Improving the Wang and Mendel's Fuzzy Rule Learning Method by Inducing Cooperation Among Rules [1]

J. Casillas
DECSAI, University of Granada
18071 – Granada, Spain
casillas@decsai.ugr.es

O. Cordón
DECSAI, University of Granada
18071 – Granada, Spain
ocordon@decsai.ugr.es

F. Herrera
DECSAI, University of Granada
18071 – Granada, Spain
herrera@decsai.ugr.es

Abstract

Nowadays, Linguistic Modeling (LM) is considered to be one of the most important areas of application for Fuzzy Logic. It is accomplished by descriptive Fuzzy Rule-Based Systems (FRBSs), whose most interesting feature is the interpolative reasoning they develop. This characteristic plays a key role in the high performance of FRBSs and is a consequence of the cooperation among the fuzzy rules involved in the FRBS.

A large quantity of automatic techniques has been proposed to generate these fuzzy rules from numerical data. One of the most interesting families of techniques, due to its simplicity and quickness, is the ad hoc data-driven methods. However, its main drawback is the cooperation among the rules which is not suitably considered.

With the aim of facing up this drawback, which makes the obtained models not to be as accurate as desired, a new approach to improve the performance obtaining more cooperative rules is introduced in this paper. Following this approach, a concrete LM method based on one of the most known and widely used ad hoc data-driven methods, the Wang and Mendel's method, is also presented. Its operation mode is composed of two stages: generation of the candidate rule set and combinatorial search of the rule set with best cooperation. Its behavior in the solving of two different applications will also be shown.

Keywords: Linguistic system models, fuzzy rule-based systems, learning, cooperative linguistic rules, simulated annealing

[1]This research has been supported by CICYT under project PB98-1319.

1 Introduction

At present, one of the most important areas for the application of Fuzzy Set Theory as developed by Zadeh in 1965 [17] are Fuzzy Rule-Based Systems (FRBSs). These kinds of systems constitute an extension of classical Rule-Based Systems, because they deal with fuzzy rules instead of classical logic rules. Thanks to this, they have been successfully applied to a wide range of problems from different areas presenting uncertainty and vagueness in different ways [1, 6, 10, 16]. One of the most important applications of FRBSs is system modeling [1, 10], 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 [11]. In this framework, one of the most important areas is Linguistic Modeling (LM), where the interpretability of the obtained model is the main requirement.

A typical FRBS is composed of two main components: the *Knowledge Base* (KB) and the *Inference System*. The KB —composed of Data Base (DB) and Rule Base (RB)— stores the available knowledge about the problem in the form of linguistic *"IF-THEN"* rules. The Inference System puts into effect the inference process on the system inputs making use of the information stored in the KB.

Several tasks have to be performed in order to design an FRBS for a concrete application. One of the most important and difficult ones is to obtain an appropriate KB about the problem being solved. The difficulty presented by the human experts to express their knowledge in the form of fuzzy rules has made researchers develop automatic techniques for performing this task.

In this sense, a large quantity of methods has been proposed for automatically generating fuzzy rules from numerical data. Usually they make use of complex rule generation mechanisms such as neural networks [7, 8], genetic algorithms [3, 4, 12], fuzzy clustering [2, 15], etc. Opposite to them, a family of efficient and simple methods, called "*Ad Hoc Data-Driven Methods*", has been proposed in the literature [1, 9, 14]. Under this name, those methods are collected which are based on processes where the learning of the fuzzy rules is guided by covering criteria of the data in the example set. The way to address this process can not be included in any well-known optimization or searching technique but it is specifically developed for this purpose. They are characterized by being methods based on a short time-consuming iterative procedure.

The advantages of the ad hoc data-driven methods are clear. On the one hand, their simplicity makes them be easily understandable and implementable. On the other hand, and thanks to its quickness, they are very suitable to be used as a first stage of the modeling process to obtain a preliminary fuzzy model, which can be subsequently refined by other methods.

One of the most known and widely used ad hoc data-driven methods is the Wang and Mendel's method (WM-method) [14]. Although its high performance has been clearly demonstrated, it has a problem related to the way of selecting the rules: the method looks for the rules with the best individual performance. This sometimes causes KBs with a low degree of accuracy to be obtained that, due to the interpolative reasoning developed by FRBSs, make the results not to be as accurate as desired.

