Article

VSFCM: A Novel Viewpoint-Driven Subspace Fuzzy C-Means Algorithm

Yiming Tang [1,2,3], Rui Chen [1,2,*] and Bowen Xia [1,2]

1 School of Computer and Information, Hefei University of Technology, Hefei 230601, China; ymtang@hfut.edu.cn (Y.T.); xbw810hfut@163.com (B.X.)
2 Anhui Province Key Laboratory of Affective Computing and Advanced Intelligent Machine, Hefei 230601, China
3 Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6R 2V4, Canada
* Correspondence: cr200002fighting@163.com

Abstract: Nowadays, most fuzzy clustering algorithms are sensitive to the initialization results of clustering algorithms and have a weak ability to handle high-dimensional data. To solve these problems, we developed the viewpoint-driven subspace fuzzy c-means (VSFCM) algorithm. Firstly, we propose a new cut-off distance. Based on this, we establish the cut-off distance-induced clustering initialization (CDCI) method and use it as a new strategy for cluster center initialization and viewpoint selection. Secondly, by taking the viewpoint obtained by CDCI as the entry point of knowledge, a new fuzzy clustering strategy driven by knowledge and data is formed. Based upon these, we put forward the VSFCM algorithm combined with viewpoints, separation terms, and subspace fuzzy feature weights. Moreover, compared with the symmetric weights obtained by other subspace clustering algorithms, the weights of the VSFCM algorithm exhibit significant asymmetry. That is, they assign greater weights to features that contribute more, which is validated on the artificial dataset DATA2 in the experimental section. The experimental results compared with multiple advanced clustering algorithms on the three types of datasets validate that the proposed VSFCM algorithm has the best performance in five indicators. It is demonstrated that the initialization method CDCI is more effective, the feature weight allocation of VSFCM is more consistent with the asymmetry of experimental data, and it can achieve better convergence speed while displaying better clustering efficiency.

Keywords: fuzzy clustering; fuzzy c-means; cluster center initialization; machine learning; image processing; fuzzy sets

1. Introduction

Clustering algorithms have received widespread attention and applications in various fields, such as pattern recognition, biology, engineering systems, image processing, and so forth [1–7]. For clustering algorithms, the data points in a given data set are divided into several clusters, and the similarity between the data points in the same cluster is greater than that in other clusters; meanwhile, there are strong differences among the clusters, which contribute to the presentation of asymmetric data structure. In the early stage, hard (Boolean) clustering was mainly studied, where every object strictly belonged to a single cluster. For example, The DPC algorithm [8] is an excellent representative of this category.

Among numerous fuzzy clustering algorithms, the fuzzy c-means (FCM) algorithm is one of the most commonly used methods [9–14]. In general, FCM attached equal importance to all features of data, establishing a symmetrical structure, which might often be inconsistent with that of the original data. Therefore, the use of weighted processing has become an important development direction. The weighted FCM (WFCM) algorithm [15] grouped data according to the weighted categories of the separated features, and the algorithm incorporated feature weights into commonly used Euclidean distances for clustering. The feature-weighted fuzzy k-means (FWFKM) algorithm [16] was based on the

fuzzy k-prototypes algorithm and a supervised algorithm. However, it still required two objective functions to optimize the data partition and feature weights. The attribute weight algorithm (AWA) [17] was a fuzzy weighted subspace clustering algorithm, which could effectively find the important features of each cluster, that is, discover the asymmetry of data. However, the disadvantage of the AWA algorithm was that it did not work when the standard deviation of certain attributes was zero, as zero might be used as the denominator in the learning rules. Improved versions have been proposed to overcome this weakness, including the fuzzy weighted k-means (FWKM) algorithm [18] and the fuzzy subspace clustering (FSC) algorithm [19,20]. In the objective function of FWKM, a small constant was added when calculating the distance, which effectively avoided the problem caused by the zero standard deviation of some attributes in AWA. Gan et al., proposed the FSC algorithm, which used a strategy similar to the FWKM algorithm discussed earlier. However, they had a significant difference in the method of parameter setting. The constant parameter introduced in FSC should be set manually, while that in FWKM was set through a predefined formula.

However, when clustering is completed in a high-dimensional space, the traditional clustering algorithms will expose obvious drawbacks [21–23]. For example, for any given pair of data points in the cluster of a high-dimensional space, these points may be far apart. Due to the lack of multi-dimensional space thinking, traditional algorithms may have some deviations in calculating the distance, resulting in unsatisfactory clustering performance. For most traditional clustering algorithms, a key challenge is that in many real-world problems, data points in different clusters are often related to different feature subsets; that is, clusters can exist in different subspaces [24–26]. Frigui and Nasraoui [27] proposed a new approach called simultaneous clustering and attribute discrimination (SCAD). It used continuous feature weighting, providing a richer feature correlation representation than feature selection. Moreover, it also independently learned the associative representation of features of each cluster in an unsupervised way. Later, Deng [28] studied the use of intra-class and inter-class information, and proposed a new clustering method, called enhanced soft subspace clustering (ESSC). However, many irrelevant data would affect the clustering performance of fuzzy clustering algorithms. In other words, different characteristic features should have different importance in clustering. Yang and Nataliani [29] proposed a feature-reduction FCM (FRFCM) algorithm, which automatically calculated the weight of a single feature while reducing the influence of these unrelated features. Tang et al. [30] proposed a new kernel fuzzy clustering method called viewpoint-based kernel fuzzy clustering with weight information granules (VWKFC). Because its new initialization algorithm is more efficient in selecting cluster centers, namely the kernel-based hypersphere density initialization algorithm, combined with weight information particles and viewpoint induction mechanisms, VWKFC is significantly superior to the other eight existing related algorithms in processing high-dimensional data.

In the early stage, the clustering process was entirely data-driven. In fact, domain knowledge could be used to assist the development of clustering. W. Pedrycz et al. [31] introduced knowledge into a data-driven process, where knowledge was embodied through viewpoints, thus giving a viewpoint-based fuzzy clustering algorithm V-FCM. Tang et al. [32] proposed a new knowledge and data-driven fuzzy clustering algorithm, which was called the density viewpoint-induced possibilistic fuzzy clustering algorithm (DVPFCM). Thereinto, a new calculation method of density radius was proposed. Based upon this, the hypersphere density-based cluster center initialization (HDCCI) algorithm was established to obtain the initial cluster centers in densely sampled areas. Then, the high-density points obtained by the HDCCI method were used as new viewpoints, and they would be integrated into the DVPFCM algorithm.

The current problems of fuzzy clustering algorithms are mainly as follows:

- Sensitive to cluster initialization

Most fuzzy clustering algorithms are sensitive to the initial results of clustering. For example, FCM, V-FCM, SCAD, ESSC, and FRFCM all rely on the result of the initialization

method. Compared with them, the initialization processing mechanism in the DVPFCM algorithm is better; that is, the initialization method HDCCI is based on the DPC algorithm. However, the HDCCI algorithm still has a shortcoming existing in the calculation method of cut-off distance. It uses a fixed and strict formula, which lacks a solid foundation and proof process and can not adapt to various data sets.

- Weak adaptability to high-dimensional data

With the arrival of the big data era, the volume of information is constantly increasing, and the dimensions are getting higher simultaneously, which forces clustering algorithms to have the ability to deal with high-dimensional data properly. However, most clustering algorithms are still weak in this aspect. When faced with high-dimensional data, FCM, V-FCM, and DVPFCM have no corresponding measures. Obviously, none of them can do this. SCAD, ESSC, and FRFCM all use subspace processing methods with different weights, which are relatively, better. However, the algorithms mentioned above are purely data-driven algorithms, and the clustering efficiency and accuracy of high-dimensional data cannot reach the ideal level.

In this study, we put forward the VSFCM algorithm and the following is a brief introduction.

On the one hand, we put forward a new optimized cut-off distance calculation method based on the DPC algorithm, which can select the point with the highest density as the viewpoint to induce the algorithm to find the cluster center more accurately. That is, the cut-off distance-induced clustering initialization method CDCI provides a new initialization strategy and perspective in this field. Under the guidance of this initialization method, the number of iterations of the algorithm is reduced, and the accuracy of the algorithm is also improved. On the other hand, we introduce viewpoints into subspace clustering to improve the convergence speed of each subspace. Moreover, the separation term is added between the clusters to minimize the compactness of the subspace clusters and maximize the projection subspace of each cluster. On the basis of the CDCI and the viewpoint, combined with the above fuzzy feature weight processing mode, the viewpoint-driven subspace fuzzy c-means algorithm VSFCM is established.

The innovations of this study are reflected in the following aspects. First of all, a new cut-off distance is proposed, improving the cluster center initialization effect and also serving as a new viewpoint selection method, which can better adapt to the clustering data structure and speed up the clustering process. Secondly, the fuzzy weight is introduced, that is, a weight that is added to each cluster and dimension to reflect the contribution degree of each feature to the clustering. Among them, smaller weights are assigned to features with a larger proportion of noise point values to reduce their participation in clustering and indirectly weaken their impact on clustering results. Finally, a new method of separation between clusters is given. The average value of the initialized cluster centers is used as the reference point for separation between clusters. Additionally, combined with the optimized weight allocation, the distance between clusters can be effectively increased.

The paper is organized as follows. Section 2 reviews existing related algorithms. Section 3 describes the proposed clustering initialization method CDCI and the viewpoint-driven subspace fuzzy c-means algorithm VSFCM in detail. Section 4 presents the experimental results of the VSFCM algorithm and several other relevant algorithms on artificial data sets and some data sets of machine learning. Section 5 gives a summary and outlook for further research.

2. Related Work

In this section, we review several clustering algorithms closely related to our work.

Assuming that the data set $X = \{x_j\}_{j=1}^n$ is a set of n samples, we divide the data into c $(2 \leq c \leq n)$ clusters to get a set of cluster centers $V = \{v_i\}_{i=1}^c$. Each sample x_j and cluster center v_i are positioned in R^l space where l is the data dimension.

The objective function of the FCM algorithm [33] is as follows:

$$J_{\text{FCM}} = \sum_{j=1}^{n} \sum_{i=1}^{c} u_{ij}^m d_{ij}^2 \tag{1}$$

Among them, u_{ij} is the degree of membership, which ranges from 0 to 1, and needs to satisfy the constraint $\sum_{i=1}^{c} u_{ij} = 1 (j = 1, 2, \cdots, n)$. $d_{ij}^2 = \left\| x_j - v_i \right\|^2$ represents the Euclidean distance between the i-th cluster center and the j-th sample, and $m \in (1, +\infty)$ is a fuzzy coefficient.

Frigui and Nasraoui [27] established two versions of the SCAD algorithm. SCAD1 attempts to balance the two terms of a composite objective function and introduces a penalty term to determine the optimal attribute-related weight. Moreover, SCAD2 introduces a fuzzy weighted index to minimize the single-item criterion. In subsequent experiments, we choose to compare our algorithm with SCAD1, so only its expression is shown here:

$$J_{\text{SCAD1}} = \sum_{j=1}^{n} \sum_{i=1}^{c} u_{ij}^m \sum_{k=1}^{l} w_{ik} d_{ijk}^2 + \sum_{i=1}^{c} \theta_i \sum_{k=1}^{l} w_{ik}^2, \tag{2}$$

Thereinto, the distance formula is $d_{ijk} = \left| x_{jk} - v_{ik} \right|$, θ_i is a weighted constraint expressed as

$$\theta_i^{(t)} = K \frac{\sum_{j=1}^{n} \left(u_{ij}^{t-1} \right)^m \sum_{k=1}^{l} w_{ik}^{t-1} d_{ijk}^{t-1}}{\sum_{k=1}^{l} \left(w_{ik}^{t-1} \right)^2}. \tag{3}$$

Here K is a constant, and the superscript $t - 1$ is the previous iteration of the t-th. Frigui and Nasraoui proved that SCAD1 and SCAD2 had similar behavior and yielded similar clustering results.

