



Selecting Parameters of the Fuzzy Possibilistic Clustering Algorithm

Mohamed Fadhel Saad
REsearch Groups on Intelligent Machines,
University of Sfax, National Engineering
School of Sfax (ENIS), BP 1173, Sfax, 3038,
Tunisia

Adel M. Alimi
REsearch Groups on Intelligent Machines,
University of Sfax, National Engineering
School of Sfax (ENIS), BP 1173, Sfax, 3038,
Tunisia

ABSTRACT

Clustering has been widely used in pattern recognition, image processing, and data analysis. It aims to organize a collection of data items into clusters, such that items within a cluster are more similar to each other than they are in other clusters. The Fuzzy Possibilistic C-Means (FPCM) is one of the most popular clustering methods based on minimization of a criterion function. So the implementation of this algorithm requires a priori selection of some parameters: the fuzzy and the typical exponent, initialization of cluster centers. But the definition of these parameters at the moment is fixed in advanced and the initialization of centers is random; so the algorithm can give results not consistent. The determination of an optimal value for these parameters and the cluster centers at the beginning are problematic and remains an open problem. New procedures for choice of the optimal values of parameters and for initialization of centers were developed. Numerical results using data sets are used to illustrate the simplicity and effectiveness of the proposed procedures.

General Terms

Pattern Recognition, Clustering, Algorithms et. al.

Keywords

Fuzzy c-means, Possibilistic c-means, Fuzzy possibilistic c-means, K-means++.

1. INTRODUCTION

Cluster analysis is a technique used for classifying data, i.e. for dividing a given data set into a set of classes or clusters. The goal is to divide the data set in such a way that two cases from the same cluster are as similar as possible and two cases from different clusters are as dissimilar as possible [1, 2]. The conventional clustering methods consider each point of the data set to match exactly one cluster. Since 1965, Zadeh has proposed fuzzy sets in order to come closer to the physical world [3]. Zadeh introduced the idea of partial memberships described by membership functions. Fuzzy sets could allow membership functions to all clusters in a data set so it was very suitable for cluster analysis. In 1973, Bezdek submitted a doctoral thesis on fuzzy math for classification [4] which is necessary conditions for the minimization of the general criterion that defines the family of clustering algorithms known as the fuzzy c-means (FCM) [5]. The FCM used the probabilistic constraint that the memberships of a data point across classes summed to one. While this was useful in creating partitions, the memberships resulting from FCM and its derivatives, however, don't always generally correspond to the intuitive concept of degree of belongingness or compatibility. Moreover, the FCM was sensitive to noise. To

mitigate such an effect, Krishnapuram and Keller (1993) throw away the constraint of memberships in FCM and proposed the possibilistic c-means (PCM) algorithm [6]. The advantages of PCM were that it overcomes the need to specify the number of clusters and were highly robust in a noisy environment. However, a few weaknesses remained in the PCM, i.e. it highly depended on a good initialization and had the undesirable tendency to produce coincident clusters. Pal et al. deduced that to classify a data point, cluster centroid had to be very close to the data point; it is the role of membership. Moreover, for estimating the centroids, the typicality was used for alleviating the undesirable effect of outliers. So Pal defined a clustering algorithm called fuzzy possibilistic c-means (FPCM) that combines the characteristics of both fuzzy and possibilistic c-means [7]. Although FPCM is much less prone to the problems of both FCM and PCM just described, the possibility values are very small when the size of the data set increases.

One can regard the different variants of fuzzy clustering algorithms as an iterative process involving successive computations of the prototypes and the partition matrix. The values of the parameters are set up in advance. They consist of the following items: the number of clusters c , the distance function, the fuzzification factor or the fuzzy exponent (m), the typical exponent (η). The algorithms can give erroneous results, due to the choice of parameters. Also different algorithms use a random initialization, which can deviate results. To eliminate this, one will define an algorithm initialization for them.

The remainder of this paper is organized as follows. In section 2, fuzzy clustering methods are presented; some drawbacks of them are also mentioned. In Section 3, the initialization of cluster centers is discussed. The problem of fuzzy and the typical exponents is presented; new procedure for choice an optimal fuzzy and typical exponents is mentioned in Section 4. The proposed approach named Improved Fuzzy Possibilistic C-Means (IFPCM) can solve the mentioned drawbacks, and obtain better quality clustering results. In Section 6, we present several examples to assess the performance of proposed approach. The comparisons are made between FCM, FPCM and IFPCM. Finally, conclusions are made in Section 7.

2. FUZZY CLUSTERING METHODS

2.1 Fuzzy c-means clustering

The fuzzy c-means (FCM) can be seen as the fuzzified version of the k-means algorithm [8]. Ruspini first proposed fuzzy c-partitions as a fuzzy approach to clustering. Later, the fuzzy c-means (FCM) algorithm with a weighting exponent $m = 2$

proposed by Dunn, and then generalized by Bezdek with $m > 1$ became popular [9]. The FCM used the probabilistic constraint that the memberships of a data point across classes summed to one. While this was useful in creating partitions, the memberships resulting from FCM and its derivatives, however, do not always generally correspond to the intuitive concept of degree of belongingness or compatibility. The FCM algorithm is an iterative algorithm clustering that produces an optimal c partitions by minimizing the weights within group sum of squared error objective function J_{FCM} :

$$J_{FCM}(U, V) = \sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^m d^2(x_j, v_i) \quad (1)$$

With $1 < m < +\infty$

Where $X = \{x_1, x_2, \dots, x_n\}$ is the set of n unlabeled feature vectors in R^p , p is the number of data items, c is the number of clusters with $2 \leq c \leq n - 1$. $V = \{v_1, v_2, \dots, v_c\} \subset R^{c \times p}$ is the c centers or prototypes of the clusters, v_i is the p -dimension center of the cluster i , and $d^2(x_j, v_i)$ is a distance measure between an object x_j and a cluster centre v_i . $U = [\mu_{ij}]$ represents a fuzzy partition matrix with $u_{ij} = u_i(x_j)$ is the degree of membership of x_j in the i th cluster; x_j is the j th of p -dimensional measured data. With the following constraints (2) and (3):

$$\sum_{i=1}^c \mu_{ij} = 1, \forall j \in \{1, \dots, n\} \quad (2)$$

$$\sum_{j=1}^n \mu_{ij} \leq n, \forall i \in \{1, \dots, c\} \quad (3)$$

The parameter m is a fuzzy exponent, on each fuzzy membership and determines the amount of fuzziness of the resulting classification; it is a fixed number greater than one.

