F.S.A.I., Vol. 11, No. 1-3, pp. 33-44, 2005

Expert Rules for Partitive Clustering Methods Using Potential Function - Based Algorithms

Viorel Nicolau[*], Dorel Aiordachioaie[*], Gheorghe Puscasu[**], Constantin Miholca[*]

[*] Department of Electronics and Telecommunications, Faculty of Electrical Engineering,
"Dunarea de Jos" University of Galati, 47 Domneasca Street, Galati 800008, Romania
Email: Viorel.Nicolau@ugal.ro
[**] Department of Automatic Control, Faculty of Computer Science,
"Dunarea de Jos" University of Galati, 47 Domneasca Street, Galati 800008, Romania

Abstract. The clustering algorithms based on potential functions are capable of clustering a set of data, making no implicit assumptions on the cluster shapes and without knowing in advance the number of clusters. Also, they do not use any prototype vectors of the clusters. In this paper, some expert rules of potential function-based algorithms (PFBA) are generated based on their intrinsic properties and clustering tendency. These rules can be used in selecting process of parameter values or to generate a fuzzy classifier, so that the best clustering to be obtained with less seeking efforts.

Keywords. expert rules, potential function, partitive clustering methods, potential function-based algorithm.

1 Introduction

In general, a clustering **M** means partitioning a data set *S* into a set of clusters M_i, $i = 1...C$, based on a particular relationship between the vectors of the data set, using unsupervised methods.

Each data vector can belong to exactly one cluster, like in crisp clustering, or it can have different values of membership degrees to different clusters, like in fuzzy clustering [3]. In addition, mixture models can be used, which assumes that data are generated by several parameterized distributions [6].

A data set clustering can be done in two main ways: hierarchical and partitive approaches. The hierarchical methods include agglomerative and divisive algorithms, corresponding to bottom-up and top-down strategies to build a hierarchical clustering tree, which can be used for interpretation of the data structure [9].

Partitive clustering algorithms divide a data set into a number of clusters according to a generic inter-point measure of similarity or dissimilarity, trying to obtain an optimum value of a performance criterion. Density-based methods and distance-based methods are the most commonly used classes of partitive algorithms. These methods are better than hierarchical ones in the sense that they

do not depend on previously found clusters, but some of them make implicit assumptions on the form and number of clusters.

In general, the distance-based methods are used for unsupervised clustering problems. The methods include algorithms based on distance of a point to the prototype vectors and algorithms based on potential functions.

The first type of algorithms, such as *k-means* and *ISODATA*, uses a measure of dissimilarity, which is the distance between the points of the data set and the prototype vectors. These algorithms try to find spherical clusters and use prototype vectors. Also, the number of clusters is usually predefined.

The second type, called potential function-based algorithms (*PFBA*), uses a measure of similarity created with a function between two points of the data set, called potential function. This is a non-increasing function with the distance between the points. *PBFA* are capable of clustering a set of data, making no implicit assumptions on the cluster shapes and without knowing in advance the number of clusters. In addition, they do not use any prototype vectors of the clusters. The main drawback of these algorithms is that they are intensive computational. Therefore, it is very important to understand the essence of *PBFA* for developing more efficient algorithms.

The goal of this paper is to generate some expert rules of potential function-based algorithms (*PFBA*) based on their intrinsic properties and clustering tendency. The knowledge base can be used in selecting process of clustering parameter values or to generate a fuzzy classifier, even with adaptive properties [1], so that the best clustering to be obtained with less seeking efforts.

The paper is organized as follows. Section 2 describes the potential function-based algorithms and their validity indices. In section 3, some intrinsic properties of *PFBA* are analyzed and expert rules are generated. In section 4, the clustering tendency is illustrated and more rules are added to the knowledge base. Conclusions are presented in section 5.

2 Potential Function-Based Algorithms

2.1 Optimal Clustering and Validity Indices

In general, optimal clustering means partitioning a data set into a set of clusters, which minimizes distances within- and maximizes distances between-clusters. However, within- and between- cluster distances can be defined in several ways.

