A Method to Enhance the 'Possibilistic C-Means with Repulsion' Algorithm based on Cluster Validity Index

Juan Wachs, Oren Shapira and Helman Stern

Dept. of Industrial Engineering and Management, Ben-Gurion University of the Negev
Be'er Sheva 84105 Israel

Abstract. In this paper, we examine the performance of fuzzy clustering algorithms as the major technique in pattern recognition. Both possibilistic and probabilistic approaches are explored. While the Possibilistic C-Means (PCM) has been shown to be advantageous over Fuzzy C-Means (FCM) in noisy environments, it has been reported that the PCM has an undesirable tendency to produce coincident clusters. Recently, an extension of the PCM has been presented by Timm *et al.,* by introducing a repulsion term. This approach combines the partitioning property of the FCM with the robust noise insensibility of the PCM. We illustrate the advantages of both the possibilistic and probabilistic families of algorithms with several examples and discuss the PCM with cluster repulsion. We provide a cluster validity function evaluation algorithm to solve the problem of parameter optimization. The algorithm is especially useful for the unsupervised case, when labeled data is unavailable.

Keywords: Possibilistic and probabilistic fuzzy clustering; Fuzzy C-Means; Cluster validity index, Robust methods.

1 Introduction

Cluster analysis is the process of classifying objects into subsets that have meaning in the context of a particular problem. The objects are thereby organized into an efficient representation that characterizes the population being sampled. (Jain and Dubes 1988). Clustering has a place of honor in many engineering fields such as pattern recognition, image processing, system modeling, data mining, and so on. The process of clustering unlabeled data consists of dividing a set of n observations $X=\{x_1,x_2,\ldots,x_n\}$ into $1<c<n$ subgroups such that each subgroup represents a "natural" substructure in X. Hard clustering methods assume that each data vector belongs to one class, however in practice clusters may overlap, and data vectors belong partially to several clusters. This scenario can be modeled properly using fuzzy set theory (Zadeh 1965), in which the membership degree, u_{ik} of a pattern x_k to the i-th cluster is a value in the interval [0,1]. Bezdek (1982) explicitly formulated this approach oriented to clustering by introducing the Fuzzy-C-Mean (FCM) clustering algorithm. Unfortunately, this method is sensitive to noise and outliers in the data. To reduce this undesirable effect, a number of approaches have been proposed, but the most remarkable has been the possibilistic approach first introduced by Krishnapuram and Keller (1993), with their Possibilistic - C-Means (PCM) algorithm. In this algorithm the membership is interpreted as the compatibilities of the datum to the class prototypes (typicalities) which correspond to the intuitive concept of degree of belonging or compatibility. These typicality-based memberships automatically reduce the effect of noise and outliers, and improve the solution. Nevertheless, the main drawback with this approach consists on the quality of the initializations. In the case of poor initializations, it is possible that the PCM will converge to a "worthless" partition where part or all the clusters are identical (coincident) while other clusters go undetected. To avoid the undesirable tendency to produce coincident clusters, a mixed c-Means approach was proposed (Pal et al. 1997) called Fuzzy-Possibilistic C-Means (FPCM). This algorithm suggests an iterative alternating optimization approach to find local minima of both objective functions. Upon closer examination of the basic architecture of the FPCM, two conclusions arise. Similar results would be obtained by applying the FCM proceeded by the PCM, or any other fuzzy clustering algorithm that yield a rich initial partition of the datum. The other conclusion is that the mutual evaluation of both membership functions (typicalities and fuzzy membership) is unnecessary, since the main approach of the possibilistic theory is to present the 'belongness' of a point to a dataset as a typicality only. The evaluation of the fuzzy membership values are not of informational contribution at all. Recently, a new scheme has been proposed, in order to overcome the problem of cluster mutual attraction forces, by introducing a supplementary term for cluster repulsion (Timm et al. 2001). By the use of cluster repulsion, a good separation between clusters is obtained, as with the FCM, while keeping the intuitive con-

cept and the noise insensibility introduced by the PCM. The goal of this paper is to establish a connection between the possibilistic approach and cluster validity indices, such that the quantitative superiority of the 'PCM with repulsion' over other methods is established. Finally, we study how to select the internal parameter of this last algorithm, and provide a procedure to find it .

The organization of this paper is as follows. In Section 2, we review the probabilistic clustering approach and its two main limitations, i. e., sensibility to noise environments and unintuituitive interpretation of membership values. In Section 3, we analyze the possibilistic approach by Krishnaparum and Keller (1993) developed to cope with the problem of noise and concept of compatibility, which lacks of clusters discrimination. In Section 4, we discuss a unified view of both probabilistic and possibilistic approaches, which are severely compromised by the multiple outputs and unspecified correct interactive parameter range. In Section 5, we review the recently proposed method by Timm et. al. (2001) based on repulsion between clusters, and report the difficulty of choosing the proper value of the weighting factor γ. In Section 6, we compare all the four clustering techniques using several examples and suggest a graphical method to obtain the optimal weighting factor γ; Finally, Section 7 presents our summary and conclusions.