Deng et al. [28] proposed the enhanced fuzzy weighted soft subspace clustering (ESSC) algorithm. The most prominent advantage of this algorithm is its ability to minimize the intra-class distance while maximizing the inter-class distance. The objective function of the algorithm is as follows:

$$J_{\text{ESSC}} = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m \sum_{k=1}^{l} w_{ik} \left\| x_{jk} - v_{ik} \right\|^2 + \gamma \sum_{i=1}^{c} \sum_{k=1}^{l} w_{ik} \ln(w_{ik}) - \eta \sum_{i=1}^{c} \left(\sum_{j=1}^{n} u_{ij}^m \right) \sum_{k=1}^{l} w_{ik} (v_{ik} - v_{0k})^2. \tag{4}$$

Here η and γ are constants, $\eta \in [0, 1]$, $\gamma \in [1, +\infty]$. Moreover, w_{ik} is the feature weight, which satisfies $\sum_{i=1}^{c} w_{ik} = 1 (k = 1, 2, \cdots, l)$. The third term is the weighted inter-class separation term, v_0 represents the center point of this data set, and the expression is

$$v_0 = \frac{\sum_{j=1}^{n} x_j}{n} \tag{5}$$

Because the ESSC algorithm considers both distances within and between classes and uses the idea of entropy information, it has certain advantages in soft subspace clustering algorithms belonging to the same category. One disadvantage of ESSC is that more parameters need to be set manually, and how to select the appropriate parameters of it is still an open question.

Yang and Nataliani [29] came up with the feature-reduction FCM (FRFCM) algorithm. It calculates a new weight for each feature by adding feature-weighted entropy to the FRFCM objective function. Then, during the iteration, the cluster centers and the

fuzzy membership matrix are updated with these new weights. It not only improves the performance compared with the FCM algorithm but also can select important features by weighting and reducing the feature dimension by discarding unimportant features. Therefore, the algorithm can automatically reduce the feature dimension and achieves a good clustering effect.

The V-FCM [31] algorithm optimizes the FCM algorithm by introducing the viewpoints. The viewpoints are represented by typical characteristic data of artificial selection, such as average value, maximum value, and minimum value. The objective function of the algorithm is as follows:

$$J_{\text{V-FCM}} = \sum_{j=1}^{n} \sum_{i=1}^{c} \sum_{k=1}^{l} u_{ik}^{m} \parallel x_{jk} - g_{ik} \parallel^{2}. \tag{6}$$

The V-FCM algorithm still has problems that it is not suitable for processing high-dimensional data, is sensitive to the initialization of cluster centers, and the selection of viewpoint types will also affect the clustering results.

The DPC algorithm [8] assumes that the cluster center is surrounded by neighbors with lower local density, and they are at a relatively large distance from any points with a higher local density. For all data points x_j, it is necessary to calculate their local density and the minimum distance δ_j between itself and other points with higher local density ρ_k. The formulas are ($j \in \{1, 2, \cdots, n\}$):

$$\rho_j = \sum f(d_{jk} - r), \tag{7}$$

$$f(x) = \begin{cases} 1, & x = d_{jk} - r < 0 \\ 0, & \text{other} \end{cases}, \tag{8}$$

$$\delta_j = \min\left\{ d_{jk} \middle| \rho_k > \rho_j, \ k \in \{1, 2, \cdots, n\} \right\}. \tag{9}$$

Here d_{jk} is the distance between two data points and r is the density radius. The local density reflects the number of data points within the radius r. For the point with the largest local density value, its $\delta_j = \max\left\{ d_{jk} \middle| k \neq j, \ k \in \{1, 2, \cdots, n\} \right\} (j \in \{1, 2, \cdots, n\})$. In the DPC algorithm, according to the $\rho - \delta$ distribution map, the data point with high density and a larger distance from the data point with a higher density are regarded as the cluster center. The reason is that the true cluster center has a large value of ρ_j and δ_j, while the δ of the noise point is small.

Recently, Tang et al. [32] proposed the density viewpoint-induced possibilistic fuzzy c-means (DVPFCM) algorithm. It provides a new initial method called the hypersphere density-based cluster center initialization (HDCCI) method, which is served for the viewpoint selection. On the basis of it, and combined with the advantages of FCM and PFCM, the DVPFCM algorithm is proposed. The specific objective function of DVPFCM is expressed as follows:

$$J_{\text{DVPFCM}} = \sum_{j=1}^{n} \sum_{i=1}^{c} \left(a u_{ij}^{m} + b t_{ij}^{p} \right) \|x_j - h_i\|^2 + \frac{\sigma^2}{m^2 c} \sum_{i=1}^{c} \sum_{j=1}^{n} \left(\varphi_j - t_{ij} \right)^p. \tag{10}$$

where

$$h_i = \begin{cases} v_i, & i \neq q \\ x_d, & i = q \end{cases}, \tag{11}$$

Here m and p are the fuzzy coefficients and the typical matrix fuzzy coefficient, respectively. With the help of the viewpoint, its convergence speed has been improved, and its robustness has also been strengthened.

Unfortunately, the above six algorithms, FCM, SCAD1, FRFCM, V-FCM, ESSC, and DVPFCM algorithms, all have the problem that they are sensitive to the initialization of

cluster centers. An improper initial value may cause the result to converge to a local optimal value or cause a slow clustering process, which has a great impact on the clustering result and also contributes to the unstable result of the algorithm.

Moreover, these algorithms also have other problems. The FCM algorithm is the most classic algorithm, but the actual efficiency and accuracy of processing high-dimensional data of it are usually not ideal. Although the V-FCM algorithm combines viewpoints to simplify the algorithm process and improve the convergence speed of the algorithm, its anti-noise ability is still weak. The ESSC algorithm proposes a weighted subspace clustering objective function based on entropy, which greatly increases the intra-cluster compactness and inter-cluster separation. However, more parameters need to be manually set, which increases the uncertainty of the algorithm. The characteristic of FRFCM is to select important features by weighting and to reduce the feature dimension by discarding unimportant features, but it fails to take the spatial features of the data into account.

The DPC algorithm is a hard clustering algorithm with a fast running speed, but its density radius is difficult to determine. The cluster center needs to be obtained by observing the density–distance distribution graph, for which it is easy to generate human error. The DVPFCM algorithm is slightly weaker when processing high-dimensional data. In addition, $r = \frac{\max(d_{ij})}{c}$ is proposed in DVPFCM, which is used to calculate the density cut-off radius, but there is no rigorous proof that the radius obtained is appropriate. Actually, this formula cannot achieve a good result in many cases, and the robustness is not very ideal. The comparison of these algorithms is summarized in Table 1.

Table 1. Comparison of advantages and disadvantages of each algorithm.

Algorithm	Advantage	Disadvantage
FCM	Classic algorithm can automatically find cluster centers.	Sensitive to cluster center initialization, poor noise immunity.
V-FCM	Simplified clustering process, fast convergence.	Poor noise immunity, sensitive to cluster center initialization.
DPC	Can quickly determine cluster centers.	The density radius is difficult to determine; cluster centers are not easy to obtain, and there are often human errors.
SCAD	Fuzzy weighted index is introduced to obtain better weight value.	Sensitive to cluster center initialization.
ESSC	Taking the distance between clusters into account, using entropy information.	Sensitive to cluster center initialization, more parameters need to be set manually.
FRFCM	Select important features by weighting, and reduce feature dimension by discarding unimportant features.	Sensitive to cluster center initialization without considering the spatial characteristics of the data.
DVPFCM	With the help of new viewpoints and typical values, there is relatively stronger robustness; there is a better initialization strategy.	Processing high-dimensional data appears weak, and the cut-off distance is not perfect.

In a word, the existing algorithms still have great defects in the initialization of cluster centers and the processing of high-dimensional data. For this reason, we focus on solving these two types of problems.

3. Proposed VSFCM Algorithm

In this section, a new cluster initialization method named the cut-off distance-induced clustering initialization (CDCI) is proposed first. Based on the CDCI, the viewpoint-driven subspace fuzzy c-means algorithm (VSFCM) is established subsequently, which has introduced the subspace clustering mode and fuzzy feature weight processing mechanism

and combined with the separation formula between clusters optimized with the viewpoint. Its algorithm idea is shown in Figure 1.
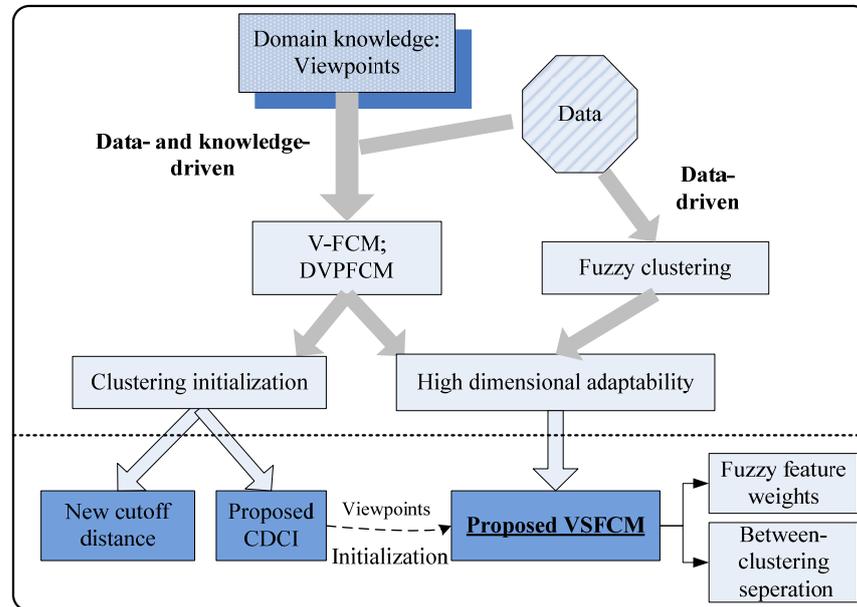

Figure 1. The idea of the proposed VSFCM algorithm.

3.1. Cluster Initialization Method Induced by Cut-Off Distance

Here is a new cluster center initialization method. The following is the corresponding flowchart (Figure 2).

The DPC algorithm is used as the start point, where the local density ρ of the data point and its minimum distance δ between itself and other points with higher local density still use the previous calculation formulas, namely (7) and (9).

In this study, we put forward a new cut-off distance as follows:

$$cd = \min((cd_1 + cd_2)/2, cd_1). \tag{12}$$

Among them,

$$cd_1 = D_{position}, cd_2 = \frac{d_{\max}}{2c}. \tag{13}$$

We arrange d_{kj} in ascending order, and we might as well write the resulting ordered sequence as D ($d_1 \leq d_2 \leq \ldots \leq d_{\max}$). Moreover, n is the number of data points, and c is the number of clusters. d_{kj} is the distance from the data point x_k to x_j. The cut-off distance recommended by the DPC algorithm should make the average number of neighbors of each data point about 1–2% of the total data. So, we choose the cut-off distance according to the upper limit ratio of 2%, which can be taken as $position = round(2\% \times M)$, where $M = \frac{1}{2}n(n-1)$. Hence we get $D_{position}$, namely cd_1. This is the first factor.

Then, the data are divided into c clusters in the form of containing the maximum distance $d_{\max}$ as the radius. Additionally, the reference number is $\frac{d_{\max}}{2c}$, which is recorded as cd_2. In the actual processing, cd_1 and cd_2 are combined as another factor $(cd_1 + cd_2)/2$.

