applied sciences

MDPI

Article

Double-Constraint Fuzzy Clustering Algorithm

Shiyuan Zhu, Yuwei Zhao and Shihong Yue *

School of Electrical Engineering and Automation, Tianjin University, Tianjin 300072, China; shiyuanzhu@tju.edu.cn (S.Z.); zyww@tju.edu.cn (Y.Z.)
* Correspondence: shyue1999@tju.edu.cn

Abstract: Given a set of data objects, the fuzzy c-means (FCM) partitional clustering algorithm is favored due to easy implementation, rapid response, and feasible optimization. However, FCM fails to reflect either the importance degree of the individual data objects or that of the clusters. Numerous variants of FCM have been proposed to address these issues. However, most of them cannot effectively apply the available information on data objects or clusters. In this paper, a double-constraint fuzzy clustering algorithm is proposed to reflect the importance degrees of both individual data objects and clusters. By incorporating double constraints into each data object and cluster, the objective function of FCM is reformulated and its realization equation is mathematically conducted. Consequently, the clustering accuracy of FCM is improved by applying the available information on both data objects and clusters. Especially, the proposed algorithm effectively addresses the limitations inherent in the existing variants of FCM. The experimental results validate the effectiveness, implementation, and robustness of the new fuzzy clustering algorithm.

Keywords: fuzzy clustering; c-means algorithm; data constraint; cluster constraint

check for updates

Citation: Zhu, S.; Zhao, Y.; Yue, S. Double-Constraint Fuzzy Clustering Algorithm. *Appl. Sci.* **2024**, *14*, 1649. https://doi.org/10.3390/app14041649

Academic Editor: Rodolfo Dufo-López

Received: 25 October 2023
Revised: 5 February 2024
Accepted: 7 February 2024
Published: 18 February 2024

1. Introduction

The c-means (CM) algorithm proposed by MacQueen [1] is the most commonly used clustering algorithm over various research fields, but it cannot accurately partition data objects in which the membership to any specific cluster is uncertain [2]. As a general extension of CM, fuzzy clustering has been proposed to address this problem [3]. Most fuzzy clustering algorithms originate from Bezdek's fuzzy c-means (FCM) algorithm [4], which has been successfully applied in numerous applications including image segmentation [5], feature extraction [6], pattern recognition [7,8], etc.

However, the computed membership degrees in FCM are relative numbers. For a given data object, its membership degrees corresponding to all clusters in the fuzzy partition matrix must sum to 1 to avoid a trivial solution. The sum makes FCM noise-sensitive and unsuitable for applications in which membership degrees show a typicality or compatibility of data points with clusters under flexible constraints [9,10]. Consequently, the clustering results of FCM are often inaccurate. To address these problems, various variants of FCM have been developed. Krishnapuran et al. [11] proposed a possibilistic c-means method by giving up relative number constraints in FCM and thus presented the typicality or compatibility from data objects to clusters. Pedrycz et al. [12] assigned an importance degree to individual data points, thereby significantly mitigating the influence of relative numbers in FCM. More recently, numerous FCM-type algorithms have been proposed. Yu [13] proposed a general c-means algorithm by extending the definition of means from statistical analysis. This algorithm generalized most variants of the fuzzy clustering method to a common model. Huang et al. [14] proposed a feature-weighted c-means algorithm to address the difficulty that FCM faces in detecting clusters distributed across various subspaces.

Despite progress, these algorithms mainly focus on finding clusters by the inherent distribution and distance among data objects [15,16], but they cannot provide an assessment

utilizing the information on each data object and cluster. Especially, in various clustering applications, the number of data points in specific cluster is a mandatory constraint that the clustering results of any clustering algorithm must obey. If these constraints cannot be met, the clustering results may be unacceptable [17,18]. To address this problem, Ng et al. [19] proposed a constrained c-means (CCM) algorithm that assigns a fixed number of data points to each cluster, which means that each column in the partition matrix is constrained by an equation. Nevertheless, there are at least three unsolved problems in CCM. First, CCM must apply the solution of the transportation problem as a practical form, but this approach is not feasible for large-scale datasets due to high computational complexity. Secondly, CCM is a CM-type clustering algorithm, rather than a fuzzy clustering algorithm, and thus the issues that can be solved by FCM are not addressed by CCM. Finally, CCM cannot combine the importance degree of both clusters and data objects [20,21]. More recently, efforts have been made to determine the number of clusters in the fuzzy clustering process. There are two approaches to determine the optimal number of clusters. One utilizes one or several clustering indices to determine the optimal number through a trial-and-error method across all possible numbers of clusters [22,23]. The other attempts to solve the number of clusters in the iterative process of fuzzy clustering, such as the proposed Bayesian probabilistic model and inference algorithm for fuzzy clustering [24], which provides expanded capabilities compared to traditional FCM. However, these algorithms fail to address the problem of typicality or compatibility inherent in FCM.

In this study, to address the typicality or compatibility of both data points and clusters, we incorporate dual constraints into each data point and cluster. Therefore, the objective function of FCM is reformulated and then its realization equation is mathematically conducted. The proposed method is easily operated as FCM and requires minimal additional parameters. We discuss the theoretical framework of the algorithm and analyze the clustering effectiveness of representative patterns in each cluster. In the proposed method, we enhance clustering qualities by satisfying the mandatory constraints on data objects and clusters. Our experimental results validate the effectiveness of the proposed algorithm and demonstrate its applicability and limitations.