2 Probabilistic Fuzzy Clustering

The most widely used prototype-based clustering method for data partition is probably the ISODATA or FCM algorithm (Bezdek 1982). Given a set of *n* data patterns, $X = x_1 ,...., x_k ,..., x_n$, the algorithm minimizes a weighted within group sum of squared error objective function, *J(U,V)*.

$$J(U,V) = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik}^{m} d^{2}(x_k, v_i) \tag{1}$$

Where x_k is the *k*-th p-dimensional data vector (or pattern), v_i is the prototype of the center of cluster *i*, u_{ik} is the degree of membership of x_k in the *i*-th cluster, *m* is a weighting exponent on each fuzzy membership, $d(x_k,v_i)$ is a distance measure between data pattern x_k and cluster center v_i, *n* is the number of data patterns, and *c* is the number of clusters. The objective function *J(U,V)* is minimized via an iterative process in which the degrees of membership u_{ik} and the cluster centers v_i are updated:

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^{m} X_k}{\sum_{k=1}^{n} u_{ik}^{m}} \quad , \quad u_{ik} = \frac{1}{1 + \sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{ij}} \right)^{\frac{2}{m-1}}} \tag{2}$$

Where, the u_{ik} satisfies

$$0 < \sum_{k=1}^{n} u_{ik} < n, \quad \forall i \qquad u_{ik} \in [0,1], \quad \forall i, k \qquad \sum_{i=1}^{c} u_{ik} = 1, \quad \forall k \tag{3}$$

In the proposed methodology, the FCM algorithm is provided with a training set of samples. Once the clusters have been created, they are manually labeled. (i.e., assigned a linguistic description). This is conducted manually. However there are some novel methods for labeling customization (Wachs et. al. 2003). The cluster center, v_i is a prototype pattern for cluster *i*, x_k is the vector of the *k*-th exemplar in the training set, u_{ik} is the 'degree of belonging' (membership value) of the *k*-th vector to cluster *i*, *c* is the number of clusters, and *n* is the number of samples in the training set.

FCM-type algorithms share the problem of sensibility to noise and outliers as do all the least squares approaches (LS). Low values of membership associated to noise points may express noise contamination, however, as can be seen from (2), the membership values generated due to the third constraint in (3), are relative numbers (Dave and Krishnapuram, 1997). This means that noise points and outliers will have at least a value of 1/c to all the remaining clusters. Any increase of this membership value implies a decrease to the belonging degree of that point to any other cluster. The last fact, leads to the idea, that noise points may have high membership values and therefore they will affect the prototype parameter estimates.

The second main deficiency of the FCM, as pointed out in Krishnapuram and Keller (1993), is that due to the last constraint in (3), this membership is interpreted as degrees of sharing, but not as degrees of possibility of a point belonging to a class. A degree of typicality or possibility of belonging may be better suited for classical fuzzy set theory.

3 Possibilistic Fuzzy Clustering

As expressed in Section 2, constraint (3) assures relative values for the degree of membership, and therefore is not suitable for applications where memberships are supposed to represent typicalities or compatibilities. Thus, in the FCM the memberships in a given cluster of two points that are equidistant from the prototype of the cluster can be significantly different and memberships of two points in a given cluster can be equal even thought the two points are arbitrarily far away from each other (Krishnapuram and Keller 1996). This situation is illustrated in Figure 1.

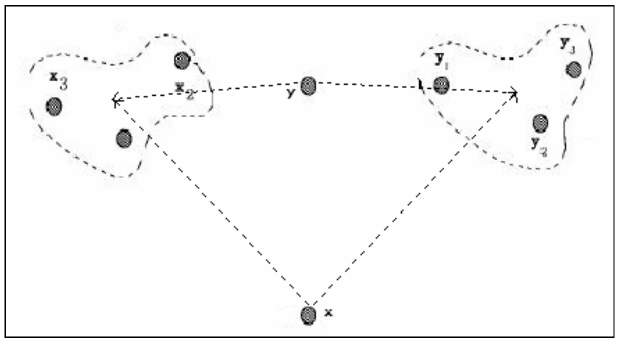


Figure 1. Example of dataset with a noise and outlier points

In this example, there are two clusters and a pair of points *x,y* ; which represents an outlier and a noise point respectively. Intuitively, point *y* shouldn't have a high degree value of membership (in the sense of compatibility) value for any cluster. Point *x*, should have a smaller degree of membership to both clusters, since it it is farther away from them than point y. Both points *x* and *y* will be assigned a membership value of 0.5 to both clusters by the FCM. One concludes that this membership values are unrepresentative of the degree of 'belonging', but also they cannot discriminate between an outlier datum and noise.

