Fuzzy-Possibilistic Product Partition: A Novel Robust Approach to c-Means Clustering

László Szilágyi

Sapientia - Hungarian Science University of Transylvania,
Faculty of Technical and Human Science, Tîrgu-Mureş, Romania
lalo@ms.sapientia.ro

Abstract. One of the main challenges in the field of c-means clustering models is creating an algorithm that is both accurate and robust. In the absence of outlier data, the conventional probabilistic fuzzy c-means (FCM) algorithm, or the latest possibilistic-fuzzy mixture model (PFCM), provide highly accurate partitions. However, during the 30-year history of FCM, the researcher community of the field failed to produce an algorithm that is accurate and insensitive to outliers at the same time. This paper introduces a novel mixture clustering model built upon probabilistic and possibilistic fuzzy partitions, where the two components are connected to each other in a qualitatively different way than they were in earlier mixtures. The fuzzy-possibilistic product partition c-means (FP^{3}CM) clustering algorithm seems to fulfil the initial requirements, namely it successfully suppresses the effect of outliers situated at any finite distance and provides partitions of high quality.

Keywords: fuzzy c-means algorithm, probabilistic partition, possibilistic partition, robust clustering.

1 Introduction

Robustness in c-means clustering refers to the stability or reproducibility of the achieved partition, and insensitivity to several kinds of noise including severely outlier data. The fuzzy c-means (FCM) clustering introduced by Bezdek [3] is a very popular clustering model due to the fine partitions it makes and its easy comprehensible alternating optimization (AO) scheme. However, the probabilistic constraints involved in FCM makes it sensitive to outlier data. To combat this problem, several solution have been proposed that produce a relaxation of this probabilistic constraint.

An early solution was given by Davé [4], who introduced an extra, specially treated noisy class to attract feature vectors situated far from all normal cluster prototypes. This theory was later improved by Menard et al [7]. Alternately, Krishnapuram and Keller came up with the possibilistic c-means algorithm (PCM) [6], which distributes the partition matrix elements based on statistical rules. This approach seemed to have solved the sensitivity to outliers, but it cannot be called a robust algorithm due to the coincident clusters it frequently

V. Torra et al. (Eds.): MDAI 2011, LNAI 6820, pp. 150–161, 2011.

produces [2]. Timm et al [10] set up a repulsive force between all couples of cluster prototypes of PCM, the strength of which decreased with distance. Their method succeeded in avoiding coincident clusters, but failed to correctly treat cases when two clusters are really close to each other. Two versions of fuzzy-possibilistic partition mixtures were proposed by Pal et al [8,9], out of which the second one appears to be a reliable clustering model. Recently, Xie et al introduced a novel possibilistic c-means clustering [12] algorithm that produces a gap between fuzzy memberships with respect to winner and non-winner clusters, similarly to the symmetrical margin between classes provided by support vector machines [11] in supervised classification problems.

All the endeavors during the last three decades failed to create a clustering model that would suppress the effect of outliers similarly to gravity systems. If we add a distant object to any working gravity system, the strength of its effect will be in reversed proportion with distance. A very distant object would hardly be observable, it would hardly influence anything within the system. On the other hand, in all existing clustering models, if we increase the distance between an outlier input vector and normal input vectors, at a certain threshold distance the partitioning will fail. It would be an excellent achievement to create a clustering model that behaves similarly to gravity systems, and would totally suppress the effect of distant outliers, while keeping or even improving the accuracy in the absence of outliers. The total suppression of the outliers' effect would mean that the further the outlier stands, the less effect it has on the normal clusters.

In this paper we introduce the novel fuzzy-possibilitistic product partition c-means clustering model (FP^{3}CM), in which the degrees of membership are given as the product of a probabilistic and a possibilistic term. This new approach can eliminate all adverse effects of distant outliers, while producing high quality partitions. The algorithm uses a reduced number of parameters, making it easily adjustable to various scenarios.

The rest of this paper is structured as follows. Section 2 summarizes the background works and counter candidates of our approach. Section 3 introduces the novel FP^{3}CM clustering model. Section 4 produces a numerical analysis of the proposed and earlier methods. Conclusions are given in the last section.

2 Preliminaries

2.1 Fuzzy c-Means Clustering

The conventional FCM partitions a set of object data into a number of c clusters based on the minimization of a quadratic objective function, formulated as:

$$J_{\mathrm{FCM}} = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m} ||\boldsymbol{x}_k - \boldsymbol{v}_i||^2 = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m} d_{ik}^2 \ , \tag{1}$$

constrained by the probabilistic condition $\sum_{i=1}^{c} u_{ik} = 1 \ \forall k = 1 \dots n$ where $\boldsymbol{x}_k$ represents the input data ($k = 1 \dots n$), $\boldsymbol{v}_i$ represents the prototype or centroid value or representative element of cluster i ($i = 1 \dots c$), $u_{ik} \in [0, 1]$ is the fuzzy membership function showing the degree to which input vector $\boldsymbol{x}_k$ belongs to cluster i, $m > 1$ is the fuzzyfication parameter, and $d_{ik} = ||\boldsymbol{x}_k - \boldsymbol{v}_i||$.

