

Relevance and consistency in rule-based systems

Witold Pedrycz* and George Vukovich**

*Department of Electrical & Computer Engineering
University of Alberta, Edmonton, Canada
(pedrycz@ee.ualberta.ca)

and

Systems Research Institute, Polish Academy of Sciences
01-447 Warsaw, Poland

and

**Canadian Space Agency, Spacecraft Engineering
6767 Route de l'Aéroport
Saint-Hubert, Quebec J3Y 8Y9, Canada

Abstract *We discuss a problem of synthesis and analysis of rules based on experimental numeric data. Two descriptors of the rules being viewed individually and en block are introduced. The relevance of the rules is quantified in terms of the data being covered by the antecedents and conclusions standing in the rule. While this index describes each rule individually, the consistency of the rule deals with the quality of the rule viewed vis-à-vis other rules. It expresses how much the rule “interacts” with others in the sense that its conclusion is distorted by the conclusion parts coming from other rules. We show how the rules are formed by means of fuzzy clustering and their quality evaluated by means of the above indexes. Global characteristics of a set of rules are also discussed and related to the number of information granules formed in the space of antecedents and conclusions.*

1. Introduction

Rule-based systems are granular and highly modular models [1][4][5][6][7][8][9][11][12][13][14]. The granularity of rules becomes fully reflected in the form of the antecedents and conclusions and is quantified in the language of sets, fuzzy sets, rough sets, probability, to name the main formal vehicles. The evident transparency of these systems is their genuine asset. Several design issues remain still open and tend to become even more profound as we move toward developing larger systems

- ✓ The origin of the rules and their eventual quantification (e.g., confidence or relevance of relationships where the confidence measure is expressed in relation to experimental data)
- ✓ The dimensionality problem becomes of concern when the number of variables in the rules increases (then the number of rules tend to explode at the exponential rate)

Fuzzy clustering gives rise to fuzzy sets and fuzzy relations being thus constructed in a fully algorithmic fashion. They are examples of information granules regarded as a basic building canvass of numerous fuzzy models. For given collections of such information granules, studying and quantifying relationships between them leads to the emergence of the rule-based models. The preliminary development step of this nature is referred to as a link analysis – a phase in which we reveal and quantify dependencies between information granules before proceeding with further detailed construction and refinement of the rule-based architecture.

More descriptively, one can consider the outcome of the link analysis to arise as a web of connections between the granules. As fundamentally we distinguish between input and output variable(s), there are two essential facets of the analysis that deal with the following aspects

- (a) expressing strength between the granules in the input and output space (relevance of the rule), and
- (b) completing a consistency (crosstalk) analysis in which we quantify an interaction between the given link and the links that are invoked owing to the interaction (overlap) between the information granules. In the language of the calculus of rule-based systems, the notion of

relevance is linked with the notion of strength of the rule whereas the second aspect of consistency is concerned with an interaction between the information granules.

The material is arranged into 6 sections. First, in Section 2 we discuss clustering as a basic means of information granulation and pose the problem of constructing rules. Section 3 covers a discussion of two descriptors of the rules, namely rule relevance and rule consistency. Experimental studies are covered in Section 4 while conclusions are included in Section 5.

2. Clustering as a vehicle of information granulation and a quest for a web of information granules

Quite often, the Fuzzy C-Means (FCM) algorithm arises as a basic vehicle of data granulation. As the method is well-known in the literature, cf. [2], we will not discuss it here but rather clarify the notation and cast the method in the setting of the problem at hand.

