

Research Article

A Weight Possibilistic Fuzzy C-Means Clustering Algorithm

Jiashun Chen ¹, Hao Zhang,² Dechang Pi,³ Mehmed Kantardzic,⁴ Qi Yin,¹ and Xin Liu¹

¹School of Computer Engineering, Jiangsu Ocean University, Lianyungang, Jiangsu 222003, China

²School of Mathematics and Information Engineering, Lianyungang Normal College, Lianyungang, Jiangsu 222003, China

³College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics Nanjing, Jiangsu 210016, China

⁴J. B Speed School of Engineering, University of Louisville, KY 40208, USA

Correspondence should be addressed to Jiashun Chen; 1976139@163.com

Received 31 March 2021; Accepted 1 June 2021; Published 11 June 2021

Academic Editor: Cristian Mateos

Copyright © 2021 Jiashun Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Fuzzy C-means (FCM) is an important clustering algorithm with broad applications such as retail market data analysis, network monitoring, web usage mining, and stock market prediction. Especially, parameters in FCM have influence on clustering results. However, a lot of FCM algorithm did not solve the problem, that is, how to set parameters. In this study, we present a kind of method for computing parameters values according to role of parameters in the clustering process. New parameters are assigned to membership and typicality so as to modify objective function, on the basis of which Lagrange equation is constructed and iterative equation of membership is acquired, so does the typicality and center equation. At last, a new possibilistic fuzzy C-means based on the weight parameter algorithm (WPFCM) was proposed. In order to test the efficiency of the algorithm, some experiments on different datasets are conducted to compare WPFCM with FCM, possibilistic C-means (PCM), and possibilistic fuzzy C-means (PFCM). Experimental results show that iterative times of WPFCM are less than FCM about 25% and PFCM about 65% on dataset X_{12} . Resubstitution errors of WPFCM are less than FCM about 19% and PCM about 74% and PFCM about 10% on the IRIS dataset.

1. Introduction

Clustering is a method of unsupervised learning and had been applied in various fields, including data mining, pattern recognition, computer vision, and bioinformatics. Cluster methods might be summarized as follows: partition-based [1, 2], hierarchical-based [3], density-based [4–6], and grid-based [7]. Partition methods included hard partition [8, 9] and soft partition [10–12]. Soft partition is represented by using fuzzy membership, and membership value lies in interval [0, 1]. Many fuzzy clustering algorithms had been developed and widely used in a variety of areas [13–16], such as data mining and pattern recognition. Ruspini [17] regarded fuzzy C-means (FCM) as a clustering algorithm, and DUNN [18] analyzed fuzzy exponent m and determined that the value of m was equal to 2. Bezdek generalized fuzzy exponent $m > 1$. The constraint of FCM ($\sum_{i=1}^c u_{ij} = 1$) might cause that membership conflicted with intuitive belonging

degree; furthermore, it made the clustering results be sensitive to noise. In order to overcome this defect, Krishnapuram and Keller [19] relaxed the constraint and proposed a new algorithm named possibilistic C-means (PCM) [20] which reduced influence of noise on clustering and had good robustness. However, PCM relied on initialization condition and might produce coincident clusters [21]. Many algorithms were developed to overcome coincident problem. For example, studies [22, 23] modified the PCM objective function adding an inverse function of the distances between cluster centers to resolve coincident problem. The study [23] proposed a new model named the fuzzy possibilistic C-means (FPCM) which introduced membership and typicality value t_{ij} subjected to $\sum_{i=1}^c t_{ij} = 1$ for unlabeled data. FPCM reduced sensitivity to noise in FCM and resolved coincident problem in PCM; however, typicality value became very small as dataset scale increasing for the reason of row sum constraints.

The study by Pal [24] proposed new algorithms named possibilistic fuzzy C-means (PFCM), which was a hybridization of PCM and FCM and overcame problems of PCM, FCM, and FPCM. PFCM solves the noise sensitivity. So, PFCM has been widely applied in many fields [25–27] and solved some problems well. PFCM added coefficients a and b for membership and possibility, which measured the relative importance in the computation of centroids; however, the values of a and b were simply fix at 1, which meant that membership and possibility had the same importance during the process of computing centroids. This setting made clustering results become less evident in some clustering. PFCM did not give a scientific and rational method to compute parameters. The main objective of this study was to generalize FCM, PCM, and PFCM algorithms and propose a new algorithm named weight possibilistic fuzzy C-means (WPFCM). We designed a new objective function on basis of PFCM. According to the requirement of minimizing objective function, iterative functions of membership, typicality, and centroid were obtained by constructing Lagrange function and deriving its derivation.

This study is quite different from literature [28]. First, this study focussed on clustering on possibilistic fuzzy C-means, while the study by Schneider [28] aimed to the possibilistic C-means algorithm. The research algorithms were different. Second, the key point was that designing weight parameter was diverse. Algorithm in this study could allocated weight value to samples inlier and outlier automatically according to the calculation method of weight parameters, which made membership value maximizing and outlier reduce influence of estimation. Weight parameter could satisfy the optimization objective function, make it iterate faster, and avoid the coincident problem. The method in the study by Schneider [28] could not have these advantages.

Experiments on different datasets show that new algorithm not only makes clustering results obvious but also partitions overlapping data better and also reduces iterative times and speeds up convergence. The rest of this study is organized as follows: Section 2 reviews the FCM, PCM, FPCM, and PFCM clustering algorithm. Section 3 provides a new method on the computation of parameters and WPFCM was proposed. Section 4 experimentally demonstrates the improvement of performance of WPFCM on some UCI database. Section 5 offers conclusion.

2. Related Works

Since fuzzy set theory was introduced by Zadeh, this method was applied in the clustering algorithm rapidly. FCM is one of the most famous algorithms and obtains clustering results by minimizing objective function and iterating membership and centroid. The objective function of FCM is designed as follows:

$$J_{\text{FCM}}(u, v) = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m D_{ij}^2, \quad m > 1, \quad (1)$$

where fuzzy exponent m is subjected to $m > 1$ and Euclidean distance is defined as $D_{ij}^2 = \|x_j - v_i\|^2$. Membership can be obtained by minimizing objective function (1). The following equations are iterative functions of membership and centroid.

$$u_{ij} = \left[\sum_{k=1}^c \left(\frac{D(x_j, v_i)}{D(x_j, v_k)} \right)^{(2/m-1)} \right]^{-1}, \quad (2)$$

$$v_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m}.$$

The clustering performance is better; however, the algorithm subjects to the following three constraints: $u_{ij} \in [0, 1]$, $(i = 1, 2, \dots, c, j = 1, 2, \dots, n)$, $\sum_{i=1}^c u_{ij} = 1$, $(j = 1, 2, \dots, n)$, and $0 < \sum_{i=1}^c u_{ij} < n$, $(j = 1, 2, \dots, n)$, which make the algorithm be sensitive to noise and usually lead to center deviation for individual anomalous data points.

