# Viewpoint-Based Kernel Fuzzy Clustering With Weight Information Granules

Yiming Tang, *Member, IEEE*, Zhifu Pan, Witold Pedrycz, *Life Fellow, IEEE*, Fuji Ren, *Senior Member, IEEE*, and Xiaocheng Song

***Abstract*—Domain knowledge can be introduced into fuzzy clustering with the aid of information granules, embodied by the concept of viewpoints. For such kind of fuzzy clustering methods, the strategy of acquisition of viewpoints has not been fully developed. Furthermore a way of determining the related information granules deserves more attention. Having these problems in mind, in this study, the density Viewpoint-based Weighted Kernel Fuzzy Clustering (VWKFC) algorithm is proposed. First, the kernel-based hypersphere density initialization (KHDI) algorithm is presented as a certain prerequisite, in which the kernel distance is utilized instead of the Euclidean distance. Besides, a novel density radius is put forward. Second, the concept of the weight information granule is established, which incorporates two parts. The feature weight matrix is provided, where different weights are assigned to different features to reduce the influence of unrelated features. Meanwhile a sample weight is assigned to each data point, thus the influence of noise and outliers on clustering can be reduced to a certain extent. Third, the data point with the highest local density obtained by KHDI is regarded as the density viewpoint. Then we combine kernel mechanism, density viewpoints, weight information granules and a maximum entropy regularization to design the VWKFC algorithm, and prove its convergence. Experimental results validate that VWKFC is superior over eight related clustering algorithms with regard to five evaluation indexes, especially when processing high-dimensional data. It has been shown that VWKFC makes the selection of initialized cluster centers and viewpoints more reasonable, and obtains better clustering results, and achieves higher convergence speed.**

***Index Terms*—Feature weighting, fuzzy c-means, fuzzy clustering, granular computing, information granules.**

Manuscript received 17 January 2022; revised 2 July 2022 and 10 August 2022; accepted 19 August 2022. Date of publication 2 September 2022; date of current version 27 March 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFC1523100, in part by the National Natural Science Foundation of China under Grants 62176083, 62176084, 61877016, and 61976078, in part by the Key Research and Development Program of Anhui Province under Grant 202004d07020004, in part by the Natural Science Foundation of Anhui Province under Grant 2108085MF203, and in part by the Fundamental Research Funds for Central Universities of China under Grant PA2021GDSK0092. *(Corresponding author: Witold Pedrycz.)*

Yiming Tang is with the Anhui Province Key Laboratory of Affective Computing and Advanced Intelligent Machine, School of Computer and Information, Hefei University of Technology, Hefei, Anhui 230601, China, and also with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6R 2V4, Canada (e-mail: tym608@163.com).

Zhifu Pan and Xiaocheng Song are with the School of Computer and Information, Hefei University of Technology, Hefei, Anhui 230601, China (e-mail: tofu6010@163.com; xcsong120@163.com).

Witold Pedrycz is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6R 2V4, Canada, and also with the Systems Research Institute, Polish Academy of Sciences, 01-447 Warsaw, Poland (e-mail: wpedrycz@ualberta.ca).

Fuji Ren is with the School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan 610056, China (e-mail: ren@is.tokushima-u.ac.jp).

This article has supplementary downloadable material available at https://doi.org/10.1109/TETCI.2022.3201620, provided by the authors.

Digital Object Identifier 10.1109/TETCI.2022.3201620

## I. Introduction

CLUSTERING divides similar data points into a single group (cluster) as far as possible through the similarity calculation of data points in the dataset, and makes the data among distinct clusters as different as possible [1], [2], [3], [4]. As an unsupervised machine learning algorithm, it has the evident advantage of processing a large amount of data without prior training, so it plays an important role in many fields [5], [6], [7], [8].

In the early days, clustering analyses mainly referred to hard clustering. The hard clustering algorithm strictly classified data objects into a certain cluster. Recently, Rodriguez and Laio [9] proposed a clustering algorithm RLM to quickly find the peak density, which was a typical hard clustering algorithm.

In addition to hard clustering, another important clustering method is fuzzy clustering, which is a generalization of hard clustering. In 1969, Ruspini introduced fuzzy sets into clustering analyses, which resulted in the fuzzy clustering method. Among many fuzzy clustering algorithms, the Fuzzy C-Means (FCM) algorithm [10], [11], [12] was most widely used. By minimizing the objective function, the FCM algorithm conveniently calculated the cluster centers and the membership degree of each data object. In this way, it made the same cluster as compact as possible and separated different clusters as far as possible. However, the FCM algorithm was very sensitive to the initialization of cluster centers. Meanwhile a small number of outliers and noise points in the dataset could easily affect its clustering results.

Later, many scholars proposed improvements to the generic version of FCM. Krishnapuram and Keller [13] proposed the Possibilistic C-Means (PCM) algorithm, which removed the constraint of membership degrees in FCM. It reduced the influence of possible noise and outliers. However, PCM was easy to produce the problem of consistent clustering. In [14], a new possibilistic fuzzy clustering algorithm called PFCM was proposed by combining possibilistic value and membership degree, which eliminated the clustering consistency problem of PCM. In [15], [16], [17], the kernel function was introduced into clustering, which was used to replace Euclidean distance for calculation

2471-285X © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Chulalongkorn University provided by UniNet. Downloaded on January 17,2024 at 08:32:14 UTC from IEEE Xplore. Restrictions apply.

and the Kernel-based FCM (KFCM) algorithm was proposed. Chen and Zhang [15] employed the idea of kernel method to calculate the distance between the sample and the cluster center in fuzzy clustering. They calculated these distances using the Gaussian kernel function (instead of the Euclidean distance) and showed that they were more robust to noise and outliers. Zhou et al. [18] raised the maximum-entropy-regularized weighted fuzzy c-means (EWFCM) algorithm, which introduced a feature weight matrix to represent the weights of different features. Yang and Nataliani [19] proposed a feature weight entropy-based feature reduction FCM (FRFCM) algorithm, which used a feature reduction mechanism. This made that clustering irrelevant features could not continue to play an important role in the clustering process. However, EWFCM and FRFCM were still sensitive to the initialization of cluster centers and easily affected by noise, although they realized the selection of features in the clustering process. Verma et al. [20] introduced the particle swarm optimization (PSO) idea into FCM, and put forward a hybrid FCM-PSO (H-FCM-PSO for short) algorithm by combining the advantages of FCM and PSO. It avoided local minima trapping problem since the centroid of its cluster was computed by using PSO through candidate solution.

Domain knowledge can be introduced into fuzzy clustering with the aid of information granules. The Viewpoint-based Fuzzy C-Means (V-FCM) algorithm was proposed by Pedrycz et al. [21]. Here the viewpoints were the entry points of domain knowledge, and the core mechanism here embodied the processing of information granules. Usually, the viewpoint was determined by the user. Through the setting of viewpoint, users could observe the structure of data from some suitable perspective. The V-FCM algorithm realized the control of the cluster centers in the iterative process through the viewpoint and led to good results. Following this point, Tang et al. [22] proposed the Density Viewpoint-induced Possibilistic Fuzzy C-Means (DVPFCM) algorithm, which presented the Hypersphere Density-based Clustering Center Initialization (HDCCI) method to generate the cluster centers by looking for the density peak points. It overcame the sensitivity to the initial cluster center and took the local cluster center with the largest density as the viewpoint. This approach made the cluster center less susceptible to noise and outliers.

There are two key problems in the current viewpoint-based fuzzy clustering methods.

- The extraction strategy of viewpoints is imperfect. How to choose ideal viewpoints becomes a critical problem. The viewpoints in V-FCM were artificially specified and appeared to be arbitrary to some extent. Although the DVPFCM algorithm utilized the RLM algorithm as a strategy to extract viewpoints, it still appeared rough in terms of specific processing techniques.
- In the original expression, the description and calculation of information granules are not elaborate enough. In particular, a more detailed expression mode is needed for the effect of each feature attribute of the data under the environment of information granules.

Aiming at these problems, in this study, we put forward the density Viewpoint-based Weighted Kernel Fuzzy Clustering (VWKFC) algorithm. Fig. 1 shows the main idea of the method. Its originality is embodied as follows. 1) A novel computing method is presented for density viewpoints, in which a new density radius is provided. 2) The kernel distance is employed instead of the Euclidean distance to calculate local density of data points, and then a more reasonable initialization strategy of the cluster centers is established. 3) The concept of the weight information granule is established, which includes a feature weight matrix for different features and a sample weight for each data point.

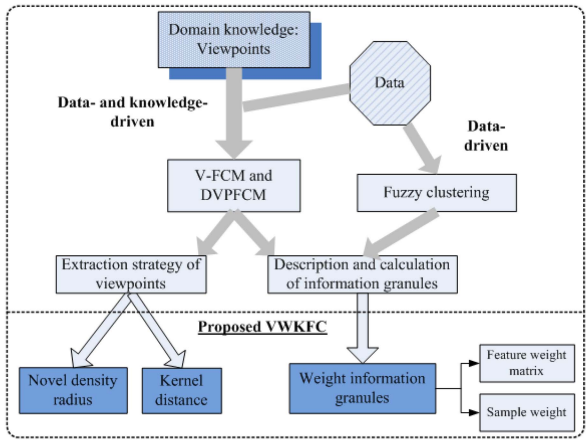


Fig. 1. Overall research idea of the VWKFC algorithm.

The paper is arranged as below. In Section II, we propose the Kernel-based Hypersphere Density Initialization (KHDI) algorithm as a strategy of pretreatment, and then come up with a new weighted fuzzy clustering algorithm called VWKFC, which is driven by the kernel mechanism, density viewpoints as well as weight information granules. Section III shows the experimental studies to verify the performance of the VWKFC algorithm. Section IV gives a summary and outlook.

