

Fuzzy Clustering of Fuzzy Systems^{*}

Paulo Salgado

CETAV-UTAD University
Quinta dos Prados
Vila Real, 5000-911, Portugal
psal@utad.pt

Paulo J. Garrido

Industrial Electronics Department
University of Minho
Guimarães, Portugal.
pgarrido@dei.uminho.pt

Abstract - A fuzzy system entirely characterizes one region of the input-output product space $S = U \times V$ through a relation expressed by a set of fuzzy rules. Effectively, the fuzzy system establishes a fuzzy map, which assigns for each input fuzzy set in U an output fuzzy set in V . The partition of this product space may be made through the decomposition of the relation. The fuzzy clustering of fuzzy rules, here proposed, as well as clustering of data, leads to a fuzzy partition of the S space. The result is a set of fuzzy sub-systems, one for each cluster that will be conveniently linked in a new structure. This paper proposes a new recursive clustering algorithm for the partition of a fuzzy system into a hierarchical collaborative structure. The global response of the hierarchical collaborative structure is identical to the input fuzzy system.

Keywords: fuzzy system, fuzzy clustering, hierarchical model.

1 Introduction

Fuzzy modeling has recently been applied with success to a variety of problems, especially in control engineering [1][2]. Fuzzy concepts are suited for modeling based on data as well as for modeling based on knowledge acquisition. In both cases, information or knowledge about the system being modeled is captured as IF-THEN rules with fuzzy predicates that establish relations between the relevant system variables.

However, knowledge acquisition is not a trivial task. Experts are not always available, and when they are, their knowledge is not always consistent, systematic, complete and well organized [9]. A similar situation happens when automated modeling strategies are used, where the high number of rules generated leads to a not readable and an opaque fuzzy model.

In many real situations, it is necessary to re-organize the fuzzy system in order to improve its readability [3]. In this context, departing from a fuzzy system $f(x)$ a set of n fuzzy sub-systems $f_1(x), f_2(x), \dots, f_n(x)$ will be obtained. If the re-organization is made according to the so-called

Separation of Linguistic Information Methodology, SLIM [4], the obtained sub-systems will contain information describing particular aspects of the system $f(x)$ [5].

Fuzzy clustering algorithms [6]–[12] are effective methods for exploring the structure of complex real data when grouping of overlapping and vague elements is necessary. The same idea will be applied, as a generalization process, to a set of fuzzy sets and its relations.

This work addresses this fundamental goal of fuzzy modeling by using an algorithm that implements Fuzzy Clustering of Fuzzy Rules. The proposed algorithm allows decomposing the fuzzy relation into sub-relations, through a process of unfolding the fuzzy rules. The obtained sub-rules are then grouped into c subgroups (clusters), by similarity association. Both tasks are implemented simultaneously by the proposed algorithm, which is a generalization of the Probabilistic Clustering Algorithm, traditionally known as the fuzzy c-means algorithm.

The results of this Fuzzy Clustering of Fuzzy Rules Algorithm, FCFRA, are c clusters with distinct fuzzy sub-rules. These clusters will be used to compose the new fuzzy sub-models. As a result, the original fuzzy relation is layered through the c levels of a hierarchical fuzzy system. In other words, the application of the FCFR algorithm in the clustering of a flat fuzzy system leads to distributing the information carried by the system among various layers of a hierarchical collaborative structure, HCS.

The paper is organized as follows. In section 2, a brief introduction of fuzzy models is made. In section 3, the concept of relevance of a set of fuzzy rules is reviewed. The HCS structure and the correspondent fuzzy system are presented. In section 4, the FCFR algorithm is proposed. A benchmark example is presented in section 5. Finally, the main conclusions are outlined in section 6.

2 Fuzzy models

A fuzzy rule-based system approximates an unknown function by covering its graph with fuzzy rules. It has been shown by various authors [13][14][15] that fuzzy systems are universal approximators. In general, the more precise the approximation is required to be, the more rules are needed and, thus, complexity and cost increase.

