Linguistic Modeling with Weighted Double-Consequent Fuzzy Rules Based on Cooperative Coevolutionary Learning *

Rafael Alcalá[1], Jorge Casillas[2], Oscar Cordón[2], Francisco Herrera[2]

[1]Dept. of Computer Science
University of Jaén, E-23071 – Jaén, Spain
e-mail: alcala@ujaen.es

[2]Dept. of Computer Science and Artificial Intelligence
University of Granada, E-18071 – Granada, Spain
e-mail: {casillas,ocordon,herrera}@decsai.ugr.es

Abstract

This paper presents an evolutionary learning process for linguistic modeling with weighted double-consequent fuzzy rules. These kinds of fuzzy rules are used to improve the linguistic modeling, with the aim of introducing a trade-off between interpretability and precision.

The use of weighted double-consequent fuzzy rules makes more complex the modeling and learning process, increasing the solution search space. Therefore, the cooperative coevolution, an advanced evolutionary technique proposed to solve decomposable complex problems, is considered to learn these kinds of rules. The proposal has been tested with different problems achieving good results.

Keywords: Fuzzy linguistic modeling, double-consequent fuzzy rules, weighted fuzzy rules, genetic algorithms, cooperative coevolution.

*Supported by the Spanish CICYT, project PB98-1319.

R. Alcalá, J. Casillas, O. Cordón, F. Herrera. Linguistic modeling with weighted double-consequent fuzzy rules based on cooperative coevolutionary learning. Integrated Computer Aided Engineering 10 (4), 2003, 343-355.

1 Introduction

One of the problems associated with Linguistic Modeling (LM) is its lack of accuracy when modeling some complex systems. It is due to the inflexibility of the concept of linguistic variable, which imposes hard restrictions to the fuzzy rule structure [1]. A way to improve the LM accuracy losing interpretability but not to a high degree is to extend the usual linguistic model structure to be more flexible.

Two specific possibilities for modifying the rule structure have been considered in the literature:

- double-consequent fuzzy rules [5, 15], where each combination of antecedents may have two consequents,
- and weighted fuzzy rules [4, 16, 25], where an importance degree is considered for each rule in the fuzzy reasoning process.

In the same way, even more flexible fuzzy rules may be obtained combining both approaches to design linguistic models based on *weighted double-consequent fuzzy rules*, thus involving a potential improvement of the accuracy and maintaining an acceptable description level. This task could be performed by Genetic Fuzzy Systems (GFSs) [6], usually based on Genetic Algorithms (GAs) [14]. However, the use of weights and double-consequents makes more complex the modeling process as it increases the solution search space since new parameters are considered in addition to the traditional approach.

Recently, an advanced evolutionary technique, cooperative coevolution [17, 20], has arised to solve problems with a large search space by independently evolving two or more species which together comprise solution structures. We will use this novel technique to generate linguistic models with the said extended structure by using a preliminary simple linguistic model with a large number of simple and double-consequent rules and coevolving two species, the subset of rules best cooperating and the weights associated to them.

Notice that, this contribution proposes the use of weighted double-consequent fuzzy rules to improve simple linguistic fuzzy models by means of a cooperative coevolutionary algorithm. *This can be intended as a meta-method over any other linguistic rule generation method, developed to improve simple linguistic fuzzy models by only reinforcing the modeling of those problem subspaces with more difficulties while the use of rule weights improves the way in which they interact.* Depending on the combination of this technique with different inductive methods, different learning approaches arise. In this work, we will consider the Wang and Mendel's method [23] (WM) for this purpose.

The paper is organized as follows. In Section 2, the said specific ways to relax the linguistic model structure are presented in depth. Section 3 briefly introduces the concepts of GAs, GFSs and cooperative coevolution. In Section 4, the weighted double-consequent fuzzy rule structure together with the coevolutionary genetic learning method to derive these kinds of rules are proposed. Experimental results are shown in Section 5, whilst some concluding remarks are pointed out in Section 6. Appendix A presents a table with the used acronyms.

2 Flexibilizing the Fuzzy Linguistic Model

Nowadays, system modeling is one of the most important applications in the framework of the Fuzzy Rule-Based Systems (FRBSs) [12, 13, 18, 26]. It may be considered as an approach used to model a system making use of a descriptive language based on fuzzy logic with fuzzy predicates [22]. In this kind of modeling the *accuracy* and the *interpretability* of the obtained model are contradictory properties directly depending on the learning process and/or the model structure. Traditionally, according to the rule structure it is possible to distinguish between two modeling approaches clearly opposed: *LM* and *Fuzzy Modeling*, with the interpretability and the accuracy of the model being their main requirement, respectively.

LM is developed by means of linguistic FRBSs, typically called Mamdani-type FRBSs [12, 13], which are composed of input and output linguistic variables [27] taking values from a linguistic term set with a real-world meaning. Therefore, each rule may be clearly interpreted by human beings.

Improvements in the LM can be accomplished to make more flexible the learning and/or the model structure [3]. Two specific possibilities to relax the model structure are the following:

- *Use of double-consequent fuzzy rules*, which involves allowing the model to present rules where each combination of antecedents may have two consequents associated when it improves the model accuracy [5, 15].
- *Consideration of weighted fuzzy rules* in which modifying the linguistic model structure an importance factor (weight) is considered for each rule [4, 16, 25]. By means of this technique, the way in which these rules interact with their neighbor ones could be indicated.

It is clear that both possibilities will improve the capability of the model to perform the *interpolative reasoning* and, thus, its performance. This is one of the most interesting features of FRBSs and plays a key role in their high performance, being a consequence of the cooperative action of the linguistic rules existing in the fuzzy rule base.

On the other hand, simply new implicit granularity levels are added with these kinds or rules. Therefore, weighted double-consequent fuzzy models are less interpretable than the classical linguistic ones but, in any case, these kinds of FRBSs can be interpreted to a high degree, and also make use of human knowledge and a deductive process. Moreover, notice that this paper will propose a linguistic fuzzy modeling technique that only consider the use of double-consequents when it is really needed to improve the accuracy of the system.

In the following subsections, these two approaches to relax the model structure will be analyzed.

