A Hierarchical Possibilistic Clustering

Mehdi Salkhordeh Haghighi[1], Hadi Sadoghi Yazdi[2], Abedin Vahedian[3]

1

Abstract: **In this paper we propose to combine two clustering approaches, namely fuzzy and possibilistic c-means. While fuzzy c-means algorithm finds suitable clusters for groups of data points, obtained memberships of data, however, encounters a major deficiency caused by misinterpretation of membership values of data points. Therefore, membership values cannot correctly interpret compatibility or degree to which data points belong to clusters. As a result, noisy data will be misinterpreted by incorrect memberships assigned, as sum of memberships of each noisy data to all clusters is constrained to be equal to 1. To overcome this, a possibilistic approach has been proposed which removes this constraint. It has, however, caused another shortcoming as cluster centers converge to an identical point. Therefore, possibilities cannot correctly interpret the degrees of compatibilities. To correct this problem, a number of works have been carried out which all try to change possibilistic objective function proposed by Krishnapuram and James M. Keller.**

In this work, a hierarchical approach has been proposed based on properties of both fuzzy and possibilistic approaches to overcome this deficiency. Sensitivities of both methods have been studied together with analyzing results obtained by both methods. Superiority of the proposed method as opposed to conventional possibilistic c-means is shown to be conspicuous.

***Index Terms*—Hierarchical clustering, possibilistic, fuzzy c-means, sensitivity analysis**

I. Introduction

Clustering can be considered as the most important unsupervised learning algorithm; As any other problem of this kind, it deals with finding a structure in a collection of unlabeled data. Data may be images, patterns, words, documents and so on. Many applications use clustering, such as:

- Data mining: DNA analysis.
- Text mining: text type clustering, document classification; clustering weblog data to discover groups of similar access patterns.
- Information retrieval: document clustering, image and video retrieval.
- Marketing: finding groups of customers with similar behavior given a large database of customer's data containing their properties and past buying records.
- Biology: classification of plants and animals given their features.
- Libraries: book ordering.
- Insurance: identifying groups of motor insurance policy holders with a high average claim cost; identifying frauds.
- City-planning: identifying groups of houses according to their type, cost and geographical location;
- Earthquake studies: clustering observed earthquake epicenters to identify dangerous zones; and
- Statistical computational linguistic, corpus-based computational lexicography.

Fuzzy c-means (FCM) is a method of clustering allowing one segment of data to belong to two or more clusters. This method (developed by Joe Dunn in 1973 [1] and improved by Jim Bezdek in 1981 [2]) is frequently used in pattern recognition. FCM and possibilistic version of c-means are reviewed in the following sub-sections.

A. Fuzzy clustering algorithm

Clustering, as a basic approach, on some unbalanced data set $X = \{x_1, x_2, \ldots, x_n\}$ is partitioning x into c subsets such that 1 < c < n. Each point in x is a vector in n-dimensional space. In most of the clustering methods, each data point belongs to at most one cluster. The primary goal of these clustering methods is to determine to which cluster, each data point belongs. One of the algorithms proposed by J. B. McQueen in 1967 was k-means. K-means is a crisp algorithm, meaning that each data point belongs to at most one cluster. We define c-partition of x as a $c \times n$ matrix representing memberships of each data point to all clusters. We show the matrix as $U_{c*n} = \{u_{i,j}\}$, i=1..c , j=1..n. In k-means algorithm, U is defined by equation (1).

$$U_{i,j} = \{u_{i,j} \mid u_{i,j} = 0 \text{ or } u_{i,j} = 1, \forall i = 1{:}c, j = 1{:}n\} \qquad (1)$$

In many situations, it is difficult to determine to which cluster a data point belongs to exactly. In 1973 a fuzzy version of k-means named FCM was proposed by Joe Dunn in 1973 [1] and improved by Jim Bezdek in 1981 [2]. In FCM, each data point belongs to all clusters based on a membership value proportion to Euclidian distance of data point xj from cluster center βi.. In FCM, matrix U is defined by equation (2). In equation (2), Ui,j is membership of data point xj to cluster j.

$$U_{i,j} = \{u_{i,j} \mid 0 \le u_{i,j} \le 1 \ , \ \textstyle\sum_{i=1}^{c} u_{i,j} = 1\} \qquad (2)$$

FCM objective function, which should be minimized, is shown in equation (3).

$$J_{FCM} = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{i,j}^{m} \, d_{i,j}^{2} \qquad (3)$$

$$d_{i,j}^{2} = (x_j - \beta_i)^T A_i (x_j - \beta_i) \qquad (4)$$

[1] Mehdi Salkhordeh Haghighi is with the Computer Department, Ferdowsi University of Mashhad, Iran, (haghighi@ieee.org)
[2] Hadi Sadoghi Yazdi is with the Computer Department,Ferdowsi University of Mashhad, Iran, (sadoghi@sttu.ac.ir)
[3]Abedin Vahedian is with the Computer Department, Ferdowsi University of Mashhad, Iran,(vahedian@um.ac.ir)

Another constraint considered for FCM is $\sum_{j=1}^{n} u_{i,j}^{m} > 0$. This constraint ensures that no cluster is empty. In equation 3, $d_{i,j}^2$ is squared distance of data point j from cluster center i shown in equation 4, m determines degree of fuzziness that usually has the condition m $\in$ [1, ∞]. As a probabilistic method, sum of memberships of each data point to all clusters is equal to 1. Because of this constraint, it cannot properly describe degree of compatibility or possibility of data points to clusters. Moreover, the algorithm has considerable troubles in noisy environments.