The constraints of FCM require data points to consider the relation to other points in current cluster and in other clusters; therefore, membership might conflict with intuitive belonging degree and does not directly reflect real clustering results. The FCM algorithm is sensitive to noise and obtains poor clustering results in noisy data environment. Krishnapuram and Keller [19] improved FCM and proposed the possibilistic C-means algorithm which relaxed constraint. Objective function is designed as follows:

$$J_{\text{PCM}}(t, v) = \sum_{i=1}^c \sum_{j=1}^n t_{ij}^q D_{ij}^2 + \sum_{i=1}^c \eta_i \sum_{j=1}^n (1 - t_{ij})^q, \quad q > 1, \quad (3)$$

where η_i is the scaling parameter of the i^{th} class and defined $\eta_i = K (\sum_{j=1}^n u_{ij}^q D_{ij}^2 / \sum_{j=1}^n u_{ij}^q)$ (common $K = 1$), and exponent q subjects to constraint $q > 1$, and Euclidean distance is defined $D_{ij}^2 = \|x_j - v_i\|^2$. Iterative functions of typicality and centroid are obtained by minimizing objective function (3). Equations (4) and (5) are the iterative functions.

$$t_{ij} = \frac{1}{\left(1 + (D^2(x_j, v_i) / \eta_i)^{(1/q-1)} \right)}, \quad (4)$$

where $0 < i < c$, $0 < j < n$,

$$v_i = \frac{\sum_{j=1}^n t_{ij}^q x_j}{\sum_{j=1}^n t_{ij}^q}. \quad (5)$$

In equation (3), t_{ij} is not the membership but possibility, and clustering results are easy to interpret. PCM [29] relaxes the constraint $\sum_{i=1}^c u_{ij} = 1$, $(j = 1, 2, \dots, n)$ and defines the constraint $\sum_{i=1}^c t_{ij} \leq c$, $(j = 1, 2, \dots, n)$, so the rows and columns are independent, and data structure becomes loose. Therefore, the algorithm is insensitive to noise and could deal with the dataset including outlier; on the other side, there is another weakness. Experiments show that PCM's clustering results depend on initialization and generate the coincident problem. Pal [30] held that clustering centroids

were closed to data centers due to the effect of membership. Pal proposed a new algorithm FPCM on the basis of FCM and PCM. FPCM used data center as the clustering center. It is feasible to a great extent. Membership is a good method when data points need to be marked clearly because it is natural to assign a point to cluster whose prototype is the nearest to the point, while possibility is important to estimate clustering centers, and effectively reduces the influence that was brought by abnormal data point. The objective function was designed as follows:

$$J_{\text{FPCM}}(u, t, v) = \sum_{i=1}^c \sum_{j=1}^n (u_{ij}^m + t_{ij}^q) D_{ij}^2, \quad (6)$$

where membership subjects to constraint $\sum_{i=1}^c u_{ij} = 1$, ($j = 1, \dots, n$), and the typicality subjects to constraint $\sum_{i=1}^n t_{ij} = 1$ ($i = 1, \dots, c$), and other constraints are $m > 1$, $q > 1$, and $0 < u_{ij}, t_{ij} < 1$, and Euclidean distance was defined $D_{ij}^2 = \|x_j - v_i\|^2$. Iterative function of membership and typicality and prototype can be obtained by minimizing objective function. The equations (7)–(9) are the iterative functions, respectively.

$$u_{ij} = \left[\sum_{i=1}^c \left(\frac{D(x_j, v_i)}{D(x_j, v_k)} \right)^{(2/m-1)} \right]^{-1}, \quad (7)$$

$$t_{ij} = \left[\sum_{i=1}^c \left(\frac{D(x_j, v_i)}{D(x_k, v_i)} \right)^{(2/q-1)} \right]^{-1}, \quad (8)$$

$$v_i = \frac{\sum_{j=1}^n (u_{ij}^m + t_{ij}^q) x_j}{\sum_{j=1}^n (u_{ij}^m + t_{ij}^q)}. \quad (9)$$

Although FPCM overcomes weakness of PCM and FCM, the typicality value becomes very small with sample data increase. Typicality value is limited by FPCM. On a large sample dataset, the typicality value is inconsistent with real value due to constraint row sum. Pal [24] improved algorithm FPCM relaxed the typicality constraint row sum, retained membership constraint column sum, and proposed a new algorithm named possibilistic fuzzy C-means (PFCM). The objective function is designed as follows:

$$J_{\text{PFCM}}(t, u, v) = \sum_{i=1}^c \sum_{j=1}^n (au_{ij}^m + bt_{ij}^q) D_{ij}^2 + \sum_{i=1}^c \sum_{j=1}^n (1 - t_{ij})^q, \quad (10)$$

where parameters subject to constraints $m > 1$, $q > 1$, $\eta > 0$, $0 < u_{ij}$, $t_{ij} < 1$, $1 \leq i \leq c$, and $1 \leq j \leq n$. Euclidean distance is defined $D_{ij}^2 = \|x_j - v_i\|^2$, and parameters a and b are the constants. Iterative function of membership and typicality and prototype can be obtained by minimizing objective function. The equations (11)–(13) are iterative functions.

$$u_{ij} = \left[\sum_{k=1}^c \left(\frac{D(x_j, v_i)}{D(x_j, v_k)} \right)^{(2/m-1)} \right]^{-1}, \quad (11)$$

$$t_{ij} = \frac{1}{\left(1 + (bD^2(x_j, v_i)/\eta_i)^{(1/q-1)} \right)}, \quad (12)$$

$$v_i = \frac{\sum_{j=1}^n (au_{ij}^m + bt_{ij}^q) x_j}{\sum_{j=1}^n (au_{ij}^m + bt_{ij}^q)}, \quad (13)$$

where η_i is defined $\eta_i = K (\sum_{j=1}^n u_{ij}^m D_{ij}^2 / \sum_{j=1}^n u_{ij}^m)$, and usually, K is a constant ($K = 1$).