2. Related Work

Let $X = \{x_i \mid i = 1, 2, \ldots, n\}$ be a dataset with n data objects distributed in c clusters, $x_i \in R^d$ in a d-dimensional data space. Four typical fuzzy clustering algorithms are reviewed as follows.

(1) Bezdek's FCM: The objective function in FCM can be stated as

$$\min J(U,V) = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m} d_{ij}^{2},\ \ s.t.\ \sum_{i=1}^{c} u_{ij} = 1,\ j = 1,2,\ldots,n, 0 < \sum_{j=1}^{n} u_{ij} \le n, \tag{1}$$

where $d_{ij} = ||x_j - v_i||$, v_i is the prototype (center) of ith cluster, u_{ij} the membership degree of jth point to ith cluster, m a fuzziness exponent, ranging in the interval [1, 3]. By the Lagrange multiplier optimization algorithm [25], the optimal membership and prototype functions of (1) is

$$u_{ij} = \left(\sum_{r=1}^{c} d_{ij}^{2/(m-1)} / d_{rj}^{2/(m-1)}\right)^{-1} \text{ and } v_i = \sum_{j=1}^{n} u^{m}{}_{ij} x_j / \sum_{j=1}^{n} u^{m}{}_{ij}. \tag{2}$$

All fuzzy membership degrees consist of a $n \times c$ fuzzy partition matrix $U = [u_{ij}]$. FCM has frequently been criticized as it cannot show the typicality (importance) or compatibility of points with clusters [6,12], and thus the following algorithm was proposed.

(2) Pedrycz's conditional FCM (CFCM): Let w_j be the importance degree of jth point, Equation (1) in CFCM turns into

$$\min J(U,V) = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m} d_{ij}^{2},\ \ s.t.\ \sum_{i=1}^{c} u_{ij} = w_j,\ j = 1,2,\ldots,n,\ 0 < \sum_{j=1}^{n} u_{ij} \le n. \tag{3}$$

The membership degree of jth point to ith cluster in FCM is conducted as

$$u_{ij} = w_j / (\sum_{r=1}^{c} d_{ij}^{2/(m-1)} / d_{rj}^{2/(m-1)}). \quad (4)$$

The computation equation of the center v_i in CFCM is the same as in the case of FCM, $i = 1, 2, \ldots, c$. The use of CFCM can enhance the typicality of different clusters and increase the accuracy of FCM. But CFCM only focuses on the typicality of points rather than clusters.

(3) Krishnapuran et al.'s possibilistic c-means (PCM): Along the objective function of FCM, PCM is formulated as

$$\min J(U, V) = \sum_{i=1}^{c} \sum_{j=1}^{n} \left\{ u_{ij}^{m} d_{ij}^{2} + (1 - u_{ij})^{m} \eta_i \right), \quad s.t.\ 0 < \sum_{j=1}^{n} u_{ij} \leq n. \quad (5)$$

From (5), the optimal membership function is

$$u_{ij} = \left\{ 1 + \left({d_{ij}}^2 / \eta_i \right)^{1/(m-1)} \right\}^{-1}, \ i = 1, \ 2, \ \ldots, \ c \quad (6)$$

where η_i is associated with the size of each cluster, it is computed as

$$\eta_i = \sum_{j=1}^{n} u_{ij}^{m} d_{ij}^{2} / \sum_{j=1}^{n} u_{ij}^{m}. \quad (7)$$

PCM can effectively stress the typicality of points but must strongly depend on an initialization procedure. In practice, FCM can realize this purpose [12].

(4) Ng et al.'s constrained CM (CCM): Equation (1) in FCM is turned into

$$\min J(U, V) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^{m} d_{ij}^{2}, \ \ s.t. \sum_{j=1}^{n} u_{ij} = w_i, i = 1, 2, \ldots, c, \ u_{ij} = 0 \ or \ 1. \quad (8)$$

From (8), the clustering center is

$$v_i = \sum_{j=1}^{n} u_{ij} x_j / \sum_{j=1}^{n} u_{ij}, \ i = 1, 2, \ldots, c. \quad (9)$$

The optimal membership function u_{ij} in CCM turns into the typical transportation problem (see [19]) that can be solved by a set of existing algorithms. But the computational complexity of these algorithms is too large to be applied in large dataset.

These algorithms have their own applicable ranges and limitations but cannot provide an assessment that utilizes the information on both data objects and clusters. Nevertheless, these constraints are not only helpful to boost clustering quality but also meet mandatory application requirements. In this paper, we propose a new method to solve these problems after conducting the iterative equation along a solid mathematical optimization process.

3. Double-Constraint Fuzzy Clustering

Let $X = \{x_j\}$ be a dataset with n data objects that are distributed in c clusters in a d-dimensional data space, $x_j \in R^d$. According to the fuzzy partition matrix U in FCM, we define two symbols as follows:

$$p_i = \sum_{j=1}^{n} u_{ij} \text{ and } q_j = \sum_{i=1}^{c} u_{ij}, \ i = 1, 2, \ldots, c; \ j = 1, 2, \ldots, n \quad (10)$$

The value of p_i is the constraint for the ith cluster if the constraint can be known a priori, whereas the value of q_j is the constraint for jth data object.

The meaning of q_j in PFCM is illustrated as follows:

1. $q_j < 1$, the jth data object is under sparse distribution, likely being a noisy point or outlier;
2. $q_j = 1$, the jth data object does not have an additional importance degree, and thus the point has the same membership degree value as it has in FCM;

3. $q_j > 1$, the jth cluster may be an aggregation of data with high density such as a clustering center and so on; these points act as the main structure of various clusters.