B. Introduction to possibilistic clustering approaches

In 1993 Raghu Krishnapuram and James M. Keller [3] proposed a possibilistic clustering approach named PCM to overcome the limitations of FCM method. Instead of probabilistic memberships, the resulting partition of data can be interpreted as a possibilistic partition, and each membership value may be interpreted as possibility or degree of compatibility. Equation (5) represents definition of matrix U based on possibilities.

$$U_{i,j} = \{u_{i,j} \mid 0 \le u_{i,j} \le 1,\ \forall i = 1{:}c, j = 1{:}n;\ \forall i\, \exists j \mid u_{i,j} > 0\} \quad (5)$$

By removing restriction $\sum_{i=1}^{c} u_{i,j} = 1$ in (2) to indicate the degree of compatibility, a trivial solution would be 0 i.e., the criterion function is minimized by assigning all memberships to 0. We would clearly prefer that the memberships for representative data points to be as high as possible, while unrepresentative points should have low memberships in all clusters. The objective function (which satisfies the requirements) obtained by adding a penalty term to the objective function of FCM in equation (3). The objective function proposed by Krishnapuram and Keller [3] is shown in equation (6). By minimizing the objective function, update formulas for ui,j , βi (center for cluster i) are indicated in equations (7,8,9).

$$J = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{i,j}^{m}\ d_{i,j}^2 + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{n} (1-u_{i,j})^m \quad (6)$$

$$u_{i,j} = \frac{1}{1 + \left(\frac{d_{i,j}^2}{\eta_i}\right)^{\frac{1}{m-1}}} \quad (7)$$

$$\beta_i = \frac{\sum_{j=1}^{n} u_{i,j}^{m}\, x_j}{\sum_{j=1}^{n} u_{i,j}^{m}} \quad (8)$$

$$\eta_i = K \frac{\sum_{j=1}^{n} u_{i,j}^{m} d_{i,j}^2}{\sum_{j=1}^{n} u_{i,j}^{m}} \quad (9)$$

In equation (9), care should be taken about choosing K. A discussion on choosing value of K together with sensitivity of FCM to these values is presented later.

The PCM algorithm proposed in [3] solved a number of problems of FCM it, however, caused some other problems. In 1996, Raghu Krishnapuram and James M. Keller [12] explained analytically one shortcoming of PCM reported by Barni, Cappellini, and Mecocci [4], that it exhibited an undesirable tendency to converge to coincidental clusters.

In 2001 Heiko Timm, Christian Borgelt, Christian Doring, and Rudolf Kruse [7] pointed out another issuee in PCM method, that objective function of PCM is minimized only if all cluster centers are identical. Nevertheless, if there was a single optimum location for cluster centers, all cluster centers would be attracted to the point. Therefore, by assigning random values to membership matrix U, the algorithm tried to move centers to the optimum point. To overcome the problem, they added a repulsion term to the objective function of PCM that would be small if clusters were far from each other. Then the attraction of the data points can compensate the repulsion if only the clusters are sufficiently spread out. The repulsion term is shown in equation (10).

$$\gamma \sum_{i=1}^{c} \sum_{k=1, k\neq i}^{c} \frac{1}{d^2(c_i, c_k)} \quad (10)$$

The objective function proposed by Heiko Timm [7] is indicated in equation (11).

$$J = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{i,j}^{m}\ d_{i,j}^2 + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{n} (1-u_{i,j})^m + \gamma \sum_{i=1}^{c} \sum_{k=1,k\neq i}^{c} \frac{1}{d(\beta_i, \beta_k)^2} \quad (11)$$

Where βi , βk are centers of clusters i and k and γ is a weighting factor. Also, $\sum_{j=1}^{n} u_{i,j} > 0\ \forall i \in \{1..c\}$. Minimization of equation (10) with respect to cluster prototypes leads to an iterative computing shown in equation (12).

$$\beta_i = \frac{\sum_{j=1}^{n} u_{i,j} x_j - \gamma \sum_{k=1,k\neq i}^{c} \beta_k \frac{1}{||\beta_k - \beta_i||^2}}{\sum_{j=1}^{n} u_{i,j} - \gamma \sum_{k=1,k\neq i}^{c} \frac{1}{||\beta_k - \beta_i||^2}} \quad (12)$$

Where, βi is new center for cluster i. Iteration continues while $\left|\beta_i^{(new)} - \beta_i^{(old)}\right| < \varepsilon$. For βi on the right side of the above equation, values of previous iteration are used. Update formula 11 shows the effect of the repulsion between clusters. A cluster is attracted by the data assigned to it and repelled by the other clusters.