3. WPFCM Algorithm

This section includes three paragraphs. The motivation of weight parameters was first introduced, and then, the calculation method of weight parameters is presented in the second part, and the last part gives objection function and the steps of algorithm.

3.1. Motivation Weight Parameters. PFCM integrates merit of PCM with FCM, which includes membership and typicality. PFCM reduces the sensitivity to noise in FCM and overcomes coincident problem in PCM and can deal with the problem that typicality value becomes very small with data points increasing in FPCM. After analyzing the parameters a and b , we found that the values of a and b have influence on membership and typicality and then affect the clustering results. If parameter a is greater than b , prototype is affected more by membership than by typicality; on the contrary, if parameter b is higher than a , prototype is affected more by typicality than by membership. Therefore, if we want to reduce influence of clustering results caused by outlier, we should select values of a lower than b . How to determine values of parameters is difficult. Usually, values of parameters are all fixed at 1, which means that membership and typicality have the same importance to clustering results. At this time, setting two parameters a and b becomes meaningless. In many situations, we do not know whether values fit to parameters, and then, it depends on experience to determine values of parameters a and b . Assigning values of parameters a and b lacks mathematical basis, so it is occasional and unscientific in PFCM. The clustering results become unstable. There is another weakness in PFCM that all vector data share the same value of parameters in the cluster process; however, different vectors have various importance for clustering. So it is unreasonable for parameters a and b to be fixed at 1. In order to overcome these weaknesses, we proposed a new method to compute values of weight parameters which replace parameters a and b in PFCM. New parameters consider the importance of each

sample data in the process of clustering. New calculation method is more reasonable. The importance of parameter lies in the fact that values of a and b directly affect typicality value t_{ij} and centroid value v_i , affect membership indirectly, and then influence on the clustering results.

3.2. Calculation Method of Weight Parameters. Many literatures have mentioned methods for calculating weight parameters [31–33]. The study by Fan et al. [32] assigned weights to properties according to the importance of each property to the cluster process. For example, in dataset IRIS [34], the third property and the fourth property are beneficial to get obvious clustering results, so they are assigned a high weight value and others are assigned a low weight value. The premise is that we must know which property is important and unimportant. To an unknown dataset, this method is inappropriate and cannot be applied. The study by Nock and Nielsen [33] estimated all samples' probability density by using the analogy method. This method needs a great deal of computation. The study by Hung [29] gave a prototype-driven learning of parameter α which is based on exponential separation strength between clusters and updated each iteration to improve the performance of FCM. Equation (14) is the definition of parameter of α .

$$\alpha = \exp\left(-\min_{i \neq j} \left(\frac{\|v_i - v_k\|^2}{\beta}\right)\right), \quad (14)$$

where parameter β is defined as the distance from data point x_j to sample mean. Parameter β is defined as follows:

$$\beta = \frac{\sum_{j=1}^n \|x_j - \bar{x}\|^2}{n}, \quad (15)$$

and \bar{x} can be defined as sample mean:

$$\bar{x} = \frac{\sum_{j=1}^n x_j}{n}. \quad (16)$$

Definition 1. A given sample set to be classified is denoted by X ($X = \{x_1, x_2, \dots, x_n\} \subset F(X)$), and X is partitioned c ($0 < c < n$) fuzzy subsets, and c is the number of clustering.

Definition 2. According to the importance of the data point x_j ($x_j \in X$) during the clustering process, weight parameter can be defined as γ_{ij} , which is the weight of x_j w. r. t. class i . The following equation is the calculation method.

$$\gamma_{ij} = \exp\left(\frac{\sum_{j=1}^n \|x_j - \bar{x}\|^2 \times c}{-\|x_j - v_i\|^2 \times n}\right). \quad (17)$$

Theorem 1. Distance from x_j to center v_i can be regarded as weight; if the distance is long, then the value of weight will be high; on the contrary, if the distance is short, then the value of weight will be low.

Proof. \bar{x} is the mean value of sample, and the difference between x_j and \bar{x} is reflected by the distance from data point x_j to \bar{x} , which is constant. The smaller the value of D_{ij}^2 , the shorter the distance from x_j to class i . We can deduce that the larger the value of $-\|x_j - v_i\|^2$, the smaller the value of $(\sum_{j=1}^n \|x_j - \bar{x}\|^2 \times c / -\|x_j - v_i\|^2 \times n)$, and value of $\exp(-\|x_j - v_i\|^2 \times n / \sum_{j=1}^n \|x_j - \bar{x}\|^2 \times c)$ will be small; on the contrary, the long the distance from x_j to class i , the larger the value of $\exp(-\|x_j - v_i\|^2 \times n / \sum_{j=1}^n \|x_j - \bar{x}\|^2 \times c)$. Optimization of objective function requires a minimum value. Weight parameter should satisfy the optimization objective; so long distance should get large γ_{ij} . It is appropriate to use γ_{ij} as the weight parameter. \square

3.3. Design Objective Function. According to the rule of classification, there is little difference among all data in the same class and great difference in different classes. During the process of designing objective function, in order to assign data point, the nearest distance from data point to center should be selected, which is denoted by the maximizing membership value. The typicality value can be used to reduce influence of estimation caused by outlier. New objective function should meet two requirements: on the one hand, role of membership in objective function should be increased when the sample is inlier; on the other hand, the role of the typicality value should be increased in objective function when the sample is outlier. Therefore, the objective function is designed as in the following equation, which include two parts: the first part is fuzzy function denoted by fuzzy weight parameter and the second part is typicality function denoted the by typicality weight parameter.