The input and output data spaces are denoted by \mathbf{X} and \mathbf{Y} , respectively. The numeric data set under discussion assumes a form of input – output pairs $(\mathbf{x}(k), \mathbf{y}(k))$, $k=1,2,\dots,N$ where $\mathbf{x}_k \in \mathbf{R}^n$ and $\mathbf{y}_k \in \mathbf{R}^m$. The results of clustering carried out for \mathbf{X} and \mathbf{Y} separately are discrete fuzzy sets (more precisely, fuzzy relations) defined over these data sets. In general, we end up with “c[1]” fuzzy relations in \mathbf{X} , say $A_1, A_2, \dots, A_{c[1]}$ and “c[2]” fuzzy relations in \mathbf{Y} , namely $B_1, B_2, \dots, B_{c[2]}$. Technically, they are organized in a form of partition matrices so we have

$$U[1] = \begin{bmatrix} A_1 \\ A_2 \\ \dots \\ A_{c[1]} \end{bmatrix} \quad U[2] = \begin{bmatrix} B_1 \\ B_2 \\ \dots \\ B_{c[2]} \end{bmatrix}$$

Our ultimate goal is a formation of a thorough and constructive description of a web of directed links between A_i and B_j . In a nutshell, such task gives rise to a collection of rules

$$- \quad \text{if } A_i \text{ then } B_j$$

The techniques of fuzzy clustering, no matter what type of objective function we employ, share several main features whose understanding is of importance in the framework of this problem. Clusters are *direction-free* (relational) constructs: in the clustering process there are no provisions as to a possible direction between the variables. This implies that the resulting granules do not accommodate any request that relates to the mapping itself and are not formed in a way they reflect a directionality component. In contrast, any mapping (function) is a *directional* construct. If we establish a link between A_i and B_j , its nature needs to be inspected with respect to the directionality of the rule obtained in this manner. More formally, if there is no directionality component (or it is ignored), we are essentially assessing properties of a Cartesian product of the two information granules, namely $A_i \times B_j$

Fuzzy clusters (information granules) are sound building blocks of rule-based systems as they provide an answer to the two challenges outlined at the beginning of the previous section. First, they reflect the structure of the data so we anticipate that any cluster comes with enough experimental evidence behind it (otherwise it would not have been formed in the first place). Second, as we are concerned with fuzzy relations rather than fuzzy sets, the dimensionality problem does not arise. For instance if $c[1]=c[2]=c$ then we have (potentially) “c” meaningful rules in spite of the potentially high dimensionality of the input and output space. Otherwise, as it happens quite often when dealing with the individual fuzzy sets, we end up with a combinatorial explosion of the collection of the rules produced at this level.

In what follows, by proceeding with information granules formed by clustering we synthesize a collection of rules and analyze their properties. The crux of this development is as follows. With each A_i $i=1,2,\dots,c[1]$ we associate a single granule in the conclusion space, that is we form a mapping

$$A_i \rightarrow B_j$$

(1)

(the way of determining the associations will be discussed later on). Following this procedure, we build a collection of rules, \mathcal{R} , where its cardinality is equal to $c[1]$. In a concise way, we can summarize the rules as a list of pairs of indexes of the respective information granules

$$(1, j_1) (2, j_2) \dots (c[1], j_{c[1]})$$

where j_k is in the range of integers from 1 to $c[2]$.

The characteristics of this form of assignment between the information granules are quantified in the next Section.

3. Characteristics of rules

There are two main descriptors of the rules. The first one reflects the experimental evidence behind the rule (association). The second captures the relationships between the rules discussed as a collection of entities.

3.1. Relevance of the rule

Let us view a certain rule as a Cartesian product) meaning that we do not consider the “direction” of the rule but look at it as an entity linking two information granules defined in the two different spaces (input-output). The rule (1) comes with the (experimental) relevance equal to

$$\text{rel}(A_i \times B_j) = \sum_{k=1}^N A_i(x_k) \text{t} B_j(y_k) \quad (2)$$

where “t” is a t-norm viewed here as a model of an *and* logical connective. If $\text{rel}(\)$ attains higher values, we say that the rule comes with more experimental relevance (in other words, it is more justifiable from the experimental point of view). For fixed “i”, we order all associations (rules) $A_i \rightarrow B_1, A_i \rightarrow B_2, \dots, A_i \rightarrow B_{c[2]}$ according to the associated relevance level. The highest value of the relevance identifies a rule of the form

$$A_i \rightarrow B_{j(i)}.$$

In light of the above realization (properties of t-norms), we come up with a straightforward monotonicity property, namely: If $A_i \subseteq A'_i$ and $B_j \subseteq B'_j$ then $\text{rel}(A_i \times B_j) \leq \text{rel}(A'_i \times B'_j)$. Intuitively, note that if we increase the size of the information granules, this change contributes to the increasing level of relevance of the particular rule as in this way we tend to “cover” more data and thus elevate experimental evidence of this rule.

