum. Intermediate values of α (twelve rules) mean they use minimum when any of the antecedents have got a match over α_i, and a connective between product and minimum in other cases. Thus, each rule works with its own particular connective.
- Some rules seem to be highly penalized with very low weights ($\beta_i \approx 0$). On the other hand, their consequents have not been changed by the evolutionary cooperative mechanism. Hence, we think perhaps this rules have a low importance or may not be necessary. The version of COR methodology uses the antecedent parts of the rules generated by WM-method.

V. CONCLUDING REMARKS

In this work we have proposed an evolutionary learning model where the linguistic RB and the aggregation connector parameters are learnt together. This fact allows them to cooperate, they are not the best choice locally, but they are a good choice to work together.

TABLE XVIII
LEVENE TEST OF HOMOGENEITY OF VARIANCE, FOR APPLICATION E_1

Levene Statistic	df1	df2	Sig.
11.131	11	323	0.000

This methodology improves accuracy in comparison with the linguistic RB learning process. Its accuracy has been shown in practice with three different applications, performing a statistical study.

In the framework of the tradeoff between precision and interpretability of linguistic FSs, the positive synergy between the different components is a helpful tool. We obtain the linguistic RB with specific conjunction operators and defuzzification parameters per rule, and as we have mentioned before, even though parameters sacrifice a part of the system interpretability, the overall objective is to get the best accuracy with the lowest loss of interpretability.

Finally, we would like to point out future studies for high dimensional problems. The following two recent contributions [28], [19] deal with the scaling up of two genetic learning algorithms for high dimensional classification problems. The present proposal has been analyzed with three applications that use 2, 4 and 6 input variables respectively. As future work we are interested in the analysis of the behavior of the cooperative evolutionary learning proposal with high dimensional problems, where it might be necessary to include a feature selection component into the evolutionary learning approach.

APPENDIX
STATISTICAL STUDY. DESCRIPTION AND RESULTS

The statistical analysis developed begins with the computation of some descriptive statistics, collected in Tables XVII, XIX and XXI. Below, we describe the columns of these tables.

TABLE XIX
DESCRIPTIVE STATISTICS, FOR APPLICATION E_2

Algorithms	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimun	Maximun
					Lower Bound	Upper Bound		
WM	5	56359.418	5238.58609	2342.76692	49854.8543	62863.9817	49048.29	62719.16
COR	30	54584.764	6580.16861	1201.36893	52127.6895	57041.8402	48863.69	66479.02
WM+D_W	30	34879.160	4196.81911	766.23083	33312.0420	36446.2780	29416.93	41563.66
WM+D_{SLIDE}	30	49422.800	2000.67263	365.27118	48675.7371	50169.8640	46456.78	54920.46
WM+C	30	36845.509	2119.73601	387.00908	36053.9866	37637.0314	32421.87	39390.57
WM+C+D_W	30	25016.474	3832.21705	699.66391	23585.5011	26447.4479	18362.29	30374.72
WM+C+D_{SLIDE}	30	31567.105	2367.63417	432.26888	30683.0163	32451.1946	25657.16	36123.49
D_W - COR	30	29640.101	2817.59324	514.41979	28587.9947	30692.2079	23912.77	35023.84
D_{SLIDE} - COR	30	48944.545	4827.16905	881.31646	47142.0506	50747.0397	42577.07	57750.34
C - COR	30	29605.329	2878.02169	525.45247	28530.6580	30680.0000	24883.61	34210.16
C – D_W - COR	30	19855.881	2520.71694	460.21784	18914.6302	20797.1325	14715.53	23057.68
C – D_{SLIDE} - COR	30	29389.165	2688.52518	490.85530	28385.2539	30393.0775	23611.15	32512.08
Total	335	35744.245	11337.70999	619.44527	34525.7395	36962.7512	14715.53	66479.02