2.1 Double-Consequent Fuzzy Linguistic Rules

More flexible linguistic models may be obtained by allowing them to present fuzzy rules where each combination of antecedents may have two consequents (linguistic terms of the output variable) associated. The consideration of these kinds of rules may be intended as a local reinforcement of the problem space zones presenting high complexity. Therefore, as shown in [5, 15], considering some rules with multiple consequents could improve the global system behavior. The rule structure so obtained is:

$$\textbf{IF } X_1 \textbf{ is } A_1 \textbf{ and } \ldots \textbf{ and } X_n \textbf{ is } A_n \textbf{ THEN } Y \textbf{ is } \{B_1, B_2\},$$

with X_i (Y) being the linguistic input (output) variables, A_i being the linguistic label used in the i-th input variable, and B_1 and B_2 the two linguistic terms associated to the output variable.

The use of two consequents has no influence on the linguistic model inference system. Since each double-consequent fuzzy rule can be decomposed into two simple rules with the same antecedent and different consequent, the usual plain fuzzy inference system can be applied [5]. The only restriction imposed is to use the FITA (First Infer, Then Aggregate) scheme [7] and considering the matching degree of the rules fired. In other case, by using the FATI (First Aggregate, Then Infer) scheme [7] or a deffuzzification strategy not considering the matching, the influence of one of the two rule consequents will be canceled. For example, the *center of gravity weighted by the matching degree* defuzzification strategy [7] may be used:

$$P_i = \frac{\int_{\mathbf{V}} Y \cdot \mu_{B'_i}(Y) \cdot dY}{\int_{\mathbf{V}} \mu_{B'_i}(Y) \cdot dY},$$

$$y_0 = \frac{\sum_i m_i \cdot P_i}{\sum_i m_i},$$

with y_0 being the crisp value obtained from the defuzzification process, m_i being the matching degree of the i-th rule, V being the universe of discourse of the output variable Y, and P_i being the characteristic value —*center of gravity*— of the output fuzzy set inferred from the i-th rule, B'_i.

We should note these kinds of rules do not constitute an inconsistency from the LM point of view but only a shift of the main labels making the final output of the rule lie in an intermediate zone between both consequents. In this case, simply new implicit granularity levels are added with these kinds or rules (similar to the use of linguistic modifiers where the interpretability is slightly lost but being interpretable to an acceptable degree). Indeed, the said double-consequent fuzzy rule structure may be interpreted as follows [5]:

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

whose output is exactly the middle point between these two consequents,

$$y_0 = \frac{m_i \cdot P_{B_1} + m_i \cdot P_{B_2}}{m_i + m_i} = \frac{P_{B_1} + P_{B_2}}{2}.$$

Of course, notice that when these double-consequent rules interact with their neighbor ones, the output is not the middle point between both consequents but it is shifted according to the firing strengths of those neighbor rules.

An example of learning process

The consideration of this structure to generate advanced linguistic models was initially proposed in [15]. Another approach, according to the Accurate Linguistic Modeling (ALM) methodology, is introduced in [5]. In this work, the use of double-consequent is considered when it is really needed to improve the accuracy of the system. This methodology consists of two steps:

1. Firstly, two rules, the primary and secondary in importance, are obtained in each fuzzy input subspace considering a specific generation process. In this contribution, the generation process proposed by Wang and Mendel [23] is considered. Thus, the process involves dividing the input and output spaces into fuzzy regions, generating the rule best covering each example, and finally selecting the two rules with the highest covering degree for each fuzzy input subspace (if there is more than a single rule on it).

2. Then, after decomposing each double-consequent rule into two independent simple ones [1], the selection process proposed in [10] is employed to select the subset of rules best cooperating. It is based on a binary-coded GA where each gene indicates if the corresponding rule is considered or not to belong to the final fuzzy rule base.

2.2 Weighted Fuzzy Linguistic Rules

Using rule weights [4, 16, 25] has been usually considered to improve the way in which the rules interacts, improving the accuracy of the learned model. In this way, rule weights suppose an effective extension of the conventional fuzzy reasoning system that allow the tuning of the system to be developed at the rule level [4, 16].

When weights are applied to complete rules, the corresponding weight is used to modulate the firing strength of a rule in the process of computing the defuzzified value. From human beings, it is very near to consider this weight as an importance degree associated to the rule, determining how this rule interacts with its neighbor ones. We will follow this approach, since the interpretability of the system is appropriately maintained. In addition, we will only consider weight values in $[0, 1]$

[1]The preliminary rule base is derived to simple rules only to be considered in the selection process. If one of the two simple rules obtained from decomposing a double-consequent rule is removed by the selection process, this fuzzy input subspace will have just a single consequent associated.

since it preserves the model readability. In this way, the use of rule weights represents an ideal framework for extended LM when we search for a trade-off between accuracy and interpretability.

In order to do so, we will follow the weighted rule structure and the inference system proposed in [16]:

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

where w is the real-valued rule weight, and *with* is the operator modeling the weighting of a rule.

With this structure, the fuzzy reasoning must be extended. The classical approach is to infer with the FITA (First Infer, Then Aggregate) scheme [7] and to compute the defuzzified output as the following *weighted sum*:

$$y_0 = \frac{\sum_i m_i \cdot w_i \cdot P_i}{\sum_i m_i \cdot w_i},$$

with m_i being the matching degree of the i-th rule, w_i being the weight associated to it, and P_i being the characteristic value of the output fuzzy set corresponding to that rule. In this contribution, the center of gravity will be considered as characteristic value [7] (see the previous subsection).

An example of learning process

A simple approximation for weighted rule learning would consist of the following two steps —we will use this process in our experiments for comparison purposes, calling it WRL—:

1. Firstly, a preliminary fuzzy rule set is derived considering a specific generation process. In this work, the generation process proposed by Wang and Mendel [23] is considered.
2. Then, a learning algorithm is used to derive the associated weights of the previously obtained rules. A real-coded GA where each gene indicates the corresponding rule weight may be considered as learning algorithm.

3 GAs, GFSs and Coevolution

Considering both approaches —weighted and double-consequent rules— together, makes more complex the modeling process thus increasing the solution search space.

GFSs have been successfully applied to learn fuzzy systems in the last years. They have been usually based on GAs although other evolutionary algorithms have been also considered. On the other hand, more sophisticated evolutionary approaches, as cooperative coevolution [17, 20], could be considered to solve this complex modeling process. These concepts are introduced in this section.