Considering that the average number of neighbors is about 1–2% of the total data, and selecting the smaller one of the two factors, (12) is obtained. Then, naturally, we use the new local density calculation formula:

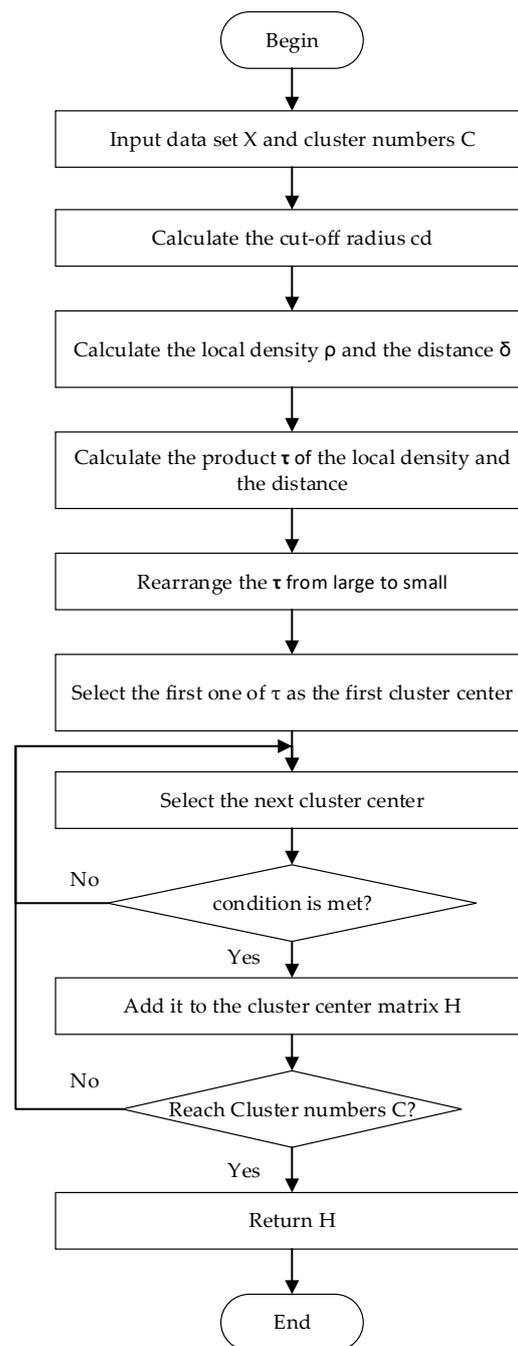$$\rho_k = \sum f(d_{kj} - cd). \tag{14}$$

Figure 2. The flowchart of the initial method CDCI.

Next, we introduce parameters τ_j $(j = 1, \cdots, n)$ to calculate the initial cluster centers directly. The formula is as below:

$$\tau_j = \rho_j \times \delta_j. \tag{15}$$

The traditional DPC algorithm uses the $\rho - \delta$ distribution map to subjectively select the cluster centers, which is easy to cause human error. In contrast, the initialization method we proposed can automatically select a more appropriate cut-off radius so that the selected initial cluster center is closer to the true value. Specifically, we calculate the parameters τ_j $(j = 1, \cdots, n)$ and sort them from small to large, and then select the point of the largest value of τ_j as the first initial cluster center. In the next high-density point selection process, we limit the distance between the current cluster center and other selected cluster centers to

be greater than the cut-off distance, ensuring that we can choose the initial cluster centers and viewpoints more conveniently, efficiently, and accurately.

3.2. The Mechanism of the VSFCM Algorithm

Here, we show the main idea of the viewpoint-driven subspace fuzzy c-means (VS-FCM) algorithm. Its flowchart is shown below (Figure 3).
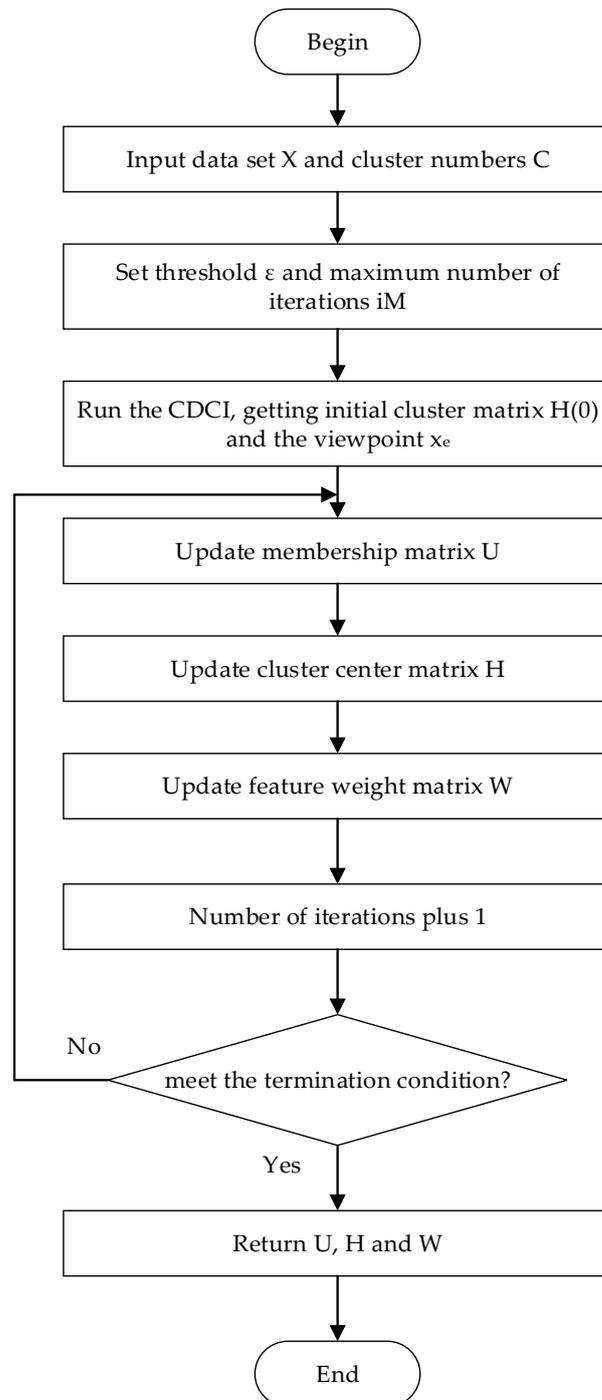

Figure 3. The flowchart of the clustering algorithm VSFCM.

The first initial cluster center selected by the CDCI method is recorded as x_e (i.e., the point of the largest value of τ_j), which is taken as the viewpoint. The position of our viewpoint is constantly changing with iteration. The row number of the viewpoint in the cluster center matrix is $q = \arg(\min(d_{qe}))$ with $d_{qe} = \parallel v_q - x_e \parallel$. That is, we replace the cluster center closest to the viewpoint as the viewpoint.

We use three parts to complete the construction of the objective function. In the first part, the fuzzy feature weight is combined with the classical FCM algorithm, and the cluster centers and the viewpoint are integrated together. The second part is an adaptive fuzzy weight penalty term, which uses the parameter φ_i that can be automatically calculated. The third part represents the inter-cluster separation term, in which the fuzzy feature weight and the cluster center that is integrated with the viewpoint are used, and the centroid of the initialization cluster center serves as the reference point for inter-cluster separation. The objective function is expressed as follows:

$$J_{\text{VSFCM}} = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m \sum_{k=1}^{l} w_{ik}^t \left\| x_{jk} - h_{ik} \right\|^2 + \sum_{i=1}^{c} \varphi_i \sum_{k=1}^{l} w_{ik}^t - \eta \sum_{i=1}^{c} (\sum_{j=1}^{n} u_{ij}^m) \sum_{k=1}^{l} w_{ik}^t \| v_{0k} - h_{ik} \|^2. \tag{16}$$

Among them $(i = 1, \ldots, c)$.

$$h_i = \begin{cases} v_i, & i \neq q \\ x_e, & i = q \end{cases}, v_0 = \sum_{i=1}^{c} v_i / c. \tag{17}$$

The following constraints are imposed $(i = 1, \ldots, c, j = 1, \ldots, n)$

$$\sum_{i=1}^{c} u_{ij} = 1, \quad \sum_{k=1}^{l} w_{ik} = 1. \tag{18}$$

Here, h_i is the cluster center fused with the viewpoint. When $i = q$, h_i is replaced with the point of maximum value of τ, namely the viewpoint. The reference point of separation between clusters is the above (17). The fuzzy weight w_{ik}^t is the weight of the k-th feature of the i-th cluster, in which the fuzzy coefficient t is used, and generally $t > 1$. φ_i is the parameter used to implement the fuzzy weight penalty. Parameter η is used to adaptively adjust the value of the separation term between clusters.

Note that (16) can be transformed into:

$$J = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m \sum_{k=1}^{l} w_{ik}^t [(x_{jk} - h_{ik})^2 - \eta (v_{0k} - h_{ik})^2] + \sum_{i=1}^{c} \varphi_i \sum_{k=1}^{l} w_{ik}^t. \tag{19}$$

The solution process is given below. We use the Lagrangian multiplier method for (19), and it becomes an optimization problem of the following formula:

$$J' = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m \sum_{k=1}^{l} w_{ik}^t [(x_{jk} - h_{ik})^2 - \eta (v_{0k} - h_{ik})^2] + \sum_{i=c}^{c} \varphi_i \sum_{k=1}^{l} w_{ik}^t - \sum_{j=1}^{n} \lambda_j (\sum_{i=1}^{c} u_{ij} - 1) - \sum_{i=1}^{c} \zeta_i (\sum_{k=1}^{l} w_{ik} - 1). \tag{20}$$

In order to minimize (20), the following three partial derivative relations need to be satisfied:

$$\frac{\partial J'}{\partial u_{ij}} = 0, \ \frac{\partial J'}{\partial h_{ik}} = 0, \ \frac{\partial J'}{\partial w_{ik}} = 0, \ i = 1, \ldots, c, \ j = 1, \ldots, n, \ k = 1, \ldots, l. \tag{21}$$

For the convenience of expression, we set

$$D_{ij} = \sum_{k=1}^{l} w_{ik}^t [(x_{jk} - h_{ik})^2 - \eta (v_{0k} - h_{ik})^2]. \tag{22}$$

Firstly, we give the solution process of u_{ij}. Starting from $\frac{\partial J'}{\partial u_{ij}} = 0$, we get

$$\frac{\partial J'}{\partial u_{ij}} = m u_{ij}^{m-1} D_{ij} - \lambda_j = 0. \tag{23}$$

From (23), we obtain

$$u_{ij} = \left(\frac{\lambda_j}{m D_{ij}}\right)^{\frac{1}{m-1}}. \tag{24}$$

Because $1 = \sum\limits_{i=1}^{c} u_{ij} = \sum\limits_{i=1}^{c} \left(\frac{\lambda_j}{m D_{ij}}\right)^{\frac{1}{m-1}}$, it follows that

$$\left(\frac{\lambda_j}{m}\right)^{\frac{1}{m-1}} = \frac{1}{\sum\limits_{l=1}^{c} D_{lj}^{-\frac{1}{m-1}}}. \tag{25}$$

Substituting (25) into (24), we have

$$u_{ij} = \frac{D_{ij}^{-\frac{1}{m-1}}}{\sum\limits_{l=1}^{c} D_{lj}^{-\frac{1}{m-1}}}. \tag{26}$$

The iterative formula of membership degree is obtained, and it involves the value of η. Note that when η is very large, D_{ij} may become negative, which is obviously not what we want. For this reason, we can naturally give the following constraints ($i = 1, \ldots, c$, $j = 1, \ldots, n$, $k = 1, \ldots, l$):

$$(x_{jk} - h_{ik})^2 - \eta(v_{0k} - h_{ik})^2 \geq 0. \tag{27}$$

Therefore, we can get

$$\eta = \alpha_0 \min_{i,j,k} \frac{(x_{jk} - h_{ik})^2}{(v_{0k} - h_{ik})^2}. \tag{28}$$

Here α_0 is a constant, and $\alpha_0 \in [0, 1]$.

