A Fuzzy Clustering Model of Data and Fuzzy c-Means

S. Nascimento[1], B. Mirkin[2] and F. Moura-Pires[1]

[1] CENTRIA, Departamento de Informática

Faculdade Ciências e Tecnologia-Universidade Nova de Lisboa

PORTUGAL

[2] Department of Computer Science, Birkbeck College,

London, UK

Abstract— **The Multiple Prototype Fuzzy Clustering Model (FCMP), introduced by Nascimento, Mirkin and Moura-Pires (1999), proposes a framework for partitional fuzzy clustering which suggests a model of how the data are generated from a cluster structure to be identified. In the model, it is assumed that the membership of each entity to a cluster expresses a part of the cluster prototype reflected in the entity.**

In this paper we extend the FCMP framework to a number of clustering criteria, and study the FCMP properties on fitting the underlying proposed model from which data is generated. A comparative study with the Fuzzy c-Means algorithm is also presented.

Keywords: Fuzzy model identification; prototype; fuzzy membership function; alternating minimization.

I. Introduction

Partitional clustering essentially deals with the task of partitioning a set of entities into a number of homogeneous clusters, with respect to a suitable similarity measure. Due to the fuzzy nature of many practical problems, a number of fuzzy clustering methods have been developed following the general fuzzy set theory strategies outlined by Zadeh, [1]. The main difference between the traditional hard clustering and fuzzy clustering can be stated as follows. While in hard clustering an entity belongs only to one cluster, in fuzzy clustering entities are allowed to belong to many clusters with different degrees of membership.

The most known method of fuzzy clustering is the Fuzzy c-Means method (FCM), initially proposed by Dunn [2] and generalized by Bezdek [3],[4] and other authors [5],[6] (see [7] for an overview). Usually, membership functions are defined based on a distance function, such that membership degrees express proximities of entities to cluster centers (i.e. prototypes). By choosing a suitable distance function (see [6],[8], [9]) different cluster shapes can be identified. However, these approaches typically fail to explicitly describe how the fuzzy cluster structure relates to the data from which it is derived.

Nascimento, Mirkin and Moura-Pires [10] proposed a framework for fuzzy clustering based on a model of how the data is generated from a cluster structure to be identified. In this approach, the underlying fuzzy c-partition is supposed to be defined in such a way that the membership of an entity to a cluster expresses a part of the cluster's prototype reflected in the entity. This way, an entity may bear 60% of a prototype A and 40% of prototype B, which simultaneously expresses the entity's membership to the respective clusters. The prototypes are considered as offered by the knowledge domain. This idea was initially proposed by Mirkin and Satarov as the so-called ideal type fuzzy clustering model [11] (see also [12]), such that each observed entity is defined as a convex combination of the prototypes, and the coefficients are the entity membership values.

In our work, we consider a different way for pertaining observed entities to the prototypes: any entity may independently relate to any prototype, which is similar to the assumption in FCM criterion. The model is called the Fuzzy Clustering Multiple Prototype (FCMP) model.

In this paper we extend the FCMP framework to a number of clustering criteria and present the main results of the study of the FCMP as a model-based approach for clustering as well as its comparison with the FCM algorithm.

The paper is organized as follows. Section 2 introduces fuzzy partitional clustering with the FCM algorithm. In section 3, the FCMP model for fuzzy clustering is described as well as a clustering algorithm to fit the model. Three versions of criteria to fit the model are described: a generic one, FCMP-0, and two "softer" versions, FCMP-1 and FCMP-2. To study the properties of the FCMP model in a systematical way, a data generator has been designed. Section 4 discusses the results of an experimental study using generated data. Conclusion is in section 5.

II. Fuzzy c-Means Algorithm

The fuzzy c-means (FCM) algorithm [3] is one of the most widely used methods in fuzzy clustering. It is based on the concept of fuzzy c-partition, introduced by Ruspini [13], summarized as follows.