3.1 Genetic Algorithms

GAs are general-purpose global search algorithms that use principles inspired by natural population genetics to evolve solutions to problems. The basic principles of the GAs were first laid down rigorously by Holland [11] and are well described in many texts as [14].

The basic idea is to maintain a population of knowledge structures that evolves over time through a process of competition and controlled variation. Each structure in the population represents a candidate solution to the specific problem and has an associated *fitness* to determine which structures are used to form new ones in the process of competition.

In this way, a subset of relatively good solutions are selected for reproduction to give offspring that replace the relatively bad solutions which die. Usually, offspring replace their parents for the next generation (generational approach). These new individuals are created by using genetic operators such as crossover and mutation. The crossover operator combines the information contained into the parents increasing the average quality of the population (exploitation), while the mutation operator randomly changes the new individuals helping the algorithm to avoid local optima (exploration).

3.2 Genetic Fuzzy Systems

During the 90s, a large amount of work has been devoted to add learning capabilities to FRBSs. The automatic design of FRBSs can be considered in many cases as an optimization or search process on the space of potential solutions. Since GAs are well known and widely used global search techniques, a large number of publications explored the use of GAs for designing FRBSs, thus obtaining the so-called GFSs [6]. Figure 1 depict this idea.

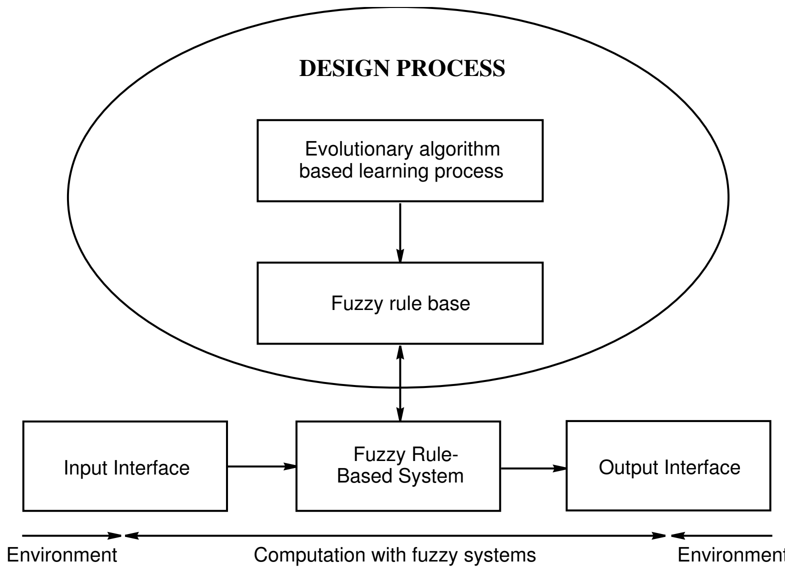

Figure 1: Genetic fuzzy systems.

Nowadays, as the field of GFSs matures and grows in visibility, there is an increasing concern about the integration of these two topics from a novel more sophisticated perspective. Indeed, as David Goldberg stated in [9], the integration of single methods into hybrid intelligent systems goes beyond simple combinations. For him, the future of Computational Intelligence "*lies in the careful integration of the best constituent technologies*" and subtle integration of the abstraction power of fuzzy systems and the innovating power of genetic systems requires a design sophistication that goes further than putting everything together. This is the case of our contribution, where we use a cooperative coevolutionary model for learning the weighted double-consequent fuzzy rules.

3.3 Cooperative Coevolutionary Algorithms

Coevolutionary algorithms [17] are advanced evolutionary techniques proposed to solve decomposable complex problems. They involve two or more species (populations) that permanently interact among them by a coupled fitness. Thereby, in spite of each species has its own coding scheme and reproduction operators, when an individual must be evaluated, its goodness is calculated considering some individuals of the other species. This coevolution makes easier to find good solutions to complex problems.

Different kinds of interactions may be considered among the species according to the dependencies existing among the solution subcomponents. Generally, we can distinguish two different kinds of interaction:

- *Competitive coevolutionary algorithms* [21]: Those where each species competes with the remainder. In this case, increasing the fitness of an individual in a species implies decreasing the fitness of the ones in the other species, i.e., the success of somebody else entails the personal failure.

- *Cooperative* or *symbiotic coevolutionary algorithms* [20]: Those where all the species cooperate to build the problem solution. In this case, the fitness of an individual depends on its ability to cooperate with individuals from other species.

The use of cooperative coevolutionary algorithms is recommendable when the following issues arise [19]:

1. the search space is huge,
2. the problem may be decomposable in subcomponents,
3. different coding schemes are used, and
4. there are strong interdependencies among the subcomponents.

In Figure 2, the evolutionary process for S species of a cooperative coevolutionary system is illustrated. Each individual being evaluated could be combined with one or more *cooperators* of the other species.

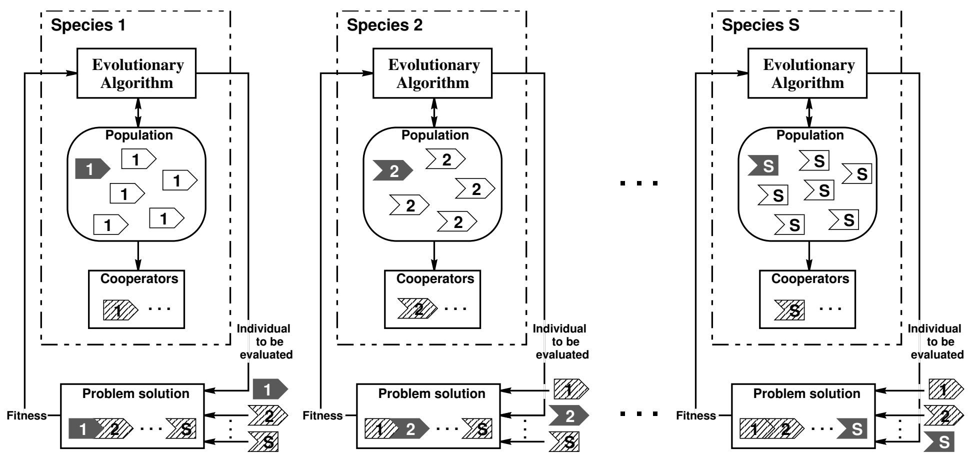

Figure 2: Cooperative coevolutionary system for S species.