In 2002, Timm and Kruse [8] solved the deficiency of PCM by adding a term to objective function of PCM proportion to inverse distance functions between cluster centers. In 2004, Jiang-She Zhang and Yiu-Wing Leung [13] changed the possibilistic approach given in [3] to overcome its drawbacks. The main idea was to integrate a fuzzy approach into their objective function, so that the improved algorithm could determine proper clusters via fuzzy method while it can achieve robustness through possibilistic approach. Their proposed approach defined two types of membership. First, a possibilistic membership which measured absolute degree of typicality of a point in any particular cluster. Second, a fuzzy membership that measured the relative degree of sharing of a point amongst all clusters. To integrate these two approaches, they interpret equation (6) as squared distance; they, then applied fuzzy clustering based on equation (6) to define their objective function shown in equation (13).

$$J = \sum_{i=1}^{C}\sum_{j=1}^{n} \left(u_{i,j}^{(f)}\right)^{m_f} \left((u_{i,j}^{(p)})^{m_p} d_{i,j}^2 + \eta_i (1-u_{i,j}^{(p)})^{m_p}\right. \quad (13)$$

In equation (13), $u^{f}_{i,j}$ and $u^{p}_{i,j}$ are fuzzy and possibilistic memberships of xj to cluster i respectively. They also proposed another equation for ηi to incorporate both fuzzy and possibilistic memberships, as represented in equation (14).

Update formula for $u^{p}_{i,j}$ is the same as equation (7) while update formula for $u^{f}_{i,j}$ is according to equation (15).

$$\eta_i = \frac{\sum_{j=1}^{n} (u_{i,j}^{(p)})^{m_p} (u_{i,j}^{(f)})^{m_f} d_{i,j}^2}{\sum_{j=i}^{n} (u_{i,j}^{(p)})^{m_p} (u_{i,j}^{(f)})^{m_f}} \qquad (14)$$

$$u_{i,j}^{(f)} = \frac{1}{\sum_{k=1}^{C} \frac{(u_{i,j}^{(p)})^{(m_p-1)/2} d_{i,k}}{(u_{k,j}^{(p)})^{(m_p-1)/2} d_{k,j}})^{\frac{2}{(m_f-1)}}} \qquad (15)$$

Cluster center β_i is also computed by equation (16).

$$\beta_i = \frac{\sum_{j=1}^{n} (u_{i,j}^{(p)})^{m_p} (u_{i,j}^{(f)})^{m_f} x_j}{\sum_{j=1}^{n} (u_{i,j}^{(p)})^{m_p} (u_{i,j}^{(f)})^{m_f}} \qquad (16)$$

As it appears from equations (13-16), the only difference of the method with PCM is using a two part ui,j.

In 2005, Nikhil R. Pal, Kuhu Pal, James M. Keller, and James C. Bezdek [11] proposed another PCM method. They believed that memberships (or relative typicality) and possibilities (or absolute typicality) were both important for correct interpretation of data substructures. Therefore, they retained the constraint $\sum_{i=1}^{c} u_{i,j} = 1$ to show memberships while relaxed the constraint row sum = 1 on the typicality values by adding a term $b\,t_{i,k}^{n}$ to objective function where $t_{i,k} \leq 1$. This leads to objective function (17):

$$J = \sum_{i=1}^{c}\sum_{j=1}^{n} (a u_{ij}^{m} + b t_{i,j}^{\eta})\, d_{i,j}^2 + \sum_{i=1}^{c} \gamma_i \sum_{j=1}^{n} (1 - t_{i,j})^{\eta} \qquad (17)$$

Subject to the constraints $\sum_{i=1}^{c} u_{i,j} = 1 \;\forall\, k$, $0 \leq u_{i,j}$, $t_{i,j} \leq 1$. Also $a > 0$, $b > 0$, $m > 1$, $\eta > 1$. In equation (17), $\gamma i > 0$ is an arbitrary constant. The constants a and b define the relative importance of fuzzy membership and typicality values in the objective function. Note that in (8), ui,j has the same meaning of membership as that in FCM. Similarly, ti,j has the same interpretation of typicality as in PCM. By minimizing objective function 17, they found equations (18,19,20) for updating ui,j , ti,j,, βi.