Definition 3. The new objective function is designed as following which is based on FCM, PCM, and PFCM.

$$J_{\text{WPFM}}(U, T, V) = \sum_{j=1}^n \sum_{i=1}^c [(1 - \gamma_{ij})u_{ij}^m + \gamma_{ij}t_{ij}^q] D_{ij}^2 + \sum_{i=1}^c \sum_{j=1}^n (1 - t_{ij})^q, \quad (18)$$

where γ_{ij} ($0 < \gamma_{ij} < 1$) denotes the weight between data point x_j and class i , which comes from equation (17). Different data points have various weight values, and then, clustering results are more reasonable by using different weight parameters and avoid the coincident problem. U , T , and V denote membership matrix ($c \times n$), typicality matrix ($c \times n$), and centroid matrix ($c \times 1$), respectively. Here, u_{ij} ($0 < u_{ij} < 1$) is the membership of feature point x_j in cluster c_i and t_{ij} ($0 < t_{ij} < 1$) is the typicality of x_j in cluster c_i . $D_{ij}^2 = \|x_j - v_i\|^2$ is the Euclidean distance between data point x_j and v_i . The parameters m ($m > 1$) and q ($q > 1$) are the fuzzy exponents. The parameter η_i ($\eta_i > 0$) is a constant, which is defined by $\eta_i = K (\sum_{j=1}^n u_{ij}^m D_{ij}^2(x_j, v_i)) / \sum_{j=1}^n u_{ij}^m$, where K usually is fixed at 1.

According to the analysis of the preceding context, we know that the nearer the distance between data point x_j and cluster c_i , the smaller the value of weight parameter γ_{ij} . The

distance from data point x_j to cluster c_i is near, which shows that data point x_j belongs to the i^{th} cluster. The weight parameter value of membership should be increased and be set as $(1-\gamma_{ij})$. On the contrary, the further the distance between x_j and the i^{th} cluster, the greater the difference between x_j and the i^{th} cluster, and x_j may be an anomalous point. Typicality weight parameter should be increased to reduce the effect of x_j on clustering. The typicality weight parameter is set as γ_{ij} . With increase (decrease) of weight parameter of membership, typicality weight parameter will be decreased (increased). The weight parameter is calculated on the basis of different sample date points, which overcomes the unreasonable value of a and b in PFCM and resolves coincident problem which is caused by small value of a and poor initialization centroid.

According to Definition 3, the Lagrangian multiplier method was used to construct the Lagrange equation. In order to minimize equation (18), the partial derivatives of u_{ij} and t_{ij} were computed according to constraints $\sum_{i=1}^c u_{ij} = 1$ and $\sum_{j=1}^n t_{ij} = 1$ and acquired the membership u_{ij} and typicality t_{ij} and centroid v_i as follows:

$$u_{ij} = \left[\sum_{i=1}^c \left(\frac{D(x_j, v_i)}{D(x_j, v_k)} \right)^{(2/m-1)} \right]^{-1}, \quad (19)$$

$$t_{ij} = \frac{1}{\left(1 + \left((\gamma_{ij}/\eta_i) D^2(x_j, v_i) \right)^{(1/q-1)} \right)}, \quad (20)$$

$$v_i = \frac{\sum_{j=1}^n \left[(1-\gamma_{ij}) u_{ij}^m + \gamma_{ij} t_{ij}^q \right] x_j}{\sum_{j=1}^n \left[(1-\gamma_{ij}) u_{ij}^m + \gamma_{ij} t_{ij}^q \right]}. \quad (21)$$

3.4. WPFCM Algorithm. According to the objective function, steps of algorithms were provided as following:

4. Experiments

In order to validate the algorithm efficiency, some experiments on different datasets were carried out. The initial value of parameters is set as follows: $\varepsilon = 0.000001$, the maximum iterative times $\text{max_iter} = 100$, constant $K = 1$, the number of class $\text{Cluster_n} = 2$ for dataset X_{12} , and the number of class $\text{Cluster_n} = 3$ for dataset IRIS [34].

Experiment 1. Dataset X_{12} [35]; algorithm: FCM, PCM, PFCM, and WPFCM; initialization:

$$V_0 = \begin{bmatrix} 0.2722 & 0.0153 \\ 0.1988 & 0.7468 \end{bmatrix}, \quad (22)$$

X_{12} is a two-dimensional dataset with 12 data points. The coordinates of X_{12} are given in Table 1. Figure 1 shows the coordinate distribution of dataset X_{12} . There are ten points forming two clusters with five points each on the left and the right sides of the axis y . Data points x_6 and x_{12} are considered as noise, and each has the same distance to two clusters.

TABLE 1: Coordinate values of dataset X_{12} .

Dataset	x	Y
x_1	-5.00	0.00
x_2	-3.34	1.67
x_3	-3.34	0.00
x_4	-3.34	-1.67
x_5	-1.67	0.00
x_6	0.00	0.00
x_7	1.67	0.00
x_8	3.34	1.67
x_9	3.34	0.00
x_{10}	3.34	-1.67
x_{11}	5.00	0.00
x_{12}	0.00	10.00

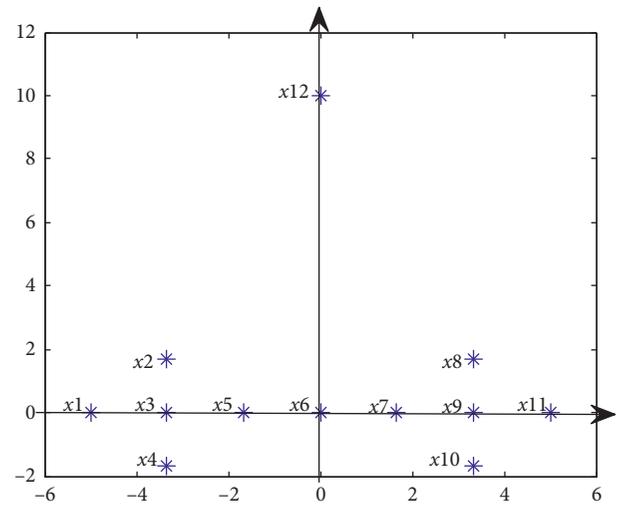


FIGURE 1: Coordinate distribution of dataset X_{12} .

