Chapter 1

DIFFERENT PROPOSALS TO IMPROVE THE ACCURACY OF FUZZY LINGUISTIC MODELING

Oscar Cordón, Francisco Herrera
Dept. of Computer Science and A. I. - University of Granada. Granada. Spain
{ocordon,herrera}@decsai.ugr.es

María José del Jesus
Dept. of Computer Science - University of Jaen. Jaen. Spain
mjjesus@ujaen.es

Pedro Villar
Dept. of Computer Languages and Systems - University of Vigo. Ourense. Spain
pvillar@uvigo.es

Igor Zwir
Dept. of Computer Science - University of Buenos Aires. Buenos Aires. Argentina
zwir@dc.uba.ar

Abstract

Nowadays, Linguistic Modeling is considered one of the most important applications of Fuzzy Set Theory, along with Fuzzy Control. Linguistic models have the advantage of providing a human-readable description of the system modeled in the form of a set of linguistic rules. In this chapter, we will analyze several approaches to improve the accuracy of linguistic models while maintaining their descriptive power. All these approaches will share the common idea of improving the way in which the Fuzzy Rule-Based System performs interpolative reasoning by improving the cooperation between the rules in the linguistic model Knowledge Base.

Introduction

Nowadays, Linguistic Modeling (LM) is considered one of the most important applications of Fuzzy Set Theory, along with Fuzzy Control. Linguistic models have the advantage of providing a human-readable description of the system modeled in the form of a set of linguistic rules [29], which is a desirable characteristic in many modeling problems. Unfortunately, their accuracy is sometimes not as high as desired when dealing with complex modeling problems, thus causing the designer to discard them and replace them by other kinds of more accurate but less interpretable models. This drawback is due to some problems related to the inflexibility of the concept of the linguistic variable, which is the one involved in the fuzzy rule structure.

In this chapter, we review several approaches to improve the accuracy of linguistic models while maintaining their descriptive power. All these approaches will share the common idea of improving the way in which the Fuzzy Rule-Based System (FRBS) performs interpolative reasoning by improving the cooperation between the rules in the linguistic model Knowledge Base (KB).

The rule cooperation may be induced in four different FRBS components, namely the Inference System (IS), the KB as a whole and both KB components in isolation, the Data Base (DB) and the Rule Base (RB). All of them will be analyzed. To be precise, we will deal with the following aspects:

- Genetic tuning of the membership functions.
- Simulated Annealing-based Learning of the DB from examples.
- Genetic selection of fuzzy rules.
- The Accurate Linguistic Modeling paradigm, based on a double-consequent linguistic rule generation and selection.
- The Hierarchical Accurate Linguistic Modeling paradigm, based on a hierarchical linguistic rule generation and selection.
- Cooperative Fuzzy Reasoning Methods for classification problems.

The behaviour of the first five methods in solving the real-world Spanish electrical distribution problem shown in the Appendix will be analyzed. On the other hand, the performance of the last one, the only dealing with classification problems, will be tested with the IRIS and

PIMA data sets. In every experiment, the same basic rule generation process will be considered, the Wang and Mendel's one (WM-method) [30]. Two variants of this method to deal with modeling and classification problems are also introduced in the Appendix.

In order to put this into effect, this chapter is set up as follows. In Section 1., the framework is presented, i.e., System Modeling with FRBSs, structure of linguistic models and problems of LM. Section 2. describes our proposals to improve the accuracy of LM, by presenting a short study of rule cooperation in FRBSs and a brief description of the different approaches. Sections 3., 4. and 5. present the specific proposals to induce cooperation from the DB, the RB and the KB respectively. On the other hand, our proposals to induce cooperation from the IS, including their own experiments, are introduced in Section 6. Finally, a summary of the chapter is presented in 7., and an Appendix describing the WM-method and the electrical problem used as benchmark is included.

1. FRAMEWORK

In this section, some preliminary concepts will be presented. First, System Modeling with FRBSs will be introduced and the two different existing approaches will be reviewed. The section will focus then on LM and the basic structure of two different kinds of linguistic models, for regression and classification problems, will be described. Finally, the problems of LM will be analyzed.

1.1 SYSTEM MODELING WITH FRBSS

One of the most important applications of FRBSs is *system modeling* [5, 28], 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 [29]. In this kind of modeling we may usually find two contradictory requirements, the accuracy and the interpretability of the model obtained.

It is possible to distinguish between two types of modeling when working with FRBSs: *Linguistic Modeling* and *Fuzzy Modeling*, according to the fact that the main requirement is the interpretability or the accuracy of the model, respectively. The former is developed by means of descriptive Mamdani-type FRBSs, which use fuzzy rules composed of linguistic variables [34] 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 behaviour of the system being modeled [29]. On the other hand, Fuzzy Modeling is put into effect by means of approximate Mamdani-type FRBSs [3, 5, 10], systems in which the fuzzy rules are composed of fuzzy predicates without a linguistic meaning, i.e., the variables forming the rules do not take as a value a linguistic term with a fuzzy set associated defining their meaning, but a fuzzy set directly.

Therefore, a *linguistic model* is a system description in the form of a linguistic rule set interpretable by human beings, which is a desirable characteristic in some problems.

1.2 STRUCTURE OF A LINGUISTIC MODEL

The basic structure of a linguistic model [31] is showed in Fig. 1.1.

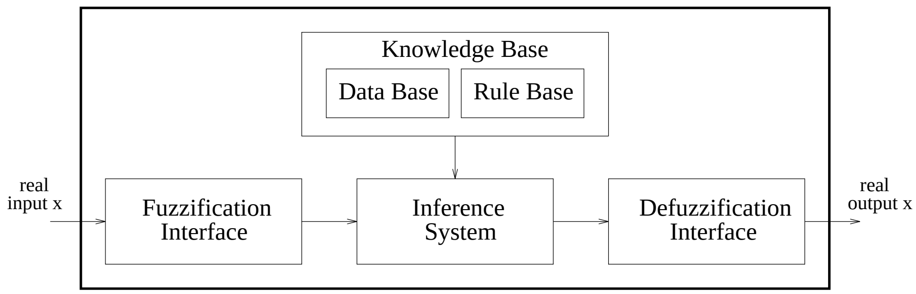

Figure 1.1 Structure of a linguistic model

The *Knowledge Base* (KB) is the component containing the knowledge about the system modeled in the form of linguistic rules. It is composed of two components:

- *Rule Base* (RB): Collection of linguistic rules:

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

  with x_j and y being linguistic system variables, and with A_{ij} and B_i being the linguistic labels associated with fuzzy sets specifying their meaning.

- *Data Base* (DB): Semantics of the linguistic labels (Fig. 1.2).

The *Fuzzification Interface* has the function of computing the firing degree of each single rule in the KB with respect to the current system input. This is done by computing the matching degree between the input and the rule antecedents, considering a conjunctive operator (a t-norm) when there is more than one input variable. The *Inference System* (IS) performs then the fuzzy reasoning process by applying the Compositional Rule of Inference [33] on each individual rule in the KB.

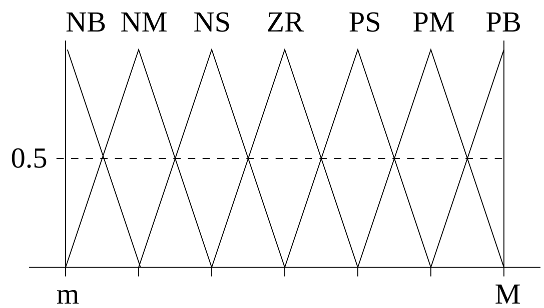

Figure 1.2 Example of semantics of the linguistic labels

The output obtained from the latter process is a number of fuzzy sets equal to the number of rules fired. The *Defuzzification Interface* works by aggregating these individual fuzzy sets in a single one and transforming it into a real number, the final output of the FRBS. For more information on the reasoning process, refer to [9].