Let $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ be a set of given data, where each data point $\mathbf{x}_k$ $(k = 1, \dots, n)$ is a vector in $\Re^p$, U_{cn} be a set of real $c \times n$ matrices, and c be an integer, $2 \leq c < n$. Then, the fuzzy c-partition space for X is the set

$$M_{fcn} = \{U \in U_{cn} : u_{ik} \in [0,1] \quad \sum_{i=1}^{c} u_{ik} = 1, \quad 0 < \sum_{k=1}^{n} u_{ik} < n \ \}, \tag{1}$$

where u_{ik} is the membership value of $\mathbf{x}_k$ in cluster i ($i = 1, \ldots, c$).

The aim of the FCM algorithm is to find an optimal fuzzy c-partition and corresponding prototypes minimizing the objective function

$$J_m(U, \mathbf{V}; X) = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^m \left\| \mathbf{x}_k - \mathbf{v}_i \right\|^2. \tag{2}$$

In (2), $V = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_c)$ is a matrix of unknown cluster centers (prototypes) $\mathbf{v}_i \in \Re^p$, $\|\cdot\|$ is the Euclidean norm, and the weighting exponent m in $[1, \infty)$ is a constant that influences the membership values.

To minimize criterion J_m, under the fuzzy constraints defined in (1), the FCM algorithm is defined as an *alternating minimization* algorithm (cf.[3] for the derivations), as follows. Choose a value for c, m and ε, a small positive constant; then, generate randomly a fuzzy c-partition U^0 and set iteration number $t = 0$. A two-step iterative process works as follows. Given the membership values $u_{ik}^{(t)}$, the cluster centers $\mathbf{v}_i^{(\mathbf{t})}$($i = 1, \ldots, c$) are calculated by

$$\mathbf{v}_i^{(t)} = \frac{\sum_{k=1}^{n} \left(u_{ik}^{(t)} \right)^m \mathbf{x}_k}{\sum_{k=1}^{n} \left(u_{ik}^{(t)} \right)^m}. \tag{3}$$

Given the new cluster centers $\mathbf{v_i^{(t)}}$, update membership values $u_{ik}^{(t)}$:

$$u_{ik}^{(t+1)} = \left[\sum_{j=1}^{c} \left(\frac{\left\| \mathbf{x}_k - \mathbf{v}_i^{(t)} \right\|^2}{\left\| \mathbf{x}_k - \mathbf{v}_j^{(t)} \right\|^2} \right)^{\frac{2}{m-1}} \right]^{-1} . \tag{4}$$

The process stops when $\left| U^{(t+1)} - U^{(t)} \right| \leq \varepsilon$, or a predefined number of iterations is reached.

III. The Multiple Prototype Fuzzy Clustering Model

A. The Generic Model

Let the data matrix X be preprocessed into Y by shifting the origin to the data gravity center and scaling features by their ranges. Thus, $Y = [y_{kh}]$ is a $n \times p$ entity-to-feature data table where each entity, described by p features, is defined by the row-vector $\mathbf{y}_k = [y_{kh}] \in \Re^p$ ($k = 1 \cdots n$; $h = 1 \cdots p$). This data set can be structured according to a fuzzy c-partition which is a set of c clusters, any cluster i ($i = 1, \cdots, c$) being defined by: 1) its prototype, a row-vector $\mathbf{v}_i = [v_{ih}] \in \Re^p$, and 2) its membership values $\{u_{ik}\}$ ($k = 1 \cdots n$), so that the following constraints hold:

$$0 \leq u_{ik} \leq 1, \quad \forall\, i, k; \tag{5a}$$

$$\sum_{i=1}^{c} u_{ik} = 1, \quad \forall\, k. \tag{5b}$$

Let us assume that each entity $\mathbf{y}_k = [y_{kh}]$ of Y is related to each prototype $\mathbf{v}_i = [v_{ih}]$ ($i = 1, \cdots, c$) up to its membership degree u_{ik}; that is, u_{ik} expresses that part of $\mathbf{v}_i$ which is present in $\mathbf{y}_k$ in such a way that approximately $y_{kh} = u_{ik} v_{ih}$ for every feature h. More exactly, we suppose that

$$y_{kh} = u_{ik} v_{ih} + \varepsilon_{ikh}, \tag{6}$$

where the residual values ε_{ikh} are as small as possible.