Secondly, we show the solution process of h_{ik}. Starting from $\frac{\partial J'}{\partial h_{ik}} = 0$, we have ($i = 1, \ldots, c$, $k = 1, \ldots, l$)

$$\frac{\partial J'}{\partial h_{ik}} = 2 \sum_{j=1}^{n} u_{ij}^m w_{ik}^t \left[(1 - \eta)h_{ik} - x_{jk} + \eta v_{0k}\right] = 0. \tag{29}$$

It can be further calculated, and we can get

$$(1 - \eta) \sum_{j=1}^{n} u_{ij}^m w_{ik}^t h_{ik} = \sum_{j=1}^{n} u_{ij}^m w_{ik}^t \left(x_{jk} - \eta v_{0k}\right). \tag{30}$$

Regarding the solution of h_{ik}, there are two cases:
Case 1: From (17) we know that $h_{ik} = x_{ek}$ when $i = q$ ($k = 1, \ldots, l$).
Case 2: When $i \neq q$, we can get from (30):

$$h_{ik} = v_{ik} = \frac{\sum\limits_{j=1}^{n} \left(x_{jk} - \eta v_{0k}\right) u_{ij}^m}{\sum\limits_{j=1}^{n} (1 - \eta) u_{ij}^m}. \tag{31}$$

In summary, we can get

$$
h_{ik} = \begin{cases} x_{ek}, & i = q, \\ \dfrac{\sum\limits_{j=1}^{n} (x_{jk} - \eta v_{0k}) u_{ij}^m}{\sum\limits_{j=1}^{n} (1-\eta) u_{ij}^m}, & i \neq q. \end{cases}
$$

(32)

Finally, we provide the solving process of w_{ik}. Starting from $\frac{\partial J'}{\partial w_{ik}} = 0$ ($i = 1, \ldots, c$, $k = 1, \ldots, l$), we can get

$$
\frac{\partial J'}{\partial w_{ik}} = t w_{ik}^{t-1} \left(\sum_{j=1}^{n} u_{ij}^m [(x_{jk} - h_{ik})^2 - \eta (v_{0k} - h_{ik})^2] + \varphi_i \right) - \zeta_i = 0.
$$

(33)

Similarly, for the convenience of expression, we set

$$
T_{ik} = \sum_{j=1}^{n} u_{ij}^m [(x_{jk} - h_{ik})^2 - \eta (v_{0k} - h_{ik})^2] + \varphi_i.
$$

(34)

From (33), we obtain

$$
w_{ik} = \left(\frac{\zeta_i}{t T_{ik}} \right)^{\frac{1}{t-1}}.
$$

(35)

Note that

$$
1 = \sum_{k=1}^{l} w_{ik} = \sum_{k=1}^{l} \left(\frac{\zeta_i}{t T_{ik}} \right)^{\frac{1}{t-1}}.
$$

(36)

From (36), we have

$$
\left(\frac{\zeta_i}{t} \right)^{\frac{1}{t-1}} = \frac{1}{\sum\limits_{p=1}^{l} T_{ip}^{-\frac{1}{t-1}}}.
$$

(37)

Substituting (37) into (35), we obtain

$$
w_{ik} = \frac{T_{ik}^{-\frac{1}{t-1}}}{\sum\limits_{p=1}^{l} T_{ip}^{-\frac{1}{t-1}}}.
$$

(38)

Among them $t \in (1, +\infty)$, φ_i is a penalty term parameter, which reflects the contribution of each attribute to the cluster centers. Moreover, the selection of its value is critical to the performance of the clustering. In actual processing, we can define φ_i as the ratio of the sum of the previous part of (19) and the fuzzy feature weight:

$$
\varphi_i = K \frac{\sum\limits_{j=1}^{n} u_{ij}^m D_{ij}}{\sum\limits_{k=1}^{l} w_{ik}^t}
$$

(39)

Here K is a positive constant.

So far, the derivation process of cluster centers, membership degree matrix, and weight matrix of the VSFCM algorithm have been fully explained.

3.3. Framework of the VSFCM Algorithm

The execution process of the CDCI method and the VSFCM algorithm is shown respectively in Algorithms 1 and 2.

Algorithm 1 Cut-off Distance-induced Clustering Initialization (CDCI)

Input: Data set $X = \{x_k\}_{k=1}^N$, number of clusters C.

Output: cluster center matrix $H = \{h_i\}_{i=1}^C$.

procedure CDCI (Data X, Number C)

1. $H = [\ ]$;
2. Calculate the cut $-$ off radius cd according to (12);
3. Calculate the local density ρ_j of each point according to (14);
4. Calculate the distance δ_j of each point according to (9);
5. Calculate τ_j according to (15);
6. Rearrange $\tau = \{\tau_j\}_{j=1}^n$ from large to small, and get the corresponding X'

after the original data set X is re-sorted by τ;

7. Select τ_1 corresponding to x_1' as the first cluster center x_e, and let $H = H \cup x_e$;
8. Let $tt = 1$, $k = 2$;
9. Repeat
10.   while $\|x_k' - H\| < cd$

           //If the distance between x_k' and selected cluster center is

smaller than cd,

           //then skip directly

11. $k = k + 1$;
12. $H = H \cup x_k'$;
13. $tt = tt + 1$;
14. Until $tt = C$
15. return H
16. **end procedure**

Algorithm 2 Viewpoint-driven subspace fuzzy C-means (VSFCM) algorithm

---

Input: Data set $X = \{x_j\}_{j=1}^n$, number of clusters C.

Output: Membership matrix $U = \{u_{ij}\}_{i,j=1}^{c,n}$, cluster center matrix $H = \{h_i\}_{i=1}^c$, weight matrix $W = \{w_{ik}\}_{i,k=1}^{c,n}$.

procedure DVPFCM (Data X, Number C)

1. Set threshold ε and maximum number of iterations iM;
2. Run Algorithm 1, and get $H^{(0)}$ and the point x_e with the highest density;
3. Do
4. $iter = iter + 1$;
5.   Update $U^{(iter)} = [u_{ij}]$ by calculating memberships u_{ik} using (26);
6.   Update $H^{(iter)} = [h_i]$ by calculating centers h_i using (32);
7.   Update $W^{(iter)} = [w_{ik}]$ by calculating weights w_{ik} using (38);
8. while $\left\| H^{(iter)} - H^{(iter-1)} \right\| \geq \varepsilon$ and $iter \leq iM$;
9. return $U^{(iter)}$, $H^{(iter)}$, $W^{(iter)}$;

end procedure

---

4. Experimental Results

In this section, we validate the clustering ability of the proposed VSFCM algorithm through a series of experiments. In the comparative experiment section, five relevant algorithms are selected, including V-FCM, SCAD, ESSC, FRFCM, and DVPFCM. Among the two algorithms of SCAD, the structure of the first one is more complex and closer to the algorithm in this paper, so we chose it to compare with ours. In terms of initialization methods, we compare our algorithm with the HDCCI algorithm and present it visually.

The testing data sets include two artificial data sets, 10 UCI machine-learning data sets, and the Olivetti face database. The artificial data sets DATA1 and DATA2 are composed of Gaussian distribution points obtained by generation tools. The tested UCI data sets [34]

include Iris, Wireless Indoor Localization, Wine, Breast Cancer Wisconsin, Seeds, Letter Recognition (A, B), Ionosphere, SPECT heart data, Aggregation, and Zoo. These UCI data sets are popular and representative of the field of machine learning. The Olivetti face database corresponds to a collection of 40 people and 10 pictures per person. We select 200 pictures corresponding to 20 people from it, with a total of 1024 attributes.

Table 2 counts the basic information of two artificial data sets and 10 UCI machine learning data sets, which include a total number of instances, features, and reference clusters. For all experiments, the default values are selected for the parameters, and the specific settings are as follows: $m = 2$, $t = 2$, $\varepsilon = 10^{-5}$. For convenience, the selection of cd in the CDCI algorithm is calculated based on (12).

Table 2. Testing data sets.

ID	Name	Instances	Features	Clusters
D1	Iris	150	4	3
D2	Wireless Indoor Localization	2000	7	4
D3	Wine	178	13	3
D4	Breast cancer	569	30	2
D5	Seeds	210	7	3
D6	Letter Recognition (A, B)	1155	16	2
D7	Ionosphere	351	33	2
D8	SPECT heart data	267	22	2
D9	Aggregation	788	2	7
D10	Zoo	101	17	7
D11	DATA1	300	2	3
D12	DATA2	180	3	3

4.1. Evaluation Indicator

We use two types of evaluation indicators: hard clustering effectiveness indicators and soft clustering effectiveness indicators. Moreover, the superscript "(+)" indicates that the larger the value of the indicator, the better the clustering performance. The superscript "(−)" indicates the opposite meaning.

Since fuzzy clustering algorithms divide data points into the clusters with the highest corresponding membership, hard clustering indicators can also be used to evaluate their clustering effects.

The hard clustering effectiveness indicators selected are the following three kinds.

(1)   Classification rate

The classification rate (CR) [32] reflects the proportion of data that are correctly classified. The formula is:

$$\mathrm{CR}^{(+)} = \frac{\sum\limits_{i=1}^{C} e_i}{n}. \tag{40}$$

in which e_i is the number of objects found correctly in the i-th cluster, and n is the number of all objects in the data set.

(2)   Normalized mutual information

Normalized mutual information (NMI) [35] reflects the statistical information shared between two clusters:

$$\mathrm{NMI}^{(+)}(W, V) = \frac{\sum\limits_{i=1}^{I} \sum\limits_{j=1}^{J} q(i,j) \log \frac{q(i,j)}{q(i)q(j)}}{\sqrt{G(W)G(V)}}. \tag{41}$$

Here W, V are the two distributions of the data set. Supposing that W, V have I and J clusters, respectively. In addition, $q(i) = \frac{|w_i|}{n}$, and $|W_i|$ is the number of objects contained

in the cluster W_i. $q(i,j) = \frac{|W_i \cap V_j|}{n}$, and $G(W) = -\sum\limits_{i=1}^{I} q(i) \log q(i)$. The structure of the calculation formula of $G(V)$ is similar.

(3)   Calinski–Harabasz indicator

The Calinski–Harabasz (CH) indicator [36] is a measure from the perspective of distance within a cluster and dispersion between clusters:

$$\text{CH}^{(+)} = \frac{\sum\limits_{i=1}^{c} n_i \times d(v_i, \overline{v})}{c-1} \Bigg/ \frac{\sum\limits_{i=1}^{c} \sum\limits_{k=1}^{n} d(x_k, v_i)}{n-c}. \tag{42}$$

Here n_i corresponds to the number of objects contained in cluster i, and $\overline{v}$ is the average of the cluster centers.

The following two kinds of soft clustering effectiveness indicators are used.

(4)   The extension indicator of ARI

The EARI [37] indicator is a fuzzy extension of the adjusted rand indicator (ARI) [38,39], and its purpose is to describe the similarity of two clustering results.

Assuming that R and Q are two hard partitions, corresponding to k and v clusters, respectively, there are some definitions in the ARI indicator as follows:

e represents the number of pairs of data belonging to the same class in R and to the same cluster in Q meanwhile.

f represents the number of pairs of data belonging to the same class in R and to the different clusters in Q meanwhile.

g represents the number of pairs of data belonging to the different classes in R and to the same cluster in Q meanwhile.

h represents the number of pairs of data belonging to the different classes in R and to the different clusters in Q meanwhile.

According to the above definition, given two membership matrices B1 and B2, it is obvious that e, f, g, and h can be rewritten as below when r and q are two soft partitions:

$$e = |W \cap Y| = \sum_{j_2=2}^{n} \sum_{j_1=1}^{j_2-1} W(j_1, j_2) \otimes Y(j_1, j_2), \tag{43}$$

$$f = |W \cap Z| = \sum_{j_2=2}^{n} \sum_{j_1=1}^{j_2-1} W(j_1, j_2) \otimes Z(j_1, j_2), \tag{44}$$

$$g = |X \cap Y| = \sum_{j_2=2}^{n} \sum_{j_1=1}^{j_2-1} X(j_1, j_2) \otimes Y(j_1, j_2), \tag{45}$$

$$h = |X \cap Z| = \sum_{j_2=2}^{n} \sum_{j_1=1}^{j_2-1} X(j_1, j_2) \otimes Z(j_1, j_2). \tag{46}$$

Among them,

$$W = \left\{ W(j_1, j_2) \Big| W(j_1, j_2) = \oplus_{i=1}^{k} \left(r_{ij_1} \otimes r_{ij_2} \right), j_2 = 2, \cdots, N, j_1 = 1, \cdots, j_2 - 1 \right\} \tag{47}$$

is the set of data pairs belonging to a cluster in B1. Similarly, Y is the set of data pairs belonging to the same cluster in B2. And

$$X = \left\{ X(j_1, j_2) \Big| X(j_1, j_2) = \oplus_{i_1, i_2 \in [1,k] | i_1 \neq i_2} \left(r_{i_1 j_1} \otimes r_{i_2 j_2} \right), j_2 = 2 \cdots N, j_1 = 1 \cdots j_2 - 1 \right\} \tag{48}$$

is a set of data pairs that do not belong to the same cluster in B1. Similarly, Z is a collection of data pairs that do not belong to the same cluster in B2. $\otimes, \oplus$ are t-norm and s-norm,

and min, and max are often used for actual processing, respectively. EARI is specifically obtained by the following formula:

$$EARI^{(+)} = \frac{e - \frac{(e+f)(e+g)}{e+f+g+h}}{\frac{(e+f)+(e+g)}{2} - \frac{(e+f)(e+g)}{e+f+g+h}}. \tag{49}$$

(5) Xie–Beni indicator

The Xie–Beni (XB) indicator [40] is a highly recognized indicator of the effectiveness of fuzzy clustering, whose formula is as below:

$$XB = \frac{\sum\limits_{i=1}^{c} \sum\limits_{k=1}^{n} u_{ik}^{m} d(x_k, v_i)}{n \times \min_{i \neq j} d(v_i, v_j)}. \tag{50}$$

4.2. Artificial Data Sets

Table 2 gives the basic information of all the data sets used in the experiment. Let us first discuss the artificial data sets DATA1 and DATA2. Figure 4 is a data distribution diagram of DATA1.
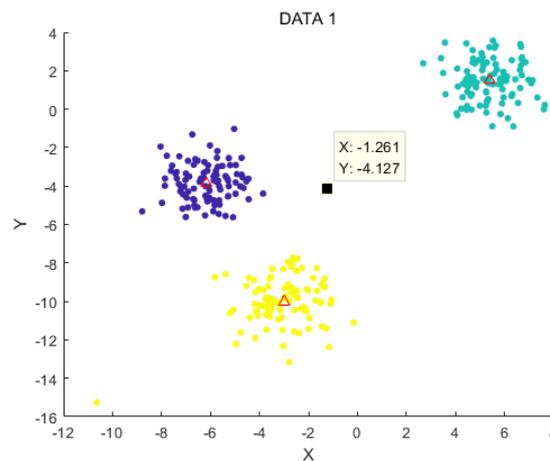

Figure 4. DATA1 data distribution map.