Alternatively, the meaning of p_i in CCM is defined as the number of data points in the ith cluster, and it is usually a mandatory requirement of the clustering results of any clustering algorithm. To date, no existing fuzzy clustering algorithm can combine the constraints of both data objects and clusters to enhance clustering quality.

Since q_j reflects the importance of each point whereas p_i reflects that of each point, they can represent the typicality or compatibility of points and clusters together. According to the constraints on both point and cluster importance degrees, a double-constraint fuzzy clustering algorithm is proposed, abbreviated as DFCM. The objective function of DFCM is formulated as

$$\min J(U,V) = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m d_{ij}^2,\ \ s.t.\ \sum_{j=1}^{n} u_{ij} = p_i,\ \sum_{i=1}^{c} u_{ij} = q_j,\ i = 1,2,\ldots,c;\ j = 1,2,\ldots,n \tag{11}$$

Taking $n + c$ Lagrange multipliers: λ_j, j = 1, 2,..., n; u_i, i = 1, 2, ..., c, the typical alternative optimization way [25] is used to solve (11). And the Lagrange function is formulated as

$$F_m = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m (x_j - v_i)^2 + \sum_{j=1}^{n} \lambda_j \left(\sum_{i=1}^{c} u_{ij} - q_j\right) + \sum_{i=1}^{c} \mu_i \left(\sum_{j=1}^{n} u_{ij} - p_i\right), \tag{12}$$

Equation (12) is solved by the following two alternative optimization problems.

Problem P1:

Fix u_{ij}: solve the kth cluster center v_k. The derivative of F_m on v_k is

$$\begin{aligned}\frac{\partial J_m}{\partial v_k} &= \sum_{i=1}^{c}\sum_{j=1}^{n} \frac{\partial u_{ij}^m d_{ij}^2}{\partial v_k} - \frac{\partial}{\partial v_k}\left(\sum_{j=1}^{n} \lambda_j \left(\sum_{i=1}^{c} u_{ij} - q_j\right)\right) - \frac{\partial}{\partial v_k}\left(\sum_{i=1}^{c} \mu_i \left(\sum_{j=1}^{n} u_{ij} - p_i\right)\right) \\ &= \sum_{i=1}^{c}\sum_{j=1}^{n} \frac{\partial u_{ij}^m (x_j - c_i)^2}{\partial v_k} = -2\sum_{j=1}^{n} u_{kj}^m (x_j - v_k) = 0\end{aligned},$$

Thus

$$v_k = \sum_{j=1}^{n} u_{kj} x_j \Big/ \sum_{j=1}^{n} u_{kj},\ \ s.t.,\ k = 1,\ 2,\ldots,c, \tag{13}$$

Problem P2:

Fix v_k, solve u_{kg} from the gth data vector to v_k. The derivative of F_m on u_{kg} yields

$$\partial F_m / \partial u_{kg} = m u_{kg}^{m-1} (x_j - v_i)^2 + \lambda_g + \mu_k = 0,$$

it is

$$\mu_{kg}^{m-1} = (-\lambda_g - \mu_k)/(m(x_g - v_k)^2). \tag{14}$$

Taking it into $\sum_{j=1}^{n} u_{ij} = p_i$, $\sum_{i=1}^{c} u_{ij} = q_j$, they are

$$\begin{cases} \sum\limits_{i=1}^{c} u_{ig} = \sum\limits_{i=1}^{c} \left(\frac{-\lambda_g - \mu_k}{m(x_g - v_i)^2}\right)^{1/(m-1)} = q_g,\ g = 1,2,\ldots,n \\ \sum\limits_{j=1}^{n} u_{kj} = \sum\limits_{i=1}^{c} \left(\frac{-\lambda_j - \mu_k}{m(x_j - v_k)^2}\right)^{1/(m-1)} = p_k,\ k = 1,2,\ldots,c \end{cases}. \tag{15}$$

Since the number of equations in (15) is $(n + c)$ and is equal to the number of both variables μ_k and λ_g, its solution is thus uniformly determined. But the power of $1/(m-1)$ limits its analytic solution. We turn to solve it iteratively using an iteratively numerical optimization process. Note that

$$\begin{aligned}\frac{\partial}{\partial \lambda_g}\sum_{k=1}^{c}\left(\frac{-\lambda_g - \mu_k}{m(x_g - v_i)^2}\right)^{1/(m-1)} &= -\sum_{k=1}^{c}\frac{1}{m-1}\left(\frac{-\lambda_g - \mu_k}{m(x_g - v_k)^2}\right)^{\frac{2-m}{m-1}}\frac{1}{m(x_g - v_k)^2} \\ \frac{\partial}{\partial \mu_g}\sum_{g=1}^{n}\left(\frac{-\lambda_g - \mu_k}{m(x_g - v_k)^2}\right)^{1/(m-1)} &= -\sum_{g=1}^{n}\frac{1}{m-1}\left(\frac{-\lambda_g - \mu_k}{m(x_g - v_k)^2}\right)^{\frac{2-m}{m-1}}\frac{1}{m(x_g - v_k)^2}\end{aligned}.$$