The meaning of a prototype, $\mathbf{v}_i$, according to equation (6): this is a “model” or “ideal” point such that any entity, $\mathbf{y}_k$, bears a proportion of it, u_{ik}, up to the residuals. The proportion, u_{ik}, is considered as the value of membership of $\mathbf{y}_k$ to the cluster i whose prototype is $\mathbf{v}_i$, which allows us to refer to this as to the proportional membership function.

A clustering criterion according to (6) can be defined as fitting of each data point to each of the prototypes up to the degree of membership. This goal is achieved by minimizing all the residual values via the least-squares criterion

$$E_0(U, \mathbf{V}; Y) = \sum_{i=1}^{c} \sum_{k=1}^{n} \sum_{h=1}^{p} (y_{kh} - u_{ik} v_{ih})^2 \tag{7}$$

with regard to the constraints (5a) and (5b).

The equations in (6) along with the least-squares criterion (7) to be minimized by unknown parameters U and $V = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_c) \in \Re^{cp}$ for Y given, will be referred to as the generic *fuzzy clustering multiple prototypes model*, FCMP-0, for short. In this model, the principle of the least-squares criterion [14] present in FCM criterion (2), is extended to a data-to-cluster model framework, which inevitably leads to a more complex form of the criterion.

B. Relaxing Restrictions of FCMP-0

In real domains of application such as clinical findings for typical scenarios of diseases, personality traits in psychology, types of consumer in market research, the concept of prototype is meaningful in such a way that data entities can be described as sharing parts of prototypes. This is the idea underlying FCMP. However, the requirement of FCMP-0 that each entity be expressed as a part of each prototype is obviously too strong and unrealistic. The intuition leads us to consider that only meaningful sharings, those expressed by high membership values, should be influential in the equations (6).

There are two ways to implement this idea in the FCMP framework: in a hard manner and in a smooth one. A "hard" version should deal only with those equations in (6) that involve rather large values of u_{ik}. By specifying a threshold, β between 0 and 1, only those differences ε_{ikh} are left in the criterion (7) that satisfy the inequality, $u_{ik} \geq \beta$. In such a model, FCMP$_\beta$, introduced in [15], entities may relate to as few prototypes as we wish, which leads to what is called "soft clustering" [16], an intermediate between crisp clustering and fuzzy clustering.

In this paper we concentrate on a different, smooth, approach to modifying FCMP-0 criterion. In order to smooth the influence of small memberships u_{ik}, let us weight the squared residuals in (7) by a power m ($m = 1, 2$) of corresponding u_{ik}:

$$E_m(U,V;Y) = \sum_{i=1}^{c}\sum_{k=1}^{n}\sum_{h=1}^{p} u_{ik}^m (y_{kh} - u_{ik}v_{ih})^2, \quad (8)$$

subject to the fuzziness constraints (5a) and (5b).

The models corresponding to each of these criteria will be denoted as FCMP-1, for $m = 1$, and FCMP-2, for $m = 2$.

C. Minimizing FCMP Criteria

The alternating minimization algorithm FCM for fuzzy c-means clustering can be extended for minimization of FCMP criteria subject to the fuzzy constraints (5a) and (5b).

Each iteration of the algorithm consists of two steps as follows. First, given prototype matrix V, the optimal membership values are found by minimizing criteria (8) for $m = 0, 1, 2$, respectively. In contrast to FCM, minimization of the criteria subject to constraints (5a) and (5b) is not an obvious task; it requires an iterative solution on its own.

Upon preliminarily experimenting with several options, the gradient projection method [17], [18] has been selected for finding optimal membership values (given the prototypes). It can be proven that the method converges fast for FCMP-0 with a constant (anti) gradient stepsize.