In order to face up this problem, we propose a modification of the WM-method to improve the rule cooperation and thus the accuracy of the obtained models. Basically, our proposal involves performing a search in the set of candidate rules to find the combination of them that obtain the best accuracy jointly, even though the rules are not individually the best in their fuzzy input subspaces.

To do so, the paper is organized as follows. In Section 2, the WM-method will be explained and its drawbacks will be analyzed. Section 3 will show the approach proposed to improve the accuracy by means of more cooperative rules. In Section 4, a method based on the WM-method following the cooperative approach will be proposed to face up the mentioned drawbacks. In Section 5, the behavior of the new process, compared with WM-method and Nozaki et al.'s method [9], will be analyzed in two different applications, the fuzzy modeling of a simple three-dimensional function and the problem of rice taste evaluation. Finally, in Section 6 some concluding remarks will be pointed out.

2 Wang and Mendel's Method

The ad hoc data-driven RB generation process proposed by Wang and Mendel in [14] has been widely known because of its simplicity and good performance. It is based on working with an input-output data set $E = \{e_1, \ldots, e_p\}$, $e_l = (x_1^l, \ldots, x_n^l, y^l)$, representing the behavior of the problem being solved, using a previous definition of the DB composed of the input and output primary fuzzy partitions.

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 available) 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. In our case, we will work with symmetrical fuzzy partitions of triangular membership functions (see Figure 1).

2. *Generate a candidate linguistic rule set*: This set will be formed by the rule best covering each example (input-output data pair) contained in E. Thus, p candidate linguistic rules will be obtained. 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

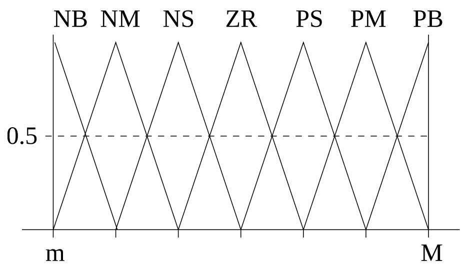

Figure 1: Graphical representation of a uniform fuzzy partition

values), and setting each one of the variables to the linguistic label (associated 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, $l = 1, \ldots, p$. 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 candidate linguistic rule set*: To do so, the p candidate rules are first grouped in g different groups, each one of them composed of all the candidate rules presenting the same antecedent. We will note by R_{ij} the j-th rule in the i-th group. To compose the final RB, the rule with the highest importance degree is chosen in each group i, $i = 1, \ldots, g$. Hence, g will be both the number of different antecedent combinations in the candidate rule set and the number of linguistic rules in the final RB generated.

The good behavior of the WM-method has been clearly demonstrated. However, sometimes the method does not perform as good as desired. It is due to a problem related to the way in which the rules are selected.

One of the most interesting features of a 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 it is known, the output obtained from a 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 system input to any degree.

However, the operation mode followed by WM-method is to bracket the example data set into fuzzy subspaces (the antecedent combinations mentioned in step 4 of the algorithm) according to the covering degree, and to obtain afterwards the rule with the best performance in each subspace. Therefore, the global interaction among the rules of the KB is not considered. This causes the finally obtained rule set, in spite of presenting a good local behavior, not to cooperate suitably. Moreover, the fact of locally processing these rules makes the method be more sensitive to noise.

A proposal modifying the WM-method to include cooperation in the generated RB will be presented in Section 4. Previously, the approach followed by this proposal to induce cooperation among the fuzzy rules will be explained in the next Section.

3 A New Approach to Improve the Accuracy Obtaining More Cooperative Rules

With the aim of addressing the drawbacks presented by the WM-method, a new approach to improve the accuracy obtaining best cooperation among the rules is proposed: the Cooperative Rules (COR) approach.