## II. The VWKFC Algorithm

### A. The KHDI Algorithm

Most of the previous fuzzy clustering algorithms are sensitive to the initialization of the cluster centers, and are easy to be disturbed by noise (or outliers). It is difficult to find correct clustering structure when processing high-dimensional datasets. To overcome the sensitivity to the initialization of the cluster centers, here we put forward a new cluster center initialization method.

Assume that the dataset is $X = \{x_j\}_{j=1}^{N}$. Commonly the Euclidean distance is taken as the non-similarity measurement between different data points, namely $d_{ij} = \|x_i - x_j\|^2$ $(i, j \in \{1, 2, \ldots, N\})$. Given that in some cases it is difficult to find a proper function to partition the data in the original space, then it is difficult to obtain the ideal effect by using the Euclidean distance. Then a good guidance tool is the kernel function, which can transform the data points in the original feature space into the higher-dimensional feature space.

Authorized licensed use limited to: Chulalongkorn University provided by UniNet. Downloaded on January 17,2024 at 08:32:14 UTC from IEEE Xplore. Restrictions apply.

Based on this idea, we use the kernel distance to replace the commonly used Euclidean distance (where $\phi$ is a mapping from the original space to the high-dimensional space, which embodies the kernel function):

$$d_{ij} = \|\phi(x_i) - \phi(x_j)\|^2 = K(x_i, x_i) - 2K(x_i, x_j) + K(x_j, x_j). \tag{1}$$

Here the Gaussian radial basis function (RBF) is adopted for calculation, and the kernel is expressed as

$$K(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}. \tag{2}$$

$\sigma$ is the width parameter to control the range of the kernel function. Because the distribution of this function is the width parameter to control the radial range of the kernel function. The distance is given as

$$d_{ij} = 2 - 2K(x_i, x_j) = 2 - 2e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}. \tag{3}$$

Here the kernel distance is utilized instead of the Euclidean distance, which is able to make the clustering algorithm more robust to noise and outliers [15], [16]. Based on such idea, we propose the Kernel-based Hypersphere Density Initialization (KHDI) algorithm.

Due to the introduction of the Gaussian RBF, to facilitate the calculation, we first need to conduct standardized processing to the data. Here, we adopt range transformation to normalize the data. Suppose that $X = \{x_1, x_2, \ldots, x_N\}$ is the data to be processed. Here for $x_j$ $(j = 1, 2, \ldots, N)$, there is $x_j = \{x_{j1}, x_{j2}, \ldots, x_{jL}\}$, and $L$ is the feature number of the data points. We can get an original data matrix:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1L} \\ \cdots & \cdots & \cdots & \cdots \\ x_{N1} & x_{N2} & \cdots & x_{NL} \end{bmatrix}. \tag{4}$$

For $x_{jl}$ $(j = 1, 2, \ldots, N;\ l = 1, 2, \ldots, L)$ in the matrix, normalization is performed as follows:

$$x'_{jl} = \frac{x_{jl} - \min_k(x_{kl})}{\max_k(x_{kl}) - \min_k(x_{kl})} \ (k \in \{1, 2 \cdots, N\}). \tag{5}$$

Here we reveal the effect of using (5). By means of range transformation, the influence of different variables (features) in the data is eliminated. The features of different orders of magnitude are transformed to the $[0, 1]$ interval, so that all the features play the same role in the clustering process. Another advantage of standardizing is that the data is mapped to the interval of $[0, 1]$ after transformation. Then we only need to set $\sigma$ in the RBF as 1 to achieve good results.

After standardized processing, we calculate the local density $\rho_i$ of each data point $x_i$ with the following formula $(i = 1, 2 \cdots, N)$:

$$\rho'_i = \sum_{j=1}^{N} \chi(d_{ij} - r), \ \rho_i = \frac{\rho'_i - \min_i(\rho'_i)}{\max_i(\rho'_i) - \min_i(\rho'_i)}. \tag{6}$$

Here $\chi(x) = \begin{cases} 1, & \text{if } x < 0 \\ 0, & \text{otherwise} \end{cases}$ and $d_{ij}$ is calculated according to (3) and $r$ denotes the density radius.

In the DVPFCM algorithm, they treat the distribution of the whole dataset as a hypersphere and its diameter is the maximum distance between points. Then the density radius in DVPFCM is calculated by:

$$r = \max(d_{ij})/2C \ (i, j \in \{1, 2 \cdots, N\}). \tag{7}$$

Here $C$ is the number of cluster centers. However, we find that the pivotal distance also need to be taken into account, as it reflects the degree of compactness within the data. Therefore in this study, the value of density radius is changed into:

$$r = (\max(d_{ij}) - \min(d_{ij}))/C \ (i, j \in \{1, 2 \cdots, N\}). \tag{8}$$

Here (8) takes into account both max and min, which is more comprehensive and detailed than (7).

To determine the initial density center, we need to compute the pivotal distance $\delta_i$ derived from the data points with larger local densities, which is expressed as

$$\delta'_i = \min\{d_{ij} | \rho_j > \rho_i, \ j = 1, 2, \ldots, N\}. \tag{9}$$

If $x_k$ has the highest local density, then we employ the pivotal distance as $\delta'_k = \max\{d_{kj} | j = 1, 2, \ldots, N\}$. Then $\delta'$ is also normalized to the interval [0,1], i.e., $\delta_i = \frac{\delta'_i - \min_i(\delta'_i)}{\max_i(\delta'_i) - \min_i(\delta'_i)}$ $(i = 1, 2 \cdots, N)$.

The work here is based on the idea of which the cluster centers are determined by the search of density peak. If a data point $x_j$ is the cluster center, then it should have a high local density $\rho_j$ and a large pivotal distance $\delta_j$. Therefore, we sort the data with the following parameter $\tau_j$:

$$\tau_j = \rho_j \times \delta_j. \tag{10}$$

The larger the parameter $\tau_j$, the more likely that the corresponding data point $x_j$ is the cluster center. If it is not the center point of clustering, it may have a high local density $\rho_j$, but its distance $\delta_j$ will be relatively small. If it is a noise or outlier point, it may have a large pivotal distance $\delta_j$, but its local density $\rho_j$ will be small. As a result, the mechanism of large $\tau_j$ can help resist noise and outliers.

Then, we arrange the data in descending order according to the parameter $\tau_j$, and select the data points that are likely to be the cluster centers. In order to prevent the selected cluster centers from being too close to each other, a sound way is to ensure that the distance between the selected initial cluster centers is greater than the following distance $d_c$:

$$d_c = \frac{\max(d_{ij})}{aC} \ (i, j \in \{1, 2 \cdots, N\}). \tag{11}$$

Here parameter $a$ is used to adjust the distance between centers.

We employ the idea of viewpoints in [21], which was directly provided by the user. But the selection method of viewpoints here is different. In detail, after obtaining the initialized cluster centers, we take the cluster center with the largest parameter $\tau_j$ (namely the first data point in descending order) as a viewpoint $x_d$. Because such data point $x_d$ has a high local density and a large pivotal distance, we can take it as a real cluster center and explore the whole data structure from this point of view.

For the KHDI algorithm, the introduction of the kernel-based distance (instead of the Euclidean one) has a certain effect to

Authorized licensed use limited to: Chulalongkorn University provided by UniNet. Downloaded on January 17,2024 at 08:32:14 UTC from IEEE Xplore. Restrictions apply.

TABLE I
THE KHDI ALGORITHM

**Inputs**: A dataset $X = \{x_j\}_{j=1}^{N}$, cluster number $C$, the width control parameter $\sigma$ of Gaussian RBF.
**Outputs**: Cluster center $V = \{v_i\}_{i=1}^{C}$ and viewpoint $x_d$.
**Procedure**: KHDI (Data $X$, Number $C$)
$V = []$;
The original data are normalized through (5);
The density radius $r$ is determined by binary search method;
The local density of data points is computed by (6);
The pivotal distance is calculated by (9) and normalized to [0,1];
The parameters $\tau_j$ of the data points are computed by (10);
The distance $d_c$ between centers is calculated by (11);
$\tau = \{\tau_j\}_{j=1}^{N}$ is arranged in descending order, and get the corresponding dataset $X'$ according to the descending order of $\tau_j$;
The data point $x'_1$ is selected corresponding to $\tau_1$ , and is taken as the first cluster center $v_1$, and we let $V = V \cup v_1$;
The first selected cluster center is taken as the viewpoint, i.e., $x_d = v_1$;
Set $num = 1$ , $index = 2$;
**repeat**
**while** $||x'_k - V|| < d_c$
$index = index + 1$;
$V = V \cup x'_k$ ;
$num = num + 1$ ;
**until** $num = C$;
**return** $V$, $x_d$;
**end procedure**

resist noise and outliers. Moreover, with the help of (10), the mechanism of large $\tau_j$ allows the viewpoint and the initial cluster centers (obtained by KHDI) to avoid noise and outliers as much as possible. To sum up, on the strength of the kernel-based distance and the mechanism of large $\tau_j$, the KHDI algorithm can reduce the influence of noise and outliers to a certain extent.

The operational flow of KHDI is shown as Table I.

### B. The VWKFC Algorithm

We put forward the concept of the weight information granule. This includes two phases. On the one hand, we give a feature weight matrix, in which different weights are assigned to different features of any data. On the other hand, we provide a sample weight to each data in a dataset. Such structure for dataset is called a weight information granule. As a result, in the process of clustering, the original data is developed into a weight information granule.