The stationary points of the objective function J_{FCM} can be found by adjoining the constraint (2). We minimized under the constraint of U . Specifically, taking of J_{FCM} with respect to μ_{ij} and v_i and zeroing then respectively, two necessary but not sufficient conditions for J_{FCM} to be at its local extrema will be as the following:

$$\mu_{ij} = \left[\sum_{k=1}^c \left(\frac{d^2(x_j, v_i)}{d^2(x_j, v_k)} \right)^{\frac{2}{m-1}} \right]^{-1}, \quad 1 \leq i \leq c, 1 \leq j \leq n \quad (4)$$

$$v_i = \frac{\sum_{k=1}^n \mu_{ik}^m x_k}{\sum_{k=1}^n \mu_{ik}^m}, \quad 1 \leq i \leq c \quad (5)$$

Although FCM is a very useful clustering method, its memberships do not always correspond well to the degree of belonging of the data, and may be inaccurate in a noisy environment, because the real data unavoidably involves some noises.

2.2 Possibilistic c-means clustering

Krishnapuram and Keller proposed a possibilistic approach of c -means called possibilistic c -means, or PCM [6]. Their approach is expected to lead to better performance in the presence of noise. This method permits to obtain clusters that correspond more closely to the intuitive concept of typicality

or compatibility, and this by the discovery of the fuzzy partitions that do not satisfy the fuzzy constraint. Krishnapuram and Keller relaxed the constrained condition (2) of the fuzzy c -partition to obtain a possibilistic type of membership function and proposed PCM for unsupervised clustering. The component generated by the PCM corresponds to a dense region in the data set; each cluster is independent to other clusters. The objective function of the PCM can be formulated as follows:

$$J_{PCM}(U, V, X) = \sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^m d^2(x_j, v_i) + \sum_{i=1}^c \eta_i \sum_{j=1}^n (1 - \mu_{ij})^m \quad (6)$$

$$\text{Where } \eta_i = \frac{\sum_{j=1}^n \mu_{ij}^m d^2(x_j, v_i)}{\sum_{j=1}^n \mu_{ij}^m} \quad (7)$$

η_i is the vector of parameters that determine the "zone of influence" of the clusters, it is a scale parameter at the i th cluster. The parameter is evaluated for each cluster separately; it determines the distance at which the membership degree equals 0.5. Using (7), η_i is proportional to the average fuzzy intracluster distance of cluster v_i .

The new membership update equation is:

$$\mu_{ij} = \frac{1}{1 - \left(\frac{d^2(x_j, v_i)}{\eta_i} \right)^{\frac{1}{m-1}}} \quad (8)$$

is the possibilistic typicality value of training sample x_j belonging to the cluster i . $m \in [1, +\infty]$ is a weighting factor called the possibilistic parameter or typical exponent.

So PCM algorithm, as advocated drops the requirement that the sum of membership grades must be equal to 1, and the results are acceptable in the case of noisy data. But the second algorithm allows data points to behave almost independently of the other data in X ; also it is very sensitive to initialization, and it has coincident clusters, because the columns and rows of the typicality matrix are independent from the other clusters.

2.3 Fuzzy possibilistic c-means clustering

The fuzzy possibilistic c -means (FPCM) can be seen as the typical version of the fuzzy c -means (FCM) algorithm. It is a clustering method which allows one piece of a data to belong to two or more clusters. This algorithm is an iterative clustering technique that produces optimal c partitions by minimizing an error objective function:

$$J_{FPCM}(U, T, V) = \sum_{i=1}^c \sum_{j=1}^n (\mu_{ij}^m + t_{ij}^n) d^2(x_j, v_i) \quad (9)$$

With $1 < m < +\infty$ and $1 < n < +\infty$

$U = [\mu_{ij}]$ represents a fuzzy partition matrix with $\mu_{ij} = \mu_i(x_j)$ is the degree of membership of x_j in the i th cluster; x_j is the j th of p -dimensional measured data. $T = [t_{ij}]$ represents a typical partition matrix with $t_{ij} = t_i(x_j)$ is the

degree of typicalities of x_j in the i th cluster alone. With the following constraints (17) and (18):

$$\sum_{i=1}^c \mu_{ij} = 1, \forall j \in \{1, \dots, n\} \quad (10)$$

$$\sum_{j=1}^n t_{ij} = 1, \forall i \in \{1, \dots, c\} \quad (11)$$

FPCM generates memberships and possibilities at the same time, together with the usual point prototypes or cluster centers for each cluster.

A solution of the objective function can be obtained via an iterative process where the degrees of membership, typicality and the cluster centers are update via:

$$\mu_{ij} = \left[\sum_{k=1}^c \left(\frac{d^2(x_j, v_i)}{d^2(x_j, v_k)} \right)^{\frac{2}{m-1}} \right]^{-1}, 1 \leq i \leq c, 1 \leq j \leq n \quad (12)$$

$$t_{ij} = \left[\sum_{k=1}^n \left(\frac{d^2(x_j, v_i)}{d^2(x_j, v_k)} \right)^{\frac{2}{\eta-1}} \right]^{-1}, 1 \leq i \leq c, 1 \leq j \leq n \quad (13)$$

$$v_i = \frac{\sum_{k=1}^n (\mu_{ik}^m + t_{ik}^\eta) x_k}{\sum_{k=1}^n (\mu_{ik}^m + t_{ik}^\eta)}, 1 \leq i \leq c \quad (14)$$

The above equations show that membership μ_{ij} is the same as FCM, but resulted values may be different and it is affected by all c cluster centers, while possibility t_{ij} is affected only by the i -th cluster center v_i .

The principle of the algorithm is as follows:

S1:

- Given the data set $X = \{x_1, x_2, \dots, x_n\} \subseteq \mathbb{R}^p$.
- Choose the number of clusters $1 < c < n$.
- Choose the fuzzy exponent $m > 1$.
- Choose the typical exponent $\eta > 1$.
- Choose the iteration limit T_{max} .
- Choose the termination tolerance $\varepsilon > 0$.
- Pick Norm $\|x\|_A$.
- Initialize the cluster centers $V(0)$ randomly.

For $t = 1, 2, \dots, T_{max}$

S2: Compute $U_t = \{\mu_{ij}\}$, using the formula (12).

S3: Compute $T_t = \{t_{ij}\}$, using the formula (13).

S4: Update the centers vectors $V_t = [v_i]$, using the formula (14).

S4: Compute $E_t = \|V^t - V^{t-1}\|$, if $(E_t \leq \varepsilon$ or $t \geq T_{max})$, Stop; Else $t = t + 1$.