In this chapter, the membership functions considered in the DB will always be triangular-shaped. The minimum t-norm will be used as conjuctive and implication operators, while the Center of Gravity weighted by the matching will be the defuzzification method considered [9].

We use a global error measure, the mean square error (MSE), as evaluation measure for our proposals. The MSE will allow us to determine the accuracy of the linguistic model obtained, which directly depends on the cooperation levels of the rules existing in the KB. The MSE over a training data set, E_p, is represented by the following expression:

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

where $S(ex^l)$ is the output value obtained from the FRBS when the input variable values are $ex^l = (ex_1^l, \ldots, ex_n^l)$, and ey^l is the known desired value.

1.3 FUZZY RULE-BASED CLASSIFICATION SYSTEMS

In this section, a special type of linguistic models is introduced: Fuzzy Rule-Based Classification Systems (FRBCSs), the FRBSs used for classification problems.

The structure of an FRBCS is very similar to the one of an FRBS. In an FRBCS, two components are distinguished: 1) The KB, composed of RB and DB as in every linguistic model, and 2) an a Fuzzy Reasoning

Method (FRM), an inference procedure which derives conclusions from a fuzzy rule set and an example.

The FRBCS design implies finding both components, and this process is carried out through a supervised learning process, which starts with a set of correctly classified examples (training examples) and whose ultimate objective is to design a Classification System, assigning class labels to new examples with a minimum error. Finally, the system performance on the test data is computed, to gain an estimate about the FRBCS real error.

The composition of the DB is the usual one in LM. The main difference between an usual linguistic model and an FRBCS lies on the structure of the linguistic rule considered in the latter and, more concretely, on the form of the consequent of the fuzzy classification rule. Three different fuzzy classification rules have been proposed in the specialised literature with the consequent being: a class [2, 21], a class and a certainty degree associated to the classification of that class [24], and the certainty degree associated to the classification of each one of the possible classes [26].

In this work, we will consider FRBCSs composed of RBs of the former two types:

$$R_k: \text{ If } x_1 \text{ is } A_1^k \text{ and } \ldots \text{ and } x_N \text{ is } A_N^k \text{ then } Y \text{ is } C_j$$
$$R_k: \text{ If } x_1 \text{ is } A_1^k \text{ and } \ldots \text{ and } x_N \text{ is } A_N^k \text{ then } Y \text{ is } C_j \text{ with } r^k$$

where:

- $x_1, \ldots, x_N$ are the selected features for the classification problem,
- $A_1^k, \ldots, A_N^k$ are linguistic labels used to discretise the continuous variable domain,
- Y is the class $C_j \in \{C_1, \ldots, C_M\}$ to which the example belongs, and
- r^k is the classification certainty degree in the class C_j for an example belonging to the fuzzy subspace defined by the rule antecedent.

Focusing on the FRM, the classical approach, called *maximum matching*, considers the rule with the highest association degree to make the final decision. This FRM classifies the pattern with the class of this rule. Graphically, this method could be seen as depicted in Fig. 1.3, where the rule R_k would show the highest association degree.

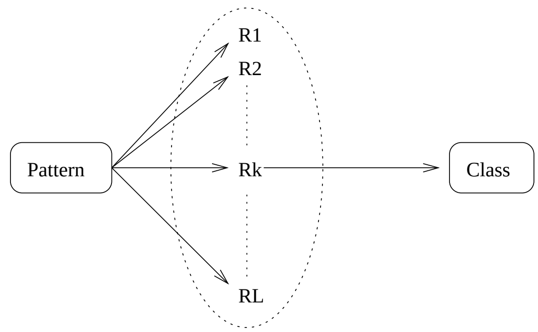

Figure 1.3 FRM that uses only the winner rule

1.4 PROBLEMS OF LINGUISTIC MODELING

As said, interpretability and accuracy are usually contradictory requeriments in System Modeling. Linguistic models present sometimes a lack of accuracy in complex modeling problems. As Zadeh pointed out in his *principle of incompatibility* [33], *"as the complexity of a system increases, our ability to make precise and yet significant statements about its behaviour diminishes ..."*. Thus, although the use of Fuzzy Logic-based techniques, specifically of FRBSs, 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.

The lack of accuracy 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 [6, 7]:

- 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 RB.

Hence, in many cases the linguistic model designed is not accurate to a sufficient degree and has to be discarded and replaced by other less interpretable but more accurate model. In this chapter, some proposals allowing us to improve the accuracy of linguistic models while maintaing their descriptive power will be introduced.

2. HOW TO IMPROVE THE ACCURACY OF LINGUISTIC MODELING

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 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 (Fig. 1.4).

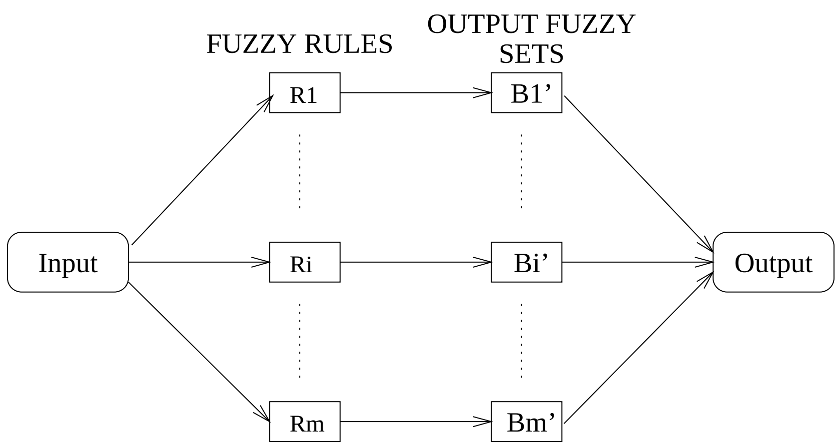

Figure 1.4 Cooperation among the fuzzy rules

Improving the cooperation among the fuzzy rules in the KB can be a good way to improve the accuracy of linguistic models. All our proposals will be based on this idea.

There are two components in an FRBS having a significant influence on the rule cooperation, the IS and the the KB. The accuracy of the FRBS can be increased, while its descriptive nature can be preserved, improving the cooperation among rules in the KB by dealing with four different aspects: the IS, the KB as a whole, and its two components, the DB and the RB, in isolation.

We propose six different approaches acting on these four different components, that are not isolated and can be combined among them:

- *Approaches to induce cooperation from the DB:*
  - Genetic tuning of the membership functions [10]
  - Simulated Annealing-based Learning of the fuzzy partition granularity [17]
- *Approaches to induce cooperation from the RB:*
  - Genetic selection of fuzzy rules [10, 19, 23]
  - Accurate Linguistic Modeling paradigm: Double-consequent linguistic rule generation and selection [12, 13]
- *Approaches to induce cooperation from the KB:*
  - Hierarchical Accurate Linguisitc Modeling Paradigm [18]
- *Approaches to induce cooperation from the IS:*
  - Cooperative Fuzzy Reasoning Methods for Classification Problems [14]

In the next sections, all of these proposals will be analyzed in depth. The specific search procedures considered will not be introduced in the chapter with the aim of not extending it excessively. The reader can refer to [20, 27] and [1] for clear and wide descriptions on Genetic Algorithms (GAs) and Simulated Annealing (SA) respectively.

3. APPROACHES TO INDUCE COOPERATION FROM THE DATA BASE

In the last few years, many approaches have been presented to automatically learn the RB from numerical information (input-output data pairs representing the system behaviour). However, there is not much information about the way to derive the DB and most of these RB learning methods need of the existence of a previous definition for it.

