COR: A Methodology to Improve *Ad Hoc* Data-Driven Linguistic Rule Learning Methods by Inducing Cooperation Among Rules

Jorge Casillas, Oscar Cordón, and Francisco Herrera

***Abstract*—This paper introduces a new learning methodology to quickly generate accurate and simple linguistic fuzzy models, the cooperative rules (COR) methodology. It acts on the consequents of the fuzzy rules to find those best cooperating.**

Instead of selecting the consequent with the highest performance in each fuzzy input subspace as *ad hoc* data-driven methods usually do, the COR methodology considers the possibility of using another consequent, different from the best one, when it allows the fuzzy model to be more accurate thanks to having a rule set with best cooperation.

Our proposal has shown good results solving three different applications when compared to other methods.

***Index Terms*—Accuracy improvement, cooperative rules, linguistic fuzzy rule-based modeling, simulated annealing.**

I. Introduction

At present, system modeling is one of the main applications of fuzzy rule-based systems (FRBSs) [1], [2]. It may be considered as an approach to model a system making use of a descriptive language based on fuzzy logic with fuzzy predicates [3]. In this framework, one of the most interesting areas is *linguistic modeling*, where the interpretability of the obtained model is the main requirement. This task is developed by means of linguistic FRBSs, which use fuzzy rules composed of linguistic variables [4] that take values in a term set with a real-world meaning. Thus, the linguistic model consists of a set of linguistic descriptions regarding the behavior of the system being modeled [3].

Several tasks have to be performed in order to design an FRBS (linguistic model) for a concrete application. One of the most important and difficult ones is to derive an appropriate knowledge base (KB) about the problem being solved. The KB stores the available knowledge in the form of fuzzy linguistic IF–THEN rules. It is composed of the rule base (RB), constituted by the collection of rules in their symbolic forms, and the data base (DB), which contains the linguistic term sets and the membership functions defining their meanings.

The difficulty presented by human experts to express their knowledge in the form of fuzzy rules has made researchers develop automatic techniques to perform this task. In this sense, a large amount of methods has been proposed to automatically generate fuzzy rules from numerical data. Usually, they use complex rule generation mechanisms such as neural networks [5], [6] or genetic algorithms [7]. Opposite to them, a family of efficient and simple methods guided by covering criteria of the data in the example set, called *ad hoc data-driven methods*, has been proposed in the literature [8]–[11].

These methods come with interesting advantages:

1) they are easily understandable and implementable thanks to their simplicity;
2) they perform the learning process very quickly;
3) thanks to the two said advantages, they are very suitable to be used as a first stage of the modeling process to obtain a preliminary fuzzy model, which can be subsequently refined by other techniques [12] or be integrated within a meta-learning process [13].

Although the high performance of *ad hoc* data-driven methods has been clearly demonstrated, they have a problem related to the way of selecting the rules: *these methods usually look for the rules with the best individual performance*. Due to this, KBs with bad cooperation among the rules composing them are sometimes obtained. This causes the results not to be as accurate as desired because of the interpolative reasoning developed by FRBSs.

In order to face this problem, we propose a new *ad hoc* data-driven methodology to improve the rule cooperation and thus the accuracy of the obtained models, the cooperative rules (COR) methodology. Once the rule antecedents (defining the fuzzy subspaces) have been obtained, the operation mode will be composed of two stages: generation of a candidate consequent set for each subspace; and search of the consequents with the best global performance.

The paper is organized as follows. Section II introduces *ad hoc* data-driven methods and analyzes the two existing generation approaches, Section III presents the proposed methodology, Section IV analyzes the behavior of our proposal and other *ad hoc* data-driven methods in three different applications, and finally, Section V outlines some concluding remarks.

Moreover, three appendixes have been included. Appendix I is devoted to introduce the particular aspects and parameter values considered in the search technique used in COR, Appendix II shows the composition of a learning method used in the experimental study, and Appendix III collects the notation considered in the paper.

Manuscript received February 1, 2000; revised December 3, 2001. This work was supported by CICYT under Project PB98-1319. This paper was recommended by Associate Editor K. R. Pattipati.

The authors are with the Department of Computer Science and Artificial Intelligence, E.T.S.I. Informática, University of Granada, Granada E-18071, Spain (e-mail: casillas@decsai.ugr.es; ocordon@decsai.ugr.es; herrera@decsai.ugr.es).

Publisher Item Identifier S 1083-4419(02)04352-2.

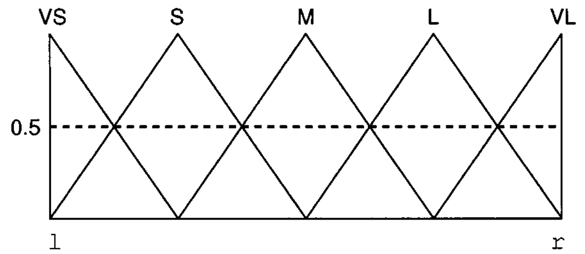


Fig. 1. Graphical representation of a fuzzy partition, standing S for *small*, M for *medium*, L for *large*, V for *very*, and with $[l, r]$ being the corresponding variable domain.

II. Ad Hoc Data-Driven Linguistic Rule Learning Methods

Ad hoc data-driven linguistic rule learning methods [8]–[11] are characterized by four main features.

1) *They are based on working with an input–output data set*—$E = \{e_1, \ldots, e_l, \ldots, e_N\}$, with $e_l = (x_1^l, \ldots, x_n^l, y_1^l, \ldots, y_m^l)$, N being the data set size, n being the number of input variables, and m being the number of output variables—*representing the behavior of the problem being solved*. In this paper, we will work with multiple-input single-output (MISO) systems, i.e., $m = 1$, $e_l = (x_1^l, \ldots, x_n^l, y^l)$.
2) *They consider a previous definition of the DB composed of the input and output primary fuzzy partitions.* It may be obtained from expert information (if it is available) or by a normalization process. We will consider symmetrical fuzzy partitions of triangular membership functions crossing at height 0.5 (as shown in Fig. 1).

   The set of linguistic terms of the i-th input variable will be denoted by $\mathcal{A}_i$—with $i \in \{1, \ldots, n\}$—and the set of linguistic terms of the output variable will be denoted by $\mathcal{B}$, with $|\mathcal{A}_i|$ ($|\mathcal{B}|$) being the number of linguistic terms of the i-th input (output) variable.
3) *The generation of the linguistic rules is guided by covering criteria of the data in the example set* (hence the name *data-driven*).
4) *The learning mechanism is not based on any well-known optimization or search technique but it is specifically developed for this purpose* (hence the name *ad hoc*).

We can distinguish between two different approaches to obtain the linguistic rules with *ad hoc* data-driven methods: guided by *examples* and guided by *fuzzy grid*. Both families of methods will be analyzed, showing a specific method for each, in Sections II-A and B. Then, some relations between them and a generic scheme of *ad hoc* data-driven methods will be respectively presented in Sections II-C and D. This preliminary analysis will help us to introduce our COR methodology in Section III.

A. Ad Hoc Data-Driven Linguistic Rule Learning Methods Guided by Examples

The linguistic rule generation process guided by examples obtains each rule from a specific example in the data set. In this way, the rule R^l is obtained from the example e_l. Actually, these rules belong to a candidate rule set, since after this generation stage a selection process is performed to derive the final RB. Fig. 2 graphically illustrates this rule generation process.

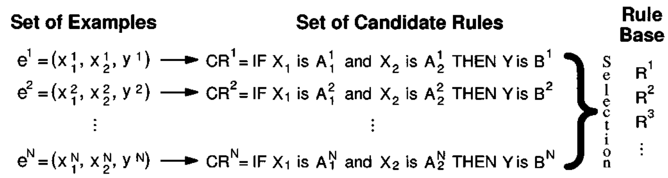


Fig. 2. Rule generation process followed by the methods guided by examples.

