An ε-insensitive fuzzy c-means clustering

Jacek M. Łęski

Institute of Electronics.

Silesian University of Technology.

Akademicka 16, 44-100 Gliwice. Poland.

Abstract

Fuzzy clustering helps to find natural vague boundaries in data. The fuzzy c-means method is one of the most popular clustering methods based on minimization of a criterion function. However, one of the greatest disadvantage of this method is sensitivity for noises and outliers in the data. This paper introduces a new ε-insensitive Fuzzy C-Means (εFCM) clustering algorithm. As a special case, this algorithm include the well-known Fuzzy C-Medians method (FCMED). A new clustering algorithm is experimentally compared with the Fuzzy C-Means (FCM) using a synthetic data with outliers.

I. Introduction

Clustering plays an important role in many engineering fields such as pattern recognition, system modeling, image analysis, communication, data mining, and so on. The clustering methods divide a set of N observations (input vectors) $\mathbf{x}_1$,$\mathbf{x}_2$,...,$\mathbf{x}_N$ into c groups denoted $\Omega_1, \Omega_2, ..., \Omega_c$ so that members of the same group are more similar to one another than to members of other groups. The number of clusters may be pre-defined or it may be set by the method.

Generally, clustering methods can be divided into [6]: hierarchical, graph theoretic, decomposing of density function, minimizing of criterion function. In this paper clustering by minimization of criterion function will be considered.

Usually, the clustering methods assume that each data vector belongs to one and only one class. This method can be natural for clustering compact and well-separated groups of data. However, in practice clusters overlap, and some data vectors belong partially to several clusters. The fuzzy set theory [15] is a natural way to describe this situation. In this case, the membership degree of a vector $\mathbf{x}_k$ to the i-th cluster (u_{ik}) is a value from $[0, 1]$ interval.

This idea was first introduced by Ruspini [12] and used by Dunn [5] to construct a fuzzy clustering method based on the criterion function minimization. In [1] Bezdek generalized this approach to an infinite family of fuzzy c-means algorithm using a weighted exponent on the fuzzy memberships.

Fuzzy c-means clustering algorithm is successfully applied to a wide variety of problems [1]. However, one of the greatest disadvantage of this method is sensitivity for noises and outliers in the data. Computed cluster centers can be placed away from the true values. In literature there are a number of approaches to reduce the effect of outliers, including the possibilistic clustering [11], the fuzzy noise clustering [2], L_p norm clustering $(0 < p < 1)$ [7] and L_1 norm clustering [10]. In this paper the last approach is of special interest.

In real applications, the data have noises and outliers, and assumed (for simplicity) models are only approximators to reality. For example, if we assume that distribution of data in clusters are Gaussian, then using weighted (by membership degree) mean should not cause a bias. In this case L_2 norm is used as dissimilarity measure.

Noises and outliers existing in the data follows that clustering methods need to be robust. According to Huber [9], a robust method can have following properties: i) it should have a reasonably good accuracy at the assumed model, ii) small deviations from the model assumptions should impair the performance only by a small amount, iii) larger deviations from the model assumptions should not cause a catastrophe. In literature there are many robust estimators [3], [9]. In this paper Vapnik's ε-insensitive estimator is of special interest [14].

The goal of this paper is to establish a connection between fuzzy c-means clustering and robust statistics using Vapnik's ε-insensitive estimator. This paper presents a new fuzzy clustering method based on robust estimator. The fuzzy c-medians can be obtained as a special case of introduced clustering method.

This paper is organized as follows: Section II presents a short description of clustering methods based on criterion function minimization. A novel clustering algorithm is described in Section III. Section IV presents simulation results of a synthetic data with outliers clustering and a comparative study with the fuzzy c-means method. Finally, conclusions are drawn in Section V.