A common way to proceed involves considering uniform fuzzy partitions with the same number of terms (usually an odd number between three and seven) for all the linguistic variables existing in the problem. Therefore, this operation mode makes the DB have a significant influence on the FRBS performance. This is why some approaches try to improve the preliminary DB definition considered once the RB have been derived. To put this into effect, a tuning process considering the whole KB obtained (the preliminary DB and the derived RB) is used a posteriori to adjust the membership function parameters. Our first proposal to improve the accuracy of LM is to do with this idea: given a complete KB, a *genetic tuning of the membership functions*. Nevertheless, the tuning process usually only adjusts the membership functions shapes and not the number of linguistic terms in each fuzzy partition, which remains fixed from the begining of the FRBS design process. Our second proposal has a different starting point, it is a DB learning method and tries to *learn an adequate fuzzy partition granularity for each linguistic variable* using Simulated Annealing.

3.1 GENETIC TUNING OF THE MEMBERSHIP FUNCTIONS

The genetic tuning process [10] is based on the existence of a previous complete KB, that is, an initial DB definition and an RB constituted by m fuzzy rules. The chromosomes only encode the primary fuzzy partitions constituting the DB in order to adjust the linguistic labels membership functions for all the fuzzy rules contained in the RB.

The GA designed presents a real coding issue that allows us to maintain the FRBS descriptive nature. Each chromosome encodes a different DB definition. A primary fuzzy partition is represented as an array composed by $3 \cdot N$ real values, with N being the number of terms forming the linguistic variable term set. The complete DB for a problem, in which m linguistic variables are involved, is encoded into a fixed length real coded chromosome C_r built by joining the partial representations of each one of the variable fuzzy partitions as is shown in the following:

$$
\begin{gathered}
(a_j, b_j, c_j) \rightarrow \text{ 3-tuple encoding of the fuzzy set } j \\
C_{ri} = (a_{i1}, b_{i1}, c_{i1}, \ldots, a_{iN_i}, b_{iN_i}, c_{iN_i}) \rightarrow \text{ fuzzy partition} \\
\text{of the linguistic variable } i. \\
C_r = C_{r1}\ C_{r2}\ \ldots\ C_{rm} \rightarrow \text{whole DB definition.}
\end{gathered}
$$

The initial gene pool is created making use of the initial DB definition. This one is encoded directly into a chromosome, denoted as C_1. The remaining individuals are generated by associating an interval of performance, $[c_h^l, c_h^r]$ to every gene c_h in C_1, $h = 1 \ldots \sum_{i=1}^{m} N_i \cdot 3$. Each interval of performance will be the interval of adjustment for the corresponding gene, $c_h \in [c_h^l, c_h^r]$.

If (t mod 3) = 1 then c_t is the left value of the support of a fuzzy number. The fuzzy number is defined by the three parameters (c_t, c_{t+1}, c_{t+2}) and the intervals of performance are the following:

$$c_t \in [c_t^l, c_t^r] = [c_t - \frac{c_{t+1}-c_t}{2}, c_t + \frac{c_{t+1}-c_t}{2}]$$

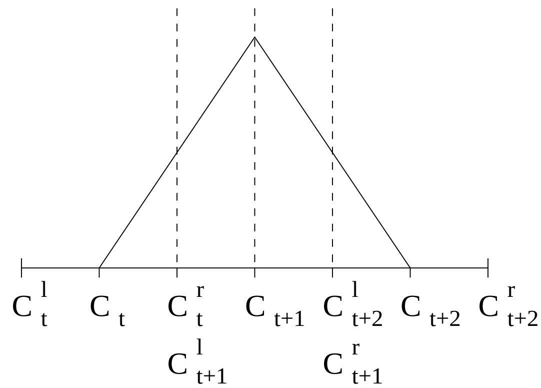

Figure 1.5 Membership function and intervals of performance for the tuning process

$$c_{t+1} \in [c^l_{t+1}, c^r_{t+1}] = [c_{t+1} - \frac{c_{t+1} - c_t}{2}, c_{t+1} + \frac{c_{t+2} - c_{t+1}}{2}]$$
$$c_{t+2} \in [c^l_{t+2}, c^r_{t+2}] = [c_{t+2} - \frac{c_{t+2} - c_{t+1}}{2}, c_{t+2} + \frac{c_{t+3} - c_{t+2}}{2}]$$

Fig. 1.5 shows these intervals. Therefore, we create a population of chromosomes containing C_1 as its first individual and the remaining ones initiated randomly, with each gene being in its respective interval of performance.

The GA designed uses the stochastic universal sampling as selection procedure together with an elitist scheme. The operators employed for performing the individual recombination and mutation are Michalewicz's non-uniform mutation [27] and the max-min-arithmetical crossover [22]. The MSE introduced in Section 1.2 plus a criterion penalyzing the lack of the completeness property compose the fitness function. Further information about this approach can be found in [10].

Table 1.1 collects the results of a brief experimentation where the KB obtained from the WM-method for the electrical problem shown in the Appendix is refined by means of the introduced tuning process. The large accuracy improvement can be clearly seen.

Table 1.1 Results obtained with the genetic tuning process

Method	*Granularity*	# Rules	$MSE_{training}$	MSE_{test}
WM-method	7 7 7	24	222622	240018
DB Tuning	7 7 7	24	144510	173167

3.2 LEARNING AN ADEQUATE FUZZY PARTITION GRANULARITY

As said, most of RB learning methods needs a previous definition of the DB, and the consideration of uniform fuzzy partitions with the same number of terms significatively affects the linguistic model accuracy. To solve this problem, we have developed a method to learn a good fuzzy partition granularity for a determinated problem [17]. We try to learn the number of linguistic terms for each variable, maintaining uniform fuzzy partitions. Since the exhaustive exploration of the search space is a very time consuming task, we consider the SA heuristic local search technique to perform the search.

In our case, given an RB generating method and an specific problem, each candidate solution is a concrete granularity level for each problem variable (number of labels), and the cost function is based on the MSE of the FRBS obtained with the WM-method using a DB with that granularity.

Three stopping criteria have been considered in order to reduce the run time of the procedure:

- The maximum number of iterations allowed without global improvement is reached.
- No solution was accepted in the last iteration.
- The maximum number of solutions have been generated.

It is interesting to point out that in all the runs done in [17] the procedure finished due to the first or second stopping criteria.

The basic operation mode of SA, adapted to our problem, is described in the next algorithm, with L being the number of possible values for the labels (seven in our case, $\{3, \ldots, 9\}$), with N being the number of problem variables, with α being the decreasing factor of the temperature, and with T_o, T being respectively the initial temperature, and the temperature in successive iterations.

SA (T_0, α, N, L):
  $T \leftarrow T_0$;
  $S_{act} \leftarrow Generate\_Initial\_Solution$;
  $S_{best} \leftarrow S_{act}$;
  $while\,(solutions \leq L^N)$ and $(iterations\_without\_improv. < N)$
   and $not(iteration\_without\_accepted\_solution)$ do
    $begin$
     $count \leftarrow 0$

$while\,(count < N^3)\ and\ (accepted\_solution\_number < N^2)\ do$
$begin$
$S_{cand} \leftarrow Generate\_Solution\_N(S_{act});$
$\delta \leftarrow cost(S_{cand}) - cost(S_{act});$
$if\ (U(0,1) < e^{(-\delta/T)})\ or\ (\delta < 0)$
$then\ S_{act} \leftarrow S_{cand};$
$if\ cost(S_{act}) < cost(S_{best})$
$then\ S_{best} \leftarrow S_{act};$
$count \leftarrow count + 1;$
$end;$
$T \leftarrow \alpha(T);$
$end;$
{Write as final solution, S_{best} }

The implementation of our SA procedure incorporates a taboo record of explored solutions, along with their cost, in order to eliminate the possibility of redundant executions of the RB generating method, with the consequent saving of run time. In fact, only 32 of the 59 solutions generated in the experiment developed in this chapter were evaluated. The results obtained are shown in Table 1.2, where the linguistic model generated by means of the granularity learning process can be compared with the one generated by the WM-method when considered the same number of labels (seven) for the three problem variables.

For more details about the SA procedure used and a wider experimentation, refer to [17].