An example of within-cluster distance is centroid distance:

$$d_c = \frac{\sum_{i=1}^{N_k} d(x_i, c_k)}{N_k}, \tag{1}$$

where N_k represents the number of vectors in cluster Q_k. Also, $x_i \in Q_k$ and c_k is the center of gravity of Q_k:

$$c_k = \sum_{i=1}^{N_k} \frac{x_i}{N_k} \tag{2}$$

Corresponding to the centroid distance, there is a between-cluster distance, named centroid linkage:

$$D_c = d(c_i, c_k), \tag{3}$$

where c_i, c_k are the center of gravity of clusters Q_i and Q_k, respectively.

To select the best one from many partitions, a validity index can be used to evaluate them. Different validity indices can be defined [2], depending on which distances are considered. For example, the Davies-Bouldin index uses d_c as within-cluster distance and D_c as between-cluster distance. In this case, the best clustering minimizes the expression:

$$I_{DB} = \frac{1}{C} \cdot \sum_{i=1}^{C} \max_{i \neq k} \left(\frac{d_c(Q_i) + d_c(Q_k)}{D_c(Q_i, Q_k)} \right), \tag{4}$$

where C is the number of clusters. This index is suitable for evaluation of partitions with spherical clusters, the best partition being indicated by the index with the minimum value.

2.2 Potential Functions

Consider a data set S of N input vectors into a d-dimensional space:

$$S = \left\{ x_i \mid x_i = (x_{1i}, x_{2i}, ..., x_{di})^T \in \Re^d, \ i = \overline{1, N} \right\} \tag{5}$$

A potential function $K(x_i, x_k)$ associated with the vector $x_i \in S$ defines a positive value, called potential of the point x_i to the reference point $x_k \in \Re^d$. The potential depends on distance between the points x_i and x_k, denoted $d_{ik} = d(x_i, x_k)$, and it is a non-increasing function with d_{ik}. Two potential functions are commonly used:

$$K_1(x_i, x_k) = \frac{1}{1 + \alpha \cdot d_{ik}^2} \tag{6}$$

$$K_2(x_i, x_k) = \exp(-\alpha \cdot d_{ik}^2), \tag{7}$$

where parameter α controls the slope of the function.

The potential values belong to the range (0, 1] and the maximum value is obtained for $d_{ik}=0$. The distance d_{ik} can be the general Minkovski distance:

$$d(x,y)=\sqrt[p]{\sum_{i=1}^{d}\left|x_i-y_i\right|^p}\ ,\quad x,y\in\Re^d\ , \tag{8}$$

where for *p*=2 Euclidean distance is obtained, which is considered in this paper.

2.3 The Algorithm Stages

A potential function-based algorithm uses a measure of similarity, which characterizes the membership of a point to a group of points, based on a potential function [5].

Consider a group of points M from S, M $\subset$ S and a point $x_i\in S$, $x_i\notin M$. A similarity measure of x_i to M can be defined as the average A_i of the potential values of the point x_i to all points of the group M:

$$A_i=A(x_i,M)=\frac{1}{N_M}\cdot\sum_{x_j\in M}K(x_i,x_j)\ , \tag{9}$$

where N_M represents the number of points in M.

Using this measure of similarity, the points of the data set S can be arranged in a certain order, starting from a specified point, pursuant to the following rule:

- select the starting point, let it be $x^1\in S$, form the first group $M_1=\{x^1\}$ and denote $A_1=1$, which represents the maximum potential value;
- find in S/M_1 the point x^2 with the maximum measure of similarity to M_1 in the meaning of (9), which is:

$$A_2=A\left(x^2,M_1\right)=\max_{x\in S/M_1}\left(A(x,M_1)\right) \tag{10}$$

- form a new group $M_2=\{M_1,x^2\}=\{x^1,x^2\}$;
- repeat the previous step until all the points of the data set S are assigned, by finding the points x^k with maximum measure of similarity to M_{k-1}:

$$A_k=A\left(x^k,M_{k-1}\right)=\max_{x\in S/M_{k-1}}\left(A(x,M_{k-1})\right) \tag{11}$$

• form the groups $M_k = \{M_{k-1}, x^k\}$.