II. Clustering by minimization of criterion function

A very popular way of data clustering is to define a criterion function (scalar index) that measures the quality of a partition. In fuzzy approach [1] the set of all possible fuzzy partitions of N, p-dimensional vectors into c clusters is defined by:

$$\Im_{fc} = \left\{ \mathbf{U} \in \Re_{cN} \,\middle|\, \underset{\substack{1\le i\le c\\1\le k\le N}}{\forall} \quad u_{ik} \in [0,1],\ \sum_{i=1}^{c} u_{ik} = 1,\ 0 < \sum_{k=1}^{N} u_{ik} < N \right\}. \tag{1}$$

$\Re_{cN}$ denote a space of all real $(c \times N)$-dimensional matrices. The fuzzy c-means criterion function has the form [1]:

$$J_m(\mathbf{U},\mathbf{V}) = \sum_{i=1}^{c}\sum_{k=1}^{N} (u_{ik})^m \, d_{ik}^2, \tag{2}$$

where $\mathbf{U} \in \Im_{fc}, \mathbf{V} \in \Re_{pc}$. d_{ik} is the inner product induced norm:

$$d_{ik}^2 = \|\mathbf{x}_k - \mathbf{v}_i\|_{\mathbf{A}}^2 = (\mathbf{x}_k - \mathbf{v}_i)^T \mathbf{A} (\mathbf{x}_k - \mathbf{v}_i), \tag{3}$$

where $\mathbf{A}$ is a positive definite matrix and m is a weighting exponent in $[1,\infty)$. Criterion (2) for $m = 2$ was introduced by Dunn [5]. An infinite family of fuzzy c-means criterions for $m \in [1,\infty)$ were introduced by Bezdek. Using Lagrange multipliers the following theorem can be proved, via obtaining necessary conditions for minimization of (2) [1]:

Theorem 1: If m and c are fixed parameters, and $I_k, \widetilde{I}_k$ are sets defined as:

$$\underset{1\le k\le N}{\forall} \begin{cases} I_k = \{ i \,|\, 1 \le i \le c;\ d_{ik} = 0 \}, \\ \widetilde{I}_k = \{1,2,...,c\} \setminus I_k, \end{cases} \tag{4}$$

then $(\mathbf{U},\mathbf{V}) \in (\Im_{fc} \times \Re_{pc})$ may be globally minimal for $J_m(\mathbf{U},\mathbf{V})$ only if:

$$\underset{\substack{1\le i\le c\\1\le k\le N}}{\forall} \quad u_{ik} = \begin{cases} (d_{ik})^{\frac{2}{1-m}} \Big/ \left[\sum\limits_{j=1}^{c} (d_{jk})^{\frac{2}{1-m}} \right], & I_k = \emptyset, \\ \begin{cases} 0, & i \notin I_k \\ \sum\limits_{i \in I_k} u_{ik} = 1, & i \in I_k \end{cases}, & I_k \neq \emptyset, \end{cases} \tag{5}$$

and

$$\underset{1\le i\le c}{\forall} \quad \mathbf{v}_i = \left[\sum_{k=1}^{N} (u_{ik})^m \mathbf{x}_k \right] \Big/ \left[\sum_{k=1}^{N} (u_{ik})^m \right]. \tag{6}$$

The optimal partition is a fixed point of (5) and (6), and the solution is obtained from the Piçard algorithm. This algorithm is called fuzzy ISODATA or Fuzzy C-Means (FCM), and can be described as:

1° Fix c $(1 < c < N)$, $m \in (1, \infty)$. Initialize $\mathbf{V}^{(0)} \in \Re_{pc}$, $j = 1$,

2° Calculate fuzzy the partition matrix $\mathbf{U}^{(j)}$ for j-th iteration using (5),

3° Update the centers for j-th iteration $\mathbf{V}^{(j)} = [\mathbf{v}_1^{(j)} \mathbf{v}_2^{(j)} \cdots \mathbf{v}_c^{(j)}]$ using (6) and $\mathbf{U}^{(j)}$,

4° If $\left\| \mathbf{U}^{(j+1)} - \mathbf{U}^{(j)} \right\|_F > \xi$ then $j \leftarrow j + 1$ and go to 2°.