Table 1.2 Results obtained with the SA-based fuzzy partition granularity learning

Method	*Granularity*	# Rules	$MSE_{training}$	MSE_{test}
WM-method	7 7 7	24	222622	240018
SA-based granularity learning	8 9 9	28	192980	230675

4. APPROACHES TO INDUCE COOPERATION FROM THE RULE BASE

In this section, two methods to improve the FRBS performance by increasing the cooperation among the rules belonging to the RB are proposed: a *genetic selection process of fuzzy rules* and the *Accurate Linguistic Modeling paradigm*, based on double-consequent fuzzy rules.

4.1 GENETIC SELECTION OF FUZZY RULES

The operation mode of many RB generation methods means that, in each input subspace, the rules are created individually from the examples in the input-output data set without taking into account the cooperation existing between them to give the final model output. That is, no information about the neighbour rules is considered in order to generate them. Because of this, the generated RB may present redundant or unnecessary rules making the model using this KB less accurate. In order to avoid this fact, a rule selection genetic process is proposed in [10, 23] with the aim of simplifying the initial linguistic rule set by removing the unnecessary rules from it and generating a KB with good cooperation.

The selection of the subset of linguistic rules best cooperating is a combinatorial optimization problem. 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. Another process solving the problem of selecting rules by means of the same technique is to be found in [25].

The 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.

The coding scheme generates fixed-length chromosomes. Considering the rules contained in the initial linguistic rule set counted from 1 to m, an m-bit string $C = (c_1, ..., c_m)$ represents a subset of candidate rules to form the RB finally obtained, B^s, such that,

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

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 the MSE of the FRBS using the RB encoded in the chromosome over the training data set as well as a criterion penalyzing the lack of the completeness property of the said RB.

A possible improvement of this method is the *genetic multiselection process* [19], which obtains different simplified RBs for modeling and classification problems. It selects the rules cooperating best from the previous RB, by working as follows:

- The basic rule selection genetic process is run several times.
- Each time a simplified rule set is generated, the space zone where it is located is penalized by means of a genotypic sharing function (niching GAs [20]).
- The process ends when the desired number of simplified RBs is generated.

Results for this proposal will not be presented in this chapter, since the WM-method generates a small rule set which does not verify the completeness property. For some results obtained when applying the genetic selection process to RBs generated from other learning methods, refer to [10, 19, 23]. On the other hand, we will see that the genetic selection process is considered in the other approach proposed to induce cooperation from the RB (next subsection) and on the one presented to induce it from the whole KB (Section 5.).

4.2 THE ACCURATE LINGUISTIC MODELING PARADIGM

The *Accurate Linguistic Modeling* (ALM) [12, 13] is a methodology to obtain more cooperative RBs for linguistic models. It is based on the following two aspects:

- The usual linguistic model structure is extended allowing the RB to present rules where each combination of antecedents may have two consequents (the primary and secondary in importance) associated when it is necessary to improve the model accuracy. It is clear that this will improve the capability of the model to perform the interpolative reasoning and, thus, its performance.

  We should note that this operation mode does not constitute an inconsistency from the interpolative reasoning point of view but only a shift of the main labels making that the final output of the rule lie in an intermediate zone between them both. Hence, it may have the following linguistic interpretation. Let us suppose that a specific combination of antecedents, "x_1 is A_1 and ... and x_n is A_n", has two different consequents associated, B_1 and B_2. From a LM point of view, the resulting double-consequent rule may be interpreted as follows:

  IF x_1 is A_1 and ... and x_n is A_n THEN y is *between* B_1 *and* B_2

- The previous point deals with the improvement of the fuzzy reasoning in an input subspace defined by a specific combination of antecedents. On the other hand, the second aspect deals with the cooperation between the rules in the KB, i.e., with the overlapped space zones that are covered by different linguistic rules. Hence, it is considered an operation mode based on generating a preliminary fuzzy rule set in which single and double-consequent rules coexist and selecting the subset of them best cooperating. It is important to remark that each double-consequent rule is decomposed in two simple ones in the selection process. Thus, this stage will specify which double-consequent rules in the preliminary rule set will remain in the final RB, that is, those fuzzy input subspaces whose **two simple rules associated have been finally selected**.

On the other hand, it should be noted that the said operation mode gives more freedom to the RB generation process. As is known, the generation of the best fuzzy rule in each subspace does not ensure that the FRBS designed will perform well, due to the fact that the rules composing the KB may not cooperate suitably. The rule selection considered in ALM 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 RB.

In [12], two specific generation processes based on the ALM methodology are introduced. Both of them are based on two stages: double-consequent rule generation and rule selection. In the following, one of these processes is briefly described:

1. *A linguistic rule generation method* from examples based on a modification of the WM-method that involves generating the two most important consequents for each combination of antecedents (instead of only the most important one, as this method usually do). All the WM-method steps shown in the Appendix remains the same but the fourth one (*Obtain a final RB from the preliminary fuzzy rule set*). Whilst 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, thus creating a double-consequent rule.

   Of course, a combination of antecedents may not have rules 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 rule selection genetic process*, introduced in Section 4.1, that selects the subset of rules in the preliminary linguistic set cooperating best working in the said way.

Another important characteristic of ALM is that it has no influence on the linguistic model inference system. The only restricition imposed is that the defuzzification method must consider the matching degree of the rules fired. In this chapter we work with the *Center of Gravity weighted by the matching degree* [9].

The inference mechanism designed will perform in the way shown next when it receives an input $x_0 = (x_1, \ldots, x_n)$:

1. *For each rule R_i, $i = 1, \ldots, T$, in the KB*:

   (a) Compute the matching degree, h_i, of the rule:

   $$h_i = Min(\mu_{A_{i1}}(x_1), \ldots, \mu_{A_{in}}(x_n))$$

   (b) Apply the Minimum t-norm in the role of implication operator to obtain the fuzzy set resulting from the application of the inference process on that rule, B'_i:

   $$\mu_{B'_i}(y) = Min(h_i, \mu_{B_i}(y))$$

2. *Obtain the Center of Gravity for each individual fuzzy set B'_i*:

$$y_i = \frac{\int_{\mathbf{V}} y \cdot \mu_{B'_i}(y) \cdot dy}{\int_{\mathbf{V}} \mu_{B'_i}(y) \cdot dy}$$

3. *Compute the final output given by the system as output, y_0, by aggregating the partial actions obtained by means of the matching degree weighted average*:

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

The results obtained by the ALM-based process proposed in the solving of the electrical application tackled are showed in Table 1.3. In order to analyze the influence of the genetic selection process introduced in Section 4.1, two different rows will be associated to the ALM process in the table, each one collecting the results obtained after the application of each stage composing it. As can be seen, the linguistic model obtained is simpler and more accurate to a high degree than the WM-method one. We should note that the number of rules showed (20) stands for simple rules, i.e., the double-consequent rules existing in the RB have been counted twice for comparison purposes.

Table 1.3 Results obtained with ALM

Method	*Granularity*	# Rules	$MSE_{training}$	MSE_{test}
WM-method	7 7 7	24	222622	240018
ALM (generation)	7 7 7	34	231174	260067
ALM (selection)	7 7 7	20	155866	178601

5. APPROACHES TO INDUCE COOPERATION FROM THE KNOWLEDGE BASE

A single method will be introduced belonging to the group of approaches inducing cooperation from the whole KB (i.e., both from the DB and the RB), the *Hierarchical System of Linguistic Rules (HSLR) learning methodology*. In HSLRs, the linguistic variables involved in the fuzzy rules are defined in linguistic partitions with different granularity levels, thus making the rules belong to different hierarchical levels [18].

To do so, the KB structure of linguistic models is extended by introducing the concept of "*layers*". In this extension, which is also a generalization, the KB is composed of a set of layers where each one contains linguistic partitions with different granularity levels and linguistic rules whose linguistic variables take values in these partitions. This KB is called Hierarchical Knowledge Base (HKB), and it is formed by a Hierarchical Data Base (HDB) and a Hierarchical Rule Base (HRB), containing linguistic partitions of the said type and linguistic rules defined over them, respectively.