Let us denote the set of membership vectors satisfying conditions (5a) and (5b) by Q. The calculations of the membership vectors $\mathbf{u}_k^{(t)} = \left[u_{ik}^{(t)}\right]$ are based on vectors $\mathbf{d}_k^{(t)} = \left[d_{ik}^{(t)}\right]$:

$$d_{ik}^{(t)} = u_{ik}^{(t-1)} - \alpha(\langle v_i, v_i\rangle\, u_{ik}^{(t-1)} - \langle y_k, v_i\rangle), \quad (9)$$

where α is a stepsize parameter of the gradient method. Then, $\mathbf{u}_k^{(t)}$ is to be taken as the projection of $\mathbf{d}_k^{(t)}$ in Q, denoted by $P_Q(\mathbf{d}_k^{(t)})$[1]. The process stops when the condition $\left|U^{(t)} - U^{(t-1)}\right| \leq \varepsilon$ is fulfilled.

Finding membership vectors by minimizing FCMP-1 and FCMP-2 is performed similarly.

Second, given membership matrix U, the optimal prototypes are determined according to the first-degree optimum conditions as

$$v_{ih}^{(t)} = \frac{\sum_{k=1}^{n}\left(u_{ik}^{(t)}\right)^a y_{kh}}{\sum_{k=1}^{n}\left(u_{ik}^{(t)}\right)^{a+1}}, \quad (10)$$

where parameter a takes value $a = 1, 2, 3$ for FCMP-0, FCMP-1 and FCMP-2, respectively. This equation resembles that for FCM, (3), though here the denominator is powered by $a + 1$, not a as the numerator. In practice, the FCM equation (3) is used with the power equal to 2, which corresponds to $a = 2$, the case of FCMP-1.

Thus, the algorithm consists of "major" iterations of updating matrices U and V and "minor" iterations of recalculation of membership values in the gradient projection method within each of the "major" iterations.

The algorithm starts with a set $V^{(0)}$ of c arbitrarily or expertly specified prototype points in $\Re^p$ and U^0; it stops when the difference between successive prototype matrices becomes small.

The algorithm converges only locally (for FCMP-1 and FCMP-2). Moreover, with a "wrong" number of clusters pre-specified, FCMP-0 may not converge at all since FCMP-0 may shift some prototypes to infinity (see [10] for an explanation of this phenomenon). In our experiments, the number of major iterations in FCMP algorithms when they converge is small, which is exploited as a stopping condition: when the number of major iterations in an FCMP run goes over a large number (in our calculations, over 100), that means the process does not converge.

IV. Experimental Study

The main goal of this experimental study is twofold. First, to study FCMP as a model-based data clustering approach. More precisely, to analyze the ability of FCMP to restore the original prototypes from which data are generated. Second, to compare FCMP algorithms with FCM (with its parameter $m = 2$).

In order to study characteristics of the FCMP model in identifying a cluster structure, the model should be tested on data exhibiting its own cluster structure (a *cluster tendency* [14]). To accomplish this, a random data generator was constructed as follows.

1. The dimension of the space (p), the number of clusters (c_0) and numbers $n_1, \ldots, n_{c_0}$ are randomly generated within prespecified intervals. The data set cardinality is defined as $n = \sum_{i=1}^{c_0} n_i$.

[1] The projection $P_Q(\mathbf{d}_k^{(t)})$ is based on an algorithm we developed for projecting a vector $\mathbf{d}_k^{(t)}$ over the simplex of membership vectors $\mathbf{u}_k^{(t)}$; its description is omitted here.

2. c_0 cluster directions are defined as follows: vectors $\mathbf{o}_i \in \Re^p$ $(i = 1, \cdots, c_0)$ are randomly generated within a prespecified cube; then, their gravity center $\mathbf{o}$ is calculated.
3. For each i, define two p-dimensional sampling boxes: one within bounds $A_i = [0.9\mathbf{o}_i, 1.1\mathbf{o}_i]$ and the other within $B_i = [\mathbf{o}, \mathbf{o}_i]$;then generate randomly $0.1n_i$ points in A_i and $0.9n_i$ points in B_i.
4. The data generated are normalized by centering to the origin and scaling by the range.

To visualize generated data, they are projected into a 2D/3D space of the best principal components.