The relevance defined above exhibits a close analogy to the notion of rule support encountered in data mining that is articulated in the language of probability theory and reads in the form

$$\text{Support}(A \rightarrow B) = \text{Prob}(A \times B)$$

3.2. Consistency of the rule

The associations we have constructed so far were totally isolated. We have not expressed and quantified possible interactions between the rules. Nevertheless, this interaction *does* exist owing to the nature of the overlapping fuzzy sets (relations). Considering two rules, we note that their antecedents and conclusions overlap to a certain degree. The overlap at the condition end could be very different than the one encountered at the conclusion end. The differences between the overlap levels shed light on an important issue of consistency of the rules. In turn, this leads to a detection of conflicting rules (where the term of conflict itself is rather continuous than binary, so we talk about a degree of conflict, or equivalently, a degree of consistency).

The problem of conflicting rules is well known in the literature and has a long path in the research in rule – based systems, especially those in the realm of fuzzy controllers, cf [10]. Bearing in mind the origin of the control knowledge, this effect was attributed mainly to some deficiencies of knowledge acquisition when working with a human expert. As in this study, we are concerned with an automated vehicle of information granulation, the effect of conflict is a result of incompatibility of information granules in the spaces of conditions and conclusions. Before we move on with a detailed quantification of this effect, let us concentrate on Table 1. It summarizes four different scenarios of interaction occurring between two rules $A_1 \rightarrow B_1$ and $A_2 \rightarrow B_2$.

	B ₁ and B ₂ similar	B ₁ and B ₂ different
A ₁ and A ₂ similar	rules are redundant	rules are conflicting
A ₁ and A ₂ different	rules exhibit similar conclusion	rules are different

Table 1. Four possible scenarios of interaction between two rules and an evaluation of their of interaction

Noticeably, there is a single entry in the table that is of significant concern, namely the rules exhibiting very similar conditions (A₁ and A₂ are similar) with different conclusions (B₁ and B₂). What it really means is that almost the *same* condition triggers two different (if not opposite) actions. In all other cases in Table 1, there are no worries from the conceptual standpoint (obviously one may question some of them as to the computational efficiency and structural redundancy of the rulebase).

We develop an index that captures the effect of consistency of two rules. In its construction, we follow the observations coming from Table 1. The consistency measure is developed in two steps by

- ✓ expressing a measure of consistency of the rules for a single data point (x(k), y(k))
- ✓ constructing a global performance measure over all data (X, Y)

Note that the term of similarity is invoked at the level of the information granules rather than original data, so in essence we are looking at A₁(x(k)) and A₂(x(k)) along with B₁(y(k)) and B₂(y(k)). To express a degree of similarity, we use the formula that is deeply rooted in the language of logic and set theory: we say that two sets are equal if the first is included in the another and vice versa. The continuous version of this statement being realized in the framework of membership grades of the respective fuzzy sets reads as

$$A_1(x(k)) \equiv A_2(x(k)) = (A_1(x(k)) \rightarrow A_2(x(k))) \wedge (A_2(x(k)) \rightarrow A_1(x(k))) \quad (3)$$

(note that the implication in logic corresponds to the inclusion operation in set theory). The implication is implemented in the form of the pseudo-residuation operation implied by a certain t-norm, namely

$$a \rightarrow b = \sup\{c \in [0,1] \mid a \wedge c \leq b\}, a, b \in [0,1]$$

The consistency is low only if A₁ and A₂ are similar and B₁ and B₂ are different. This naturally leads to the following expression as a measure of consistency

$$A_1(x(k)) \equiv A_2(x(k)) \rightarrow (B_1(y(k)) \equiv B_2(y(k))) \quad (4)$$

To gain a better insight into the character of this expression, we plot it for the pseudo-residuation generated by a product and minimum, refer to Figure 1.