One of the most well known and widely used example-based methods is the Wang and Mendel's method (**WM-method**) [11]. This method puts into effect the RB generation by means of the steps shown in Fig. 3.

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 contained 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, $(A_1^l, \ldots, A_n^l, B^l)$, with $A_i^l \in \mathcal{A}_i$ and $B^l \in \mathcal{B}$.
3. *Give an importance degree to each rule*—Let CR^l = IF X_1 is A_1^l and ... and X_n 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:
   $$CV_\Pi(CR^l, e_l) = \mu_{A_1^l}(x_1^l) \cdot \ldots \cdot \mu_{A_n^l}(x_n^l) \cdot \mu_{B^l}(y^l).$$
4. *Obtain a final 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.

Fig. 3. WM-method.

B. Ad Hoc Data-Driven Linguistic Rule Learning Methods Guided by Fuzzy Grid

Another possibility to generate the linguistic rules is to bracket the examples according to a fuzzy grid, and then to obtain a rule for each group (subspace) taking into account all of them. The fuzzy grid is obtained by the Cartesian product of the linguistic terms existing in the input primary fuzzy partition for each input variable, $\mathcal{A}_1 \times \cdots \times \mathcal{A}_n$.

Fig. 4 graphically shows this rule generation process, where $S_s = (A_1^s, \ldots, A_i^s, \ldots, A_n^s)$—with $s \in \{1, \ldots, N_s\}$, $N_s = \prod_{i=1}^n |\mathcal{A}_i|$ being the number of multidimensional fuzzy input subspaces, and $A_i^s \in \mathcal{A}_i$—denotes a particular fuzzy input subspace and R^s is the corresponding linguistic rule.

An example of this second approach can be found in [8], where Cordón and Herrera adapt the fuzzy-grid-based learning method of Ishibuchi *et al.* [9] for simplified Takagi-Sugeno-type rules [14] (also known as singleton rules, i.e., rules with a single

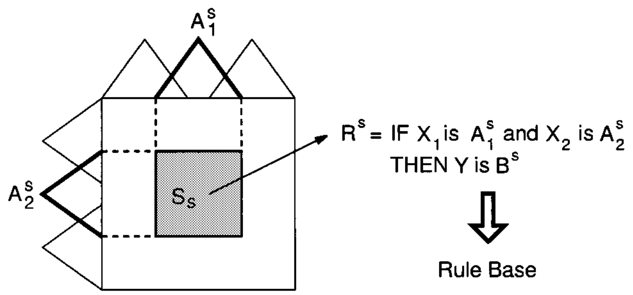


Fig. 4. Rule generation process followed by the methods guided by fuzzy grid.

1. *Consider a fuzzy partition of the variable spaces.*
2. *For each n-dimensional fuzzy input subspace, S_s, do*:
   2.1. Build the set E_s' composed of the input-output data pairs that are located in this subspace, i.e., $E_s' = \{e_{l^s} = (x_1^{l^s}, \ldots, x_n^{l^s}, y^{l^s}) \in E \text{ such that } \mu_{A_1^s}(x_1^{l^s}) \cdot \ldots \cdot \mu_{A_n^s}(x_n^{l^s}) \neq 0\}$.
   2.2. If $|E_s'| \neq 0$, i.e., if there is any data in this space zone, then:
      2.2.1. Let $B^s = \{B_k \in \mathcal{B},\ k \in \{1, \ldots, |\mathcal{B}|\} \text{ such that } \exists e_{l^s} \in E_s' \text{ with } \mu_{B_k}(y^{l^s}) \neq 0\}$ be the set of linguistic labels in the output variable term set which contain examples belonging to E_s', and let $c_s = |B^s|$ be the cardinality of B^s.
      2.2.2. For each linguistic label $B_{k^s}^s \in B^s$—with $k^s \in \{1, \ldots, c_s\}$—compute the covering value, CV_T, of the linguistic rule built using this term as a value in the consequent, $R_{k^s}^s$ = IF X_1 is A_1^s and ... and X_n is A_n^s THEN Y is $B_{k^s}^s$, over each example $e_{l^s} \in E_s'$ as follows:
      $$CV_T(R_{k^s}^s, e_{l^s}) = T(\mu_{A_1^s}(x_1^{l^s}), \ldots, \mu_{A_n^s}(x_n^{l^s}), \mu_{B_{k^s}^s}(y^{l^s})),$$
      with T being a t-norm. In this paper, we will work with the *Minimum*.
      2.2.3. Add the rule $R_{k^s}^s$ that presents the highest value in the *rule valuation function*, $RVF(\cdot)$, to the RB.

   Otherwise, let $B^s = \emptyset$ and do not generate any rule in that multidimensional fuzzy input subspace.

Fig. 5. CH-method.

point instead a fuzzy set in the consequent) to allow it to generate linguistic rules (**CH-method**). The method is described in Fig. 5.

Many different choices may be considered for the rule valuation function, $\mathrm{RVF}(\cdot)$, used in the algorithm. In this paper, we will work with the three following ones.

1) *Covering of the example best covered*. The absolute covering degree of $R_{k^s}^s$ over the best covered example in E_s' is calculated:
$$\mathrm{RVF}_1(R_{k^s}^s) = \max_{l^s=1}^{|E_s'|} \{\mathrm{CV}_T(R_{k^s}^s, e_{l^s})\}.$$

2) *Mean covering over the example set*. The mean covering degree of $R_{k^s}^s$ over all the examples in E_s' is calculated:
$$\mathrm{RVF}_2(R_{k^s}^s) = \frac{\sum_{l^s=1}^{|E_s'|} \mathrm{CV}_T(R_{k^s}^s, e_{l^s})}{|E_s'|}.$$

3) *Combination of both criteria, best and mean covering*:
$$\begin{aligned}\mathrm{RVF}_3(R_{k^s}^s) &= \mathrm{RVF}_1(R_{k^s}^s) \cdot \mathrm{RVF}_2(R_{k^s}^s) \\ &= \max_{l^s=1}^{|E_s'|} \{\mathrm{CV}_T(R_{k^s}^s, e_{l^s})\} \\ &\quad \cdot \frac{\sum_{l^s=1}^{|E_s'|} \mathrm{CV}_T(R_{k^s}^s, e_{l^s})}{|E_s'|}.\end{aligned}$$

C. Relation Between Example-Based and Fuzzy-Grid-Based Methods

While in an example-based method each example generates a single rule, the generation process performed by a fuzzy-grid-based method involves that an example may contribute to the generation of several rules. Indeed, we may draw an analogy between both approaches considering the use of different kinds of grids (Fig. 6 graphically shows this fact).

1) *Example-based method*—Since the antecedent of the rule generated from each example is obtained by taking the linguistic labels best covering the input values of this example, it can be seen that an example-based method partitions the example set according to a *crisp grid* bounded by the cross points between labels. Thus, an example may only belong to one subspace.
2) *Fuzzy-grid-based method*—However, with the fuzzy grid used by this latter approach, an example may belong to more than one subspace. In Fig. 6(b), the examples lying in white zones have an influence on the generation of one rule, those lying in light grey zones influence two rules, and the ones lying in dark grey zones influence four rules.

It is not possible to determine what approach is the best. Notice that the example approach always obtains an equal number or fewer rules than the fuzzy grid one, as in the fuzzy grid approach the examples have an influence on a wider region, thus generating more rules. However, this fact may make the model obtained by an example-based method not to be as accurate as desired sometimes.

D. Generic Scheme of Ad Hoc Data-Driven Linguistic Rule Learning Methods

According to the relations between example-based and fuzzy-grid-based methods found in Section II-C, it is possible to establish a generic scheme of *ad hoc* data-driven linguistic rule learning methods as seen in Fig. 7.