As shown in Figure 4, the clusters of three colors correspond to three classes of the DATA1, the red triangle "$\triangle$" represents the cluster center of each cluster, and the black solid square"$\blacksquare$" represents the reference point for inter-cluster separation. The cluster centers are $V_1 = [-6.1545, -3.8668]$, $V_2 = [5.3800, 1.4988]$, and $V_3 = [-3.0078, -10.0121]$. After calculation, we get $X_0 = [-1.2809, -4.1416]$, $V_0 = [-1.2607, -4.1267]$. We express the weight of the full space as $w_1' = w_2' = w_3' = [0.5, 0.5]$, three subspace weights are expressed as $w_1 = [0.9, 0.1], w_2 = [0.4, 0.6], w_3 = [0.6, 0.4]$. The corresponding separation between clusters can be expressed respectively as:

$$J_{\text{full\_space}} = \sum_{i}^{3} ((w_{i1}')^2 (V_{i1} - V_{01})^2 + (w_{i2}')^2 (V_{i2} - V_{02})^2) = 34.36 \tag{51}$$

$$J_{\text{ESSC\_subspace}} = \sum_{i}^{3} ((w_{i1})^2 (V_{i1} - X_{01})^2 + (w_{i2})^2 (V_{i2} - X_{02})^2) = 44.38 \tag{52}$$

$$J_{\text{VSFCM\_subspace}} = \sum_{i}^{3} ((w_{i1})^2 (V_{i1} - V_{01})^2 + (w_{i2})^2 (V_{i2} - V_{02})^2) = 44.49 \tag{53}$$

Among them, $v_0 = \sum\limits_{i=1}^{c} v_i/c$, $x_0 = \sum\limits_{j=1}^{n} x_j/n$. Here v_i is obtained by using our proposed CDCI method.

In this example, the separation between clusters in the subspace is significantly greater than the separation between clusters in the full space, which means that the subspace clustering has higher inter-cluster separation and better clustering effect than the full-space clustering. In particular, the separation between full-space clusters and the separation between subspace clusters can also be used as separations under different distance metrics.

Compared with the ESSC algorithm, which is also a subspace algorithm, our algorithm has been improved to some extent. That is, replace x_0 with v_0 in its formula. Moreover, the effect of subspace separation of our algorithm is slightly better than it, and ours is more stable. When faced with more complex data sets or more scattered and irregular data sets, our algorithms can still maintain a good result. In a word, our proposed algorithm can get better separation between clusters, which can get better results.

Figure 5 is the distribution diagram of each dimension of DATA2, and Table 3 is the distribution of each algorithm on DATA2.
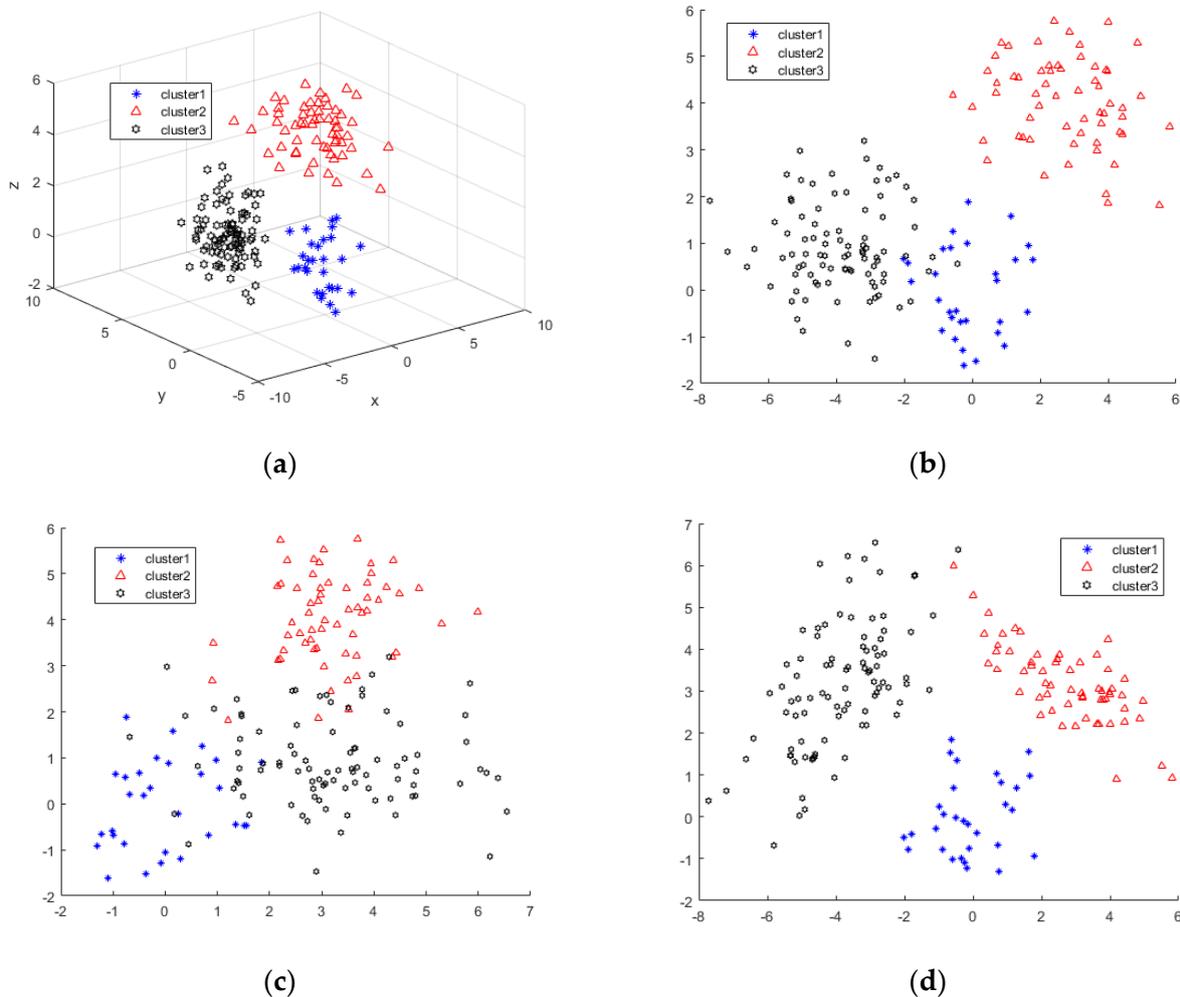

(a)

(b)

(c)

(d)

Figure 5. (**a**) DATA2 data distribution map. (**b**) DATA2: x-y dimensional data distribution map. (**c**) DATA2: x-z data distribution map. (**d**) DATA2: y-z data distribution map.

Table 3. Dimensional weight distribution of algorithms on DATA2.

Algorithms	Clusters	Weight of the Detected Three Features in Each Cluster		
VSFCM	Cluster1	0.4394	0.2980	0.2626
	Cluster2	0.3645	0.3285	0.3068
	Cluster3	0.4380	0.2089	0.3531
FRFCM		0.3860	0.1960	0.4180
ESSC	Cluster1	0.3433	0.3321	0.3246
	Cluster2	0.3896	0.2573	0.3532
	Cluster3	0.3299	0.3400	0.3301
SCAD1	Cluster1	0.3284	0.3417	0.3299
	Cluster2	0.3500	0.3096	0.3404
	Cluster3	0.3393	0.3327	0.3280

From Table 3, we can find that the weight of the ESSC and SCAD1 algorithm does not fluctuate very much around 0.3333, but our VSFCM algorithm has a large difference in the attribution of weights. Taking Cluster 1 as an example, three weights, 0.3433, 0.3321, and 0.3246 of ESSC, are very close. However, the weight values obtained by the VSFCM algorithm are 0.4394, 0.2980, and 0.2626. From the VSFCM algorithm, it can be seen that the first feature has a significant contribution to the clustering result, followed by the second feature, and the worst is the third feature.

Table 4 shows the clustering results of several algorithms on two artificial data sets, DATA1 and DATA2. Obviously, the performance of our algorithm is better than other algorithms. Moreover, mainly due to the contribution of weight distribution, the values of the five clustering indicators of the subspace clustering algorithms are obviously higher than that of other algorithms.

Table 4. Clustering results on two artificial data sets.

Datasets	Algorithms	CH(+)	NMI(+)	EARI(+)	CR(+)	XB(−)
DATA1	V-FCM	26.1104	0.8529	0.8959	0.9400	0.3092
	DVPFCM	26.2273	0.8997	0.9517	0.9667	0.2579
	SCAD1	27.1581	0.9488	0.9795	0.9867	0.1760
	ESSC	27.1549	<u>0.9702</u>	<u>0.9954</u>	0.9800	<u>0.1716</u>
	FRFCM	<u>27.1592</u>	**0.9830**	0.9944	<u>0.9933</u>	0.1739
	VSFCM	**27.1597**	**0.9830**	**0.9957**	**0.9967**	**0.1663**
DATA2	V-FCM	12.7069	0.7933	0.7735	0.8333	0.4738
	DVPFCM	12.9600	0.9264	0.8037	0.8333	0.4636
	SCAD1	13.6113	<u>0.9368</u>	0.9484	0.9667	0.3002
	ESSC	14.1086	0.9222	0.9714	0.9778	<u>0.3000</u>
	FRFCM	<u>14.1692</u>	<u>0.9368</u>	<u>0.9793</u>	<u>0.9833</u>	0.3330
	VSFCM	**14.2811**	**0.9534**	**0.9858**	**0.9889**	**0.2881**

The bold represents the best value among all algorithms, and the underlined represents the second best.

4.3. UCI Data Sets

The UCI data sets adopted here include Iris, Wireless Localization, Wine, Seeds, breast cancer Wisconsin, letter recognition (A, B), SPECT heart data, Aggregation, and Zoo. Among them, Breast Cancer data is a common medical data set in machine learning, which can be divided into two clusters. Figure 6 shows its Sammon mapping, where the yellow and purple point sets represent two classes of the dataset respectively.
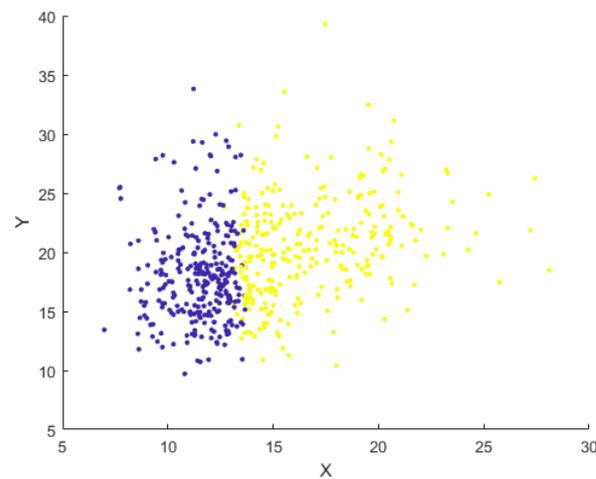
Figure 6. Breast Cancer data (Sammon mapping).

