

PROBABILISTIC CLUSTERING ALGORITHMS FOR FUZZY RULES DECOMPOSITION

Paulo Salgado¹ and Getúlio Igrejas²

¹ CETAV-Universidade de Trás-os-Montes e Alto Douro, 5001-801, Vila Real, Portugal

² ESTiG - Instituto Politécnico de Bragança, 5301-857, Bragança, Portugal

Abstract: The fuzzy *c*-means (FCM) clustering algorithm is the best known and used method in fuzzy clustering and is generally applied to well defined set of data. In this paper a generalized Probabilistic fuzzy *c*-means (FCM) algorithm is proposed and applied to clustering fuzzy sets. This technique leads to a fuzzy partition of the fuzzy rules, one for each cluster, which corresponds to a new set of fuzzy sub-systems. When applied to the clustering of a flat fuzzy system results a set of decomposed sub-systems that will be conveniently linked into a Parallel Collaborative Structures. *Copyright © 2007 IFAC*

Keywords: Clustering algorithms; Fuzzy System; Fuzzy C-means, Relevance.

1. INTRODUCTION

Cluster analysis is primarily a tool for discovering previously hidden structure in the set of unordered objects, where we assume that a natural grouping exists in the data. Cluster analysis is a technique for classifying data, i.e., to divide a given set of objects into a set of classes or clusters based on similarity. The goal is to divide the data set in such a way that cases assigned to the same cluster should be as similar as possible whereas two objects from different clusters should be as dissimilar as possible. It is an approach towards unsupervised learning as well as one of the major techniques in pattern recognition.

The conventional (hard) clustering methods restrict each point of the data set to exactly one cluster. These methods yield exhaustive partitions of the example set into non-empty and pairwise disjoint subsets. Fuzzy cluster analysis, therefore allows gradual memberships of data points to clusters in $[0, 1]$. This gives the flexibility to express that data points belong to more than one cluster at the same time. Furthermore, these membership degrees offer a much finer degree of detail of the data model.

One of the most popular object data clustering algorithms is the FCM algorithm, proposed by Dunn

(1973) and extended by Bezdek (1981), which can be applied if the objects of interest are represented as points in a multi-dimensional space. FCM relates the concept of object similarity to spatial closeness and finds cluster centres as prototypes. Several examples of application of FCM to real clustering problems have proved the good characteristics of this algorithm with respect to stability and partition quality. Further, its convergence has been formally demonstrated (Bezdek, 1987; Hathaway et. al., 1988). From this method a large variety of clustering techniques was derived with more complex prototypes, which are mainly interesting in data analysis applications. However, the generalization of these techniques to clustering imprecisely or uncertainly data or objects is not yet explored. Moreover, in the real-world applications, transaction data are usually composed of quantitative values. Designing a sophisticated data-mining algorithm to deal with different types of data turns a challenge in this research topic.

Recently, fuzzy set theory is more and more frequently used in intelligent systems, because of its simplicity and similarity to human reasoning. The theory has been successfully applied to many fields such as manufacturing, engineering, diagnosis, economics, and others (Höppner, 1999).

In this context, a generalization of the previously methods in order to be used in clustering of fuzzy data (or fuzzy numbers) would be a meritorious research. In this work, a new fuzzy relational clustering algorithm, based on the fuzzy c -means algorithm is proposed to clusters fuzzy data, which is used in the antecedent and the consequents parts of the fuzzy rules. This clustering process divides the fuzzy rules of a Fuzzy System into a set of classes or clusters of fuzzy rules based on similarity. From this new strategy, a flat fuzzy system $f(x)$ can be organized into a hierarchical structure of fuzzy systems (Salgado, 2005a and 2007b).

Hierarchical fuzzy modelling is a promising method to identify fuzzy models of target systems with many input variables or/and with different complexity interrelation. Partitioning a fuzzy system reduces its complexity, which simplifies the identification problem, improves the computation times and saves resources, such as memory space. Moreover, with the organization of the fuzzy system into a new hierarchical structure, the model readability and transparency can be improved. In this context, we propose a new technique, the Probabilistic Fuzzy Clustering of Fuzzy Rules (FCFR), based on cluster methodology, to decompose a flat fuzzy system $f(x)$ into a set of n fuzzy sub-systems $f_1(x), f_2(x), \dots, f_n(x)$, organized in a collaborative structure. Each of these clusters may contain information related with particular aspects of the system $f(x)$. The proposed algorithm allows grouping a set of rules into c subgroups (clusters) of similar rules. It is a generalization of the Probabilistic Clustering Algorithm (FCM), here applied to rules instead of points. With this algorithm, the system obtained from the data is transformed into a new system, organized into several subsystems, in PCS structures (Salgado, 2005b and 2007a).