To do so, a *Combinatorial Search of Cooperative Rules* is performed over the set of candidate rules to find the rule set best cooperating. Instead of selecting the consequent with the highest importance degree in each subspace like WM-method does, the COR approach considers the possibility of using another consequent different from the best one, when it allows the FRBS to be more accurate thanks to having a KB with best cooperation. To do so, COR performs a combinatorial search among the candidate rules looking for the set of consequents with the best accuracy.

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

- The first one accomplishes a full search through the set of possible combinations. Although with this technique we ensure the obtaining of the best solution, it may take a long time or simply be unapproachable in terms of run time, when there is a great number of combinations. Therefore, this technique is only recommended in confined spaces.

- On the other hand, when the use of an explicit enumeration is not possible, an approximate technique is needed. Any search technique can be used. However, since one of the main advantages of the ad hoc data-driven methods is their ability to find good fuzzy models quickly, the search technique should also be quick. We propose to use the Simulated Annealing (SA) technique since it is easily implementable and it finds good solutions quickly.

  SA is a numerical optimization technique based on the analogy with the physical annealing process of solids [13]. The SA-based algorithm begins with an initial solution and generates a neighbor of this solution by means of a suitable mechanism. If the adaptation of the latter is better than the former, the current solution is replaced by the generated neighbor; otherwise, this replacement is accomplished with a specific probability that will be decreased during the algorithm progress. This process is iterated a large number of times.

Following the COR approach, an specific method based on the WM-method will be presented in the next Section.

4 A Proposal Based on the Wang and Mendel's Method Following the Cooperative Rules Approach

In this Section, a modification of WM-method following the presented philosophy is introduced: the WM-based COR-method (COR-WM-method). It is composed of two stages: in the first one, the set of candidate rules is obtained (in the usual WM-method way); in the second one, the set of rules with the best cooperation is obtained by means of a SA technique.

The algorithm is made up of the following steps:

1. *Consider a fuzzy partition of the variable spaces* obtained from the expert information (if it is available) or by a normalization process, as shown in step 1 of the WM-method.

2. *Generate a candidate linguistic rule set from the examples* like in step 2 of the WM-method.

3. *Group the obtained rules according to the antecedent combinations*, which is equivalent to applying the first part of step 4 in the WM-method, *and select the possible consequent values for each one of them.* To do so, the consequents B^{ij} of the candidate linguistic rules R_{ij} are analyzed and g different possible consequent term sets S_i, $i = 1, \ldots, g$, are obtained by taking only the different ones in each group. From now on, the l-th member of S_i will be noted by c^i_l.

4. *Run the SA-based algorithm to look for the combination* $\{c^1_{l_1}, \ldots, c^g_{l_g}\}$ *with the best degree of accuracy.* The initial solution is obtained by generating a possible combination at random. The neighbor generation mechanism randomly selects a specific $s \in \{1, \ldots, g\}$ and changes $c^s_{l_s}$ by $c^s_{l'_s} = \{c^s_1, \ldots, c^s_{l_s-1}, c^s_{l_s+1}, \ldots, c^s_{|S_s|}\}$. To evaluate the quality of each solution, an index measuring the cooperation degree of the encoded rule set is considered. In this case, the algorithm uses a global error function called *mean square error* (MSE), which defined as

$$MSE = \frac{1}{2 \cdot N} \sum_{i=1}^{N} (y'_i - y_i)^2 \ ,$$

   with N being the data set size, y' being the output obtained from the FRBS, and y being the known desired output. The closer to zero the measure is, the greater the performance is and, thus, the better the rule cooperation is.

5 Experimental Study

In order to analyze the behavior of the proposed process, we have chosen two different applications: the fuzzy modeling of a three-dimensional function, and the problem of rice taste evaluation, proposed in [9]. In both cases, we will compare the accuracy of the linguistic models generated from our process with the ones designed by means of the WM-method and the Nozaki et al.'s method (N-method) [9].

An initial DB constituted by a primary fuzzy partition for each variable will be considered in each case. Every partition is formed by *seven linguistic terms* for the three-dimensional function, and *three linguistic terms* for the rice taste evaluation problem, with triangular-shaped equally distributed fuzzy sets giving meaning to them (as shown in Figure 1), and the appropriate scaling factors to translate the generic universe of discourse into the one associated with each problem variable.