In this section, one assumes that fuzzy systems are multi-input-single-output systems $y: U \mapsto V$, where $U = U_1 \times \dots \times U_n \subset \mathbb{R}^n$ is the input space and $V \subset \mathbb{R}$ is the output space. A multi-output system can be separated into a group of single-output systems. A set D of input-output data pairs experimentally derived from an unknown function or system f is to be modeled as a fuzzy system.

Consider a fuzzy system that comprises four principal components: fuzzifier, fuzzy rule base, fuzzy inference engine, and defuzzifier. In this paper, we assume that the fuzzifier is the most commonly used singleton fuzzifier and the fuzzy rule base consists of M fuzzy rules in the following form

$$R^{(l)}: \text{IF } x_1 \text{ is } A_1^l \text{ and } \dots \text{ and } x_n \text{ is } A_n^l \text{ THEN } y \text{ is } B^l. \quad (1)$$

In (1) $R^{(l)}$ represents the l -th rule ($l=1, \dots, M$), and the fuzzy sets $A_i^l \subset U_i$ and $B^l \subset V$ are linguistic terms characterized by fuzzy membership functions $A_i^l(x_i)$ and $B^l(y)$ respectively. The inference engine uses these fuzzy rules to determine a mapping from fuzzy sets in the input universe of discourse $U \subset \mathbb{R}^n$ to fuzzy sets in the output universe of discourse $V \subset \mathbb{R}$, based on fuzzy logic principles. Each R^l can be viewed as a fuzzy implication or relation $A^l = A_1^l \times A_2^l \times \dots \times A_n^l \mapsto B^l$, which is a fuzzy set in $U \times V = U_1 \times \dots \times U_n \times V$ with membership function:

$$R_{A \mapsto B}^l(\mathbf{x}, y) = A^l(\mathbf{x}) \otimes B^l(y). \quad (2)$$

In (2) “ \otimes ” is an operator of fuzzy implication, usually the prod or min implication operator [16], which emerges from the generalized rules of *modus ponens*, *modus tollens* or hypothetical syllogism. In this paper, the product operation is adopted as the implication operator.

Let A' be an arbitrary fuzzy set in U ; then each R^l of (1) determines a fuzzy set $B_l'(y)$ in V based on the sup-star composition

$$B^l(y) = \sup_{\mathbf{x} \in U} [A'(\mathbf{x}) \star R_{A \mapsto B}^l(\mathbf{x}, y)]. \quad (3)$$

In (3) \star could be any operator in the class of t-norms. In fuzzy control systems, the fuzzy set A' is the result of the fuzzification process.

Finally, the output of the fuzzy inference engine is the combination of the M fuzzy sets $\{B_1', \dots, B_l', \dots, B_M'\}$ by union:

$$B' = \bigcup_{l=1}^M B_l'. \quad (4)$$

Let one suppose that \mathbf{x} is a crisp input to the fuzzy system. When A' is a fuzzy singleton (i.e., $A'(\mathbf{x}') = 1$ for $\mathbf{x} = \mathbf{x}'$ and $A'(\mathbf{x}') = 0$ $\mathbf{x} \neq \mathbf{x}'$), we have from (2) and (3) that each rule R^l leads to an output fuzzy set:

$$B^l(y) = \sup_{\mathbf{x} \in U} [A'(\mathbf{x}) \star R_{A \mapsto B}^l(\mathbf{x}, y)] = A^l(\mathbf{x}) \otimes B^l(y). \quad (5)$$

This happens because the “sup” is archived at $\mathbf{x}' = \mathbf{x}$ and $A'(\mathbf{x}) = 1$.

The defuzzifier performs a mapping from the fuzzy set in V to crisp points in V . In this paper, a center-average defuzzifier [17] is assumed and the fuzzy system can be expressed as the following:

$$y = \frac{\sum_{l=1}^M A^l(\mathbf{x}) \cdot \theta^l}{\sum_{l=1}^M A^l(\mathbf{x})} \quad (6)$$

In (6) θ^l is the point in V at which $B^l(y)$ achieves its maximum value, when $B^l(y)$ is a normal fuzzy set and have equal volume for all $l=1, \dots, M$.