The paper is organized as follows: firstly, a brief introduction to fuzzy systems is presented. The concept of relevance of a set of rules and of fuzzy system is reviewed. The PCS structure is described in section 3. In section 4 the FCFR strategy is proposed. An example is presented in section 5. Finally, the main conclusions are outlined in section 6.

2. RELEVANCE OF FUZZY SYSTEM

A generic fuzzy model is presented as a collection of fuzzy rules in the following form:

R^i : IF x_1 is A_{1l} and ... and x_n is A_{nl} THEN $y=z_l(\vec{x})$

where $\vec{x} = (x_1, x_2, \dots, x_n)^T \in X$ and $y \in Y$ are linguistic variables, A_{ij} are fuzzy sets of the universes of discourse $X_i \in R$, and $z_l(\vec{x})$ is a function of the input variables. Typically, z can take one of the following three forms: fuzzy set (Mamdani type fuzzy systems), singleton (Takagi-Sugeno) or polynomial function (Takagi-Sugeno-Kang, TSK) type fuzzy systems. Takagi-Sugeno fuzzy systems with centre average defuzzification, product-inference rule and singleton fuzzification are represented by:

$$f(\vec{x}_k) = \sum_{l=1}^M p^l(\vec{x}_k) \cdot \theta^l \quad (1)$$

where $p^l(\vec{x}) = \mu^l(\vec{x}) / \sum_{l=1}^M \mu^l(\vec{x})$ is the fuzzy basis

functions (FBF), M represent the number of rules, θ^l is the point at which the output fuzzy set l achieves its maximum value, and μ^l is the membership of the antecedent of rule l . The defuzzified output y of the fuzzy model is calculated as a weighted average (Roventa et al., 2003) of all fuzzy rules outputs.

Fuzzy Logic Systems, FLS, are based on a set of rules that map regions in an input space, X , into regions in an output space, Y , describing a region in a product space $S = X \times Y$. The fuzzy rules are fuzzy relations in the product space S described by a set of rules \mathfrak{F} , which create a power set of fuzzy rules $\tilde{P}(\mathfrak{F})$. In the traditional systems, as equation (1), all the rules are considered as having the same contribution in the characterization of the fuzzy system. However, they will have different importance in different regions of space or in modelling fundamental relationships. For the characterization of the relative importance of sets of rules, in the modelling process, it is essential to define a relevance function.

The relevance is a measure of the relative importance of the rules that describe the region S and is a special fuzzy measure that involves the relativity of a support region, which we see as a fuzzy measure only if the support of rules agrees with region S . Depending on the context where the relevance is to be measured, different metrics may be defined.

Definition 1: The relevance of the rule $R \in \tilde{P}(\mathfrak{F})$ on a region S can be characterized by a real positive value. The normalized relevance function maps the power set of fuzzy rules $\tilde{P}(\mathfrak{F})$ on the real interval $[0, 1]$, i.e.: $\mathfrak{R}_s(R) \in [0, 1]$.

In the context of fuzzy systems there are many definitions of relevance of fuzzy rules. Next, we propose one of them for the fuzzy system (1).

Definition 2: Let \mathfrak{F} be a set of rules that map X into Y , describing completely the region S . The relevance of a rule $R^l \in \mathfrak{F}$, of fuzzy system (1) in S space is defined as:

$$\mathfrak{R}_l(\vec{x}_k) = \frac{\mu^l(\vec{x}_k)}{\sum_{l=1}^M \mu^l(\vec{x}_k)} \quad (2)$$

i.e., the relevance in (\vec{x}, y) is the maximum of the ratio between the output membership function value of rule l in (\vec{x}, y) , and the union (sum) value of all membership functions in (\vec{x}, y) .

Let one consider the Fuzzy Systems that obey to definition 3.