According to the Newton iteration method [25], u_{kg} is iteratively solved as

$$\begin{cases} \lambda_g^{t+1} = \lambda_g^t - \left(\sum_{k=1}^{c} \left(\frac{-\lambda_g-\mu_k}{m(x_g-v_k)^2}\right)^{1/(m-1)} - q_g\right) / \left(-\sum_{k=1}^{c} \frac{1}{m-1}\left(\frac{-\lambda_g-\mu_k}{m(x_g-v_k)^2}\right)^{\frac{2-m}{m-1}} \frac{1}{m(x_g-v_k)^2}\right) \\ \mu_k^{t+1} = \mu_k^t - \left(\sum_{g=1}^{n} \left(\frac{-\lambda_g-\mu_k}{m(x_g-v_i)^2}\right)^{1/(m-1)} - p_k\right) / \left(-\sum_{g=1}^{n} \frac{1}{m-1}\left(\frac{-\lambda_g-\mu_k}{m(x_g-v_k)^2}\right)^{\frac{2-m}{m-1}} \frac{1}{m(x_g-v_k)^2}\right) \end{cases}, \quad (16)$$

where t is the iteration time. In this way, DFCM is alternatively optimized as follows. Given the initial (v^0, λ^0, μ^0), (λ^1, μ^1) can be calculated by (16); then, u^1 is calculated by (14); v^1 is obtained by (13). Then, (v^1, λ^1, μ^1) is used to calculate (λ^2, μ^2), and the above process is repeated until a stop criterion is met.

Especially, when m = 2, (15) reduces to the following form:

$$\begin{cases} \lambda_g = -\left(\sum_{k=1}^{c} \frac{\mu_k}{(x_g-v_k)^2} + q_g\right) / \sum_{k=1}^{c} \frac{1}{(x_g-v_k)^2}, \; g = 1, 2, \ldots, n \\ \mu_k = -\left(\sum_{g=1}^{n} \frac{\lambda_g}{(x_g-v_i)^2} + p_k\right) / \sum_{g=1}^{n} \frac{1}{(x_g-v_k)^2}, \; k = 1, 2, \ldots, c \end{cases}. \quad (17)$$

According to (17), DFCM can be iteratively solved as follows. Given the initial v^0, both λ^0 and μ^0 are solved. Subsequently, v^1 is determined using (13), followed by the computation of λ^1 and μ^1, and this sequence continues iteratively. The process is repeated until a stop criterion is met. According to the algorithm optimization principle, the convergence of this process is guaranteed.

In practice, the weighting value of each point q_j can be evaluated by the importance degree of each point. However, the weighting value p_i cannot be directly associated with any cluster since these clusters are unknown before the clustering process is completed.

To settle this problem, we implement DFCM in two steps: coarse partitioning and fine partitioning. In the first step, the use of FCM can obtain c clusters, C_1, C_2, …, and C_c, subject to $|C_1| < |C_2| < \ldots < |C_c|$.

In the second step, p_1, p_2, …, p_c are actually the number of data vectors in all clusters such that $p_1 < p_2 < \ldots < p_c$. We add these constraints of p_1, p_2, …, p_c to their corresponding c clusters in FCM. Beginning with these clustering centers in FCM, DFCM is used to partition C_1, C_2, …, C_c and obtain the final clustering results.

The computational time of DFCM includes the computation of the weighting value of p_i as well as the coarse-tuning and fine-tuning steps. However, these two steps account for the majority of the runtime in the entire clustering process. But DFCM begins with the clustering results (centers) derived from FCM, enabling it to reach an optimal solution more rapidly and effectively.

Algorithm 1. DFCM algorithm

Input: Dataset X, number of clusters c, exponent indexes m, and acceptable error ε
Output: Partitioned clusters from X

Method:
(1) Determine p_1, p_2, …, p_c and q_1, q_2, …, q_n;
(2) Partition X to C_1, C_2, …, C_c by FCM;
(3) Determine p_1, p_2, …, p_c from $|C_1|$, $|C_2|$, …, $|C_c|$;
(4) Initialize the clustering center in DFCM by v_1, v_2, …, v_c from FCM;
(5) Solve u_{ij} of jth point to cluster by (16) or (17), i = 1~c, j = 1~n;
(6) Solve v_i by (13), i = 1~c;
(7) Stop if $\|U^{s+1} - U^s\| \leq \varepsilon$,otherwise go to step (5);
(8) Partition X to C_1, C_2, …, C_c by their final membership degrees.

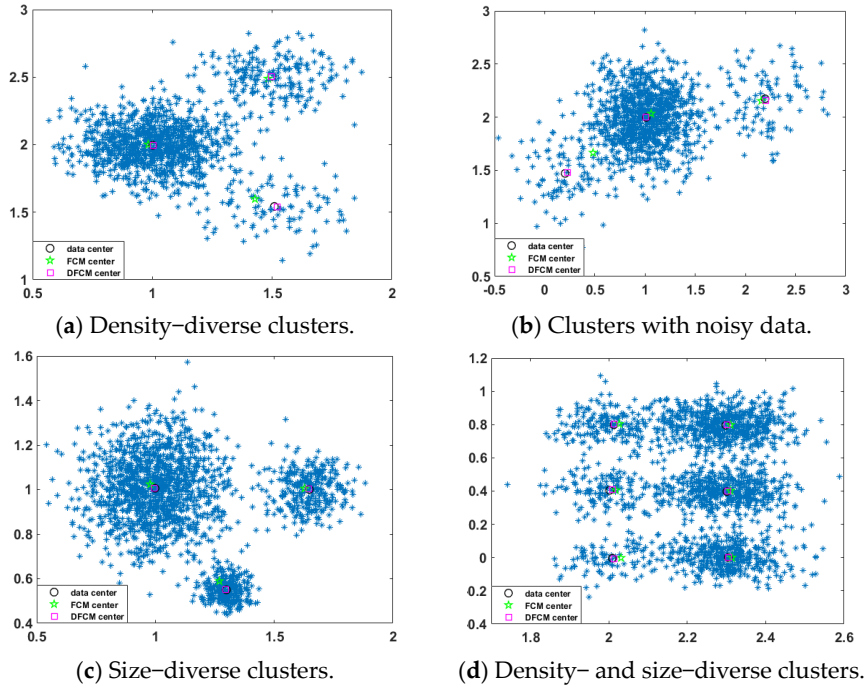
4. Experiment

Four synthetic low-dimensional datasets with different clustering features (e.g., density, size, and overlap) and eight actual datasets from UCI were used to assess the effectiveness and efficiency of DFCM. We applied DFCM to partition all data in these datasets and compared the results with three typical clustering algorithms: FCM, PCM, and CFCM.