$\|\cdot\|_F$ denotes Frobenius norm and ξ pre-set parameters. In this algorithm, the parameter m influences the fuzziness of the clusters; a larger m results in fuzzier clusters. For $m \longrightarrow 1^+$, the fuzzy c-means solution becomes a hard one, and for $m \longrightarrow \infty$ the solution is as fuzzy as possible: $u_{ik} = 1/c$, for all i, k. Usually $m = 2$ is chosen, because there is no theoretical basis for the optimal selection of m.

III. A NEW CLUSTERING ALGORITHM

In clustering algorithm described in previous section, we used a quadratic loss function as a dissimilarity measure between the data and the cluster centers. The reason of using this measure is mathematical, that is, for simplicity and low computational burden. However this approach is sensitive to the noises and outliers. In literature there are many robust loss functions [9], but due to its simplicity Vapnik's ε-insensitive loss function [14] is specially interesting. If we denote an error as t, then ε-insensitive loss function has the form:

$$|t|_\varepsilon = \begin{cases} 0, & |t| \le \varepsilon, \\ |t| - \varepsilon, & |t| > \varepsilon, \end{cases} \tag{7}$$

where ε denotes insensitivity parameter. The well-known absolute error loss function is a special case of (7) for $\varepsilon = 0$.

Using the ε-insensitive loss function the fuzzy c-means criterion function (2) takes the form:

$$J_{m\varepsilon}(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^{c} \sum_{k=1}^{N} (u_{ik})^m \ |\mathbf{x_k} - \mathbf{v_i}|_\varepsilon , \tag{8}$$

where

$$|\mathbf{x_k} - \mathbf{v_i}|_\varepsilon = \sum_{l=1}^{p} |x_{kl} - v_{il}|_\varepsilon . \tag{9}$$

If $\mathbf{V} \in \Re_{cp}$ is fixed, then columns of $\mathbf{U}$ are independent, and the minimization of (8) can be performed term by term:

$$J_{m\varepsilon}(\mathbf{U}, \mathbf{V}) = \sum_{\mathbf{k}=1}^{\mathbf{N}} \mathbf{g_k}(\mathbf{U}) , \tag{10}$$

where

$$\underset{1\le k\le N}{\forall} \quad g_k(\mathbf{U}) = \sum_{i=1}^{c} (u_{ik})^m \ |\mathbf{x_k} - \mathbf{v_i}|_\varepsilon . \tag{11}$$

The Lagrangian of (11) with constraints from (1) is:

$$\underset{1\le k\le N}{\forall} \quad G_k(\mathbf{U}, \boldsymbol{\lambda}) = \sum_{i=1}^{c} (u_{ik})^m \ |\mathbf{x_k} - \mathbf{v_i}|_\varepsilon - \lambda[\sum_{i=1}^{c} u_{ik} - 1], \tag{12}$$

where λ is the Lagrange multiplier. Setting the Lagrangian's gradient to zero we obtain:

$$\underset{1\le k\le N}{\forall} \quad \frac{\partial G_k(\mathbf{U}, \boldsymbol{\lambda})}{\partial \lambda} = \sum_{i=1}^{c} u_{ik} - 1 = 0, \tag{13}$$

and:

$$\underset{\substack{1\le s\le c\\ 1\le k\le N}}{\forall} \quad \frac{\partial G_k(\mathbf{U}, \boldsymbol{\lambda})}{\partial u_{sk}} = m\,(u_{sk})^{m-1} |\mathbf{x_k} - \mathbf{v_s}|_\varepsilon - \lambda = 0. \tag{14}$$

From (14) we get:

$$u_{sk} = \left(\frac{\lambda}{m}\right)^{\frac{1}{m-1}} (|\mathbf{x_k} - \mathbf{v_s}|_\varepsilon)^{\frac{1}{1-m}} . \tag{15}$$

From (15), (13) we obtain:

$$\left(\frac{\lambda}{m}\right)^{\frac{1}{m-1}} \sum_{j=1}^{c} \left(|\mathbf{x_k} - \mathbf{v_j}|_\varepsilon\right)^{\frac{1}{1-m}} = 1. \tag{16}$$