Through the KHDI algorithm, we get the initialized cluster centers $V^{(0)}$ and viewpoint $x_d$. Here we use the viewpoint $x_d$ to replace one cluster center. There are several cluster centers, and then we need to determine which cluster center should be replaced by the viewpoint $x_d$. Aiming at such problem, we can find the cluster center $v_q$ closest to viewpoint $x_d$ in the obtained cluster center matrix. Then $x_d$ replaces $v_q$ and occupies the position $q$. As a result, such position $q$ is used to represent the row position of the viewpoint in the cluster center matrix. In detail, we use $q = \arg(\min(d_{qd}))$ to represent the position of the viewpoint in the cluster center matrix. $d_{qd}$ represents the distance between $x_d$ and $v_q$. That is, we find the cluster center $v_q$ with the smallest distance to viewpoint $x_d$, and replace $v_q$ with viewpoint $x_d$. In this case, we calculate it by the kernel distance $d_{qd} = \|\phi(x_d) - \phi(v_q)\|^2$. During the operation of the algorithm, since the value of $q$ will constantly change, the position of the viewpoint will also constantly change.

Here we put forward the density Viewpoint-based Weighted Kernel Fuzzy Clustering (VWKFC) algorithm. Its objective function is shown as follows (where $\phi$ is the mapping in (1)):

$$J = \sum_{i=1, i\neq q}^{C}\sum_{j=1}^{N}\alpha_j u_{ij}^m \sum_{l=1}^{L} w_{il}\|\phi(x_{jl}) - \phi(v_{il})\|^2 + \sum_{j=1}^{N}\alpha_j u_{qj}^m \sum_{l=1}^{L} w_{ql}\|\phi(x_{jl}) - \phi(x_{dl})\|^2 + \gamma^{-1}\sum_{i=1}^{C}\sum_{l=1}^{L} w_{il}\ln w_{il}. \quad (12)$$

Among them,

$$\alpha_j = \frac{1}{C}\sum_{i=1}^{C}\exp\left(-\|x_j - v_i^{(0)}\|\right). \quad (13)$$

And the constraint conditions are ($j = 1, 2, \ldots, N$, $i = 1, 2, \ldots, C$):

$$\sum_{i=1}^{C} u_{ij} = 1;\ \sum_{l=1}^{L} w_{il} = 1;\ w_{il} > 0. \quad (14)$$

In (12), $C$ represents the number of clusters, and $N$ denotes the number of data points to be processed. $L$ is the number of features of the data, and $q$ indicates the row position of the viewpoint. Besides, $u_{ij}$ stands for the membership degree of $j$-th data point $x_j$ to the $i$-th cluster, and its range is $[0, 1]$. $m$ is the fuzzy coefficient, and its value range is $(1, +\infty)$. $w_{il}$ signifies the weight of the feature, indicating the importance of the $l$-dimension feature of the data to the $i$-th cluster center in the clustering process. $\gamma$ is a positive regularized parameter.

On the basis of our experience, we have found that each sample $x_j$ (i.e., the data point $x_j$ in the dataset $X$) may have different importance for the clustering process ($j \in \{1, 2, \ldots, N\}$). For example, if some samples are located at or near the center of a bunch of data, then these samples are obviously more important for clustering. In consequence, we wish to design a weight related to each sample $x_j$ in the objective function of the proposed algorithm. Here we call it the sample weight, which is denoted by $\alpha_j$ for every $x_j$ ($j \in \{1, 2, \ldots, N\}$). Note that $v_i^{(0)}$ represents the $i$-th initial cluster center obtained by KHDI. The initial cluster centers $V^{(0)}$ obtained by KHDI is basically close to the reference cluster centers. Therefore, we can measure the importance of each data point $x_j$ through the initial cluster centers $V^{(0)}$. Note that the kernel distance is used here, which is more exquisite than the Euclidean distance, especially for high-dimensional data. In view of this point of view, we can employ the average kernel distance between $x_j$ and all initial cluster centers $v_i^{(0)}$ to express the importance, i.e., using (13) to characterize the sample weight $\alpha_j$. Then, the closer $x_j$ is to the initial cluster centers, the larger the sample weight $\alpha_j$ is, and the greater the role this point $x_j$ plays in clustering. As a consequence, the weight of noise points and outliers far

Authorized licensed use limited to: Chulalongkorn University provided by UniNet. Downloaded on January 17,2024 at 08:32:14 UTC from IEEE Xplore. Restrictions apply.

away from the initial cluster centers will be reduced as much as possible, and thus the influence and role of noise points and outliers on the clustering result will be greatly weakened.

Note that it is a minimization problem for this objective function expressed by (12), while our goal is to find the minimum value of (12). As for the underlying design idea, this objective function is divided into three items. The first item is calculated when the cluster center is for the case of non-viewpoint, in which it corresponds to the situation of $i \in \{1, 2, \ldots, q-1, q+1, \ldots, C\}$. Here the sample weight $\alpha_j$ and the feature weight $w_{il}$ are utilized, which constitute a weight information granule. The second item is computed when the cluster center is a viewpoint (obtained by the KHDI algorithm), which focuses on the case of $i = q$. It reflects the effect of the viewpoint mentioned above. Therefore, the second item does not need to include the cluster number $C$. As for the third item, i.e., $\gamma^{-1} \sum_{i=1}^{C} \sum_{l=1}^{L} w_{il} \ln w_{il}$, it reflects the negative entropy of feature weight $w_{il}$. Note that the feature weight $w_{il}$ characterizes the probability of the $l$-dimension contributing to clustering results ($i = 1, \ldots, C$; $l = 1, \ldots, L$). The third item can stimulate more dimensions to contribute to the identification of clusters. Through this kind of strategy, we are able to avert the trouble of recognizing clusters by few dimensions when we are faced with sparse data. As a consequence, the third item can optimize the distribution of feature weights, and identify the important features in the data. This makes the objective function more scientific and effective. Moreover, the first and second items express the same connotation, just for different $i$. In fact the first two items can form a whole. Finally, by choosing the appropriate value of $\gamma$, we can balance the first two terms and the third term, so as to achieve a better clustering effect.

Furthermore, we can combine the first term and the second term of (12) and unify the viewpoints and non-viewpoints in the cluster centers to express them in $G = \{g_{il}\}_{i=1,l=1}^{C,L}$. The simplified objective function is expressed as follows:

$$\sum_{i=1}^{C}\sum_{j=1}^{N} \alpha_j u_{ij}^m \sum_{l=1}^{L} w_{il} \|\phi(x_{jl}) - \phi(g_{il})\|^2 + \gamma^{-1} \sum_{i=1}^{C}\sum_{l=1}^{L} w_{il} \ln w_{il}, \tag{15}$$

$$\text{in which} \qquad g_{il} = \begin{cases} x_{il}, & i \neq q, \\ x_{dl}, & i = q. \end{cases} \tag{16}$$

$x_{dl}$ represents the $l$-dimension feature of viewpoint $x_d$ obtained by KHDI.

### C. Iterative Formulas of VWKFC

Through the use of the Lagrange multiplier method, the updated iterative formulas of VWKFC are as follows ($i = 1, 2, \ldots, C$; $j = 1, 2, \ldots, N$; $l = 1, 2, \ldots, L$):

$$u_{ij} = \frac{\left(\sum_{l=1}^{L} w_{il} \|\phi(x_{jl}) - \phi(g_{il})\|^2\right)^{-\frac{1}{m-1}}}{\sum_{s=1}^{C} \left(\sum_{l=1}^{L} w_{sl} \|\phi(x_{jl}) - \phi(g_{sl})\|^2\right)^{-\frac{1}{m-1}}}. \tag{17}$$

TABLE II
THE VWKFC ALGORITHM

**Inputs**: Data $X = \{x_j\}_{j=1}^{N}$, cluster number $C$, $m$, $\sigma$, $\gamma$, iteration stop threshold $\varepsilon$ and maximum number of iterations $iM$.
**Outputs**: Membership matrix $U = \{u_{ij}\}_{i,j=1}^{C,N}$, cluster center matrix $G = \{g_i\}_{j=1}^{C}$, and feature weight matrix $W = \{w_{il}\}_{i,l=1}^{C,L}$.
**Procedure**: VWKFC ($X$, $C$, $m$, $\sigma$, $\gamma$, $\varepsilon$, $iM$)
The cluster centers $G^{(0)}$ is initialized by running KHDI, and the density viewpoint $x_d$ is obtained;
Set $w_{il}$ uniformly as $1/L$ ( $L$ is the feature number of data);
Let $iter = 0$;
**repeat**
$iter = iter + 1$;
Compute $u_{ij}$ by (17), and update $U^{(iter)} = [u_{ij}]$;
Calculate $g_i$ by (19), and renew $G^{(iter)} = [g_i]$;
Figure up $w_{il}$ by (18), and update $W^{(iter)} = [w_{il}]$;
**until** $||G^{(iter)} - G^{(iter-1)}|| < \varepsilon$ or $iter > iM$;
**return** $U^{(iter)}$, $G^{(iter)}$, $W^{(iter)}$;
**end procedure**

$$w_{il} = \frac{\exp\{-\gamma \sum_{j=1}^{N} \alpha_j u_{ij}^m \|\phi(x_{jl}) - \phi(g_{il})\|^2\}}{\sum_{s=1}^{L} \exp\{-\gamma \sum_{j=1}^{N} \alpha_j u_{ij}^m \|\phi(x_{js}) - \phi(g_{is})\|^2\}}. \tag{18}$$

$$g_{il} = \begin{cases} x_{dl}, & i = q, \\ \dfrac{\sum_{j=1}^{N} \alpha_j u_{ij}^m w_{il} e^{-\frac{(x_{jl} - g_{il})^2}{2\sigma^2}} x_{jl}}{\sum_{j=1}^{N} \alpha_j u_{ij}^m w_{il} e^{-\frac{(x_{jl} - g_{il})^2}{2\sigma^2}}}, & i \neq q. \end{cases} \tag{19}$$

We summarize the execution process of the VWKFC algorithm, which is displayed in Table II.