TABLE XX
LEVENE TEST OF HOMOGENEITY OF VARIANCE, FOR APPLICATION E_2

Levene Statistic	df1	df2	Sig.
10.260	11	323	0.000

It begins with N which is the number of subjects sampled assigned to each algorithm. The Mean is the sum of all scores divided by the number of scores, that is, the arithmetic average. The Standard Deviation informs about how far the scores are from the mean on average, so it shows the degree of observations tending to cluster near the center of the distribution. The Standard Error is an estimation of the Standard Deviation of the Mean if repeated samples of the same size were taken from the same population. It can be calculated by dividing a sample's standard deviation by the square root of the number in the sample. It is used in calculating the 95% of confidence interval for the simple mean. The 95% confidence interval for Mean have been obtained from the sample mean, standard deviation, and sample size. This confidence interval means that if we were to repeatedly perform the study and computed the confidence intervals for each simple drawn, on average, 95 out of each 100 confidence intervals would contain the true population mean. It combines measures of both central tendency (mean) and variation (standard error) to provide information about where we should expect the population mean to fall. Finally, the Minimun and Maximun are the lowest and highest values of all scores.

ANOVA (analysis of variance) test is performed in order to determine whether there are differences in the means between groups or across different conditions. It is used to determine if the means are far enough apart to be considered "significantly" different.

The basic logic of significance testing is that we will assume that the population groups have the same mean (null hypothesis), then determine the probability of obtaining a sample with group mean differences as large (or larger) as what we find in our data. To make this assessment the amount of variation among the group means (between-group variation) is compared to the amount of variation among the observations within each group (within-group variation). Assuming that in the population the group means are equal (null hypothesis), the only source of variation among the sample means would be the fact that the groups are composed of different individual observations. Thus, the ratio of the two sources of variation (between-group/within-group) should be about one when there are no population differences. When the distribution of the individual observations within each group follows the normal curve, the statistical distribution of this ratio is known (F distribution) and we can make a probability statement about the consistency of our data with the null hypothesis. The final result is the probability of obtaining sample differences as large (or larger) as what we found, if there were no population differences. If this probability is sufficiently small (usually less than 0.05, i.e., less than 5 chances in 100), we conclude the population groups differ.

The Levene test shows the assumption of homogeneity of variance. It indicates that the variances do not differ across groups. First, the test will assume that the population groups have the same variance (null hypothesis). The df1 y df2 column contains information about the degrees of freedom. The F column contains information about Levene statistic to decide if it is far enough from one to say that the group variances are not equal. The significance (Sig.) column indicates that under the null hypothesis of no group differences. If this probability is sufficiently small (usually less than 0.05, i.e., less than 5 chances in 100) we conclude the population groups differ.

The purpose of post hoc testing is to determine exactly which groups differ from each other in terms of mean differences. This

TABLE XXI
DESCRIPTIVE STATISTICS, RICE TASTE EVALUATION PROBLEM

Algorithms	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimun	Maximun
WM	10	0.01311870	0.004468455	0.001413050	0.00992216	0.01631524	0.007504	0.020754
COR	30	0.00824400	0.001102983	0.000201376	0.00783214	0.00865586	0.006908	0.010177
WM+D_W	30	0.00316090	0.000795976	0.000145325	0.00286368	0.00345812	0.001468	0.004605
WM+D_{SLIDE}	30	0.00627030	0.002866194	0.000523293	0.00520005	0.00734055	0.002386	0.011530
WM+C	30	0.00197833	0.000405931	0.000074112	0.00182676	0.00212991	0.001371	0.003072
WM+C+D_W	30	0.00188150	0.000351720	0.000064215	0.00175017	0.00201283	0.001210	0.002347
WM+C+D_{SLIDE}	30	0.00228750	0.000558351	0.000101941	0.00207901	0.00249599	0.001444	0.003155
D_W - COR	30	0.00293447	0.000798646	0.000145812	0.00263625	0.00323269	0.001253	0.004227
D_{SLIDE} - COR	30	0.00362810	0.001084425	0.000197988	0.00322317	0.00403303	0.001982	0.005687
C - COR	30	0.00172357	0.000401888	0.000073374	0.00157350	0.00187363	0.001106	0.002527
C – D_W - COR	30	0.00180220	0.000370321	0.000067611	0.00166392	0.00194048	0.001193	0.002557
C – D_{SLIDE} - COR	30	0.00184717	0.000421635	0.000076980	0.00168973	0.00200461	0.001114	0.002919
Total	340	0.00354096	0.002901653	0.000157364	0.00323143	0.00385050	0.001106	0.020754