The minimization of the objective function is reached by alternately applying the optimization of J_{FCM} over $\{u_{ik}\}$ with $\boldsymbol{v}_i$ fixed, and the optimization of J_{FCM} over $\{\boldsymbol{v}_i\}$ with u_{ik} fixed, [3]. During each cycle, the optimal values are computed from the zero gradient conditions, and obtained as follows:

$$u_{ik}^{\star} = \frac{d_{ik}^{-2/(m-1)}}{\sum_{j=1}^{c} d_{jk}^{-2/(m-1)}} \quad \forall\, i = 1 \ldots c, \forall\, k = 1 \ldots n \ , \tag{2}$$

$$\boldsymbol{v}_i^{\star} = \frac{\sum_{k=1}^{n} u_{ik}^{m} \boldsymbol{x}_k}{\sum_{k=1}^{n} u_{ik}^{m}} \quad \forall\, i = 1 \ldots c \ . \tag{3}$$

According to the alternating optimization scheme, Eqs. (2) and (3) are alternately applied, until cluster prototypes stabilize.

2.2 Possibilistic *c*-Means Clustering

In order to avoid the sensibility of the probabilistic partition to outlier data, Krishnapuram and Keller [6] introduced the possibilistic c-means algorithm. The elements of the possibilistic partition are denoted by t_{ik}, $i = 1 \ldots c$, $k = 1 \ldots n$. The value of t_{ik} characterizes the compatibility of data vector $\boldsymbol{x}_k$ with the cluster represented by prototype v_i.

The objective function of the PCM algorithm is

$$J_{\mathrm{PCM}} = \sum_{i=1}^{c} \sum_{k=1}^{n} \left[t_{ik}^{p} d_{ik}^{2} + (1 - t_{ik})^{p} \eta_i \right] \ . \tag{4}$$

constrained by $0 \leq t_{ik} \leq 1$ $\forall i = 1 \ldots c$, $\forall k = 1 \ldots n$, and $0 < \sum_{i=1}^{c} t_{ik} < c$ $\forall k = 1 \ldots n$, where $p > 1$ represents the possibilistic exponent, and parameters η_i are the penalty terms that control the variance of the clusters.

The iterative AO algorithm, that results from zero gradient conditions of the objective function, repeatedly applies the following formulas until convergence is reached:

$$t_{ik}^{\star} = \left[1 + \left(\frac{d_{ik}^{2}}{\eta_i} \right)^{1/(p-1)} \right]^{-1} \quad \forall\, i = 1 \ldots c, \forall\, k = 1 \ldots n \ , \tag{5}$$

$$\boldsymbol{v}_i^{\star} = \frac{\sum_{k=1}^{n} t_{ik}^{p} \boldsymbol{x}_k}{\sum_{k=1}^{n} t_{ik}^{p}} \quad \forall\, i = 1 \ldots c \ . \tag{6}$$

In the probabilistic fuzzy partition, the degrees of membership assigned to an input vector $\boldsymbol{x}_k$ with respect to cluster i depends on the distances of the given vector to all cluster prototypes: d_{1k}, d_{2k}, ..., d_{ck}. On the other hand, in the possibilistic partition, the typicality value assigned to input vector $\boldsymbol{x}_k$ with respect to any cluster i depends on only one distance: d_{ik}.

PCM efficiently suppresses the effects of outlier data, at the price of frequently producing coincident cluster prototypes. This latter is the result of the highly independent clusters [2].

2.3 Existing Fuzzy-Possibilistic Mixture Partitions

In order to avoid the pitfalls of independently computed possibilistic partitions, several solutions have been proposed. The most remarkable ones are the possibilistic-fuzzy mixture clustering models proposed by Pal et al in [8] and [9].