End.

2.4 Remarks

The clustering reduces to consolidate the examples of the most natural way possible. This desire to classify naturally is of course ambiguous and often formalized by the objective to define groups of objects such as the distance between objects in the same class, which is minimized and the distance between objects of different classes which is maximized. This vision of clustering therefore is compelled to have a defined distance to the descriptive language of objects. It is in the case where the feature space objects is a digital vector space in which each dimension corresponds to a separate attribute. Each object is described by a vector of attributes with real values. The c -means algorithm and its variants are frequently used in the literature of clustering, this algorithm performs the grouping by minimizing intragroup variance. The use of c -means in several areas, does not neglect the fact that it has several disadvantages: The enormous execution time due to its iterative nature. The fuzzification and typical parameters are fixed in advance m and η ; who possess in most cases the value 2. This value has no mathematical or empirical interpretation. The optimal number of classes c giving the best description of data structure is defined.

3. INITIALIZATION OF CLUSTER CENTERS

3.1 Background

The problem of classification by FPCM or other variant is expressed as a minimization of a functional, under certain constraints. The underlying algorithms don't ensure the optimality of the solution, because a local minimum can be found that stops the iterations. The initialization step conditioning the research of a minimum is therefore fundamental. Many strategies have been proposed in the literature, the simplest is to ask an expert to determine interest regions representing the centers of classes; the algorithm converges to an acceptable solution, but the unsupervised aspect is lost. Pena et al. [10] have compared four methods to initialize the partition matrix U . These four methods are: method completely random, the Forgy approach [8], MacQueen approach and Kaufman approach [11]. Pena et al. [10] assert that the best initialization scheme in terms of robustness and quality of partition is the one proposed by Kaufman. These propose to select iteratively the cluster centers until c vectors are chosen. The first vector selected is the closest to the gravity center of the set of vectors. The next center is selected by the heuristic rule of choosing a promised element to have around him a maximum of vectors not yet selected. Using an unsupervised initialization has however the disadvantage of not necessarily converge to a correct solution. Thus, Velthuizen et al. [12] bring disappointing recognition rate results of pathological brain tissue (tumor) by FCM. The result depends in large part on the initialization of the algorithm chosen. According to the work of Christian Borgelt [1] in his thesis, the initialization of cluster centers for better classification is done in three ways. First, initialization methods independent of the data: These are statistical methods that make random initialization of the points from a probability distribution. The choice of the probability distribution depends on the information data. However, the uniform distribution is the most used. The advantage of its methods is its simplicity because it is unnecessary to perform calculations by the computer, but its disadvantage is that the initial positions of the points may be over far or noise from the data. After, initialization methods dependent data

(simple): the first choice center is done directly by selecting data points, or indirectly by the use of average data points. Then the other centers were selected based on the first center, the method most used in this case is Maxmindist (MAXimize MINImal DISTance). Finally, one has sophisticated initialization methods; these initialization methods are methods of classification, so he is talking about mixed classification that is to say a classifier to initialize and then applying our approach. So to initialize FCM, one uses the results of HCM (Hard C-Means); also initialize PCM, FPCM, FPCM; one uses the results of FCM. David Arthur and Sergei Vassilvskii [13] proposed an approximation algorithm named k-means++ that select the initial values for the K-means algorithm. k-means++ algorithm is the following:

1. Choose randomly a center class from points of data set.
2. For each point x , calculate the distance between x and the nearest center.
3. Choose a new center from data set $v_i = \bar{x} \in X$, by computing a weighted probability

$$\frac{D(\bar{x})^2}{\sum_{x \in X} D(x)^2} \quad (15)$$

4. Repeat steps 2 and 3, until k centers are chosen.

3.2 Improved K-means++

The most recent and efficient algorithm for initialization is k-means++. This algorithm solved the problem of random initialization of cluster centers. But step 1 of the algorithm is a choice of the first center, which is done by random manner, or this choice can be a noise or overlap point in the data set, which can give incorrect results. Yet our objective in this thesis is to eliminate all random treatment. We will present a modified version of k-means++ by using an efficient method for selecting the first center. This method is based on the density of points. A density for each point is calculated as follows:

$$PD(i) = \sum_{j=1}^n \exp(-\|x_i - x_j\|^2) \quad (16)$$

After, one chooses the point which has the largest value of PD (point density) as the center of class. So we have the following function:

```
Function initialize_center(X: Data set)→v1
Begin
  For i=1 to n Do
    Compute PD(i) by Eq(16)
  v1← indexmax(PD)
End.
```

The improved K-Means++ algorithm is the following:

1. $v1 \leftarrow \text{initialize_center}(X)$
2. For each point x , calculate the distance between x and the nearest center
3. Choose a new center from data set, by computing a

weighted probability

4. Repeat steps 2 and 3, until k centers are chosen.

3.3 Some results

The proposed initialization algorithm gave good results for synthetic base that they are presented from the examples studied.

3.1.1 Example 1

In the first experiment, one uses a three-cluster data set without an outlier shown in Figure 1. To demonstrate the quality of classification of FPCM in relation to the random initialization. The clustering results of the algorithm are shown in Figure. 1(a)~1(b) respectively, where the points from the clustering algorithm are with symbols “o” and clusters centers are with symbols “*”.

The image in question is composed of three well separated classes, the algorithm with or without using K-means++ succeeded in finding the correct centers.

It can be concluded that the random initialization does not have a major impact if the points are well separated. But another problem that can be posed is the exact number of classes.

3.1.2 Example 2

In the second experiment, one uses a two-cluster data set with outlier shown in Figure.2.

The image is composed of two symmetrical point's clouds which are added noise point (top image). The algorithm with K-means++ seems almost not influenced by the intruder and therefore gives almost correct centers and the other case gives remote centers of true clusters.

3.1.3 Example 3

The image is composed of three classes, two of them are overlapped which gives the impression that it takes two classes instead of three (Figure 3). The figure shows that the algorithm isn't influenced by the random initialization.