Definition 3: The fuzzy system relevance in the point $\vec{x}_k \in S$ is the sum of the relevance of all rules point $\vec{x}_k \in S$ and equal to one:

$$\mathfrak{R}_s(\vec{x}_k) = \sum_{i=1}^M \mathfrak{R}_i(\vec{x}_k) = 1 \quad (3)$$

3. THE PARALLEL COLLABORATIVE STRUCTURE

A clustering algorithm is used in this work to implement the separation of information among the various subsystems, which are organized into a Parallel Collaborative Structure, PCS. Each of these subsystems may contain information related with particular aspects of the system or merely collaborates to the performance of $f(x)$. A PCS structure with n sub models fuzzy systems is depicted in Fig. 1. Each fuzzy system model i has two outputs: an output variable y_i and the correspondent fuzzy system relevance $\mathfrak{R}_i(x)$.

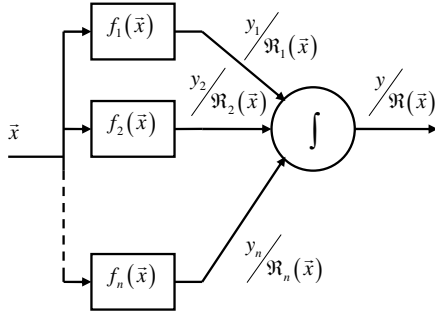


Fig. 1. Structure of Hierarchical Collaborative Fuzzy System

This fuzzy system architecture describes the strength of mind collaboration among the different fuzzy models. Therefore, the output of the SLIM model is the integral of the individual contributions of each fuzzy subsystem:

$$f(x) = \int_{i=1}^n f_i(x) \cdot \mathfrak{R}_i(x) \quad (4)$$

where $\mathfrak{R}_i(x)$ represents the relevance function of the i^{th} fuzzy subsystem covering the point x of the Universe of Discourse, and the \int is an aggregation operator. The relevance $\mathfrak{R}_i(x)$ reveals the effective contribution (or belief of its contribution) to the respective fuzzy system. This variable should be considered in the aggregation of all collaborative systems.

With the same meaning of its congener sub-systems, the relevance of an aggregated system is given by:

$$\mathfrak{R}_i(x) = \bigcup_{i=1}^n \mathfrak{R}_i(x) \quad (5)$$

Naturally, if the i^{th} fuzzy subsystem covers appropriately the region of point x , its relevance value is high (very close to one), otherwise the relevance value is low (near zero or zero).

4. THE PROBABILISTIC CLUSTERING ALGORITHM OF FUZZY RULES

4.1 The FCM algorithm

Clustering is well established as a way to separate a set $X = \{x_1, x_2, \dots, x_{np}\}$ into c subsets that represent (sub)structures of X . A partition can be described by a $c \times n$ partition matrix U . Each element u_{ik} , $i=1, \dots, c$, $k=1, \dots, n$ of the partition matrix represents the membership of $x_k \in X$ in the i^{th} cluster. We distinguish a particular set of partition matrices:

$$M_{fcm} = \left\{ U \in [0,1]^{cn} \mid \sum_{i=1}^c u_{ik} = 1, k=1, \dots, np; i=1, \dots, c \right\} \quad (6)$$

FCM is defined as the following problem: Given the data set X , any norm $\|\cdot\|$ on \mathbb{R}^p and a fuzziness parameter $m \in (1, \infty)$, minimize the objective function

$$J(U, V) = \sum_{k=1}^n \sum_{i=1}^c u_{ik}^m \cdot d_{ik}, \quad 1 < m \leq \infty \quad (7)$$

where $d_{ik} = \|x_k - v_i\|^2$; $U \in M_{fcm}$ and $V = \{v_1, \dots, v_c\} \subset \mathbb{R}^p$ is a set of prototype points (cluster centers).

It can be shown that the following algorithm may lead the pair (U^*, V^*) to a minimum, using alternating optimization (Hathaway et. al, 1988), which result is resumed as follows:

Probabilistic Fuzzy C-Means Algorithm

Step 1– For a set of points $X = \{x_1, x_2, \dots, x_{np}\}$, with $x_i \in \mathbb{R}^n$, keep c , $2 \leq c < np$, and initialize $U(0) \in M_{fcm}$.

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

$$v_i^{(r)} = \frac{\sum_{k=1}^{np} \left(u_{ik}^{(r)} \right)^m \cdot x_k}{\sum_{k=1}^{np} \left(u_{ik}^{(r)} \right)^m}, \quad i=1, 2, \dots, c. \quad (8)$$

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

$$u_{ik}^{(r+1)} = 1 / \left(\sum_{j=1}^c \left(\frac{d_{ik}}{d_{jk}} \right)^{\frac{1}{m-1}} \right) \quad (9)$$

for, $1 \leq i \leq c$, $1 \leq k \leq np$, where $\eta_k \in \mathbb{R}$.