TABLE XXII
LEVENE TEST OF HOMOGENEITY OF VARIANCE, FOR RICE TASTE EVALUATION PROBLEM

Levene Statistic	df1	df2	Sig.
32.077	11	328	0.000

is usually done after the original ANOVA test indicates that all groups are not identical. The Tamhane test determines the Multiple Comparisons between algorithms for each one with others. This test uses the Welch procedure for determining degrees of freedom for the square error of the contrast. It uses Student's t distribution, and the Sidak procedure to find the alpha level. It is appropriated when variances are unequal or when variances and group sizes are unequal. Due to the great space occupied by the Tamhane results tables for multiple comparisons for each application, we decided to put in this work the summary Tables XI, XII and XIII.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their valuable comments.

REFERENCES

[1] J. Alcala-Fdez, F. Herrera, F. Márquez, and A. Peregrín, "Increasing fuzzy rules cooperation based on evolutionary adaptive inference systems," *Int. J. Intell. Syst.*, 2006, In press.

[2] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York: Wiley, 1984.

[3] J. J. Buckley and Y. Hayashi, "Can approximate reasoning be consistent?," *Fuzzy Sets Syst.*, vol. 65, no. 1, pp. 13–18, 1994.

[4] J. Casillas, O. Cordón, and F. Herrera, "COR: A methodology to improve ad hoc data-driven linguistic rule learning methods by inducing cooperation among rules," *IEEE Trans. Syst., Man, Cybern.—Part B: Cybern.*, vol. 32, no. 4, pp. 526–537, 2002.

[5] J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, *Interpretability Issues in Fuzzy Modeling*. New York: Springer-Verlag, 2003.

[6] J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, *Accuracy Improvements in Linguistic Fuzzy Modeling*. New York: Springer-Verlag, 2003.

[7] J. S. Cho and D. J. Park, "Novel fuzzy logic control based on weighting of partially inconsistent rules using neural network," *J. Intel. Fuzzy Syst.*, vol. 8, pp. 99–100, 2000.

[8] O. Cordón, F. Herrera, F. Hoffmann, and L. Magdalena, *Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases*. Singapore, Republic of Singapore: World Scientific, 2001.

[9] O. Cordón, F. Herrera, F. A. Marquez, and A. Peregrín, "A study on the evolutionary adaptive defuzzification methods in fuzzy modelling," *Int. J. Hybrid Intell. Syst.*, vol. 1, no. 1, pp. 36–48, 2004.

[10] O. Cordón, F. Gomide, F. Herrera, F. Hoffmann, and L. Magdalena, "Ten years of genetic fuzzy systems: Current framework and new trends," *Fuzzy Sets Syst.*, vol. 41, pp. 5–31, 2004.

[11] O. Cordón, F. Herrera, and A. Peregrín, "Applicability of the fuzzy operators in the design of fuzzy logic controllers," *Fuzzy Sets Syst.*, vol. 86, pp. 15–41, 1997.

[12] O. Cordón, F. Herrera, and L. Sánchez, "Solving electrical distribution problems using hybrid evolutionary data analysis techniques," *Appl. Intell.*, vol. 10, pp. 5–24, 1999.

[13] K. A. Crockett, Z. Bandar, J. Fowdar, and J. O'Shea, "Genetic tuning of fuzzy inference within fuzzy classifier systems," *Exp. Syst.*, vol. 23, pp. 63–82, 2006.

[14] L. Davis, *Handbook of Genetic Algorithms*. New York: Van Nostrand Reinhold, 1991.

[15] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*. New York: Wiley, 2001.

[16] M. R. Emami, J. B. Türksen, and A. A. Goldenberg, "Development of a systematic methodology of fuzzy logic modelling," *IEEE Trans. Fuzzy Syst.*, vol. 6, pp. 346–361, 1998.

[17] L. J. Eshelman, G. J. E. Rawlins, Ed., "The CHC adaptive search algorithm: How to have safe search when engaging in Nontraditional genetic recombination," *Found. Genetic Algor.*, vol. 1, pp. 265–283, 1991.