### D. Proof for the Convergence of the VWKFC Algorithm

Zangwill gave a point-to-set function $\Psi : Y \to \mathcal{Q}(Y)$, where $\mathcal{Q}(Y)$ denoted the power set of $Y$ and a closed point-to-set map was defined. But the VWKFC algorithm is a point-to-point function while the "closed" attribute is "continuity" for the situation of point-to-point function. Let $\mathbb{R}$ be the set of all real numbers.

*Theorem 1: (Zangwill's convergence theorem, [23])Let the point-to-point function $\Psi : Y \to \mathcal{Q}(Y)$ derive $\{z_k\}_{k=0}^{\infty}$ with $z_{k+1} = \Psi(z_k)$. Suppose that a solution set $S \subseteq Y$ is provided and*

*i) there exists a continuous function $J : Y \to \mathbb{R}$ such that, if $y \notin S$, then $J(\Psi(y)) < J(y)$, while if $y \in S$, then $J(\Psi(y)) \leq J(y)$;*

*ii) $\Psi$ is continuous at $y$ if $y \notin S$;*

*iii) all elements $z_k$ are in a compact set of $S \subseteq Y$.*

If these conditions are satisfied then the algorithm shall stop at the solution set denoted by $\Omega$ or the limit of any convergent subsequence shall be in $\Omega$.

Denote $M_f = \{\mathbf{U} = [u_{ij}]_{C\times N} | \sum_{i=1}^{C} u_{ij} = 1, \sum_{j=1}^{N} u_{ij} > 0, u_{ij} \geq 0\}$ and $M_w = \{\mathbf{W} = [w_{il}]_{C\times L} | \sum_{l=1}^{L} w_{il} = 1,\ w_{il} > 0\ (i = 1, 2, \ldots, C)\}$, and $M_V = \{\mathbf{G} = [g_{il}]_{C\times L}\}$.

In order to verify the convergence of VWKFC, we need to define VWKFC operator $\Psi_*$. Then we could discover the sufficient and necessary condition for a strict minimizer of the VWKFC objective function by discussing the Jacobian matrix and the bordered Hessian matrix.

Authorized licensed use limited to: Chulalongkorn University provided by UniNet. Downloaded on January 17,2024 at 08:32:14 UTC from IEEE Xplore. Restrictions apply.

Let $U_T : (\mathbb{R}^L)^C \times M_w \to M_f$ with $U_T(\mathbf{G}, \mathbf{W}) = \mathbf{U}^* = [u_{ij}]_{C\times N}$, where $u_{ij}$ is computed by (17).

Let $W_T : M_f \times (\mathbb{R}^L)^C \to M_w$ with $W_T(\mathbf{U}, \mathbf{G}) = \mathbf{W}^* = [w_{il}]_{C\times L}$, where $w_{il}$ is determined by (18).

Let $G_T : M_f \times (\mathbb{R}^L)^C \times M_w \to (\mathbb{R}^L)^C$ with $G_T(\mathbf{U}, \mathbf{G}, \mathbf{W}) = \mathbf{G}^* = [g_{il}]_{C\times L}$ where $g_{il}$ is obtained by (19).

*Definition 1: The VWKFC operator* $\Psi_* : M_f \times (\mathbb{R}^L)^C \times M_w \to M_f \times (\mathbb{R}^L)^C \times M_w$ *is defined by* $\Psi_* = \Psi_3 \circ \Psi_2 \circ \Psi_1$ *where* $\Psi_1 : M_f \times (\mathbb{R}^L)^C \times M_w \to M_f \times (\mathbb{R}^L)^C \times M_w$ *with* $\Psi_1(\mathbf{U}, \mathbf{G}, \mathbf{W}) = (U_T(\mathbf{G}, \mathbf{W}), \mathbf{G}, \mathbf{W})$, $\Psi_2 : M_f \times (\mathbb{R}^L)^C \times M_w \to M_f \times (\mathbb{R}^L)^C \times M_w$ *with* $\Psi_2(\mathbf{U}, \mathbf{G}, \mathbf{W}) = (\mathbf{U}, G_T(\mathbf{U}, \mathbf{G}, \mathbf{W}), \mathbf{W})$ *and* $\Psi_3 : M_f \times (\mathbb{R}^L)^C \times M_w \to M_f \times (\mathbb{R}^L)^C \times M_w$ *with* $\Psi_3(\mathbf{U}, \mathbf{G}, \mathbf{W}) = (\mathbf{U}, \mathbf{G}, W_T(\mathbf{U}, \mathbf{G}))$. *Then, one has*

$$\begin{aligned}
&\Psi_*(\mathbf{U}, \mathbf{G}, \mathbf{W}) \\
&= (\Psi_3 \circ \Psi_2 \circ \Psi_1)(\mathbf{U}, \mathbf{G}, \mathbf{W}) = \Psi_3(\Psi_2(\Psi_1(\mathbf{U}, \mathbf{G}, \mathbf{W}))) \\
&= \Psi_3(\Psi_2(U_T(\mathbf{G}, \mathbf{W}), \mathbf{G}, \mathbf{W})) = \Psi_3(\mathbf{U}^*, G_T(\mathbf{U}^*, \mathbf{G}, \mathbf{W}), \mathbf{W}) \\
&= (\mathbf{U}^*, \mathbf{G}^*, W_T(\mathbf{U}^*, \mathbf{G}^*)) = (\mathbf{U}^*, \mathbf{G}^*, \mathbf{W}^*).
\end{aligned}$$

*where*

$$\mathbf{U}^* = U_T(\mathbf{G}, \mathbf{W}), \mathbf{G}^* = G_T(\mathbf{U}, \mathbf{G}, \mathbf{W}), \mathbf{W}^* = W_T(\mathbf{U}, \mathbf{G}).$$

*Theorem 2: (Lagrange's theorem, [24]) Suppose that* $f : D_f \to \mathbb{R}$ *(where* $D_f \subseteq \mathbb{R}^n$*) and* $h_i : D_{h_i} \to \mathbb{R}$ *(where* $D_{h_i} \subseteq \mathbb{R}^n$*,* $i = 1, \ldots, t$*,* $t < n$*) are continuously partially differentiable and* $x^0 = (x_1^0, \ldots, x_n^0) \in D_f$ *is a local extreme point of* $f$ *such that* $h_i(x_1, \ldots, x_n) = 0$ *holds* $(i = 1, \ldots, t)$*. Let* $La(x; \lambda) = f(x_1, \ldots, x_n) + \sum_{i=1}^t \lambda_i h_i(x_1, \ldots, x_n)$ *and*

$$|J| = \begin{vmatrix} \frac{\partial h_1(x)}{\partial x_1} & \cdots & \frac{\partial h_1(x)}{\partial x_t} \\ \vdots & & \vdots \\ \frac{\partial h_t(x)}{\partial x_1} & \cdots & \frac{\partial h_t(x)}{\partial x_t} \end{vmatrix} \neq 0$$

*at* $x^0$ *(where* $|J|$ *means the determinant of* $J$*). Then, one has that the gradient of* $La(x; \lambda)$ *at* $(x^0; \lambda^0)$ *is 0, i.e.,* $\bigtriangledown LA(x^0; \lambda^0) = 0$*.*

*Theorem 3: (Local sufficient conditions, [24]) Let* $f : D_f \to \mathbb{R}$ *(where* $D_f \subseteq \mathbb{R}^n$*), and* $D_{h_i} \subseteq \mathbb{R}^n$ *(in which* $i = 1, \ldots, t$*,* $t < n$*) be twice continuously partially differentiable and let* $(x^0; \lambda^0)$ *with* $x^0 \in D_f$ *be a solution of the system* $\bigtriangledown La(x^0; \lambda^0) = 0$*. Let*

$$H_{La}(x, \lambda) = \begin{pmatrix} 0 & \cdots & 0 & \frac{\partial^2 La}{\partial \lambda_1 \partial x_1} & \cdots & \frac{\partial^2 La}{\partial \lambda_1 \partial x_n} \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & \frac{\partial^2 La}{\partial \lambda_t \partial x_1} & \cdots & \frac{\partial^2 La}{\partial \lambda_t \partial x_n} \\ \frac{\partial^2 La}{\partial x_1 \partial \lambda_1} & \cdots & \frac{\partial^2 La}{\partial x_1 \partial \lambda_t} & \frac{\partial^2 La}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 La}{\partial x_1 \partial x_n} \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial^2 La}{\partial x_n \partial \lambda_1} & \cdots & \frac{\partial^2 La}{\partial x_n \partial \lambda_t} & \frac{\partial^2 La}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 La}{\partial x_n \partial x_n} \end{pmatrix}$$

*be the bordered Hessian while take into account its leading principle minors* $|\bar{H}_r(x^0; \lambda^0)|$ *of* $r = 2t + 1, 2t + 2, \ldots, n + t$ *at* $(x^0; \lambda^0)$*. As a result, the following conclusion can be obtained.*

*i) If all leading principle minors* $|\bar{H}_r(x^0; \lambda^0)|, 2t + 1 \leq r \leq n + t$*, possess the sign* $(-1)^t$*, then* $x^0 = (x_1^0, \ldots, x_n^0)$ *is a local minimum point of* $f$ *satisfying* $h_i(x) = 0$ *(*$i = 1, \ldots, t$*).*

*ii) If the signs of all* $|\bar{H}_r(x^0; \lambda^0)|$*,* $r = 2t + 1, 2t + 2, \ldots, n + t$*, are alternated and the sign of* $|\bar{H}_{n+t}(x^0; \lambda^0)| = |\bar{H}_L(x^0; \lambda^0)|$ *is that of* $(-1)^n$*, then* $x^0 = (x_1^0, \ldots, x_n^0)$ *is a local maximum element of* $f$ *satisfying* $h_i(x) = 0$ *(*$i = 1, \ldots, t$*).*

*iii) If neither the criteria of i) nor those of ii) hold, then* $x^0$ *is not a local extreme element of* $h$ *satisfying* $h_i(x) = 0$ *(*$i = 1, \ldots, t$*). Here, the situation in which one or several leading principal minors have a value of zero is not considered a violation of i) or ii) (unnumered Equation)shown at the bottom of next page.*

*Theorem 4: (Generalized Heine-Borel theorem, [25]) Each closed and basically bounded subset of a topological space is compact.*

For convenience, we denote $\Theta_{ij} = \sum_{l=1}^L w_{il} \|\phi(x_{jl}) - \phi(g_{il})\|^2$ ($i = 1, 2, \ldots, C$ and $j = 1, 2, \ldots, N$).