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. ε is a small real positive constant.

The equation (9) defines the probabilistic (FCM) membership function for cluster i in the universe of discourse of all data vectors X .

4.2 Probabilistic Clustering Algorithm of fuzzy rules

In this section, one assumes that fuzzy systems are multi-input-single-output systems $y: X \mapsto Y$, where $X = X_1 \times \dots \times X_n \subset \mathbb{R}^n$ is the input space and $V \subset \mathbb{R}$ is the output space of type (1), which has been clearly recognized as an attractive alternative to functional approximation schemes, since it is able to realize nonlinear mappings of any continuous function (Wang, 1992). Conceptually, the functional relationships between input-output variables, mathematically called dependent-independent variables, are expressed by fuzzy rules base through an inference process.

The fuzzy rules are relationships between fuzzy sets (or fuzzy numbers) that portioned the antecedent and consequent space.

The objective of fuzzy clustering partition is to separate a set of fuzzy rules $\mathfrak{F}=\{R_1, R_2, \dots, R_M\}$ in c clusters in the antecedent space and e clusters in the consequent space, according to a ‘‘similarity’’ criterion. This process allows finding the optimal clusters centres, V and Z , respectively in the input and output space, the partition matrix, U , of combined input-output partition and the matrix W of scalars values. Each value u_{ijl} represents the membership degree of the l^{th} rule, R_l , belonging to the i^{th} cluster of the input space and j^{th} cluster of the output space. w_{jl} is a value that express the translation of the consequent of the l^{th} rule fuzzy sets in direction of the center of j^{th} the output center of cluster. So, the center of each rule l in the cluster j is θ_l^j , with $\theta_l^j = w_{jl}\theta^l$ and is expectable that:

$$\sum_{j=1}^e w_{jl} = 1, \quad l = 1, \dots, M \quad (10)$$

with $w_{jl} \in \mathbb{R}$.

Let $x_k \in S$ be a point covered by one or more fuzzy rules. Naturally, the membership degree of point x_k belonging to $(ij)^{\text{th}}$ cluster is:

$$\sum_{i=1}^c \sum_{j=1}^e u_{ijl} = 1, \quad \forall x_k \in S \quad (11)$$

and the relevance of the rules l in x_k point:

$$\sum_{l=1}^M \mathfrak{R}_l(x_k) = 1, \quad \forall x_k \in S \quad (12)$$

The rule decomposition into $c \times e$ sub-relations will lead to an output fuzzy set decomposition as well. For fuzzy probabilistic clustering, each rule and x_k point, must obey simultaneously to equations (6) and (11). This requirements and the relevance condition of equation (6) are completely satisfied in equation (11). So, for the Fuzzy Clustering of Fuzzy Rules Algorithm (FCFRA) the objective is to find $U=[u_{ijl}]$, $V=[v_1, \dots, v_c] \in R^{n \times c}$ and $Z=[z_1, \dots, z_e] \in R^e$ where:

$$J = \sum_{k=1}^n \sum_{l=1}^M \sum_{i=1}^c \sum_{j=1}^e u_{ijl}^m \mathfrak{R}_l^m(x_k) \left[(x_k - v_i)^2 + (\theta_l w_{jl} - z_j)^2 \right] \quad (13)$$

is minimized, with a weighting constant $m > 1$, with equation (10), (11) and (12) as a constraint.

It can be shown that the following algorithm may lead the pair (U^*, V^*, W^*) to a minimum. The results can be expressed by the following algorithm:

Probabilistic Fuzzy Clustering algorithms of fuzzy rules – FCAFR

Step 1– For a set of points $X=\{x_1, \dots, x_n\}$, with $x_i \in S$, and a set of rules $\mathfrak{F}=\{R_1, R_2, \dots, R_M\}$, with relevance $\mathfrak{R}_l(x_k)$, $k = 1, \dots, M$, keep c , $2 \leq c < np$, and initialize $U(0) \in M_{fcm}$.