To compare FCM and FCMP algorithms, the emphasis will be done with regard to the clustering results rather than the performance of the algorithms. It is quite obvious that our criteria (7), (8) are more complex than that of FCM, (2), and thus require more calculations.

In our experiments, each of the algorithms, FCM, FCMP-0, FCMP-1 and FCMP-2 has been run on the same data set (with the same initial setting) for different values of c $(c = 2, 3, 4 \ldots)$.

The clustering solutions found by FCMP algorithms have been characterized by the following features: 1) number of clusters found, c'; 2) cluster separability; 3) proximity to the FCM found prototypes; and 4) proximity to the original prototypes of generated data. The separability index was also calculated for FCM solutions as well as their proximity to the original prototypes.

The separability index, $B_c = 1 - \frac{c}{c-1}\left(1 - \frac{1}{n}\sum_{k,i}(u_{ik})^2\right)$, assesses the fuzziness of partition U; it takes values in the range $[0, 1]$ such that $B_c = 1$ for hard partitions and $B_c = 0$ for the uniform memberships (cf. [3], pp. 157, for a detailed description).

The difference between FCMP prototypes ($\mathbf{v}'_i$) and "reference" prototypes ($\mathbf{v}_i$) (in our experiments, the original prototypes or FCM ones), is defined by

$$D_{\mathbf{v}} = \frac{\sum_{i=1}^{c}\sum_{h=1}^{p}(v'_{ih} - v_{ih})^2}{\sum_{i=1}^{c}\sum_{h=1}^{p}v_{ih}^2 + \sum_{i=1}^{c}\sum_{h=1}^{p}v_{ih}'^2}, \qquad (11)$$

which measures the squared relative quadratic mean difference between corresponding prototypes $\mathbf{v}_i$ and $\mathbf{v}'_i$. Matching between prototypes in $\mathbf{v}$ and $\mathbf{v}'$ is determined according to smallest distances. When the number of prototypes c' found is smaller than c, only c' prototypes are considered in (11).

For fixed p and c, a group of 15 data sets were generated with different numbers of entities and different prototypes. The experiments comprised 7 such groups with p ranging from 5 to 180 and c, from 3 to 6.

For each group of data sets of the same dimension (p) and generated prototypes (c_0), the four algorithms have been compared based on the number of major iterations (t_1), number of prototypes found (c'), separability coefficient B_c, distance (D_{FCM}) to FCM prototypes and distance (D_0) to the original prototypes.

The results of running FCM and FCMP algorithms for those data sets lead us to distinguish between low and high dimensions of the data (with several hundred of entities, $\frac{p}{c_0} \leq 5$ is considered small and $\frac{p}{c_0} \geq 25$ is considered high). Tables I and II display the (average) results of running FCM and FCMP algorithms for two groups of the data sets: one of small dimension ($p = 5$, $c_0 = 3$), and the other of high dimension ($p = 180$, $c_0 = 6$).

TABLE I

Average results of running FCM and FCMP algorithms for 15 data sets of small dimension space ($p = 5, c_0 = 3$), generated with $c_0 = 3$ prototypes.

c_0=3 p=5	t_1	c'	B_c	D_{FCM} (%)	D_0 (%)
FCM	12	3	0.61	-	14.7
FCMP-0	10	3	0.84	0.49	12.2
FCMP-1	11	3	0.80	0.89	10.2
FCMP-2	11	3	0.43	7.10	2.3

TABLE II

Average results of running FCM and FCMP algorithms for 15 data sets in a high dimension space ($p = 180, c_0 = 6$), generated with $c_0 = 6$ prototypes. **A)** FCMP-0 converges to the original number of prototypes; **B)** FCMP-0 underestimates the number of prototypes.

c=c_0=6 p= 180	t_1	c'	B_c	D_{FCM} (%)	D_0 (%)
			A		
FCM	27	1	0.01	-	96.8
FCMP-0	60	6	0.78	143.5	11.7
FCMP-1	11	6	1.00	94.2	15.8
FCMP-2	20	6	0.30	97.2	0.45
			B		
FCM	37	1	0.03	-	97.9
FCMP-0	101	5	0.78	39.5	6.3
FCMP-1	12	6	1.00	77.4	15.3
FCMP-2	20	6	0.30	73.4	0.47