*Lemma 1: Suppose that* $\mathbf{G} = \hat{\mathbf{G}}$ *and* $\mathbf{W} = \hat{\mathbf{W}}$ *are fixed, then* $J_{VWKFC}(\mathbf{U}, \hat{\mathbf{G}}, \hat{\mathbf{W}})$ *subject to* $\sum_{i=1}^C u_{ij} = 1$ *is locally minimized at* $\mathbf{U}^* = [u_{ij}^*]_{C\times N}$ *if and only if* $u_{ij}^*$ *satisfies (17).*

*Proof:* Because for any $i$, the constraints $\sum_{i=1}^C u_{ij} = 1$ are the same, viz., $h_i(x_1, \ldots, x_N) = \ldots = h_N(x_1, \ldots, x_N)$, we can only analyze a specific $i$. Therefore, $\forall i$, the Lagrangian function is

$$\begin{aligned}
J^* = &\sum_{i=1}^C \sum_{j=1}^N \alpha_j u_{ij}^m \sum_{l=1}^L w_{il} \|\phi(x_{jl}) - \phi(g_{il})\|^2 \\
&+ \gamma^{-1} \sum_{i=1}^C \sum_{l=1}^L w_{il} \ln w_{il} + \lambda \left( \sum_{i=1}^C u_{ij} - 1 \right),
\end{aligned}$$

in which $\lambda$ is embodied as a Lagrangian multiplier. Using the gradient of $J^*$ w.r.t. $u_{ij}$ and $\lambda$, one has

$$\frac{\partial J^*}{\partial u_{ij}} = \alpha_j m u_{ij}^{m-1} \sum_{l=1}^L w_{il} \|\phi(x_{jl}) - \phi(g_{il})\|^2 + \lambda = 0, \text{ and}$$

$$\frac{\partial J^*}{\partial \lambda} = \sum_{i=1}^C u_{ij} - 1 = 0.$$

It can be concluded that:

$$u_{ij} = \left( \frac{-\lambda}{\alpha_j m \sum_{l=1}^L w_{il} \|\phi(x_{jl}) - \phi(g_{il})\|^2} \right)^{\frac{1}{m-1}}.$$

From the constraint condition of membership degree, we achieve:

$$\sum_{i=1}^C u_{ij} = \sum_{i=1}^C \left( \frac{-\lambda}{\alpha_j m \sum_{l=1}^L w_{il} \|\phi(x_{jl}) - \phi(g_{il})\|^2} \right)^{\frac{1}{m-1}} = 1.$$

Then, from these two formulas, one has

$$u_{ij} = \frac{\left( \sum_{l=1}^L w_{il} \|\phi(x_{jl}) - \phi(g_{il})\|^2 \right)^{-\frac{1}{m-1}}}{\sum_{s=1}^C \left( \sum_{l=1}^L w_{sl} \|\phi(x_{jl}) - \phi(g_{sl})\|^2 \right)^{-\frac{1}{m-1}}}.$$

As a result, "only if" condition is proved.

Authorized licensed use limited to: Chulalongkorn University provided by UniNet. Downloaded on January 17,2024 at 08:32:14 UTC from IEEE Xplore. Restrictions apply.

Moreover, we validate the case for "if" condition from Theorem 3. If $\mathbf{G}=\hat{\mathbf{G}}$ and $\mathbf{W}=\hat{\mathbf{W}}$ are fixed, then

$$\frac{\partial^2 J^*}{\partial u_{ij}\partial u_{rh}} = \Im_{ij}\Im_{rh}\alpha_j m(m-1)u_{ij}^{m-2} \times \sum_{l=1}^{L} w_{il}\|\phi(x_{jl})-\phi(g_{il})\|^2$$

and

$$\frac{\partial^2 J^*}{\partial u_{ij}\partial \lambda} = \frac{\partial^2 J^*}{\partial \lambda \partial u_{ij}} = 1.$$

Here $\Im$ is a Kronecker index with

$$\Im_{ij}=\begin{cases}1, & i=j,\\ 0, & i\neq j\end{cases},\ \Im_{rh}=\begin{cases}1, & r=h,\\ 0, & r\neq h\end{cases}.$$

Thus, the bordered Hessian matrix w.r.t. $u_{ij}$ and $\lambda$ is

$$H_{La}(u_{ij},\lambda)=\begin{pmatrix}0 & 1 & 1\cdots & 1\\ 1 & \frac{\partial^2 J^*}{\partial u_{i1}\partial u_{i1}} & 0\cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 1 & 0 & 0\cdots & \frac{\partial^2 J^*}{\partial u_{CN}\partial u_{CN}}\end{pmatrix}.$$

Then every leading principle minors are analyzed as unnumbered equation shown at the bottom of this page. Hence, it follows from Theorem 3 that $J_{\text{VWKFC}}(\mathbf{U},\hat{\mathbf{V}},\hat{\mathbf{W}})$ satisfying $\sum_{i=1}^{C}u_{ij}=1$ is locally minimized at $\mathbf{U}^*=[u_{ij}^*]_{C\times N}$ with $\forall i,j$

$$u_{ij}^*=\frac{\left(\sum_{l=1}^{L} w_{il}\|\phi(x_{jl})-\phi(g_{il})\|^2\right)^{-\frac{1}{m-1}}}{\sum_{s=1}^{C}\left(\sum_{l=1}^{L} w_{sl}\|\phi(x_{jl})-\phi(g_{sl})\|^2\right)^{-\frac{1}{m-1}}}.$$

□

*Lemma 2: Suppose that* $\mathbf{U}=\hat{\mathbf{U}}$ *and* $\mathbf{G}=\hat{\mathbf{G}}$ *are fixed, then* $J_{VWKFC}(\hat{\mathbf{U}},\hat{\mathbf{G}},\mathbf{W})$ *satisfying* $\sum_{l=1}^{L}w_{il}=1$ *and* $w_{il}>0$ *is minimized at* $\mathbf{W}^*=[w_{il}^*]_{C\times L}$ *if and only if* $w_{il}^*$ *satisfies (18).*

*Proof:* The Lagrangian function is

$$J_1^*=\sum_{i=1}^{C}\sum_{j=1}^{N}\alpha_j u_{ij}^m\sum_{l=1}^{L}w_{il}\|\phi(x_{jl})-\phi(g_{il})\|^2 + \gamma^{-1}\sum_{i=1}^{C}\sum_{l=1}^{L}w_{il}\ln w_{il}+\lambda_1\sum_{i=1}^{C}(1-\sum_{l=1}^{L}w_{il}),$$

in which $\lambda_1$ is utilized as a Lagrangian multiplier. The partial derivative of feature weight $w_{il}$ is calculated and the result is set to zero, and the following result is obtained ($i=1,2,\ldots,C; l=1,2,\ldots,L$):

$$\sum_{j=1}^{N}\alpha_j u_{ij}^m\|\phi(x_{jl})-\phi(g_{il})\|^2+\gamma^{-1}(1+\ln w_{il})-\lambda_1=0.$$

After derivation, we get:

$$w_{il}=e^{(\lambda_1\gamma-1)}\exp\left\{-\gamma\sum_{j=1}^{N}\alpha_j u_{ij}^m\|\phi(x_{jl})-\phi(g_{il})\|^2\right\}.$$

From the constraint condition of feature weight, one has:

$$e^{(\lambda_1\gamma-1)}\sum_{l=1}^{L}\exp\left\{-\gamma\sum_{j=1}^{N}\alpha_j u_{ij}^m\|\phi(x_{jl})-\phi(g_{il})\|^2\right\}=1.$$

Then we obtain the updated expression of $w_{il}$ :

$$w_{il}=\frac{\exp\{-\gamma\sum_{j=1}^{N}\alpha_j u_{ij}^m\|\phi(x_{jl})-\phi(g_{il})\|^2\}}{\sum_{s=1}^{L}\exp\{-\gamma\sum_{j=1}^{N}\alpha_j u_{ij}^m\|\phi(x_{js})-\phi(g_{is})\|^2\}}.$$

Hence, the "only if" condition is validated.