The description of the HKB and the relation between its components is studied next, and the methodology to automatically design an HSLR from generic linguistic rule generating methods is introduced later on. For more details about HSLR methodology, refer to [18].

5.1 HIERARCHICAL KNOWLEDGE BASE

This HKB is composed of a set of layers. We define a layer by its components in the following way:

$$layer(t, n) = DB(t, n) + RB(t, n)$$

with $DB(t, n)$ being the DB which contains the linguistic partitions with granularity level n of layer t, and with $RB(t, n)$ being the RB formed by those linguistic rules whose linguistic variables take values in the former partitions. From now on and for the sake of simplicity, we are going to refer to the components of a $DB(t,n)$ and $RB(t,n)$ as *n-linguistic partitions* and *n-linguistic rules*, respectively.

This set of layers is organized as a hierarchy, where the order is given by the granularity level of the linguistic partition defined in each layer. That is, given two succesive layers t and $t+1$, then the granularity level of the linguistic partitions of layer $t+1$ is greater than the ones of layer t. This causes a refinement of the previous layer linguistic partitions.

As a consequence of the previous definitions, we could now define the HDB as the union of the DBs of every layer t:

$$HDB = \cup_t DB(t,n)$$

and by the same token, the HRB is defined as:

$$HRB = \cup_t RB(t,n)$$

Focusing again on the HDB, we should note that, in this work, we are using *n-linguistic partitions* with the same number of linguistic terms for all input-output variables, composed of triangular-shaped, symetrical and uniformly distributed membership functions.

In order to build the HDB, we develop an strategy which satisfies two main requirements:

- To preserve all possible fuzzy set structures from one layer to the next in the hierarchy.
- To make smooth transitions between successive layers.

Hence, to build a new linguistic partition in the DB of the layer $t+1$ from a *n-linguistic partition* of the layer t with the minimum change between their granularity levels, we just add a new linguistic term between each two consecutive terms of the *n-linguistic partition*, after reducing the support of these linguistic terms in order to keep place for the new one, which is located in the middle of them. An example of the correspondence between a *3-linguistic partition* and *5-linguistic partition* is shown in Fig. 1.6.

Generically, we could say that a DB from a layer $t+1$ is obtained as:

$$DB(t,n) \rightarrow DB(t+1, 2 \cdot n - 1)$$

which means that an n-linguistic *partition* in $DB(t,n)$ with n linguistic terms becomes a *($2{\cdot}n-1$)-linguistic partition* in $DB(t+1, 2 \cdot n - 1)$.

As regards the HRB, the *n-linguistic rules* contained in $RB(t,n)$ are those rules whose linguistic variables take values from the *n-linguistic partitions* contained in $DB(t,n)$.

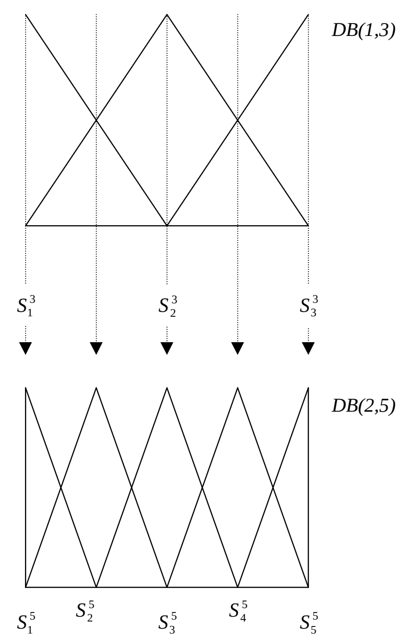

Figure 1.6 Transition from a partition in $DB(1,3)$ to another one in $DB(2,5)$

The main purpose of developing an HRB is to model the space of the problem in a more accurate way. To do so, those *n-linguistic rules* that model a subspace with bad performance are expanded in a set of *(2·n−1)-linguistic rules*, which become their image in $RB(t+1, 2\cdot n-1)$. This set of rules models the same subspace that the former one and replaces it.

We should note that not all *n-linguistic rules* are to be expanded. Only those *n-linguistic rules* which model a subspace of the problem with a significant error become the ones that are involved in this rule expansion process to build the $RB(t+1, 2\cdot n-1)$. The remaining rules preserve their location in $RB(t, n)$.

An explanation for this behaviour could be found in the fact that it is not always true that a set of rules with a higher granularity level perform a better modeling of a problem than other set composed of linguistic rules with a lower granularity level. Moreover, this is not true for all kinds of problems, and what is more, it is also not true for all linguistic rules that model a problem [16]. In an attempt to put this idea into effect, we consider a three-stage process to perform the mentioned rule expansion:

- Selection of those bad performance *n-linguistic rules* from $RB(t, n)$ that are going to be expanded in $RB(t+1, 2 \cdot n - 1)$.
- Selection of those terms from $DB(t+1, 2 \cdot n - 1)$ that are going to be contained in the *(2·n − 1)-linguistic rules,* considered as an image of the bad rules.
- Accomplishment of the *(2·n − 1)-linguistic rule* generation process, based on the previously selected term sets.

5.2 STRUCTURE OF THE HSLR LEARNING METHODOLOGY

Our HSLR learning methodology is composed of three main processes which will be described in detail in the following subsections [18]:

- The first process generates the HKB following the descriptions given previously.
- The second process performs a genetic rule selection task that removes the redundant or unnecessary rules from the HRB in order to select a subset of rules that cooperate better.
- In the third process, a user evaluation process extends this approach to an iterative process, where he could adapt many parameters and re-execute the processes to achieve better results.

5.3 HIERARCHICAL KNOWLEDGE BASE GENERATION PROCESS

In this subsection we present our metodology to generate an HKB. It is based on an inductive linguistic rule generation method (LRG-method), that in this chapter will be the WM-method. It also takes as a base a set of input-output data E_p and a previously defined $DB(t, n)$.

Our HKB generation process has three main steps, that are listed below:

1. **RB(t,n) generation process,** where the rules from the present $DB(t, n)$ are generated.

   An LRG-method is run with the terms defined in the present partitions, that are in $DB(t, n)$, denoted as $LRG(DB(t, n), E_p)$.

2. **RB(t+1,2·n-1) generation process,** where the linguistic rules from layer t+1 are generated taking into account $RB(t,n)$, $DB(t,n)$ and $DB(t+1, 2\cdot n-1)$.

   (a) *Calculate the error of RB(t,n)*: Compute $MSE(E_p, RB(t,n))$

   (b) *Calculate the error of each individual n-linguistic rule*: Compute $MSE(E_i, R_i^n)$.

   (c) *Select the set of n-linguistic rules with bad performance*: Select those bad *n-linguistic rules* which are going to be expanded:
   $IF\ MSE(E_i, R_i^n) \geq \alpha \cdot MSE(E_p, RB(t,n))\ THEN\ R_i^n \in RB_{bad}(t,n)\ ELSE\ R_i^n \in RB_{good}(t,n)$
   For example, $\alpha = 1.1$ means that an *n-linguistic rule* with an MSE a 10 % higher than the MSE of the entire $RB(t,n)$ should be expanded.

   (d) *Obtain the DB(t+1,2·n−1):* create DB_{x_j}*(t+1,2·n−1)* for all input linguistic variables x_j ($j = 1,...,m$), and DB_y*(t+1,2·n−1)* for the output linguistic variable y.

   (e) *Select the (2·n − 1)-linguistic partition terms:* Obtain those terms in DB(t+1,2·n − 1) that are considered in the fuzzy input and output subspaces of the bad rules that are to be expanded: $I(R_i^n), \forall\, R_i^n \in RB_{bad}(t,n)$.

   (f) *Combine the selected (2·n − 1)-linguistic partition terms to perform (2·n − 1)-linguistic rules:* For each $R_i^n \in RB_{bad}(t,n)$, compute $LRG(I(R_i^n), E_i)$.