The PCM formulation relaxes the objective function (1) by dropping the sum to 1 constraint (3). In order to avoid a trivial solution of $u_{ik} = 0$ for all i, a penalty term is added which forces u_{ik} to be as large as possible. This was done by modifying the objective function in (1) as follows:

$$J(U,V) = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik}^{m} d^{2}(x_k, v_i) + \sum_{i=1}^{c} \eta_i \sum_{k=1}^{n} (1 - u_{ik})^{m} \quad (4)$$

Where η_i is a positive number, and $u_{ik} \in [0,1]$. The new membership update equation is:

$$u_{ik} = \frac{1}{1 + \left[\dfrac{d^{2}(x_k, v_i)}{\eta_i} \right]^{1/(m-1)}} \quad (5)$$

The parameter η_i is evaluated for each cluster separately; it determines the distance at which the membership degree equals 0.5.

$$\eta_i = K \sum_{k=1}^{n} u_{ik}^{m} d^{2}(x_k, v_i) \Big/ \sum_{k=1}^{n} u_{ik}^{m} \quad (6)$$

Using (6) η_i is proportional to the average fuzzy intracluster distance of cluster v_i. Usually K is chosen to be 1.

The main drawback of this promising approach appears when the objective function in (4) is truly minimized, and this occurs only if all cluster centers are identical (coincident centroids). This failure is due to the reason that the membership degrees (5) depend only on the distance between the point to the cluster, and not on its relative distance to other clusters. Usually only part of the centroids are coincident, since the algorithm converges to a local minimum of the objective function (4). Still this represents undesirable behavior for a clustering algorithm (Barni et. al. 1996).

4 Mixed c-Means Clustering

The model proposed by Pal et. al.(1997) called fuzzy-possibilistic c-means (FPCM) establishes a connection between the possibilistic and probabilistic approaches. The FPCM creates both memberships (in the sense of relative belonging) and typicalities, along with the usual centroids through a standard alternating optimization process. The current approach solves the noise susceptibility characteristic of the FCM, and also overcomes the coincident clusters problem of the PCM. A modified version of the objective function (1) is obtained by adding a possibilistic term, within the typicality, based on all the n data points rather on all the c centroids:

$$J(U,V)=\sum_{i=1}^{c}\sum_{k=1}^{n}M_{ik}d^2(x_k,v_i) \tag{7}$$

Where $M_{ik}=u_{ik}^m+t_{ik}^{\eta}$, given positive reals m and η .

The membership formula constraint remains the same as in (2) and (3) respectively. Through distributing the typicalities w.r.t. all n samples, we get:

$$\sum_{k=1}^{n}t_{ik}=1 \ , \quad \forall i \tag{8}$$

$$t_{ik}=\left(\sum_{j=1}^{c}\left(\frac{d^2(x_k,v_i)}{d^2(x_k,v_j)}\right)^{2/(\eta-1)}\right)^{-1}, \quad \forall i,k \tag{9}$$

$$v_i=\frac{\sum_{k=1}^{n}M_{ik}x_k}{\sum_{k=1}^{n}M_{ik}}, \quad \forall i \tag{10}$$

This approach lacks of justification for the computation of both membership and typicalities values; and the correct range for the interactive parameters (m , η) are not specified.

5 Possibilistic Fuzzy Clustering with Repulsion

Recently, the Possibilistic Fuzzy Clustering with Repulsion was proposed to address the drawbacks associated with the FCM and the PCM. The undesirable behavior of the PCM can be explained by the fact that the objective function (4) is truly minimized only if all the centroids are identical (coincident), since the typicality of a point to a cluster, depends only on the distance between the point to that cluster.

This method aims to minimize the intracluster distances (Bezdek and Pal, 1998), while maximizing the intercluster distances, without using implicitly the restriction (3), but by adding a cluster repulsion term to the objective function (4).

$$J(U,V)=\sum_{k=1}^{n}\sum_{i=1}^{c}u_{ik}^m d^2(x_k,v_i)+\sum_{i=1}^{c}\eta_i\sum_{k=1}^{n}(1-u_{ik})^m+\gamma\sum_{i=1}^{c}\sum_{k=1,k\neq i}^{c}\frac{1}{d^2(v_i,v_k)} \tag{11}$$

Where γ is a weighting factor, and u_{ik} satisfies:

$$u_{ik}\in[0,1] \quad ,\forall i \tag{12}$$

The repulsion term is relevant if the clusters are close enough. With growing distance it becomes smaller until it is compensated by the attraction of the clusters. On the other hand, if the clusters are sufficiently spread out, and the intercluster distance decreases (due to the first two terms), the attraction of the cluster can be compensated only by the repulsion term.

Minimization of (11) with respect to cluster prototypes leads to:

$$v_i = \frac{\sum_{j=1}^{n} u_{ij} x_j - \gamma \sum_{k=1,k\neq i} v_k \frac{1}{d^2(v_k, v_i)}}{\sum_{j=1}^{n} u_{ij} - \gamma \sum_{k=1,k\neq i} \frac{1}{d^2(v_k, v_i)}} \tag{13}$$

Singularity occurs when one or more of the distances $d^2(v_k, v_i) = 0$ at any iteration. In such a case, (13) cannot be calculated. When this happens, assign zeros to each nonsingular class (all the classes except *i*) and assign 1 to class *i,* in the membership matrix *U*. Similar as for the PCM algorithm, the formula for updating the membership degrees u_{ik} is obtained using (5).