For the "if" condition, if $\mathbf{U}=\hat{\mathbf{U}}$ and $\mathbf{G}=\hat{\mathbf{G}}$ are fixed, then one has $\frac{\partial J_{\text{VWKFC}}^2}{\partial w_{il}\partial w_{rl}}=\Im_{ir}\frac{1}{\gamma w_{il}}$. Here $\Im$ is a Kronecker index with $\Im_{ir}=\begin{cases}1, & i=r,\\ 0, & i\neq r\end{cases}$.

---

Leading principle minors of $H_{La}(u_{ij},\lambda)$:

$$|\bar{H}_3(u_j^*,\lambda^*)|=\begin{vmatrix}0 & 1 & 1\\ 1 & \alpha_j m(m-1)u_{1j}^{m-2}\Theta_{1j} & 0\\ 1 & 0 & \alpha_j m(m-1)u_{2j}^{m-2}\Theta_{2j}\end{vmatrix}_{\mu_j=\mu_j^*,\lambda=\lambda^*}$$

$$=-[\alpha_j m(m-1)u_{1j}^{m-2}\Theta_{1j}+\alpha_j m(m-1)u_{2j}^{m-2}\Theta_{2j}]_{u_j=u_j^*,\lambda=\lambda^*}<0,$$

$$\left|\bar{H}_4(u_j^*,\lambda^*)\right|=-\left(\sum_{i=1}^{3}\prod_{\substack{p=1\\p\neq i}}^{3}\alpha_j m(m-1)u_{pj}^{m-2}\Theta_{pj}\right)_{u_j=u_j^*,\lambda=\lambda^*}<0,$$

$\ldots$ and

$$\left|\bar{H}_{C+1}(u_j^*,\lambda^*)\right|=-\left(\sum_{i=1}^{C}\prod_{\substack{p=1\\p\neq i}}^{C}\alpha_j m(m-1)u_{pj}^{m-2}\Theta_{pj}\right)_{u_j=u_j^*,\lambda=\lambda^*}<0.$$

Authorized licensed use limited to: Chulalongkorn University provided by UniNet. Downloaded on January 17,2024 at 08:32:14 UTC from IEEE Xplore. Restrictions apply.

Hence the Hessian matrix of $J_{\text{VWKFC}}(\hat{\mathbf{U}}, \hat{\mathbf{G}}, \mathbf{W})$ w.r.t. $w_{il}$ is

$$diag\left(\frac{1}{\gamma w_{1l}}, \frac{1}{\gamma w_{2\text{-}l}}, \ldots, \frac{1}{\gamma w_{Cl}}\right)$$

and evidently, the Hessian matrix is positive definite. So $J_{\text{VWKFC}}(\hat{\mathbf{U}}, \hat{\mathbf{G}}, \mathbf{W})$ is minimized at $W^* = [w_{il}^*]_{N\times L}$ with $\forall i, l$

$$w_{il}^* = \frac{\exp\{-\gamma \sum_{j=1}^{N} \alpha_j u_{ij}^m \|\phi(x_{jl}) - \phi(g_{il})\|^2\}}{\sum_{s=1}^{L} \exp\{-\gamma \sum_{j=1}^{N} \alpha_j u_{ij}^m \|\phi(x_{js}) - \phi(g_{is})\|^2\}}.$$

□

*Lemma 3: Suppose that* $\mathbf{U} = \hat{\mathbf{U}}$, $\mathbf{G} = \hat{\mathbf{G}}$ *and* $\mathbf{W} = \hat{\mathbf{W}}$ *are fixed, and the guidance of the density viewpoint is employed during clustering, then* $J_{VWKFC}(\hat{\mathbf{U}}, \hat{\mathbf{G}}, \hat{\mathbf{W}})$ *is minimized at* $\mathbf{G}^* = [g_{il}^*]_{C\times L}$ *if and only if* $g_{il}^*$ *satisfies (19).*

*Proof:* When calculating the distance between the data points and the cluster centers, we use Gaussian RBF, i.e., $\|\phi(x_{jl}) - \phi(g_{il})\|^2 = 2 - 2e^{-\frac{(x_{jl}-g_{il})^2}{2\sigma^2}}$.

According to (16), when $i = q$ occurs, the value of the cluster center is the value of the density viewpoint, i.e., $g_{il} = x_{dl}$. When $i \neq q$, we take the partial derivative as $\frac{\partial J_{\text{VWKFC}}}{\partial g_{il}} = \sum_{j=1}^{N} 2\alpha_j u_{ij}^m w_{il} e^{-\frac{(x_{jl}-g_{il})^2}{2\sigma^2}} \frac{(g_{il}-x_{jl})}{\sigma^2}$. Then set $\frac{\partial J_{\text{VWKFC}}}{\partial g_{il}} = 0$ $(i = 1, 2, \ldots, C;\ l = 1, 2, \ldots, L)$. After derivation, we obtain the updated expression:

$$g_{il} = \begin{cases} x_{dl}, & i = q, \\ \dfrac{\sum_{j=1}^{N} \alpha_j u_{ij}^m w_{il} e^{-\frac{(x_{jl}-\hat{g}_{il})^2}{2\sigma^2}} x_{jl}}{\sum_{j=1}^{N} \alpha_j u_{ij}^m w_{il} e^{-\frac{(x_{jl}-\hat{g}_{il})^2}{2\sigma^2}}}, & i \neq q. \end{cases}$$

Here we use $\hat{g}_{il}$ on the right-hand side of the formula above to differentiate. In fact, $g_{il}$ corresponds to current iteration; while $\hat{g}_{il}$ corresponds to the result of the last iteration and is regarded as constant in current iteration. Such mode is illuminated by the processing strategy of KFCM. Hence the "only if" condition is validated.

Moreover, the proof of the "if" condition is as below.

If $\mathbf{U} = \hat{\mathbf{U}}$, $\mathbf{G} = \hat{\mathbf{G}} = [\hat{g}_{il}]_{C\times L}$ and $\mathbf{W} = \hat{\mathbf{W}}$ are fixed, then one has $\frac{\partial^2 J_{\text{VWKFC}}}{\partial g_{il} \partial g_{rt}} = \sum_{j=1}^{N} \frac{2}{\sigma^2} \alpha_j u_{ij}^m w_{il} e^{-\frac{(x_{jl}-\hat{g}_{il})^2}{2\sigma^2}}$ (noting $\hat{g}_{il}$ is regarded as constant in current iteration). Here $\Im$ is a Kronecker index as $\Im_{ir} = \begin{cases} 1, & i = r, \\ 0, & i \neq r \end{cases}$.

The Hessian matrix of $J_{\text{VWKFC}}(\hat{\mathbf{U}}, \hat{\mathbf{G}}, \hat{\mathbf{W}})$ is

$$diag\Bigg( \sum_{j=1}^{N} \frac{2}{\sigma^2} \alpha_j u_{1j}^m w_{1\text{-}l} e^{-\frac{(x_{jl}-\hat{g}_{1l})^2}{2\sigma^2}}, \sum_{j=1}^{N} \frac{2}{\sigma^2} \alpha_j u_{2j}^m w_{2\text{-}l} e^{-\frac{(x_{jl}-\hat{g}_{2\text{-}l})^2}{2\sigma^2}}, \ldots, \sum_{j=1}^{N} \frac{2}{\sigma^2} \alpha_j u_{Cj}^m w_{Cl} e^{-\frac{(x_{jl}-\hat{g}_{Cl})^2}{2\sigma^2}} \Bigg).$$

Evidently, the Hessian matrix is positive definite. In other words, $J_{\text{VWKFC}}(\hat{\mathbf{U}}, \hat{\mathbf{G}}, \hat{\mathbf{W}})$ is minimized at which $g_{il}^*$ is obtained from (19). □

*Lemma 4:* $J_{VWKFC}$ *is continuous on* $M_f \times (\mathbb{R}^L)^C \times M_w$.

*Proof:* In $J_{\text{VWKFC}}$, $\{g_{il} \to \|\phi(x_{jl}) - \phi(g_{il})\|^2\}$, $\{u_{ij} \to u_{ij}^m\}$, and $\{w_{il} \to \ln w_{il}\}$ are continuous. The sum of products of $\{w_{il} \to w_{il}\}$ and $\{g_{il} \to \|\phi(x_{jl}) - \phi(g_{il})\|^2\}$ is continuous. Besides, the sum of products of $\{u_{ij} \to u_{ij}^m\}$, $\{w_{il} \to w_{il}\}$ and $\{g_{il} \to \|\phi(x_{jl}) - \phi(g_{il})\|^2\}$ is likewise continuous. Moreover, the sum of products of $\{w_{il} \to w_{il}\}$ and $\{w_{il} \to \ln w_{il}\}$ is continuous. Consequently, $J_{\text{VWKFC}}$ is continuous on $M_f \times (\mathbb{R}^L)^C \times M_w$. □

Let $\Omega_{VWKFC}$ be the solution set of $J_{\text{VWKFC}}$.

*Lemma 5: For any* $(\mathbf{U}, \mathbf{G}, \mathbf{W}) \notin \Omega_{VWKFC}$, *one has that* $J_{VWKFC}(\Psi_*(\mathbf{U}, \mathbf{G}, \mathbf{W})) = J_{VWKFC}(\mathbf{U}^*, \mathbf{G}^*, \mathbf{W}^*) < J_{VWKFC}(\mathbf{U}, \mathbf{G}, \mathbf{W})$.