3. **HRB summarization process,** where the linguistic rules from the both RBs are joined to obtain the HRB.

   Obtain a set of linguistic rules, joining the group of the new generated (2·n − 1)-linguistic rules and the former good performance n-linguistic rules:

   $$HRB = RB_{good}(t,n) \cup RB(t+1, 2\cdot n-1)$$

5.4 HIERARCHICAL RULE BASE SELECTION PROCESS

As has been seen in previous sections, the operation mode of the proposed generation method means that in each input subspace, the *n-linguistic rules* are created individually from the examples in the input-output data set. This happens without taking into account the cooperation existing among the rules which gives the final model output. That is, no information about the neighbor rules is considered in order to generate them.

In the HRB -where there are coexisting rules with different granularity levels- it may happen that a complete set of *(2·n − 1)-linguistic rules,* which replaces an expanded rule, does not produce good results. This means that there will be higher errors. However, a subset of this set of *(2·n − 1)-linguistic rules* may work properly, with less rules that have good cooperation between them. Thus, the HRB generated may present redundant or unnecessary rules making the model using this HKB less accurate.

In order to avoid this fact, we will use the linguistic rule selection genetic process described in Section 4.1. with the aim of simplifying the initial linguistic rule set by removing the unnecessary rules from it and generating an HKB with good cooperation.

5.5 USER EVALUATION PROCESS

The application of our metodology could be also considered as an user controlled iterative process. In this sense, the user could adapt the granularity of the initial linguistic partitions and/or the threshold α which determines if an *n-linguistic rule* will be expanded in a set of *(2·n − 1)-linguistic rules*, and apply again the process in order to obtain a better model.

This process works in this way: if the error measure of the obtained model (i.e. global error) does not satisfy the user requirements, then the user can adapt the parameter α –item 2.c in the HKB Generation Process– and/or reinitialize the process with a different granularity level for the initial partition of the linguistic variables domain.

5.6 EXPERIMENTAL STUDY

The results obtained by applying the HSLR learning process introduced in this section to the electrical problem are shown in Table 1.4. For a complete experimentation to solve this and other problems, refer to [18]. In view of the data showed in the table, it can be seen that the model generated performs significantly better than the WM-method one, with a short intrepretability lose (28 rules in the KB instead of 24).

Table 1.4 Results obtained with the HSLR methodology

Method	*Granularity*	# Rules	$MSE_{training}$	MSE_{test}
WM-method	7 7 7	24	222622	240018
granularity learning	(WM,5,9)	28	157755	180488

6. APPROACHES TO INDUCE COOPERATION FROM THE INFERENCE SYSTEM

In this section, we propose a new fuzzy reasoning model for a concrete type of linguistic models: FRBCSs, the FRBSs for classification problems. Some specific FRMs included in the general model will be also introduced with the aim of improving the rule cooperation in these kinds of systems.

6.1 IMPROVING THE RULE COOPERATION IN FRBCSS

By using the classical FRBCS reasoning method shown in Section 1.3, the information provided by the other rules that also are compatible (have also been fired) with the example is not considered. In this section, we propose to use FRMs that combine the information given by the different rules fired by a pattern.

To do so, a general reasoning model for FRBCSs [14] is introduced, that in this paper is particularised to an RB composed of rules with a class and its certainty degree in the consequent. This model is described in the following.

In the classification of an example $E^t = (e_1^t, \ldots, e_N^t)$, the RB $R = \{R_1, \ldots, R_L\}$ is divided into M subsets according to the class indicated by its consequent,

$$R = R_{C_1} \cup R_{C_2} \cup \ldots \cup R_{C_M}$$

and the next scheme is followed:

1. **Compatibility degree.** The compatibility degree of the antecedent with the example is computed for all the rules in the RB, applying a t-norm over the membership degree of the values of the example (e_i^t) to the corresponding fuzzy subsets.

$$R^k(E^t) = T(\mu_{A_1^k}(e_1^t), \ldots, \mu_{A_N^k}(e_N^t)), \quad k = 1, \ldots, L$$

2. **Association degree.** The association degree of the example E^t with the M classes is computed according to each rule in the RB.

$$b_i^k = h(R^k(E^t), r^k), \quad k = 1, \ldots, |R_{C_i}| \quad i = 1, \ldots, M$$

3. **Weighting function.** The values obtained are weighted by means of a function g. An expression which promotes the highest values and penalizes the smallest ones seems to be the most adequate choice for this function.

$$B_i^k = g(b_i^k), \quad k = 1, \ \ldots, |R_{C_i}| \quad i = 1, \ldots, M$$

4. **Pattern classification soundness degree for all classes.** To compute this value, an aggregation operator is used which combines, for each class, the positive association degrees computed in the previous step

$$Y_i = f(B_i^k, \ k = 1, \ldots, |R_{C_i}| \ \text{ and } \ B_i^k > 0), \quad i = 1, \ldots, M$$

with f being an aggregation operator that returns a value between the minimum and the maximum.

5. **Classification.** A decision function F is applied to the classification degrees of the example. This function will return the class label corresponding to the maximum value.

$$C_l = F(Y_1, \ldots, Y_M) \ \text{ such that } \ Y_l = \max_{j=1,\ldots,M} Y_j$$

We should note that, in this general model, if we select the function f in the fourth step as the maximum operator, we have the classical FRM:

$$f_0(a_1, \ldots, a_s) = \max_{i=1,\ldots,s} a_i$$

with $a_1, \ldots, a_s$ being the values to aggregate for an example E^t with respect to a class C_j.

According to the general reasoning model, we propose a new kind of inference models. The difference lies on the choice of function $f(\cdot)$ in step 4, due to the fact that we consider FRMs that integrate all fuzzy rules to derive conclusions from a set of fuzzy classification rules and a pattern. This idea is graphically represented in Fig. 1.7.

Some proposals for the function f in FRBCSs belonging to this family are described in Table 1.5. An analysis of them, as well as a review of previous applications of the first function f_1, is to be found in [14].

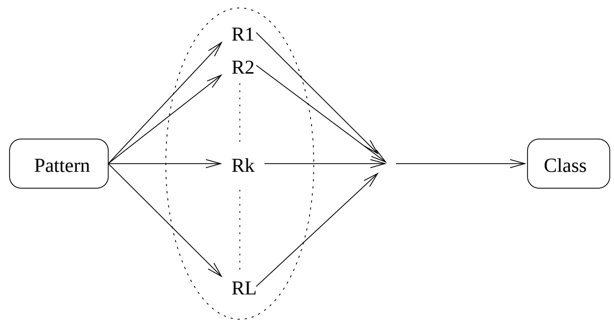

Figure 1.7 FRM integrating all fuzzy rules

Table 1.5 Different proposals for the aggregation function f in step 4

Normalized Sum $f_1(a_1,\ldots,a_s) = \frac{\sum_{i=1}^s a_i}{f_{1_{max}}}$ $f_{1_{max}} = \max_{j=1,\ldots,M} \sum_{i=1}^s a_i$	**Sowa and-like** $f_4(a_1,\ldots,a_s) = \alpha \cdot a_{min} + (1-\alpha)\frac{1}{s}\sum_{i=1}^s a_i$ $\alpha \in [0,1], \quad a_{min} = min\{a_1,\ldots,a_s\}$
Arithmetic Mean $f_2(a_1,\ldots,a_s) = \frac{\sum_{i=1}^s a_i}{s}$	**Sowa or-like** $f_5(a_1,\ldots,a_s) = \alpha \cdot a_{max} + (1-\alpha)\frac{1}{s}\sum_{i=1}^s a_i$ $\alpha \in [0,1], \quad a_{max} = max\{a_1,\ldots,a_s\}$
Quasiarithmetic Mean $f_3(a_1,\ldots,a_s) = H^{-1}\left[\frac{1}{s}\sum_{i=1}^s H(a_i)\right]$ $H(x) = x^p, \quad p \in \mathrm{R}$	**Badd** $f_6(a_1,\ldots,a_s) = \frac{\sum_{i=1}^s a_i^{\alpha+1}}{\sum_{i=1}^s a_i^{\alpha}}, \quad \alpha \in \mathrm{R}$
with $a_1,\ldots,a_s$ being the values to aggregate for pattern E^t with respect to a class C_j	

6.2 EXPERIMENTAL STUDY

For this brief study, two well known sets of samples, IRIS and PIMA, have been considered. The IRIS data base is a set of 150 examples of iris flowers with three classes and four attributes. PIMA is a set of 768 solved cases of diagnostics of diabetes where eight variables are taken into account and there are two possible classes (having or not having the disease).