Combination of (15) and (16) yields:

$$\underset{\substack{1\le s\le c\\ 1\le k\le N}}{\forall} \quad u_{sk} = (|\mathbf{x_k} - \mathbf{v_s}|_\varepsilon)^{\frac{1}{1-m}} \Big/ \left[\sum_{j=1}^{c} \left(|\mathbf{x_k} - \mathbf{v_j}|_\varepsilon\right)^{\frac{1}{1-m}}\right] . \tag{17}$$

If $I_k \neq \emptyset$, then choosing $u_{ik} = 0$ for $i \notin I_k$ and $\sum_{i\in I_k} u_{ik} = 1$ for $i \in I_k$ results in minimization of the criterion value in (8), because partition matrix elements are zeros for non-zero distances, and non-zero for zero distances. Finally, the necessary conditions for minimization of (8) with respect to $\mathbf{U}$ can be written in the following form:

$$\underset{\substack{1\le i\le c\\ 1\le k\le N}}{\forall} \quad u_{ik} = \begin{cases} (|\mathbf{x_k} - \mathbf{v_i}|_\varepsilon)^{\frac{1}{1-m}} \Big/ \left[\sum_{j=1}^{c} \left(|\mathbf{x_k} - \mathbf{v_j}|_\varepsilon\right)^{\frac{1}{1-m}}\right], & I_k = \emptyset, \\ \begin{cases} 0, & i \notin I_k \\ \sum_{i\in I_k} u_{ik} = 1, & i \in I_k \end{cases}, & I_k \neq \emptyset. \end{cases} \tag{18}$$

A key problem in the new clustering method is to obtain necessary conditions for prototypes matrix $\mathbf{V}$. Combination (8) and (9) yields:

$$J_{m\varepsilon}(\mathbf{U},\mathbf{V}) = \sum_{i=1}^{c}\sum_{k=1}^{N}(u_{ik})^m \sum_{l=1}^{p}|x_{kl}-v_{il}|_\varepsilon = \sum_{i=1}^{c}\sum_{l=1}^{p} g_{il}(v_{il}), \tag{19}$$

where

$$g_{il}(v_{il}) = \sum_{k=1}^{N}(u_{ik})^m \; |x_{kl}-v_{il}|_\varepsilon . \tag{20}$$

Our problem of minimization of (8) with respect to the prototypes can be decomposed to $c \cdot p$ minimization problems of (20) for $i = 1,...,c$; $l = 1,...,p$. In general case, not for all data x_{kl} following inequalities are satisfied: $|x_{kl}-v_{il}| \le \varepsilon$, $|v_{il}-x_{kl}| \le \varepsilon$. If we introduced slack variables $\xi_k^+, \xi_k^- \ge 0$, then for all data x_{kl} we can write:

$$\begin{cases} v_{il} - x_{kl} \le \varepsilon + \xi_k^+, \\ x_{kl} - v_{il} \le \varepsilon + \xi_k^- . \end{cases} \tag{21}$$

Now, criterion (20) can be written in the form:

$$g_{il}(v_{il}) = \sum_{k=1}^{N}(u_{ik})^m \; \left(\xi_k^+ + \xi_k^-\right), \tag{22}$$

subject to the constraints (21) and $\xi_k^+ \ge 0$, $\xi_k^- \ge 0$. The Lagrangian function of (22) with the above constraints is:

$$\begin{aligned} G_{il}(v_{il}) \;=\; & \sum_{k=1}^{N}(u_{ik})^m \; \left(\xi_k^+ + \xi_k^-\right) - \sum_{k=1}^{N}\lambda_k^+\left(\varepsilon + \xi_k^+ - v_{il} + x_{kl}\right) \\ & - \sum_{k=1}^{N}\lambda_k^-\left(\varepsilon + \xi_k^- + v_{il} - x_{kl}\right) - \sum_{k=1}^{N}\left(\mu_k^+\xi_k^+ + \mu_k^-\xi_k^-\right), \end{aligned} \tag{23}$$