An alternative repulsion term for (11) was also suggested by the same authors, in order to minimize the objective function:

$$J(U,V) = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik}^{m} d^2(x_k, v_i) + \sum_{i=1}^{c} \eta_i \sum_{k=1}^{n} (1-u_{ik})^m + \gamma \sum_{i=1}^{c} \sum_{k=1,k\neq i}^{c} e^{-d^2(v_k, v_i)} \tag{14}$$

The main difference between both terms, are the manner in which the repulsion term decreases with growing interdistances. The strategy consisting of adding a repulsion term, avoids the problems of the possibilistic cluster analysis as described in Section 3. The weighting factor γ is used to balance the attraction and repulsion forces, i.e., minimizing the intradistances inside clusters and maximizing the interdistances between clusters. The central problem of this algorithm, as with the previous one, is that it requires a resolution parameter (in this case the weighting factor γ) and no clue is given about the correct range of this parameter (Dave and Krishnapuram, 1997), in order to get "better" partitions and noise avoidance.

The problem of unknown bounds of the weighting factor is related to cluster validity methods. Cluster validity studies the "goodness" of a partition generated by a clustering algorithm. The sum of intracluster distances, over the minimum of the intercluster distances is one of the most commonly used validity measures because of its analytical simplicity. This formulation is known as the Xie-Beni index v_{XB}, and is defined as:

$$v_{XB}(U,V;X) = \frac{\sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^2 \parallel x_k - v_i \parallel^2}{n(\min_{i\neq j}\{\parallel v_i - v_j \parallel^2\})} \tag{15}$$

A good *(U,V)* pair should produce a small value of (15) because u_{ik} is expected to be high when $||x_k\text{-}v_i||$ and well-separated v_i's will produce a high value in the denominator of (15). Consequently, the minimum of v_{XB} , is the most desirable partition.

6 Comparative Tests

In this section, we use several examples to show the comparative properties of the four algorithms discussed above. First a simple example is to illustrate the noise immunity property of the possibilistic approach and it advantages over the probabilistic approach. Later, a more realistic example is presented and the performances are compared between the fuzzy, possibilistic and unified algorithms based on accuracy measures. Last, a validity measure will be employed as an estimate for the optimal selection of the weighting factor γ.

The first example illustrates two well-separated noise-free clusters of 30 points each, drawn from two components, each one from a normal distribution, with μ_{1x}=2, μ_{1y}=3, μ_{2x}=5, μ_{2y}=3 and σ=0.5, see Figure 2.a. In Figure 2.b, the crisp partition for the FCM, PCM, 'FPCM', and 'PCM with repulsion' are similar, the centroids are almost the same. Each point is assigned to the cluster which it has the highest membership for the FCM, and for the possibilistic algorithms, the highest typicality was used. Ties are broken arbitrary. The parameters used were: *m*=2, *c*=2, ε=0.0001, η=2, γ=10.

After adding 28 points of random noise to one cluster (the cluster on the right of Figure 1.a) of the data set with μ_{2x}=6.5, μ_{2y}=4.5 and σ=1, the crisp partition presents significant differences; the FCM and 'FPCM' performs the worst, while the other two present similar partitions, see Figures 2.c, 2.d, 2.e and 2.f. The FCM algorithm was used to obtain a good initialization for the PCM. The noise addition to the original data affected significantly the cluster centers in the probabilistic based algorithms, while in the possibilistic based algorithms, the cluster centers are virtually unchanged.