*Proof:* Suppose that $(\mathbf{U}, \mathbf{G}, \mathbf{W}) \notin \Omega_{VWKFC}$. Then, one has $J_{\text{VWKFC}}(\Psi_*(\mathbf{U}, \mathbf{G}, \mathbf{W})) = J_{\text{VWKFC}}(\Psi_3 \circ \Psi_2 \circ \Psi_1(\mathbf{U}, \mathbf{G}, \mathbf{W})) = J_{\text{VWKFC}}(\mathbf{U}^*, \mathbf{G}^*, \mathbf{W}^*) < J_{\text{VWKFC}}(\mathbf{U}^*, \mathbf{G}, \mathbf{W}^*)$ by Lemma 3. It implies that $J_{\text{VWKFC}}(\mathbf{U}^*, \mathbf{G}, \mathbf{W}^*) < J_{\text{VWKFC}}(\mathbf{U}^*, \mathbf{G}, \mathbf{W})$ by Lemma 2, and that $J_{\text{VWKFC}}(\mathbf{U}^*, \mathbf{G}, \mathbf{W}) < J_{\text{VWKFC}}(\mathbf{U}, \mathbf{G}, \mathbf{W})$ by Lemma 1. As a result, $J_{\text{VWKFC}}(\Psi_*(\mathbf{U}, \mathbf{G}, \mathbf{W})) = J_{\text{VWKFC}}(\mathbf{U}^*, \mathbf{G}^*, \mathbf{W}^*) < J_{\text{VWKFC}}(\mathbf{U}, \mathbf{G}, \mathbf{W})$ for any $(\mathbf{U}, \mathbf{G}, \mathbf{W}) \notin \Omega_{VWKFC}$. □

*Lemma 6: The VWKFC operator* $\Psi_*$ *is continuous on* $M_f \times (\mathbb{R}^L)^C \times M_w$.

*Proof:* In VWKFC, one has $\Psi_1(\mathbf{U}, \mathbf{G}, \mathbf{W}) = (U_T(\mathbf{G}, \mathbf{W}), \mathbf{G}, \mathbf{W})$, where

$$U_T(\mathbf{G}, \mathbf{W}) = \frac{\left(\sum_{l=1}^{L} w_{il} \|\phi(x_{jl}) - \phi(g_{il})\|^2\right)^{-\frac{1}{m-1}}}{\sum_{s=1}^{C} \left(\sum_{l=1}^{L} w_{sl} \|\phi(x_{jl}) - \phi(g_{sl})\|^2\right)^{-\frac{1}{m-1}}} = u_{ij},$$

Note that $\{g_{il} \to \|\phi(x_{jl}) - \phi(g_{il})\|^2\}$ and $\{\phi(g_{sl}) \to \|\phi(x_{jl}) - \phi(g_{sl})\|^2\}$ are continuous. The sum of products of $\{w_{il} \to w_{il}\}$ and $\{g_{il} \to \|\phi(x_{jl}) - \phi(g_{il})\|^2\}$ is continuous, while the sum of products of $\{w_{il} \to w_{il}\}$ and $\{g_{sl} \to \|\phi(x_{jl}) - \phi(g_{sl})\|^2\}$ is continuous. Besides, $\{(\sum_{l=1}^{L} w_{il} \|\phi(x_{jl}) - \phi(g_{il})\|^2) \to (\sum_{l=1}^{L} w_{il} \|\phi(x_{jl}) - \phi(g_{il})\|^2)^{-\frac{1}{m-1}}\}$ is continuous while $\{(\sum_{l=1}^{L} w_{sl} \|\phi(x_{jl}) - \phi(g_{sl})\|^2) \to (\sum_{l=1}^{L} w_{sl} \|\phi(x_{jl}) - \phi(g_{sl})\|^2)^{-\frac{1}{m-1}}\}$ is continuous. Moreover, the quotient of two continuous functions, is likewise continuous. So $\Psi_1$ is continuous on $M_f \times (\mathbb{R}^L)^C \times M_w$.

In a similar way, as for $\Psi_2(\mathbf{U}, \mathbf{G}, \mathbf{W}) = (\mathbf{U}, G_T(\mathbf{U}, \mathbf{G}, \mathbf{W}), \mathbf{W})$, we can obtain $\Psi_2$ is continuous on $M_f \times (\mathbb{R}^L)^C \times M_w$. Moreover, as for $\Psi_3(\mathbf{U}, \mathbf{G}, \mathbf{W}) = (\mathbf{U}, \mathbf{G}, W_T(\mathbf{U}, \mathbf{G}))$, we can achieve $\Psi_3$ is continuous on $M_f \times (\mathbb{R}^L)^C \times M_w$.

All in all, the operator $\Psi_{T*} = \Psi_3 \circ \Psi_2 \circ \Psi_1$ is continuous on $M_f \times (\mathbb{R}^L)^C \times M_w$. □

*Lemma 7: Suppose that* $[conv(X)]^C$ *is the C-fold Cartesian product of the convex hull of X, and that*

Authorized licensed use limited to: Chulalongkorn University provided by UniNet. Downloaded on January 17,2024 at 08:32:14 UTC from IEEE Xplore. Restrictions apply.

*$(U_T(\mathbf{G}^{(0)},\mathbf{W}^{(0)}),\mathbf{G}^{(0)},\mathbf{W}^{(0)})$ is the starting position of iteration using $\Psi_*$. Then $\Psi_*^{(t)}(U_T(\mathbf{G}^{(0)},\mathbf{W}^{(0)}),\mathbf{G}^{(0)},\mathbf{W}^{(0)}) \in M_f \times [conv(X)]^C \times M_w$ is compact in $M_f \times (\mathbb{R}^L)^C \times M_w$.*

*Proof:* Suppose that $(U_T(\mathbf{G}^{(0)},\mathbf{W}^{(0)}),\mathbf{G}^{(0)},\mathbf{W}^{(0)})$ is the starting position of iteration using $\Psi_*$ where $U_T(\mathbf{G}^{(0)},\mathbf{W}^{(0)}) = (U_{T11}(\mathbf{G}^{(0)},\mathbf{W}^{(0)}), U_{T21}(\mathbf{G}^{(0)},\mathbf{W}^{(0)}),\ldots,U_{TNC}(\mathbf{G}^{(0)},\mathbf{W}^{(0)})$ with

$$U_{Tij}(\mathbf{G}^{(0)},\mathbf{W}^{(0)}) = \frac{\left(\sum_{l=1}^{L} w_{il}^{(0)}\|\phi(x_{jl}) - \phi(g_{il}^{(0)})\|^2\right)^{-\frac{1}{m-1}}}{\sum_{s=1}^{C}\left(\sum_{l=1}^{L} w_{sl}^{(0)}\|\phi(x_{jl}) - \phi(g_{sl}^{(0)})\|^2\right)^{-\frac{1}{m-1}}} = u_{ij}^{(0)}.$$

Then

$$g_{il}^{(1)} = G_{Til}(\mathbf{U}^{(0)},\mathbf{G}^{(0)},\mathbf{W}^{(0)}) = \begin{cases} x_{dl}, & i = q, \\ \dfrac{\sum_{j=1}^{N}\alpha_j (u_{ij}^{(0)})^m w_{il}^{(0)} e^{-\frac{(x_{jl}-g_{il}^{(0)})^2}{2\sigma^2}} x_{jl}}{\sum_{j=1}^{N}\alpha_j (u_{ij}^{(0)})^m w_{il}^{(0)} e^{-\frac{(x_{jl}-g_{il}^{(0)})^2}{2\sigma^2}}}, & i \neq q. \end{cases}$$

It can be divided into two cases. On the one hand, suppose that $i = q$. Then $g_{il}^{(1)} = x_{dl}$. Note that $x_d \in X$ from the KHDI algorithm, and obviously $g_{il}^{(1)} \in [conv(X)]^C$. On the other hand, suppose that $i \neq q$. Let

$$\wp_{ij} = \frac{\alpha_j (u_{ij}^{(0)})^m w_{il}^{(0)} e^{-\frac{(x_{jl}-g_{il}^{(0)})^2}{2\sigma^2}}}{\sum_{j=1}^{N}\alpha_j (u_{ij}^{(0)})^m w_{il}^{(0)} e^{-\frac{(x_{jl}-g_{il}^{(0)})^2}{2\sigma^2}}}, \ \forall i.$$

Hence, $0 \leq \wp_{ij} \leq 1, \ \forall i, j$, and $g_{il}^{(1)} = \sum_{j=1}^{N} \wp_{ij} x_{jl}$ with

$$\sum_{j=1}^{N} \wp_{ij} = \sum_{j=1}^{N} \frac{\alpha_j (u_{ij}^{(0)})^m w_{il}^{(0)} e^{-\frac{(x_{jl}-g_{il}^{(0)})^2}{2\sigma^2}}}{\sum_{j=1}^{N}\alpha_j (u_{ij}^{(0)})^m w_{il}^{(0)} e^{-\frac{(x_{jl}-g_{il}^{(0)})^2}{2\sigma^2}}} = 1.$$

As a result, $g_{il}^{(1)} \in [conv(X)]^C$. Summarizing above, $\mathbf{G}^{(1)} \in [conv(X)]^C$. Persistently recursively, $\mathbf{G}^{(t)} \in [conv(X)]^C$, $\forall t \geq 1$.