[18] L. J. Eshelman, K. E. Mathias, and J. D. Schaffer, R. Belew, M. Vose, and M. Kaufmann, Eds., "Convergence controlled variation," *Found. Genetic Algor.*, vol. 4, pp. 203–224, 1997.

[19] A. González and R. Pérez, "An analysis of the scalability of an embedded feature selection model for classification problems," in *Proc. Eleventh Int. Conf. Inf. Processing Manage. Uncertainty in Knowledge-Based Syst.*, 2006, pp. 1949–1956.

[20] M. Gupta and J. Qi, "Theory of T-norms and fuzzy inference methods," *Fuzzy Sets Syst.*, vol. 40, pp. 431–450, 1991.

[21] F. Herrera, M. Lozano, and J. L. Verdegay, "Tackling real-coded genetic algorithms: Operators and tools for behavioural analysis," *Artif. Intell. Rev.*, vol. 12, no. 4, pp. 265–319, 1998.

[22] F. Herrera, F. A. Marquez, and A. Peregrín, "Genetic adaptation of rule connectives and conjunction operators in fuzzy rule based systems: An experimental comparative study," in *Proc. Third Int. Conf. European Soc. Fuzzy Logic Tech.*, 2003, pp. 100–104.
[23] C. Y. Huang, C. Y. Chen, and B. D. Liu, "Current-mode fuzzy linguistic hedge circuits," *Anal. Integr. Circuits Signal Process.*, vol. 19, pp. 225–278, 1999.
[24] H. Ishibuchi and T. Nakashima, "Effect of rule weights in fuzzy rule-based classification systems," *IEEE Trans. Fuzzy Syst.*, vol. 9, no. 4, pp. 506–515, 2001.
[25] H. Ishibuchi, K. Nozaki, H. Tanaka, Y. Hosaka, and M. Matsuda, "Empirical study on learning in fuzzy systems by rice taste analysis," *Fuzzy Sets Syst.*, vol. 64, pp. 129–144, 1994.
[26] H. Ishibuchi and T. Yamamoto, "Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining," *Fuzzy Sets Syst.*, vol. 141, pp. 59–88, 2004.
[27] H. Ishibuchi and T. Yamamoto, "Rule weight specification in fuzzy rule-based classification systems," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 4, pp. 428–435, 2005.
[28] H. Ishibuchi and Y. Nojima, "Tradeoff between accuracy and rule length in fuzzy rule-based classification systems for high-dimensional problems," in *Proc. Eleventh Int. Conf. Inf. Processing Manage. Uncertainty in Knowledge-Based Syst.*, 2006, pp. 1936–1943.
[29] D. Kim, Y. Choi, and S. Lee, "An accurate COG defuzzifier design using lamarckian co-adaptation of learning and evolution," *Fuzzy Sets Syst.*, vol. 130-2, pp. 207–225, 2002.
[30] B.-D. Liu, C. Y. Chen, and J. Y. Tsao, "Design of adaptative fuzzy logic controller based on linguistic-hedge concepts and genetic algorithms," *IEEE Trans. Syst., Man, Cybern.*, vol. 31, pp. 32–53, 2001.
[31] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolutionary Programs*. New York: Springer, 1992.
[32] M. Mizumoto, "Pictorial representations of fuzzy connectives, part I: Cases of T-norms, T-conorms and averaging operators," *Fuzzy Sets Syst.*, vol. 31, pp. 217–242, 1989.
[33] D. Nauck and R. Kruse, "How the learning of rule weights affects the interpretability of fuzzy systems," in *Proc. Seventh IEEE Int. Conf. Fuzzy Syst.*, 1998, pp. 1235–1240.
[34] K. Nozaki, H. Ishibuchi, and H. Tanaka, "A simple but powerful heuristic method for generating fuzzy rules from numerical data," *Fuzzy Sets Syst.*, vol. 86, no. 3, pp. 251–270, 1997.
[35] N. R. Pal and K. Pal, "Handling of inconsistent rules with an extended model of fuzzy reasoning," *J. Intell. Fuzzy Syst.*, vol. 7, pp. 55–73, 1999.
[36] L. Rutkowski and K. Cpalka, "Flexible neuro-fuzzy systems," *IEEE Trans. Neural Net.*, vol. 14, no. 3, pp. 554–574, 2003.
[37] L. Rutkowski, *Flexible Neuro-Fuzzy Systems: Structures, Learning and Performance Evaluation*. New York: Kluwer, 2004.
[38] L. Rutkowski and K. Cpalka, "Designing and learning of adjustable quasi-triangular norms with applications to neuro-fuzzy systems," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 1, pp. 140–151, 2005.
[39] I. SPSS, *SPSS 11.0 Advanced Models*. New York: SPSS Inc., 1999.
[40] M. Sugeno and T. Yasukawa, "A fuzzy-logic-based approach to qualitative modeling," *IEEE Trans. Fuzzy Syst.*, vol. 1, pp. 7–31, Feb. 1993.
[41] E. Trillas and L. Valverde, "On implication and indistiguishability in the setting of fuzzy logic," in *Management Decision Support Systems Using Fuzzy Sets and Possibility Theory*, J. Kacpryzk and R. Yager, Eds. New York: Verlag TüV Rheinland, 1985, pp. 198–212.
[42] L.-X. Wang and J. M. Mendel, "Generating fuzzy rules by learning from examples," *IEEE Trans. Syst., Man, Cybern.*, vol. 22-6, pp. 1414–1427, Nov. 1992.
[43] H. Wang, S. Kwong, Y. Jin, W. Wei, and K. F. Man, "Multi-objective hierarchical genetic algorithm for interpretable fuzzy rule-based knowledge extraction," *Fuzzy Sets Syst.*, vol. 149, pp. 149–186, 2005.
[44] H. Wu and J. M. Mendel, "On choosing models for linguistic connector words for mamdani fuzzy logic systems," *IEEE Trans. Fuzzy Syst.*, vol. 4, pp. 29–44, Feb. 2004.
[45] R. R. Yager and D. P. Filev, "SLIDE: A simple adaptive defuzzification method," *IEEE Trans. Fuzzy Syst.*, vol. 1, pp. 69–78, 1993.
[46] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning," *Inf. Sci.*, vol. 8, pp. 301–357, 1975.