$$u_{i,j} = \left(\sum_{k=1}^{c} \left(\frac{d_{i,j}}{d_{k,j}}\right)^{2/(m-1)}\right)^{-1} \qquad (18)$$

$$1 \leq i \leq c \;,\; 1 \leq k \leq n$$

$$t_{i,j} = \frac{1}{1 + \left(\frac{b}{\gamma_i} d_{i,j}^2\right)^{1/(\eta-1)}} \qquad (19)$$

$$1 \leq i \leq c \;,\; 1 \leq k \leq n$$

$$\beta_i = \frac{\sum_{j=1}^{n} (a u_{i,j}^{m} + b t_{i,j}^{\eta}) x_j}{\sum_{j=1}^{n} (a u_{i,j}^{m} + b t_{i,j}^{\eta})} \qquad (20)$$

$$1 \leq i \leq c$$

C. Structure of paper

Aforementioned works have been carried out to overcome weaknesses of possibilistic approach proposed in [3]. They all attempt to change objective function so that ui,j correctly describe degree of compatibility of each data point to all clusters. In our research, we intensively evaluate the behavior of PCM method proposed in [3]. We also examine sensitivity of the PCM to its parameters (e.g. K) by running the algorithm over some known sample data sets such as Iris, Wine, Glass, Breast, Wpbc. We next present a new approach to overcome the drawback of PCM which moves cluster centers to an identical point while incorrectly finds degrees of compatibilities in some situations. Finally, we analyze results of our new method on the same data sets. Dominant notes of the proposed approach are:

- Evaluating behavior of PCM algorithm
- Simplicity of the proposed method
- Having FCM ability as a pre-clustering tool
- Tuning of membership to form degree of compatibility using PCM procedure
- Acceptable results over standard data sets.

The remaining of this paper is as follows: In section 2, we evaluate thoroughly the PCM approach and examine the weaknesses of the approach both analytically and experimentally. In section 3, our new approach (HPCM) is presented to overcome the limitations of PCM together with results of the new method analyzed on some known sample data sets. In this section, we also compare the results obtained by performing HPCM and PCM algorithms on Iris, Wine, Glass, Breast, Wpbc sample data sets and analyze sensitivity of the two algorithms on some parameters. Finally, in section 4 a conclusion about the results we achieved in this work is presented.

II. ANALYZING PCM METHOD

We first analyze the PCM method proposed by Raghu Krishnapuram and James M. Keller to analytically explain the key problem of the method which converges to local minimum so that all cluster centers will be identical at the end of the algorithm, resulting in misinterpretation of the degrees of compatibilities. Then, the sensitivity of the method to parameter K is examined. This analysis also includes running the algorithm on some known sample data sets to highlight the mentioned problems.

A. Issues with PCM

As mentioned earlier, evaluating objective function (6) reveals that, it is minimized only if all cluster centers are very close or even the same. The reason is that the equation (7) for updating membership degree of a data point to a cluster depends only on distance of the data point to the cluster, while it does not depend on the distance to other clusters.

Nevertheless, if there is a single optimum location for cluster centers, all cluster centers will be attracted to the point. So, by assigning random values to membership matrix U, the algorithm attempts to move centers to the optimum point. This is resulted because the algorithm is trapped in local minimum of the objective function.

B. Analytical description of PCM problem

It is possible to analytically explain why cluster centers gradually approach to a single optimum point in PCM. Using equation (8) for updating cluster centers, if we run the algorithm many times, for each cluster center βi we compute expectation or ensemble average of βi by equation (16) to indicate where the cluster center converges to.

If we run the algorithm with valuable deterministic data, for a two cluster case, possibility of data point j (xj) to first cluster is α1,j and possibility of xj to second cluster is α2,j . Since $0 < \alpha_{1,j} < 1$, $0 < \alpha_{2,j} < 1$, we conclude that $0 < \alpha_{1,j} + \alpha_{2,j} < 2$.

On the other hand, there is no other constraint for α1, j, α2, j. It is also known that for valuable amounts of data samples, ensemble average of α1, j where $0 < \alpha_{1,j} < 1$ is $\frac{1}{2}$. This is also true for α2,j if $0 < \alpha_{2,j} < 1$. Obviously the ensemble average of $\alpha_{1,j} + \alpha_{2,j}$ where $0 < \alpha_{1,j} + \alpha_{2,j} < 2$ is also 1. Therefore, we conclude that $\{\alpha_{1,j}\} = \frac{1}{2}$, $E\{\alpha_{2,j}\} = \frac{1}{2}$. Finally, because there is no constraint on ui,j except $0 < u_{i,j} < 1$, ensemble average of ui,j is equation (21).

$$E\{u_{i,j}^m\} = \frac{1}{2} \qquad (21)$$

From equation (21), equation (22) is concluded:

$$\frac{\sum_{j=1}^{n} u_{i,j}^m}{n} = \frac{1}{2} \Rightarrow \sum u_{i,j}^m = \frac{n}{2} \qquad (22)$$

Since the set of all data points (x_j) are deterministic, we compute expectation of βi as equation (23).

$$E\{\beta_i\} = E\left\{\frac{\sum_{j=1}^{n} u_{i,j}^m x_j}{\sum_{j=1}^{n} u_{i,j}^m}\right\} = E\left\{\frac{\sum_{j=1}^{n} u_{i,j}^m x_j}{\frac{n}{2}}\right\} = \frac{2}{n} E\{\sum_{j=1}^{n} u_{i,j}^m x_j\} \qquad (23)$$

Simplifying equation (23) yields equation (24):

$$E\{\beta_i\} = E\{x_j\} = \theta \qquad (24)$$

where, θ is a constant vector. Therefore, ensemble average of cluster centers converges to θ when PCM algorithm runs many times. This is true for all clusters since xi for i=1...n are deterministic values for all clusters. It, however, does not happen for FCM because it has another constraint on ui,j, therefore, we cannot conclude equations (22-24).