From this new point of view, it is intuitive to think about possible combinations between both methods using the candidate rule generation process (step 2) of one of them and the rule selection (step 3) of the other. For example, it would be possible to use the rule valuation functions, $\mathrm{RVF}(\cdot)$, considered by the CH-method to give an importance degree to each rule in each of

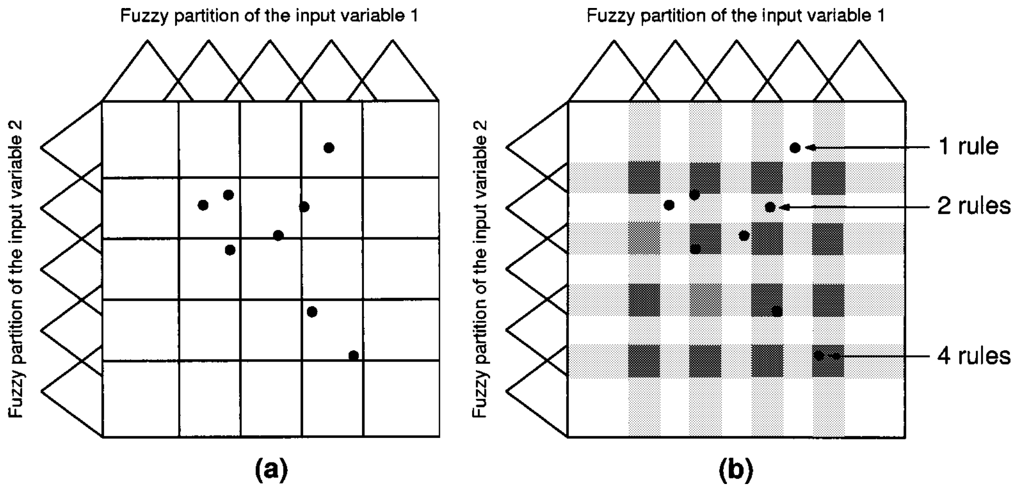


Fig. 6. (a) Example-based methods (crisp grid)—an example only contributes to the generation of one rule and (b) fuzzy-grid-based methods—an example may contribute to the generation of several rules.

1. *Consider a fuzzy partition of the variable spaces.*
2. *Generate candidate rules in each fuzzy input subspace according to the generation approach followed (example or fuzzy grid).*

   After this stage, a set of linguistic labels,

   $$B^s = \{B_1^s, \ldots, B_{k^s}^s, \ldots, B_{c_s}^s\},$$

   as well as rules with the different possible consequents,

   $$R_{k^s}^s = \text{ IF } X_1 \text{ is } A_1^s \text{ and ... and } X_n \text{ is } A_n^s \text{ THEN } Y \text{ is } B_{k^s}^s,$$

   will be available in each subspace S_s.
3. *Select the rule with the best consequent in each fuzzy input subspace with respect to some covering criterion.*

Fig. 7. Generic scheme of *ad hoc* data-driven methods.

the groups obtained by the WM-method, and to select the rule with the best value. Another alternative would be to use the importance degree considered in the WM-method to select the best rule in each subspace defined by the CH-method.

Actually, some of these combinations are implicitly defined since $\mathrm{RVF}_1(\cdot)$ coincides with the criterion used in the WM-method to select the rule. Nevertheless, two new algorithms arise from the use of $\mathrm{RVF}_2(\cdot)$ and $\mathrm{RVF}_3(\cdot)$ to determine the best rule in each of the groups obtained by the WM-method, in the following called $\mathbf{WM + RVF_2}$**-method** and $\mathbf{WM + RVF_3}$**-method**, respectively.

III. COR: A Methodology to Improve the Accuracy by Obtaining Cooperative Rules

A. Cooperative Rules Methodology

One of the most interesting features of an FRBS is the interpolative reasoning it develops. This characteristic plays a key role in the high performance of FRBSs and is a consequence of the cooperation among the linguistic rules composing the KB. It is a well-known fact that the output obtained from an FRBS is not usually due to a single linguistic rule but to the cooperative action of several linguistic rules that have been fired because they match the system input to any degree.

However, the global interaction among the rules of the KB is not considered in *ad hoc* data-driven methods (since they select the rule with the best performance in each subspace). This causes the final RB obtained, in spite of presenting a good local behavior, not to cooperate suitably. Moreover, the fact of locally

1. *Consider a fuzzy partition of the variable spaces.*
2. *Generate candidate rules in each subspace.*
3. *Select the most cooperative rule in each subspace.* This stage is performed in two steps:

   3.1. ***Obtain the set of possible consequents for each subspace containing examples.***
   To do that, let $C = \{r_j \in \{1, \ldots, N_s\}$ *such that* $B^{r_j} \neq \emptyset\}$ — with $N_C = |C|$ being the number of subspaces which contain examples and $j \in \{1, \ldots, N_C\}$ — be the set of identifications of those candidate consequent term sets which contain at least one element.

   3.2. ***Run the SA-based algorithm to look for the combination*** $\{B_{k^{r_1}}^{r_1}, \ldots, B_{k^{r_j}}^{r_j}, \ldots, B_{k^{r_{N_C}}}^{r_{N_C}}\}$ ***with the best accuracy.***
   The initial solution is obtained by generating a possible combination at random.
   The neighbor generation mechanism randomly selects a specific $r_j \in C$ such that $c_{r_j} \geq 2$, and changes $B_{k^{r_j}}^{r_j}$ by $B_{k^{r_j\prime}}^{r_j}$, with $k^{r_j\prime}$ randomly generated in $\{1, \ldots, c_{r_j}\}$ such that $k^{r_j\prime} \neq k^{r_j}$.
   To evaluate the quality of each solution, an index measuring the cooperation degree of the encoded rule set is considered. In this case, the algorithm uses a global error function called *mean square error* (MSE), which is defined as

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

   with Y^l being the output obtained from the FRBS when the example e^l is used, and y^l being the known desired output. The closer to zero the measure, the greater the global performance and, thus, the better the rule cooperation.

Fig. 8. Learning method based on SA following the COR methodology.

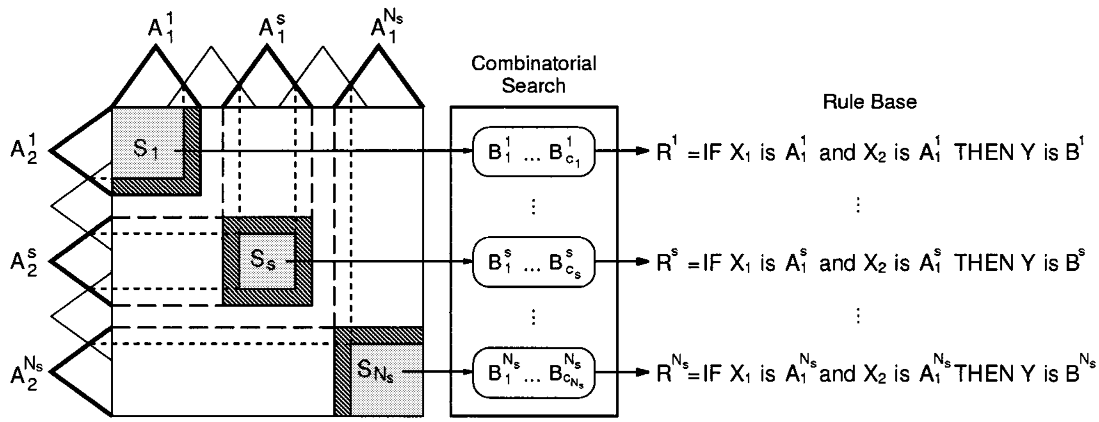


Fig. 9. Rule generation process followed by the COR methodology.

processing these rules makes these methods more sensitive to noise.

With the aim of addressing these drawbacks, we propose a new methodology to improve the accuracy of linguistic models obtaining best cooperation among the rules: the COR methodology. 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 KB 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.

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

1) The former accomplishes a full search through the set of possible combinations. Although with this technique we ensure 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 recommended 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.