Francisco Alfredo Márquez received the M.Sc. degree in computer science from the University of Sevilla, Sevilla, Spain, in 1999.

He is currently an Assistant Professor with the Department of Electronic Engineering, Computer Science and Automatics, University of Huelva, Huelva, Spain. His research interests include decision making problems in fuzzy environment, fuzzy rule-based systems, genetic algorithms, and genetic fuzzy systems.

Antonio Peregrín received the M.Sc. and Ph.D. degrees in computer science, both from the University of Granada, Granada, Spain, in 1995 and 2000, respectively.

Currently, he is an Associate Professor with the Department of Electronic Engineering, Computer Science and Automatics, University of Huelva, Huelva, Spain. His research interests include genetic fuzzy systems, fuzzy rule-based systems, evolutionary algorithms, linguistic modeling and metaheuristics.

Francisco Herrera received the M.Sc. degree in mathematics in 1988 and the Ph.D. degree in mathematics in 1991, both from the University of Granada, Spain.

He is currently a Professor in the Department of Computer Science and Artificial Intelligence at the University of Granada. He has published over 100 papers in international journals and he is coauthor of the book "*Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases*" (Singapore: World Scientific, 2001). As edited activities, he has co-edited three international books and co-edited 15 special issues in international journals on different Soft Computing topics, such as, "*Preference Modelling*", "*Computing with Words*", "*Genetic Algorithms*" and "*Genetic Fuzzy Systems*". He currently serves on the editorial boards of the the following journals: *Soft Computing, Fuzzy Sets and Systems, Mathware and Soft Computing*, *International Journal of Hybrid Intelligent Systems*, *International Journal of Computational Intelligence Research*, *International Journal of Information Technology and Intelligent and Computing and Evolutionary Intelligence*. His current research interests include computing with words, preference modeling and decision making, data mining and knowledge discovery, data preparation, genetic algorithms, fuzzy rule-based systems, and genetic fuzzy systems.