A Proposal for Improving the Accuracy of Linguistic Modeling

Oscar Cordón and Francisco Herrera

***Abstract*—In this paper, we propose *accurate linguistic modeling*, a methodology to design linguistic models that are accurate to a high degree and may be suitably interpreted. This approach will be based on two main assumptions related to the interpolative reasoning developed by fuzzy rule-based systems: a small change in the structure of the linguistic model based on allowing the linguistic rule to have two consequents associated and a different way to obtain the knowledge base based on generating a preliminary fuzzy rule set composed of a large number of rules and then selecting the subset of them best cooperating. Moreover, we will introduce two variants of an automatic design method for these kinds of linguistic models based on two well-known inductive fuzzy rule generation processes and a genetic process for selecting rules. The accuracy of the proposed methods will be compared with other linguistic modeling techniques with different characteristics when solving of three different applications.**

***Index Terms*—Descriptive Mamdani-type fuzzy rule-based systems, double-consequent linguistic rules, genetic algorithms, inductive fuzzy rule generation, linguistic modeling, rule selection.**

I. Introduction

NOWADAYS, one of the most important applications of fuzzy rule-based systems (FRBS's) is *system modeling*[1], [2], which in this field may be considered as an approach used to model a system making use of a descriptive language based on fuzzy logic with fuzzy predicates [3]. When a descriptive Mamdani-type FRBS [4]—a kind of fuzzy system whose knowledge base (KB) is comprised by fuzzy rules composed of linguistic variables [5] that take values in a term set with a real-world meaning—is considered to compose the model structure, the linguistic model so obtained consists of a set of linguistic descriptions regarding the behavior of the system being modeled [3]. Hence, the research field developing system modeling with these kinds of FRBS's is usually called *linguistic modeling*.

One of the problems associated with linguistic modeling is its lack of accuracy in some cases. As Zadeh pointed out in his *principle of incompatibility* [6], "*as the complexity of a system increases, our ability to make precise and yet significant statements about its behavior diminishes . . ..*" Thus, although the use of descriptive Mamdani-type FRBS's allows us to deal with the modeling of systems in which a certain degree of imprecision is involved, building a linguistic model clearly interpretable by human beings, the accuracy obtained is not always as good as desired and we prefer a loss in the model description ability to obtain an improvement in the overall model performance. The choice between how interpretable and how accurate the model must be usually depends on the user's needs for the specific problem and will condition the kind of FRBS selected to generate it. When the accuracy is the main modeling requirement, other kinds of FRBS's, such as Takagi–Sugeno–Kang (TSK) [7] or approximate Mamdani-type ones [1], [8], [9], can be considered but the final model obtained will have associated the drawback of an important loss in its descriptive power.

In this paper, we introduce *accurate linguistic modeling* (ALM), a linguistic modeling approach which will allow us to improve the accuracy of linguistic models without losing its interpretability to a high degree. This approach will be based on two main assumptions relating to the interpolative reasoning developed by FRBS's:

- a small change in the structure of the linguistic model to locally improve the model accuracy: the coexistence of single and double-consequent rules in the KB;
- a different way to derive the KB to globally improve the rule cooperation based on the generation of a preliminary fuzzy rule set with a large number of single- and double-consequent rules and the selection of the subset of them best cooperating.

To do so, this paper is set up as follows. In Section II, the basis of ALM will be introduced. In Section III, two variants of an automatic design method to generate linguistic models of this new kind based on two well-known inductive fuzzy rule generation processes and a genetic process for selecting rules will be proposed. In Section IV, the behavior of both processes will be analyzed for solving of three different applications, the fuzzy modeling of a three-dimensional (3-D) function, the problem of rice taste evaluation and an electrical engineering distribution problem. The results obtained will be compared with other processes with different characteristics. Finally, in Section V, some concluding remarks will be pointed out.

II. ALM: An Approach to Generate Accurate Linguistic Models

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 FRBS's and is a consequence of the *cooperation among the fuzzy rules composing the KB*. As is known, the output obtained from an FRBS is not usually due to a single fuzzy rule but to the cooperative action of several fuzzy rules that have been fired because they match the input to the system to some degree.

ALM will deal with the way in which the linguistic model makes inference in order to improve its accuracy while not losing its description. Hence, it will be based on two main aspects that will be described in the two following subsections. The remaining one in this section analyzes some interesting remarks of the proposed approach.

Manuscript received June 12, 1998; January 3, 2000. This work was supported by Project CICYT TIC96-0778.

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

Publisher Item Identifier S 1063-6706(00)05119-5.

A. New Descriptive Knowledge Base Structure to Locally Improve the Model Accuracy

1) Problems of the Usual Linguistic Modeling Structure: The KB structure usually employed in the field of linguistic modeling is composed of a collection of Mamdani-type linguistic rules with the form

$$R_i: \text{ IF } x_1 \text{ is } A_{i1} \text{ and } \cdots \text{ and } x_n \text{ is } A_{\text{in}} \text{ THEN } y \text{ is } B_i$$

with $x_1, \cdots, x_n$ and y being the input linguistic variables and the output one, respectively; and with $A_{i1}, \cdots, A_{\text{in}}, B_i$ being linguistic labels from different term sets with fuzzy sets associated defining their meaning. These fuzzy sets are defined on the universes of discourse $U_1, \cdots, U_n, V$ and are characterized by their membership functions

$$\mu_{A_{ij}}(\mu_{B_i}): U_j(V) \rightarrow [0, 1], \quad j = 1, \cdots, n.$$

This KB structure has the drawback of its lack of accuracy when working with very complex systems. This fact is due to some problems relating to the fuzzy rule structure considered, which are a consequence of the inflexibility of the concept of linguistic variable. A brief summary of these problems is shown as follows [9], [10].

- Lack of flexibility due to the rigid partitioning of the input and output spaces.
- The homogeneous partitioning of these spaces when the input–output mapping varies in complexity within the space is inefficient and does not scale to high-dimensional spaces.
- Dependent input variables are very hard to partition.
- Limitation on the size of the rule base (RB).