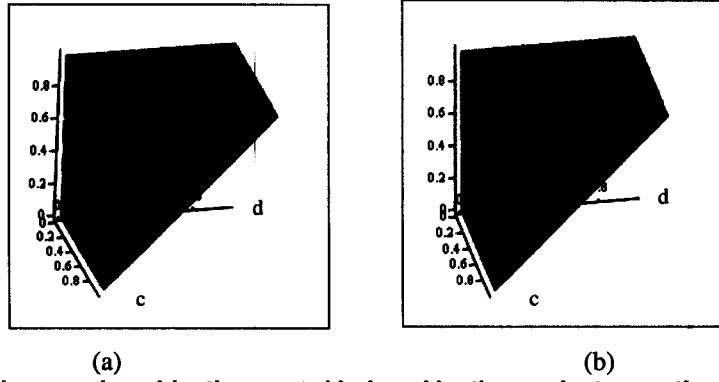


Figure 1. Plot of the pseudo-residualization $c \rightarrow d$ induced by the product operation (a) and minimum (b) (“c” and “d” denote a level of similarity in the space of antecedents and conclusions).

The above expression concerns a single data point. Naturally, a sum over the entire data is a legitimate global measure of consistency of the two rules (rule-1 and rule-2),

$$\text{cons}(1, 2) = \frac{1}{N} \sum_{k=1}^N A_1(\mathbf{x}(k)) \equiv A_2(\mathbf{x}(k)) \rightarrow (B_1(\mathbf{y}(k)) \equiv B_2(\mathbf{y}(k))) \quad (5)$$

Likewise we may like to express a consistency of a given rule versus all other rules in the ruleset \mathcal{R} . This leads to the expression

$$\text{Cons}(i, \mathcal{R}) = \sum_{\substack{j=1 \\ j \neq i}} \text{cons}(i, j) \quad (6)$$

where “i” and “j” are indexes of the rules in \mathcal{R} .

4. Experimental studies

In this section, we are concerned with a synthesis and analysis of fuzzy rules for a selected data set available on the WWW (<ftp://ftp.ics.uci.edu/pub/machine-learning-databases/>), namely fuel consumption. We use the FCM method set up in the same way across all experiments. The fuzzification factor (m) standing in the standard objective function is equal to 2,

$$Q = \sum_{i=1}^c \sum_{k=1}^N u_{ik}^m d_{ik}^2$$

where $c = c[1]$ or $c[2]$.

The dissimilarity between the patterns is expressed in terms of a weighted Euclidean distance. A partition matrix is initialized randomly. Once the information granules (clusters) have been generated, the analysis of the rules is completed in terms of their relevance and consistency. Furthermore we carry out a global analysis as to the number of rules and quantify them as an overall collection and derive characteristics related to their suitability in the description of data. As to the implementation details of t-norms, we use a product operation. The implication is also induced by the same t-norm.

This dataset comes from the StatLib library that is maintained at Carnegie Mellon University. The data concerns a city-cycle fuel consumption that is expressed in miles per gallon and has to be predicted on a basis of 7 attributes (number of cylinders, displacement, horsepower, weight, acceleration, model year, and

origin). All but the fuel consumption are treated as inputs. City-cycle fuel consumption is an output variable (conclusion).

We choose the number of clusters (information granules) to be equal to $c[1] = c[2] = 7$ (this is done for illustrative reasons; we discuss this selection in more detail later on). The relevance of the rules are shown in Figure 2.