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

$$v_i^{(r)} = \frac{\sum_{k=1}^n \left(\sum_{l=1}^M U_{il}^m \cdot \mathfrak{R}_l^m(x_k) \right) \cdot x_k}{\sum_{k=1}^n \left(\sum_{l=1}^M U_{il}^m \cdot \mathfrak{R}_l^m(x_k) \right)} \quad (14)$$

where $U_{il}^m = \sum_{j=1}^e u_{ijl}^m$, $i=1, 2, \dots, e$ and.

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

$$u_{ijl}^{(r+1)} = \frac{1}{\sum_{r=1}^c \sum_{s=1}^e \left(\frac{\sum_{k=1}^n \mathfrak{R}_l^m(x_k) \cdot D_{ijk}}{\sum_{k=1}^n \mathfrak{R}_l^m(x_k) \cdot D_{rsk}} \right)^{\frac{1}{m-1}}} \quad (15)$$

where $D_{ijk} = \left[(x_k - v_i)^2 + (\theta_l w_{jl} - z_j)^2 \right]$, with $1 \leq i \leq c$, $1 \leq l \leq M$.

Step 4 – Compute the new partition matrix $W(r+1)$ with the expression:

$$w_{jl}^{(r+1)} = \frac{1 - \hat{\theta}_l^r \sum_{r=1}^e V_r}{\sum_{r=1}^e \left(\frac{\bar{U}_{jl}}{\bar{U}_{rl}} \right)^m} + \hat{\theta}_l^r V_j \quad (16)$$

with $\hat{\theta}_l = \theta_l / (\theta_l^T \theta_l)$ and $\bar{U}_{jl}^m = \sum_{i=1}^c u_{ijl}^m$.

Step 5 – Compute z_j with:

$$z_j^{(r+1)} = \frac{\sum_{l=1}^M [U_{jl}^m \cdot \bar{\mathfrak{R}}_l^m w_{jl} \theta_l]}{\sum_{l=1}^M [U_{jl}^m \cdot \bar{\mathfrak{R}}_l^m]} \quad (17)$$

where $\bar{\mathfrak{R}}_l^m = \sum_{k=1}^n \mathfrak{R}_l^m(x_k)$.

Step 6– If $\|U(r+1) - U(r)\| < \varepsilon$ then the process ends. Otherwise let $r = r + 1$ and go to step 2.

More details about this method can be found in (Salgado, 2007b).

5. EXPERIMENTAL RESULTS

In this section, an example is given to illustrate the proposed strategy for possibilistic clustering in “fuzzy rules domain”. Fig. 2 shows a volcano’s surface generated with 40×40 data points. The exercise is to capture in a PCS system the description of the function, through the clustering decomposition of a flat fuzzy system (FS). The original structure of FS is identified from the data points using the Nearest Neighborhood Identification method, with a radius of 1.2 and a negligible error. A set of 380 fuzzy rules was generated. It is general perception that the volcano function, $W=F(U,V)$, can be generated by the following three level of PCS structure (or 3 collaborative fuzzy models), each one has the task to model, in collaborative contribution, a particular representation of the Vulcan surface. So, it is natural to have the following sub-system:

Level 1 (Mountain): IF (U,V) is very close to $(5,5)$ THEN W is quasi null;

Level 2 (Hall): IF (U,V) is close to $(5,5)$ THEN W is high;

Level 3 (Background): IF U and V are anything THEN W is low;

Now, we begin building the PCS structure in line with the SLIM-PCS Algorithm. As mentioned, in the first step, the system is modelled by a set of rules, which is an accuracy modelling of the identified system. The output of the system at this stage is practically identical of the one shown in Fig. 2.

The second step consists in the decomposition of the fuzzy rules of the FS into 3 clusters ($m = 1.2$). Each one of these clusters represents a fuzzy system in a PCS structure. Fig. 3 to Fig. 5 shows the individual output response of each hierarchical fuzzy model. The original image can be described as the aggregation (equation (4)) of these three clusters surfaces. So, the use of the FCAFR algorithm makes the stratification of the early flat fuzzy system into a PCS structure. The membership values of the fuzzy rules for each cluster are shown in Fig. 6 to Fig. 8. (note that the membership functions for each cluster are represented by a surface instead of its discrete values). From these figures we can observe where each cluster is “relevant” in the description of the various regions of the surface. It must be noted that the 1st cluster identifies the mountain of the volcano without the interior cavity and this last one is modelled by the 2nd cluster. The 3rd cluster identifies the foot of the mountain.