Taking into account the characteristics of the example sets, fuzzy partitions constituted by five triangular fuzzy sets have been considered to define the DB in both cases. As regards the RB, two different kinds of fuzzy classification rules have been considered for the experimentation, the ones with only a class and with a class and its certainty degree in the consequent. The RB has been generated by means of the adaption of the WM-method to classification problems shown in the Appendix in both cases.

To calculate an error estimation of an FRBCS, random resampling [32] with five random partitions of the sample bases in training and test sets (70% and 30% respectively) have been considered.

The best results obtained with our different proposals of FRMs are showed in Tables 1.6 and 1.7 for the two types of rules considered. The classification percentages obtained by the classical reasoning method are also shown for comparison purposes. For a complete experimental study including all the FRMs proposed and comparing against different classification techniques, refer to [14].

Table 1.6 Results obtained when using rules with a class in the consequent

Iris			**Pima**		
FRM	Tra	Test	FRM	Tra	Test
Classical (f_0)	90.97	88.25	Classical (f_0)	89.51	64.88
$f_1\ g_2$	98.56	**94.38**	$f_1\ g_1$	83.97	**72.11**
$f_5\ g_2$	90.64	92.27	$f_6\ g_2$	91.32	67.38
$f_3\ g_2$	89.73	91.78	$f_5\ g_2$	85.78	64.56

Table 1.7 Results obtained when using rules with a class and a certainty degree in the consequent

Iris			**Pima**		
FRM	Tra	Test	FRM	Tra	Test
Classical (f_0)	97.31	94.32	Classical (f_0)	85.81	73.23
$f_3\ g_1$	97.31	94.32	$f_6\ g_2$	85.89	**73.53**
$f_5\ g_1$	97.31	94.32	$f_6\ g_1$	85.85	**73.53**
$f_6\ g_1$	97.31	94.32	$f_3\ g_2$	85.78	73.44

In view of the results obtained, the cooperative FRMs have demonstrated a good behaviour with RBs obtained from rule generation processes not considering them. However, it is possible to obtain better results by including them in the FRBCS learning process. To do so, a three-stage genetic fuzzy rule-based classification system considering the rule cooperation induction during the learning stage was introduced in [11].

7. SUMMARY

This chapter has been devoted to the problem of improving the accuracy of linguistic models while maintaining its descriptive power. As shown, linguistic models are human-readable rule-based descriptions of the system modeled but sometimes they are not as accurate as desired due to some problems of the linguistic rule structure.

With this aim, the possibility of improving the way in which the linguistic model performs interpolative reasoning by improving the cooperation between the rules in the KB has been analyzed. Several approaches to do so have been introduced, classified in four different groups according to the FRBS component from which the cooperation is induced: IS, KB, DB and RB. To be precise, the following six approaches have been studied:

- Genetic tuning of the membership functions (DB).
- SA-based Learning of the DB from examples (DB).
- Genetic selection of fuzzy rules (RB).
- The ALM paradigm, based on a double-consequent linguistic rule generation and selection (RB).
- The HALM paradigm, based on a hierarchical linguistic rule generation and selection (KB).
- Cooperative FRMs for classification problems (IS).

The behaviour of the first five has been analyzed in solving a real-world Spanish electrical distribution problem, where all of them have obtained good results, being more accurate than the basic linguistic model generated from the WM-method. On the other hand, the last one has shown good performance with the classical IRIS and PIMA data sets.

As mentioned in the chapter, the different approaches proposed are not mutually exclusive and can be combined to obtain better linguistic models. In fact, the ALM and HALM paradigms make use of other of the approaches introduced, the rule selection genetic process, as one of their components. This fact makes us think that the combination of the different approaches can be a promising research field and our future work will be focused on studying it.

Acknowledgments

This research has been suported by the "Ministerio de Educación y Ciencia" of Spain under project **PB98-1319**.

Appendix: The Wang and Mendel Rule Generation Method

The Wang and Mendel's RB generation method (WM-method) [30] is one of the simplest and most known LM design methods (for more information about the different learning techniques considered for this task and some specific approaches, refer to [4]). In this inductive method, the generation of the RB is put into effect by means of the following steps:

1. *Consider a fuzzy partition of the input variable spaces*: It may be obtained from the expert information (if it is availaible) or by a normalization process. If the latter is the case, perform a fuzzy partition of the input variable spaces dividing each universe of discourse into a number of equal or unequal partitions, select a kind of membership function and assign one fuzzy set to each subspace.

2. *Generate a preliminary linguistic rule set*: This set will be formed by the rule best covering each example (input-ouput data pair) contained in the tarining data set. The structure of these rules is obtained by taking a specific example, i.e., an $n+1$-dimensional real array (n input and 1 output values) and setting each one of the rule variables to the linguistic label associated to the fuzzy set best covering every array component.

3. *Give an importance degree to each rule*: Let $R_l =$ *IF* x_1 *is* A_1 *and* $\ldots$ *and* x_n *is* A_n *THEN* y *is* B be the linguistic rule generated from the example $e_l = (x_1^l, \ldots, x_n^l, y^l)$. The importance degree associated to it will be obtained as follows:

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

4. *Obtain a final RB from the preliminary fuzzy rule set*: The rule with the highest importance degree is chosen for each combination of antecedents.

Appendix: The Wang and Mendel Rule Generation Method for Classification

In [8], an extension of the WM-method was proposed to deal with classification problems. This process starts with a set of input-output data pairs (the training data set) with the following structure:

$$E^1 = (e_1^1, \ldots, e_N^1, o^1),\ E^2 = (e_1^2, \ldots, e_N^2, o^2),\ \ldots,\ E^p = (e_1^p, \ldots, e_N^p, o^p)$$

Table 1.C.1 Variables of the electrical problem

Symbol	Meaning
x_1	Number of inhabitants of the town
x_2	Distance from the center of the town to the three furthest clients
y	Total length of low voltage line installed

where o^h is the class label for the pattern E^h.

The task here is to generate a set of linguistic classification rules from the training data set that describes the relationship between the system variables and determines a mapping D between the feature space S^N and the class set $C = \{C_1, \ldots, C_M\}$.

The method consists of the following steps:

- *Fuzzifying the feature space.* Finding the domain intervals of the attributes and partition each domain into X_i regions ($i = 1, \ldots, N$). A membership function is adopted for each fuzzy region.
- *Generating fuzzy rules from given data pairs.* For each training data $E^h = (e_1^h, \ldots, e_N^h, o^h)$, we have
  - To determine the membership degrees of e_i^h in different input fuzzy subsets.
  - To assign the input $e_1^h, \ldots, e_N^h$ to the region with the maximum membership degree.
  - To produce a fuzzy rule from E^h, with the if-part that represents the selected fuzzy region and the consequent with the class determined by o^h. Repeated fuzzy rules are not considered.

Appendix: Total low voltage line length installed in a rural town

The problem considered is that of finding a model that relates the total length of low voltage line installed in Spanish rural towns [15]. This model will be used to estimate the total length of line being maintained by an electrical company. We were provided with a sample of 495 towns in which the length of line was actually measured and the company used the model to extrapolate this length over more than 10.000 towns with these properties. We will limit ourselves to the estimation of the length of line in a town, given the inputs showed in Table 1.C.1. To develop the different experiments in this chapter, the sample has been randomly divided in two subsets, the training and test ones, with an 80%-20% of the original size respectively. Thus, the training set contains 396 elements, whilst the test one is composed by 99 elements.