In this way, the set S is ordered, $S = \{x^1, x^2, ..., x^N\}$ and a new series is obtained: $A_1, A_2, ..., A_N$.

All potential function-based algorithms compute the new series $A_1, A_2, ..., A_N$, which contains the necessary information for clustering. The analysis of this series differs from algorithm to algorithm. For example, the algorithm considered in this paper [4] has the following stages:

- select the starting point;
- arrange the points of the data set S, using the rule described above;
- compute the ratios $R_1, ..., R_N$, where

$$R_1 = 1, \quad R_k = \frac{A_{k-1}}{A_k}, \quad k = \overline{2, N} \tag{12}$$

- compute the mean value m_R and the standard deviation σ_R of the ratios R_k;
- consider a threshold $p = r \cdot c \cdot \sigma_R$, where r is a variable integer (r = 1...20) and c is a weighting scalar, c ∈ [0.3, 1];
- the clustering decision is made comparing the difference $R_k - R_{k-1}$ with the threshold and a new cluster begin if $R_k - R_{k-1} > p$. Thus, a new partition is obtained;
- compute other partitions for different threshold values, by increasing r, until $p > R_k - R_{k-1}$ for all differences.

The clustering result is considered the partition, which remains unchanged for the greatest number of *r*-values.

3 Intrinsic Properties Of *Pfba*

3.1 Relative Variation between Potential Functions

The potential functions are smoothly if parameter α has small values. The function variations with d_{ik} for three different values of α are illustrated in Figure 1. The potential functions K_2 are represented with continuous lines and K_1 functions are represented with dotted lines.

In Figure 1, for the same α value, the two potential functions have similar values if distances between the points are small. In this case, the potential values are high and the clustering performances are similarly for both functions. Therefore, if the input data sets are normalized, the distances are smaller than unity value and the potential functions produce similar clustering performances.

To define the distance range where the two potential functions are similarly, the relative variation between them (ΔK) is computed, by imposing a maximum value of its tolerance (*tol*):

$$\Delta K(d_{ik}) = \frac{K_1(d_{ik}) - K_2(d_{ik})}{K_1(d_{ik})} \leq tol \tag{13}$$

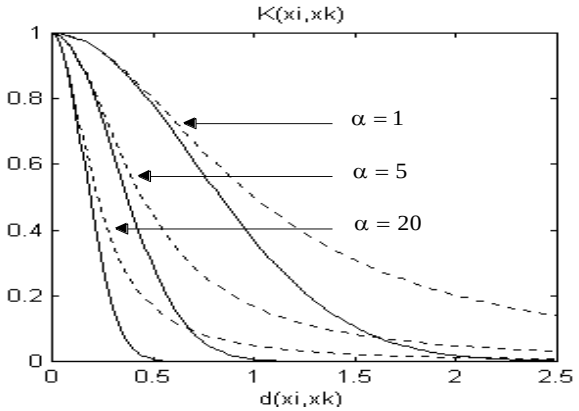

Fig. 1. Potential functions for three values of α

An analytical condition of (13) can be obtained if the exponential function K_2 is approximated by a rational function, such as Pade approximation:

$$\exp(-x^2) \cong \frac{2 - x^2}{2 + x^2} \tag{14}$$

The approximation (14) generates a relative error smaller than 1%, if the argument is within the range:

$$x^2 \leq 0.5 \quad \Rightarrow \quad x \leq 0.7 \tag{15}$$

Computing the inequality (13) with K_2 replaced by (14), it results the condition for the distance d_{ik} :

$$\alpha \cdot d_{ik}^2 \leq \frac{tol + \sqrt{tol \cdot (8 + tol)}}{2} \tag{16}$$

For example, by choosing $tol = 0.1 = 10\%$, it results:

$$\alpha \cdot d_{ik}^2 \leq \frac{1}{2} \quad \Rightarrow \quad d_{ik} \leq \sqrt{\frac{1}{2 \cdot \alpha}} \tag{17}$$

It can be observed that the distance range which verifies the condition (17) depends on the value of α parameter.