As a result, all cluster centers gradually converge to an identical point. This property of PCM causes to misinterpret degrees of compatibilities of data points to clusters. On the other hand, the algorithm is sensitive to initial values of U. If initial values of U are not properly assigned, the algorithm rapidly converges to a local minimum of its objective function so that degrees of compatibilities are computed based on identical cluster centers.

We have tested the PCM algorithm on different known sample data sets. Results of executing the algorithm on Iris, Glass, Breast, Wine, Wpbc data sets are shown in table 1. The algorithm was executed on these sample data sets for 10 times. Then, average distance between cluster centers was measured as a criterion to show how far cluster centers are located from each other. Table 1 indicates the results as average distance between clusters in each execution. Since the algorithm is sensitive to parameter K, as shown in equation (9), in table 1, K is assumed to be 1.

$$\frac{\sum_{i=1}^{C-1} \sum_{j=i+1}^{C} d_{i,j}}{\frac{C!}{2 \times (C-2)!}} \qquad (25)$$

Since the major problem of PCM is that the cluster centers converge to an identical point, it appears a good idea to measure average distances between each two cluster centers as a metric for evaluating effectiveness of HPCM vs. PCM to find cluster centers as far as possible to maximize between class parameter. To do this, two values are needed. First, sum of all distances between each two center ($\sum d_{i,j}$, $i = 1:C-1, j = i+1:C$) where, di,j is Euclidean distance between centers i,j. Second, total number of paired centers ($\frac{C!}{2 \times (C-2)!}$) where, C is the number of clusters. Finally, equation 25 gives total average distances between centers as a metric for comparing HPCM vs. PCM. Equation 25 is a metric defined to show efficiency of PCM vs. HPCM for finding cluster centers. In Table 1, each row indicates results of computing equation (25) for each sample data set, 10 times. By evaluating results shown in Table 1, it is evident that cluster centers in all examples are very close to each other, and almost identical. This undesired effect, which is due to minimizing the objective function shown in equation (6), causes misinterpretation of compatibilities of data points to clusters. In Table 1, average distances between cluster centers for wine and Wpbc data sets seem to be more than in other data sets. This is because scale factors of data in Wine and Wpbc data sets are different from the data in other data sets. As a result, cluster centers gradually converge to each other in PCM as seen in equations (21-24). The values obtained in Table 1 represent this effect.

Table 1. Average distance between cluster centers.

execution	Glass PCM	Glass HPCM	Iris PCM	Iris HPCM	Wine PCM	Wine HPCM	Breast PCM	Breast HPCM	Wpbc PCM	Wpbc HPCM
1	0.0013	0.3892	0.0046	2.7943	4.0920	359.9852	0.0018	0.4136	8.2634	627.3964
2	0.0008	0.3892	0.0396	2.7943	5.6096	359.9852	0.0017	0.4136	7.3064	627.3964
3	0.0009	0.3892	0.0149	2.7943	4.3165	359.9852	0.0018	0.4136	5.2621	627.3964
4	0.0011	0.3892	0.0105	2.7943	4.2276	359.9852	0.0035	0.4136	2.0706	627.3964
5	0.0004	0.3892	0.0162	2.7943	7.3930	359.9852	0.0009	0.4136	4.0297	627.3964

III. THE PROPOSED HIERARCHICAL CLUSTERING METHOD

By analyzing many different samples and recalling equations (20-24), it is obvious that, if based on some criteria, initial values of U are selected correctly, the PCM algorithm works well. By randomly initializing membership values, cluster centers converge to some identical point to minimize equation (6). This effect is shown in Table 1 by executing PCM algorithm on some known data sets. The effect of convergence of cluster centers to an identical point will not happen for FCM method because it has another limitation presented by equation (2). Therefore, this property of FCM can be used to overcome the problem mentioned, as convergence of cluster centers to an identical point, in PCM. On the other hand, fuzzy clustering is sensitive to noise because misinterprets membership values of noise points. In our new method, we introduce a hierarchical approach for finding more accurate centers for clusters with more precise values for U so that these values can better describe compatibilities of data points to clusters with less sensitivity to noise. Another key element in our approach is to determine value of K based on some criteria. In some cases, the PCM approach is highly sensitive to K. We first analyze the sensitivity on some known data sets to describe what causes the sensitivity. We then compare sensitivity of our new approach to show its lower sensitivity to the value of K.