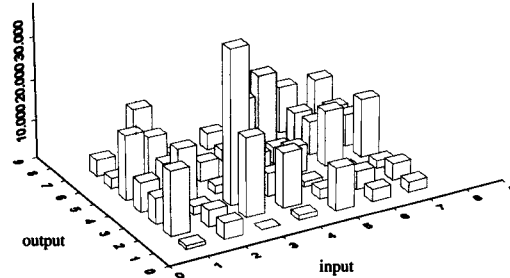


Figure 2. Relevance of the rules

In each row of the matrix, Figure 3, we have at least one fairly dominant association. By selecting the dominant links, we end up with the following seven rules

1 → 5 2 → 7 3 → 3 4 → 5 5 → 6 6 → 4 7 → 4

where the above is a schematic summary of the links (associations) between the information granules in the input and output space. Noticeably all rules are quite similar in terms of their relevance. One may anticipate this to be a result of using clusters as generic building blocks of the rules that comes with a similar experimental evidence behind the fuzzy relations).

The consistency levels of these rules are as follows

rule no.	consistency
1	2.707643
2	2.962191
3	3.455391
4	2.634168
5	2.604724
6	2.667142
7	2.687571

Again, these levels of consistency are fairly similar across all the rules with a single exception (rule 3 → 3 is the most consistent with the corresponding value of the consistency level of 3.45).

Now we investigate a situation where there is a significantly different number of the information granules in the input and output spaces. More specifically, we analyze (a) $c[1] = 3$ and $c[2] = 10$ and (b) $c[1] = 10$ and $c[2] = 3$. The results are visualized in Figure 3 and Tables 2 and 3.

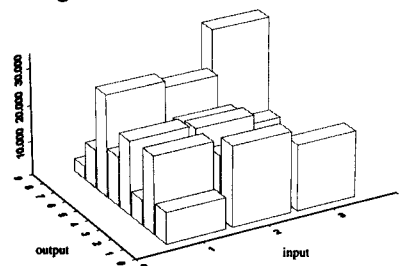


Figure 3. Relevance of the rules for $c[1]=3$ and $c[2]=10$

rule	relevance	consistency
1 → 6	28.63	0.76
2 → 7	23.85	1.02
3 → 8	33.98	1.20

Table 2. Rules (associations) formed by the information granules for $c[1]=3$ and $c[2]=10$

rule	relevance	consistency
1 → 1	39.32	3.26
2 → 3	18.98	2.23
3 → 1	33.16	3.43
4 → 3	18.62	2.27
5 → 1	23.54	2.35
6 → 3	19.16	2.25
7 → 2	26.36	3.54
8 → 3	18.20	2.25
9 → 1	22.02	2.30
10 → 3	22.23	2.31

Table 3. Rules (associations) formed by the information granules for $c[1]=3$ and $c[2]=10$

The most striking is a fact of increasing consistency of the rules with the increasing number of the information granules in the condition space. The consistency goes down significantly when the values of $c[2]$ get lower. The change in $c[2]$ from 10 to 3 decreases the level of consistency by a factor of 2.

In the above series of experiments, the number of the rules is fixed in advance (as we picked up some specific values of $c[1]$ and $c[2]$). Evidently, the number of information granules (both in the input and output space) and a way in which they are distributed across the spaces induce a certain quality of the rules. The performance of the rules can be quantified in the two terms we have already used, namely the relevance and consistency of the rules. These characteristics were computed for some specific configurations of the values of $c[1]$ and $c[2]$. Importantly, these two are crucial to the overall collection of the rules. We complete calculations of the relevance and the consistency of the entire system of rules treating them as a function of $c[1]$ and $c[2]$. The plots of the overall summary of the performance of the rules, $rel(\mathcal{R})$ and $cons(\mathcal{R})$ are shown in Figure 4 and 5, respectively,

$$rel(\mathcal{R}) = \sum_{k:\text{all rules}} rel(rule_k) \quad (7)$$

$$cons(\mathcal{R}) = \sum_{k:\text{all rules}} cons(rule_k) \quad (8)$$

They reveal important an important feature: the total relevance of the rules depends very much on the values of $c[2]$ and it drops down radically once we move beyond a certain number of the information granules. The confidence level of all rules changes radically with respect to $c[1]$.