As regards the FRBS reasoning method used, we have selected the *minimum t-norm* playing the role of the implication and conjunctive operators, and the *center of gravity weighted by the matching* strategy acting as the defuzzification operator [5].

5.1 Fuzzy modeling of a simple three-dimensional function

For this first experiment, a simple unimodal three-dimensional mathematical function (see Figure 2) is considered to be modeled. The mathematical function and the variable universes of discourse are the following:

$$F(x_1, x_2) = x_1^2 + x_2^2$$
$$x_1, x_2 \in [-5, 5], \quad F(x_1, x_2) \in [0, 50] \ .$$

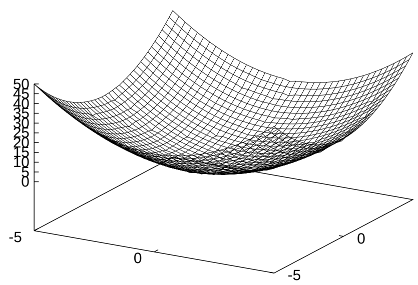

Figure 2: Graphical representation of the three-dimensional function F

A training data set uniformly distributed in the three-dimensional definition space has been obtained experimentally. In this way, a set with 1681 values has been generated taking 41 values for each one of the two input variables considered to be uniformly distributed in their intervals. Another data set has been generated for its use as test set to evaluate the performance of the learning method, avoiding any possible bias related to the data in the training set. The size of this data set (168) is ten percent of the training set one. The data is obtained generating the input variable values at random in the concrete universes of discourse for each of them, and computing the associated output variable value.

The results obtained by the three methods considered are collected in Table 1, where #R stands for the number of rules, and MSE_{tra} and MSE_{tst} for the values obtained over the training and test data sets respectively. The best results are shown in boldface.

Table 1: Results obtained modeling F

Method	**#R**	**MSE$_{tra}$**	**MSE$_{tst}$**
WM-method	**49**	2.048137	2.255928
N-method	98	2.465487	1.768125
COR-WM-method	**49**	**1.605482**	**1.132797**

Analyzing these results, we may note the good behavior presented by our COR-WM-method. The linguistic model generated from it clearly outperforms the ones designed by the other two methods. Compared with WM-method, a significantly more accurate model is obtained thanks to the cooperative rule consideration. Of course, the same number of rules is generated. Opposite to N-method, moreover of improving the performance to a high degree, the model obtained by our process presents a very much simpler KB (49 rules against 98), which is a very important aspect in LM where the interpretability is the main requirement.

5.2 Rice taste evaluation

Subjective qualification of food taste is a very important but difficult problem. In the case of the rice taste qualification, it is usually put into effect by means of a subjective evaluation called the *sensory test*. In this test, a group of experts, usually composed of 24 persons, evaluate the rice according to a set of characteristics associated to it. These factors are: *flavor*, *appearance*, *taste*, *stickiness*, and *toughness* [9].

Because of the large quantity of relevant variables, the problem of rice taste analysis becomes very complex, thus leading to solve it by means of modeling techniques capable of obtaining a model representing the non-linear relationships existing in it. Moreover, the problem-solving goal is not only to obtain an accurate model, but to obtain a user-interpretable model as well, capable of putting some light on the reasoning process performed by the expert for evaluating a kind of rice in a specific way. Due to all these reasons, in this Section we deal with obtaining a linguistic model to solve the said problem.

In order to do so, we are going to use the data set presented in [9]. This set is composed of 105 data arrays collecting subjective evaluations of the six variables in question (the five mentioned and the overall evaluation of the kind of rice), made up by experts on this number of kinds of rice grown in Japan (for example, Sasanishiki, Akita-Komachi, etc.). The six variables are normalized, thus taking values in the real interval [0,1].

With the aim of not biasing the learning, we have randomly obtained ten different partitions of the mentioned set, composed by 75 pieces of data in the training set and 30 in the test one, to generate ten linguistic models in each experiment.

Table 2 shows the results obtained modeling this application. The values shown in columns MSE_{tra} and MSE_{tst} have been computed as an average of the MSE values obtained in the approximation of the training and test data sets, respectively, by the ten linguistic models generated in each case. The column #R stands for the average number of linguistic rules in the KBs of the ten models generated by each process.