4. FUZZY AND TYPICAL PARAMETERS

4.1 Background

The parameter m is the exponent for each fuzzy membership, which determines the degree of fuzziness, $1 < m < 1$. Bezdek [5] suggested that for the FCM clustering algorithm, m should be in the range 1 to 30, with the ranger 1.5 to 3 giving good results. The authors noted no strong theoretical or empirical evidence for these choices. They also gave an interesting interpretation of the special case where $m = 2$. Cannon et al [15] also noted no theoretical basis for choosing a good value suggesting $1.1 \leq m \leq 5$ as the most useful range. McBratney and Moore [16] investigated the choice of m for the FCM algorithm. They observed that the objective function value, J_{FCM} , decreases monotonically with increasing number of groups c and increasing values of m , and that its rate of change with changing m is not constant; and they found a value of m approximately 2 optimal but dependent on the number of classes. Choe and Jordan [17] have found a method for determining m based on the concept of fuzzy decision theory. They noted that the FCM algorithm is insensitive to the value of m in the range 1.1 to 30 and the value $m = 12$ is optimal. Foody [18] used the value $m = 1.25$ for the supervised Mahalanobis Distance fuzzy classifier



stating that this value was qualitatively assessed to give good results. Deer and Eklund [19] presented an investigation of the value of the fuzzy exponent using the criterion that fuzzy set memberships reflect class proportions in the mixed pixels of a remotely sensed image. Okeke and Karmieli [20] present a new and efficient procedure for determining a local optimal value for the fuzzy exponent in the implementation of fuzzy classification technique. In the proposed procedure, the output of the fuzzy classification is used to predict the original data set. Then the differences between the predicted data and the original data set, for a range of fuzzy exponent values, are inspected. The fuzzy exponent value that corresponds to the least distance between the predicted data set and the original data set becomes the optimal value. The method has the benefit of providing a unified method of identifying optimal fuzzy exponent.

This is achieved only with the products of classification and without considering the algorithm of the fuzzy classification itself. Thus the method is applicable to all cases (original fuzzy c-means algorithm and its extensions) wherever the value of the fuzzy exponent is required. This algorithm assumes that the optimal number of classes in the data set must have been determined. Optimal number of classes might be known a priori as in the case of supervised classification in remote sensing or determined by classification validity measures. Thus, assuming one has determined the optimal number of classes c . Krishnapuram and Keller gave some recommendations based on their findings and raised by Barni et al [21]. They said that 'fuzzifier' m is different for both FCM and PCM. The interpretation of the fuzzy exponent is also different in FCM and typical exponent in PCM. In the former, increasing values of fuzzy exponent represent increased sharing of points among all clusters, whereas in the latter increasing values of typical exponent represent increased possibility of all points completely belonging to a given cluster. Therefore, the value m that gives satisfactory results can be different in the two clustering algorithms. While an usual value for FCM is $m = 2$, Krishnapuram and Keller state that, in case of Gaussian clusters, an appropriate choice for PCM would be $m = 1.5$. Indeed, with lower values of m , the membership functions used by Krishnapuram and Keller decrease more rapidly when the when the distance of the centers. It would permit to delimit territories returning to every center, and would avoid the different center bringing together. However in most general cases, it is necessary to choose values of m as close as possible to 1 for realist centers. Barra [22] in his thesis has developed a procedure for

determination of the optimal value of the possibilistic exponent in a medical context (MRI image). He proved that every value of m included in the interval $[1.5,3]$ gives a weak rate of error classification. Also the value of m is not the same for different images.

To summarize, one can say that there is no method to optimize fuzzy and typical parameters in general; each problem requires a choice depending on the nature of the data. A value in the range $[1.5,3]$ is generally accepted and Tucker (1987) suggests in his theorem take $m \geq n / (n - 2)$ (n is the number of object) to ensure convergence of the algorithm.

Finally for the FPCM approach, it doesn't exist methods or studies that discussed the optimal values of m and η . But as in the two previous algorithms (FCM and PCM), a value 2 is given for both parameters. This value is not the optimal, but it is used for the mathematical simplifications or by practice.

4.2 New procedure for choice an optimal fuzzy and typical exponents

Before using the FPCM algorithm, the following parameters must be specified: the numbers of clusters c , the fuzziness exponent m , and the typical exponent η . As says already, one is going to study in this paper the exponents.

Researchers have long studied the fuzzy parameter, and they found different ranges and single values for optimal fuzzy exponent m . All these results can be improved and may give other values better than previous with certain criteria. Fuzzification factor influences the shape of the clusters. Typically, its value is set to 2; it is the most popular for the FCM processing of any data type. By changing the values of m , one can make the clusters look more Boolean so that seen more membership grades close to 0 or 1. This happens when m approaches to 1. On the other hand, if m increases, then the resulting membership grades lead to spike-like functions. So value 2 would not be good enough and may be misleading. This has been demonstrated in the detailed empirical and theoretical investigation of Deer and Eklund [19], Choe and Jordan [17] and Okeke and Karmieli [20], pointed out that the appropriate fuzzy exponent would depend on the complexity of the data structure; or the structure of the data set is unknown a priori and may vary considerably from one point to the other within the same data.