Table 2 presents centroids which are generated by running FCM, PFCM, and WPFCM on X_{12} . Suppose distance equation $\text{Dist}_X = \|V_{X_{12}} - V_X\|^2$, which denotes the distance from real centroid to centroid V_X generated by algorithms. The following are the centroids:

$$\begin{aligned} V_{\text{FCM}} &= \begin{bmatrix} 2.9854 & 0.5435 \\ -2.9853 & 0.5435 \end{bmatrix}, \\ V_{\text{PFCM}} &= \begin{bmatrix} 2.8367 & 0.3643 \\ -2.8367 & 0.3643 \end{bmatrix}, \\ V_{\text{WPFCM}} &= \begin{bmatrix} 3.0059 & 0.2052 \\ -3.0059 & 0.2052 \end{bmatrix}. \end{aligned} \quad (23)$$

The real centroid of X_{12} is $V_{X_{12}} = \begin{bmatrix} 3.34 & 0 \\ -3.34 & 0 \end{bmatrix}$. We compute distance of each algorithm by using distance equation as follows: $\text{Dist}_{\text{FCM}} = 0.4212$, $\text{Dist}_{\text{PFCM}} = 0.3860$, and $\text{Dist}_{\text{WPFCM}} = 0.1537$. Comparing three distances, although each algorithm can get good result, $\text{Dist}_{\text{WPFCM}}$ has the minimum distance, that is, to say V_{WPFCM} is nearer to real centroid than other V_X . V_{WPFCM} reflects real cluster center better.

TABLE 2: Prototypes of FCM, PFCM, and WPFCM on X_{12} .

FCM ($m=2$)	PFCM ($m=2, q=2, a=1, b=1$)	WPFCM ($m=1.5, q=5$)
$\begin{bmatrix} 2.9854 & 0.5435 \\ -2.9853 & 0.5435 \end{bmatrix}$	$\begin{bmatrix} 2.8367 & 0.3643 \\ -2.8367 & 0.3643 \end{bmatrix}$	$\begin{bmatrix} 3.0059 & 0.2052 \\ -3.0059 & 0.2052 \end{bmatrix}$

TABLE 3: Iterative times of FCM, PFCM, and WPFCM on X_{12} .

Algorithm	FCM ($m=2$)	PFCM ($m=2, q=2, a=1, b=1$)	WPFCM ($m=1.5, q=5$)
Iterative times	12	26	9

Table 3 provides the minimum iteration times of FCM, PFCM, and WPFCM with optimal given parameters. Iteration times of WPFCM are slight less than FCM and far less than PFCM. Therefore, WPFCM has less running time in large datasets and has a high speed of convergence.

Table 4 presents the membership value by running FCM, PFCM, and WPFCM. By comparison, membership values of WPFCM are better than the other two algorithms; especially for data points x_3 and x_9 , the membership values are equal to one. Data points x_3 and x_9 are the center of two clusters, which show that WPFCM is easier to recognize the cluster center. Membership value cannot tell noisy data point x_6 and x_{12} , but noisy data are identified by using the typicality value in Table 5. By analyzing data in Table 5, typicality values of WPFCM are greater than PFCM. If one data point has larger typicality value, the data point is more likely to belong to the cluster. One of typicality values of x_3 and x_9 are up to 1 in Table 5, which show data point, respectively, belongs to two clusters with large possibility.

Membership values of noisy data x_6 and x_{12} are equal to 0.5. Figure 1 shows that distance from x_6 to two clustering centers is far less than x_{12} , but Table 4 cannot show this difference. Table 5 shows that typicality values of ten data points are greater than 0.9 except for x_6 and x_{12} . Typicality values of x_6 and x_{12} are far less than others, so we consider the data points x_6 and x_{12} are noise. We also find the typicality value of x_{12} is far less than x_6 in Table 5, which shows noisy data x_{12} belong to two clusters with less possibility than x_6 and which reflects distribution of x_6 and x_{12} in Figure 1. WPFCM improves the defect of FCM. From Table 5, we also find the typicality value of WPFCM is better than PFCM, so WPFCM can get more obvious clustering results.

Table 6 presents centroids and iteration times by running WPFCM on dataset X_{12} with different parameters. Clustering results of WPFCM are better than FCM and PFCM as a whole. Iteration times of WPFCM are a bit less than FCM and PFCM. When the value of m keeps unchanged and value of q varied from 2 to 5, there is an increasing tendency of membership values, but not obvious; however, typicality values have an evident decreasing tendency. Clustering centers also have great changes, and the values are increasing and nearer to real centroid. Iteration times are decreasing. With increase of q , the influence of weight parameter γ_{ij} on clustering results is increasing. Weight parameter γ_{ij} is generated in iterative procedure, and initial centroid is generated randomly, so WPFCM overcomes the defect of

TABLE 4: Membership of FCM, PFCM, and WPFCM on X_{12} .

Dataset	FCM ($m=2$)		PFCM ($m=2, q=2, a=1, b=1$)		WPFCM ($m=1.5, q=5$)	
	U_1	U_2	U_1	U_2	U_1	U_2
x_1	0.0636	0.9364	0.0725	0.9275	0.0039	0.9961
x_2	0.0327	0.9673	0.0468	0.9532	0.0028	0.9972
x_3	0.0103	0.9897	0.0100	0.9900	0.0000	1.0000
x_4	0.1006	0.8994	0.0941	0.9059	0.0068	0.9932
x_5	0.0844	0.9156	0.0681	0.9319	0.0069	0.9931
x_6	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
x_7	0.9156	0.0844	0.9319	0.0681	0.9931	0.0069
x_8	0.9673	0.0327	0.9532	0.0468	0.9972	0.0028
x_9	0.9897	0.0103	0.9900	0.0100	1.0000	0.0000
x_{10}	0.8994	0.1006	0.9059	0.0941	0.9932	0.0068
x_{11}	0.9364	0.0636	0.9275	0.0725	0.9961	0.0039
x_{12}	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000

TABLE 5: Typicality values of PFCM and WPFCM on X_{12} .

Dataset	FCM ($m=2$)		PFCM ($m=2, q=2, a=1, b=1$)		WPFCM ($m=1.5, q=5$)	
	T_1	T_2	T_1	T_2	T_1	T_2
x_1	—	—	0.1142	0.6225	0.4160	0.9271
x_2	—	—	0.1660	0.8021	0.4598	0.9891
x_3	—	—	0.1717	0.9536	0.4659	1.0000
x_4	—	—	0.1580	0.6437	0.4561	0.9438
x_5	—	—	0.2796	0.8416	0.5527	0.9961
x_6	—	—	0.4924	0.4924	0.7383	0.7383
x_7	—	—	0.8416	0.2796	0.9961	0.5527
x_8	—	—	0.8021	0.1660	0.9891	0.4598
x_9	—	—	0.9536	0.1717	1.0000	0.4659
x_{10}	—	—	0.6437	0.1580	0.9438	0.4561
x_{11}	—	—	0.6225	0.1142	0.9271	0.4160
x_{12}	—	—	0.0729	0.0729	0.3730	0.3730

random selection a and b and improves the vulnerability of uncertain clustering results. Clustering results in Table 7 are better than in Table 6; however, we find that membership values greatly reduce but iteration times increase. Comprehensive consideration suggests that the value of m and q are 1.5 and 5, respectively.