   In that contribution we propose to use the simulated annealing (SA) technique [15] for this purpose. SA is a numerical optimization technique based on the analogy with the physical annealing process of solids. The SA-based algorithm begins with an initial solution and generates a neighbor of this solution by means of a suitable mechanism. If the latter is better than the former, the current solution is replaced by the generated neighbor; otherwise, this replacement is accomplished with a specific probability that will be decreased during the algorithm progress. This process is iterated a large number of times.

   Another approach to perform the COR methodology with the ant colony system algorithm is introduced in [16].

The proposed COR methodology only impinges on the third stage of the said *ad hoc* data-driven method generic scheme (Fig. 7). Fig. 8 presents a COR-based algorithm using the SA technique. The specific aspects of the SA-based algorithm used in this paper are shown in Appendix A. Following this methodology, two new algorithms arise combining COR with the two analyzed methods: the COR-based WM-method (**COR-WM-method**) and COR-based CH-method (**COR-CH-method**).

B. Analysis of the Cooperative Rules Methodology

Fig. 9 graphically shows the rule generation process based on the COR methodology using the COR-WM-method or the COR-CH-method to generate the candidate consequents for each subspace (second step in the algorithm). In this way, whilst the COR-WM-method considers the light grey zones, the COR-CH-method also takes into account the dark grey zones. This fact will have repercussions in the number of rules generated (as commented on in Section II-C) and in the sets of candidate consequents.

Since in the CH-method the examples contribute to generate several rules, usually more diversity of candidate consequents will be generated in each subspace and, therefore, the search space tackled by the combinatorial search will be larger. This fact has pros and cons.

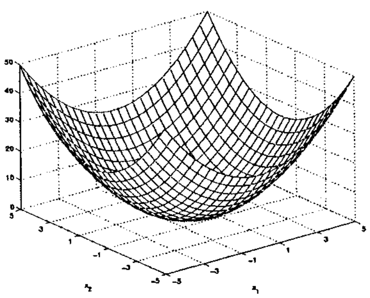
1) on the one hand, the algorithm handles more possible combinations, thus being able to find better results,
2) on the other hand, the larger the search space, the more difficult to find good solutions.

The suitability of one approach or the other will depend on the problem nature, represented by the example set (number of examples and their distribution), which will determine the diversity of the candidate consequent set in each subspace depending on the approach.

Finally, we should note that COR-based methods are halfway between the simplicity and speed of the classic *ad hoc* data-driven methods, and the complexity and slowness of the remaining rule generation mechanisms, being closer to the

TABLE I
METHODS CONSIDERED IN THIS EXPERIMENTAL STUDY

Method	Approach
WM-method	Guided by examples
WM+RVF_2-method	Guided by examples
WM+RVF_3-method	Guided by examples
CH-method (RVF_1)	Guided by fuzzy grid
CH-method (RVF_2)	Guided by fuzzy grid
CH-method (RVF_3)	Guided by fuzzy grid
COR-WM-method	Guided by examples
COR-CH-method	Guided by fuzzy grid
NIT-method	Guided by fuzzy grid with two rules per subspace



$$F(x_1, x_2) = x_1^2 + x_2^2$$

$$x_1, x_2 \in [-5, 5],$$
$$F(x_1, x_2) \in [0, 50]$$

Fig. 10. Graphical representation, mathematical expression, and variable domains of the three-dimensional surface F.

TABLE II
RESULTS OBTAINED IN THE MODELING OF F

Method	#R	MSE_{tra}	MSE_{tst}
WM-method	**49**	2.048137	2.287129
WM+RVF_2-method	**49**	2.740901	2.505504
WM+RVF_3-method	**49**	2.048137	2.287129
CH-method (RVF_1)	**49**	2.048137	2.287129
CH-method (RVF_2)	**49**	3.755082	3.393716
CH-method (RVF_3)	**49**	2.740901	2.505504
COR-WM-method	**49**	**1.605482**	**1.175941**
COR-CH-method	**49**	1.609891	1.302117
NIT-method	98	2.465487	1.751173

TABLE III
WM-METHOD VS. COR-WM-METHOD MODELING F

X_2 \ X_1	ES	VS	S	M	L	VL	EL
ES	{L,VL,EL}	{M,L,VL}	{S,M,L}	{S,M}	{S,M,L}	{M,L,VL}	{L,VL,EL}
VS	{M,L,VL}	{S,M,L}	{VS,S,M}	{VS,S}	{VS,S,M}	{S,M,L}	{M,L,VL}
S	{S,M,L}	{VS,S,M}	{ES,VS,S}	{ES,VS}	{ES,VS}	{VS,S,M}	{S,M,L}
M	{S,M}	{VS,S}	{ES,VS}	{ES}	{ES,VS}	{VS,S}	{S,M}
L	{S,M,L}	{VS,S,M}	{ES,VS}	{ES,VS}	{ES,VS}	{VS,S,M}	{S,M,L}
VL	{M,L,VL}	{S,M,L}	{VS,S,M}	{VS,S}	{VS,S,M}	{S,M,L}	{M,L,VL}
EL	{L,VL,EL}	{M,L,VL}	{S,M,L}	{S,M}	{S,M,L}	{M,L,VL}	{L,VL,EL}

(a) Candidate consequent sets in each subspace following the example approach

X_2 \ X_1	ES	VS	S	M	L	VL	EL
ES	EL	L	M	M	M	L	EL
VS	L	M	S	VS	S	M	L
S	M	S	VS	ES	VS	S	M
M	M	VS	ES	ES	ES	VS	M
L	M	S	VS	ES	VS	S	M
VL	L	M	S	VS	S	M	L
EL	EL	L	M	M	M	L	EL

(b) WM-method's RB
($MSE_{tra/tst} = 2.0481/2.2871$)

X_2 \ X_1	ES	VS	S	M	L	VL	EL
ES	EL	L	***L***	M	M	***VL***	EL
VS	L	M	***VS***	VS	S	***S***	L
S	***L***	***VS***	VS	ES	***ES***	S	M
M	M	VS	ES	ES	ES	VS	M
L	M	S	***ES***	ES	VS	***VS***	***L***
VL	L	***S***	S	VS	***VS***	M	L
EL	EL	***VL***	M	M	***L***	L	EL

(c) COR-WM-method's RB
($MSE_{tra/tst} = 1.6055/1.1760$)

former ones. Its performance will depend to a great extent on the combinatorial search technique considered and the search space tackled.

IV. EXPERIMENTAL STUDY

This experimental study will be devoted to analyze the performance of our two COR-based processes. With this aim, we have chosen three different applications:[1] the modeling of a three-dimensional surface [12], the problem of rice taste evaluation [9], [10], and an electrical distribution network real-world problem [17].

In the three cases, we will analyze the accuracy of the linguistic models generated from the processes introduced in Sections II and III as well as by the fuzzy grid-based method proposed by Nozaki *et al.* (**NIT-method**) in [10] (its description can be consulted in Appendix B). Table I shows the 9 methods and the approach followed by each of them.

Primary fuzzy partitions with triangular-shaped equally distributed fuzzy sets will be considered for each problem (as shown in Fig. 1). With respect to the FRBS 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 [18].

A. Modeling of the Three-Dimensional Surface F

1) Problem Description: The aim in this first problem will be to model the surface of the two-dimensional mathematical function F shown in Fig. 10. To do so, a training data set with 1,681 values uniformly distributed in the three-dimensional definition space has been obtained experimentally [12].

Another data set has been generated for its use as test set to evaluate the performance of the learning method, avoiding any possible bias related to the data in the training set. These data (420 examples, 20% of the total number of examples—training and test sets together) are obtained generating the input variable values at random in the concrete universes of discourse for each of them, and computing the associated output variable value.

Seven labels are used in each linguistic variable fuzzy partition for this experiment.

2) Obtained Results and Analysis: The results obtained by the nine methods analyzed are collected in Table II, where $\#R$ stands for the number of rules, and $\mathrm{MSE_{tra}}$ and $\mathrm{MSE_{tst}}$ for the error obtained over the training and test data sets, respectively. The best results are shown in boldface.