Due to these reasons, we consider obtaining a new more flexible KB structure that allows us to improve the accuracy of linguistic models without losing their interpretability.

2) Considering Double-Consequent Rules in the Knowledge Base: In [11], an attempt was made to put this idea into effect first by designing a fuzzy model based on simplified TSK-type rules, i.e., rules with a single point in the consequent, and then transforming it into a linguistic model, which has to be as accurate as the former. To do so, they introduce a secondary KB, in addition to the usual KB, and propose an inference system capable of obtaining an output result from the combined action of both fuzzy rule bases. Hence, what the system really does is to allow a specific combination of antecedents to have two different consequents associated, the first and second in importance, thus avoiding some of the said problems associated to the linguistic rule structure. On the one hand, they achieve that the maximum number of rules is not prefixed by the granularity of the fuzzy partitions associated to the input variables and, on the other hand, the inflexibility derived from the use of these kinds of partitions is avoided. The fact that the result of the inference on the two resulting rules is an interpolation of their individual outputs allows us to obtain a more accurate model without losing its interpretability.

Taking this idea as a starting point, we allow a specific combination of antecedents to have two consequents associated but only in those cases in which it is really necessary to improve the model accuracy in this subspace and not in all the possible ones as in [11]. Therefore, the existence of a primary and a secondary fuzzy rule base is avoided, and the number of rules in the single KB is decreased, that makes easier to interpret the model.

These double-consequent rules will locally improve the interpolative reasoning performed by the model allowing a shift of the main labels making the final output of the rule lie in an intermediate zone between the two consequent fuzzy sets. Hence, this rule structure allows us to avoid three of the four problems of classical linguistic rules analyzed in the previous subsection: the inflexibility derived from the rigid partitioning of the input and output spaces, the difficulty to scale to complex spaces without increasing the fuzzy partition granularity to a high degree and the limitation on the size of the RB.

On the other hand, we should note that this operation mode does not constitute an inconsistency. Let us suppose that a specific combination of antecedents, "x_1 is A_1 and $\ldots$ and x_n is A_n," has two different consequents associated: B_1 and B_2. The resulting double-consequent rule may be linguistically interpreted as follows:

$$\begin{array}{l}\text{IF } x_1 \text{ is } A_1 \text{ and } \cdots \text{ and } x_n \text{ is } A_n \\ \text{THEN } y \text{ is between } B_1 \text{ and } B_2.\end{array}$$

Finally, two other advantages of our approach with respect to Nozaki *et al.* is that we do not need the existence of a previous TSK fuzzy model to generate the linguistic KB and that we do not need a specific inference system to perform reasoning with the system. The only restriction in our case is to use any defuzzification method working in mode B-FITA (first infer, then aggregate)—a strategy based on first defuzzifying the fuzzy output inferred from each individual rule and then aggregating the characteristic values so obtained— and considering the matching degree of the rules fired (for a review of different defuzzification methods working in this way, refer to [12]. This is due to the fact that *each double-consequent rule is considered as two simple rules with the same antecedent and different consequent to perform the inference process*. Hence, the consideration of a defuzzification method working in mode A-FATI (first aggregate, then infer)—first aggregating the fuzzy outputs inferred from each rule (usually by means of the maximum or the minimum) into a single global fuzzy set and then defuzzifying it to obtain the final output— or of a defuzzification strategy not considering the matching will cancel the influence of one of the two-rule consequents.

In this contribution, we will use the Minimum t-norm in the role of conjunctive and implication operator and the *center of gravity weighted by the matching degree* [12] as defuzzification strategy. In the latter, the final system output y_0 is computed by means of the expression:

$$y_0 = \frac{\sum_{i=1}^{T} h_i \cdot y_i}{\sum_{i=1}^{T} h_i}$$

with y_i being the center of gravity of the fuzzy set infered from rule R_i in the KB and h_i being the matching degree between the system input and the antecedent of R_i.

B. New Way to Generate Fuzzy Rules to Globally Improve the Cooperation Among Them

The second aspect deals with the cooperation among the rules in the KB, i.e., with the overlapped space zones that are covered by different linguistic rules. As is known, the generation of the best fuzzy rule in each input subspace does not ensure that the FRBS will perform well due to the fact that the rules composing the KB may not cooperate suitably. Many times, the accuracy of the FRBS may be improved if other rules different than the primary ones are generated in some subspaces because they cooperate in a better way with their neighbor rules.

Hence, ALM will consider an operation mode based on generating a preliminary fuzzy rule set composed of a large number of rules, which will be single- or double-consequent ones depending on the complexity of the specific fuzzy subspace —no rules will be generated in the subspaces where the system is not defined. Then, all these fuzzy rules will be treated as single-consequent ones (each double-consequent rule will be broken down into two simple rules) and the subset of them with best cooperation level will be selected in order to compose the final RB. Thus, this stage will specify which double-consequent rules in the preliminary rule set will remain in the final RB; that is, those fuzzy subspaces whose two simple rules associated have been finally selected and also will remove those single-consequent rules that are unnecessary due to the cooperative action of neighbor rules on their input subspace.

The said operation mode gives more freedom to the RB generation process since the rule selection process can make the final RB present single-consequent rules not being the best ones in their fuzzy input subspaces in order to improve the cooperation of the global KB.

C. Some Important Remarks About ALM

On the one hand, in view of the assumptions presented in the previous subsections, any ALM process will be composed of the two following methods:

1) *linguistic rule generation method*, which will generate a preliminary rule set where single and double-consequent linguistic rules will coexist according to the problem complexity;
2) *rule selection process*, that will select the subset of rules cooperating best from the preliminary fuzzy rule set generated in the previous step.