Experiment 2. Dataset: IRIS; algorithm: FCM, PCM, PFCM, and WPFCM

TABLE 6: Results generated by WPFCM for different values of the parameters with X_{12} .

Dataset	$m = 1.5, q = 2$	$m = 1.5, q = 3$	$m = 1.5, q = 4$	$m = 1.5, q = 5$
Centroids	$\begin{bmatrix} 2.7487 & 0.2067 \\ -2.7487 & 0.2067 \end{bmatrix}$	$\begin{bmatrix} 2.8711 & 0.2161 \\ -2.8711 & 0.2161 \end{bmatrix}$	$\begin{bmatrix} 2.9570 & 0.2108 \\ -2.9570 & 0.2108 \end{bmatrix}$	$\begin{bmatrix} 3.0059 & 0.2052 \\ -3.0059 & 0.2052 \end{bmatrix}$
Iteration times	12	12	11	9

TABLE 7: Results generated by WPFCM for different values of the parameters with X_{12} .

Dataset	$m = 2, q = 2$	$m = 2, q = 3$	$m = 2, q = 4$	$m = 2, q = 5$
Centroids	$\begin{bmatrix} 2.7917 & 0.1784 \\ -2.7917 & 0.1784 \end{bmatrix}$	$\begin{bmatrix} 2.9167 & 0.1880 \\ -2.9167 & 0.1880 \end{bmatrix}$	$\begin{bmatrix} 3.0108 & 0.1823 \\ -3.0108 & 0.1823 \end{bmatrix}$	$\begin{bmatrix} 3.0652 & 0.1761 \\ -3.0652 & 0.1761 \end{bmatrix}$
Iteration times	12	13	12	11

TABLE 8: Centroids generated by FCM, PCM, PFCM, and WPFCM for different initial centroids with IRIS.

Initial centroids		FCM ($m = 2$)			PCM ($m = 2, q = 2$)			
0.5841	0.1078	0.9063	5.0036	5.8892	6.7751	5.0038	6.1730	6.1730
0.8797	0.8178	0.2607	3.4030	2.7612	3.0524	3.4113	2.8806	2.8806
0.5944	0.0225	0.4253	1.4850	4.3642	5.6468	1.4668	4.7632	4.7633
0.3127	0.1615	0.1788	0.2515	1.3974	2.0536	0.2375	1.6046	1.6047

TABLE 9: Centroids generated by FCM, PCM, PFCM, and WPFCM for different initial centroids with IRIS.

Initial centroids		PFCM ($m = 2, q = 2, a = 1, b = 1$)			WPFCM ($m = 1.5, q = 1.5$)			
0.5841	0.1078	0.9063	5.0034	5.9253	6.6157	5.0091	5.8912	6.5713
0.8797	0.8178	0.2607	3.3966	2.7901	3.0121	3.4092	2.7589	3.0028
0.5944	0.0225	0.4253	1.4869	4.4012	5.4510	1.4851	4.2622	5.5790
0.3127	0.1615	0.1788	0.2500	1.4094	1.9860	0.2532	1.3174	2.0347

TABLE 10: Distance values between different centroids generated by FCM, PCM, PFCM, and WPFCM with IRIS.

Dist _{ij}	FCM ($m = 2$)	PCM ($m = 2, q = 2$)	PFCM ($m = 2, q = 2, a = 1, b = 1$)	WPFCM ($m = 1.5, q = 1.5$)	Real center
Dist ₁₂	10.7991	14.3839	11.055	10.0458	10.2938
Dist ₁₃	23.8293	14.3848	21.4751	22.5394	22.5954
Dist ₂₃	2.9453	2.0000e-008	1.9605	2.7705	2.6166

IRIS is a four-dimensional dataset including three classes: setosa, versicolor, and virginica. Each cluster has 50 data points, adding up to 150 data points. The first cluster setosa has good separation from the other two clusters without overlapping. There are some overlaps between versicolor and virginica.

Data in Tables 8 and 9 were acquired by running FCM, PCM, PFCM, and WPFCM many times on IRIS. Each algorithm got good clustering centroid. Compared with other algorithms, WPFCM acquired more obvious membership and typicality values and better separation. The two centroids of versicolor and virginica got by running the PCM algorithm almost overlap. It is difficult to find separation between clusters in Tables 8 and 9. In order to compare separation between different classes, we defined the distance between classes as $Dist_{ij} = ||V_i - V_j||^2$, which denotes the distance from i^{th} cluster to the j^{th} cluster.

Table 10 provides distance values between different centroids generated by FCM, PCM, PFCM, and WPFCM with IRIS. $Dist_{12}$ and $Dist_{13}$ that are calculated by using $Dist_{ij} = ||V_i - V_j||^2$ in WPFCM, FCM, and PFCM reflect the fact that setosa separates from the other two classes versicolor and virginica. However, in PCM, $Dist_{12}$ and $Dist_{13}$ have almost identical results and $Dist_{23}$ is nearly zero, so the results do not reflect the features of dataset which is caused by coincidence of PCM. Although FCM, PFCM, and WPFCM all reflect separation of setosa from other two class and overlapping between versicolor and virginica, by comparing $Dist_{23}$, we find that $Dist_{23}$ in WPFCM is the nearest to the real value. We conclude that WPFCM reflects the characteristic of dataset better than other algorithms and easily get good partition especially for class versicolor and class virginica.

Table 11 provides the distance values between different centroids and real centroid generated by FCM, PCM, PFCM,

TABLE 11: Distance between different centroids and real centroid of FCM, PCM, PFCM, and WPFCM on IRIS.

	FCM ($m=2$)	PCM ($m=2, q=2$)	PFCM ($m=2, q=2, a=1, b=1$)	WPFCM ($m=1.5, q=1.5$)
Dist $_{X_1}$	0.0011	1.4262e-004	0.0014	0.0010
Dist $_{X_2}$	0.0186	0.4055	0.0284	0.0016
Dist $_{X_3}$	0.0554	0.9650	0.0140	0.0022
Dist $_X$	0.0750	1.3706	0.0437	0.0049

TABLE 12: Iterative times of four algorithms on IRIS.