Analyzing these results, we may note that the COR-based methods significantly improve (from a 22% better in the least favorable case to a 113% in the most) the accuracy of the models compared with non cooperative methods like the WM and CH ones. In Table III we show how, from the candidate consequent set generated in each subspace following the example approach [Table III(a)], the WM-method performs the selection considering the best rule in each subspace [Table III(b)], whereas the

[1]The training and test data partitions used for the three problems are available at http://decsai.ugr.es/~casillas/FMLib/

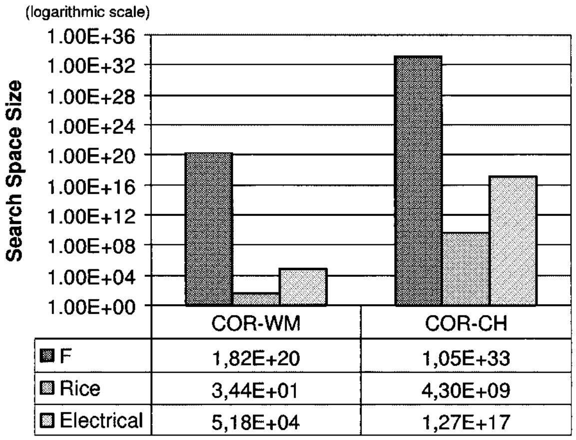


Fig. 11. Search space tackled by the two COR-based algorithms for each problem.

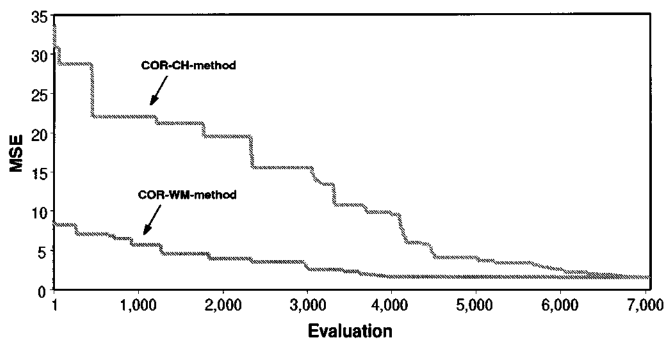


Fig. 12. Evolution chart of the COR-WM-method and COR-CH-method in the F problem.

COR-WM-method makes use of a cooperative criterion that takes into account the global behavior of the rules [Table III(c)]. Hence, the latter method selects a different consequent from the locally best one in the cells shown in boldface and italics, making the accuracy of the obtained model notably better (22% and 49% for training and test, respectively).

The COR-WM-method obtains a more accurate model than the COR-CH-method. The slight difference between both algorithms may depend on the size of the search space tackled by the corresponding SA processes (Fig. 11 graphically shows these sizes for the problems considered), larger in the COR-CH-method.

This fact is clearly shown when analyzing the evolution charts of both methods (Fig. 12). Indeed, we may observe that the COR-CH-method only obtains a solution as good as the one generated by the COR-WM-method at the end of the algorithm run.

The main difference between the COR-CH-method and the fuzzy grid-based NIT-method is that the latter makes use of the two best consequents in each subspace giving a certainty factor to each while the former analyzes what rule provides the higher global accuracy and only makes use of such a rule. Thus, the model obtained by the COR-CH-method is more accurate and presents a significantly simpler KB (49 rules against 98, a 50% lesser) than the NIT-method one, which is a very important aspect in linguistic modeling where the interpretability is the main requirement. Therefore, in this problem, using the most globally cooperative rule in each subspace seems to be more suitable than taking the two best local rules.

B. The Rice Taste Evaluation Problem

1) Problem Description: 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 persons, evaluate the rice according to a set of characteristics associated with it. These factors are *flavor*, *appearance*, *taste*, *stickiness*, and *toughness* [9], [10].

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. Moreover, the problem-solving goal is not only to obtain an accurate model, but to obtain a user-interpretable model as well, capable of putting some light on the reasoning process performed by the expert for evaluating a kind of rice in a specific way. Due to all these reasons, in this section we try to obtain a linguistic model to solve the said problem.

To do so, we use the data set presented in [10]. 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 $[0, 1]$.

Because of the small number of examples used, there is a high risk of biasing the learning process. With the aim of making a fair comparison of the 9 fuzzy rule learning algorithms analyzed, 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, thus generating 10 different linguistic models in each experiment. This is the same experimental procedure developed by the authors in the paper where the example data set is presented [10].

Two labels are considered to partition each linguistic variable domain.

2) Obtained Results and Analysis: Table IV shows the arithmetic mean and standard deviation values of the 10 linguistic models generated by each method in this application. Due to the small search space tackled by the COR-WM-method in the rice problem (see Fig. 11), an explicit enumeration instead the simulated annealing is used in this case.

In view of the obtained results, both COR-WM-method and COR-CH-method clearly outperform the corresponding WM and CH methods. Significantly more accurate models (from a 28% better to a 65%) are obtained thanks to the cooperative rule consideration, of course maintaining the same number of rules. Moreover, the low standard deviation values show the robustness of the COR-based algorithms.

Table V shows the candidate consequent sets generated by the example approach and the RBs obtained by the WM-method and COR-WM-method for a data set partition. The fact of using only two consequents different from those generated by the WM-method (shown in boldface and italics) makes the

TABLE IV
RESULTS OBTAINED IN THE RICE TASTE EVALUATION PROBLEM

Method	#R $\bar{x}$	#R σ	MSE$_{tra}$ $\bar{x}$	MSE$_{tra}$ σ	MSE$_{tst}$ $\bar{x}$	MSE$_{tst}$ σ
WM-method	**15**	1	0.013284	0.005987	0.013119	0.004239
WM+RVF_2-method	**15**	1	0.014841	0.006291	0.015127	0.005754
WM+RVF_3-method	**15**	1	0.017049	0.006589	0.017320	0.005377
CH-method (RVF_1)	32	**0**	0.020296	0.006479	0.021184	0.006010
CH-method (RVF_2)	32	**0**	0.011567	0.002698	0.011927	0.003899
CH-method (RVF_3)	32	**0**	0.009803	0.002142	0.010834	0.002209
COR-WM-method	**15**	1	0.007979	0.001887	0.008244	**0.001084**
COR-CH-method	32	**0**	**0.007076**	0.000571	**0.008012**	0.001766
NIT-method	64	**0**	0.008626	**0.000345**	0.009851	0.001931

$\bar{x}$ = arithmetic mean, σ = standard deviation

TABLE V
WM-METHOD ($\mathrm{MSE}_{\mathrm{tra/tst}} = 0.014\,704/0.016\,700$) VS.
COR-WM-METHOD ($\mathrm{MSE}_{\mathrm{tra/tst}} = 0.008\,232/0.007\,546$) IN THE FIRST DATA SET PARTITION OF THE RICE PROBLEM

Rule	*Flav.*	*App.*	*Taste*	*Stickiness*	*Tough.*	*Evaluation*: Candidate consequents	WM	COR-WM
R_1	bad	bad	bad	not-sticky	tender	{low}	low	low
R_2	bad	bad	bad	not-sticky	tough	{low,high}	low	low
R_3	bad	good	bad	not-sticky	tender	{low}	low	low
R_4	bad	good	good	not-sticky	tender	{low,high}	low	***high***
R_5	bad	good	good	sticky	tender	{low}	low	low
R_6	good	bad	bad	not-sticky	tender	{low}	low	low
R_7	good	bad	bad	not-sticky	tough	{low}	low	low
R_8	good	bad	good	not-sticky	tender	{low}	low	low
R_9	good	good	bad	not-sticky	tender	{low,high}	high	high
R_{10}	good	good	bad	not-sticky	tough	{low,high}	low	***high***
R_{11}	good	good	bad	sticky	tender	{high}	high	high
R_{12}	good	good	good	not-sticky	tender	{high}	high	high
R_{13}	good	good	good	not-sticky	tough	{low,high}	high	high
R_{14}	good	good	good	sticky	tender	{low,high}	high	high
R_{15}	good	good	good	sticky	tough	{high}	high	high

Flav. = Flavor, ***App.*** = Appearance, ***Tough.*** = Toughness

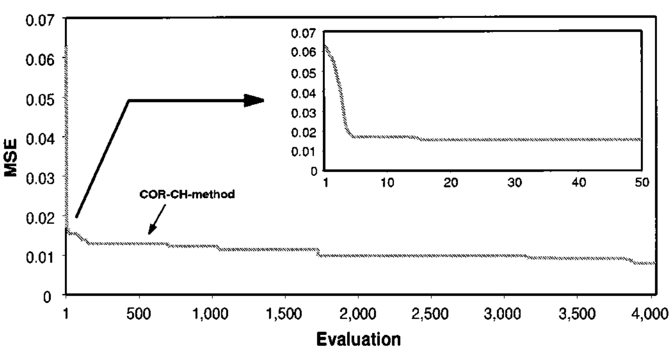


Fig. 13. Evolution chart of the COR-CH-method in the first data set partition of the rice problem.