The so-called fuzzy-possibilistic c-means (FPCM) algorithm, introduced by Pal et al [8], minimizes the following objective function

$$J_{\mathrm{FPCM}} = \sum_{i=1}^{c}\sum_{k=1}^{n}[u_{ik}^m + t_{ik}^p]d_{ik}^2 \ , \tag{7}$$

constrained by two probabilistic conditions

$$\sum_{i=1}^{c} u_{ik} = 1 \quad \forall k = 1\dots n \quad \text{and} \quad \sum_{k=1}^{n} t_{ik} = 1 \quad \forall i = 1\dots c \ . \tag{8}$$

Using the zero gradient conditions of the above cost function, we obtain the following optimization formulas for the iterative AO scheme of the algorithm:

$$u_{ik}^\star = \frac{d_{ik}^{-2/(m-1)}}{\sum_{j=1}^{c} d_{jk}^{-2/(m-1)}} \quad \forall\, i = 1\dots c, \forall\, k = 1\dots n \ , \tag{9}$$

$$t_{ik}^\star = \frac{d_{ik}^{-2/(p-1)}}{\sum_{l=1}^{n} d_{il}^{-2/(p-1)}} \quad \forall\, i = 1\dots c, \forall\, k = 1\dots n \ , \tag{10}$$

$$\boldsymbol{v}_i^\star = \frac{\sum_{k=1}^{n}[u_{ik}^m + t_{ik}^p]\boldsymbol{x}_k}{\sum_{k=1}^{n}[u_{ik}^m + t_{ik}^p]} \quad \forall\, i = 1\dots c \ . \tag{11}$$

FPCM has the main advantage of not using the penalty terms η_i, thus making the parameter adjustment easier. However, Eq. (10) suggests that the possibilistic effect of the algorithm loses its strength as the number of input vectors grows. In case of thousands of vectors, FPCM practically reduces to FCM, regardless of the value of the exponent p.

Later, Pal et al [9] proposed another mixture clustering model, called possibilistic-fuzzy c-means (PFCM) clustering, which minimizes the objective function

$$J_{\mathrm{PFCM}} = \sum_{i=1}^{c}\sum_{k=1}^{n}[a u_{ik}^m + b t_{ik}^p]d_{ik}^2 + \sum_{i=1}^{c}\eta_i\sum_{k=1}^{n}(1-t_{ik})^p \ , \tag{12}$$

constrained by the conventional probabilistic and possibilistic conditions of FCM and PCM, respectively.

Here a and b are two tradeoff parameters that control the strength of the possibilistic and probabilistic term in the mixed partition. All other parameters are the same as in FCM and PCM.

The minimization formulas include Eq. (2) for updating the probabilistic fuzzy partition, further on

$$t_{ik}^\star = \left[1 + \left(\frac{b d_{ik}^2}{\eta_i}\right)^{1/(p-1)}\right]^{-1} \quad \forall\, i = 1\dots c, \forall\, k = 1\dots n \ , \tag{13}$$

is the update formula for typicality values, while cluster prototypes are computed as:

$$\boldsymbol{v}_i^\star = \frac{\sum_{k=1}^n [a u_{ik}^m + b t_{ik}^p] \boldsymbol{x}_k}{\sum_{k=1}^n [a u_{ik}^m + b t_{ik}^p]} \quad \forall\, i = 1 \dots c \ . \tag{14}$$

This latter algorithm was found accurate and robust, but as we will see in later sections, it is still sensitive to outlier data.

3 Methods

3.1 Intuition

In a probabilistic fuzzy partition, any outlier input vector $\boldsymbol{x}_{\text{out}}$ receives high membership values with respect to all clusters, that is, $u_{i,\text{out}} \approx 1/c$, which strongly influence all cluster prototypes.

On the other hand, in a possibilistic approach, outlier input vectors receive very low typicality values with respect to all clusters.

In our opinion, it would be a robust solution to have an objective function whose zero gradient conditions give the following cluster prototype update formula:

$$\boldsymbol{v}_i^\star = \frac{\sum_{k=1}^n \mu_{ik}^m \tau_{ik}^p \boldsymbol{x}_k}{\sum_{k=1}^n \mu_{ik}^m \tau_{ik}^p} \quad \forall\, i = 1 \dots c \ . \tag{15}$$

where $\mu_{ik}, i = 1 \dots c, k = 1 \dots n$ describe a probabilistic fuzzy partition that is not necessarily equivalent with the FCM's one, and $\tau_{ik}, i = 1 \dots c, k = 1 \dots n$ stand for the elements of a possibilistic partition matrix. We will attempt to propose such an objective function in the next subsection.

3.2 The Proposed Clustering Model

Now let us introduce the fuzzy-possibilistic product partition c-means clustering model, which minimizes

$$J_{\text{FP}^3\text{CM}} = \sum_{i=1}^{c} \sum_{k=1}^{n} \mu_{ik}^m \left[\tau_{ik}^p d_{ik}^2 + (1 - \tau_{ik})^p \eta_i \right] \ , \tag{16}$$

constrained by the conventional probabilistic condition written as $\sum_{i=1}^c \mu_{ik} = 1$ $\forall k = 1 \dots n$, and the conventional possibilistic conditions $0 \leq \tau_{ik} \leq 1$ $\forall i = 1 \dots c$, $\forall k = 1 \dots n$, and $0 < \sum_{i=1}^c \tau_{ik} < c$ $\forall k = 1 \dots n$. The only parameters of FP^{3}CM are the fuzzy exponent $m > 1$, the possibilistic exponent $p > 1$, and the conventional penalty terms of the possibilistic partition denoted by η_i, $i = 1 \dots n$.