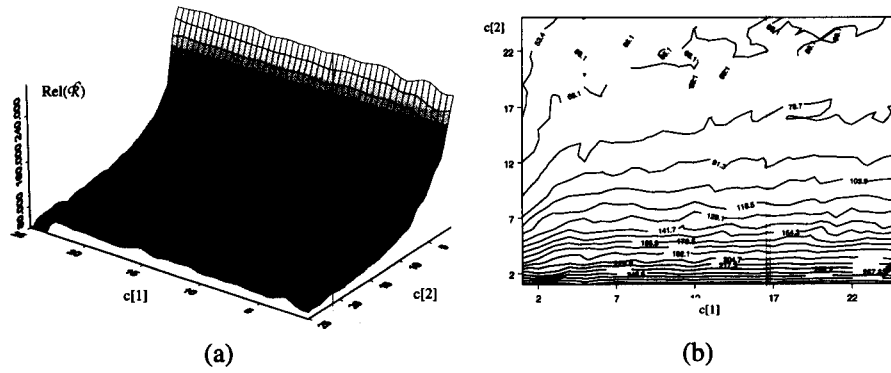


Figure 4. Relevance $rel(\mathcal{R})$ as a function of $c[1]$ and $c[2]$: 3D plot (a) and contour plot (b)

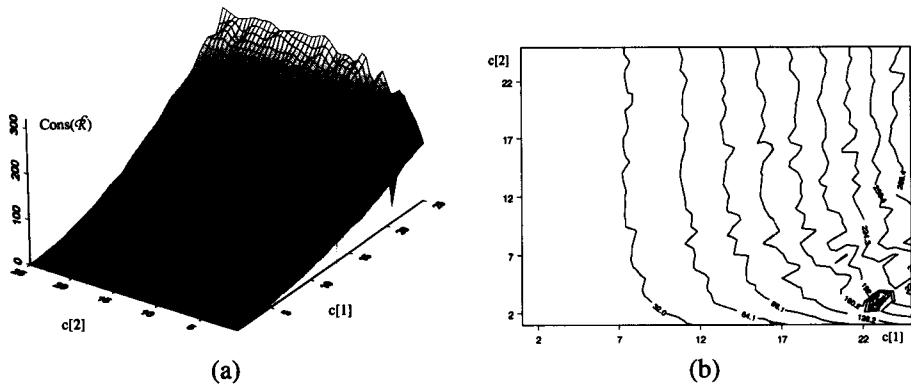


Figure 5. Consistency $cons(\mathcal{R})$ as a function of $c[1]$ and $c[2]$: 3D plot (a) and contour plot (b)

If $c[2]$ is too high, we note a significant drop in the relevance level of the rules. The second number, $c[1]$ does not impact the relevance to the same extent. The situation is different for the consistency of the rules: here $c[1]$ plays a dominant role and too low values have an evident impact on the reduced consistency of the ruleset. The changes in $c[2]$ affect the consistency but the pattern of changes is not too evident. In some region we can see that lower values of $c[2]$ contribute to higher consistency values.

This experimentation leads to some interesting design guidelines of rule-based systems as to the selection of the size of the vocabulary of information granules in the input and output spaces. To assure sufficient relevance and consistency, a feasible region in the $c[1] - c[2]$ plane is delineated in Figure 6. Obviously, some detailed cutoff values of $c[1]$ and $c[2]$ are data-specific yet a general tendency as to the behavior of these two characteristics still holds.