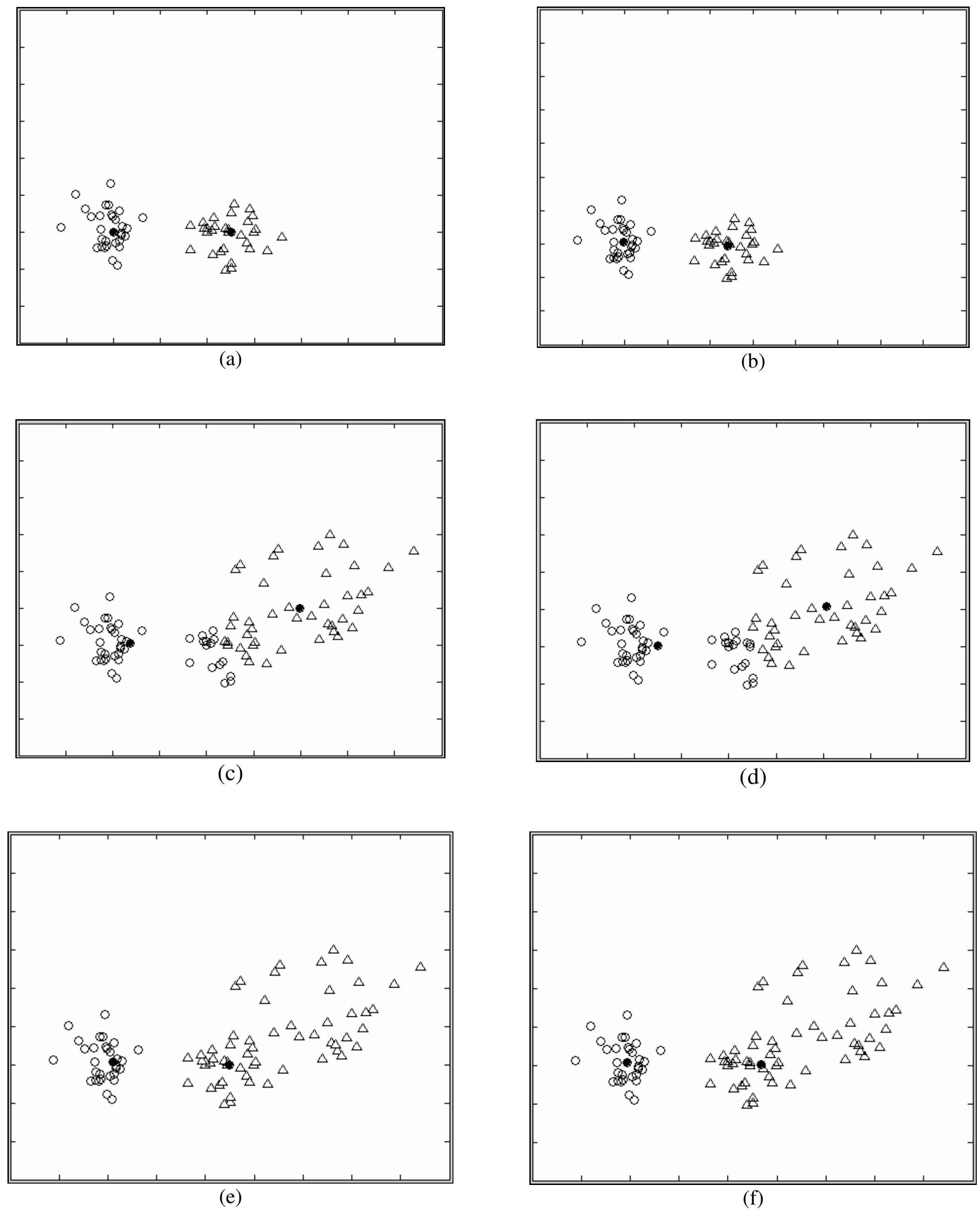


Figure 2. Partition of the synthetic data set: (a) the original data; (b) the partition originated by the FCM, FPCM, PCM and 'PCM with repulsion'; (c) the partition of the FCM with noise; (d) the partition of the FPCM with noise; (e) the partition of the PCM with noise; (f) the partition of the 'PCM with repulsion', with noisy data set.

Table 1, shows the centroids of the clusters, as seen by the different four algorithms. The last column presents the allocation deviation of the centroids from their original location, after adding the noisy data. The performance of the PCM and 'PCM with repulsion' are acceptable. The FCM algorithm gave the poorest estimates for the centroids, since its deviation is the highest.

The second example shows a more realistic example with the well-known IRIS data set (Fisher 1936). IRIS consists of 150 points in four dimensions that represent three physical classes each with 50 points.

The numerical representation of two classes has substantial overlap, while the third is well separated of the other two, therefore a three clusters classification is recommended. The algorithms discussed above have been tested on the IRIS data set, only petal length and width attributes have been used, since they carry the most information about the distribution of the iris flowers. Several runs of the four algorithms on IRIS data set were made, with different initializations and different (m, η, γ) tuples. Here we report results only for the initialization in Table 2. The performance of each algorithm (classification accuracy) is measured as the ratio between numbers of correct classification, and the total sample set. The number of mistakes is based on comparing the hardened version of the membership and typicalities matrices, to the physically correct crisp 3-partition of IRIS.

Table 1. Centroids estimation using the FCM, PCM, FPCM, and 'PCM with Repulsion' for Figure 2

	Fuzzy c-Means	Possibilistic	Mixed c-Means	Possiblistic with Repulsion
Original Data Set	(1.89 3.08) (4.47 2.91)	(1.96 3.04) (4.41 2.93)	(1.89 3.08) (4.47 2.91)	(2.01 3.06) (4.37 2.93)
With Noise	(2.49 3.02) (6.08 4.06)	(2.10 3.05) (4.48 2.98)	(2.29 3.04) (5.90 3.94)	(1.92 3.06) (4.69 3.01)
Deviation	2.06	0.16	1.81	0.34

Table 2. Classification accuracy on the IRIS data using the FCM, PCM, FPCM, and 'PCM with Repulsion'

Parameters			Accuracy				Iterations			
m	η	γ	FCM	FPCM	PCM	rep.PCM *	FCM	FPCM	PCM	rep. PCM *
2	3	0.1	82%	82%	64%	53%	26	12	26	13
2	3	1				78%				
2	3	15				97%				
2	3	50				89%				
2	3	100				76%				
2	3	200				62%				
2	2	0.1	82%	83%	64%	53%	26	12	26	13
2	2	1				78%				
2	2	15				97%				
2	2	50				89%				
2	2	100				76%				
2	2	200				62%				
2	5	0.1	82%	83%	64%	53%	26	12	26	13
2	5	1				78%				
2	5	15				97%				
2	5	50				89%				
2	5	100				76%				
2	5	200				62%				
3	2	0.1	90%	95%	70%	58%	25	13	25	14
3	2	1				80%				
3	2	15				97%				
3	2	50				66%				
3	2	100				65%				
3	2	200				62%				

From Table 2, we find that the FPCM and 'PCM with repulsion' give higher or same accuracies than FCM for best γ cases, which indicate that typicality based [*]classification represents better the physical data-partition, than membership values. Note that typicality-based classification accuracy failed only for the PCM case, where two cluster coincidence affected the performance of the algorithm in all four cases presented. The 'PCM with repulsion', using a good selection of parameter γ, shows significant superiority over the other algorithms studied here. Fortunately, Table 2, also shows that the number of iterations required for the PCM 'with repulsion' is similar to the FPCM, and is about half of that of the PCM and FCM.