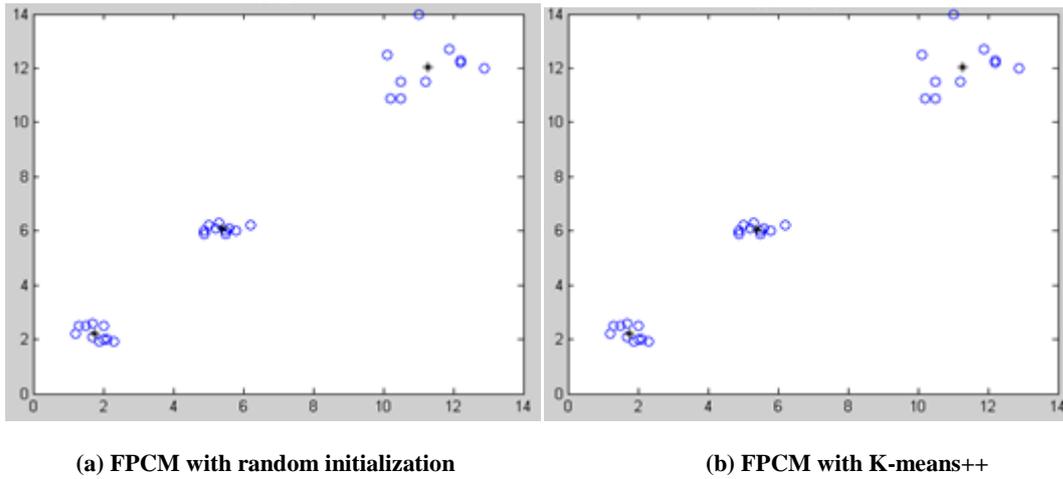


Fig 1: FPCM clustering results for the three-cluster data set

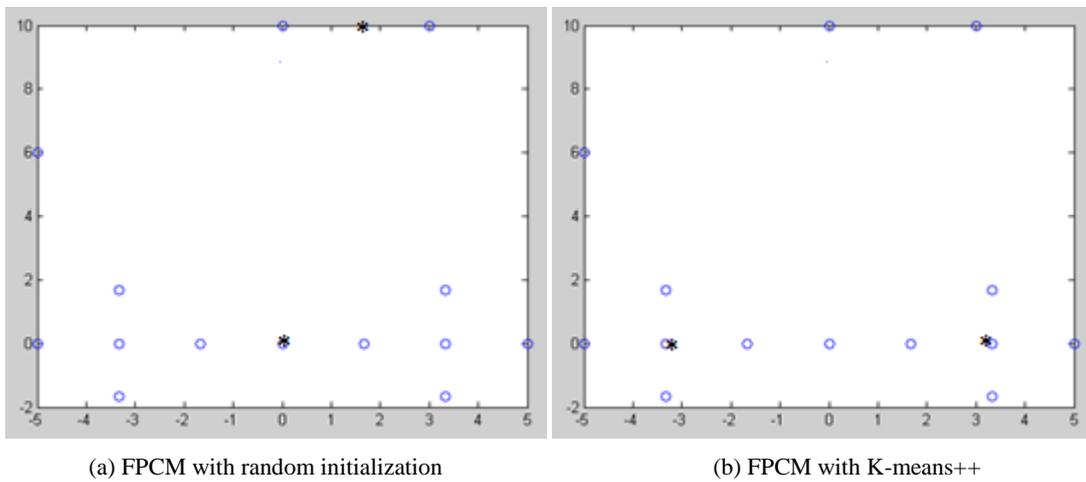


Fig 2: FPCM clustering results for the two-cluster data set with an outlier.

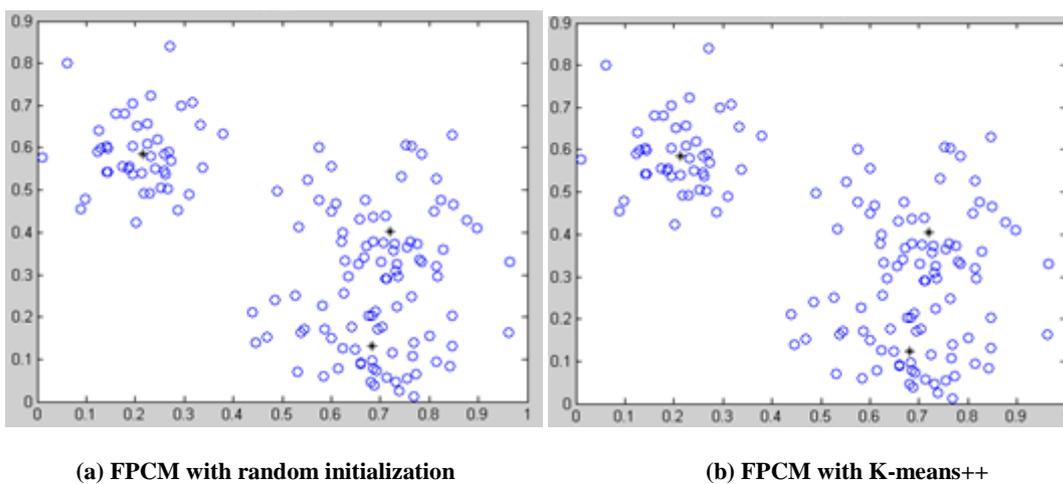


Fig 3: FPCM clustering results for the three-overlapped cluster data set with an outlier

Therefore, since every data set has a unique data structure, the optimal value of m should be peculiar to each particular data set and should be sought from within the data structure of each data set.

Different ranges and single values for optimal typical exponent, η , used by different researchers, have been stated in the previous section. For these ranges of values, there are still possibilities of making more precise choices of η -value within

the range. Like the optimal fuzzy exponent, it depends on the complexity of the data structure. The only difference through typical and fuzzy exponent, is that if increasing values of fuzzy exponent represent increased sharing of points among all clusters, whereas in the latter increasing values of typical exponent represent increased possibility of all points completely belonging to a given cluster.

The fuzzy possibilistic c-means, FPCM, is based on the fusion of fuzzy and the possibility concepts, so the objective function of this approach, is composed of two terms:

$$J_{FPCM}(U, T, V) = \sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^m d^2(x_j, v_i) + \sum_{i=1}^c \sum_{j=1}^n t_{ij}^\eta d^2(x_j, v_i) \quad (17)$$

The first term is the fuzzy function; the second term is the typical function. So to determine the optimal values of fuzzy and typical exponent in FPCM, you must find the optimal value of fuzzy exponent in FCM and optimal value in PCM.

To have an optimal value of fuzzy exponent m in FPCM; you must apply FCM algorithm, and take the products of the classification process on X to be the U_{FCM} and V_{FCM} . Also to have an optimal value of Typical exponent η in FPCM; you must apply PCM algorithm, and take the products of the classification process on X to be the U_{PCM} and V_{PCM} .

In order to reconstruct or predict the original data set from the outputs of the FCM algorithm, Okeke and Karmieli [20] proposed a new formula by using the idea of linear mixture modelling. So the formula of predict object is:

$$\tilde{X} = \tilde{V}\tilde{U} \quad (18)$$

Where \tilde{X} represents the $(1 \times n)$ vector of predicted Data set values of the data set. \tilde{V} represents the $(1 \times c)$ vector of the FCM or PCM output prototypes (centers) and \tilde{U} represents the $(c \times n)$ matrix of the FCM or PCM output degree of membership of each prototype to the data set values.

The evaluation metric used by all algorithms of clustering is the root mean squared error (RMSE). The RMSE is calculated by the root of the averaging all squared errors between the original data (X) and the corresponding predicted values data (\tilde{X}).

$$RMSE = \sqrt{\frac{\sum_{j=1}^n \sum_{i=1}^c (x_{ij} - \tilde{x}_{ij})^2}{n}} \quad (19)$$

where x_{ij} and \tilde{x}_{ij} are the actual and predicted rating values data respectively; n is the number of data and c is the number of clusters.

In order to assess the quality of a classifier on the validation data, one needs an evaluation measure to know how 'good' it is. Performance index is criterion to choose a good clustering. Optimal clusters should minimize distance within clusters (intracluster or cluster compactness) and maximize distance between clusters (intercluster or cluster separation).