The minimization formulas are obtained using zero gradient conditions, aided by Lagrange multipliers in case of the probabilistic term. We will compute the partial derivatives of the functional:

$$\mathcal{L} = J_{\text{FP}^3\text{CM}} + \sum_{k=1}^{n} \lambda_k \left(1 - \sum_{i=1}^{c} \mu_{ik} \right) \ , \tag{17}$$

where λ_k stands for the Lagrange multipliers. The zero crossing of the partial derivatives with respect to τ_{ik}, $\forall i = 1 \dots c$, $\forall k = 1 \dots n$, leads to:

$$\frac{\partial \mathcal{L}}{\partial \tau_{ik}} = 0 \quad \Rightarrow \quad \mu_{ik}^m \left[p\tau_{ik}^{p-1} d_{ik}^2 - \eta_i p (1 - \tau_{ik})^{p-1} \right] = 0.$$

If $\mu_{ik} = 0$, the value of τ_{ik} does not make a difference. Otherwise we get

$$\left(\frac{1 - \tau_{ik}}{\tau_{ik}} \right)^{p-1} = \frac{d_{ik}^2}{\eta_i} \quad \Rightarrow \quad \frac{1}{\tau_{ik}} - 1 = \left(\frac{d_{ik}^2}{\eta_i} \right)^{1/(p-1)} ,$$

which finally leads to a formula that is identical with Eq. (5):

$$\tau_{ik}^\star = \left[1 + \left(\frac{d_{ik}^2}{\eta_i} \right)^{1/(p-1)} \right]^{-1} \qquad \forall\, i = 1 \dots c, \forall\, k = 1 \dots n \ . \tag{18}$$

Further on, let us examine the zero crossing of partial derivatives with respect to μ_{ik}. For any $i = 1 \dots c$ and any $k = 1 \dots n$ we get

$$\frac{\partial \mathcal{L}}{\partial \mu_{ik}} = 0 \quad \Rightarrow \quad m\mu_{ik}^{m-1} \left[\tau_{ik}^p d_{ik}^2 + \eta_i (1 - \tau_{ik})^p \right] = \lambda_k,$$

which implies

$$\mu_{ik} = \left(\frac{\lambda_k}{m} \right)^{1/(m-1)} \times \left[\tau_{ik}^p d_{ik}^2 + \eta_i (1 - \tau_{ik})^p \right]^{-1/(m-1)} \quad . \tag{19}$$

The probabilistic condition says $\sum_{j=1}^c \mu_{jk} = 1$, which by the means of Eq. (19) becomes

$$1 = \left(\frac{\lambda_k}{m} \right)^{1/(m-1)} \times \sum_{j=1}^{c} \left[\tau_{jk}^p d_{jk}^2 + \eta_j (1 - \tau_{jk})^p \right]^{-1/(m-1)} \quad . \tag{20}$$

Dividing Eq. (19) by Eq. (20) term by term, leads to

$$\mu_{ik}^\star = \frac{[\tau_{ik}^p d_{ik}^2 + \eta_i (1 - \tau_{ik})^p]^{-1/(m-1)}}{\sum_{j=1}^c [\tau_{jk}^p d_{jk}^2 + \eta_j (1 - \tau_{jk})^p]^{-1/(m-1)}} \ , \tag{21}$$

which holds for any $i = 1 \dots c$, and any $k = 1 \dots n$. Finally, let us investigate the zero crossings of the partial derivatives with respect to $\boldsymbol{v}_i$, $i = 1 \dots n$:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{v}_i} = 0 \quad \Rightarrow \quad -2 \sum_{k=1}^{n} \mu_{ik}^m \tau_{ik}^p (\boldsymbol{x}_k - \boldsymbol{v}_i) = 0 \ ,$$

which implies

$$\boldsymbol{v}_i \sum_{k=1}^{n} \mu_{ik}^m \tau_{ik}^p = \sum_{k=1}^{n} \mu_{ik}^m \tau_{ik}^p \boldsymbol{x}_k \quad \Rightarrow \quad \boldsymbol{v}_i^\star = \frac{\sum_{k=1}^n \mu_{ik}^m \tau_{ik}^p \boldsymbol{x}_k}{\sum_{k=1}^n \mu_{ik}^m \tau_{ik}^p} \ , \tag{22}$$

valid for any $i = 1 \dots c$, exactly as we wished in Eq. (15). Let us remark the followings:

1. The possibilistic memberships τ_{ik} are established exactly the same way, as in the PCM algorithm. This is why the penalty terms η_i can be set as recommended by Krishnapuram and Keller [6].
2. The probabilistic partition is somewhat similar to FCM's partition, but distances are distorted, and the partition is influenced by the possibilitic penalty terms η_i.
3. Outlier input vectors $\boldsymbol{x}_k$ are indicated by the algorithm with a low value of $\max\{\sqrt[m+p]{\mu_{ik}^m \tau_{ik}^p}, i = 1 \dots c\}$.
4. The defuzzyfication of the final partition should be performed according to the following rule: $\boldsymbol{x}_k$ is assigned to cluster with index w_k, where

$$w_k = \arg\max_j \left(\mu_{jk}^m \tau_{jk}^p | j = 1 \dots c \right) \ . \tag{23}$$

In case of equal η_i values, for any $i = 1 \dots c$, the rule becomes more simple: $w_k = \arg\max_j (\mu_{jk} | j = 1 \dots c)$.

3.3 The Alternative Optimization Algorithm of FP^{3}CM

Let us summarize the optimization algorithm of the proposed clustering model:

1. Set fuzzy exponent m and possibilistic exponent p, both greater than 1.
2. Set possibilistic penalty terms η_i, $i = 1 \dots c$, as recommended by Krishnapuram and Keller in [6].
3. Update possibilistic membership values using Eq. (18).
4. Update probabilistic membership values using Eq. (21).
5. Update cluster prototypes using Eq. (22).
6. Repeat steps 3-5 until cluster prototypes converge.
7. If it is necessary for the application, perform defuzzyfication of the obtained product partition as indicated in Eq. (23).

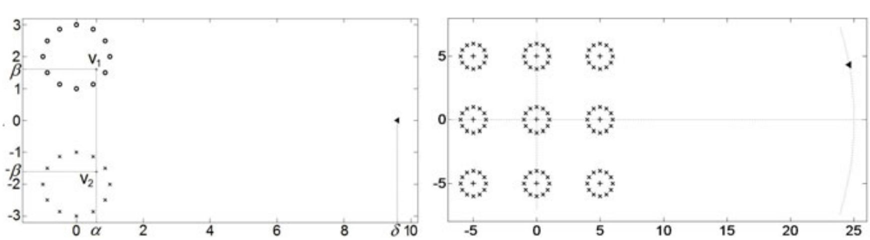

Fig. 1. Two scenarios for the numerical test of robustness: (left) two clusters and an outlier, (right) nine clusters and an outlier. We investigate the positions of cluster prototypes and the resulting partition accuracy, versus the outlier's position.

4 Results and Discussion

In the followings, we will perform some numerical tests to evaluate the robustness and accuracy of the proposed algorithm. We will compare its performances with counter candidates like FCM, FPCM, and PFCM. The pure possibilistic PCM algorithm is excluded from these tests due to its frequently coincident cluster prototypes.

4.1 Two Clusters and One Outlier Input Vector

Let us consider two sets of ν data points each, uniformly distributed along unit-radius circles: $\boldsymbol{x}_k = (\cos\frac{2k\pi}{\nu}, 2 + \sin\frac{2k\pi}{\nu})^T$ and $\boldsymbol{x}_{\nu+k} = (\cos\frac{2k\pi}{\nu}, -2 + \sin\frac{2k\pi}{\nu})^T$, $\forall k = 1 \ldots \nu$.

The input data set also includes an outlier, situated at $\boldsymbol{x}_{2\nu+1} = (\delta, 0)^T$. We will attempt to classify these $n = 2\nu + 1$ vectors into $c = 2$ clusters, setting the initial cluster prototypes in the middle of the two circles: $\boldsymbol{v}_1 = (0, 2)^T$ and $\boldsymbol{v}_2 = (0, -2)^T$.

During the iterative optimization of all tested algorithms, the cluster prototypes will be attracted by the outlier vector. As long as the outlier cannot tear off any of the two prototypes, $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ will behave symmetrically, having their coordinates $\boldsymbol{v}_1 = (\alpha, \beta)^T$ and $\boldsymbol{v}_2 = (\alpha, -\beta)^T$. A graphical representation of the problem is shown in Fig. 1(left).

The question is, how α and β will depend on the outlier's position δ in case of all tested algorithms, and how far the outlier vector can go without tearing off one of the cluster prototypes.