A. A new possibilistic approach

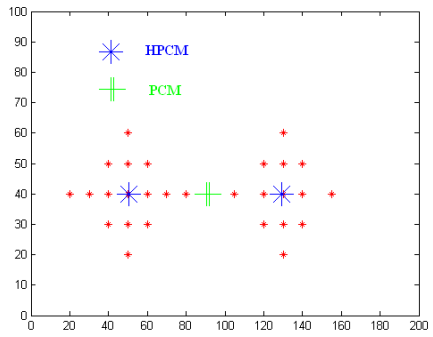
The proposed approach for possibilistic clustering uses properties of both fuzzy and possibilistic clustering to solve the problem described earlier. Fuzzy clustering has one unique constraint which enables us to overcome the limitation of PCM. Equation (2) shows that sum of all membership values of each data point to all clusters is equal to 1. This constraint prohibits cluster centers to converge to an identical point. On the other hand, fuzzy clustering finds cluster centers near optimal in many cases. However, values of U are not properly defined. In our new approach, however, we hierarchically use fuzzy followed by possibilistic algorithms to first, find initial values of U based on standard fuzzy algorithm, while finding cluster centers near optimal point. In this way, the fuzzy algorithm finds cluster centers near optimal, but membership values may describe data points incorrectly, while noise points are misinterpreted. To describe correct memberships as degrees of compatibilities while adjusting cluster centers together with describing noise points correctly, we run possibilistic algorithm. We initialize U with values computed in previous level by fuzzy algorithm. In other ways, initial values of U for PCM, are those ones computed by FCM in previous step. This hierarchical approach, then finds values of U so that they correctly describe degrees of compatibilities, while keeping cluster centers as separate as possible. In the next level of the hierarchical approach, where PCM adjusts degrees of compatibilities, the number of iterations needed to do this is very little compared with execution of PCM or FCM separately.

B. Analysis of HPCM vs. PCM

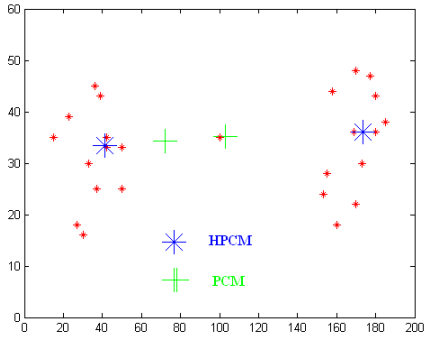
Our test results together with analysis of the results are based on the sample data sets listed in Table 1 and Table 2. In Table 2 we compare results obtained by PCM vs. HPCM, i.e. results obtained by our hierarchical approach. To compare results obtained by the methods, we used average Euclidean distance between clusters based on equation (24).

According to these tables, in all sample data sets, cluster centers found by PCM are very close together, while cluster centers found by HPCM are not only far from each other, but also they are identical in each execution, resulting in stability of the new algorithm. Therefore, initial value of U has no major effect on positions of cluster centers and possibilities. Execution of HPCM and PCM on the sample data sets tested by [3] revealed that PCM was highly sensitive to initial values of U. In each execution, therefore, PCM finds different centers for clusters, while HPCM is not sensitive to initial values of U. As a result, it finds identical cluster centers in each execution. Fig.1 shows execution of the two algorithms on two sample data sets presented in [3].

As shown in Fig. 1(a), centers found by HPCM are completely separate while position of each one is in center of each cluster. But, centers found by PCM are so close together that cannot be separated, so, misinterpret possibilities. The same is true for Fig. 1(b).



(a) First sample



(b) Second sample

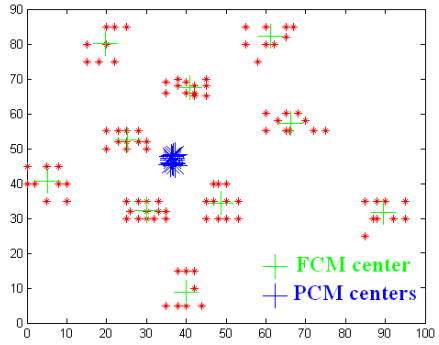
Fig. 1. Comparing centers found by PCM and HPCM.

Fig. 2 indicates another example of PCM problem based on Solomon's C101 data set. Fig. 2.a also shows cluster centers found by FCM algorithm together with cluster centers found by PCM algorithm presented by [3]. It is obvious that FCM found correct cluster centers while, PCM makes cluster centers to converge to an identical point. Fig. 2(b) depicts the result on running our new method. It is clear that cluster centers are adjusted by our new algorithm with respect to FCM.