Fukuyama and Sugeno tried to model the cluster validation by exploiting the compactness and the separation [23]. It consists of the difference of the compactness and the separation. The minimal value of index designates a 'good'

clustering in relation to others. In below equations, J_m is a compactness measure, and K_m is a degree of separation between each cluster (v_i) and the mean (\bar{v}) of cluster centroids. Then the performance index is formed as follows:

$$\text{Performance Index} = PI = J_m - K_m \quad (20)$$

For example for FCM, $J_m = J_{FCM}$ and

$$K_m = \sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^m d^2(v_i - \bar{v}) \quad (21)$$

It is clear that for compact and well-separated clusters one expects small values for PI. The first term in brackets measures the compactness of the clusters while the second one measures the distances of the clusters representatives. When the value of PI is negative, the classification is good.

Our approach for selecting the optimal value of fuzzy and typical exponents is based on the RMSE that measures the difference between the original data set and the reconstructed or predicted data set; and the PI that defines the index of performance of the algorithm. It must have a minimal value of RMSE and PI. We can combine the two coefficients [24] and we have a new criterion that is formed as follows:

$$CTR = \frac{PI + RMSE}{\sqrt{(PI^2 - RMSE^2)}} \quad (22)$$

The majority of the values of this index is between -1 and 1. And for different cases the minimal values is the better.

The algorithm for determining the optimal value of fuzzy exponent is as follows:

Procedure Optimal_fuzzy_typical_parameters

1. Initialization: Set $m = 1.1$, c and choose maximum value for m
2. Apply FCM algorithm to compute the fuzzy prototypes \tilde{V} and fuzzy membership \tilde{U}
3. Reconstruct the original data set using Eq. (18)
 Compute RMSE using Eq. (19)
 Compute PI using Eq. (20)
 Compute the criteria using Eq (22)
 Increment the m value; if $m \leq m_{max}$ go to 2
8. Determine the optimal value that has the minimal value of CTR.

For choose the optimal value of typical exponent, we can apply the same algorithm and modifying the FCM by PCM.

4.3 Some results

To show the feasibility of the methodology proposed in this paper, one performs some experiments to compare the performances of the algorithm with some numerical data sets. All algorithms are implemented under the same initial values and stopping conditions.

4.3.1 Example 1

In the first experiment, one uses a two-cluster data set (X11) as presented in [25] shown in Figure 4. To demonstrate the quality of classification with the optimal values of m and η in relation to the other values for the algorithm in a case data set

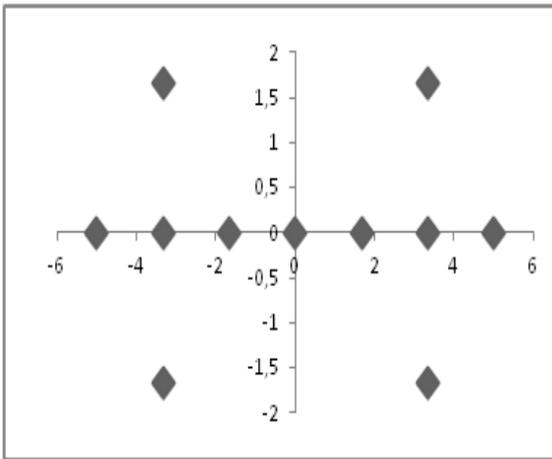
without outlier. So the change of PI, RMSE or CTR, is caused by a big modification of η and small modification of m . According to our approach, one finds an optimal value for $m = 1.6$ and for $\eta = 1.6$.

Table 1 presents the results produced by FPCM on X11 for different values of m and η . One notes that the fuzzy exponent value m has great influence over the value of typical exponent η for classification. For different values of m , we have different results; but this is not true for η . If one sees the CTR for $(m = 1.4, \eta = 1.1)$ and $(m = 1.4, \eta = 2)$, one notices that this coefficient has not changed for both situations. One can say the environment does not noise.

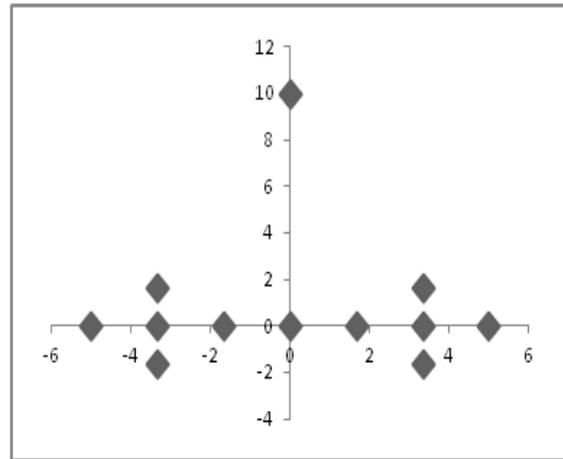
Turning now to the execution time. Table 3 presents the processing time, with the usual values and $(m = 2, \eta = 2)$; the execution time is 0.031. Contrariwise, using the procedure, the time greatly increases, the value is 1.025. So

this procedure increases the calculation and the time of running of the clustering; but it greatly increases the quality of the clustering. With evolved material resources, this execution time is not a problem; and one can say in terms of clustering quality that that our procedure gave a value for this type of problem.

In the second experiment, one uses a two-cluster data set with outlier (X12) as presented in [25] shown in Figure 4. The clustering results of the algorithm are shown in Table II. In this case one has a change in the data set, so one must have a modification in the optimal values of m and η . The optimal found are $m = 2.5$ and for $\eta = 1.3$. The processing time presented in Table III has a large difference between the algorithm without and with the new procedure.



(a) Input features vectors X11



(b) Input features vectors X12

Fig 4: Input features vectors

5. PROPOSED APPROACH

This selected method (FPCM) is an iterative algorithm clustering that is composed of two parts. The first part is fixed during the execution of the algorithm, it is an initial choices of some parameters: number of clusters, random initialization of cluster centers, maximum number of iterations, fuzzy and typical exponents. The second part is the treatment that permits to update the degrees of membership, typicality and the cluster centers until the convergence of the algorithm. This method is based on the objective function. The evolution of the method must feel the objective function and parameters that have a direct influence on results.

We propose a new architecture of the algorithm described as follows:

S1:

- Given the data set $X = \{x_1, x_2, \dots, x_n\} \subseteq R^p$.
- Choose the number of clusters $1 < c < n$.
- Choose the optimal values of fuzzy and typical parameters using the procedure:

optimal_fuzzy_typical_parameters

-- Choose the iteration limit T_{max} .

-- Choose the termination tolerance $\epsilon > 0$.

-- Pick Norm $\|x\|_A$.

-- Initialize the cluster centers $V(0)$ using the Improved K-Means++ algorithm.

For $t = 1, 2, \dots, T_{max}$

S2: Compute $U_t = \{\mu_{ij}\}$, using the formula (12).