The center-average defuzzifier is a simplified defuzzifying method because it does not require one to define and calculate the linguistic “or” of equation (4) [18]. By this fact, equation (6) is the most commonly used expression of fuzzy systems.

Given the input vector $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$, the degree of activation is calculated as

$$A^l(\mathbf{x}) = \prod_{i=1}^n A_i^l(x_i), \quad l=1, 2, \dots, M. \quad (7)$$

In (7) $A_i^l(x_i)$ is the membership function of the fuzzy set A_i^l for input variable x_i in the premise of the l^{th} rule, see (1).

The decomposition of the fuzzy relations $R_{A \rightarrow B}^l(\mathbf{x}, y)$ leads to the decomposition of the raw fuzzy rules in new sub-rules. This task is here realized by the FCFRA algorithm.

3 The collaborative fuzzy model

Hierarchical fuzzy modeling is a promising method to identify fuzzy models of target systems with many input variables and/or with different complexity interrelation among subsystems. Partitioning a fuzzy system reduces its complexity, increases its readability, and simplifies the identification problem. It must be noticed that not all fuzzy systems can be divided in a set of fuzzy-sets by a hard-partition. In most cases, this last circumstance results from the fact that the subsystems are not disjoint or independent. However, it is always possible to make a fuzzy partition (clustering) of the flat fuzzy model, by signaling the fuzzy subsystems with a value of relevance. This idea is similar to what happens in many clustering problems where it is not possible to find a hard clustering of the data, yet the problems are solvable by a fuzzy-clustering strategy.

The structure that is best adapted to the clustering of fuzzy rules is the HCS structure, as is illustrated in figure 1, where each block is a fuzzy system that represents one cluster. Differently from what happens in traditional fuzzy systems, all subsystems have two outputs: the relevance output and the defuzzified output, as defined in [3].

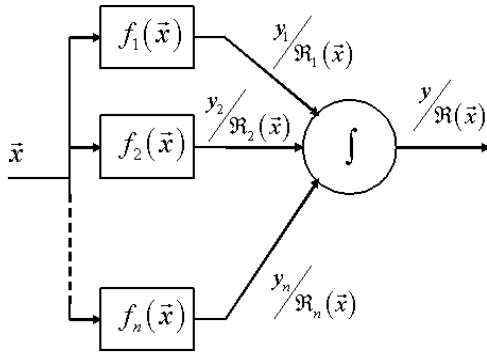


Figure 1: Hierarchical collaborative structure

Relevance is a characterization of the relative importance of a fuzzy system in the context of an application and may be defined as follows.

Definition 1: Relevance is a fuzzy measure of the relative importance of a fuzzy system, f , in a given context S , and its signature is expressed as

$$\mathfrak{R}_S : \tilde{P}(f) \rightarrow [0, 1] \quad (8)$$

where $\tilde{P}(f)$ is the power-set of all fuzzy systems in the context S .

It is to be noted that relevance is a characterization of a fuzzy system in the perspective of one specific context. In the fuzzy modeling process, it is natural to consider that the context S distributes over the regions of the input-output space. A fuzzy system that describes conveniently one region S' of space will have a higher level of relevance in the model.

The relevance of a fuzzy system f is the result of an additive process over the relevance values of its fuzzy rules. The relevance measure obeys a set of axioms defined in [3]. The axioms give relevance a nice set of properties that can be resumed as follows.

P1. Let $\mathfrak{S} = \{R^{(1)}, \dots, R^{(l)}, \dots, R^{(M)}\}$ the set of all rules of the rule base of f and $\tilde{P}(\mathfrak{S})$ the power set of \mathfrak{S} . For any $A, B \in \tilde{P}(\mathfrak{S})$, if $A \subseteq B$, then $\mathfrak{R}_S(A) \leq \mathfrak{R}_S(B)$.