The results of the experiments can be summarized as follows.
1. For the low dimension data sets, FCMP-0 almost always finds the "natural" number of clusters present in data (i.e. corresponding to the number of the original prototypes, c_0). In particular, when the algorithm is run for a higher number of prototypes ($c > c_0$), it removes the extra prototypes out of the data space, thus preventing it from convergence. In the high dimension spaces, FCMP-0 finds the correct number of clusters in 50% of the cases (see Table II**A**). For the other data sets (see Table II**B**), FCMP-0 underestimates the number of clusters. However, in high dimensional spaces the FCM typically leads to even smaller number of clusters, (see Tables II**A** and

IIB), making several of the initial prototypes converge to the same points. Further experiments show that this feature of FCM depends not only on the space dimension but is also influenced by the generated data structure. Specifically, for the high dimension generated data, if we increase the proportion of points generated around the original prototypes (within the boxes A_i, phase 3 of the data generator) from 0.1 to 0.8, FCM is able to identify the correct number of prototypes from which data have been generated. Issues related to high-dimensional spaces have been discussed in [19]. As to FCMP-1 and FCMP-2 algorithms, they always converge to a solution for the various values of c (i.e. $c' = c$), and seem to be not influenced by the space dimension.

2. The prototypes found by FCMP-0 and FCMP-1 almost coincide with those found by FCM when the number of centroids has been determined by FCM correctly (as in Table I); FCMP-2 found prototypes are different from FCM corresponding ones (see corresponding entries D_{FCM} in Tables I and II).

3. Concerning the proximity (D_0) to the original prototypes of generated data, FCMP-2 prototypes almost always coincide with the original ones. On the contrary, FCM found prototypes are distant from the original ones. These results suggest that FCMP-2 clustering criterion is the one that better fits the FCMP model, at least, with our data generator.

4. According to the separability coefficient, B_c, FCMP-0 and FCMP-1 partitions are more contrast than FCM ones. In particular, in high dimension spaces FCMP-1 leads to hard clustering solutions. The FCMP-2 gives the fuzziest results, typically differing from those in FCM.

5. On average, the number of major iterations (t_1) in FCMP-1 and FCMP-2 are smaller than that in FCM, while in FCMP-0 this number does not differ significantly from that in FCM (in the case of small dimensions). However, the running time is greater for FCMP algorithms, because of the time spent for minor iterations with the gradient projection method.

V. Conclusion

The FCMP framework proposes a model of how the data are generated from a cluster structure to be identified. This implies direct interpretability of the fuzzy membership values, which should be considered a motivation for introducing the model-based methods.

Based on the experimental results obtained in this research, the following can be stated. The FCMP-2 algorithm is able to restore the original prototypes from which data have been generated, and FCMP-0 can be viewed as a device for estimating the number of clusters in the underlying structure to be found. For small dimension spaces, FCMP-1 is an intermediate model between FCMP-0 and FCMP-2, and can be viewed as a model based parallel to FCM. On the high dimension data, FCMP-1 degenerates in a hard clustering approach. Also, FCM drastically decreases the number of prototypes in the high dimension spaces (at least with the proposed data generator).

This model-based clustering approach seems appealing in the sense that, on doing cluster analysis, the experts of a knowledge domain usually have conceptual understanding of how the domain is organized in terms of tentative prototypes. This knowledge may well serve as the initial setting for data based structurization of the domain. In such a case, the belongingness of data entities to clusters are based on how much they share the features of corresponding prototypes. This seems useful in such application areas as mental disorders in psychiatry or consumer behavior in marketing. However, the effective utility of the multiple prototypes model still remains to be demonstrated with real data.