4.1. Four Synthetic Datasets

Each cluster in the four synthetic datasets is generated by "*randn()*" in the Matlab® toolbox. Thus, each cluster is regular and is centralized on the center of the related function. As a result, after labeling the above "*randn()*" functions, the correct cluster label of any data point is just the label of the function that generates these data. These original cluster labels in the above datasets do not take part in any clustering process but are just used to examine the accuracy of these algorithms after the clustering process is completed. The four synthetic datasets are indicated as Set 1–Set 4.

The effectiveness of a clustering algorithm is typically assessed using datasets characterized by various features such as density difference, size difference, and noise effects. And Sets 1~4 are constructed along these features. Set 1 contains 1300 data points distributed across three clusters: a high-density cluster with 1000 data points and two low-density clusters with 100 and 200 data points each (see Figure 1a). These clusters exhibit diversity in terms of density. Set 2 contains 1200 data points distributed across slightly overlapping clusters (see Figure 1b). Set 3 contains 1900 data points distributed across three size-diverse clusters, where the largest cluster possesses a diameter twice that of the two smaller clusters (see Figure 1c). In Set 4, there are 1500 data points distributed across spherical clusters that partially overlap (see Figure 1d). In these figures, the centers derived from FCM and DFCM are marked by small green and red circles, respectively.



(**a**) Density−diverse clusters. (**b**) Clusters with noisy data.

(**c**) Size−diverse clusters. (**d**) Density− and size−diverse clusters.

Figure 1. Synthetic datasets with different features.

4.2. Eight Real Datasets from UCI

Eight actual datasets from UCI [26] were used to assess the clustering accuracy and the mandatory constraint on the volume of data in each cluster. These datasets were selected due to their representativeness of different clustering structures and characteristics. Other datasets from UCI have similar features to these eight datasets. The correct clustering labels

and the volume of data in each cluster are known a priori. These labels remain separate from the clustering process and are only used to evaluate the accuracy of various clustering results. Table 1 shows the number of clusters, the volume of data in each cluster, and the dimensionality of each dataset.

Table 1. Characteristics of 8 actual datasets from UCI.

No.	Dataset	*n*	*dim*	*c*	
1	*Iris*	3	4	150	50/50/50
2	*Seeds*	3	7	210	70/70/70
3	*Tea*	3	5	151	49/50/52
4	*Breast*	6	9	106	21/15/18/16/14/22
5	*Cancer*	2	9	683	444/239
6	*Wholesale*	2	7	440	298/142
7	*Appendicitis*	2	7	106	21/85
8	*Wisconsin*	2	9	699	444/239

Notes: The symbols "*n*", "*dim*", and "*c*" are the number of data points, dimension, and the number of clusters in each dataset, respectively.

The *Iris* dataset contains 150 data points, each characterized by four attributes and distributed across three clusters. Each cluster contains 50 data vectors. Two clusters overlap, while the third cluster is linearly separable from those two clusters. In the past decades, this dataset has frequently been used to assess the clustering results of different clustering algorithms. The, *Tea*, and *Breast* datasets exhibit characteristics similar to those of *Iris*. The *Wisconsin* dataset is high-dimensional, containing 683 instances after removing 16 instances due to missing values. Each instance has nine attributes. This dataset contains two clusters: 444 samples are categorized as "Benign" and 239 as "Malignant". These two clusters are mainly in two different hyperplanes that respond to different components (attributes). The *Cancer*, *Appendicitis*, and *Wisconsin* datasets have similar characteristics.

4.3. Clustering Results

The clustering results were assessed using three indices: accuracy, the number of data points in each cluster, and runtime. Accuracy was determined by the percentage of correctly partitioned data points in each dataset. The total number of incorrectly partitioned data points in each cluster was determined using the following index:

$$sum = \sum_{i=1}^{c} \left| \left| C_i^1 \right| - \left| C_i^2 \right| \right| / n. \tag{18}$$

where $\left|C_i^1\right|$ and $\left|C_i^2\right|$ are the actual and the computed numbers of data in the ith cluster, respectively. On the other hand, according to CFCM, the weighting value of the jth data point q_j is determined by its density as

$$q_j = \rho_j / (\sum_{k=1}^{n} \rho_k / n),\ s.t.,\ \rho_j = (\sum_{k \in N(x_j)} \left\| x_k - x_j \right\|)^{-1}, j = 1, 2, \ldots, n. \tag{19}$$

where $N(x_j)$ is the set of neighboring data points around x_j and $\|\cdot\|$ denotes the distance between any pair of points.

All clustering results are presented in Table 2 and the clustering centers of Sets 1–4 are shown in Figure 1. According to clustering *accuracy* and *Sum*, we compared four clustering algorithms: FCM, PCM, CFCM, and DFCM. In these algorithms, all fuzziness exponents are uniformly taken as 1.5 and the stop error are 10^{-3}.