Figure 7 shows the $\rho - \delta$ distribution diagram of Breast Cancer of two cluster center initialization methods. In Figure 7, we can see that the first initial cluster center is well determined, that is, the data point in the upper right corner. In the selection process of the second initial cluster center, the boundary between the optional points of the CDCI algorithm and other data points is very clear, while that of the HDCCI algorithm is fuzzy. So, when we use the HDCCI algorithm to select the initial cluster center, it is easy to choose the wrong one. In contrast, the initial cluster centers selected by our cluster center initialization method CDCI are more in line with the characteristics of the ideal cluster centers.
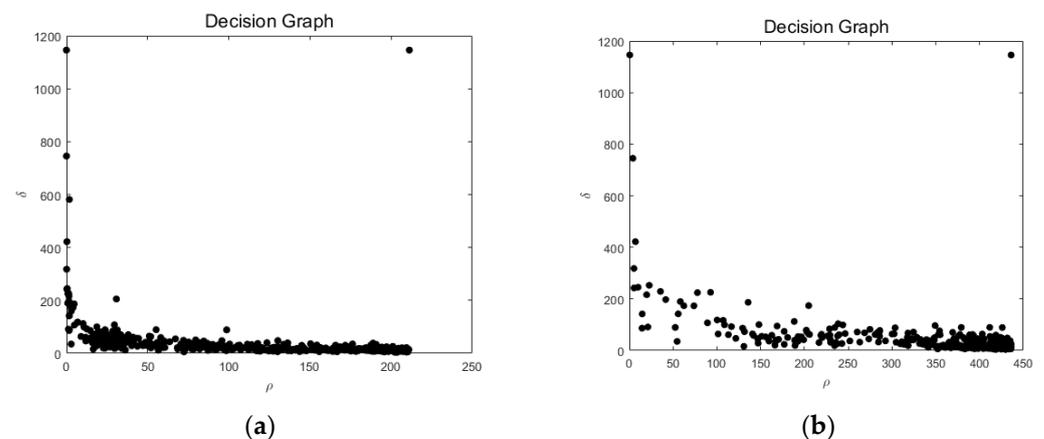

(a)

(b)

Figure 7. (**a**) CDCI $\rho - \delta$ distribution map. (**b**) HDCCI $\rho - \delta$ distribution map.

Table 5 describes the average value of the results of each clustering algorithm (including V-FCM, SCAD, ESSC, FRFCM, DVPFCM, and our algorithm) running 20 times on the selected UCI data sets. The adopted evaluation indicators here are the above-mentioned three hard indicators (CR, CH, and NMI) and two fuzzy indicators (EARI and XB). According to the above five clustering indicators, the performance of the proposed VSFCM algorithm is evaluated and compared with the existing three subspace clustering algorithms and two fuzzy clustering algorithms combined with viewpoints. In order to observe the results more conveniently, we bold the best result and underlined the second-best result.

Table 5. Clustering results on UCI data sets.

Data Sets	Algorithms	CH(+)	NMI(+)	EARI(+)	CR(+)	XB(−)
Iris	V-FCM	11.8364	0.8498	0.9312	0.9533	0.2786
	DVPFCM	11.8423	0.8705	0.9400	0.9600	0.2731
	SCAD1	11.5894	0.7277	0.8712	0.8800	0.3205
	ESSC	11.6442	0.7578	0.8861	0.9000	0.3117
	FRFCM	11.6853	0.7665	0.8895	0.9067	0.2880
	VSFCM	**12.1360**	**0.8801**	**0.9568**	**0.9667**	**0.2628**
Wireless Indoor Localization	V-FCM	66.8676	0.8113	0.8730	0.9325	0.4845
	DVPFCM	67.0487	0.8117	0.8733	0.9330	0.4764
	SCAD1	79.9004	0.8508	0.9478	0.9445	0.4262
	ESSC	80.1303	0.8549	0.9494	0.9460	0.4240
	FRFCM	80.3587	0.8620	0.9529	0.9465	0.4113
	VSFCM	**81.6420**	**0.8981**	**0.9658**	**0.9605**	**0.3988**
Wine	V-FCM	8.8531	0.6500	0.8598	0.8596	1.7169
	DVPFCM	12.3937	0.6748	0.8712	0.8708	1.5531
	SCAD1	12.5666	0.7710	0.8728	0.9270	1.4794
	ESSC	12.8969	0.7710	0.8773	0.9270	1.2942
	FRFCM	13.3068	0.7955	0.9031	0.9382	1.2786
	VSFCM	**14.4425**	**0.8212**	**0.9520**	**0.9494**	**1.2302**
Breast	V-FCM	73.4053	0.6555	0.8850	0.9385	0.3178
	DVPFCM	74.0104	0.6555	0.8850	0.9385	0.3146
	SCAD1	108.1275	0.6320	0.9250	0.9297	0.2913
	ESSC	108.7064	0.6507	0.9400	0.9332	0.2903
	FRFCM	107.3007	0.6106	0.9244	0.9262	0.2940
	VSFCM	**110.5897**	**0.7084**	**0.9825**	**0.9420**	**0.2779**
Seeds	V-FCM	16.0498	0.6423	0.8450	0.8762	0.6296
	DVPFCM	16.0668	0.6545	0.8478	0.8810	0.5999
	SCAD1	16.1163	0.6654	0.8487	0.8857	0.2508
	ESSC	16.1378	0.6795	0.8580	0.8905	0.2493
	FRFCM	16.2337	0.7026	0.8601	0.8952	0.2280
	VSFCM	**16.7541**	**0.7173**	**0.8745**	**0.9095**	**0.2269**
Ionosphere	V-FCM	19.9278	0.1036	0.4159	0.6809	0.7634
	DVPFCM	21.5964	0.1036	0.4180	0.6809	0.6566
	SCAD1	32.2422	**0.1320**	0.4762	**0.7094**	0.4649
	ESSC	31.6597	0.1271	0.4732	0.7066	0.5057
	FRFCM	32.1711	0.1292	0.4738	0.7066	0.4804
	VSFCM	**32.3488**	**0.1320**	**0.4764**	**0.7094**	**0.4635**
Letter Recognition (A, B)	V-FCM	155.3391	0.7197	0.8948	0.9408	0.6633
	DVPFCM	156.9256	**0.7249**	0.8965	**0.9415**	0.6561
	SCAD1	195.6678	0.6655	0.9181	0.9293	0.6486
	ESSC	198.4907	0.7189	0.9353	0.9395	0.6352
	FRFCM	196.1744	0.6917	0.9288	0.9344	0.6390
	VSFCM	**199.5879**	**0.7249**	**0.9385**	**0.9415**	**0.6333**
Aggregation	V-FCM	11.9580	0.9279	0.9755	0.9315	0.2790
	DVPFCM	11.9818	0.9350	0.9769	0.9365	0.2787
	SCAD1	11.8053	0.8666	0.9585	0.8985	0.2978
	ESSC	11.9525	0.8991	0.9740	0.9061	0.2803
	FRFCM	11.8163	0.9197	0.9805	0.9302	0.2763
	VSFCM	**12.0721**	**0.9460**	**0.9924**	**0.9429**	**0.2643**
Zoo	V-FCM	1.1496	0.7864	0.8973	0.6832	0.2878
	DVPFCM	1.1752	0.8029	0.9357	0.7030	0.2528
	SCAD1	1.1858	0.8146	0.9672	0.7228	0.2636
	ESSC	1.1915	0.8286	0.9795	0.7129	0.2493
	FRFCM	1.2085	0.8319	0.9768	0.7228	0.2237
	VSFCM	**1.2194**	**0.8324**	**0.9807**	**0.7426**	**0.2170**

The bold represents the best value among all algorithms, and the underlined represents the second best.

As can be seen from Table 5, the six algorithms can be divided into three grades, the worst is V-FCM and DVPFCM, the performance of FRFCM and ESSC is equivalent to or better than that of SCAD1, and the best is the proposed VSFCM algorithm. Although DVPFCM and V-FCM algorithms mentioned are generally inferior to the other three algorithms, they can achieve better clustering performance on the Iris and Letter Recognition (A, B) data sets measured by NMI and CR, respectively. This shows that for all data sets, no one algorithm is always better than others.

Through comparison, we further notice that the best clustering performance indicated by NMI and CR is not always consistent with the best clustering performance indicated by other indicators. That is, the clustering performance with higher NMI and CR values does not necessarily possess higher XB, CH, and EARI values. Therefore, it is necessary to comprehensively evaluate the performance of clustering algorithms with different metrics.

From Table 5, the following conclusions can be drawn:

First of all, due to the low dimension of small data sets, which contributes to the small impact of weights on the results, V-FCM and DVPFCM (belonging to viewpoint-oriented fuzzy clustering algorithms) can get better results on these data sets. In addition, the viewpoint can guide the clustering algorithm to run in a more correct direction, so the NMI and CR indicators of V-FCM and DVPFCM perform better on some data sets.

Secondly, for the weighted fuzzy clustering algorithms, SCAD, ESSC, and FRFCM with more complex structures have a great advantage in processing multi-dimensional data sets, owing to the efficiency and effect of clustering improved by the weight distribution.

Finally, our proposed algorithm VSFCM is obviously superior to the other algorithms mentioned above. Because our algorithm well integrates the advantages of the two types of algorithms and successfully improves the clustering effect.

In general, our proposed VSFCM algorithm is more ideal for obtaining initial cluster centers and has better performance in various clustering evaluation indicators.

Table 6 reflects the dimension weight distribution results of the subspace algorithm carried out on the Wine data set, presenting in the order of the weight values of each cluster corresponding to the 13 features in the data. It can be seen from Table 6 that different subspace clustering algorithms may have different degrees of importance to the same feature. Similar results can be obtained with other UCI data sets. For multi-dimensional data sets, the reasonable distribution of dimension weights can improve the accuracy and efficiency of clustering, so our proposed VSFCM algorithm can achieve more ideal clustering results.

Table 6. Wine's ranking results of the detected 13 features in each cluster.

Algorithms	Clusters	Ranking Results of the Detected 13 Features in Each Cluster												
VSFCM	Cluster1	12	6	9	11	5	4	2	8	7	3	1	10	13
	Cluster2	4	13	6	5	10	8	2	12	11	1	9	7	3
	Cluster3	10	12	5	7	6	8	1	13	9	11	4	3	2
FRFCM		10	2	13	12	11	7	3	6	9	5	8	4	1
ESSC	Cluster1	12	6	2	9	5	8	3	10	7	4	1	11	13
	Cluster2	9	13	2	4	3	7	5	12	6	11	8	10	1
	Cluster3	10	11	4	5	3	9	2	13	6	1	7	12	8
SCAD1	Cluster1	12	6	2	8	5	9	4	10	7	3	1	11	13
	Cluster2	9	13	2	4	3	7	5	12	6	11	8	10	1
	Cluster3	10	11	4	5	2	9	3	13	6	1	7	12	8

4.4. The Olivetti Face Database

The 400 pictures in the Olivetti face database were collected in different places, with different light intensities and different emotional states of the same person (eyes open or closed, smiling or not smiling). Our experiment uses the first 20 people's pictures.

Moreover, Figure 8 shows the histogram of its running results. Moreover, from Table 7, we can see the clustering results of different relevant algorithms on this database.
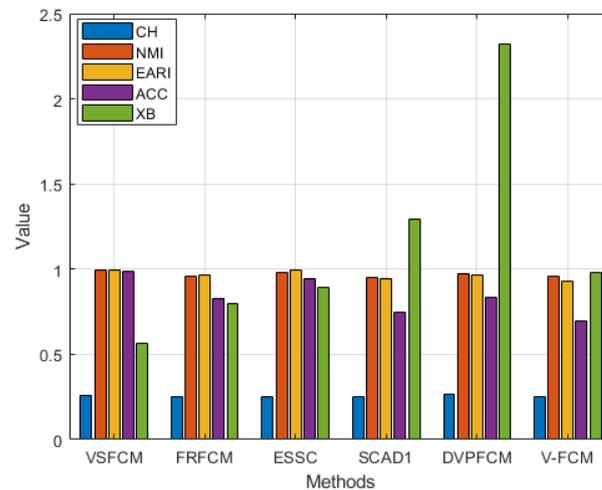

Figure 8. Histogram of the Olivetti face database running results.