On the other hand, we may draw two very important conclusions from the operation mode of ALM. First, it is possible that, although the preliminary fuzzy rule set generated has some double-consequent rules, the final KB does not contain any rule of this kind after the selection process. In this case, the linguistic model obtained has taken advantage of the way in which the fuzzy rules has been generated because many rule subsets with different cooperation levels have been analyzed. This is why it will present a KB composed of rules cooperating well, a fact that may not happen in other inductive design methods —such as the Wang and Mendel's rule generation method [13], that will be considered in next sections—which are based on directly generating the best consequent for each fuzzy input subspace.

In addition, it is possible that the KB obtained presents less rules than KB's generated from other methods thanks to two aspects: both the existence of two rules (a double-consequent rule) in the same input subspace and the generation of neighbor rules with better cooperation may mean that many of the rules in the KB are unnecessary to give the final system response.

All these assumptions will be corroborated in view of the experiments developed in Section IV.

III. Two Specific ALM Processes

In this section, two specific ALM processes will be introduced, only differing on the composition of the rule generation method. It will present two variants, both of them based on modifications made on two inductive linguistic rule generation processes, the Wang and Mendel's method (WM-method) [13] and an adaption of the Ishibuchi *et al.* simplified TSK fuzzy rule generation method [14] to allow it to generate linguistic rules with fuzzy consequents (I-method) (other two generation processes for fuzzy classification rules and TSK fuzzy rules based on the one proposed in that paper are to be found in [15] and [16], respectively).

In both cases, the modification imposed by ALM involves generating the two most important consequents for each combination of antecedents (instead of only the most important one, as usual). We should remark that an input subspace will only have two consequents associated (and, thus, a double-consequent rule) when there is a need to do so, i.e., when there is any data on it allowing two different consequents to be generated.

On the other hand, the rule selection process will be based on a genetic algorithm (GA) [17], although any other optimization technique can be used to develop this task.

Both processes will be described in depth in the next two subsections.

A. Linguistic Rule Generation Method

Our linguistic rule generation method is based on the existence of a set of input–output data (examples) $E = \{e_1, \cdots, e_l, \cdots, e_p\}$, $e_l = (x_1^l, \cdots, x_n^l, y^l)$, representing the behavior of the system being modeled, and of a previous definition of the data base (DB) composed of the fuzzy partitions considered for the input and output variables. These fuzzy partitions may be obtained from the expert information (if it is availaible) or by a normalization process, where each variable domain is divided into a number of equal or unequal partitions, a kind of membership function is selected and a fuzzy set is assigned to each subspace. In this paper, we will work with symmetrical fuzzy partitions of triangular membership functions [see Fig. 1].

As mentioned, this method will present two variants, the WM-based ALM and the I-based ALM. The following subsections introduce them both.

1) The WM-based ALM Generation Method: In this first case, the generation of the RB is put into effect by means of the following steps.

1) *Generate a preliminary linguistic rule set*: This set will be formed by the rule best covering each example e_l

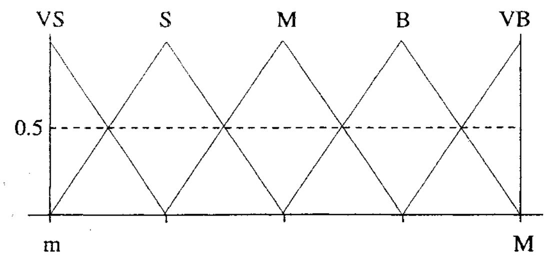

Fig. 1. Graphic representation of the fuzzy partition considered.

contained in E. The structure of these rules is obtained by setting each one of the rule variables to the linguistic label associated to the fuzzy set best covering every example component.

2) *Give an importance degree to each rule*: Let R_l : *IF* x_1 *is* A_1 *and* ... *and* x_n *is* A_n *THEN* y *is* B be the linguistic rule generated from the example e_l. Its importance degree will be obtained as follows:

$$G(R_l) = \mu_{A_1}(x_1^l) \cdot \cdots \cdot \mu_{A_n}(x_n^l) \cdot \mu_B(y^l).$$

3) *Obtain a final RB from the preliminary fuzzy rule set*: This step is the only one that is different from the original WM method. While in that method the rule with the highest importance degree is the only one chosen for each combination of antecedents, in our case, we allow two different rules, the two most important ones in each input subspace (if they exist), to form part of the RB composing a double-consequent rule. Of course, a combination of antecedents may have no rule associated (if there are no examples in that input subspace) or only one rule (if all the examples in that subspace generated the same rule). Therefore, the generation of rules with double consequent is only addressed when the problem complexity, represented by the example set, shows that it is necessary.

2) The I-Based ALM Generation Method: In [14], Ishibuchi *et al.*proposed an inductive method to design simplified TSK FRBS's. The algorithm worked in a different way from the WM method since all the examples located in each fuzzy input subspace are considered to generate a single fuzzy rule for that subspace, its consequent obtained by averaging the outputs of those examples weighted by the matching degree of their inputs with the rule antecedent.

We have developed two different modifications on the previous algorithm. First, it has been adapted to generate descriptive Mamdani-type rules. Second, we have allowed the process to generate not only one but two different consequents when it is necessary, the same as in the previous section. Thus, the generation method so obtained consists of the following steps. *For each multidimensional fuzzy input subspace obtained by combining the individual input variable subspaces using the "and" conjunction do*:

1) build the set E' composed of the examples $e_i \in E$ that are located in this subspace;

2) if $|E'| \neq 0$, i.e., if there is any data on it, then the following holds.
   a) For each linguistic label B_j in the output variable term set, $\{B_1, \cdots, B_t\}$, build a rule using the current antecedent and B_j in the consequent. Compute the covering value of each linguistic rule so generated, $R_j \in R^c = \{R_1, \cdots, R_t\}$, over each example $e_l \in E'$ as follows:

$$R_j : \text{IF } x_1 \text{ is } A_1 \text{ and} \cdots \text{and } x_n \text{ is } A_n \text{ THEN } y \text{ is } B_j$$
$$R_j(e_l) = T\left(\mu_{A_1}(x_1^l), \cdots, \mu_{A_n}(x_n^l), \mu_{B_j}(y^l)\right)$$

   with $j = 1, \cdots, t$ and T being a t-norm. In this paper, we will work with the minimum.
   b) Select the *two rules* from R^c with the highest values in the *rule selection function* $S(\cdot)$ considered.
   c) Add this double-consequent rule to the RB.