Clusters detected by the PCM as a function of gamma, using m=2, are depicted in Figure 3. For γ=0.1 only two clusters are detected because the possibilistic algorithm is not forced to divide the data, and both clusters are coincident (Figure 3.a). By incrementing γ the attraction between clusters, decreases and the centroids of the coincident clusters are driven apart. For γ=1, the distance between samples assigned to clusters and their respective centroid is minimized, and that the distance between clusters is maximized, (Figure 3.b, Figure 3.c). For further values of γ, the repulsion increases with the distance of the clusters, driving them ever farther apart, hence harming the classification accuracy (see Figure 3 (d)). As shown in Table 2, in the 'rep.PCM' column, the parameter γ affected the performance of the data partition, and therefore, cluster validity measures should give some clue about the optimal value of γ.

[*] The objective function used has been defined in (11).

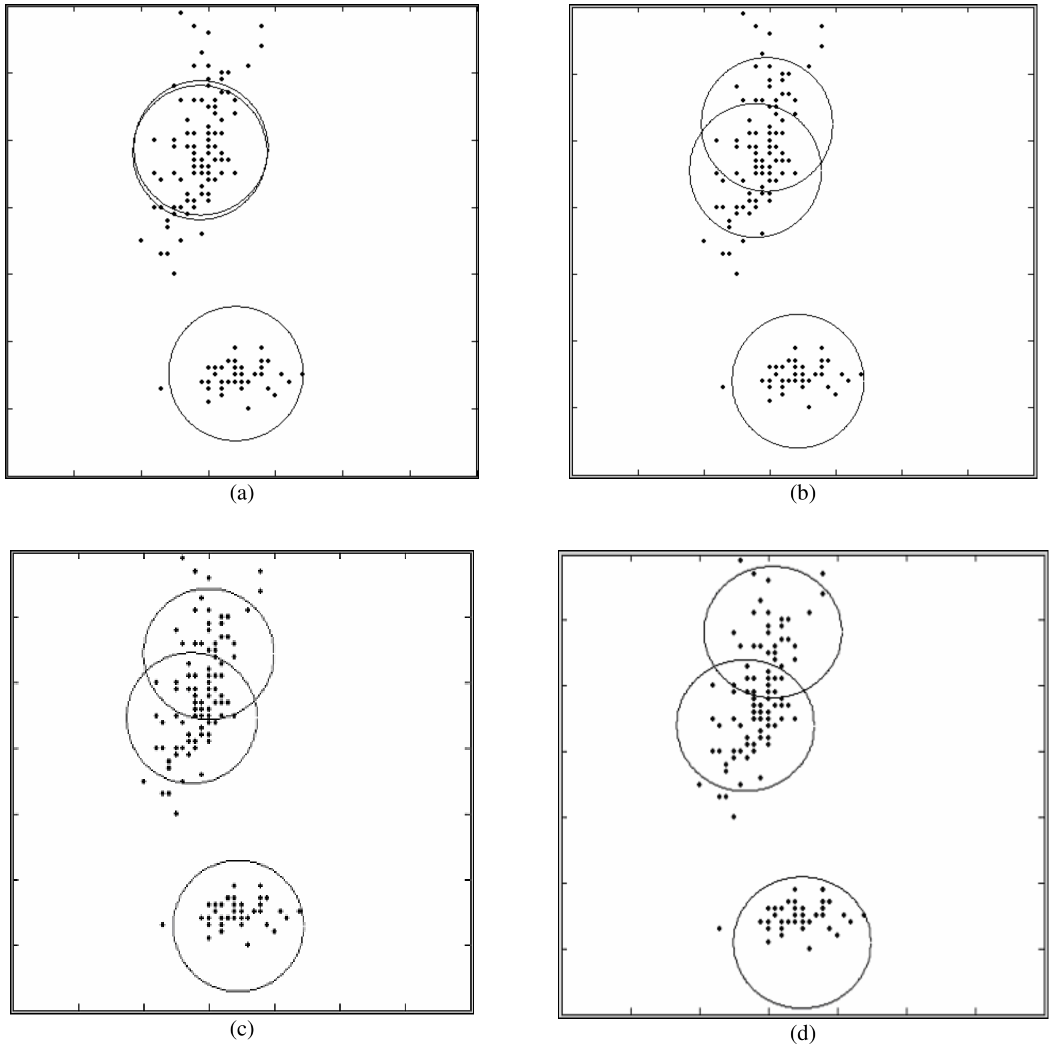
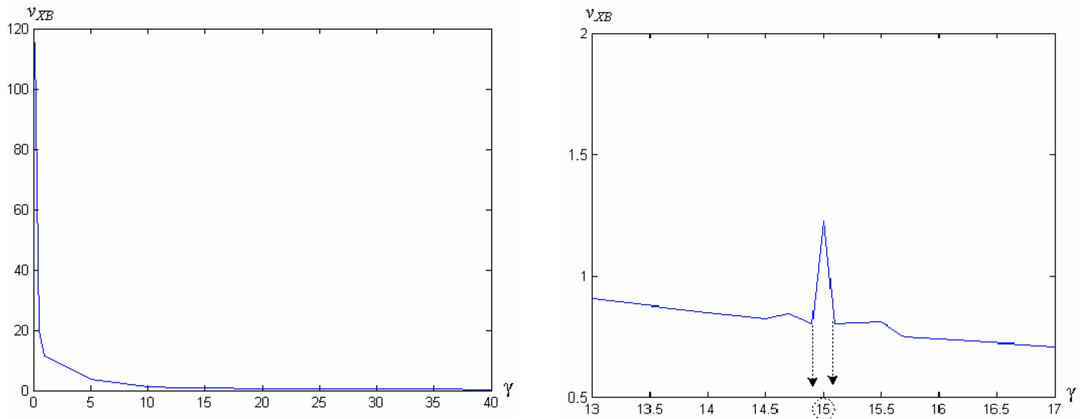