Figure 2 presents the outcome of numerical simulations performed on all mentioned algorithms in various circumstances. The α coordinate of the symmetrical cluster prototypes is shown in two different plots in Fig. 2(b) and (c). In case of all existing algorithms, the further the outlier goes, the stronger it attracts the centroids, and at a certain boundary, one of the prototypes is torn out by the outlier.

On the other hand, FP^{3}CM behaves like a gravity system: the further the outlier is situated, the weaker its effect is upon the cluster centroids. No matter how far the outlier is, the obtained partition is correct. The outlier receives such a low membership value to both clusters that it can be easily assigned to the noisy class at defuzzyfication. Figure 2(d) shows the behavior of FP^{3}CM in case of various values of possibilistic exponent p, at a constant value of fuzzy exponent $m = 2$. The plots reveal that stronger possibilistic component or lower values of p lead to more efficient rejection of the outlier effect. However, when the outlier is not too far, lower exponent values also cause stronger deviation of the cluster centroids.

4.2 Accuracy Test with Nine Regular Clusters and an Outlier

As it is shown in Fig. 1(right), the input data in this second test consists of 9 sets of vectors uniformly distributed along unit radius circles, situated in the neighborhood of the origin. Initially, the cluster prototypes are placed in the middle

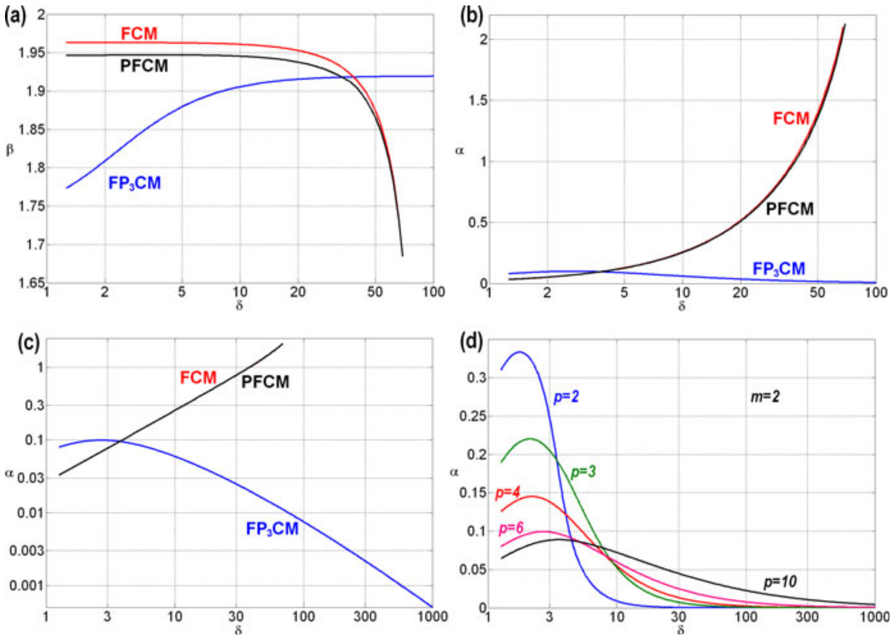

Fig. 2. (a)-(c) Position of the two symmetrical cluster prototypes at $m = 2$ and $p = 6$: (a) the β coordinate plotted against the position of the outlier δ, (b) the α coordinate plotted against the position of the outlier δ, (c) logarithmic plot of α coordinate against the distance of the outlier. Some of these graphs end at the threshold value of δ where the algorithms fail. In case of FP^{3}CM, the further the outlier wanders, the less influence it has upon cluster prototypes; (d) The α coordinate produced by the proposed algorithm FP^{3}CM, at $m = 2$ and various values of p. The algorithm manages to suppress the effect of departed outliers.

Table 1. The limit distance δ, in case of various algorithms and circumstances, where the tested algorithm fails to produce nine accurate clusters

Algo-rithm	Circumstances m	p	$\sqrt{\eta_i}$	a	b	Limit distance	Algo-rithm	Circumstances m	p	$\sqrt{\eta_i}$	a	b	Limit distance
FCM	2					361	PFCM	2	3	1.0	1	5	437
FPCM	2	5				361	PFCM	2	3	1.5	1	5	521
FPCM	2	2				367	PFCM	2	3	2.0	1	5	593
FPCM	2	1.2				401	PFCM	2	3	2.5	1	5	546
PFCM	2	2	1.0	2	3	410	PFCM	2	2	1.0	1	5	459
PFCM	2	2	1.5	2	3	479	PFCM	2	2	1.5	1	5	602
PFCM	2	2	2.0	2	3	563	PFCM	2	2	2.0	1	5	789
PFCM	2	2	2.5	2	3	649	PFCM	2	2	2.5	1	5	1001
PFCM	2	5	1.0	1	5	394	PFCM	2	2	3.0	1	5	1220
PFCM	2	5	1.5	1	5	421	PFCM	2	2	4.0	1	5	1354
PFCM	2	5	2.0	1	5	428	PFCM	2	2	5.0	1	5	1089
PFCM	2	5	2.5	1	5	370	FP^{3}CM	wide range					$+\infty$

of the nine circles. The single outlier vector moves along the big circle of radius δ, with its center in the origin. The aim of this study is to establish, which is the boundary value for δ where tested algorithms crash in various circumstances.