P2. Let $A, B \in \tilde{P}(\mathfrak{S})$. If $C = A \cup B$, then $\mathfrak{R}_S(C) \geq \max(\mathfrak{R}_S(A), \mathfrak{R}_S(B))$, i. e., $\mathfrak{R}_S(C) = s(\mathfrak{R}_S(A), \mathfrak{R}_S(B))$, where s represents any s -norm operation.

P3. If A is a set of rules which only covers the region $T \subseteq S$, then $\mathfrak{R}_S(A) = \mathfrak{R}_T(A)$.

P4. If a fuzzy system f completely describes the region S of space then $\mathfrak{R}_S(f) = 1$.

P5. The union of the relevance of all fuzzy rules of the system, \mathfrak{S} , is equal to the relevance of the fuzzy system.

This means that $\mathfrak{R}(f) = \mathfrak{R}\left(\bigcup_l R^{(l)}\right)$.

P6. Given the space S partitioned in n regions $S = S_1 \cup \dots \cup S_n$ then $\mathfrak{R}_S(f) = \mathfrak{R}\left(\bigcup_i \mathfrak{R}_{S_i}(f)\right)$.

From P2, the relevance of the fuzzy rules of one fuzzy system may be obtained by the union of the relevance of the fuzzy rules:

$$\mathfrak{R}\left(\bigcup_{l=1}^M R^{(l)}\right) = \bigcup_{l=1}^M \mathfrak{R}(R^{(l)}) \quad (9)$$

Equation (9) combined with P4 and P5 gives:

$$\mathfrak{R}(f) = \bigcup_{l=1}^M \mathfrak{R}(R^{(l)}) = 1 \quad (10)$$

In the context of this work one considers that union is an additive operation, i.e., that

$$\mathfrak{R}(f) = \sum_{l=1}^M \mathfrak{R}(R^{(l)}) = 1. \quad (11)$$

This can be designed as the probabilistic aggregation of the relevance measure.

In this work the following definition of *relevance of a fuzzy rule* in the fuzzy system (6).

Definition 2: The *relevance* of a rule $R_l \in \mathfrak{S}$, of fuzzy system (6) for an input $\mathbf{x} \in U$ is defined as

$$\mathfrak{R}_l(\mathbf{x}, y) = \mathfrak{R}(R^{(l)}(\mathbf{x}, y)) = \frac{A^l(\mathbf{x})}{\sum_{l=1}^M A^l(\mathbf{x})} \quad (12)$$

This measuring function of relevance gives the value of the relative power of firing of the rule l in (\mathbf{x}, y) point of space.

The HCS structure is appropriate for situation where the different fuzzy models must collaborate for a better global performance of the model. The output of the SLIM model is the integral of the individual contributions of each fuzzy subsystem

$$f(\mathbf{x}) = \int_{i=1}^n f_i(\mathbf{x}) \cdot \mathfrak{R}_i(\mathbf{x}), \quad (13)$$

where $\mathfrak{R}_i(\mathbf{x})$ represents the relevance function of the i^{th} fuzzy subsystem covering the point x of the universe of discourse, and \int is an aggregation operator. It should be noted that, for enhanced readability of the expressions, the reference to the f_i of the fuzzy sub-model was replaced by i .

The relevance $\mathfrak{R}_i(\mathbf{x})$ reveals the effective contribution (or belief of contribution) to the respective fuzzy system. This variable should be considered in the aggregation of all collaborative systems.

The relevance of an aggregated system is the union of all its sub-systems given by:

$$\mathfrak{R}_i(\mathbf{x}) = \bigcup_{i=1}^n \mathfrak{R}_i(\mathbf{x}) \quad (14)$$

If the i^{th} fuzzy subsystem covers appropriately the region of point \mathbf{x} , its relevance value is high (very close to

one), otherwise the relevance value is low (near zero or zero).