   Otherwise, do not generate rules in that input fuzzy subspace.

Different choices may be considered for the rule selection function, S. In this paper, we work with the following ones.

- **Maximum covering over the example set**: This index computes the *accumulated covering degree of the rule over all the examples in the example set* E'

$$S_1(R_j) = \sum_{e_l \in E'} R_j(e_l).$$

- **Maximum covering of the example best covered**: The value associated to each rule is its *covering degree over the best covered example in* E'

$$S_2(R_j) = \max_{e_l \in E'} R_j(e_l).$$

- **Average of covering degrees**: The *average of the two previous covering degrees* is calculated. To do so, the S_1 index is normalized (dividing it by the size of E', $|E'|$) to have them both in the $[0, 1]$ interval

$$S_3(R_j) = \frac{\sum_{e_l \in E'} R_j(e_l)}{|E'|} \cdot \max_{e_l \in E'} R_j(e_l).$$

We should note that unlike the generation method introduced in the previous section, in this case, it is very difficult for a fuzzy subspace to have a single linguistic rule associated. It may have none (when there is no example located on it) or two rules associated (otherwise), due to the fact that each example is almost always covered by two different fuzzy sets because of the type of output fuzzy partition considered. Therefore, *this generation process will always generate more (or, at least, the same) rules than the WM-based one*.

B. Rule Selection Genetic Process

The selection of the subset of linguistic rules best cooperating is a combinatorial optimization problem [3], [18]. Since the number of variables involved in it, i.e., the number of preliminary rules, may be very large, we consider an approximate algorithm to solve it—a GA [17].

Our rule selection genetic process is based on a binary coded GA in which the selection of the individuals is performed using the stochastic universal sampling procedure together with an elitist selection scheme and the generation of the offspring population is put into effect by using the classical binary multipoint crossover (performed at two points) and uniform mutation operators [19], [20].

The coding scheme generates fixed-length chromosomes. Considering the single-consequent rules contained in the linguistic rule set derived from the previous step counted from 1 to m, an m-bit string $C = (c_1, \cdots, c_m)$ represents a subset of candidate rules to form the RB finally obtained as this stage output B^s such that

$$\textit{If } c_i = 1 \text{ then } R_i \in B^s \text{ else } R_i \notin B^s.$$

The initial population is generated by introducing a chromosome representing the complete previously obtained rule set, i.e., with all $c_i = 1$. The remaining chromosomes are selected at random.

As regards the fitness funtion $F(C_j)$, it is based on a global error measure that determines the accuracy of the FRBS encoded in the chromosome, which depends on the cooperation level of the rules existing in the KB. We usually work with the mean square error (SE), although other measures may be used. SE over a training data set E is represented by the following expression:

$$F(C_j) = \frac{1}{2 \cdot |E|} \sum_{e_l \in E} (y^l - S(ex^l))^2$$

with $S(ex^l)$ being the output value obtained from the FRBS using the RB coded in C_j, when the input variable values are $ex^l = (x_1^l, \cdots, x_n^l)$ and y^l is the known desired value.

We should note that this basic rule selection genetic process can be extended in different ways: trying also to minimize the number of rules in the selected RB by means of a weighted average [18] or a multi-objective GA [21], forcing this RB to cover all the examples in the training set to a specific degree [19] or considering a niching GA to perform a better search in the multimodal space [22].

For the sake of simplicity, none of this extensions will be considered in this paper. Although better results could be obtained in other case, the implementation of the rule selection is not a key question in our methodology and, thus, we prefer not to include additional experimentations related to this point. Moreover, we think that the criterion considered, the SE over a data set, will be enough to obtain good results due to the fact that it is directly related to the accuracy of the linguistic model; that is significantly affected by an excessive number of rules or by the possibility of non covering any of the training examples.

IV. Examples of Application: Experiments Developed and Results Obtained

With the aim of analyzing the behavior of the proposed ALM processes, three different applications have been chosen: the fuzzy modeling of a 3-D function [19], the problem of rice taste evaluation [11], [14], and a real-world Spanish electrical engineering distribution problem [23]–[25]. In all three cases, we will compare the accuracy of the linguistic models generated from our processes with the ones designed by means of other methods: the one proposed by Nozaki *et al.* (NIT-method) in [11] that has been analyzed in Section II-A.2, and the basic WM-method and I-method ones. In addition, classical regression techniques and neural models will be considered in the electrical distribution problem.

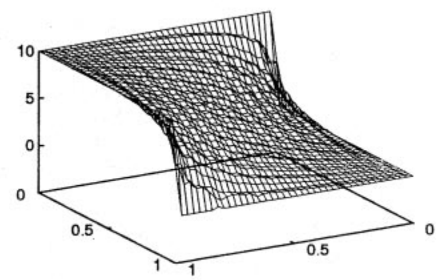

Fig. 2. Graphical representation of the function considered.

A. Fuzzy Modeling of a 3-D Function

The expression of the selected function is shown as follows, along with the universes of discourse considered for the variables [19]. It is a simple function presenting two discontinuities at the points $(0, 0)$ and $(1, 1)$. Its representation is shown in Fig. 2.

$$F(x_1, x_2) = 10 \cdot \frac{x_1 - x_1 x_2}{x_1 - 2x_1 x_2 + x_2}, \quad x_1, x_2 \in [0, 1], \quad F(x_1, x_2) \in [0, 10].$$

In order to model this function, a training data set composed of 674 pieces of data uniformly distributed in the 3-D definition space has been obtained experimentally. A test set composed of 67 examples has been randomly generated to evaluate the performance of the design methods, avoiding any possible bias related to the data in the training set.