The obtained boundary distances are summarized in Table 1. These values emphasize the fact that currently existing algorithms may have enhanced the robustness of FCM, they may have enabled the outlier to fall somewhat further (no more than by one order of magnitude) without making the clustering crash. The novel clustering model FP^{3}CM seems to efficiently suppress the influence of the outlier vector, leading to accurate partitions for any limited value of δ.

4.3 Numerical Tests Using IRIS Data

In the followings, we will analyze the accuracy and robustness of the investigated clustering models using the IRIS data set [1], which consist of 150 labeled feature vectors of four dimensions (sepal length and width, petal length and width), organized in three clusters ("setosa","versicolor", and "virginica") of fifty vectors each. It is a reported facts, that conventional clustering models like FCM produce 133-134 correct decisions when classifying IRIS data. PFCM produced the best reported accuracy with 140 correct decisions using $a = b = 1$, $m = p = 3$, and initializing $\boldsymbol{v}_i$ with terminal FCM prototypes [9]. Under less advantageous circumstances, PFCM reportedly produced 136-137 correct decisions.

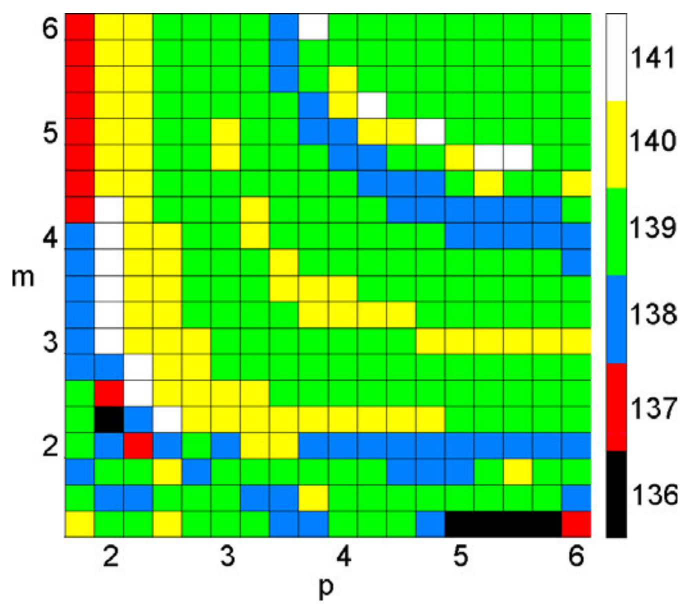

Fig. 3. Number of correct decisions (out of 150) obtained by FP^{3}CM, plotted against fuzzy exponent m and possibilistic exponent p, using $\sqrt{\eta_i} = 0.7$

Table 2. Detailed values of the final IRIS cluster prototypes: a high quality partition

Correct decisions	Centroid vector	Sepal length	Sepal width	Petal length	Petal width
141	$\boldsymbol{v}_1$	5.0443	3.4307	1.4641	0.2337
	$\boldsymbol{v}_2$	6.0729	2.9104	4.5693	1.4485
	$\boldsymbol{v}_3$	6.4794	2.9876	5.2934	1.9687

Table 3. Partition accuracies and confusion matrices in various scenarios

Circumstances	IRIS type	FCM $\boldsymbol{v}_1$	FCM $\boldsymbol{v}_2$	FCM $\boldsymbol{v}_3$	PFCM $\boldsymbol{v}_1$	PFCM $\boldsymbol{v}_2$	PFCM $\boldsymbol{v}_3$	FP^{3}CM $\boldsymbol{v}_1$	FP^{3}CM $\boldsymbol{v}_2$	FP^{3}CM $\boldsymbol{v}_3$	Correct decisions
no outlier added	Setosa	50	0	0	50	0	0	50	0	0	FCM → 136
	Versicolor	0	47	3	0	47	3	0	48	2	PFCM → 136
	Virginica	0	11	39	0	11	39	0	7	43	FP^{3}CM → 141
outlier added at 20	Setosa	50	0	0	50	0	0	50	0	0	FCM → 134
	Versicolor	0	50	0	0	50	0	0	47	3	PFCM → 135
	Virginica	0	16	34	0	15	35	0	7	43	FP^{3}CM → 140
outlier added at 30	Setosa	50	0	0	50	0	0	50	0	0	FCM → 128
	Versicolor	1	49	0	1	49	0	0	47	3	PFCM → 131
	Virginica	0	21	29	0	18	32	0	7	43	FP^{3}CM → 140
outlier added at 50 or 10^6	Setosa	50	0	0	50	0	0	50	0	0	FCM crashes
	Versicolor	3	47	0	3	47	0	0	47	3	PFCM crashes
	Virginica	0	50	0	0	50	0	0	7	43	FP^{3}CM → 140