In the four synthetic datasets, Figure 1 illustrates that the cluster centers derived from FCM deviate from the actual centers. This deviation is attributed to FCM's limited capacity to discern the typicality or compatibility of points for effective clustering. Contrary to FCM, DFCM was able to determine the clustering centers more accurately. Table 3 presents a detailed comparative analysis of the four clustering algorithms. In terms of clustering

accuracy, DFCM surpasses the other three algorithms, with CFCM ranking the second and FCM achieving the lowest accuracy. The results demonstrate that DFCM is effective and valuable. The incorporation of dual constraints in DFCM enhances the accuracy of the clustering centers and the corresponding membership degrees. In contrast, PCM and CFCM exhibit slight deviations, whereas DFCM demonstrates negligible deviations. Hence, the value of *Sum* of DFCM is the lowest among the three algorithms. PCM and CFCM rank as intermediate, while FCM has the highest value. Furthermore, the value of *Sum* derived from CFCM is nearly the same as DFCM, as shown in Table 2. However, FCM has the shortest runtime when the number of clusters is fixed, and both PCM and DFCM depend on FCM due to their initialization processes.

Table 2. Clustering results of all tested datasets.

	Algorithm	**FCM**		**PCM**		**CFCM**		**DFCM**	
	Dataset	***A*(%)/*Sum***	***Time* (*s*)**	***A*(%)/*Sum***	***Time* (*s*)**	***A*(%)/*Sum***	***Time* (*s*)**	***A*(%)/*Sum***	***Time* (*s*)**
Synthetic datasets	Set1	95.9/0.075	0.0283	96.2/0.071	0.1222	97.8/0.039	4.8718	98.5/0.006	4.9533
	Set2	87.8/0.238	0.0216	80.0/0.403	0.1336	89.8/0.128	4.1582	98.9/0.012	4.3201
	Set3	95.5/0.092	0.0206	98.4/0.032	0.1817	97.1/0.058	10.995	99.5/0.003	11.218
	Set4	94.4/0.071	0.0475	95.0/0.059	0.2010	94.8/0.062	11.315	96.2/0.005	11.320
Real datasets	Iris	88.0/0.133	0.0111	92.7/0.000	0.0826	90.7/0.027	0.6003	91.3/0.040	0.6218
	Seeds	89.0/0.067	0.0300	89.0/0.067	0.0768	90.5/0.067	0.0479	91.9/0.038	0.0520
	Tea	49.7/0.305	0.0037	52.3/0.265	0.0886	51.0/0.305	0.0562	55.0/0.132	0.0702
	Breast	51.9/0.547	0.0090	57.5/0.381	0.0846	58.5/0.528	0.1157	64.2/0.208	0.1201
	Cancer	92.8/0.081	0.0092	91.7/0.018	0.0804	93.8/0.018	0.1701	93.7/0.035	0.1761
	Wholesale	82.7/0.264	0.0131	79.8/0.277	0.1243	87.3/0.168	0.0924	88.2/0.118	0.1049
	Appendicitis	79.4/0.165	0.0088	84.5/0.062	0.1118	83.5/0.082	1.5589	88.7/0.021	1.5838
	Wisconsin	93.8/0.060	0.0090	96.0/0.063	0.1384	95.6/0.009	1.1571	95.9/0.009	1.2124

Note: For any clustering algorithm, "*A* (%)" indicates accuracy (percentage), "*Sum*" is computed by Equation (18), and "*Time* (*s*)" is the CPU runtime (second). The items underlined in red are the best results in any corresponding row.

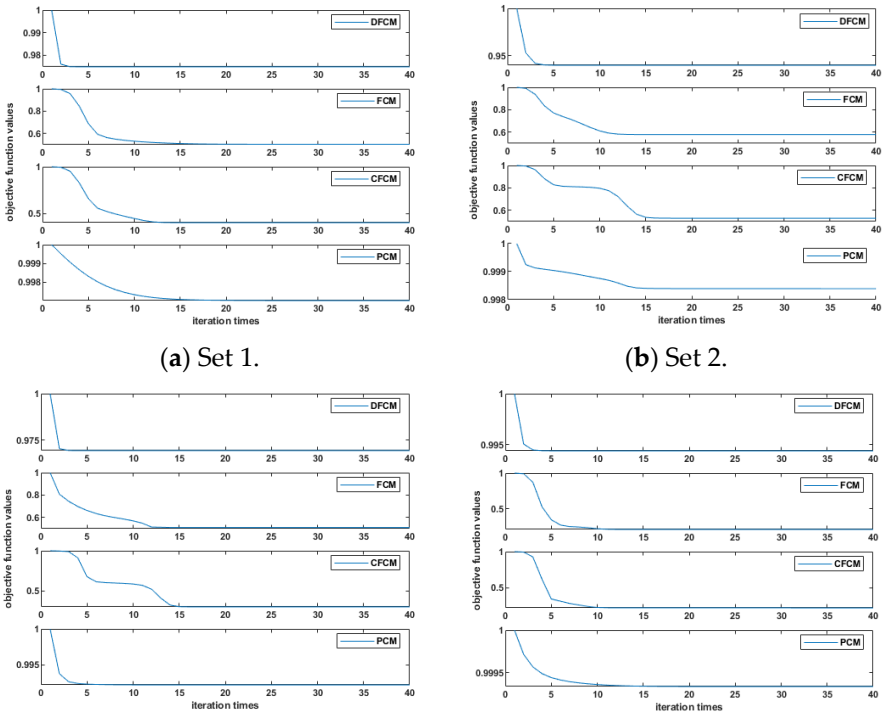
Table 3. Applicable ranges and limitations of the four clustering algorithms.