This sophisticated technique can be used within the field of GFSs. Indeed, in [19] this technique has been already applied to learn FRBSs coevolving two species, the membership functions and the fuzzy rules.

4 Genetic Fuzzy Systems to Generate Weighted Double-Consequent Fuzzy Rules

In this section the structure of the proposed weighted double-consequent fuzzy rule as well as a cooperative coevolutionary-based GFS to learn these kinds of linguistic models are explained in detail. Finally, in order to check the behavior of the coevolutionary approach a genetic model based on a standard GA is presented as a first approximation to learn these kinds of models.

4.1 Rule Structure and Learning Process

To improve the linguistic model accuracy, the use of a more flexible linguistic model structure that combine the two said approaches is proposed. Thus, the weighted double-consequent fuzzy rules present the following structure:

$$\textbf{IF } X_1 \textbf{ is } A_1 \textbf{ and } \ldots \textbf{ and } X_n \textbf{ is } A_n \textbf{ THEN } Y \textbf{ is } \{B_1 \textit{ with } [w_1], B_2 \textit{ with } [w_2]\},$$

with w_1 and w_2 being the weights associated to the consequents B_1 and B_2, respectively. Therefore, a weighted double-consequent fuzzy rule can be seen as two weighted single-consequent fuzzy rules with the same antecedent and different consequents (and so, still considering two consequents in the corresponding subspace). Thus, the fuzzy reasoning must be extended as in the case of weighted fuzzy rules, considering the matching degree of the rules fired (see Sections 2.1 and 2.2).

These kinds of rules could be interpreted by adding the characteristics of double-consequent and weighted fuzzy rules. Therefore, when double consequent fuzzy rules are obtained, the output can be interpreted as a shift of the main labels making the final output of the rule lie in an intermediate zone between both consequents. However, in this case the way in which these rules interact is known since the correspondent weights can be interpreted as their importance degree (see Section 2.2).

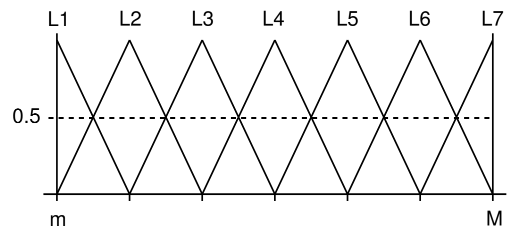

Figure 3: Graphical representation of a possible fuzzy partition.

To generate linguistic models with this new structure, we may follow an operation mode similar to the ALM methodology [5] introduced in Section 2.1, but including the weight learning. To do that, we will consider symmetrical fuzzy partitions of triangular-shaped membership functions (see Figure 3). Therefore, after performing the first step of the ALM methodology, where an initial set of numerous double-consequent rules is generated, and decomposing them to simple ones (see Section 2.1), the two following tasks must be performed:

- Genetic selection of a subset of rules presenting good cooperation.
- Genetic derivation of the weights associated to these rules.

These interdependent tasks significantly increase the search space with respect to the original methodology making the choice of the considered search technique crucial.

4.2 Evolutionary Learning of Weighted Double-Consequent Fuzzy Rules with a Cooperative Coevolutionary Model

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

4.2.1 Interaction Scheme Between Species

The objective will be to minimize the well-known *Mean Square Error* (MSE):

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

with N being the number of training data, $F_{ij}(x^l)$ being the output inferred from the model obtained by combining the individuals i and j of the species 1 (rule selection) and 2 (weight derivation) when the input $x^l = (x_1^l, \ldots, x_n^l)$ is presented, and y^l being the known desired output.

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

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

with i and j being individuals of species 1 and 2 respectively, R_1 and R_2 being the sets of the r fittest individuals in the previous generation of the species 1 and 2 respectively, and P_1 and P_2 being the sets of the p individuals selected at random from the previous generation of the species 1 and 2 respectively. This evaluation process is graphically shown in Figure 4.

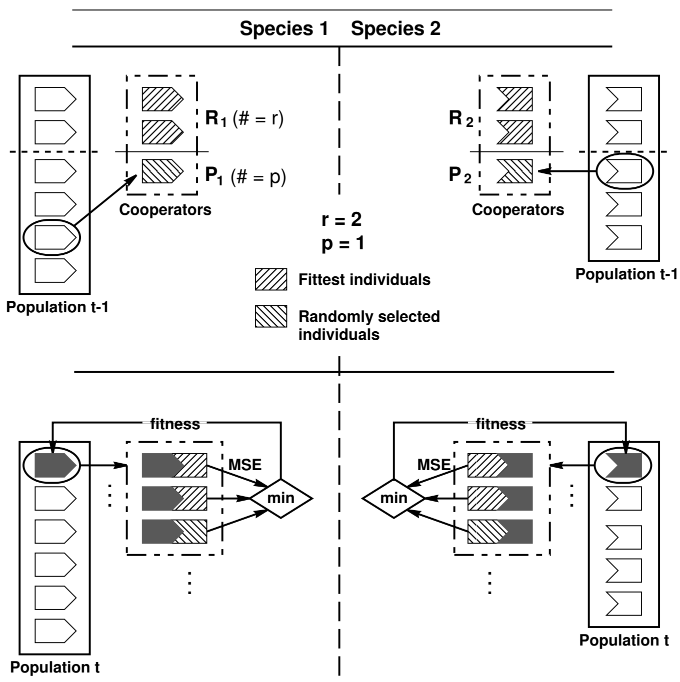

Figure 4: Fitness evaluation for species 1 and 2.

Whilst the sets $R_{1|2}$ allow the best individuals to influence in the process guiding the search towards good solutions, the sets $P_{1|2}$ introduce diversity in the search. The combined use of both kinds of sets makes the algorithm have a trade-off between exploitation ($R_{1|2}$) and exploration ($P_{1|2}$). The cardinalities of the sets $R_{1|2}$ and $P_{1|2}$ have to be previously defined by the designer.

A **generational** [14] scheme is followed in both species. Baker's stochastic universal sampling procedure [2] together with an elitist mechanism (that ensures to maintain the best individual of the previous generation) are used.

The specific operators considered in every species are described in the following.