Rule no. 1: If all the distances between the points (d_{ik}) are small enough to verify the condition (16), then the relative variation of potential functions is smaller than selected tolerance (*tol*) and the clustering performances are similarly for both functions within the given tolerance.

3.2. Selection of Function Parameter

Clustering results depend on the parameter value (α) of potential functions. For optimum value of α parameter, the distance d_{ik} corresponding to maximum variation of potential functions must be computed.

For the two potential functions, results:

$$K_1^{''}(d_{ik}) = 0 \Rightarrow d_1 = \sqrt{\frac{1}{3 \cdot \alpha}} \Rightarrow K_1(d_1) = 0.75 \tag{18}$$

$$K_2^{''}(d_{ik}) = 0 \Rightarrow d_2 = \sqrt{\frac{1}{2 \cdot \alpha}} \Rightarrow K_2(d_2) = 0.6 \tag{19}$$

It is desired to obtain maximum variation of potential functions for most distances between the points. Hence, based on average distance between data points (d_{av}), an optimum parameter value is computed, so that the sensibility of potential function with distance to be maximized. For that, the average distance is used in (18) and (19) instead of d_1 and d_2, resulting:

$$\alpha_1 = \frac{1}{3 \cdot d_{av}^2} \text{ and } \alpha_2 = \frac{1}{2 \cdot d_{av}^2} \tag{20}$$

Rule no. 2: For the input data set, the function parameter must be selected, based on the average distance between data points (d_{av}), like in (20).

3.3 Constant Potential Surfaces

A constant potential value to a reference point $x_k \in \Re^d$ is obtained by the potential function $K(x_i, x_k)$ associated with the points $x_i \in \Re^d$ for which the distance d_{ik} is constant. All points x_i generate a constant potential surface, whose shape depends on distance definition. For Euclidean distance, the constant potential surface has spherical shape around the reference point $x_k \in \Re^d$ [8].

The parameter α also affects the constant potential surface, different α values generating different potential surfaces, but their shapes are similar around the reference point. If α value increases, the potential surface is moving nearer to the reference point.

Similarly, a potential value of a point x_i to a group of reference points $M=\{x_{k1}, x_{k2}, ..., x_{km}\}$ can be defined as the average of the potential values of the point x_i to all reference points x_{kj}:

$$A_i = A(x_i, M) = \frac{1}{m} \cdot \sum_{j=1}^{m} K(x_i, x_{kj}) \tag{21}$$

In this case, a constant potential value to the group M generates a potential surface, which also depends on distance definition. The constant potential surfaces surround the reference points, but their shapes are affected by α-values and reference point positions. For example, two reference points $x_{k1}, x_{k2} \in \Re^2$ and a constant potential value $K = 0.5$ are considered. The reference points are represented with '+' in Figure 2.

For every reference point, three different constant potential surfaces are generated with potential function $K_2(x_i, x_k)$, illustrated with dotted lines, corresponding to K and three α-values: 2, 5 and 10. In Figure 2, the constant potential surfaces to a single reference point have spherical shapes around the point, being represented with dotted lines. The big circle represent the first potential surface, corresponding to $\alpha = 2$.