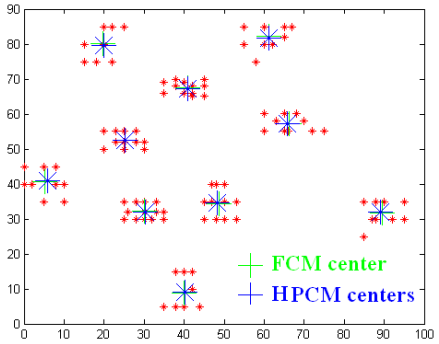
C. Analyzing sensitivity of PCM vs. HPCM with respect to K

Another important issue about PCM approach is that it is highly sensitive to K. Value of K In equation (9) together with initial value of U can determine degree of convergence of the objective function. More ever, some groups of data are more sensitive to K than others. We have tested sensitivity of PCM vs. HPCM with respect to K on some known data sets. Fig. 3 shows execution of possibilistic and our new method on Iris, Glass, Breast, Wine, and Wpbc data sets. Vertical axis in Fig. 3 represents average percent of correct classification of data points in 10 executions of the two algorithms. One of the major points that Fig. 3 indicates is that average percent of correct classification of our new algorithm in all sample data sets for almost all values of K is significantly higher than the ones in PCM algorithm. Therefore, our new algorithm assigns data points to clusters more accurately than standard possibilistic algorithm.

On the other hand, from sensitivity point of view, Fig. 3 indicates another fact: the figure compares sensitivity of five sample data points with respect to K for both algorithms. Analysis of Fig. 2 and Table 2 reveals that in our new method, one can find a relation between size of clusters and sensitivity to K. If data sets have special conditions, sensitivity of data sets to K alleviates. According to the results obtained in Table 2, it is obvious that if sizes of clusters for a data set are close together, sensitivity of the data set to K is low. For data sets where sizes of some clusters are very different from the others, sensitivity to K increases. Low sensitivity of Iris and Wine data sets is because sizes of clusters in these data sets are very close together or even identical.

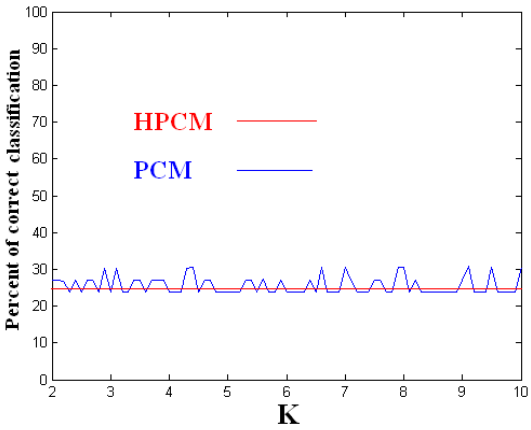


(a) execution of FCM vs. PCM

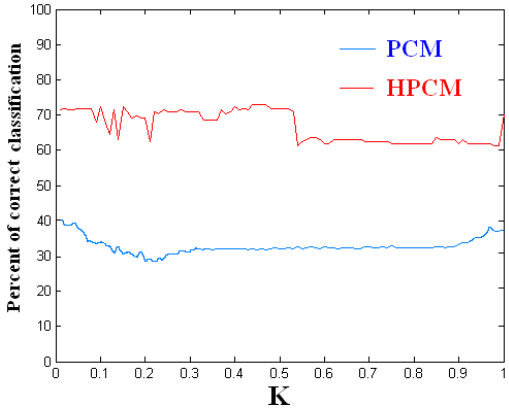


(b) execution of FCM vs. HPCM

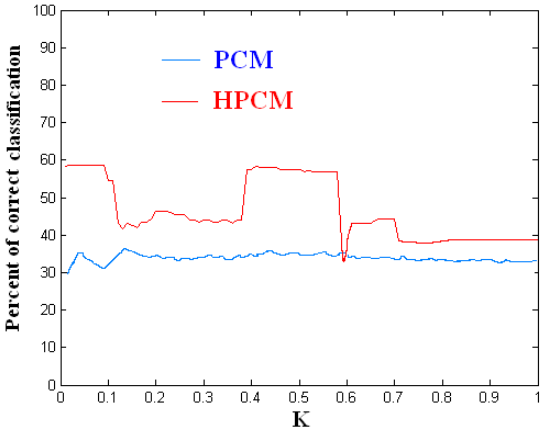
Fig. 2. Solomon's C101 data set



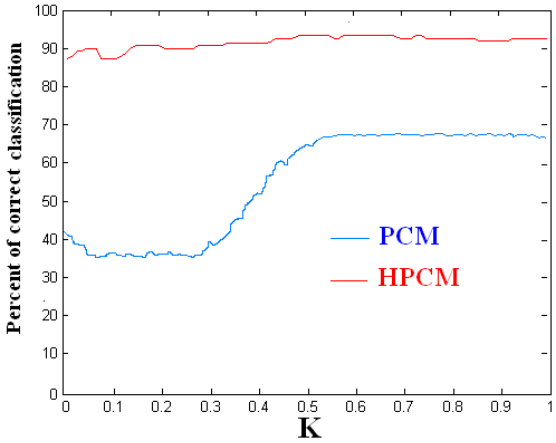
(a) comparison for Wpbc data set.