Table 2: Results obtained in the rice taste evaluation

Method	**#R**	$\mathbf{MSE}_{tra}$	$\mathbf{MSE}_{tst}$
WM-method	**23**	0.003339	0.003758
N-method	364.8	**0.002512**	**0.003221**
COR-WM-method	**23**	0.003040	0.003484

In view of the obtained results, we may again note the good behavior presented by the proposed COR-WM-method. Its linguistic model again presents a more accurate behavior than the WM-method's one. Nevertheless, the performance of our model is worse than the one obtained by N-method. However, in exchange of losing a little bit of accuracy, our process obtains a much lesser number of rules than N-method, which significantly improves the interpretability.

We should say that due to the small size of the example set, a reduced number of combinations (8.4 on average for the ten partitions) is generated in the first stage of the COR-WM-method. Therefore, the explicit enumeration has been used in this case instead of the SA procedure since the best solution can be quickly found.

6 Concluding Remarks

In this paper, a new approach to generate accurate and simple linguistic models has been proposed, the COR approach. It is based on considering the cooperation among the fuzzy rules in the generation process making good use of the interpolative reasoning developed by the finally designed FRBS.

Following this approach, a specific method modifying the WM-method has also been presented. Its operation mode is composed of two stages: generation of the candidate rule set and combinatorial search of the rule set with the best cooperation. Its behavior has been compared to other two LM methods when solving two different problems. The proposed process has obtained very good results combining accuracy and interpretability.

This leads us to conclude that the consideration of cooperative rules improves the performance of the linguistic models and the derivation of KBs by firstly generating a candidate rule set and then searching the best combination of rules is a good way to accomplish this aspect.

References

[1] A. Bárdossy, L. Duckstein, Fuzzy rule-based modeling with application to geophysical, biological and engineering systems (CRC Press, 1995).

[2] S.L. Chiu, Fuzzy model identification based on cluster estimation, *Journal of Intelligent and Fuzzy Systems* **2** (1994) 267–278.

[3] O. Cordón, F. Herrera, A three-stage evolutionary process for learning descriptive and approximate fuzzy logic controller knowledge bases from examples, *Int. Journal of Approximate Reasoning* **17**:4 (1997) 369–407.

[4] O. Cordón, F. Herrera, Hybridizing genetic algorithms with sharing scheme and evolution strategies for designing approximate fuzzy rule-based systems, *Fuzzy Sets and Systems* (to appear).

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

[6] K. Hirota (Ed.), Industrial applications of fuzzy technology (Springer-Verlag, 1993).

[7] J.-S.R. Jang, ANFIS: Adaptive-Network-Based Fuzzy Inference System, *IEEE Transactions on Systems, Man, and Cybernetics* **23**:3 (1993) 665–685.

[8] D. Nauck, R. Kruse, Neuro-fuzzy systems for function approximation, *Fuzzy Sets and Systems* **101** (1999) 261–271.

[9] 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.

[10] W. Pedrycz (Ed.), Fuzzy modelling. Paradigms and practice (Kluwer Academic Press, 1996).

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

[12] P. Thrift, Fuzzy logic synthesis with genetic algorithms, *Proc. of the 4th Int. Conf. on Genetic Algorithms*, 1991, pp. 509–513.

[13] P.J.M. van Laarhoven, E.H.L. Aarts, Simulated annealing: theory and applications (Kluwer Academic, 1987).

[14] L.X. Wang, J.M. Mendel, Generating fuzzy rules by learning from examples, *IEEE Transactions on Systems, Man, and Cybernetics* **22**:6 (1992) 1414-1427.

[15] R.R. Yager, D.P. Filev, Generation of fuzzy rules by mountain clustering, *Journal of Intelligent and Fuzzy Systems* **2** (1994) 209–219.

[16] R.R. Yager, L.A. Zadeh (Eds.), An introduction to fuzzy logic applications in intelligent systems (Kluwer Academic Press, 1992)

[17] L.A. Zadeh, Fuzzy sets, *Information and Control* **8** (1965) 338–353.