References

[1] Aarts, E.H.L., *Simulated Annealing and Boltzman Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing*, John Wiley & Sons, 1989.

[2] Abe, S., Thawonmas, R., "A Fuzzy Classifier with Ellipsoidal Regions," *IEEE Transactions on Fuzzy Systems*, Volume 5, No. 3, 1997, pp. 358-368.

[3] Alcalá, R., Casillas, J., Cordón, O., Herrera, F., "Approximate Mamdani-type Fuzzy Rule-Based Systems," *Technical Report #DECSAI-990117*, Dept. of Computer Science and A.I., University of Granada, October 1999.

[4] Alcalá, R., Casillas, J., Cordón, O., Herrera, F., Zwir, I., "Techniques for Learning and Tuning Fuzzy Rule-Based Systems for Linguistic Modeling and Their Application," in: C.T. Leondes (ed.), *Knowledge Engineering. Systems, Techniques and Applications*, Academic Press, 1999.

[5] Bardossy, A., Duckstein, L., *Fuzzy Rule-Based Modeling With Application to Geophysical, Biological and Engineering Systems*, CRC Press, 1995.

[6] Bastian, A., "How to Handle the Flexibility of Linguistic Variables with Applications," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Volume 2, No. 4, 1994, pp. 463-484.

[7] Carse, B., Fogarty, T.C., Munro, A., "Evolving Fuzzy Rule Based Controllers using Genetic Algorithms," *Fuzzy Sets and Systems*, Volume 80, 1996, pp. 273-293.

[8] Chi, Z., Wu, J., Yan, H., "Handwritten Numeral Recognition Using Self-organizing Maps and Fuzzy Rules," *Pattern Recognition*, Volume 28, No. 1, 1995, pp. 59-66.

[9] Cordón, O., Herrera, F., Peregrín, A., "Applicability of the Fuzzy Operators in the Design of Fuzzy Logic Controllers," *Fuzzy Sets and Systems*, Volume 86, 1997, pp. 15-41.

[10] Cordón, O., Herrera, F., "A Three-stage Evolutionary Process for Learning Descriptive and Approximative Fuzzy Logic Controller

Knowledge Bases from Examples," *International Journal of Approximate Reasoning*, Volume 17, No. 4, 1997, pp. 369-407.

[11] Cordón, O., del Jesus, M.J., Herrera, F., "Genetic Learning of Fuzzy Rule-based Classification Systems Cooperating with Fuzzy Reasoning Methods," *International Journal of Intelligent Systems*, Volume 13, No. 10-11, 1998, pp. 1025-1053.

[12] Cordón, O., Herrera, F., "A Proposal for Improving the Accuracy of Linguistic Modeling," *Technical Report #DECSAI-980113*, Dept. of Computer Science and A.I., University of Granada, May 1998.

[13] Cordón, O., Herrera, F., "ALM: A Methodology to Design Accurate Linguistic Models for Intelligent Data Analysis," *Third Intelligent Data Analysis Conference (IDA'99), L.N.C.S. 1642*, Amsterdam, Holland, 1999, pp. 15-26.

[14] Cordón, O., del Jesus, M.J., Herrera, F., "A Proposal on Reasoning Methods in Fuzzy Rule-based Classification Systems," *International Journal of Approximate Reasoning*, Volume 20, 1999, pp. 21-45.

[15] Cordón, O., Herrera, F., Sánchez, A., "Solving Electrical Distribution Problems Using Hybrid Evolutionary Data Analysis Techniques," *Applied Intelligence*, Volume 10, 1999, pp. 5-24.

[16] Cordón, O., Herrera, F., Villar, P., "Influence of Fuzzy Partition Granularity on Fuzzy Rule-Based System Behaviour", *EUSFLAT-ESTYLF Joint Conference*, Palma de Mallorca, Spain, 1999, pp. 159-162.

[17] Cordón, O., Herrera, F., Villar, P., "Analysis and Guidelines to Obtain a Good Uniform Fuzzy Partition Granularity for Fuzzy Rule-Based Systems using Simulated Annealing," *Technical Report #DECSAI-990116*, Dept. of Computer Science and A.I., University of Granada, September 1999.

[18] Cordón, O., Herrera, F., Zwir, I., "Linguistic Modeling by Hierarchical Systems of Linguistic Rules," *Technical Report #DECSAI-990114*, Dept. of Computer Science and A.I., University of Granada, July, 1999.

[19] Cordón, O., Herrera, F., "Hybridizing Genetic Algorithms with Sharing Scheme and Evolution Strategies for Designing Approximate Fuzzy Rule-Based Systems", *Fuzzy Sets and Systems*, Volume 111, No. 3, 2000.

[20] Goldberg, D.E., *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, 1989.

[21] González, A., Pérez, R., "Completeness and Consistency Conditions for Learning Fuzzy Rules," *Fuzzy Sets and Systems*, Volume 96, 1998, pp. 37-51.

[22] Herrera, F., Lozano, M., Verdegay, J.L., "Tuning Fuzzy Controllers by Genetic Algorithms," *International Journal of Approximate Reasoning*, Volume 12, 1995, pp. 299-315.

[23] Herrera, F., Lozano, M., Verdegay, J.L., "Fuzzy Connectives Based Crossover Operators to Model Genetic Algorithms Population Diversity," *Fuzzy Sets and Systems*, Volume 92, No. 1, 1997, pp. 21-30.

[24] Ishibuchi, H., Nozaki, K., Tanaka, H., " Distributed Representation of Fuzzy Rules and Its Application to Pattern Classification," *Fuzzy Sets and Systems*, Volume 52, 1992, pp. 21-32.

[25] Ishibuchi, H., Nozaki, K., Yamamoto, N., Tanaka, H., " Selecting Fuzzy If-then Rules for Classification Problems Using Genetic Algorithms," *IEEE Transactions on Fuzzy Systems*, Volume 3, No. 3, 1995, pp. 260-270.

[26] Mandal, D.P., Murthy, C.A., Pal, S.K., "Formulation of a Multivalued Recognition System," *IEEE Transactions on Systems, Man, and Cybernetics*, Volume 22, No. 4, 1992, pp. 607-620.

[27] Michalewicz, Z., *Genetic Algorithms + Data Structures = Evolution Programs*, Third edition, Springer-Verlag, 1996.

[28] Pedrycz, W. (Ed.), *Fuzzy Modelling: Paradigms and Practice*, Kluwer Academic Press, 1996.

[29] Sugeno, M., Yasukawa, T., "A Fuzzy-logic-based Approach to Qualitative Modeling," *IEEE Transactions on Fuzzy Systems*, Volume 1, No. 1, 1993, pp. 7-31.

[30] Wang, L.X., Mendel, J.M., "Generating Fuzzy Rules by Learning from Examples," *IEEE Transactions on Systems, Man, and Cybernetics*, Volume 22, 1992, pp. 1414-1427.

[31] Wang, L.X., *Adaptive Fuzzy Systems and Control*, Prentice-Hall, 1994.

[32] Weiss. S.M., Kulikowski, C.A., *Computer Systems that Learn*, Morgan Kaufmann Publishers, 1991.

[33] Zadeh, L.A., "Outline of a New Approach to the Analysis of Complex Systems and Decision Processes," *IEEE Transactions on Systems, Man, and Cybernetics*, Volume 3, No. 1, 1973, pp. 28-44.

[34] Zadeh, L.A., "The Concept of a Linguistic Variable and its Application to Aproximate Reasoning," *Information Science*, Part I: Volume 8, 1975, pp. 199-249, Part II: Volume 8, 1975, pp. 301-357, Part III: Volume 9, 1975, pp. 43-80.