4.2.2 Species 1: Fuzzy rule selection

For the species 1, we will use the genetic rule selection method proposed in [5]. The **coding scheme** generates binary-coded strings of length m (with m being the number of single-consequent fuzzy rules in the previously derived rule set, obtained in the first step of ALM). Depending on whether a rule is selected or not, the alleles '1' or '0' will be respectively assigned to the corresponding gene. Thus, the p-th chromosome for the species 1, C_1^p, will be a binary vector representing the subset of rules finally obtained.

The **initial pool** is generated at random except for the first individual, which represents the complete previously obtained fuzzy rule set:

$$\forall k \in \{1, \ldots, m\},\ C_1^1[k] = 1.$$

For this species, the standard two-point **crossover** operator is used. The two-point crossover involves interchanging the fragments of the parents contained between two points selected at random. As regards the **mutation** operator, it flips the value of the gene.

4.2.3 Species 2: Weight derivation

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

The **initial pool** for this species is generated with the first chromosome having all the genes with the value '1', and the remaining individuals taking values randomly generated within the variation interval $[0, 1]$:

$$\forall k \in \{1, \ldots, m\},\ C_2^1[k] = 1.0.$$

The max-min-arithmetical **crossover** operator [10] is considered. Using the max-min-arithmetical crossover, if $C_2^v = (c_1, \ldots, c_k, \ldots, c_n)$ and $C_2^w = (c'_1, \ldots, c'_k, \ldots, c'_n)$ are going to be crossed, the resulting descendents are the two best of the next four offspring:

$$\begin{aligned} &C_2^1 = aC_2^w + (1-a)C_2^v, && C_2^2 = aC_2^v + (1-a)C_2^w, \\ &C_2^3 \text{ with } c_{3k} = \min\{c_k, c'_k\}, && C_2^4 \text{ with } c_{4k} = \max\{c_k, c'_k\}, \end{aligned}$$

with $a \in [0, 1]$ being a constant parameter chosen by the GA designer.

As regards the **mutation** operator, it simply involves changing the value of the selected gene by other value obtained at random within the interval $[0, 1]$.

4.3 A Standard Genetic Algorithm

A standard GA performing the rule selection together with the derivation of weights was developed as a first approximation to the problem. The classical generational [14] scheme together with the Baker's stochastic universal sampling procedure [2] and an elitist mechanism were considered in this algorithm.

Coding scheme and initial gene pool

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

- The coding scheme for the C_1 part was introduced in Section 4.2 as the coding scheme for species 1. Thus, the corresponding part C_1^p for the p-th chromosome will be a binary vector representing the subset of rules finally obtained.
- The coding scheme for the C_2 part was also introduced in Section 4.2 as the coding scheme for species 2. Now, the corresponding part C_2^p for the p-th chromosome will be a real-valued vector representing the weights associated to the fuzzy rules considered.

The initial pool is obtained with an individual having all genes with value '1' in both parts, and the remaining individuals generated at random.

Evaluating the chromosome

The fitness function considered is the MSE. With $F_i(x^l)$ being the model inferred output for the i-th chromosome, this measure is represented by the following expression:

$$f(i) = \text{MSE}_i = \frac{1}{2 \cdot N} \sum_{l=1}^{N} (F_i(x^l) - y^l)^2,$$

Genetic operators

The **crossover** operator will depend on the chromosome part where it is applied: in the C_1 part, the standard two-point crossover is used, whilst in the C_2 part, the max-min-arithmetical crossover [10] is considered. Both operators were explained in the previous section. In this case, eight offspring are generated by combining the two ones from the C_1 part (two-point crossover) with the four ones from the C_2 part (max-min-arithmetical crossover). The two best offspring so obtained replace the two correspondent parents in the population.

As regards the **mutation** operator, it flips the gene value in the C_1 part while takes a value at random within the interval $[0, 1]$ for the corresponding gene in the C_2 part. Both operators were also explained in the previous section.

5 Experiments and Results

In this section, we will analyze the performance of the two GFSs presented in Sections 4.2 and 4.3, the proposed cooperative coevolutionary algorithm (calling it as WALM-CC) and the standard steady-state GA-based method (calling it as WALM), when solving two real-world electrical engineering distribution problems [8]. They will be compared to the models designed by the following methods: the well-known ad hoc data-driven method proposed by Wang and Mendel (calling it as WM) [23], the original ALM method [5], and the WRL method presented in Section 2.2. Except WM, all of them follow a **generational** [14] scheme.

With respect to the fuzzy reasoning method used, we have selected the *minimum t-norm* playing the role of the implication and conjunctive operators, and the *center of gravity weighted by the matching* strategy acting as the defuzzification operator [7].

Finally, the following values have been considered for the parameters of each method [2]:

- *Rule selection step of ALM*: 61 individuals, 1000 generations, 0.6 as crossover probability, and 0.2 as mutation probability per chromosome.
- *WRL*: 61 individuals, 1,000 generations, 0.6 as crossover probability, 0.2 as mutation probability per chromosome, and 0.35 for the factor a in the max-min-arithmetical crossover.
- *WALM*: 61 individuals, 1,000 generations, 0.6 as crossover probability, 0.2 as mutation probability per chromosome, and 0.35 for the factor a in the crossover operator.
- *WALM-CC*: 62 individuals (31 for each species), 1,000 generations, 0.6 and 0.2 for the crossover and mutation probabilities in both species respectively, 0.35 for the factor a in the crossover operator in the species 2, the three fittest individuals ($|R_{1|2}| = 3$) and two random individuals ($|P_{1|2}|=2$) of each species are considered for the coupled fitness.

[2]With these values we have tried easy the comparisons selecting standard common parameters that work well in most cases instead of searching very specific values for each method. Moreover, we have set a large number of generations (1000) in order to allow the compared algorithms an appropriate convergence. No significant changes were achieved by increasing that number of generations.

5.1 Estimating the Length of Low Voltage Lines

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 as the estimation of the maintenance costs of the network, which was the main goal in this application [8]. Since a direct measure is very difficult to obtain in some cases [3], the consideration of models becomes useful. In this way, the problem involves finding a model that relates the *total length of low voltage line* installed in a rural town with the *number of inhabitants* in the town and the *mean of the distances from the center of the town to the three furthest clients* in it. This model will be used to estimate the total length of line being maintained.