where $\lambda_k^+, \lambda_k^-, \mu_k^+, \mu_k^- \ge 0$ are the Lagrange multipliers. The requirements is to minimize this Lagrangian with respect to v_{il}, ξ_k^+, ξ_k^-. It must be also maximized with respect to the Lagrange multipliers. The following conditions for optimality (the saddle point of the Lagrangian) we get by differentiating (23) with respect to v_{il}, ξ_k^+, ξ_k^- and setting the results equal to zero:

$$\begin{cases} \frac{\partial G_{il}(v_{il})}{\partial v_{il}} = \sum_{k=1}^{N}\left(\lambda_k^+ - \lambda_k^-\right) = 0, \\ \frac{\partial G_{il}(v_{il})}{\partial \xi_k^+} = (u_{ik})^m - \lambda_k^+ - \mu_k^+ = 0, \\ \frac{\partial G_{il}(v_{il})}{\partial \xi_k^-} = (u_{ik})^m - \lambda_k^- - \mu_k^- = 0. \end{cases} \tag{24}$$

From the last two conditions (24) and the requirements $\mu_k^+, \mu_k^- \geq 0$ we obtain that $\lambda_k^+, \lambda_k^- \in [0, (u_{ik})^m]$. Putting conditions (24) in the Lagrangian (23) we get:

$$G_{il}(v_{il}) = -\sum_{k=1}^{N} \left(\lambda_k^+ - \lambda_k^-\right) x_{kl} - \varepsilon \sum_{k=1}^{N} \left(\lambda_k^+ + \lambda_k^-\right). \tag{25}$$

Maximization of (25) subject to constraints:

$$\begin{cases} \sum\limits_{k=1}^{N} \left(\lambda_k^+ - \lambda_k^-\right) = 0, \\ \lambda_k^+, \lambda_k^- \in [0, (u_{ik})^m], \end{cases} \tag{26}$$

is the so called Wolfe dual formulation (problem). It is well-known from optimization theory that at the saddle point, for each Lagrange multiplier, the Karush-Kühn-Tucker conditions must be satisfied:

$$\begin{cases} \lambda_k^+ \left(\varepsilon + \xi_k^+ - v_{il} + x_{kl}\right) = 0, \\ \lambda_k^- \left(\varepsilon + \xi_k^- + v_{il} - x_{kl}\right) = 0, \\ \left((u_{ik})^m - \lambda_k^+\right) \xi_k^+ = 0, \\ \left((u_{ik})^m - \lambda_k^-\right) \xi_k^- = 0. \end{cases} \tag{27}$$

From last two conditions (27) we see that $\lambda_k^+ \in (0, (u_{ik})^m) \implies \xi_k^+ = 0$ and $\lambda_k^- \in (0, (u_{ik})^m) \implies \xi_k^- = 0$. In this case from the first two conditions (27) we have:

$$\begin{cases} v_{il} = x_{kl} + \varepsilon, \quad \text{for } \lambda_k^+ \in (0, (u_{ik})^m), \\ v_{il} = x_{kl} - \varepsilon, \quad \text{for } \lambda_k^- \in (0, (u_{ik})^m). \end{cases} \tag{28}$$

Thus, we may determine cluster center v_{il} for (28) by taking any x_{kl} for which we have the Lagrange multipliers in open interval $(0, (u_{ik})^m)$. However, from numerical point of view, it is better to take mean value of v_{il} obtained for all data for which the conditions (28) are satisfied.

Using described in this section method for cluster centers calculation we obtain the algorithm that can be called as ε-insensitive Fuzzy C-Means (εFCM):

1° Fix c $(1 < c < N)$, $m \in (1, \infty)$ and $\varepsilon \geq 0$. Initialize $\mathbf{V}^{(0)} \in \Re_{pc}$, $j = 1$,

2° Calculate the fuzzy partition matrix $\mathbf{U}^{(j)}$ for j-th iteration using (18),

3° Update the centers for j-th iteration $\mathbf{V}^{(j)} = [\mathbf{v}_1^{(j)} \mathbf{v}_2^{(j)} \cdots \mathbf{v}_c^{(j)}]$ using (25), (28) and $\mathbf{U}^{(j)}$,

4° If $\left\| \mathbf{U}^{(j+1)} - \mathbf{U}^{(j)} \right\|_F > \xi$ then $j \leftarrow j + 1$ and go to 2°.