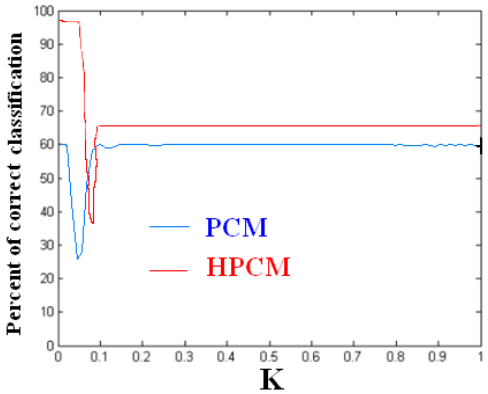
(b) comparison for Wine data set.



(c) comparison for Glass data set.



(d) comparison for Iris data set.



(e) Comparison for Breast data set.

Fig. 3. Comparing correct classification rate with respect to K.

Table 2. Sensitivity of PCM and HPCM algorithms with respect to K.

Data sets	#of data	#of clusters	PCM Sensitivity to K	HPCM sensitivity to K	#of samples in clusters
Glass	214	6	Low	High	76,70,29,17,13,9
Iris	150	3	High	Low	50,50,50
Wine	178	3	Low	Low	71,59,48
Breast	699	2	High	High	458,241
Wpbc	198	2	High	High	151,47

IV. CONCLUSION

Use In this paper, we made an investigation on PCM approach presented by Krishnapuram and James M. Keller in 1993 to overcome weaknesses of FCM method. Our analysis was driven from different aspects. One aspect was the limitation in their method misinterpreting the membership values while pushing cluster centers to an identical point. The other aspect was the sensitivity of their method to the value of K as different values of K might affect the results based on various conditions of input data. On the other hand; we presented a hierarchical approach to overcome the problems. Our method used FCM and PCM methods hierarchically so that membership values interpret data points correctly while, will not permit cluster centers to converge to an identical point. Our method, HPCM, was compared with PCM on different known data sets. The results indicate that the method is more accurate than PCM, while does not presents weaknesses of the previous method.

REFERENCES

[1] J. C. Dunn ,"A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters," Journal of Cybernetics 3: 32-57, (1973).

[2] J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algoritms," Plenum Press, New York, (1981).

[3] R. Krishnapuram and J. Keller, "A possibilistic approach to clustering," IEEE Trans. Fuzzy Syst., vol. 1, no. 2, pp. 98–110, Apr. 1993.

[4] M. Barni, V. Cappellini, and A. Mecocci, "A possibilistic approach to clustering," IEEE Trans. Fuzzy Syst., vol. 4, no. 3, pp. 393–396, Jun.1996.

[5] R. Krishnapuram and J. Keller, "The possibilistic c-Means algorithm:Insights and recommendations," IEEE Trans. Fuzzy Syst., vol. 4, no. 3,pp. 385–393, Jun. 1996.

[6] N. R. Pal, K. Pal, and J. C. Bezdek, "A mixed c-means clustering model,"in IEEE Int. Conf. Fuzzy Systems, Spain, 1997, pp. 11–21.

[7] H. Timm, C. Borgelt, C. Doring, and R. Kruse, "Fuzzy cluster analysis with cluster repulsion," presented at the Euro. Symp. Intelligent Technologies (EUNITE), Tenerife, Spain, 2001.

[8] H. Timm and R. Kruse, "A modification to improve possibilistic fuzzy cluster analysis," presented at the IEEE Int. Conf. Fuzzy Systems, FUZZ-IEEE' 2002, Honolulu, HI, 2002.

[9] H. Timm, C. Borgelt, C. Doring, and R. Kruse, "An extension to possibilistic fuzzy cluster analysis," Fuzzy Sets Syst., vol. 2004, pp. 3–16.

[10] E. E. Gustafson andW. C. Kessel, "Fuzzy clustering with a fuzzy covariance matrix," in Proc. IEEE Conf. Decision and Control, San Diego, CA,1979, pp. 761–766.

[11] Nikhil R. Pal, Kuhu Pal, James M. Keller, and James C. Bezdek, "A Possibilistic Fuzzy c-Means Clustering Algorithm", ieee transaction on fuzzy systems, vol. 13, no. 4, august 2005.

[12] Raghu Krishnapuram and James M. Keller," The Possibilistic C-Means Algorithm: Insights and Recommendations", ieee transaction on fuzzy systems, vol. 4, no. 3, august 1996.

[13] Jiang-She Zhang and Yiu-Wing Leung," Improved Possibilistic C-Means Clustering Algorithms" ieee transaction on fuzzy systems, vol. 23, no. 2, august 2004.