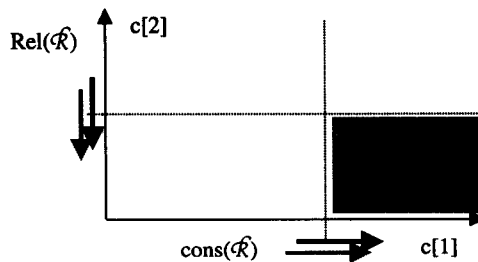


Figure 6. A feasible region of granularity in the $c[1]$ and $c[2]$ plane

5. Conclusions

The study focused on the synthesis of information granules (fuzzy sets and fuzzy relations) and generation of rules (associations) composed of them. Such associations are characterized by two indexes. The first one is about the relevance of the rule and expresses how much experimental evidence is behind the association. The second one is about a directional aspect of the construct and describes much a given rule interacts with all others and produces a crosstalk (relational) effect. We also looked into an interesting numeric quantification of the rules with respect to the size of the vocabulary of the information granules both at the antecedent and conclusion part of the rules. It was revealed that there are some cutoff values of these granules beyond which the quality of the rules drops off significantly.

The above analysis imposes a minimal level of structuralization of the rules. The rules developed here are the direct product of data summarization as we use fuzzy clustering to reveal and capture the structure of the data. It should be stressed that the cluster-based rules help avoid combinatorial explosion in cases of high dimensional spaces. The language of relations (rather than fuzzy sets) becomes instrumental in this setting.

The issue of the numeric quality of the rules has not been discussed at all (namely, a problem of expressing the quality of the rules vis-à-vis the original experimental numeric data). In other words, we have not studied the features of the transformation of the inference results (coming from the rules) into numeric representations. This phase is definitely related with the clustering mechanism itself, the number of clusters in the space of conclusion and a way of aggregation of the conclusions.

6. References

1. A. Bardossy, L. Duckstein, *Fuzzy Rule-Based Modeling with Application to Geophysical, Biological and Engineering Systems*, CRC Press, Boca Raton, 1995.
2. J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, N. York, 1981.
3. B. Bouchon-Meunier, M. Rifqi, S. Bothorel, Towards general measures of comparison of objects, *Fuzzy Sets and Systems*, 84, 2, 1996, 143-153.
4. O. Cordon, M.J. del Jesus, F. Herrera, A proposal on reasoning methods in fuzzy rule based classification systems, *Int. J. of Approximate Reasoning*, 20, 1999, 21-45.
5. O. Cordon, F. Herrera, Villar P., Analysis and guidelines to obtain a good uniform fuzzy partition granularity for fuzzy rule-based systems using simulated annealing", *Int. J of Approximate Reasoning*, Vol. 25, 3, 2000, 187-215.
6. M. Delgado, F. Gomez-Skarmeta, and F. Martin, A fuzzy clustering-based prototyping for fuzzy rule-based modeling, *IEEE Transactions on Fuzzy Systems*, 5(2), 1997, 223-233.
7. M. Delgado, A.F. Gomez-Skarmets, F. Martin, A methodology to model fuzzy systems using fuzzy clustering in a rapid-prototyping approach, *Fuzzy Sets and Systems*, vol. 97, no.3, 1998, 287-302.
8. D. Dubois, H. Prade, What are fuzzy rules and how to use them, *Fuzzy Sets and Systems*, 84, 1996, 169-185.
9. H. Ishibuchi, K. Nozaki, N. Yamamoto, H. Tanaka, Selecting fuzzy if-then rules for classification problems using genetic algorithms", *IEEE Transactions on Fuzzy Systems*, Vol. 3: 3, 1995, 260-270.
10. W. Pedrycz, *Fuzzy Control and Fuzzy Systems*, 2nd edition, Research Studies Press, Chichester, 1993.
11. E. H. Ruspini, On the semantics of fuzzy logic, *International Journal of Approximate Reasoning* 5, 1991, 45-88.
12. T. Sudkamp, Similarity, interpolation, and fuzzy rule construction, *Fuzzy Sets and Systems*, vol. 58, no. 1, 1993, 73-86.
13. T. A. Sudkamp, R. J. Hammell II, Granularity and specificity in fuzzy function approximation, in Proc. *NAFIPS-98*, 105-109, 1998.
14. R. R. Yager, D. P. Filev, *Essentials of Fuzzy Modeling and Control*. J. Wiley, New York, 1994.