6. CONCLUSION

In this work, the mathematical fundamentals for Possibilistic fuzzy clustering of fuzzy rules were presented. In the FCFR the relevance concept has a significant importance. Based on this concept, it is possible to make a possibilistic fuzzy clustering algorithm of fuzzy rules, which is naturally a generalization of possibilistic clustering algorithms.

ACKNOWLEDGMENTS

This work was supported by Fundação para a Ciência e Tecnologia (FCT) under grant POSI/SRI/41975/2001 and by CETAV-Centro de Estudos Tecnológicos do Ambiente e da Vida.

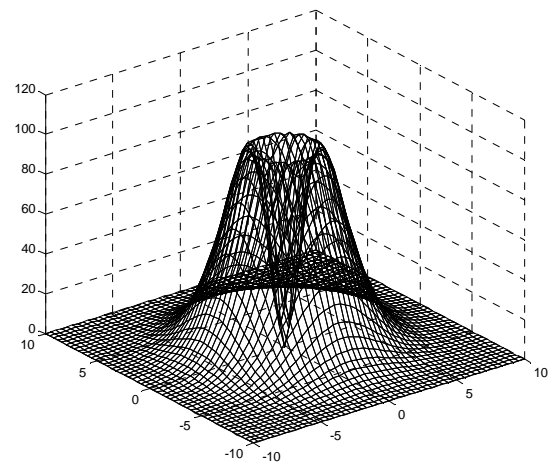


Fig. 2– Volcano surface – original system.

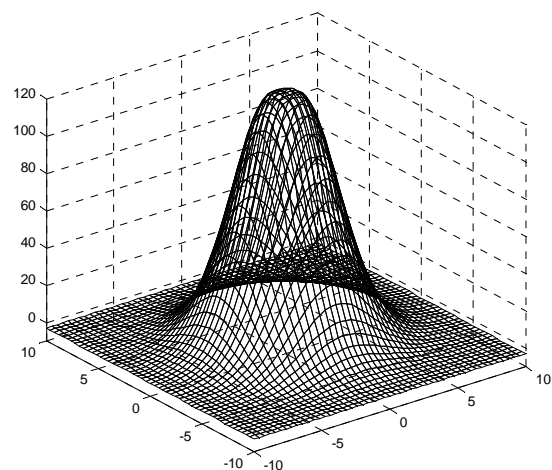


Fig. 3 – Surface generated by 1th fuzzy system cluster.

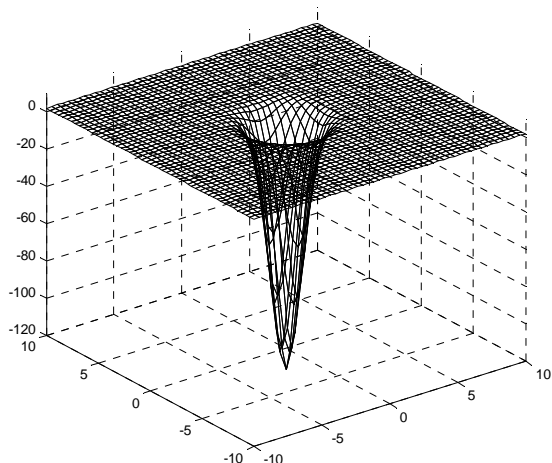


Fig. 4- Surface generated by 2nd cluster fuzzy system: the hall.

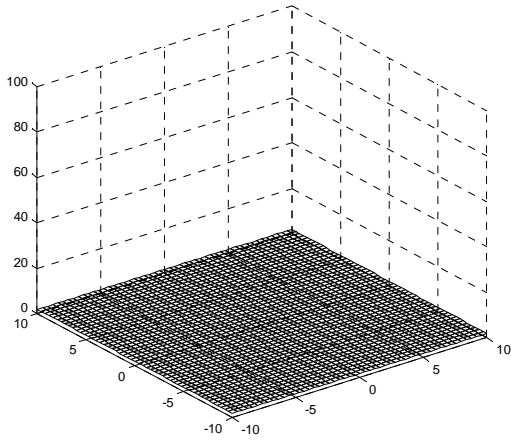


Fig. 5- Surface generated by third fuzzy system cluster – the background of surface.