To do so, a sample of 495 rural nuclei has been randomly divided into two subsets, the training set with 396 elements and the test set with 99 elements, the 80% and the 20% respectively. Both data sets considered are available at *http://decsai.ugr.es/∼casillas/fmlib/*. The linguistic partitions considered are comprised by seven linguistic terms with triangular-shaped fuzzy sets giving meaning to them (see Figure 3). The corresponding labels, $\{L_1, L_2, L_3, L_4, L_5, L_6, L_7\}$, stand for extremely small, very small, small, medium, large, very large, and extremely large respectively.

The results obtained by the five analyzed methods are presented in Table 1, where $\#R$ stands for the number of rules, and MSE$_{tra}$ and MSE$_{tst}$ for the error obtained over the training and test data respectively. The best results are shown in boldface.

Table 1: Results obtained in the length of low voltage lines estimation problem

Method	**#R** ←	(SC+DC)	**MSE**$_{tra}$	**MSE**$_{tst}$
WM	24	–	222,654	239,962
ALM	**17**	(14+3)	155,898	178,534
WRL	24	–	191,577	221,583
WALM	17	(14+3)	145,124	186,704
WALM-CC	18	(14+4)	**144,290**	**176,057**

SC = Single Consequent, DC = Double Consequent.

In this case, WM —performing classical LM— presents the worst results obtaining more rules than the remaining ones. ALM and WRL —performing improved LM— present significantly more accurate models than WM in both, approximation and generalization, with improvements of about 30% and 25% respectively.

WALM does not achieve the desired results, only improving the result obtained from ALM in the approximation capability. This fact evidences some lacks in the optimization technique considered to learn weighted double-consequent fuzzy rules, since theoretically, better results than the ones obtained by ALM should be obtained.

In this case, the model obtained by WALM-CC presents the best results with improvement rates near of the 10% in generalization (MSE$_{tst}$) with respect to the results obtained by WALM. Therefore, the necessity of considering a sophisticated technique to solve this complex learning problem seems to be interesting.

Figure 5 represents the decision table of the model obtained from WALM-CC. In the left side of this figure, each cell of the table represents a fuzzy subspace and contains its associated output consequent(s) —the primary (C1) and/or the secondary (C2) in importance—, i.e., the correspondent label(s) together with its(their) respective rounded rule weight(s). The *absolute importance weight* for each fuzzy rule has been graphically showed by means of the grey colour scale, from black (1.0) to white (0.0). In this way, we can easily see the importance of a rule with *respect to their neighbor ones* which could help the system experts to identify important rules. On the other hand, in the right side of this figure, an expert interpretation of the *relative importance* of the rules is presented as regards their influence in the modeling of the respective problem space zone. Three kinds of rules are represented in the figure:

[3] Low voltage lines installation is often very intricate since they are contained in little villages and rural nuclei.

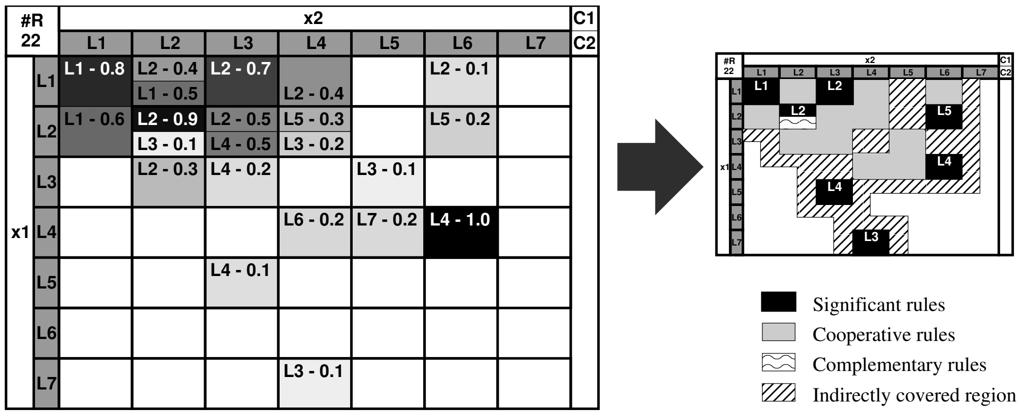

Figure 5: Decision table of the linguistic model obtained from WALM-CC for the length of low voltage lines estimation problem.

- *Significant or important rules*: Those in black, corresponding to rules that have a higher weight than their neighbors or rules that are the ones of their regions.
- *Cooperative rules*: Those in grey, representing rules that have a more or less similar weight than their neighbor ones.
- *Complementary rules*: Those in white (with waves), representing rules that have a lower weight than their neighbor ones.

Due to the kind of fuzzy partition considered (see Figure 3), there are many input subspaces which, in spite of having no rules associated, are covered by their neighbor rules, e.g., the input subspace labeled as L_3-L_4. Two different zones can be clearly distinguished in the table. The first one —located in the top-left corner— presents three significant rules with outputs about L_2 and coincides with an important concentration of training examples. The second one composes a front in the input space and presents significant rules with outputs about L_4.

In this case, only four weighted double-consequent rules have been needed (locally improving the model where it is necessary). Regarding the significant rules, we should notice that the importance of the rules does not directly depend on the weights of the rules, but also on the weights of their neighbors. Thus, rules as the one located in the input subspace labeled as L_2-L_6 become significant rules since their weights are higher than those of their neighbors.

5.2 Estimating the Maintenance Costs of Medium Voltage Lines

Estimating the maintenance costs of the medium voltage electrical network in a town [8] is an interesting problem. Since the medium voltage lines existing in a town have been installed incrementally, a direct measure is very difficult to obtain and the consideration of models becomes useful. These estimations allow electrical companies to justify their expenses. Moreover, the model must be able to explain how a specific value is computed for a certain town. Our objective will be to relate the *maintenance costs of medium voltage line* with the following four variables: *sum of the lengths of all streets in the town*, *total area of the town*, *area that is occupied by buildings*, and *energy supply to the town*. We will deal with estimations of minimum maintenance costs based on a model of the optimal electrical network for a town in a sample of 1,059 towns.