Acknowledgment

This research has been carried out in the Department of Computer Science of FCT- Universidade Nova de Lisboa and DIMACS (Center for Discrete Mathematics and Theoretical Computer Science), Rutgers University, NJ, USA. One of the authors (S.N.) was also supported by a grant from PRAXIS XXI program, of the Portuguese Foundation for Science and Technology (FCT). Dr. Fred Roberts, Director of DIMACS, is greatly acknowledged for providing B.M. and S.N. with opportunities to visit DIMACS for several times in 1998-1999. The authors thank Dr. Boris Polyak for his valuable suggestions with regard to organization of computations.

References

[1] L. Zadeh, "Fuzzy sets", *Information and Control*, **8**, pp. 338–352, 1965.

[2] J. Dunn, "A fuzzy relative of the Isodata process and its use in detecting compact, well-separated clusters", *Journal of Cybernetics*, **3**(3), pp. 32–57, 1973.

[3] J. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.

[4] J. Bezdek and R. Hathaway, "Recent convergence results for the fuzzy c-means clustering algorithms", *Journal of Classification*, **5**(2), pp. 237–247, 1988.

[5] G. Gustafson and W. Kessel, "Fuzzy clustering with a fuzzy covariance matrix", in *Proc. IEEE CDC*, pp. 761–766, San Diego, 1979.

[6] F. Klawonn and A. Keller, "Fuzzy clustering based on modified distance measures", J. Kok, D. Hand and M. Berthold (eds.), *Advances in Intelligent Data Analysis, Third International Symposium*,(IDA'99), Lecture Notes in Computer Science, 1642, Springer-Verlag, pp. 291–301, 1999.

[7] R. Kruse, F. Hoppner, F. Klawonn and T. Runkler, *Fuzzy Cluster Analysis*, John Wiley and Sons, 1999.

[8] R. Dave, "Fuzzy shell-clustering and applications to circle detection of digital images", *International Journal of General Systems*, **16**(4), pp. 343–355, 1990.

[9] L. Bobrowski and J. Bezdek, "c-means with l_1 and l_∞ norms", *IEEE Transactions on Systems, Man and Cybernetics*, **21**(3), pp. 545–554, 1991.

[10] S. Nascimento, B. Mirkin and F. Moura-Pires, "Multiple prototypes model for fuzzy clustering", in J. Kok, D. Hand and M. Berthold, (eds.), *Advances in Intelligent Data Analysis.*

Third International Symposium,(IDA'99), Lecture Notes in Computer Science, 1642, Springer-Verlag, pp. 269–279, 1999.

[11] B. Mirkin and G. Satarov, "Method of fuzzy additive types for analysis of multidimensional data: I, II", *Automation and Remote Control*, **51**(5, 6), pp. 683–688, pp. 817–821, 1990.

[12] B. Mirkin, *Mathematical Classification and Clustering*, Kluwer Academic Publishers, 1996.

[13] E. Ruspini, "A new approach to clustering", *Information and Control*, **15**, pp. 22–32, 1969.

[14] A. Jain and R. Dubes, *Algorithms for Clustering Data*, Prentice Hall, Englewood Cliffs, NJ, 1988.

[15] S. Nascimento, "Soft clustering with the multiple prototype fuzzy clustering model", in F. Costa Nicolau, H. Bacelar-Nicolau and J. Janssen, (eds.), *Applied Stochastic Models and Data Analysis. Quantitative Methods in Business and Industry Society*, Instituto Nacional de Estatística (INE), pp. 249–255, 1999.

[16] S. Selim and M. Ismail, "Soft clustering of multidimensional data: a semi-fuzzy approach", *Pattern Recognition*, **17**(5), pp. 559–568, 1984.

[17] B. Polyak, *Introduction to Optimization*, Optimization Software, Inc., New York, 1987.

[18] D. Bertsekas, *Nonlinear Programming*, Athena Scientific, Massachusetts, USA, 1995.

[19] L. Jimenez and D. Landgrebe, "Supervised classification in high-dimensional space: geometrical, statistical and asymptotical properties of multivariate data", *IEEE Transactions on Systems, Man and Cybernetics- Part C*, **28**(1), pp. 39–54, 1998.