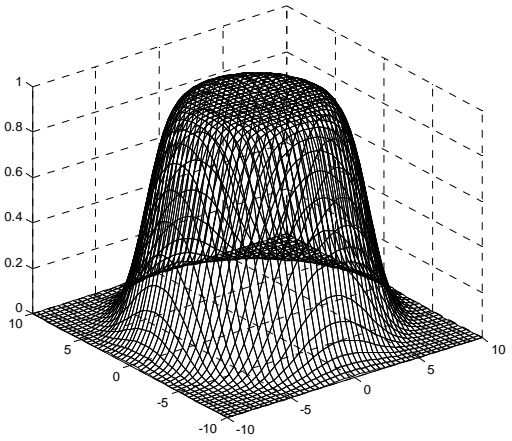


Fig. 7- Membership function u_{i1} , for cluster2.

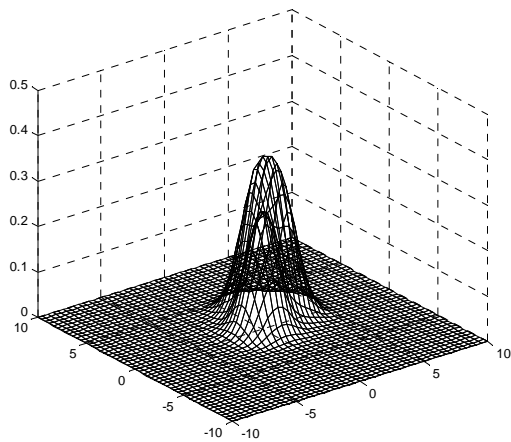


Fig. 6- Membership function u_{i1} , for cluster 1.

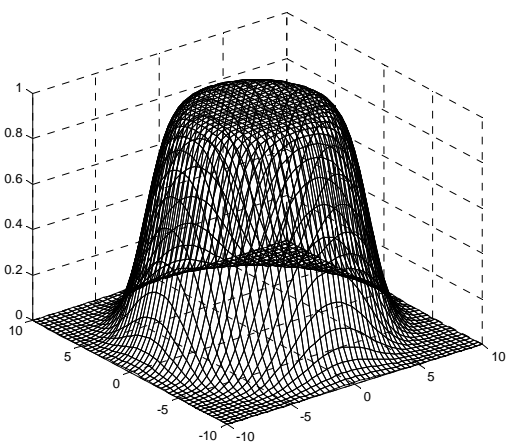


Fig. 8: Membership function u_{i1} , for cluster3.

REFERENCES

- Dunn, J.C., (1974). A fuzzy relative of the isodata process and its use in detecting compact, well separated clusters. *J. Cybernet.* **3**, 95–104.
- Bezdek, J.C.(1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, NY.
- Bezdek JC, Hathaway RJ, Sabin MJ, Tucker WT (1987) Convergence theory for fuzzy C-means: counterexamples and repairs. *IEEE Trans Syst, Man, and Cybern SMC-17(5):873–877*
- Bezdek JC, Keller J, Krisnapuram R, Pal NR (1999) *Fuzzy models and algorithms for pattern recognition and image processing*, Kluwer Academic Publishing, Boston
- Hathaway RJ, Bezdek JC (1988) Recent convergence results for the fuzzy C-means clustering algorithms. *J Class* **5:237–247**
- Höppner, F., Klawonn, F., Kruse, R. & Runkler, T. (1999). *Fuzzy Cluster Analysis—Methods for Image Recognition, Classification, and Data Analysis*. New York: Wiley.
- Roventa, E., Spircu, T. (2003). Averaging Procedures in Defuzzification Processes, *Fuzzy Sets and Systems* **136**, pp. 375-385.
- Salgado, Paulo, (2005a). Clustering and hierarchization of fuzzy systems, *Soft Computing Journal*, Vol. 9, n° 10, pp. 715-731, October, 2005, Springer Verlag.
- Salgado, P. & Boaventura, J. (2005b). Greenhouse climate hierarchical fuzzy modelling, *Control Engineering Practice*, **13** pp. 613-628.
- Salgado, P. (2007a). Rule generation for hierarchical collaborative fuzzy system, *Applied Mathematical Modelling*, (In Press).
- Salgado, P. (2007b), Hierarchical decomposition of the fuzzy systems by clustering process, accepted to publish.
- Wang, Li-Xin. and J. M. Mendel (1992). Fuzzy basis functions, universal approximation, and orthogonal least-square learning, *IEEE Trans. Neural Networks*, **3**, pp. 807–814.