To develop the different experiments, 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 847 elements, whilst the test one is composed by 212 elements. These data sets are available at `http://decsai.ugr.es/∼casillas/fmlib/`. Five linguistic terms with triangular-shaped fuzzy sets giving meaning to them are considered for each variable (see Figure 3). In these

Table 2: Results obtained in the maintenance costs of medium voltage lines estimation problem

Method	#R	← (SC+DC)	$\mathbf{MSE}_{tra}$	$\mathbf{MSE}_{tst}$
WM	66	–	71,294	80,934
ALM	**47**	(44+3)	51,714	58,806
WRL	66	–	33,639	33,319
WALM	47	(43+4)	27,719	32,455
WALM-CC	54	(49+5)	**24,961**	**28,225**

SC = Single Consequent, DC = Double Consequent.

case, the corresponding labels, $\{L_1, L_2, L_3, L_4, L_5\}$, stand for very small, small, medium, large, and very large, respectively.

In view of the obtained results, we can see once again as the ALM and WRL methods —which are respectively based on double-consequent fuzzy rules and weighted fuzzy rules— achieve significant improvements over the WM method. The model obtained from ALM is the one comprised of less fuzzy rules. However, all the methods except WM present better results than the one obtained from ALM. Apart from ALM, the two proposed GFSs obtain the models with less number of rules.

Analyzing the model obtained by the WALM-CC method, we can conclude that it presents the best performance in both approximation and generalization, with improvement rates of about 25% and 15% with respect to WRL, respectively. Moreover, WALM-CC presents improvement rates of about a 10% in approximation and generalization with respect to WALM. This fact is a consequence of the coevolutionary approach ability to tackle with decomposable complex problems.

#R: 54 (49 + 5 DC)

X1	X2	X3	X4	Y	with
L1	L1	L1	L1	L1	[0.838]
L1	L1	L1	L2	L2	[0.688]
L1	L2	L1	L1	L1	[0.386]
L1	L2	L2	L1	L2	[0.466]
L1	L2	L2	L2	L2	[0.355]
L2	L1	L1	L1	L1	[0.491]
L2	L1	L2	L2	L2	[0.573]
L2	L2	L1	L1	L1	[0.421]
L2	L2	L1	L2	L2	[0.388]
L2	L2	L2	L1	L1	[0.099]
L2	L2	L2	L2	L2,L3	[0.605,0.217]
L2	L3	L2	L1	L2	[0.803]
L2	L3	L2	L2	L2	[0.185]
L2	L3	L3	L2	L3	[0.945]
L3	L2	L1	L1	L1	[0.452]
L3	L2	L1	L2	L2	[0.994]
L3	L2	L1	L3	L2	[0.283]
L3	L2	L2	L3	L3	[0.322]

X1	X2	X3	X4	Y	with
L3	L3	L2	L2	L2,L3	[0.684,0.108]
L3	L3	L2	L3	L3	[0.476]
L3	L3	L3	L2	L3	[0.773]
L3	L3	L3	L3	L4	[0.497]
L3	L4	L3	L2	L3	[0.256]
L3	L4	L3	L3	L3	[0.716]
L3	L4	L4	L3	L4	[0.847]
L4	L2	L2	L1	L2	[0.372]
L4	L2	L2	L2	L2,L3	[0.254,0.178]
L4	L2	L2	L3	L3,L2	[0.152,0.171]
L4	L2	L2	L4	L3	[0.267]
L4	L3	L2	L1	L2	[0.798]
L4	L3	L2	L3	L3	[0.366]
L4	L3	L2	L4	L3	[0.696]
L4	L3	L3	L2	L3	[0.200]
L4	L3	L3	L3	L4	[0.166]
L4	L3	L3	L4	L4	[0.198]
L4	L4	L3	L2	L3	[0.526]

X1	X2	X3	X4	Y	with
L4	L4	L3	L3	L3	[0.257]
L4	L4	L3	L4	L4	[0.467]
L4	L4	L4	L2	L4	[0.776]
L4	L4	L4	L3	L4	[0.321]
L4	L4	L4	L4	L5	[0.786]
L4	L5	L4	L2	L3	[0.324]
L4	L5	L4	L3	L4	[0.907]
L4	L5	L4	L4	L5	[0.014]
L4	L5	L5	L2	L5	[0.434]
L4	L5	L5	L3	L5	[0.893]
L5	L2	L2	L2	L2,L3	[0.232,0.151]
L5	L2	L2	L4	L3	[0.044]
L5	L2	L2	L5	L4	[0.824]
L5	L2	L3	L2	L3	[0.985]
L5	L2	L3	L5	L4	[0.964]
L5	L4	L3	L2	L3	[0.170]
L5	L4	L3	L4	L4	[0.057]
L5	L4	L3	L5	L5	[0.399]

ECM-tra: 24962 / ECM-tst: 28226

Figure 6: Rule set of the linguistic model obtained from WALM-CC for the maintenance costs of medium voltage lines estimation problem.

Figure 6 represents the decision table of the model obtained from WALM-CC. In this case, each row represents a fuzzy subspace and contains its associated output consequent(s) —the primary and/or the secondary in importance—, i.e., the correspondent label(s) together with its(their) respective rounded rule weight(s). Once again, the *absolute importance weight* for each fuzzy rule has been graphically showed by means of the grey colour scale, from black (1.0) to white (0.0).

From the 625 (5^4) possible fuzzy rules, the obtained linguistic model is composed of only 54 fuzzy rules. In this case, it only contains five double-consequent rules. Notice that, all the double-consequent rules are very near in the four-dimensional space, representing a zone with higher complexity. Moreover, rules with weights near to 1 represents groups of important rules and do not usually appear alone.

6 Concluding Remarks

In this paper, a new linguistic model structure using weighted double-consequent fuzzy rules has been proposed with the aim of improving the performance of the so obtained models maintaining an acceptable description level. Its main interest lies in flexibilyzing the model structure in a different way from the usual one (e.g., learning the composition of the fuzzy membership functions).

A simple GA could be proposed as a first approximation to learn these kinds of improved linguistic models, but the results obtained in a preliminary experimentation were not very satisfactory. However, since the search space is too large and the problem easily decomposable, we propose in his stead the use of a more sophisticated genetic model able to solve decomposable complex problems, the cooperative coevolution. The very accurate results of the proposed coevolutive learning method, compared with other related approaches, has been contrasted when solving two electrical distribution problems.