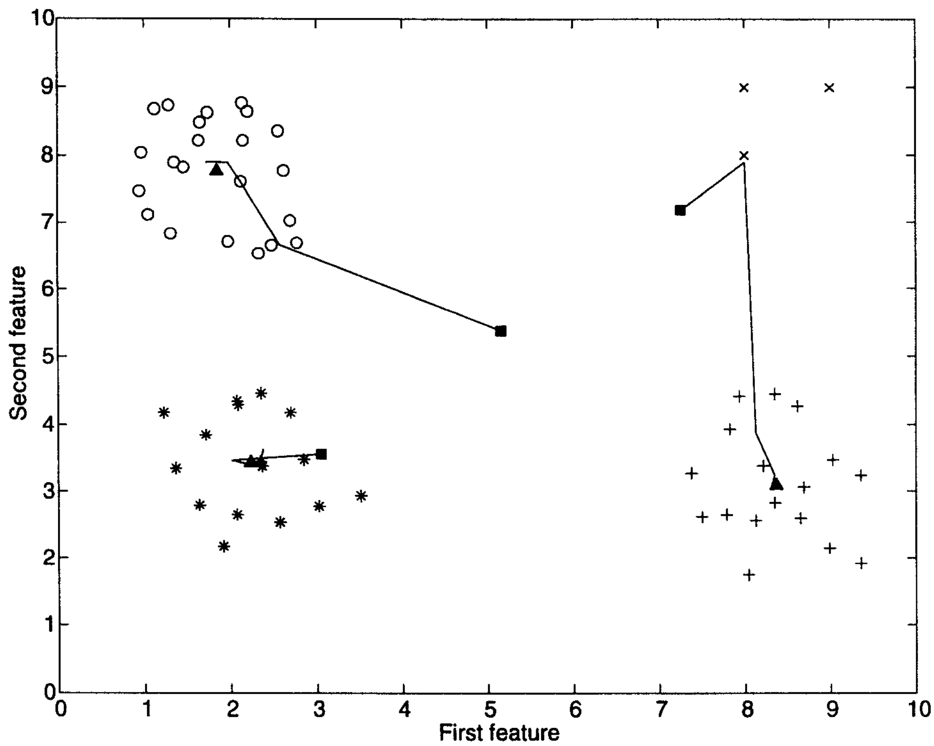

Fig. 1. Results for clustering of a synthetic data with outliers using the εFCM method.

IV. Numerical example

For all calculations in this section the k-th component of the j-th initial prototype were obtained as:

$$\underset{1 \leq j \leq c}{\forall} \quad v_{jk} = m_k + j\frac{M_k - m_k}{c+1}, \tag{29}$$

where c is a number of clusters, $m_k = \min\limits_i x_{ik}$, $M_k = \max\limits_i x_{ik}$. The computations has been stopped if in a sequential iterations the change of the minimized criterion is less than 10^{-2}. In this subsection comparison the FCM and the εFCM by calculation centroids obtained by applying these methods to synthetic two-dimensional data is presented. These data presented in Fig.1 contains three well-separated groups of data and outliers located in the upper-right corner. The true cluster centers calculated after excluding the outliers are: $\mathbf{v}_1 = [2.2379\ 3.4262]^T$, $\mathbf{v}_2 = [1.8439\ 7.7709]^T$, $\mathbf{v}_3 = [0.3627\ 3.0919]^T$. These centers are marked in Fig.1 by triangles. The FCM clustering are performed for the parameters m equal to 1.5, 2.0 and 5.0. The εFCM method is performed for the parameter m equal to 1.5, 2.0,

TABLE I

MAXIMAL ERRORS OF CLUSTER'S CENTERS CALCULATIONS.