4 The clustering algorithm

The available data samples are collected in matrix \mathbf{Z} formed by concatenating the input data matrix $\mathbf{X}^T = [\mathbf{x}_1, \dots, \mathbf{x}_N]$, the output vector $\mathbf{y} = [y_1, \dots, y_N]$, and $\mathbf{Z} = [\mathbf{X}^T, \mathbf{y}]^T$. Therefore, each observation is an $n+1$ -dimensional column vector $\mathbf{z}_k = [\mathbf{x}_k^T, y_k]^T$. One assumes that for this set of data is identified by a fuzzy model $y = f(\mathbf{x})$ with an adequate accuracy.

Through clustering, the data set is partitioned into c clusters. Contrary to what happens in the classical fuzzy clustering, here the partition of the input-output space is made through the clustering of the fuzzy rules of f . A fuzzy clustering partition has the goal of separating a set of fuzzy rules $\mathfrak{S} = \{R_1, R_2, \dots, R_M\}$ in c clusters, according to a similarity criterion. The result is a fuzzy partition matrix $U = [u_{il}]_{c \times M}$, whose element $u_{il} \in [0, 1]$ represents the degree of membership of rule k in cluster i . Clusters of different shapes can be obtained by using an appropriate definition of cluster prototypes (e.g., linear varieties) or by using different distance measures. Each sample rule l satisfies the constraint that $\sum_{i=1}^c u_{il} = 1$. The set of prototypes for the clusters is given by $\mathbf{v} = (v_1, v_2, \dots, v_c)$.

For a vector $\mathbf{z}_k \in Z$ a set of fuzzy rules is fired. The sum of the relevance of all the rules in the set is in agreement with relation (11). Combination with the above constraint results in

$$\sum_{i=1}^c \sum_{l=1}^M \mathfrak{R}_l(\mathbf{z}_k) \cdot u_{il} = 1, \quad \forall \mathbf{z}_k \in S. \quad (15)$$

The fuzzy clustering of fuzzy rules algorithm, FCFRA, minimizes the functional

$$J(U, V) = \sum_{k=1}^n \left[\sum_{i=1}^c \sum_{l=1}^M (u_{il} \mathfrak{R}_l(\mathbf{z}_k))^m d_{ik}^2 \right] \quad (16)$$

where $d_{ik}^2 = (\mathbf{z}_k - \mathbf{v}_i)^T A (\mathbf{z}_k - \mathbf{v}_i)$, with A being a semi-positive matrix. The alternating optimization, AO, method is one technique to find the minimum. The power of the membership function is called the weighting exponent. Using the memberships U , class exemplars are calculated from the data points. The class exemplars are then used to calculate new memberships. This procedure is repeated

until some form of convergence occurs. The algorithm is resumed below.

The partition should find a vector of clusters centers, V , and a partition matrix, U . Each value u_{il} of U represents the membership degree of the l^{th} rule, R_l , to the i^{th} cluster, A_i .

The Fuzzy Clustering of Fuzzy Rules algorithm – FCFRA

Step 1– For a set of points Z , with $z_k \in S$, and a set of rules $\mathfrak{S} = \{R_1, R_2, \dots, R_M\}$, with relevance $\mathfrak{R}_l(z_k)$, $k = 1, \dots, M$, keep c , $2 \leq c < n$, and initialize $U^{(0)} \in M_{cf}$.

Step 2– On the r^{th} iteration, with $r = 0, 1, 2, \dots$, compute the c mean vectors.

$$v_i^{(r+1)} = \frac{\sum_{l=1}^M \left(u_{il}^{(r)} \right)^m \cdot \sum_{k=1}^{np} \left(\mathfrak{R}_l(z_k) \right)^m \cdot z_k}{\sum_{l=1}^M \left(u_{il}^{(r)} \right)^m \cdot \sum_{k=1}^{np} \left(\mathfrak{R}_l(z_k) \right)^m} \quad (17)$$

where $\left[u_{il}^{(r)} \right] = U^{(r)}$, $i = 1, 2, \dots, c$.