Algorithm	**Overlapped Cluster**	**Different Densities**	**Different Sizes**	**Different Shapes**	**Time Complexity**
FCM	Applicable	Applicable	Inapplicable	Inapplicable	O (*ctn*)
PCM	Applicable	Applicable	Applicable	Inapplicable	O (*ctn*)
CFCM	Inapplicable	Applicable	Partially applicable	Partially applicable	O (n^3)
DFCM	Applicable	Applicable	Applicable	Partially applicable	O (n^2)

In the eight real datasets, Table 2 shows that DFCM outperformed the other three algorithms in five of the datasets, whereas it failed in terms of *A* and *Sum* in the remaining three. However, the clustering results of DFCM is very close to the best results of the three datasets. Hence, DFCM exhibits a slight superiority compared to the other three algorithms. Our conclusions based on these results are as follows: Firstly, most clusters in these eight datasets are non-spherical, but the four algorithms can in principle work well only with datasets with spherical clusters. Secondly, the complex structures of the eight datasets result in a generally low clustering accuracy of any clustering algorithm applied. In recent decades, the clustering accuracy rates of these datasets were less than 50% using existing algorithms [6].

In terms of average runtime, FCM's runtime was the shortest among the four algorithms, followed by PCM, CFCM, and DFCM. These results are consistent with those in the four synthetic datasets. Figure 2 shows the convergence of DFCM in the four synthetic datasets compared with the other three clustering algorithms, and the number of iterations is fixed at 40. The corresponding objective function value reflects the convergence speed.

Especially, as these objective functions of the four algorithms have different orders of magnitude, Figure 2 illustrates their relative values of objective function by normalizing these values to the interval [0, 1]. As shown in Figure 2, PCM has the fastest convergence speed among the other three algorithms, followed by DFCM and FCM. Moreover, CFCM shows instability across the four datasets. Given that the convergence speed of any iteration algorithm reflects its runtime and varying tendency, Figure 2 clearly illustrates the runtime of the four algorithms when applied to various datasets.



(**a**) Set 1. (**b**) Set 2.

(**c**) Set 3. (**d**) Set 4.

Figure 2. Convergence curves of the four synthetic datasets.

4.4. Comparison and Discussion

When confronted with various datasets with different characteristics, these four algorithms reveal their limitations. According to the general approach of evaluating a clustering algorithm, Table 3 summarizes their applicable ranges based on five common features: cluster overlap, density difference, size difference, cluster shape, and time complexity. The term "applicable" indicates that the given algorithm can correctly cluster these data points in the dataset. Conversely, "inapplicable" indicates that the given algorithm is completely incapable of clustering the dataset correctly, while "partially applicable" refers to limited effectiveness in clustering a dataset with the relevant features.

Furthermore, the clustering results of the four clustering algorithms are explained and discussed as follows. Although FCM, PCM, CFCM, and DFCM can all be used to cluster datasets with overlapping clusters, both CFCM and DFCM can perform better since they stress the typicality of each point. And PCM must depend on FCM as its initialization process and otherwise cannot be applied to separate overlapping clusters. Differences in density and size can cause larger errors for FCM and PCM and partially affect the clustering results of both CFCM and DFCM due to their condition of constraining various clusters. In principle, the four algorithms cannot cluster data that are distributed in arbitrarily shaped clusters. But DFCM and DFCM have smaller errors than FCM and PCM when all clusters have convex shapes. Consequently, DFCM has an advantage over the other three algorithms in terms of clustering accuracy. But DFCM has a longer runtime than FCM and PCM, but not CFCM, and FCM is the most effective.

5. Conclusions

Fuzzy clustering is now extensively employed across various research fields. Numerous applications necessitate the partition of a dataset into clusters with a fixed number of instances. Additionally, it is essential to determine the importance degree of each data instance. However, to date, no clustering algorithm can effectively satisfy these criteria due to challenges in assigning a fixed number of data points to undefined clusters and the lack of a feasible iterative formula. To address this issue, we propose a new fuzzy clustering method utilizing a feasible iteration method which can be regarded as an extension of the fuzzy clustering algorithm. The new method emphasizes the importance degree of both clusters and individual data points, satisfying the dual criteria for specific points and underlying clusters.

Despite progress, the proposed DFCM algorithm does not address two critical issues in the clustering process. Firstly, determining the appropriate number of clusters remains unsolved. Therefore, the algorithm requires prior knowledge of cluster quantity in the dataset, which is a significant practical limitation. Secondly, similar to other existing clustering algorithms, the algorithm proposed in this study is primarily designed for datasets with spherical clusters. If it is employed to partition a dataset with non-spherical clusters, the clustering results usually exhibit significant errors. Currently, various adaptations of fuzzy clustering are being explored to address these issues. Integrating these effective methods with the proposed DFCM algorithm will be our future focus. Note that fuzzy clustering has been extended to applications involving clustering arbitrarily shaped clusters [27]. In the future, an important research direction is to enhance the capability of the DFCM algorithm to accurately cluster datasets with various cluster shapes.