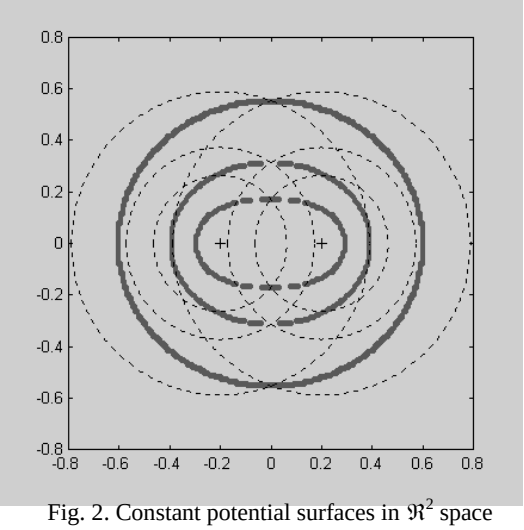

Fig. 2. Constant potential surfaces in $\Re^2$ space

The potential surfaces generated by the constant potential value K to the reference group $\{x_{k1}, x_{k2}\}$ are represented with continuous lines in Figure 2, corresponding to the same α-values. They surround the group of reference points, but the shapes depend on α-values and reference point positions. For small α-values, the shape tends to be spherical.

Rule no. 3: The shape of potential surface generated by constant potential value depends on the reference point positions. As potential value decreases, the shape tends to the one generated by general Minkovski distance.

4 Clustering Tendency

4.1 Arrangement of Data Points

Arrangement of the points in the ordered data set S depends on selections of first point, potential function type and α parameter. The tendency is to order the points in successive layers around the first points. Also, these selections affect the values of series A_k and R_k, and can affect the clustering performance [7].

To illustrate the influence of the first point selection and the ordering tendency of the points, a data set is clustered starting from two different points. Consider the data set I, with two spherical and well-separated clusters, as shown in Figure 3. The clusters have 40 and respectively 100 points, which are denoted $x_1,...,x_{140}$, marked with '+' in Figure 3.