Figure 3. Iris dataset classified using 'PCM with repulsion': (a) γ=0.1, (b) γ=1, (c) γ=20, (d) γ=40.

For an unsupervised data set it is not possible to use the classification accuracy measure, and therefore an alternative method must be employed to find the best value of the weighting factor γ. We note that for the optimal value of γ, all the clusters are overall compact, and separate to each other. A smaller value of the compactness and separation validity function (15) indicates a better partition of the data set. For small values of γ, coincident centroids appear, then the denominator of (15) is negligible, and as a result of this, the v_{XB} is very high. However it, is noted that v_{XB} as a function of γ, is monotonically decreasing when γ gets very large (the denominator of (15) grows to infinity), see Figure 4.a.

Nevertheless, a close-up of the v_{XB} as a function shows that a local minimum is reached when the centroid is placed at its optimal value. Still, if the centroid is misplaced far enough from all the others centroids, the points that should belong to clusters will be assigned to the closer centroid, and hence v_{XB} will drop abruptly. We conclude that the use of the Xie-Beni validation index is valid only around the optimal values for the centroids, where the validation function behaves convex. The method proposed is to find the optimal γ by evaluating the function v_{XB} over a range of values of γ. This function is plotted in Fig 4(a). In Fig 4(b) a zoom of the region around v_{XB}=15 clearly detects the "peak". All the γ values immediately before and after the peak are candidates to be optimal. The γ* (optimal weighting factor) is the one which yields the lower v_{XB} between all the candidates. Figure 4.b, shows the plots for the hard partition of IRIS obtained using the 'PCM with repulsion', with *m*=2. The figure illustrates two valleys around the peak in γ=15, where the optimal between them is γ*=15.1, since there the v_{XB} is the lowest (v_{XB*} =0.8).



(a) (b)

Figure 4. Plots of v_{XB} vs. γ for the IRIS data set: (a) The monotonic decreasing function. (b) A close-up of the peak around γ =15.

Observing Table 2, we see that a close to optimal classification accuracy (97%) was obtained for γ=15, therefore the result obtained graphically is satisfactory. This approach has been tested on the synthetic data set with and without noise, and we report results of 100% of classification accuracy using γ*=12 and γ*=15 respectively, see Figure 4.(b)-(c). The last test has been conducted on the Wisconsin Breast Cancer Data (Merz and Murphy, 1996), and a 95% of classification accuracy was obtained using γ*=14.9, for more details, see (Shapira and Wachs, 2004). The method, based on the validity function v_{XB}, , has been implemented in the following pseudo algorithm:

1) **Initialize** γ←0.2, Δ←0.1, m←2, K←1, $v_{XB}(U,V;X;\gamma-\Delta)$←0;
2) **Initialize** centroids v_i randomly;
3) Use (5) to **update** typicalities u_{ik} and use (2) to **update** centroids;
4) **Calculate** η_i using (6);
5) **Do** converge test; if negative goto 3;
6) **Compute** function $v_{XB}(U,V;X;\gamma)$;
7) **Repeat** steps 2,3,4,5 for $v_{XB}(U,V;X;\gamma+\Delta)$;
8) **If** $v_{XB}(U,V;X;\gamma-\Delta)< v_{XB}(U,V;X;\gamma)< v_{XB}(U,V;X;\gamma+\Delta)$ **then** γ_k= γ;
9) **If** γ=stop_value **then stop**; **else** γ= γ+2 Δ;
10) **Goto** 2;
11) **Find** $\gamma^*=\min v_{XB}(U,V;X;\gamma_k)$ over all k;

Steps 2,3,4 and 5 are the ‘PCM with repulsion’ algorithm. The convergence test is $\|J_{i+1} - J_i\| < \varepsilon$ where i is the number of iteration, J is the cost function obtained from (11) or (14), and ε is the accepted error margin. The fuzziness parameter m usually is set to 2, and the number of clusters c is fixed a priori. Initial values of v_i are datum selected randomly from the sample set. Note, the examples presented here contained a single peak, however, the algorithm will handle multiple peaks, assuming each peak is surrounded by valleys (each containing strictly a local minimum).