Author Contributions: Software, validation, formal analysis, and data curation, S.Z. and Y.Z.; conceptualization, resources, and supervision, S.Y.; methodology, visualization, investigation, and writing—original draft preparation, S.Z., Y.Z. and S.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Science Foundation of China, grant number 61973232.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The UCI datasets used in this article are from the UCI Machine Learning Repository (https://archive.ics.uci.edu (accessed on 18 October 2023)).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. MacQueen, J.B. Some Methods for Classification and Analysis of Multivariate Observations. In *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*; Project Euclid: Durham, NC, USA, 1967; pp. 281–297.
2. Detroja, K.P.; Gudi, R.D.; Patwardhan, S.C. A possibilistic clustering approach to novel fault detection and isolation. *J. Process Control.* **2006**, *16*, 1055–1073. [CrossRef]
3. Yue, S.H.; Wang, J.S.; Tao, G.; Wang, H.X. An unsupervised grid-based approach for clustering analysis. *Sci. China Inf. Sci.* **2010**, *53*, 1345–1357. [CrossRef]
4. Bezdek, J.C. *Fuzzy Models for Pattern Recognition*; Plenum Press: New York, NY, USA, 1992.
5. Lei, T.; Liu, P.; Nandi, A.K. Automatic Fuzzy Clustering Framework for Image Segmentation. *IEEE Trans. Fuzzy Syst.* **2020**, *28*, 2078–2092. [CrossRef]
6. Xu, R.; Wunsch, D. Survey of clustering algorithms. *IEEE Trans. Neural Netw.* **2005**, *16*, 645–678. [CrossRef]
7. Setnes, M. Supervised fuzzy clustering for rule extraction. *IEEE Trans. Fuzzy Syst.* **2000**, *8*, 416–424. [CrossRef]
8. Arbelaitz, O.; Gurrutxaga, I.; Muguerza, J.; Perez, J.M.; Perona, I. An extensive comparative study of cluster validity indices. *Pattern Recognit.* **2013**, *46*, 243–256. [CrossRef]
9. Hang, G.; Lu, J.; Zhang, Y. Application of spherical fuzzy c-means algorithm in clustering Chinese documents. *J. Syst. Simul.* **2004**, *3*, 516–518.
10. Wang, Z.; Wang, S.S.; Shao, Y.H. Semisupervised fuzzy clustering with fuzzy pairwise constraints. *IEEE Trans. Fuzzy Syst.* **2022**, *30*, 3797–3811.
11. Krishnapuran, R.; Keller, J.M. A possibilistic approach to clustering. *IEEE Trans. Fuzzy Syst.* **1993**, *1*, 98–110. [CrossRef]

12. Pedrycz, W. Conditional fuzzy c-means. *Pattern Recognit. Lett.* **1996**, *17*, 625–631. [CrossRef]
13. Yu, J. General c-means clustering model and its applications. In Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison, WI, USA, 18–20 June 2003; Volume 2, pp. 122–127. [CrossRef]
14. Huang, Z.; Ng, M.K.; Rong, H. Automated variable weighting in k-means type clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 657–668. [CrossRef] [PubMed]
15. Yue, S.H.; Li, P. Adaptive fuzzy clustering. *J. Zhejiang Univ. Sci.* **2004**, *6*, 49–55.
16. Lei, T.; Jia, X.; Liu, T.; Liu, S.; Meng, H.; Nandi, A.K. Adaptive morphological reconstruction for seeded image segmentation. *IEEE Trans. Image Process.* **2019**, *28*, 5510–5523. [CrossRef] [PubMed]
17. Wang, Y.X.; Chen, L.; Zhou, J.; Li, T.J.; Yu, Y.F. Pairwise constraints-based semi-supervised fuzzy clustering with multi-manifold regularization. *Inf. Sci.* **2023**, *638*, 778–785. [CrossRef]
18. Mika, S.I. Dynamic fuzzy clustering using fuzzy cluster loading. *Int. J. Gen. Syst.* **2006**, *35*, 209–230. [CrossRef]
19. Ng, M.K. A note on constrained k-means algorithms. *Pattern Recognit.* **2000**, *33*, 515–519. [CrossRef]
20. Peng, Y.; Zhu, X.; Ge, Y. Fuzzy graph clustering. *Inf. Sci.* **2021**, *571*, 38–49. [CrossRef]
21. Liu, J.W.; Xu, M.Z. Kernelized fuzzy attribute C-means clustering algorithm. *Fuzzy Sets Syst.* **2008**, *159*, 2428–2445. [CrossRef]
22. Yue, S.H.; Wang, J.P.; Wang, J.S.; Bao, X.J. A new validity index for evaluating the clustering results by partitional clustering algorithms. *Soft Comput.* **2016**, *20*, 1127–1138. [CrossRef]
23. Anderson, D.; Zare, A.; Price, S. Comparing fuzzy, probabilistic, and possibilistic partitions using the earth mover's distance. *IEEE Trans. Fuzzy Syst.* **2013**, *21*, 766–775. [CrossRef]
24. Glenn, T.C.; Zare, A.; Gader, P.D. Bayesian fuzzy clustering. *IEEE Trans. Fuzzy Syst.* **2015**, *23*, 1545–1561. [CrossRef]
25. Carter, M.W.; Price, C.C. *Operations Research*; CRC Press Inc.: Boca Raton, FL, USA, 2000.
26. UCI Dataset. Available online: http://archive.ics.uci.edu/ml/datasets.php (accessed on 12 October 2021).
27. Li, Q.; Yue, S.H.; Wang, Y.R. Boundary matching and interior connectivity-based cluster validity analysis. *Appl. Sci.* **2020**, *10*, 1337. [CrossRef]