COR-WM-method's model significantly more accurate (44% in training and 55% in test).

Opposite to the NIT-method, the COR-based methods obtain a more accurate model with a much lesser number of rules, which significantly improves the interpretability. The COR-CH-method obtains the most accurate linguistic model among all the analyzed models. Moreover, it shows a strong convergence as can be seen in Fig. 13. We can conclude again that a cooperative set of simple rules performs better than a non cooperative set of double-consequent rules.

TABLE VI
RESULTS OBTAINED IN THE ELECTRICAL APPLICATION

Method	#R $\bar{x}$	#R σ	MSE$_{tra}$ $\bar{x}$	MSE$_{tra}$ σ	MSE$_{tst}$ $\bar{x}$	MSE$_{tst}$ σ
WM-method	**22**	**1**	211,733	8,069	236,770	**12,647**
WM+RVF_2-method	**22**	**1**	207,302	8,564	239,114	35,364
WM+RVF_3-method	**22**	**1**	206,437	8,247	236,707	31,829
CH-method (RVF_1)	30	2	228,577	26,035	267,323	35,535
CH-method (RVF_2)	30	2	268,781	12,284	299,444	33,729
CH-method (RVF_3)	30	2	270,264	8,232	289,134	34,935
COR-WM-method	**22**	**1**	180,995	7,794	220,320	32,492
COR-CH-method	30	2	**171,659**	2,997	**203,050**	16,890
NIT-method	61	4	182,297	**2,764**	219,283	45,400

$\bar{x}$ = arithmetic mean, σ = standard deviation

C. The Electrical Distribution Network Problem

1) Problem Description: Sometimes, there is a need to measure the amount of electric lines that an electric company owns. This measurement may be useful for several aspects such as estimation of the maintenance costs of the network, which was the main goal of the problem presented here in Spain [17]. Low-voltage lines are contained in villages and it is very expensive to measure their length. Therefore, an indirect method to do so is needed.

The problem involves finding a model that relates the total *length of low-voltage line* installed in a rural town with the *number of inhabitants* in the town and the mean of the distances from the center of the town to the three furthest clients in it (*radius of village*) [17]. This model will be used to estimate the total length of line being maintained. Moreover, it would be preferable that the obtained solutions are not only numerically accurate in the problem-solving, but interpretable by human beings to some degree.

The data set is composed of 495 pieces of real data obtained from direct measures in this number of villages [17]. We have performed a five-fold cross-validation for this experiment.

Seven labels are regarded for each linguistic variable.

2) Obtained Results and Analysis: The arithmetic mean and standard deviation values of the 5 models generated by each method are collected in Table VI.

From an analysis of these results, we may again note the good behavior presented by the proposed COR-based methods. Their linguistic models are again more accurate than the corresponding WM and CH methods, with the model obtained by the COR-CH-method being the most accurate among all the an-

TABLE VII
CH-METHOD VS. COR-CH-METHOD IN A SPECIFIC DATA SET OF THE ELECTRICAL PROBLEM

X_2 \ X_1	ES	VS	S	M	L	VL	EL
ES	{ES,VS}	{ES,VS,S}	{VS,ES,S}	{VS,S}			
VS	{VS,ES, S,M}	{VS,S, ES,M}	{VS,S, M,ES}	{VS,S, M,ES}	{S,M}		
S	{VS,S,ES, M,L,VL}	{VS,S,M, L,ES,VL}	{M,L,VS, S,VL,ES}	{S,L,VL, M,ES,VS}	{M,S}		
M	{S,VS,M, L,ES,VL}	{L,M,S, VS,VL,ES}	{VL,S,VS, L,M,EL}	{VL,M, EL,L}	{VL,EL}		{S,M}
L	{VS,S,M, L,VL}	{M,VS,S, L,VL}	{S,M,VL, L,EL}	{EL,M, VL,S}	{VL,EL}		{S,M}
VL	{VS,S, M,L}	{L,S, VS,M}	{L,M,S}	{M,S}			
EL	{M,S}	{M,S}					

(a) Candidate consequent sets in each subspace following the fuzzy grid approach

X_2 \ X_1	ES	VS	S	M	L	VL	EL
ES	ES	ES	VS	VS			
VS	VS	VS	VS	VS	S		
S	VS	VS	M	S	M		
M	S	L	VL	VL	VL		S
L	VS	M	S	ES	VL		S
VL	VS	L	L	M			
EL	M	M					

(b) CH-method's RB ($\text{MSE}_{tra/tst}$ = 228,577/267,323)

X_2 \ X_1	ES	VS	S	M	L	VL	EL
ES	ES	ES	VS	VS			
VS	VS	VS	VS	***S***	***M***		
S	VS	***S***	***S***	***M***	***S***		
M	***VS***	***M***	***M***	***ES***	VL		S
L	***S***	M	***M***	ES	VL		***M***
VL	VS	L	***M***	***S***			
EL	M	M					

(c) COR-CH-method's RB ($\text{MSE}_{tra/tst}$ = 171,659/203,050)

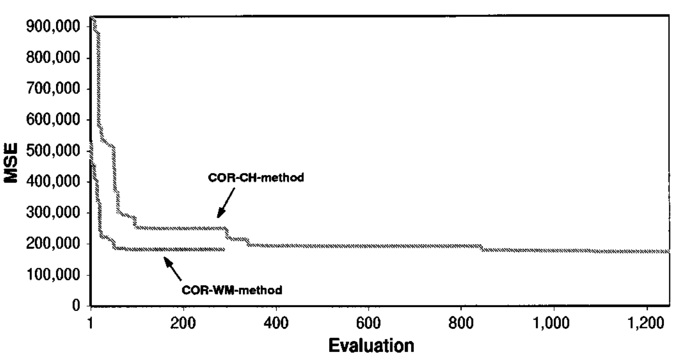


Fig. 14. Evolution chart of the COR-WM-method and COR-CH-method for a specific data set partition in the electrical problem.

alyzed models. Table VII shows the candidate consequent sets (according to the fuzzy grid approach) and the RBs generated by this method and the CH-method (with RVF_1) for a specific data set partition. In that table, X_1 stands for the population size and X_2 for the radius of village.

The differences of accuracy found between the COR-WM-method and COR-CH-method seem to be related to the sizes of the search spaces tackled, which allow the COR-CH-method to accomplish a better search thus obtaining a better performing model. This fact is clearly illustrated in Fig. 14, where the evolution charts of both processes are shown. It is also interesting to note that the consideration of cooperation among the linguistic rules lessens the noise effect observed in noncooperative methods, especially in the CH ones.

The COR-CH-based method again obtains a model with better performance than the NIT-method. Moreover, our methods obtain lesser standard deviation values in the generalization degree that again show their robustness.