Moreover, the improved linguistic models so obtained have presented a good description level. In this way, models with no more than five double-consequent rules were obtained, only considering the use of these kinds of rules when it was really needed. Moreover, significant rules have been identified studying the weights of the rules, helping us to interpret the model behavior.

A Acronyms

Table 3 presents the list of acronyms considered in this paper.

Table 3: Acronyms considered in the paper

Acronym - Meaning	**Acronym - Meaning**
ALM - Accurate Linguistic Modeling	*MSE - Mean Square Error*
FRBSs - Fuzzy Rule Based Systems	*WALM - Weighted Accurate Linguistic Modeling algorithm*
GAs - Genetic Algorithm	*WALM-CC - WALM algorithm with Cooperative Coevolution*
GFSs - Genetic Fuzzy Systems	*WM - Wang and Mendel algorithm*
LM - Linguistic Modeling	*WRL - Weighted Rule Learning*

Acknowledgements

Oscar Cordón would like to express his gratitude towards Professor Lotfi A. Zadeh for the hospitality, support and inspiration he encountered during his stay with the Berkeley Initiative in Soft Computing (BISC) at the University of California-Berkeley in 1999.

References

[1] A. Bastian, How to handle the flexibility of linguistic variables with applications, International Journal of Uncertainty, Fuzziness and Knowlegde-Based Systems 2:4 (1994) 463–484.

[2] J.E. Baker, Reducing bias and inefficiency in the selection algorithm, in: J.J. Grefenstette (Ed.), Proceedings of the 2nd International Conference on Genetic Algorithms, Lawrence Erlbaum (Hillsdale, NJ, USA, 1987) 14–21.

[3] J. Casillas, O. Cordón, F. Herrera, Can linguistic modeling be as accurate as fuzzy modeling without losing its description to a high degree?, Technical Report #DECSAI-00-01-20, Department of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain, 2000.

[4] J.-S. Cho, D.-J. Park, Novel fuzzy logic control based on weighting of partially inconsistent rules using neural network, Journal of Intelligent and Fuzzy Systems 8 (2000) 99–110.

[5] O. Cordón, F. Herrera, A proposal for improving the accuracy of linguistic modeling, IEEE Transaction on Fuzzy Systems 8:3 (2000) 335–344.

[6] O. Cordón, F. Herrera, F. Hoffmann, L. Magdalena, Genetic fuzzy systems: evolutionary tuning and learning of fuzzy knowledge bases (World Scientific, Singapore, 2001).

[7] O. Cordón, F. Herrera, A. Peregrín, Applicability of the fuzzy operators in the design of fuzzy logic controllers, Fuzzy Sets and Systems 86:1 (1997) 15–41.

[8] O. Cordón, F. Herrera, L. Sánchez, Solving electrical distribution problems using hybrid evolutionary data analysis techniques, Applied Intelligence 10 (1999) 5–24.

[9] D. Goldberg, A meditation on the computational intelligence and its future, Illigal Report #2000019, Department of General Engineering, University of Illinois at Urbana-Champaign, Foreword Proceedings of the 2000 International Symposium on Computational Intelligence.

[10] F. Herrera, M. Lozano, J.L. Verdegay, A learning process for fuzzy control rules using genetic algorithms, Fuzzy Sets and Systems 100 (1998) 143–158.

[11] J.H. Holland, Adaptation in natural and artificial systems (Ann arbor: The University of Michigan Press, 1975).

[12] E.H. Mamdani, Applications of fuzzy algorithm for control a simple dynamic plant, Proceedings of the IEEE 121 (1974) 1585–1588.

[13] E.H. Mamdani, S. Assilian, An experiment in linguistic synthesis with a fuzzy logic controller, International Journal of Man-Machine Studies 7 (1975) 1–13.

[14] Z. Michalewicz, Genetic algorithms + data structures = evolution programs, Springer-Verlag, Heidelberg, Germany, 1996.

[15] K. Nozaki, H. Ishibuchi, H. Tanaka, A simple but powerful heuristic method for generating fuzzy rules from numerical data, Fuzzy Sets and Systems 86 (1997) 251-270.

[16] N.R. Pal, K. Pal, Handling of inconsistent rules with an extended model of fuzzy reasoning, Journal of Intelligent and Fuzzy Systems. 7 (1999) 55–73.

[17] J. Paredis, Coevolutionary computation, Artificial Life 2 (1995) 355–375.

[18] W. Pedrycz (ed.), Fuzzy Modelling. Paradigms and Practice, (Kluwer Academic Press, 1996).

[19] C. A. Peña-Reyes, M. Sipper, Fuzzy CoCo: a cooperative coevolutionary approach to fuzzy modeling, IEEE Transactions on Fuzzy Systems 9:5 (2001) 727–737.

[20] M.A. Potter, K.A. De Jong, Cooperative coevolution: an architecture for evolving coadapted subcomponents, Evolutionary Computation 8:1 (2000) 1–29.

[21] C.D. Rosin, R.K. Belew, New methods for competitive coevolution, Evolutionary Computation 5:1 (1997) 1–29.

[22] M. Sugeno, T. Yasukawa, A fuzzy-logic-based approach to qualitative modeling, IEEE Transactions on Fuzzy Systems 1 (1993) 7–31.

[23] L.X. Wang, J.M. Mendel, Generating fuzzy rules by learning from examples, IEEE Transactions on Systems, Man, and Cybernetics 22 (1992) 1414–1427.

[24] D. Whitley, J. Kauth, GENITOR: A different genetic algorithm, Proceedings of the Rocky Mountain Conference on Artificial Intelligence, Denver, USA (1988) 118–130.

[25] W. Yu, Z. Bien, Design of fuzzy logic controller with inconsistent rule base, Journal of Intelligent and Fuzzy Systems 2 (1994) 147–159.

[26] L.A. Zadeh, Outline of a new approach to the analysis of complex systems and decision processes, IEEE Transactions on Systems, Man, and Cybernetics 3 (1973) 28–44.

[27] L.A. Zadeh, The concept of a linguistic variable and its applications to approximate reasoning, Information Science, Part I: 8 (1975) 199–249; Part II: 8 (1975) 301–357; Part III: 9 (1975) 43–80.