The DB used for all design methods is constituted by three primary fuzzy partitions formed by *seven linguistic terms* with triangular-shaped fuzzy sets giving meaning to them [as shown in Fig. 1]. The linguistic term set considered is $\{NB, NM, NS, ZE, PS, PM, PB\}$, standing N for negative, ZE for zero, P for positive, and S, M, and B for small, medium, and big, respectively.

Finally, the parameters considered for the rule selection genetic process are: Number of generations: 500; Population size: 61; Crossover probability: 0.6; and Mutation probability: 0.1 (per individual).

The results obtained in the experiments developed are collected in Table I where #R stands for the number of single-consequent rules in the corresponding KB, $\mathrm{SE_{tra}}$ and $\mathrm{SE_{tst}}$ for the values obtained in the SE measure computed over the training and test data sets, respectively, and SF for the rule selection function used in the I-method and I-based ALM generation method. As may be observed, the results obtained by our

TABLE I
RESULTS OBTAINED IN THE MODELING OF THE SELECTED FUNCTION

	Generation			Selection		
Method	#R	SE_{tra}	SE_{tst}	#R	SE_{tra}	SE_{tst}
NIT-method	98	0.175382	0.061249	–	—	—
WM-method	**49**	0.194386	0.044466	–	—	—
WM-based ALM	88	0.220062	0.146529	55	0.019083	0.026261
I-method (SF=S_1)	**49**	0.168582	0.076571	–	—	—
I-method (SF=S_2)	**49**	0.194386	0.044466	–	—	—
I-method (SF=S_3)	**49**	0.097398	0.043984	–	—	—
I-based ALM (SF=S_1)	98	0.254952	0.189540	55	**0.019083**	**0.022120**
I-based ALM (SF=S_2)	98	0.295414	0.218886	53	0.025444	0.030144
I-based ALM (SF=S_3)	98	0.264324	0.184874	51	0.020137	0.025837

TABLE II
RESULTS OBTAINED IN THE RICE TASTE EVALUTION USINGTWO LABELS IN THE FUZZY PARTITION

	Generation			Selection		
Method	#R	SE_{tra}	SE_{tst}	#R	SE_{tra}	SE_{tst}
NIT-method	64	0.00862	0.00985	–	—	—
WM-method	15	0.01328	0.01311	–	—	—
WM-based ALM	19.8	0.02192	0.02412	**5**	**0.00341**	**0.00398**
I-method (SF=S_1)	32	0.02129	0.02126	–	—	—
I-method (SF=S_2)	32	0.05220	0.05369	–	—	—
I-method (SF=S_3)	32	0.00910	0.01012	–	—	—
I-based ALM (SF=S_1)	64	0.02391	0.02628	8.5	0.00356	0.00508
I-based ALM (SF=S_2)	64	0.02391	0.02628	8.6	0.00358	0.00509
I-based ALM (SF=S_3)	64	0.02391	0.02628	7.9	0.00359	0.00470

processes after each individual stage, generation and selection, have been included, and the best results are shown in boldface.

In view of these results, we should underline the good performance presented by both variants of our ALM process that generate the most accurate models in the approximation of the training and test sets. Their behavior is also appropriate as regards the model complexity since our models presents only a few more rules than the ones generated from the WM-method and the I-method. For example, by only adding eight new rules to the RB generated by means of the WM-method (and by removing two), the best model is obtained by the I-based ALM process (considering the rule selection function S_1) with a significantly higher accuracy than the latter one and with a very small loss of interpretability (see both RB's in Table IV). Finally, all the ALM models are more accurate to a high degree than the NIT-method one, presenting much simpler KB's (a maximum of 55 rules against 98).

B. Rice Taste Evaluation

Rice taste evaluation is usually put into effect by means of a *sensory test*, where a group of 24 experts gives an *evaluation* of rice kinds according to five characteristics: *flavor*, *appearance*, *taste*, *stickiness*, and *toughness* [11], [14].

The modeling of this problem becomes very complex due to the large quantity of relevant variables and to the fact that the problem-solving goal is not only to obtain an accurate model, but also a user-interpretable model representing the nonlinear relationships existing in the problem as well as putting some light on the reasoning process performed by the human experts.

In this section we deal with obtaining several linguistic models to solve the rice evaluation problem by considering the data set presented in [11]. This set is composed of 105 data arrays collecting subjective evaluations of the five input and one output variables, all of them normalized in $[0, 1]$ for the same number of rice kinds.

With the aim of not biasing the learning, ten different partitions of the data set have been randomly obtained, composed by 75 (30) pieces of data in the training (test) set, to generate ten linguistic models in each experiment. We will use the same linguistic modeling processes considered in the previous section, as well as the same parameter values in the rule selection genetic process.

As was done in [11], a different number of linguistic labels has been considered for the fuzzy partitions (two and three triangular fuzzy sets). Higher granularities will not be considered since they cause the KB's in the linguistic models obtained to be excessively complex (to have many rules) and thus less interpretable.

The results obtained are collected in Tables II and III. The values shown in columns $\mathrm{SE}_{\mathrm{tra}}$, $\mathrm{SE}_{\mathrm{tst}}$, and $\#R$ have been computed as an average of the values obtained by the ten linguistic models generated in each case.