TABLE VIII
VALUES OF PARAMETERS USED IN THE TWO SA-BASED METHODS

	COR-WM			COR-CH		
	T_0	N_{max}	A_{max}	T_0	N_{max}	A_{max}
F	40	98	49	40	98	98
Rice	—	—	—	70	32	32
Electrical	500	24	24	500	32	32

V. CONCLUDING REMARKS

This paper has performed a double task:

1) analyzing the features of *ad hoc* data-driven RB learning methods by giving a taxonomy and a generic integration scheme;
2) introducing the COR methodology.

The behavior of different methods, including two proposals based on our methodology, has been analyzed when solving three specific modeling problems. In view of the obtained results, we may point out some interesting conclusions.

1) Among non cooperative methods, both approaches obtain models with similar accuracy although the example-based ones usually generate a lesser number of rules. However, sometimes these methods are not as accurate as desired due to the reduced number of rules. On the other hand, example-based methods seem to be more suitable than the fuzzy grid-based ones for real-world problems affected by noise.
2) As regards our cooperative proposals, they have obtained very good results combining accuracy and interpretability. This leads us to conclude that the consideration of cooperative rules improves the performance of the linguistic models and the derivation of RBs by firstly generating a candidate rule set and then searching the best combination of rules is a good way to accomplish this aspect. Moreover, the use of cooperative rules seems to be very robust against noise.

APPENDIX I
SA-BASED ALGORITHMS CONSIDERED IN THE EXPERIMENTAL STUDY

The SA-based algorithms used in the COR-based methods take the following aspects into account.

1) The cooling scheme used will be the exponential one proposed by Kirkpatrick [19] ($T_{k+1} = T_k \cdot C$, with $C = 0.9$).
2) The equilibrium at a specific temperature will be achieved when a *maximum number of neighbors* ($N_{\max}$) has been generated or when a *maximum number of acceptances* ($A_{\max}$) has been attained.
3) Finally, as regards the stopping criterion, the algorithm will stop when no neighbor is accepted for a specific temperature.

Table VIII shows the values of parameters considered in the two SA-based methods for each application, where T_0 stands for the *initial temperature*.

We may outline some general criteria about the values used for each parameter.

1) A logical stance to establish the value of $N_{\max}$ is to make it a multiple of the number of rules considered, i.e., the number of subspaces with examples according to the learning approach followed (N_C). Thus, the values associated with the different subspaces will have, on average, the same probability of being changed. The value of $A_{\max}$ is usually a percentage (50%–100%) of $N_{\max}$.
2) Higher values for $N_{\max}$ and $A_{\max}$ are usually recommended when the algorithm must tackle a large search space.
3) Higher values for T_0 are usually recommended when the problem has a high risk of falling in local optima.

In the following subsections, detailed descriptions of the parameter value decisions taken for each problem are shown.

A. Surface Modeling of the Two-Dimensional Function F

As the search space tackled by COR-based methods for the F problem is large (see Fig. 11), the parameters of the two SA processes have been set with the aim of exploring a great number of possibilities at each temperature. This way, $N_{\max}$ was $2 \cdot N_C$ in both cases. As for $A_{\max}$, the values were N_C and $2 \cdot N_C$ for the COR-WM-method and COR-CH-method respectively, since the former tackles a smaller search space.

The initial temperature was 40 in both cases. Nevertheless, similar accuracy results were obtained with a wide range of values. This fact shows the robustness of the SA-based algorithms for the T_0 parameter in this problem.

The COR-WM-method and COR-CH-method respectively needed 6,084 and 6,924 evaluations to find their solutions. Considering the huge search space tackled, these values are significantly small. Experiments with a genetic algorithm tackling the same search space obtained similar accuracy results with a number of evaluations in the order of 20 000.

B. The Rice Taste Evaluation Problem

For the rice problem, we should say that in the COR-WM-method case, due to the small size of the example set and to the small number of linguistic terms in the consequent linguistic variable, a reduced number of combinations (see Fig. 11) is generated in the first step of the combinatorial search. Therefore, the explicit enumeration has been used in this method instead of the SA procedure since the best solution can be quickly found.

For the COR-CH-method, which works with a medium search space size (see Fig. 11), $N_{\max}$ and $A_{\max}$ were set to N_C. The initial temperature was 70. This decision was taken because the problem has a slight risk of falling in local optima and the SA-based algorithms presented certain sensibility to this parameter. As for the number of evaluations needed to obtain the solution, the average value for the 10 models was 3,963.

C. The Electrical Distribution Network Problem

Due to the nature of the electrical problem, there is a high risk of coming across local optima. This fact makes the SA-based algorithms very sensitive to the initial temperature value—particularly the COR-CH-method because of its larger search space size—obtaining significantly different accuracy results depending on it. To face this difficulty performing a proper exploration, a high enough initial temperature should be considered. After some previous experiments, this value was set to 500 in both SA-based methods.

$N_{\max}$ and $A_{\max}$ values were set to N_C in both algorithms. The number of evaluations needed to obtain the solution was 69 and 1,129 for the COR-WM-method and COR-CH-method, respectively.

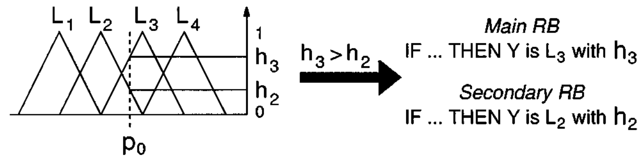


Fig. 15. Obtaining linguistic rules from singleton ones in the NIT-method.

APPENDIX II
NOZAKI, ISHIBUCHI, AND TANAKA'S LEARNING METHOD

The NIT-method—proposed by Nozaki, Ishibuchi, and Tanaka in [10]—consists of two phases.

1) *First phase, generation of singleton rules*—In a first step, a KB with singleton consequents (i.e., containing real values instead linguistic terms in the output variable) is learnt. To determine the real consequent of each rule, a weight is defined for each training data array as the result of raising the membership degree of the input to the power α (a parameter that defines a nonlinear scaling function) and of getting the product of the membership function values of each input. This additional parameter will not be considered in our experiments (we will set $\alpha = 1$) since this paper aims to analyze the rule generation process without modifying the DB definition.

   The real-valued consequent will be obtained as the weighted average of the known output value associated with each array of input values. If the weight (matching degree) is zero, the rule will not be considered (therefore, the NIT-method follows the fuzzy grid approach).
2) *Second phase, transformation into linguistic rules*—In this step, singleton representation is translated into the linguistic one keeping a similar accuracy. It is achieved thanks to the use of two linguistic RBs: the main and secondary ones. In the first place, it will be necessary to accomplish a fuzzy partition of the universe of discourse of the consequent in some linguistic labels. The authors consider uniformly distributed partitions with triangular membership functions (as shown in Fig. 1).

   Then, the membership degree of the singleton consequent (p_0) to each fuzzy set in the output variable fuzzy partition is computed. The linguistic label of the fuzzy set with the largest membership degree (L_3 in the example of Fig. 15) is taken to compose the fuzzy rule in the main RB, whereas the linguistic label with the next higher member-