Step 3– Compute the new partition matrix $U^{(r+1)}$ using the expression:

$$u_{il}^{(r+1)} = \frac{1}{\sum_{j=1}^c \left(\frac{\sum_{k=1}^{np} \left(\mathfrak{R}_l(z_k) \right)^m \cdot d_{ik}}{\sum_{k=1}^{np} \left(\mathfrak{R}_l(z_k) \right)^m \cdot d_{ik}} \right)^{\frac{2}{m-1}}} \quad (18)$$

with $1 \leq i \leq c$, $1 \leq l \leq M$.

Step 4– Compare $U^{(r)}$ with $U^{(r+1)}$: If $\| U^{(r+1)} - U^{(r)} \| < \varepsilon$ then the process ends. Otherwise let $r=r+1$ and go to step 2. A small real positive constant is denoted by ε .

The applications of the FCFRA algorithm to a fuzzy system (6) results in a fuzzy system with HCS structure, where the i^{th} fuzzy sub-model is

$$f_i(x) = \frac{\sum_{l=1}^M \left(A^l(x) \cdot u_{il} \right) \cdot \theta^l}{\sum_{l=1}^M \left(A^l(x) \cdot u_{il} \right)}. \quad (19)$$

5 Experimental results

In this section, an example is given to illustrate the proposed strategy for probabilistic clustering in the “fuzzy rules domain”. Let one consider the system

$$f(x, y) = f_1(x, y) + f_2(x, y) = -2 \frac{\sin(2\pi x)}{2\pi x} + 4e^{-9y^2} \quad (20)$$

where f is a nonlinear function, f_1 and f_2 its sub-models, all assumed to be unknown.

The aim is to partition the function f in two clusters. In the first step, the system is modeled as a set of rules, using the nearest neighborhood method [17]. The resulting system after the learning process has 100 fuzzy rules. The output of the system at this stage is shown in Figure 2.a) with an approximation error lower than 0.1. The second step consists in clustering the fuzzy model into 2 clusters with $m=2$, with each one representing a fuzzy system in a HCS structure, using the algorithm presented. Figures 2.b) and 2.c) show the individual outputs response of each fuzzy model, obtained by the clustering process.

6 Conclusions

The application of fuzzy clustering of fuzzy rules has been addressed, as a generalization of the Probabilistic Clustering Algorithm, traditionally known as fuzzy c-means algorithms. The result of the clustering process is a hierarchical structure containing sub-models. This allows an effective partitioning of the input-output space (indirectly by fuzzy rule clustering). This strategy leads to a more transparent model. The performance of the proposed modeling technique was demonstrated on a benchmark problem.

Acknowledgements

This work was supported by the Portuguese Ministry of Science and Technology (MCT), under the project SAPIENS- POCTI/33574/99.