From an analysis of these results, we may see that all the ALM models clearly outperform again the remainder linguistic models in the two experiments developed. As regards the individual results, the best ones correspond to the WM-based ALM process—which obtains very simple models with an average of five rules—when considering two labels and to the I-based one

TABLE III
RESULTS OBTAINED IN THE RICE TASTE EVALUTION USING THREE LABELS IN THE FUZZY PARTITION

	Generation			Selection		
Method	#R	SE_{tra}	SE_{tst}	#R	SE_{tra}	SE_{tst}
NIT-method	364.8	0.00251	0.00322	–	—	—
WM-method	23	0.00333	0.00375	–	—	—
WM-based ALM	25.7	0.00595	0.00736	**12.2**	0.00185	0.00290
I-method (SF=S_1)	182.4	0.00390	0.00512	–	—	—
I-method (SF=S_2)	182.4	0.00500	0.00468	–	—	—
I-method (SF=S_3)	182.4	0.00328	0.00431	–	—	—
I-based ALM (SF=S_1)	364.8	0.00695	0.00634	146	0.00146	0.00289
I-based ALM (SF=S_2)	364.8	0.00644	0.00602	150.3	0.00144	0.00344
I-based ALM (SF=S_3)	364.8	0.00713	0.00726	136.2	**0.00141**	**0.00275**

TABLE IV
RBS OF THE LINGUISTIC MODELS OBTAINED FOR THE FUNCTION BY THE WM-METHOD (TOP) AND I-BASED ALM PROCESS (BOTTOM)

x_1 \ x_2	NB	NM	NS	ZE	PS	PM	PB
NB	*NB*	*NB*	*NB*	*NB*	*NB*	*NB*	*NB*
NM	*PB*	*ZE*	*NS*	*NM*	*NM*	*NB*	*NB*
NS	*PB*	*PS*	*ZE*	*NS*	*NM*	*NM*	*NB*
ZE	*PB*	*PM*	*PS*	*ZE*	*NS*	*NM*	*NB*
PS	*PB*	*PM*	*PM*	*PS*	*ZE*	*NS*	*NB*
PM	*PB*	*PB*	*PM*	*PM*	*PS*	*ZE*	*NB*
PB	*PB*	*PB*	*PB*	*PB*	*PB*	*PB*	*NB*

x_1 \ x_2	NB	NM	NS	ZE	PS	PM	PB
NB		*NB*	*NB*	*NB*	*NB*	*NB*	*NB*
NM	*PB*	*ZE*	*NM* *NS*	*NM*	*NB* *NM*	*NB*	*NB*
NS	*PB*	*PS* *PM*	*ZE*	*NS*	*NM* *NS*	*NB* *NM*	*NB*
ZE	*PB*	*PM*	*PS*	*ZE*	*NS*	*NM*	*NB*
PS	*PB*	*PM* *PB*	*PS* *PM*	*PS*	*ZE*	*NS*	*NB*
PM	*PB*	*PB*	*PM* *PB*	*PM*	*PS*	*ZE*	*NB*
PB	*PB*	*PB*	*PB*	*PB*	*PB*	*PB*	

TABLE V
RBS OF THE BEST LINGUISTIC MODELS IN THE RICE PROBLEM GENERATED BY THE WM METHOD (TOP) AND WM-BASED ALM PROCESS (BOTTOM) USING 2 LABELS

	Flavor	App.	Taste	Stick.	Tough.	Eval.
R_1:	Good	Good	Good	Sticky	Tender	High
R_2:	Good	Good	Good	Not st.	Tender	High
R_3:	Good	Good	Good	Not st.	Tough	High
R_4:	Good	Good	Bad	Not st.	Tender	High
R_5:	Good	Good	Bad	Sticky	Tender	High
R_6:	Good	Good	Good	Sticky	Tough	High
R_7:	Good	Bad	Bad	Not st.	Tough	Low
R_8:	Good	Good	Bad	Not st.	Tough	Low
R_9:	Bad	Bad	Bad	Not st.	Tender	Low
R_{10}:	Bad	Good	Good	Sticky	Tender	Low
R_{11}:	Good	Bad	Bad	Not st.	Tender	Low
R_{12}:	Bad	Bad	Bad	Not st.	Tough	Low
R_{13}:	Good	Bad	Good	Not st.	Tender	Low
R_{14}:	Bad	Good	Good	Not st.	Tender	Low
R_{15}:	Bad	Good	Bad	Not st.	Tender	Low

	Flavor	App.	Taste	Stick.	Tough.	Eval.
R_1:	Good	Good	Good	Sticky	Tender	High
R_2:	Good	Good	Good	Sticky	Tough	High
R_3:	Bad	Good	Good	Not st.	Tender	High
R_4:	Good	Bad	Bad	Not st.	Tender	Low
R_5:	Bad	Bad	Bad	Not st.	Tough	Low
R_6:	Bad	Good	Bad	Not st.	Tender	Low

using the rule selection function S_3 when working with three (although the number of rules is somewhat large, which also makes better to choose the WM-based ALM model in order to improve the interpretability).

Even considering the generation of double-consequent rules, the RB's generated from the two ALM processes contain less rules than the ones obtained from the corresponding single-consequent generation processes in all cases. Moreover, all of them present less rules than the NIT-method KB. Overall, our models are simpler and this makes them easier to be interpreted.

In fact, none of the 20 models generated from the WM-based ALM process presents double-consequent rules in their RB's. This leads us to conclude that, as mentioned in Sections II-B and II-C, the operation mode based on generating a preliminary fuzzy rule set with a large number of rules and selecting the subset of them cooperating best allows us to obtain good results in linguistic modeling.

As an example, Table V shows the composition of the RB's of the linguistic model with the best generalization level generated from the WM-based ALM process $(\mathrm{SE_{tra}} = 0.003\,83, \mathrm{SE_{tst}} = 0.002\,85)$ when considering 2 labels, and of the one generated from the WM method using the same data set as well $(\mathrm{SE_{tra}} = 0.014\,70, \mathrm{SE_{tst}} = 0.016\,70)$.

C. Electrical Engineering Distribution Problem

Sometimes, there is a need to measure the amount of electricity lines that an electric company owns. This measurement may be useful for several aspects such us the estimation of the maintenance costs of the network, which was the main goal of the problem presented here in Spain [23]–[25]. High and medium voltage lines can be easily measured, but low voltage line is contained in cities and villages and it would be very expensive to measure it. This kind of line used to be very convoluted and, in some cases, one company may serve more than 10 000 small nuclei.