TABLE IX
NOTATION FOLLOWED IN THE PAPER

Notation	Meaning
n	Number of input variables
i	Input variable counter, $(1 \leq i \leq n)$
$\mathcal{A}_i$	Set of linguistic terms of the i-th input variable
$\mathcal{B}$	Set of linguistic terms of the output variable
k	Term counter of the output variable, $(1 \leq k \leq \vert\mathcal{B}\vert)$
E	Set of input-output data pairs (examples)
N	Number of examples, $N = \vert E\vert$
l	Example counter, $(1 \leq l \leq N)$
x_i^l	Value of the i-th input variable of the l-th example
y^l	Value of the output variable of the l-th example
e^l	An $(n+1)$-dimensional vector containing the l-th example, $e^l = (x_1^l, \ldots, x_n^l, y^l)$
A_i^l	The fuzzy set best covering x_i^l, $A_i^l \in \mathcal{A}_i$
B^l	The fuzzy set best covering y^l, $B^l \in \mathcal{B}$
CR^l	Candidate rule obtained from the l-th example, $CR^l = \textit{IF } X_1 \textit{ is } A_1^l \textit{ and} \ldots \textit{and } X_n \textit{ is } A_n^l \textit{ THEN } Y \textit{ is } B^l$
N_s	Number of multidimensional fuzzy input subspaces, $N_s = \prod_{i=1}^{n} \vert\mathcal{A}_i\vert$
s	Multidimensional subspace counter, $(1 \leq s \leq N_s)$
A_i^s	Linguistic term used in the subspace S_s for the i-th input variable, $A_i^s \in \mathcal{A}_i$
S_s	s-th n-dimensional fuzzy input subspace, $S_s = (A_1^s, \ldots, A_i^s, \ldots, A_n^s)$
E_s'	Set of input-output data pairs (examples) that are located in the subspace S_s, $E_s' = \{e_{l^s} = (x_1^{l^s}, \ldots, x_n^{l^s}, y^{l^s}) \in E \mid \mu_{A_1^s}(x_1^{l^s}) \cdot \ldots \cdot \mu_{A_n^s}(x_n^{l^s}) \neq 0\}$
e_{l^s}	l^s-th example located in the subspace S_s, $e_{l^s} \in E_s'$
B^s	Set of linguistic labels in the output variable term set which contain examples belonging to E_s', $B^s = \{B_k \in \mathcal{B} \textit{ such that } \exists e_{l^s} \in E_s' \textit{ with } \mu_{B_k}(y^{l^s}) \neq 0\}$
c_s	Number of candidate consequent fuzzy sets in the subspace S_s, $c_s = \vert B^s\vert$
k^s	Candidate consequent fuzzy set counter in the subspace S_s, $(1 \leq k^s \leq c_s)$
$B_{k^s}^s$	k^s-th candidate consequent fuzzy set in the subspace S_s, $B_{k^s}^s \in B^s$
$R_{k^s}^s$	Rule in the subspace S_s with the consequent $B_{k^s}^s$
N_C	Number of multidimensional fuzzy input subspaces which contain examples, $N_C = \vert\{E_s' \textit{ s.t. } E_s' \neq \emptyset\}\vert$
j	Counter of candidate consequent sets which contain at least one example, $(1 \leq j \leq N_C)$
r_j	Identification of the j-th candidate consequent set that contain at least one example, i.e., $r_j \in \{1, \ldots, N_s\}$ such that $B^{r_j} \neq \emptyset$
$\vert S\vert$	Cardinality of the set S
$\mu_A(x)$	Membership degree of the value x over the linguistic label (associated fuzzy set) A
$CV_T(R,e)$	Covering value of the rule $R = \textit{IF } X_1 \textit{ is } A_1 \textit{ and} \ldots \textit{and } X_n \textit{ is } A_n \textit{ THEN } Y \textit{ is } B$ over the example $e = (x_1, \ldots, x_n, y)$ using the t-norm T, $CV_T(R,e) = T(\mu_{A_1}(x_1), \ldots, \mu_{A_n}(x_n), \mu_B(y))$

ship value (L_2 in Fig. 15) is considered for the secondary RB.

Besides, the said membership degrees are considered as rule weights (h_3 and h_2 in the drawing). The inference system considered makes use of these weights to draw profit from the two RBs.

APPENDIX III
NOTATION

Table IX collects the notation used in this paper.

REFERENCES

[1] A. Bárdossy and L. Duckstein, *Fuzzy Rule-Based Modeling With Application to Geophysical, Biological and Engineering Systems*. Boca Raton, FL: CRC, 1995.

[2] W. Pedrycz, Ed., *Fuzzy Modeling: Paradigms and Practice*. Norwell, MA: Kluwer, 1996.

[3] M. Sugeno and T. Yasukawa, "A fuzzy-logic-based approach to qualitative modeling," *IEEE Trans. Fuzzy Syst.*, vol. 1, pp. 7–31, Feb. 1993.

[4] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning," *Inform. Sci.*, vol. 8, pp. 301–357, 1975.

[5] R. Fullér, *Introduction to Neuro-Fuzzy Systems*. Heidelberg, Germany: Springer-Verlag, 2000.

[6] D. Nauck, F. Klawonn, and R. Kruse, *Fundations of Neuro-Fuzzy Systems*. New York: Wiley, 1997.

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

[8] O. Cordón and F. Herrera, "A proposal for improving the accuracy of linguistic modeling," *IEEE Trans. Fuzzy Syst.*, vol. 8, pp. 335–344, June 2000.

[9] 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.

[10] 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.

[11] L.-X. Wang and J. M. Mendel, "Generating fuzzy rules by learning from examples," *IEEE Trans. Syst., Man, Cybern.*, vol. 22, pp. 1414–1427, Nov. 1992.

[12] O. Cordón and F. Herrera, "A three-stage evolutionary process for learning descriptive and approximate fuzzy logic controller knowledge bases from examples," *Int. J. Approx. Reas.*, vol. 17, no. 4, pp. 369–407, 1997.

[13] O. Cordón, F. Herrera, and P. Villar, "Analysis and guidelines to obtain a good fuzzy partition granularity for fuzzy rule-based systems using simulated annealing," *Int. J. Approx. Reas.*, vol. 25, no. 3, pp. 187–215, 2000.

[14] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its application to modeling and control," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-15, pp. 116–132, Jan. 1985.

[15] P. J. M. van Laarhoven and E. H. L. Aarts, *Simulated Annealing: Theory and Applications*. Norwell, MA: Kluwer, 1987.

[16] R. Alcalá, J. Casillas, O. Cordón, and F. Herrera, "Improvement to the cooperative rules methodology by using the ant colony system algorithm," *Math. Soft Comput.*, vol. 8, no. 3, pp. 321–335, 2001.

[17] O. Cordón, F. Herrera, and L. Sánchez, "Solving electrical distribution problems using hybrid evolutionary data analysis techniques," *Appl. Intell.*, vol. 10, no. 1, pp. 5–24, 1999.

[18] O. Cordón, F. Herrera, and A. Peregrin, "Applicability of the fuzzy operators in the design of fuzzy logic controllers," *Fuzzy Sets Syst.*, vol. 86, no. 1, pp. 15–41, 1997.

[19] S. Kirkpatrick, "Optimization by simulated annealing—Quantitative studies," *J. Stat. Phys.*, vol. 34, pp. 975–986, 1984.

Jorge Casillas was born in Granada, Spain, in 1975. He received the M.Sc. degree in computer engineering and the Ph.D. degree in computer science, both from the University of Granada, Spain, in 1998 and 2001, respectively.

He is an Assistant Professor with the Department of Computer Science and Artificial Intelligence, University of Granada. His research interests include fuzzy modeling, and learning/tuning fuzzy systems for modeling, control, and classification with different metaheuristics.

Oscar Cordón received the M.Sc. and Ph.D. degrees in computer science from the University of Granada, Spain, in 1994 and 1997, respectively.

He is an Associate Professor with the Department of Computer Science and Artificial Intelligence, University of Granada. He has published about 30 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-organized several special sessions in international conferences and co-edited some special issues in international journals on "genetic fuzzy systems" and "information retrieval." His current main research interests are in the fields of fuzzy rule-based systems, fuzzy and linguistic modeling, fuzzy classification, genetic fuzzy systems, evolutionary algorithms, ant colony optimization, and information retrieval.

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

He is an Associate Professor with the Department of Computer Science and Artificial Intelligence, University of Granada. He is coauthor of the book *Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases* (Singapore: World Scientific, 2001). He has co-edited an international book, co-organized several special sessions in international conferences, and co-edited some special issues in international journals on "genetic fuzzy systems" and "computing with words." His research interests include decision making problems in fuzzy environment, fuzzy rule-based systems, machine learning, ant colony optimization, genetic algorithms, and genetic fuzzy systems.