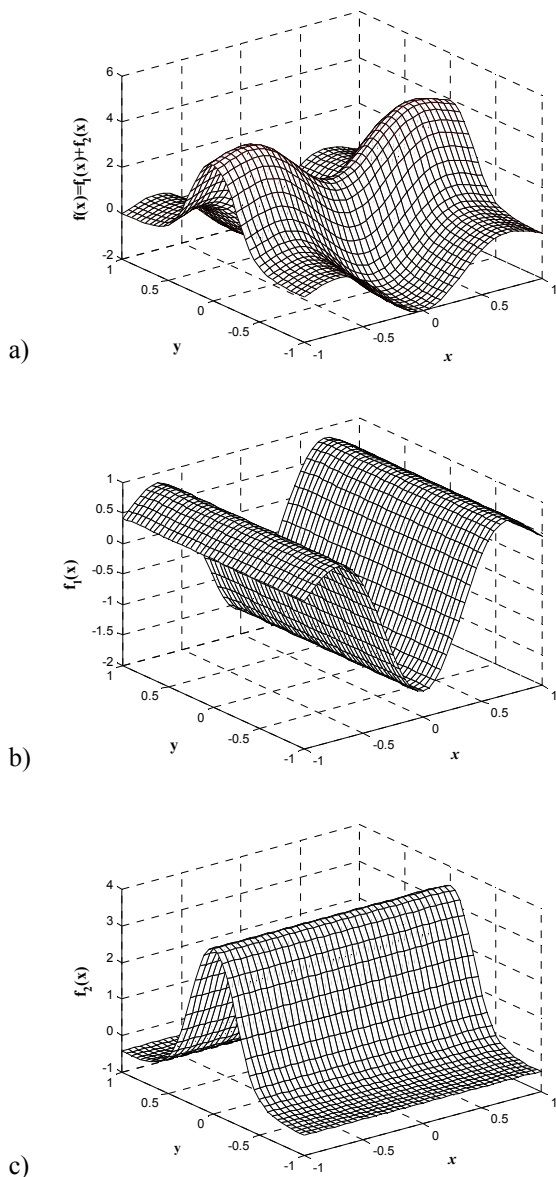


Figure 2: a) the surface of a fuzzy system and its partition in 2 clusters (b and c) by the FCFRA.

References

- [1] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-15, pp. 116-132, 1985.
- [2] R. R. Yager and D. P. Filev, *Essentials of Fuzzy Modeling and Control*, New York: Wiley, 1994.
- [3] P. Salgado, "Clustering and hierarchization of fuzzy systems", *Soft Computing Journal*, Springer Verlag, *in press*.
- [4] P. Salgado, "Relevance of fuzzy system and rules", in "Systematic Organization of Information in Fuzzy Logic", Part II - Fundamentals, NATO Advanced Studies Series, pp. 105-130, IOS Press, 2003.
- [5] P. Salgado, and J. B. Cunha, "Greenhouse Climate Hierarchical Fuzzy Modelling", *Control Engineering Practice*, Elsevier, *in press*
- [6] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, New York: Plenum, 1981.
- [7] J. C. Bezdek and S. K. Pal, *Fuzzy Models for Pattern Recognition*, New York: IEEE Press, 1992.
- [8] N. R. Pal and J. C. Bezdek, "On cluster validity for the fuzzy c-means model," *IEEE Trans. Fuzzy Syst.*, vol. 3, pp. 370-379, 1995.
- [9] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 841-847, Aug. 1991.
- [10] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, New York: Wiley, 1973.
- [11] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
- [12] R. Krishnapuram, H. Frigui, and O. Nasraoui, "Fuzzy and possibilistic shell clustering algorithms and their application to boundary detection and surface approximation-Part I," *IEEE Trans. Fuzzy Syst.*, vol. 3, pp. 29-43, Feb. 1995.
- [13] J. J. Buckley, "Sugeno type controllers are universal controllers," *Fuzzy Sets Syst.*, vol. 53, pp. 299-304, 1993.
- [14] J. L. Castro and M. Delgado, "Fuzzy systems with defuzzification are universal approximators," *IEEE Trans. Syst., Man, Cybern.- Part B*, vol. 26, pp. 149-152, Feb. 1996.
- [15] L. X. Wang and J. M. Mendel, "Generating fuzzy rules by learning from examples," *IEEE Trans. Syst., Man, Cybern.*, vol. 22, no. 6, pp.1414-1427, Nov./Dec. 1992.
- [16] C. C. Lee, "Fuzzy logic in control systems : Fuzzy logic control – part II", *IEEE Trans. Syst. Man, Cybern.*, vol. 20, n^o2, pp. 419-435, 1990.
- [17] L.-X Wang, *A course in fuzzy systems and control*, NJ, Prentice-Hall PTR, 1997.
- [18] L.-X Wang, "Stable adaptive fuzzy control of nonlinear systems," *IEEE Trans. Fuzzy Syst.*, vol 1, n 2, pp.146-155, 1993.