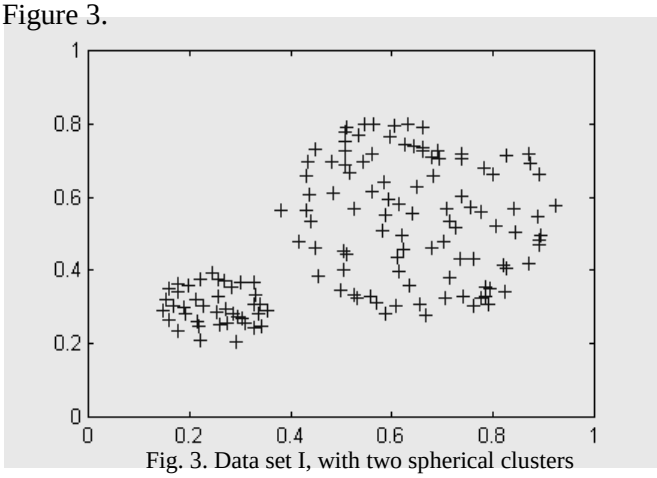

Fig. 3. Data set I, with two spherical clusters

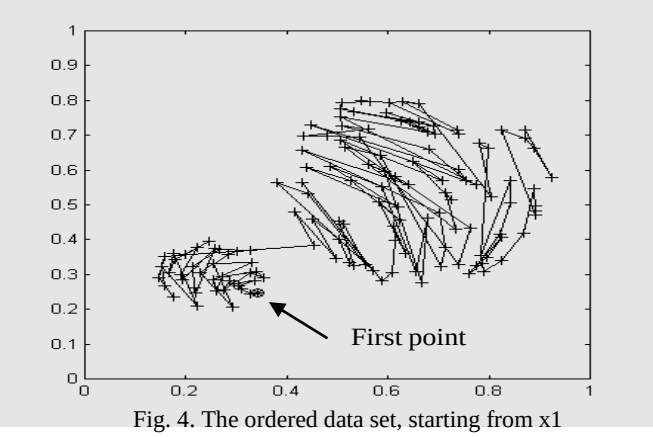

Fig. 4. The ordered data set, starting from x1

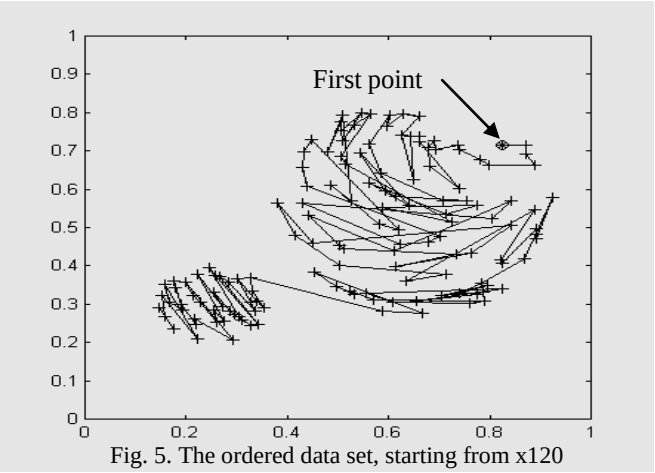

Fig. 5. The ordered data set, starting from x120

Running the clustering algorithm twice, with the first point into the small cluster and then into the big cluster, the arrangements of the data set points are different and are illustrated in Figures 4 and 5. The potential function is K_1 and the parameter α=10. In these figures, the points are marked with '+' and the start point is marked distinctly. In addition, lines are drawn between every two consecutive points in the ordered data set S.

In Figure 4, the arrangement starts with the first point of the data set x_1, which is into the small cluster, and in Figure 5 the start point is x_{120}, which is into the big cluster. It can be observed that the points are ordered in successive layers around the first ordered points.

Rule no. 4: The arrangement of the points in the ordered data set S depends on selection of first point, the tendency being to order the points in successive layers around the first points.

4.2 Influence of Function Parameter

The constant potential surface to a group of points M depends on potential function and parameter α and tends to take similar shape as the one of the cluster when α increases, even for more complex cluster.

The influence of α values on the constant potential surfaces is illustrated for two different α values, using the potential function K_2. Increasing α value, the constant potential surface will be closely to the cluster points and the new cluster will be oblong. Thus, the parameter α can be used to characterize the shape of the clusters: more compact or oblong.

Consider a complex cluster M with 199 points and a new point x_{200}, which has the measure of similarity to M denoted A_{200}. The value of the constant potential surface was chosen equal to A_{200}, which is useful to compare new additional points

with x_{200}. For $\alpha=25$, the constant potential value is $A_{200}=0.055$ and the constant potential surface is illustrated with gray color in Figure 6. The points of the cluster are marked with '+' and the last point placed on the constant potential surface is marked with 'o'.

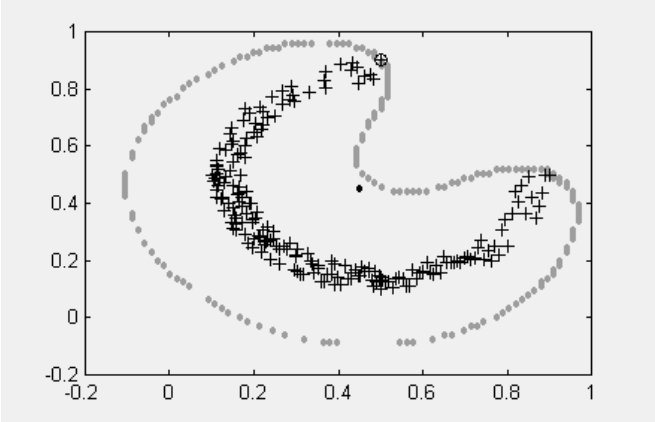

Fig. 6. Constant potential surface for α = 25

Similarly, for $\alpha=80$, the constant potential value is $A_{200}=0.029$ and the constant potential surface is illustrated with gray color in Figure 7. Additional points placed outer potential surface have measure of similarity to M smaller than A_{200} and the points are ordered after x_{200}. By contrast, any additional point placed into potential surface is ordered before x_{200}. For example, in Figures 6 and 7, the point at the (0.45, 0.45) coordinates is marked with '.'. This point is ordered before x_{200} if $\alpha=25$ and is ordered after x_{200} if $\alpha=80$.