	FCM ($m=2$)	PCM ($m=2, q=2$)	PFCM ($m=2, q=2, a=1, b=1$)	WPFCM ($m=1.5, q=1.5$)
Iterative times	23	75	74	27

TABLE 13: Resubstitution errors of four algorithms on IRIS.

	FCM ($m=2$)	PCM ($m=2, q=2$)	PFCM ($m=2, q=2, a=1, b=1$)	WPFCM ($m=1.5, q=1.5$)
U_{eX}	16	50	14	13
T_{eX}	-----	50	13	10

- (1) Initializing parameters m ($m > 1$), q ($q > 1$), and ε, c ($0 < c < 1$), setting the maximum cycle number \max_iter , setting the initial value of cycle number as 1, and randomly generating centroid V_0 .
- (2) Computing distance according to $D_{ij}^2 = |x_j - v_i|^2$
- (3) Computing the weight parameter γ_{ij} and $(1-\gamma_{ij})$ by using equation (18)
- (4) Computing membership value u_{ij} and typicality value t_{ij} by using equations (19) and (20)
- (5) Computing the objective function obj_fcn
- (6) If $|\text{obj_fcn}(i) - \text{obj_fcn}(i-1)| < \varepsilon$ or iterative times are less than \max_iter , then stop
Else $\text{obj_fcn}(i) \rightarrow \text{obj_fcn}(i-1)$
- (7) Computing centroid v_i by using equation (21) and going to step 2

ALGORITHM 1: Weight possibilistic fuzzy c-means clustering algorithm.

and WPFCM. Formula (24) is defined as the sum of distance between centroid acquired from each algorithm and real centroid. Dist_{x_i} represents distance from the i^{th} cluster center to real centroid. Each Dist_{X_i} in WPFCM is less than in other algorithms. Compared value of Dist_X , there is relation $\text{Dist}_{\text{WPFCM}} < \text{Dist}_{\text{PFCM}} < \text{Dist}_{\text{FCM}} < \text{Dist}_{\text{PCM}}$ in Table 11, which shows that there is little difference between centroid of WPFCM and real centroid.

Iterative times generated by FCM, PCM, PFCM, and WPFCM with IRIS are given in Table 12. Iterative times of WPFCM are slightly larger than FCM, but far less than PCM and PFCM. The WPFCM algorithm acquires clustering center quickly and has fast convergence speed.

The number of resubstitution errors from FCM, PCM, PFCM, and WPFCM on dataset IRIS is given in Table 13. Resubstitution errors of WPFCM are slightly less than FCM and PFCM, but far less than PCM, no matter with regard to membership value or typicality value. Table 13 includes two relations: $U_{e\text{WPFCM}} < U_{e\text{PFCM}} < U_{e\text{FCM}} < U_{e\text{PCM}}$ and $T_{e\text{WPFCM}} < T_{e\text{PFCM}} < T_{e\text{FCM}} < T_{e\text{PCM}}$. Resubstitution errors of membership and typicality are up to 50 in PCM, which are far greater than other algorithms. The reason is that the PCM algorithm has clustering consistency issues, and there are overlapping data in versicolor and virginica.

$$\text{Dist}_X = \sum_{i=1}^c \|V_{\text{IRIS}_i} - V_{X_i}\|^2. \quad (24)$$

5. Conclusions

A new possibilistic fuzzy C-mean based on weight parameters was proposed according to the importance of membership and typicality in the clustering process. First, aiming at unreasonable parameters a and b , we designed weight parameter γ_{ij} based on literature [23] and provided the concrete calculation method. Weight parameters $(1-\gamma_{ij})$ and γ_{ij} were assigned to membership and typicality. Objective function (equation (18)) was improved, and then, new algorithm idea (Algorithm 1) was provided. Experiment on different datasets show that the new algorithm has good performance in dealing with noisy data and gets better clustering results. WPFCM resolves the coincident problem and overcomes defect of sensitivity to noisy data. New algorithm discusses the influence on membership values, typicality values, and centroids with different values of exponent parameters m and q . Exponent parameters are determined by comprehensively comparing membership

values, typicality values, and centroids. Experiments compare iterative times in different algorithms. WPFCM has less iterative times and fast convergence speed. Resubstitution errors of WPFCM are near to FCM and PFCM, but far less than PCM. Comprehensively, many performance indexes suggest that WPFCM overcomes weakness of sensitivity of FCM and resolves the coincident problem of PCM and unreasonable weight parameters of PFCM. The next work is to extend the new algorithm to the nonpoint prototype clustering model such as the spherical prototype, the quadric prototype, and the shell prototype.

Data Availability

The data used to support the findings of this study are available at <http://archive.ics.uci.edu/ml/datasets/iris>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by “Petrel Program of Lianyungang Jiangsu Province, China” (KK18088), and “the Program of Science and Technology Associate Chief Engineer of Jiangsu Province of China” (FZ20200458).