Therefore, there is a need to find a relationship between some characteristics of the population and the length of line installed in it, making use of some known data, that may be employed to predict the real length of line in any other village. We will try to solve this problem by generating different kinds of models determining the unknown relationship. To do so, we were provided with the measured line length (output variable), and the number of inhabitants and the mean distance from the center of the town to the three furthest clients (input variables) in a sample of 495 rural nuclei [23].

TABLE VI
RESULTS OBTAINED IN THE ELECTRICAL APPLICATION CONSIDERING FIVE LABELS IN THE FUZZY PARTITIONS

	Generation			Selection		
Method	#R	SE_{tra}	SE_{tst}	#R	SE_{tra}	SE_{tst}
NIT-method	40	229104.8	206636.4	–	—	—
WM-method	**13**	298446.0	282058.1	–	—	—
WM-based ALM	23	263674.1	268679.9	**13**	**178571.4**	**180847.8**
I-method (SF=S_1)	20	320443.5	282657.9	–	—	—
I-method (SF=S_2)	20	310308.8	286775.1	–	—	—
I-method (SF=S_3)	20	329726.2	306325.7	–	—	—
I-based ALM (SF=S_1)	40	283205.6	288685.8	23	186928.2	183504.6
I-based ALM (SF=S_2)	40	264718.4	262801.3	22	179383.2	181284.2
I-based ALM (SF=S_3)	40	283205.6	288685.8	23	186928.2	183504.6

TABLE VII
RESULTS OBTAINED IN THE ELECTRICAL APPLICATION CONSIDERING SEVEN LABELS IN THE FUZZY PARTITIONS

	Generation			Selection		
Method	#R	SE_{tra}	SE_{tst}	#R	SE_{tra}	SE_{tst}
NIT-method	64	185383.1	170480.4	–	—	—
WM-method	24	222622.7	240018.2	–	—	—
WM-based ALM	34	231174.2	260067.3	**20**	155866.3	178601.1
I-method (SF=S_1)	32	275870.0	269601.7	–	—	—
I-method (SF=S_2)	32	239380.7	276108.4	–	—	—
I-method (SF=S_3)	32	267923.9	249523.8	–	—	—
I-based ALM (SF=S_1)	64	219533.6	191484.0	34	155611.6	211724.6
I-based ALM (SF=S_2)	64	214601.7	252990.3	38	166499.3	187903.8
I-based ALM (SF=S_3)	64	208317.0	188929.1	35	**154469.9**	**167061.3**

To compare classical methods, linguistic modeling and neural modeling we have randomly divided the sample into two sets comprising 396 and 99 samples, labeled training and test. The SE is considered again to measure the accuracy of the different models generated.

In this case, the linguistic variable fuzzy partitions are divided into five and seven fuzzy sets in the experiments developed, with the term set considered in the latter case being $\{ES, VS, S, M, B, VB, EB\}$, with S standing for small, M for medium, B for big, and E and V for extremely and very, respectively. The values of the rule selection process parameters are still the same. The results obtained with the different linguistic modeling methods considered are shown in Tables VI and VII.

In view of the results obtained, we should remark some important conclusions. First, and in the same way as in the two previous applications, the different models generated from the two variants of our ALM process clearly outperform the NIT-method and both single consequent inductive generation methods, WM-method and I-method, in the two experiments carried out, with the differences being more significant between them in this problem. For example, in the vast majority of the cases, the accuracy of the models generated from the first step of the proposed process (i.e., before applying the rule selection genetic process) is already better than the model designed from the corresponding single consequent generation process.

The WM-based ALM process generates the most accurate model with five labels but, on the contrary, the best model is obtained from the I-based ALM process (with SF S_3) when using seven labels. In the latter case, the differences are more significant, both between the models generated by the I-based ALM process and the WM-based one and among the three models generated by the former variant.

TABLE VIII
RB'S FROM THE WM-METHOD (TOP) AND WM-BASED ALM PROCESS (BOTTOM) FOR THE ELECTRICAL APPLICATION (SEVEN LABELS)

x_1 \ x_2	**ES**	**VS**	**S**	**M**	**B**	**VB**	**EB**
ES	ES	VS	VS	S		VS	M
VS	ES	VS	VS	B	S	B	
S		VS	M	S	S	B	
M		VS	S	VB	EB	M	
B			M				
VB							
EB				S			

x_1 \ x_2	**ES**	**VS**	**S**	**M**	**B**	**VB**	**EB**
ES	ES	ES VS	VS	VS		VS	M
VS	ES	VS	VS M	S B		B	
S		VS					
M			S	VB		M	
B			M				
VB							
EB				S			

As regards the model complexity, the behavior is different in the two variants, but really good in both. While the models obtained by means of the I-based ALM process present a few more rules than the ones obtained from the I-method (a max-

TABLE IX
OVERALL RESULTS OBTAINED IN THE ELECTRICAL APPLICATION

Method	SE_{tra}	SE_{tst}
Linear	287775	209656
Exponential	232743	197004
2th order polynomial	235948	203232
3rd order polynomial	235934	202991
3 layer perceptron 2-25-1	169399	167092
NIT-method	185383	170480
WM-method	222622	240018
WM-based ALM	155866	178601
I-method (SF=S_3)	267923	249523
I-based ALM (SF=S_3)	**154469**	**167061**

imum of 23 against 20, when considering five labels, and 38 against 32, when considering seven), those RB's generated from the WM-based ALM process have the same or less rules than the ones derived from the WM-method (13 in both cases when working with five labels and 20 against 24 when considering seven, as shown in Table VIII). The latter is a very important result, because we are able to obtain a simpler and more accurate model following the approach proposed in this paper (we should remember that we found the same assumption in the rice taste evaluation problem).