Table 7. Clustering results of the Olivetti face database.

Data Sets	Algorithms	V-FCM	DVPFCM	SCAD1	ESSC	FRFCM	VSFCM
	CH(+)	0.2527	**0.2664**	0.2524	0.2520	0.2512	<u>0.2570</u>
	NMI(+)	0.9573	0.9700	0.9516	<u>0.9828</u>	0.9591	**0.9956**
face	EARI(+)	0.9292	0.9654	0.9418	**0.9975**	0.9667	<u>0.9964</u>
	CR(+)	0.6950	0.8350	0.7500	<u>0.9450</u>	0.8300	**0.9850**
	XB(−)	0.9769	2.3237	1.2936	0.8954	<u>0.7964</u>	**0.5651**

The bold represents the best value among all algorithms, and the underlined represents the second best.

In Table 7, the CR value of our VSFCM algorithm on this data set can even reach 0.9850, while that of FRFCM, ESSC, SCAD1, DVPFCM, and V-FCM are significantly lower, which are 0.8300, 0.9450, 0.7500, 0.8350, and 0.6950, respectively. As can be seen from the histogram in Figure 8, the indicator values of our algorithm are obviously better than other algorithms. When processing Olivetti's face database, the indicators CR, NMI, and EARI of our algorithm are all close to 1. It reveals that the weight distribution of our algorithm plays an important role in the process of high-dimensional data (200×1024), which includes reducing the weight of some less important dimensions while increasing the weight of relatively important dimensions to improve the accuracy and efficiency of the algorithm. Moreover, the proposed algorithm introduces the separation term between clusters to increase the distance between clusters, and the initial cluster centers obtained by using the cluster center initialization method CDCI are closer to the real ones, which prevents the VSFCM algorithm from falling into the local optimal value or iterative divergence.

4.5. Time Complexity

Tables 8 and 9, respectively, present the average number of iterations of each algorithm and the average calculation time per run (each algorithm was executed 50 times). The number in parentheses indicates the ranking of the algorithm, and Arank indicates the average ranking of the algorithm. Note that there is no iterative process in the HDCCI algorithm, so its relevant statistical data is not needed here.

Table 8. Average number of iterations of each algorithm.

	V-FCM	DVPFCM	SCAD1	ESSC	FRFCM	VSFCM
D1	18 (3)	51 (6)	16 (2)	20 (4)	7 (1)	20 (4)
D2	37 (5)	121 (6)	12 (2)	13 (3)	10 (1)	19 (4)
D3	10 (2)	26 (6)	19 (5)	16 (4)	7 (1)	15 (3)
D4	7 (2)	22 (6)	13 (4)	12 (3)	5 (1)	16 (5)
D5	42 (4)	104 (5)	15 (3)	15 (3)	8 (1)	12 (2)
D6	10 (2)	56 (6)	16 (4)	19 (5)	11 (3)	9 (1)
D7	9 (2)	20 (5)	16 (4)	46 (6)	6 (1)	12 (3)
Arank	2.9 (2)	5.7 (6)	3.4 (4)	4 (5)	**1.3 (1)**	3.1 (3)

The bold represents the best value among all algorithms, and the underlined represents the second best.

Table 9. Average running time of each algorithm.

	V-FCM	DVPFCM	SCAD1	ESSC	FRFCM	VSFCM
D1	0.4836 (6)	0.1836 (5)	0.1189 (4)	0.0345 (1)	0.0934 (2)	0.1150 (3)
D2	4.7661 (5)	11.2978 (6)	1.0110 (3)	0.5590 (1)	1.2039 (4)	0.9489 (2)
D3	0.0473 (1)	0.1282 (5)	0.1839 (6)	0.0763 (2)	0.1146 (4)	0.0960 (3)
D4	0.3458 (2)	0.6581 (5)	0.6268 (4)	0.2555 (1)	1.0447 (6)	0.5627 (3)
D5	0.1581 (6)	0.0473 (1)	0.1011 (3)	0.0796 (2)	0.1507 (5)	0.1287 (4)
D6	0.1549 (1)	0.4161 (4)	0.5214 (5)	0.2669 (3)	1.6541 (6)	0.2050 (2)
D7	2.1087 (5)	4.2933 (6)	1.1066 (3)	1.3857 (4)	1.0721 (2)	0.6442 (1)
Arank	3.7 (3)	4.6 (6)	4 (4)	**2 (1)**	4.1 (5)	2.6 (2)

The bold represents the best value among all algorithms, and the underlined represents the second best.

It can be seen from the following two tables that the VSFCM algorithm ranks third in terms of the average number of iterations and second in terms of average iteration time. Compared with other clustering algorithms, the number of iterations and iteration time of the VSFCM algorithm required are generally less. This shows that the convergence speed of our proposed algorithm is relatively fast.

4.6. Discussion

This study proposes a new and more effective cut-off distance, using high-density points to achieve the initialization of the cluster center and the selection of viewpoints. This can effectively restrain the interference of noise (because the ideal cluster center has a large value of the ρ and δ, but the δ value of the noise point is small). Therefore, the cluster center initialization method CDCI in this study is better than the DPC and the HDCCI (the initialization method of DVPFCM).

In addition, we test the performance of our proposed VSFCM algorithm and other algorithms on two artificial data sets, 10 UCI data sets, and the Olivetti face database. Through comparison, it is found that the proposed VSFCM algorithm performs better than V-FCM, DVPFCM, SCAD1, ESSC, and FRFCM in five indicators, and the distribution of feature weights of our algorithm is more in line with the characteristics of the data.

Finally, we compare the average number of iterations and iteration time of each algorithm and discover that the number of iterations and iteration time required by the VSFCM algorithm is generally less, which leads to the conclusion that the VSFCM algorithm has a better convergence speed.

From a deeper perspective, the proposed VSFCM algorithm has better performance due to the following aspects:

- First, the proposed cluster center initialization method CDCI can get the initial cluster center close to the real data structure through the new cut-off distance. This not only prevents the iterative divergence of the algorithm but also speeds up the convergence of the algorithm and helps improve the accuracy of the algorithm;
- Second, as a part of the data structure of the VSFCM algorithm, the viewpoint serves for the knowledge-induced clustering process. Moreover, it is combined with data to form

a cluster driving force driven by both knowledge and data, inducing the clustering algorithm to obtain more accurate clustering results. Moreover, the viewpoint is introduced into the subspace fuzzy clustering, which can better guide the clustering process of each cluster subspace;

- Third, the fuzzy weight is introduced, which can overcome the influence of outliers on cluster analysis and speed up the convergence of clustering. Our algorithm adds weights to each data as a whole to indicate the degree of contribution to clustering and assigns smaller weights to noise points to reduce their participation in clustering, thus weakening their impact on clustering results;
- Finally, the separation term between clusters is introduced into the objective function. The reference point for separation between clusters is the average value of the initial cluster centers, which is combined with the distribution of weights to improve the accuracy of the algorithm by enhancing the ability to process data spatial distance.

Compared with the ESSC in [28], the proposed VSFCM algorithm has the following advantages:

- The reference point for inter-cluster separation has been improved. In this study, v_0 is used instead of x_0. When we initialize the cluster center, outliers are automatically removed, which weakens their influence of them on the reference points so that each cluster can be separated more clearly;
- We use the viewpoint to guide the convergence of each projection subspace of soft subspace clustering, which conforms to the data structure of clustering and accelerates the clustering process.

Compared with the DVPFCM of [32], the proposed VSFCM algorithm has the following advantages:

- The cut-off distance is optimized. The selection of cut-off distance is very important and directly affects the clustering accuracy of the algorithm. Generally, the cut-off distance should be selected so that the sphere space of the cut-off radius contains 1% to 2% of the total number of datasets. However, the calculation formula of the cut-off radius of HDCCI has no corresponding rigorous scientific proof, which makes it difficult to initialize the cluster center completely in line with the concept of DPC. In contrast, the improved cut-off distance achieved better results while meeting the DPC standards;
- Fuzzy weights are introduced. By assigning smaller weights to features with a larger proportion of noise points and outliers, the influence of them on clustering results can be indirectly weakened;
- Subspace clustering is employed. When faced with high-dimensional data, traditional clustering algorithms have no effective processing measures. However, the algorithm combined with subspace clustering has better adaptability to it. By continuously adjusting the weight of each subspace, the contribution of each subspace to the clustering can be accurately described, making the clustering results better.

5. Conclusions

This paper has proposed a new subspace clustering algorithm, which is named viewpoint-driven subspace fuzzy c-means algorithm (VSFCM), and has achieved good clustering results. The main work and contributions of this paper are summarized as follows:

First of all, in view of the problem that a large number of clustering algorithms are sensitive to the initialization of cluster centers, we propose a new cut-off distance under the system of the DPC algorithm and further provide a cut-off distance-induced clustering initialization method CDCI, which is used as an initialization strategy for cluster centers and also served as a new viewpoint selection strategy. It makes the initial cluster center closer to the true cluster center, which not only improves the clustering convergence speed, but also promotes the clustering accuracy to a certain extent.

Secondly, by taking the viewpoint obtained by CDCI as being reflective of domain knowledge, the fuzzy clustering idea driven by knowledge and data is proposed. Subsequently, we introduce the subspace clustering mode and fuzzy feature weight processing mechanism and propose the separation formula between clusters, which is optimized with the viewpoint. On the basis of these, we establish the viewpoint-driven subspace fuzzy c-means algorithm (VSFCM). The viewpoint in it helps guide the clustering algorithm to discover the real data structure and thus get better clustering results. Additionally, the introduced separation term between clusters can maximize the distance between cluster centers in real time, achieve the maximum separation between clusters, and also optimize the internal mechanism of clustering.

Finally, by applying our proposed VSFCM algorithm and comparison algorithms (V-FCM, SCAD, ESSC, FRFCM, and DVPFCM) to the three types of data sets, we can see that the proposed VSFCM algorithm performs best in terms of the five indicators, and exhibits stronger stability.

Our clustering algorithm has a certain degree of noise resistance [5,41,42], but its effectiveness is not outstanding and can be further improved in the future. Moreover, there is room for improvement in the number of iterations and iteration time of our algorithm, which can be reduced by optimizing our algorithm structure and parameter settings.

In addition, the work of this study can be extended to the clustering of granular information [30,43–45], which will offer new directions for data analysis. Moreover, we can develop our fuzzy clustering algorithm in the field of fuzzy reasoning [46,47] and carry on clustering for the fuzzy rules.

Author Contributions: Conceptualization, Y.T. and R.C.; methodology, Y.T. and B.X.; software, R.C. and B.X.; validation, Y.T., R.C. and B.X.; formal analysis, Y.T.; investigation, R.C.; resources, B.X.; data curation, R.C.; writing—original draft preparation, R.C. and B.X.; writing—review and editing, Y.T.; visualization, R.C.; supervision, Y.T.; project administration, R.C.; funding acquisition, Y.T. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

Abbreviations

Following are two abbreviation lists corresponding to the algorithm and symbols presented in this manuscript, respectively.