References

- [1] S. S. Wang, A. Gittens, and M. W. Mahoney, “Scalable kernel k -means clustering with nystrom approximation: relative-error bounds,” *Journal of Machine Learning Research*, vol. 20, pp. 1–49, 2019.
- [2] K. Zhao, X. Zhao, C. Peng et al., “Partition level multiview subspace clustering,” *Neural Networks*, vol. 122, pp. 279–288, 2020.
- [3] D. Fogli, G. Giovanni, M. Redolfi, and R. Tonoli, “A knowledge-based approach to hierarchical classification: a voting metaphor,” *Expert Systems with Applications*, vol. 161, no. 15, Article ID 113737, 2020.
- [4] Y.-A. Geng, Q. Li, R. Zheng et al., “RECOME: a new density-based clustering algorithm using relative KNN kernel density,” *Information Sciences*, vol. 436–437, pp. 13–30, 2018.
- [5] R. M. Mahboobeh, A. A. Ahmad, B. Nasersharif, and B. Raahemi, “A new density-based subspace selection method using mutual information for high dimensional outlier detection,” *Knowledge-Based Systems*, vol. 216, Article ID 106733, 2021.
- [6] M. Lia, X. H Bi, L. M. Wang, and X. M. Han, “A method of two-stage clustering learning based on improved DBSCAN and density peak algorithm,” *Computer Communications*, vol. 167, pp. 75–84, 2021.
- [7] B. Liu, Y. Y. Xia, and P. S. Yu, *Clustering through Decision Tree Construction*, pp. 20–29, CIKM2000, NewYork, NY, USA, 2000.
- [8] H. Liu, J. Wu, T. Liu, D. Tao, and Y. Fu, “Spectral ensemble clustering via weighted k -means: theoretical and practical evidence,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 5, pp. 1129–1143, 2017.
- [9] N. Gornitz, L. A. Lima, K.-R. Muller, M. Kloft, and S. Nakajima, “Support vector data descriptions and k -means clustering: one class?” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 9, pp. 3994–4006, 2018.
- [10] F. A. T. Carvalho, Y. Lechevallier, and F. M. Melo, “Partitioning fuzzy clustering algorithms based on multiple dissimilarity matrices,” *Fuzzy Sets and Systems*, vol. 215, no. 16, pp. 1–28, 2013.
- [11] F.-Q. Li, S.-L. Wang, and G.-S. Liu, “A Bayesian Possibilistic C-Means clustering approach for cervical cancer screening,” *Information Sciences*, vol. 501, pp. 495–510, 2019.
- [12] H. Yazdani, “Bounded fuzzy possibilistic method,” *Fuzzy Sets and Systems*, vol. 389, no. 15, pp. 51–65, 2020.
- [13] P. D. Pantula, S. S. Miriyala, and K. Mitra, “An Evolutionary Neuro-Fuzzy C-means Clustering Technique,” *Engineering Applications of Artificial Intelligence*, vol. 89, Article ID 103435, 2020.
- [14] A. Singh and A. K. P. Upadhyay, “A novel approach to incorporate local information in possibilistic c-Means algorithm for an optical remote sensing imagery,” *The Egyptian Journal of Remote Sensing and Space Science*, vol. 24, 2020.
- [15] D. S. Mai, L. T. Ngo, L. H. Trinh, and H. Hagrass, “A hybrid interval type-2 semi-supervised possibilistic fuzzy c-means clustering and particle swarm optimization for satellite image analysis,” *Information Sciences*, vol. 548, pp. 398–422, 2021.
- [16] S. Askari, “Fuzzy C-Means clustering algorithm for data with unequal cluster sizes and contaminated with noise and outliers: review and development,” *Expert Systems with Applications*, vol. 165, Article ID 113856, 2021.
- [17] E. R. Ruspini, “A new approach to clustering,” *Information Control*, vol. 15, no. 1, pp. 22–32, 1969.
- [18] J. C. Dunn, “A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters,” *Journal of Cybernetics*, vol. 3, no. 3, pp. 32–57, 1973.
- [19] R. Krishnapuram and J. Keller, “A possibilistic approach to clustering,” *IEEE TFS*, vol. 1, no. 2, pp. 88–110, 1993.
- [20] R. Krishnapuram and J. M. Keller, “The possibilistic C-means algorithm: insights and recommendations,” *IEEE Transactions on Fuzzy Systems*, vol. 4, no. 3, pp. 385–393, 1996.
- [21] M. Barni, V. Cappellini, and A. Mecocci, “A possibilistic approach to clustering,” *IEEE Transactions on Fuzzy Systems*, vol. 4, no. 3, pp. 393–396, 1996.
- [22] H. Timm, C. Borgelt, and R. Kruse, *Fuzzy Cluster Analysis with Cluster Repulsion* EUNITE, Tenerife, Spain, 2001.
- [23] Q. Zhang, L. T. Yang, Z. Chen, and F. Xia, “A high-order possibilistic $\$C\$$ -means algorithm for clustering incomplete multimedia data,” *IEEE Systems Journal*, vol. 11, no. 4, pp. 2160–2169, 2017.
- [24] N. R. Pal, K. H. Pal, and M. K. James, “A possibilistic fuzzy c-means clustering algorithm,” *IEEE TFS*, vol. 13, no. 1, pp. 517–530, 2005.
- [25] C. L. Chowdhary and D. P. Acharjya, “Segmentation of mammograms using a novel intuitionistic possibilistic fuzzy c-mean clustering algorithm,” *Nature Inspired Computing*, vol. 652, pp. 75–82, 2018.
- [26] S. Askari, N. Montazerin, M. H. F. Zarandi, and E. Hakimi, “Generalized entropy based possibilistic fuzzy C-Means for clustering noisy data and its convergence proof,” *Neurocomputing*, vol. 219, pp. 186–202, 2017.
- [27] C. L. Chowdhary, M. Mittal, P. A. Pattanaik, and Z. Marszalek, “An efficient segmentation and classification system in medical images using intuitionist possibilistic fuzzy C-mean clustering and fuzzy SVM algorithm,” *Sensors*, vol. 20, no. 14, 2020.
- [28] A. Schneider, “Weighted possibilistic C-means clustering algorithms,” in *Proceedings of the Ninth IEEE International*

- Conference on Fuzzy Systems*, pp. 176–180, San Antonio, TX, USA, May 2000.
- [29] W. L. Hung, M. Yang, and D. Chen, “Parameter selection for suppressed fuzzy c-means with an application to MRI segmentation,” *Pattern Recognition Letters*, vol. 27, no. 5, pp. 424–438, 2005.
 - [30] N. R. Pal, K. Pal, and J. C. Bezdek, “A mixed c-means clustering model,” in *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems*, pp. 11–21, Barcelona, Spain, July 1997.
 - [31] L. T. Cheng, S. G. Wang, and X. Wei, “New fuzzy c-means clustering model based on the data weighted approach,” *Data & Knowledge Engineering*, vol. 69, pp. 881–900, 2010.
 - [32] J. Fan, M. Han, and J. Wang, “Single point iterative weighted fuzzy C-means clustering algorithm for remote sensing image segmentation,” *Pattern Recognition*, vol. 42, no. 11, pp. 2527–2540, 2009.
 - [33] R. Nock and F. Nielsen, “On weighting clustering,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 8, pp. 1223–1235, 2005.
 - [34] E. Anderson, “The IRIS of gasp peninsula,” *Bulletin of American IRIS Society*, vol. 59, pp. 2–5, 1935.
 - [35] N. R. Pal, K. Pal, and J. C. Bezdek, “A new hybrid c-means clustering model,” in *Proceedings of the 2004 IEEE International Conference on Fuzzy Systems*, pp. 179–184, Piscataway, NY, USA, July 2004.