Finally, Table IX shows the best results obtained by all the modeling techniques considered for the problem. The parameters of the polynomial models were fitted by Levenberg–Marquardt, while exponential and linear models were fitted by linear least squares. The multilayer perceptron was trained with the quick propagation algorithm. The number of neurons in the hidden layer was chosen to minimize the test error.

As may be seen, all the linguistic models clearly outperform classical regression ones in the approximation of both data sets. Besides, the linguistic model generated from the I-based process considering the rule selection function S_3 is more accurate than the neural one, which is the second best model in view of its generalization level. Although the results are almost the same in this characteristic (167 061 versus 167 092), the value obtained by the linguistic model in the SE over the training data set shows a significant performance advantage for it over the neural network (154 469 versus 169 399). Therefore, this model is the one that best approximates the real system and that presents best generalization capabilities and, moreover, it has the advantage of being much more interpretable than the neural model.

V. CONCLUDING REMARKS

In this paper, ALM has been proposed, which is a new approach to design linguistic models accurate to a high degree and suitably interpretable by human-beings. Both ALM assumptions trying to improve the interpolative reasoning in FRBS's have proven to be effective. On the one hand, double-consequent linguistic rules have demonstrated to improve the model accuracy in some specific space zones presenting a higher complexity, while maintining the model description by the following interpretation:

$$\text{IF } x_1 \text{ is } A_1 \text{ and} \cdots \text{and } x_n \text{ is } A_n$$
$$\text{THEN } y \text{ is between } B_1 \text{ and } B_2.$$

On the other hand, the generation of a preliminary fuzzy rule set composed of a large number of single and double-consequent rules according to the problem complexity, the consideration of all these rules as simple ones and the selection of the subset cooperating best among them have allowed us to improve the cooperation among the rules in the KB and, thus, the global model accuracy in all the definition space.

ACKNOWLEDGMENT

The authors would like to thank to Dr. L. Sánchez, Oviedo University, Spain, for the electrical engineering application from Hidroeléctrica del Cantábrico and for solving it by means of classical and neural techniques.

REFERENCES

[1] A. Bardossy and L. Duckstein, *Fuzzy Rule-Based Modeling with Application to Geophysical, Biological and Engineering Systems*. Orlando, 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. X. Wang, *Adaptive Fuzzy Systems and Control*. Englewood Cliffs, NJ: Prentice-Hall, 1994.
[5] L. A. Zadeh, "The concept of a linguistic variable and its application to aproximate reasoning," *Inform. Sci.*. pt.: I, vol. 8, pp. 199–249, 1975; pt.: II vol. 8, pp. 301–357, 1975; pt.: III vol. 9, pp. 43–80, 1975.
[6] ——, "Outline of a new approach to the analysis of complex systems and decision processes," *IEEE Trans. Syst., Man, Cybern.*, vol. 3, pp. 28–44, Jan. 1973.
[7] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its application to modeling and control," *IEEE Trans. Syst., Man, Cybern.*, vol. 15, pp. 116–132, Jan. 1985.
[8] R. Alcalá, J. Casillas, O. Cordón, and F. Herrera, "Approximate Mamdani-type fuzzy rule-based systems: Features and taxonomy of learning methods," University of Granada, Tech. Rep. #DECSAI-99 117, 1999. Dept. of Computer Science and A.I..
[9] B. Carse, T. C. Fogarty, and A. Munro, "Evolving fuzzy rule based controllers using genetic algorithms," *Fuzzy Sets Syst.*, vol. 80, pp. 273–293, 1996.
[10] A. Bastian, "How to handle the flexibility of linguistic variables with applications," *Int. J. Uncertainty, Fuzziness, Knowledge-Based Syst.*, vol. 2, no. 4, pp. 463–484, 1994.
[11] 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, pp. 251–270, 1997.
[12] 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.
[13] 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./Dec. 1992.
[14] H. Ishibuchi, K. Nozaki, H. Tanaka, Y. Hosaka, and M. Matsuda, "Empirical study on learning in fuzzy systems by rice analysis," *Fuzzy Sets Syst.*, vol. 64, pp. 129–144, 1994.
[15] H. Ishibuchi, K. Nozaki, and H. Tanaka, "Distributed representation of fuzzy rules and its application to pattern classification," *Fuzzy Sets Syst.*, vol. 52, pp. 21–32, 1992.
[16] O. Cordón and F. Herrera, "A two-stage evolutionary process to design TSK fuzzy rule-based systems," *IEEE Trans. Syst., Man, Cybern.—B*, vol. 29, pp. 703–715, June 1999.
[17] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, MA: Addison-Wesley, 1989.

[18] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka, "Selecting fuzzy if-then rules for classification problems using genetic algorithms," *IEEE Trans. Fuzzy Syst.*, vol. 3, pp. 260–270, Aug. 1995.
[19] O. Cordón and F. Herrera, "A three-stage evolutionary process for learning descriptive and approximative fuzzy logic controller knowledge bases from examples," *Int. J. Approximate Reasoning*, vol. 17, no. 4, pp. 369–407, 1997.
[20] F. Herrera, M. Lozano, and J. L. Verdegay, "A learning process for fuzzy control rules using genetic algorithms," *Fuzzy Sets Syst.*, vol. 100, pp. 143–158, 1998.
[21] H. Ishibuchi, T. Murata, and I. B. Turksen, "Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems," *Fuzzy Sets Syst.*, vol. 89, pp. 135–150, 1997.
[22] O. Cordón and F. Herrera, "Hybridizing genetic algorithms with sharing scheme and evolution strategies for designing approximate fuzzy rule-based systems," Fuzzy Sets Syst., to be published.
[23] L. Sánchez, "Study of the Asturias rural and urban low voltage network," Hidroeléctrica del Cantábrico Research and Development Department, Asturias, Spain, Tech. Rep., 1997. (in Spanish).
[24] 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.
[25] L. Sánchez, "Interval-valued GA-P algorithms," IEEE Trans. Evolutionary Computat., to be published.