Abbreviations	Full Name of Algorithms
FCM	Fuzzy c-means algorithm.
DPC	Clustering by fast search and finding of density peaks.
HDCCI	Hypersphere density-based clustering centers initialization.
V-FCM	A knowledge-driven fuzzy clustering algorithm with viewpoints.
DVPFCM	Density viewpoint-induced possibilistic fuzzy c-means algorithm.
SCAD1	Simultaneous clustering and attribute discrimination—version 1.
ESSC	Enhanced soft subspace clustering.
FRFCM	A feature-reduction FCM.
VSFCM	Viewpoint-driven subspace fuzzy c-means algorithm.
Symbols	**Interpretations**
cd	Cut-off distance.
ρ	Local density of the data point.
δ	Minimum distance between itself and other points with higher local density.
τ	Product of ρ and δ.
c	Number of cluster centers.
n, l	Numbers and dimensions of the data, respectively.
X	Data set.
x_e	Viewpoint.
m, t	Fuzzy coefficients for membership degree and fuzzy weight, respectively.
h_i	Cluster center, $i = 1, 2, ..., c$.
u_{ij}	Membership degree, $i = 1, 2, ..., c$, $j = 1, 2, ..., n$.
w_{ik}	Fuzzy weight, $i = 1, 2, ..., c$, $k = 1, 2, ..., l$.
φ_i	Parameter for adaptive fuzzy weight penalty term, $i = 1, 2, ..., c$.
η	Parameter for adjusting the value of the separation term between clusters.
v_0	Reference point for inter-cluster separation.
D_{ij}	For the convenience of expression, $D_{ij} = \sum\limits_{k=1}^{l} w_{ik}^t \left[\left(x_{jk} - h_{ik} \right)^2 - \eta \left(v_{0k} - h_{ik} \right)^2 \right]$.
T_{ik}	For the same reason, $T_{ik} = \sum\limits_{j=1}^{n} u_{ij}^m [(x_{jk} - h_{ik})^2 - \eta (v_{0k} - h_{ik})^2] + \delta_i$.

References

1. Zhou, T.; Qiao, Y.Y.; Salous, S.; Liu, L.; Tao, C. Machine Learning-Based Multipath Components Clustering and Cluster Characteristics Analysis in High-Speed Railway Scenarios. *IEEE Trans. Antennas Propag.* **2022**, *70*, 4027–4039. [CrossRef]
2. Tang, Y.; Chen, R.; Xia, B. Viewpoint-driven Subspace Fuzzy C-Means Algorithm. In *Artificial Intelligence Logic and Applications*; Chen, Y., Zhang, S., Eds.; AILA 2022, Communications in Computer and Information Science; Springer: Singapore, 2022; Volume 1657, pp. 91–105.
3. Josephine, D.C.J.; Wise, D.C.J.W.; Verunathi, A.R.; SheelaLavanya, J.M.; SterlinRani, D.; Saravanan, K.G. A Novel Approach of Applying Rank Ordering Clustering Algorithm in Agricultural Data. In Proceedings of the 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 20–22 January 2022.
4. Lin, Z.P.; Kang, Z.; Zhang, L.D.; Tian, L. Multi-View Attributed Graph Clustering. *IEEE Trans. Knowl. Data. Eng.* **2023**, *35*, 1872–1880. [CrossRef]
5. Tang, Y.; Huang, J.; Pedrycz, W.; Li, B.; Ren, F. A fuzzy cluster validity index induced by triple center relation. *IEEE Trans. Cybern.* **2023**. [CrossRef]
6. Charan, J.D.; Keerthivasan, P.; Sakthiguhan, D.; Lanitha, B.; Sundareswari, K. An Effective Optimized Clustering Algorithm for Social Media. In Proceedings of the 2022 International Conference on Edge Computing and Applications (ICECAA), Tamilnadu, India, 13–15 October 2022.
7. Gong, Y.D.; Guo, X.Y.; Lai, G.M. A Centralized Energy-Efficient Clustering Protocol for Wireless Sensor Networks. *IEEE Sens. J.* **2023**, *23*, 1623–1634. [CrossRef]
8. Rodriguez, A.; Laio, A. Clustering by fast search and find of density peaks. *Science* **2014**, *344*, 1492–1496. [CrossRef]
9. Gong, J.M.; Wu, C.C.; Guo, L.F.; Liu, W.; Pei, M.J. An image fusion algorithm based on fuzzy C-means clustering. In Proceedings of the 14th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Changsha, China, 15–16 January 2022.
10. Hu, S. Fuzzy C-means Clustering Algorithm and Its Application in Mental Health Intelligent Evaluation System. In Proceedings of the 2022 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS), Dalian, China, 11–12 December 2022.
11. Husnah, M.; Novita, R. Clustering of Customer Lifetime Value with Length Recency Frequency and Monetary Model Using Fuzzy C-Means Algorithm. In Proceedings of the 2022 International Conference on Informatics Electrical and Electronics (ICIEE), Yogyakarta, Indonesia, 5–7 October 2022.

12. Kumar, D.; Agrawal, R.K.; Kumar, P. Bias-Corrected Intuitionistic Fuzzy C-Means With Spatial Neighborhood Information Approach for Human Brain MRI Image Segmentation. *IEEE Trans. Fuzzy Syst.* **2022**, *30*, 687–700. [CrossRef]

13. Barrion, M.H.C.; Bandala, A.A. Modified Fast and Robust Fuzzy C-means Algorithm for Flood Damage Assessment using Optimal Image Segmentation Cluster Number. In Proceedings of the 2023 17th International Conference on Ubiquitous Information Management and Communication (IMCOM), Seoul, Republic of Korea, 3–5 January 2023.

14. Sathyamoorthy, M.; Kumar, R.; Vanitha, C.N.; Sharma, B.; Syamraj, V.; Chowdhury, S. An Efficient Integrated approach of Fuzzy C-Means Map Reduce for Weather Forecasting Data Collection. In Proceedings of the 2023 6th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 3–4 March 2023.

15. Wang, X.Z.; Wang, Y.D.; Wang, L.J. Improving fuzzy c-means clustering based on feature-weight learning. *Pattern Recognit. Lett.* **2004**, *25*, 1123–1132. [CrossRef]

16. Li, J.; Gao, X.B.; Jiao, L.C. A new feature weighted fuzzy clustering algorithm. In Proceedings of the 10th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, Regina, SK, Canada, 31 August–3 September 2005.

17. Chan, E.Y.; Ching, W.K.; Ng, M.K.; Huang, J.Z. An optimization algorithm for clustering using weighted dissimilarity measures. *Pattern Recognit.* **2004**, *37*, 943–952. [CrossRef]

18. Jing, L.P.; Ng, M.K.; Xu, J.; Huang, J.Z. Subspace clustering of text documents with feature weighting k-means algorithm. In Proceedings of the 9th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining, Hanoi, Vietnam, 18–20 May 2005.

19. Gan, G.J.; Wu, J.H.; Yang, Z.J. A fuzzy subspace algorithm for clustering high dimensional data. In Proceedings of the 2nd International Conference on Advanced Data Mining and Applications, Xi'an, China, 14–16 August 2006.

20. Gan, G.J.; Wu, J.H. A convergence theorem for the fuzzy subspace clustering (FSC) algorithm. *Pattern Recognit.* **2008**, *41*, 1939–1947. [CrossRef]

21. Zong, P.; Jiang, J.Y.; Qin, J. Study of High-Dimensional Data Analysis based on Clustering Algorithm. In Proceedings of the 15th International Conference on Computer Science & Education (ICCSE), Delft, The Netherlands, 18–22 August 2020.

22. Alam, A.; Muqeem, M. Integrated k-means Clustering with Nature Inspired Optimization Algorithm for the Prediction of Disease on High Dimensional Data. In Proceedings of the 2022 International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 16–18 March 2022.

23. Chakraborty, S.; Das, S. Detecting Meaningful Clusters From High-Dimensional Data: A Strongly Consistent Sparse Center-Based Clustering Approach. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 2894–2908. [CrossRef]

24. Lu, C.Y.; Feng, J.S.; Lin, Z.C.; Mei, T.; Yan, S.C. Subspace clustering by block diagonal representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 487–501. [CrossRef] [PubMed]

25. Lee, Y.-H.; Kim, S.-J.; Lee, K.Y.; Nam, T. Online Robust Subspace Clustering with Application to Power Grid Monitoring. *IEEE Access* **2023**, *11*, 27816–27828. [CrossRef]

26. Zeng, S.; Duan, X.; Li, H.; Bai, J.; Tang, Y.; Wang, Z. A Sparse Framework for Robust Possibilistic K-Subspace Clustering. *IEEE Trans. Fuzzy Syst.* **2023**, *31*, 1124–1138. [CrossRef]

27. Frigui, H.; Nasraoui, O. Unsupervised learning of prototypes and attribute weights. *Pattern Recognit.* **2004**, *37*, 567–581. [CrossRef]

28. Deng, Z.H.; Choi, K.S.; Chung, F.L.; Wang, S.T. Enhanced soft subspace clustering integrating within-cluster and between-cluster information. *Pattern Recognit.* **2010**, *43*, 767–781. [CrossRef]

29. Yang, M.S.; Nataliani, Y. A feature-reduction fuzzy clustering algorithm based on feature-weighted entropy. *IEEE Trans. Fuzzy Syst.* **2018**, *26*, 817–835. [CrossRef]

30. Tang, Y.; Pan, Z.; Pedrycz, W.; Ren, F.; Song, X. Viewpoint-Based Kernel Fuzzy Clustering with Weight Information Granules. *IEEE Trans. Emerg. Top. Comput. Intell.* **2023**, *7*, 342–356. [CrossRef]

31. Pedrycz, W.; Loia, V.; Senatore, S. Fuzzy clustering with viewpoints. *IEEE Trans. Fuzzy Syst.* **2010**, *18*, 274–284. [CrossRef]

32. Tang, Y.; Hu, X.; Pedrycz, W.; Song, X. Possibilistic fuzzy clustering with high-density viewpoint. *Neurocomputing* **2019**, *329*, 407–423. [CrossRef]

33. Bezdek, J.C. *Pattern Recognition with Fuzzy Objective Function Algorithms*; Kluwer Academic Publishers: Norwell, MA, USA, 1981.

34. UCI Machine Learning Repository. Available online: http://archive.ics.usi.edu/ml/datasets.html (accessed on 5 March 2022).

35. Strehl, A.; Ghosh, J. Cluster ensembles-A knowledge reuse framework for combining multiple partitions. *J. Mach. Learn. Res.* **2002**, *3*, 583–617.

36. Calinski, T.; Harabasz, J. A dendrite method for cluster analysis. *Commun. Stat.* **1972**, *3*, 1–27.

37. Campello, R.J.G.B. A fuzzy extension of the Rand index and other related indexes for clustering and classification assessment. *Pattern Recognit Lett.* **2007**, *28*, 833–841. [CrossRef]

38. Hubert, L.; Arabie, P. Comparing partitions. *J. Classif.* **1985**, *2*, 193–218. [CrossRef]

39. Huang, H.C.; Chuang, Y.Y.; Chen, C.S. Multiple kernel fuzzy clustering. *IEEE Trans. Fuzzy Syst.* **2012**, *20*, 120–134. [CrossRef]

40. Xie, X.L.; Beni, G. A validity measure for fuzzy clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **1991**, *13*, 841–847. [CrossRef]

41. Pan, X.; Zhang, X.; Jiang, Z.; Wang, S. Anti-noise possibilistic clustering based on maximum entropy. In Proceedings of the 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), Nanjing, China, 24–26 November 2017.

42. Li, W.; Zhu, H.; Liu, W.; Chen, D.; Jiang, J.; Jin, Q. An Anti-Noise Process Mining Algorithm Based on Minimum Spanning Tree Clustering. *IEEE Access* **2018**, *6*, 48756–48764. [CrossRef]

43. Raj, A.; Minz, S. Spatial Granule based Clustering Technique for Hyperspectral Images. In Proceedings of the 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon), Mysuru, India, 16–17 October 2022.

44. Cordovil, L.A.Q.; Coutinho, P.H.S.; de Bessa, I.V.; D'Angelo, M.F.S.V.; Palhares, R.M. Uncertain Data Modeling Based on Evolving Ellipsoidal Fuzzy Information Granules. *IEEE Trans. Fuzzy Syst.* **2020**, *28*, 2427–2436. [CrossRef]

45. Stalder, F.; Denzler, A.; Mazzola, L. Towards Granular Knowledge Structures: Comparison of Different Approaches. In Proceedings of the 2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI), Herl'any, Slovakia, 21–23 January 2021.

46. Marsala, C.; Bouchon-Meunier, B. Explainable Fuzzy Interpolative Reasoning. In Proceedings of the 2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Padua, Italy, 18–23 July 2022.

47. Laha, M.; Konar, A.; Nagar, A.K. Olfactory Perceptual-Ability Assessment by Near-Infrared Spectroscopy Using Vertical-Slice Based Fuzzy Reasoning. *IEEE Access* **2023**, *11*, 17779–17792. [CrossRef]