		εFCM	
		m	
ε	1.5	2.0	2.5
	0.1047	0.1047	0.1143
0	0.2718	0.2731	0.2971
	0.2186	0.1465	0.8475
	0.0507	0.1110	0.1047
1	0.1294	0.1294	0.0000
	0.0314	0.0126	0.3403
	0.1047	0.1047	0.1047
2	0.1334	0.1792	0.1334
	0.1465	0.0126	0.0126
	0.1047	0.1047	0.1047
3	0.1334	0.1387	0.1294
	0.1928	0.1465	0.2801
		FCM	
	0.0415	0.0743	0.1793
-	0.1127	0.0843	0.2170
	0.4576	0.2372	0.0606

2.5 and the parameter ε equal to 0, 1, 2, 3.

In Table I for each combination of these parameters the maximal (by variables) absolute deviation of cluster centers from the true centers are presented. It is shown from this Table that the εFCM method terminate very nearly from the true values. The method is not very sensitive to the choice of m and ε parameters, but the best results, in the sense of a sum of absolute deviations, is obtained for $m = 2$ and $\varepsilon = 1$. The centers obtained by the εFCM method for $m = 2$ and $\varepsilon = 1$ are presented in Fig.1. From Table I it is also shown that cluster centers obtained by the FCM algorithm are far away from the true centers.

V. Conclusions

Noises and outliers in clustered data follows that these methods need to be robust. This paper establish the connection between fuzzy c-means method and robust statistics. Introduced the ε-insensitive fuzzy clustering method is based on Vapnik's ε-insensitive loss function. The new method is introduced as a constrained minimization problem of criterion function. The necessary conditions to obtain local minimum of the criterion function are shown. The existing fuzzy c-medians method can be obtained as a special case of the method proposed in this paper. Also comparative study of the ε-insensitive fuzzy c-means with traditional fuzzy c-means is included. This numerical examples showing usefulness of the method proposed in the paper in clustering data with outliers.

References

[1] J.C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms" New York, Plenum Press, 1982.

[2] R.N. Davé, "Characterization and Detection of Noise in Clustering" Pattern Recognition Lett. Vol. 12. No. 11, pp. 657-664, 1991.

[3] R.N. Davé, R. Krishnapuram, "Robust Clustering Methods: A Unified View" IEEE Trans. Fuzzy Systems, Vol. 5, No. 2, pp. 270-293, 1997.

[4] R.O. Duda, P.E. Hart, "Pattern Classification and Scene Analysis" New York, John Wiley & Sons, 1973.

[5] J.C. Dunn, "A Fuzzy Relative of the ISODATA Process and its Use in Detecting Compact Well-Separated Cluster" Journal Cybernetics Vol. 3, No.3, pp.32-57, 1973.

[6] K. Fukunaga, "Introduction to Statistical Pattern Recognition" 2nd Ed., San Diego, Academic Press, 1990.

[7] R.J. Hathaway, J.C. Bezdek, "Generalized Fuzzy c-Means Clustering Strategies Using L_p Norm Distances" IEEE Trans. Fuzzy Systems, Vol. 8, No. 5, pp. 576-582, 2000.

[8] S. Haykin, "Neural Networks. A Comprehensive Foundation" 2nd Ed., Upper Saddle River, Prentice-Hall, 1999.

[9] P.J. Huber, "Robust statistics" New York, Wiley, 1981.

[10] P.R. Kersten, "Fuzzy Order Statistics and Their Application to Fuzzy Clustering" IEEE Trans. Fuzzy Systems, Vol. 7, No. 6, pp. 708-712, 1999.

[11] R. Krishnapuram, J.M. Keller, "A Possibilistic Approach to Clustering", IEEE Trans. Fuzzy Systems, Vol. 1, No. 1, pp. 98-110, 1993.

[12] E.H. Ruspini, "A New Approach to Clustering" Inform. Control Vol. 15, No. 1 , pp.22-32, 1969.

[13] J.T. Tou and R.C. Gonzalez, "Pattern Recognition Principles" London, Adison-Wesley, 1974.

[14] V. Vapnik, "Statistical Learning Theory" New York, Wiley, 1998.

[15] L.A. Zadeh, "Fuzzy Sets" Information and Control Vol. 8, pp.338-353, 1965.