The results for the ‘PCM with repulsion’, compared to other possibilistic and probabilistic algorithms, showed to be close to optimum when the weighting factor γ was obtained using the graphical method detailed in this paper. This method is suboptimal, however, computationally expensive in high dimensions. Its performance can be improved by considering the shapes of clusters and a better validity index.

7 Conclusions

We have reviewed four algorithms for possibilistic and probabilistic clustering: FCM, PCM, FPCM, and the ‘PCM with repulsion’. We then proposed a functional evaluation method and a strategy to obtain a good range for the weighting factor γ, based on the Xie and Beni index validation measure. Computational examples on three data sets were used to compare the four algorithms described in this paper, and to support the assertion that the weighting factor obtained by our function evaluation algorithm is useful. In summary, the PCM in its original form is not very suitable for clustering due to the undesirable coincident centroids, but it provides robustness to noise and an intuitive interpretation of the membership values. The 'PCM with repulsion' overcomes this problem, and showed a satisfactory performance when a proper value for the weighting factor γ was used. Our study has shown a correspondence between the Xie-Beni validity function and the range of the weighting factor of γ. A function evaluation algorithm to search for the optimal γ has been provided .The points obtained using the validity index for discrete γ is a quantization of all the points on the curve of the validity function, and therefore the "peaks" might be missed if to coarse a quantization is used. Therefore, a limitation of this method is that the identification of the sharp peaks of the function is scale dependent. Nevertheless, when the quantization is fine, the accuracy of the estimates is rich. The tradeoff is that for fine quantization of the weighting factor, the search over the validity function requires more computation. Although the method proposed is computationally intensive, it is still better than supervising the clustering algorithm, especially when labeling knowledge is unavailable.

Further investigation and more numerical tests are required before much can be asserted about an optimal range for the weighting factor γ, in the presence of noisy environments, which may cause false "peaks" in the validity function curve. The Xie-Beni index only measures compact and separate clusters, therefore the reliability of the approach may be improved by considering better validity measures.

References

Bezdek, J. C., (1982), *Pattern recognition with fuzzy objective function algorithms*, Plenum Press, New York.

Bezdek, J. C. and Pal, N. R. (1998), "Some New Indexes of Cluster Validity," *IEEE Trans. on System, Man, and Cybernetics, Part B*, vol. 28, no.3, pp. 301-315.

Barni, M., Cappellini V., and Mecocci, A. (1996), "Comments on 'A Possibilistic Approach to Clustering'," *IEEE Trans. Fuzzy Systems*, vol. 4, pp. 393-396.

Dave, R. N. and Krishnapuram, R. (1997), "Robust clustering methods: a unified view," *IEEE Trans. Fuzzy Systems,* vol. 5 no. 2, pp.270-293.

Fisher, R. (1936), "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol.7, no. 2, pp. 179-188.

Jain, A. K. and Dubes, R. C. (1988), *Algorithms for clustering data*, Prentice Hall, New Jersey.

Krishnapuram, R. and Keller, J. (1993), "A possibilistic approach to clustering," *IEEE Trans. Fuzzy Systems, vol. 1,* pp. 98-110.

Krishnapuram, R. and Keller, J. (1996), "The Possibilistic C-Means algorithm: Insights and recommendations*". IEEE Trans. Fuzzy Systems*, vol. 4, pp. 385-393.

Merz, C. J. and Murphy, P. M. (1996), UCI Repository of Machine Learning Databases, http://www.ics.uci.edu/~mlearn/MLRepository.html, University of California, Dept. of Information and computer Science.

Pal, N. R., Pal, K. and Bezdek J. C. (1997), "A mixed c-means clustering model," in *Proc. of the Sixth IEEE International Conference on Fuzzy Systems*, vol. 1, pp. 11-21.

Shapira, O. and Wachs, J. P. (2004) "A Graphical method based on the Xie-Beni Validity index to improve the 'Possibilistic C-Means with Repulsion' Algorithm" to appear in *IMS 2004*, Sakarya, Turkey.

Timm, H., Borgelt, C. and Kruse, R. (2001), "Fuzzy cluster analysis with cluster repulsion," in *Proc. of the European Symposium on Intelligent Technologies, Hybrid Systems and their implementation on Smart Adaptive Systems*, Tenerife, Spain.

Wachs, J., Stern, H. and Edan, Y. (2003) "Parameter Search for an Image Processing Fuzzy C-Means Hand Gesture Recognition System" in *Proc. of IEEE International Conference on Image Processing*, vol. 3, Barcelona, Spain, pp. 341-345.

Zadeh, L. A. (1965), "Fuzzy Sets," *Information and Control*, vol. 8, pp. 338-353.