We have tested the proposed FP^{3}CM clustering model in a wide range of both the fuzzy and the possibilistic exponents. The resulting partition quality is summarized in Fig. 3. The best partition achieved by FP^{3}CM had 141 correct decisions, which is above any reported result. Details upon the final cluster prototypes are given in Table 2. We also need to remark, that almost any parameter setting leads to good partition quality. To make sure FP^{3}CM clusters accurately, the possibilistic term should not be too strong, it is recommendable to keep parameter $p \geq 2$.

A series of numerical tests using the IRIS data targeted the clustering robustness. We artificially inserted an outlier vector into the input data set, with coordinates $\boldsymbol{x}_{151} = (\delta, \delta, \delta, \delta)^T$, and proceeded all vectors to clustering into $c = 3$ groups. Table 3 gives us an overview upon accuracy, confusion matrices, and sensibility to the outlier's position. As we can see it in the table, most existing clustering models failed somewhere between $\delta = 30$ and $\delta = 50$, while the proposed algorithm led to high quality partition even at $\delta = 10^6$, being less affected by distant outliers. All these tests were performed at $m = 2.0$, $p = 3.5$, $\sqrt{\eta_i} = 0.7$ $\forall i = 1 \ldots c$, $a = 1$, and $b = 5$.

5 Conclusions

In this paper we proposed a novel fuzzy-possibilistic mixture clustering model, in order to combat the sensitivity of existing c-means clustering models to outlier data. We performed several numerical tests on artificially created test data and the very popular IRIS data set, to evaluate the behavior of the proposed FP^{3}CM clustering model. In the presence of distant outliers, the proposed clustering model outperforms all existing c-means approaches. Further on, even in the absence of outliers, FP^{3}CM is slightly more accurate than PFCM, and outperforms conventional approaches in partition quality.

The adaptation of the proposed methodology to detect clusters of certain predefined shapes is going to be straightforward task, along the guidelines established by Davé and Bhaswan [5].

Acknowledgment. This research was funded by CNCSIS UEFISCSU, project no. PD_667, under contract no. 28/05.08.2010.

References

1. Anderson, E.: The IRISes of the Gaspe peninsula. Bull. Amer. IRIS Soc. 59, 2–5 (1935)
2. Barni, M., Capellini, V., Mecocci, A.: Comments on a possibilistic approach to clustering. IEEE Trans. Fuzzy Syst. 4, 393–396 (1996)
3. Bezdek, J.C.: Pattern recognition with fuzzy objective function algorithms. Plenum, New York (1981)
4. Davé, R.N.: Characterization and detection of noise in clustering. Patt. Recogn. Lett. 12, 657–664 (1991)
5. Davé, R.N., Bhaswan, K.: Adaptive fuzzy c-shells clustering and detection of ellipses. IEEE Trans. Neural Netw. 3(5), 643–662 (1992)
6. Krishnapuram, R., Keller, J.M.: A possibilistic approach to clustering. IEEE Trans. Fuzzy Syst. 1, 98–110 (1993)
7. Menard, M., Damko, C., Loonis, P.: The fuzzy $c + 2$ means: solving the ambiguity rejection in clustering. Patt. Recogn. 33, 1219–1237 (2000)
8. Pal, N.R., Pal, K., Bezdek, J.C.: A mixed c-means clustering model. In: Proc. IEEE Int'l Conf. Fuzzy Systems (FUZZ-IEEE), pp. 11–21 (1997)
9. Pal, N.R., Pal, K., Keller, J.M., Bezdek, J.C.: A possibilistic fuzzy c-means clustering algorithm. IEEE Trans. Fuzzy Syst. 13, 517–530 (2005)
10. Timm, H., Borgelt, C., Döring, C., Kruse, R.: An extension to possibilistic fuzzy cluster analysis. Fuzzy Sets and Systems 147, 3–16 (2004)
11. Vapnik, V.: Statistical learning theory. Wiley, New York (1998)
12. Xie, Z., Wang, S., Chung, F.L.: An enhanced possibilistic c-means clustering algorithm. Soft. Comput. 12, 593–611 (2008)