S3: Compute $T_t = \{t_{ij}\}$, using the formula (13).

S4: Update the centers vectors $V_t = [v_i]$, using the formula (14).

S4: Compute $E_t = \|V^t - V^{t-1}\|$, if $(E_t \leq \epsilon$ or $t \geq T_{max})$, Stop; Else $t = t + 1$.

End.

The new algorithm named Improved Fuzzy Possibilistic C-means (IFPCM) contains: a procedure initialization of the centers which permits to select the first centers of classes by replacing the random initialization. And, a procedure for choice of optimal

fuzzy and typical exponents, which makes the optimal choice of parameters m and η .

We replaced the user intervention and the random by two iterative procedures. Results will be more effective, but the execution time of the method will increase.

There are other parameters that influence the algorithm and which aren't covered in this paper; these parameters are:

Number of clusters: FPCM algorithm requires the user to predefine the number of clusters (c), but it is not always possible to know this number in advance. Since the scores obtained using the c -means family algorithms depend on the choice of c , it is necessary to validate each result of the partitions once they are found. This validation is performed by a specific algorithm that allows assuming the appropriate value of the number c . We call this algorithm "validity index of the classification". It evaluates each class and determines the optimal or valid partition [23].

Termination criterion: FPCM clustering is an iterative algorithm that stops when the norm of the difference between V into two successive iterations is less than ϵ (the termination parameter). This parameter describes the stability of the cluster centers between the iterations of algorithm. Generally 0.001 is taken as the ϵ value. When ϵ tends to 0, one has more precision and more computing time.

Choice of the distance: The metric used in the algorithm, conditions the shape of point clouds to separate. In general, the distance d_2 is defined as:

$$\forall i \in [1, \dots, c], j \in [1, \dots, n] \quad d^2(x_j, v_i) = (x_j - v_i)^T A (x_j - v_i) \quad (23)$$

where A is a positive definite matrix. When A is the identity matrix, d^2 is the Euclidean distance and structure of the point cloud is spherical. Other choices are possible for A , can detect clouds of more complex form. One has already seen in previous chapters the types of distance used by variants of c -means.

6. EXPERIMENTAL RESULTS

To show the feasibility of the methodology proposed in this paper, one performs some experiments to compare the performances of the algorithm with some numerical data sets. All algorithms are implemented under the same initial values and stopping conditions. The experiments are all performed on an ACER computer with 3 GHz Pentium (11) processors using MATLAB (Mathworks, Inc. Natick, MA).

In the experiment, we tested these methods on well-known data sets from the UCI machine learning repository [26]. Both FCM and FPCM algorithms are executed with the usual parameters ($m = 2$ and $\eta = 2$), the initialization of cluster centers is performed in a random manner.

After a classifier or a cluster model has been constructed, one would like to know how good it is. Quality criteria are fairly easy to find for classifiers, or according to Borgelt [1] the quality of a clustering result is calculated while using index of performances or validity index that are used to determine the number of classes.

In Table 3. One notes that these parameters have a direct influence on the results of the algorithm. Using the Performance Index named Fukuyama-Sugeno index, who supposes that the algorithm which has the minimal value of index is the best in relation to others. Our approach proves its performance despite its running time which is strictly greater than the time of other algorithms.

7. CONCLUSION

The fuzzy and typical exponents and initialization of cluster centers are required for the implementation of the fuzzy possibilistic c -means. The questions of choice of optimal values for the parameters and good initialization have remained an open problem. This paper presents a new procedure to search the optimal fuzzy and typical exponents and another for initialization. These procedures increase the calculation and the time of running of the clustering; but this technique eliminates the arbitrarily choice and initialization at the beginning of the clustering algorithm. Numerical results using data sets are used and illustrate the simplicity and effectiveness of the proposed procedures. A comparison of the clustering algorithm FPCM with different values of parameters shows that a clustering algorithm with optimal values of fuzzy and typical exponents and a better initialization will increase the cluster compactness and the separation between classes. Finally, a numerical examples show that the clustering algorithm with optimal values (m and η) gives more accurate clustering results than the FPCM algorithm with arbitrarily parameters for typical problem. The proposed method has the benefit of providing a unified method of identifying optimal fuzzy and typical exponents. This is achieved only with the products of classification and without considering the algorithm of the fuzzy possibilistic classification itself. Thus the method is applicable to all cases wherever the values of the exponents are required. This work can be integrated with the problem of determination of the optimal value of the number of classes. So we will have a procedure for initializing the parameters of the classification method FPCM, ie, the number of the classes and the values of and typical fuzzy exponents.

Table 1. Table captions should be placed above the table

m	η	PI	RMSE	CRT
1,1	1,1	-73,353	0,188320	-0,99743598
1,1	1,4	-73,317	0,188370	-0,99743404
1,4	1,1	-75,845	0,167130	-0,99779885
1,4	2	-75,571	0,167190	-0,99779885
1,4	3	-73,833	0,167650	-0,99779009
1,6	1,1	-76,348	0,166140	-0,99773191
1,6	1,2	-76,348	0,166140	-0,99782627
1,6	1,6	-76,341	0,166140	-0,99782627
1,6	2	-76,157	0,166140	-0,99782607
1,6	3	-74,612	0,166610	-0,99782083
2	1,1	-71,673	0,168090	-0,99776947
2	1,2	-71,674	0,168090	-0,99765751
2	1,6	-71,670	0,168060	-0,99765754
2	2	-71,566	0,168180	-0,99765783
2	3	-70,260	0,169260	-0,99765276
3	1,1	-48,784	0,206400	-0,99759384
3	1,6	-48,783	0,206400	-0,99577802
3	2	-48,723	0,202700	-0,99577793
3	3	-47,401	0,204100	-0,99584837

Table 2. Results produced by FPCM for different values of the parameters m and η for X12.

m	η	PI	RMSE	CRT
1,1	1,1	18,554	0,266150	1,014300
1,1	1,4	18,944	0,266150	1,014000
1,4	1,1	6,088	0,265930	1,043700



1,4	1,4	6,669	0,265860	1,039900
1,4	2	9,781	0,265640	1,027200
1,4	3	10,846	0,265600	1,024500
2	1,3	-16,045	0,266520	-0,983390
2	1,7	-14,730	0,266370	-0,981920
2	2	-13,014	0,266190	-0,979550
2,5	1,3	-19,657	0,266840	-0,986430
2,5	1,5	-19,437	0,266810	-0,986270
2,5	1,7	-18,659	0,266700	-0,985710
2,5	2	-16,740	0,266450	-0,984080
2,5	2,5	-14,808	0,266230	-0,982020
3	1,1	-18,635	0,267160	-0,985660
3	2	-15,852	0,266710	-0,983180
3	2,5	-13,304	0,266350	-0,979980
3	3	-12,592	0,266260	-0,978860

Table 3 Performance Index generated by FCM, FPCM, IFPCM for different data sets.