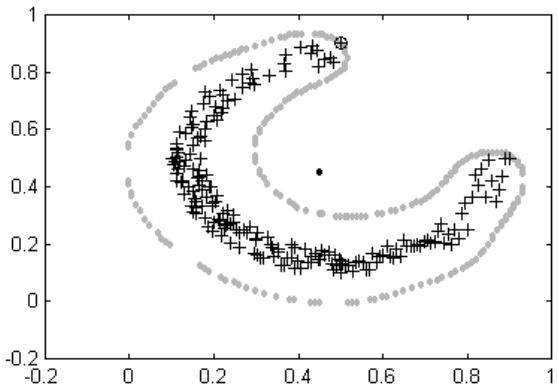

Fig. 7. Constant potential surface for α = 80

Rule no. 5: The constant potential surface to a group of points M tends to take similar shape as the one of the cluster when α increases, even for more complex cluster. Hence, the α parameter can be used to characterize the shape of the clusters: more compact or oblong.

5 Conclusions

The *PFBA* do not use any prototype vectors and they give good results even for complex shape clusters. Some expert rules of potential function-based algorithms were generated based on their intrinsic properties and clustering tendency. The potential functions give similar clustering performances if the distances between the points are small enough. The points are ordered in successive layers around the first point and the α parameter can be used to characterize the shape of the clusters: more compact or oblong. The knowledge base is useful in selecting process of parameter values or to generate a fuzzy classifier, so that the best clustering to be obtained with less seeking efforts.

References

[1] M. BALAS, V. BALAS: *The Family of Fuzzy Self Adaptive Interpolative Controllers*, in Proc. of the International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU 2004), Perugia, Italy, 2004

[2] J. C. BEZDEK: *Some new indexes of cluster validity*, IEEE Trans. Syst., Cybern., vol. **28**, pp. 301-315, 1998.

[3] J. C. BEZDEK, S.K. PAL: Fuzzy Models for Pattern Recognition: Methods that Search for Structures in Data, IEEE, New York, 1992.

[4] S. BUMBARU, E. CEANGA, I. BIVOL: *Self-learning classifier for data analysis*, in Proc of the VI-th Iugoslav International Symposium on Information Processing, vol. **H5**, pp. 1-6, Iugoslavia, 1970.

[5] A.A. DOROFEYUK: Algorithms of teaching the machine the pattern recognition without teacher based on the method of potential functions, Avtomatika i Telemechanica, vol. **10**, pp. 78-87, 1966.

[6] G.J. MCLAHLAN, K.E. BASFORD: *Mixture Models: Inference and Applications to Clustering*, Vol. **84**, Marcel Dekker, New York, 1987.

[7] V. NICOLAU, E. CEANGĂ, GHE. PUŞCAŞU: *Advantages of partitive clustering algorithms based on potential functions*, in Proc. of the SCS 13th European Simulation Symposium "Simulation in Industry", pp. 675-679, ISBN 90-77039-02-3, Marseille, France, 2001.

[8] V. NICOLAU, GHE. PUŞCAŞU, R. POPA: *Properties of potential function-based clustering algorithms*, 2000 Annals of "Dunarea de Jos" University of Galati, Fascicle III, pp. 30-35, ISSN 1221-454X, Galati, Romania, 2000.

[9] J. VESANTO, E. ALHONIEMI: *Clustering of the Self-Organizing Map*, IEEE Trans. on Neural Networks, vol. **11**, nr. 3, pp. 586-600, 2000.

This paper has been presented in a preliminary form at the IEEE SOFA 2005 Conference and has been selected for publication in F.S.A.I. based on the recommendation of the referees.