DS	ND	NC	NDI	PI FCM	PI FPCM	PI IFPCM
Iris	150	3	4	-44527	-46847	-50441
breast - cancer - wisconsin - cont	683	4	9	-6299	-6402	-11512
Wine	178	3	13	10751000	11334000	16297000
Yeast	528	11	10	117,62	119,28	-476
Auto MPG	398	8	3	-1,9721e+008	-2,0202e+008	2,13445e+005
Balance Scale	625	4	3	1698,2	1711	1326
Buta	345	7	2	81790	77319	48844
glass	214	9	6	-6,1097e+005	-6,6859e+005	-728905
hayes	132	5	3	-1,3298e+005	-1,4136e+005	-150440
Monk's Problem	432	7	2	1006,6	1013	917
Lettre Image	16e+003	16	26	57556	57457	46550

ND : Number of data

DS : Data set

NC : Number of clusters

NDI : number of data items

PI FCM : Performance Index FCM

PI FPCM : Performance Index FCM

PI IFPCM : Performance Index FCM

8. REFERENCES

- [1] Borgelt, C. Prototype-based classification and clustering. Habilitation, University of Magdeburg Germany, June 2005.
- [2] Saad, M. F. and Alimi, A. M. A New Improved Fuzzy Possibilistic C-Means Algorithm Based On Weight Degree. Intelligent Automation and Computer Engineering, 90-104, 2010.
- [3] Zadeh, L.A. Fuzzy Sets. Inf. control, vol. 8, 338-352, 1965.
- [4] Bezdek, J.C. Fuzzy mathematics in pattern classification. PH.D dissertation, Cornell Univ. Ithaca, NY, 1973.
- [5] Bezdek, J.C. Pattern recognition with fuzzy objective function algorithms. Plenum, New York, 1981.
- [6] Krishnapuram, R. and Keller, J. A possibilistic Approach to Clustering. IEEE Trans. on Fuzzy Systems, vol. 1, no. 2, May 1993.
- [7] Pal, N.R. Pal, K. and Bezdek, J. C. A mixed c-means clustering model. Proceedings of the Sixth IEEE International Conference on Fuzzy Systems, vol. 1, 11-21, Jul. 1997.
- [8] Forgy, E.W. Cluster analysis of multivariate data: Efficiency versus interpretability of classification. Biometrics, vol. 21, 768-769, 1965.
- [9] Saad, M. F. and Alimi, A. M. Modified Fuzzy Possibilistic C-means. Proceedings of the International MultiConference of Engineers and Computer Scientists, vol. 1, 177-182, 2009.
- [10] Pena, J., Lozano, J. and Larranaga, P. An empirical Comparison of four Initialization Methods for the k-means Algorithm. Pattern Recognition Letters, vol. 20, 1027-1040, 1999.
- [11] Kaufman, L. and Rousseeuw, P.J. Finding Groups in Data, an Introduction to Cluster Analysis. Wiley, Canada, 1990.
- [12] Velthuizen, R., Hall, L. and Clarke, L. Unsupervised fuzzy segmentation of 3D magnetic resonance brain images. In Proceedings International Symposium on Electronic Images: Science and Technology, 627-635, Jan 31-Feb 4, 1993.
- [13] Bahman, B., Benjamin, M., Andrea, V., Ravi, K., Sergei, V. Scalable k-means++, Proceedings of the VLDB Endowment, vol. 5, 622-633, March 2012.
- [14] Shanthi, R. and Suganya, R. Enhancement of Fuzzy Possibilistic C-Means Algorithm using EM Algorithm. International Journal of Computer Applications, vol. 61, no. 12, 10-15, January 2013.
- [15] Cannon, R. L., Dave, J. V. and Bezdek, J.C. Efficient Implementation of the Fuzzy C-Means Clustering Algorithms. IEEE Trans. Pattern Anal. Mach. Intell. vol. 8, no. 2, 248-255, 1986.
- [16] Eef, M., Marc, V.M. Philippe, D.S., Timothy, S. Mohammad, M.I., Fun, M., Ellen, V. and Gunther, G.



- Imaging a Polygonal Network of Ice-Wedge Casts with an Electromagnetic Induction Sensor. *Soil Science Society of America Journal*, vol. 75, no. 6, 2095-2100, 2011.
- [17] Choe, H. and Jordan, J.B. On the optimal choice of parameters in Fuzzy C-Means Algorithm. *Proc of the IEEE International Conference on fuzzy systems*, 349-354, 1992.
- [18] Foody, G.M. and Cox, D.P. Sub-Pixel and cover composition estimation using mixture model and fuzzy set membership. *International Journal of Remote Sensing*, vol. 10, 1823-1842, 1989.
- [19] Deer, P.J. and Eklund, P. A study of parameter values for a Mahalanobis distance fuzzy classifier. *Fuzzy Sets and Systems*, vol. 137, no. 2, 191-213, 2003.
- [20] Okeke, F. and Karnieli, A. Linear mixture approach for selecting fuzzy exponent value in fuzzy c-means algorithm. *Ecological informatics*, vol. 1, 117-124, 2006.
- [21] Barni, M., Cappellini, V. and Mecocci, A. Comments on A possibilistic approach to clustering. *IEEE Trans. Fuzzy Systems*, vol. 4, 393-396, 1996.
- [22] Barra, V. Fusion 3D Images of the Brain: Study of Models and Applications. thesis: University Blaise Pascal and University Auvergne, France, July 2000.
- [23] Saad, M. F. and Alimi, A. M. Validity Index and number of clusters. *International Journal of Computer Science Issues*, vol. 9, no. 3, January 2012.
- [24] Belton, V. and Stewart, T.J. Multiple criteria approach: an integrated perspective. Kluwer Academics Publishers, Netherlands, 2001.
- [25] Bezdek, J.C., Popescu, M. , Keller, J.M. Comparing Fuzzy, Probabilistic, and Possibilistic Partitions. *IEEE Transactions on Fuzzy Systems*. vol. 18, no. 5, 906-918, 2010.
- [26] Blake, C.L. and Merz, C.J. UCI Repository of Machine Learning Databases. University of California, Irvine, CA, USA, 1998.