Moreover, it is evident that one has

$$w_{il}^{(1)} = W_{Til}(\mathbf{U}^{(0)},\mathbf{G}^{(1)}) = \frac{\exp\{-\gamma\sum_{j=1}^{N}\alpha_j (u_{ij}^{(0)})^m \|\phi(x_{jl}) - \phi(g_{il}^{(1)})\|^2\}}{\sum_{s=1}^{L}\exp\{-\gamma\sum_{j=1}^{N}\alpha_j (u_{ij}^{(0)})^m \|\phi(x_{js}) - \phi(g_{is}^{(1)})\|^2\}}$$

and $\mathbf{W}^{(1)} \in M_w$. Also we have

$$u_{ij}^{(1)} = U_{Tij}(\mathbf{G}^{(1)},\mathbf{W}^{(1)}) = \frac{(\sum_{l=1}^{L} w_{il}^{(1)}\|\phi(x_{jl}) - \phi(g_{il}^{(1)})\|^2)^{-\frac{1}{m-1}}}{\sum_{s=1}^{C}(\sum_{l=1}^{L} w_{sl}^{(1)}\|\phi(x_{jl}) - \phi(g_{sl}^{(1)})\|^2)^{-\frac{1}{m-1}}} \in M_f,$$

and $\mathbf{U}^{(1)} \in M_f$. Persistently recursively, $\mathbf{W}^{(t)} \in M_w$ and $\mathbf{U}^{(t)} \in M_f$, $\forall t \geq 1$. Hence, $\Psi_*^{(t)}(U_T(\mathbf{G}^{(0)},\mathbf{W}^{(0)}),\mathbf{G}^{(0)},\mathbf{W}^{(0)}) \in M_f \times [conv(X)]^C \times M_w, \forall t$.

Finally, we validate that $M_f \times [conv(X)]^C \times M_w$ is compact in $M_f \times (\mathbb{R}^L)^C \times M_w$. Because $X$ is finite, every $x_j \in X$ has finite elements. As a result, the diameter of $X$ (which equals to the one of $conv(X)$) is bounded. Because $conv(X)$ is the convex hull of finitely many generators $x_j$, it is closed. Hence, $conv(X)$ is bounded and closed in $\mathbb{R}^L$, and hence $conv(X)$ is compact. On the strengths of Theorem 3, it can be obtained that $[conv(X)]^C$ is also compact. For $M_w$, noting that $\sum_{l=1}^{L} w_{il} = 1, w_{il} > 0\ (i = 1, 2, \ldots, C)$ and $w_{il} \in (\mathbb{R}^L)^C$, evidently one has $M_w$ is bounded and closed. Because $\sum_{j=1}^{N} u_{ij} > 0$ and $\sum_{i=1}^{C} u_{ij} = 1$, $M_f$ is closed and bounded. Consequently, $M_f \times [conv(X)]^C \times M_w$ is compact in $M_f \times (\mathbb{R}^L)^C \times M_w$. It implies that $\Psi_*^{(t)}(U_T(\mathbf{G}^{(0)},\mathbf{W}^{(0)}),\mathbf{G}^{(0)},\mathbf{W}^{(0)}) \in M_f \times [conv(X)]^C \times M_w$ is compact in $M_f \times (\mathbb{R}^L)^C \times M_w$. □

In the light of Lemma 4 to Lemma 7 by validating the condition of Zangwill's convergence theorem, we achieve Theorem 5 of the convergence theorem for VWKFC.

*Theorem 5: Suppose that $X = \{x_1,\ldots,x_N\}$ is bounded in $\mathbb{R}^L$ with the VWKFC objective function $J_{VWKFC}(\mathbf{U},\mathbf{G},\mathbf{W})$ satisfying $\sum_{i=1}^{C} u_{ij} = 1$ as well as $\sum_{l=1}^{L} w_{il} = 1$, and that $\Psi_* : M_f \times (\mathbb{R}^L)^C \times M_w \to M_f \times (\mathbb{R}^L)^C \times M_w$ is the VWKFC operator as characterized in Definition 1. Then, for any VWKFC convergent subsequence $\Psi_*^{(t_k)}(U_T(\mathbf{G}^{(0)},\mathbf{W}^{(0)}),\mathbf{G}^{(0)},\mathbf{W}^{(0)})$ shall tend to the optimal solution $(\mathbf{U}^*,\mathbf{G}^*,\mathbf{W}^*)$ in $\Omega_{VWKFC}$, and the VWKFC sequence $\Psi_*^{(t)}(U_T(\mathbf{G}^{(0)},\mathbf{W}^{(0)}),\mathbf{G}^{(0)},\mathbf{W}^{(0)})$ shall monotonically converge to the optimal solution $(\mathbf{U}^*,\mathbf{G}^*,\mathbf{W}^*)$ in $\Omega_{VWKFC}$.*

## III. Experimental Studies

### A. Experimental Settings

Here we use eight algorithms for comparison, which include the FCM, KFCM, RLM, VFCM, EWFCM, FRFCM, DVPFCM and H-FCM-PSO algorithms. The platform is Windows 10, and the programming languages are Matlab 2013b and Python 3.5. During the experiment, we ran the proposed algorithm on five artificial datasets, eight UCI datasets [26], the ORL face dataset [9], two high-dimensional datasets, and three datasets with various shapes. Experiments on different types of datasets help reflect the fairness of comparison.

For all algorithms, we mainly utilize default values of the parameters. The specific settings are $m = 2.5$, $a = 2$, $\gamma = 1$. For PFCM and DVPFCM, we employ $a_1 = 1$, $a_2 = 1$, $p = 2$. For VFCM, we obtain the high-density viewpoint through the HDCCI method. Considering that the performance of these algorithms depends on the initialization state, we run each experiment 30 times with different initializations, and average them to get the values of each evaluation index.

We employ five evaluation indexes, including classification accuracy (ACC) [27], normalized mutual information (NMI) [28], Calinski-Harabasz (CH) [29], ARI extension index (EARI) [30]–[32] and Xie-Beni (XB) index [33]. Among them, ACC, NMI and CH are hard clustering indexes, which can be applied to both hard and soft clustering algorithms. EARI and XB indexes are soft clustering indexes. Different evaluation indexes reflect

Authorized licensed use limited to: Chulalongkorn University provided by UniNet. Downloaded on January 17,2024 at 08:32:14 UTC from IEEE Xplore. Restrictions apply.

TABLE III
RUNNING RESULTS OF COMPARISON ALGORITHMS ON ARTIFICIAL DATASETS

| | | FCM | KFCM | RLM | V-FCM | EWFCM | FRFCM | DVPFCM | H-FCM-PSO | VWKFC |
|---|---|---|---|---|---|---|---|---|---|---|
| Data-a | ACC(+) | 0.9254 | 0.9565 | 0.9733 | 0.9867 | 0.99 | 0.9925 | 0.9933 | 0.9933 | **0.9966** |
| | CH(+) | 7.9865 | 8.4489 | 8.9765 | 9.2346 | 9.8287 | 11.4623 | 11.6342 | 11.8942 | **11.9834** |
| | NMI(+) | 0.4967 | 0.5235 | 0.5753 | 0.5997 | 0.6506 | 0.8847 | 0.9144 | 0.9488 | **0.9702** |
| | EARI(+) | 0.6035 | 0.6346 | 0.7369 | 0.7523 | 0.852 | 0.9613 | 0.9632 | 0.9824 | **0.9911** |
| | XB(-) | 0.8876 | 0.6437 | 0.3875 | 0.385 | 0.2946 | 0.2447 | 0.1672 | 0.1310 | **0.1046** |
| Data-b | ACC(+) | 0.639 | 0.689 | 0.8553 | 0.887 | 0.951 | 0.987 | 0.9883 | 0.9880 | **0.989** |
| | CH(+) | 7.9849 | 10.4737 | 11.6455 | 12.2932 | 13.7355 | 14.8743 | 14.8546 | 14.8781 | **14.8935** |
| | NMI(+) | 0.832 | 0.8554 | 0.8925 | 0.9575 | 0.962 | 0.9654 | 0.9692 | 0.9683 | **0.9692** |
| | EARI(+) | 0.6799 | 0.7533 | 0.836 | 0.8451 | 0.9158 | 0.9877 | 0.9879 | 0.9879 | **0.988** |
| | XB(-) | 0.5247 | 0.4822 | 0.594 | 0.4637 | 0.243 | 0.1945 | 0.1601 | 0.1178 | **0.1049** |

TABLE IV
DETAILS FOR UCI DATASETS

| Name | Instances | Attributes | Classes |
|---|---|---|---|
| Iris | 150 | 4 | 3 |
| Wine | 178 | 13 | 3 |
| Seeds | 210 | 7 | 3 |
| Spect | 267 | 22 | 2 |
| Breast_cancer | 569 | 30 | 2 |
| Letter_AB | 1555 | 16 | 2 |
| Wifi Localization | 2000 | 7 | 4 |
| Statlog | 6435 | 36 | 6 |

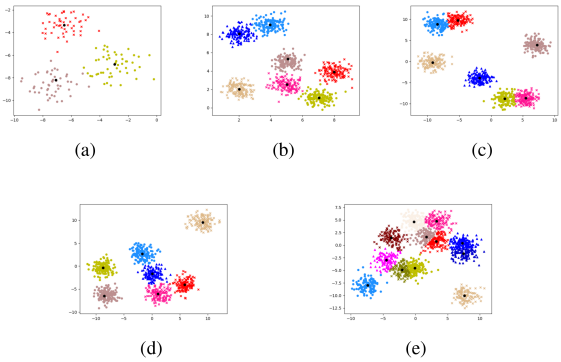


Fig. 2. Running results of VWKFC on artificial datasets. (a) Data-a. (b) Data-b. (c) Data-c. (d) Data-d. (e) Data-e.

the performance of the clustering method at disparate perspectives. Using five different indexes can more comprehensively reflect the clustering results, thus making the comparison more reliable.

Note that the RLM algorithm is a Boolean clustering algorithm. Here we supplement its membership matrix whose membership degrees are made up of zeros and ones. Then we can also provide the values of EARI and XB for RLM.

As for choosing the values of parameters, we firstly select some values through some past experience. Secondly, we run a clustering algorithm by trial and error with these basic values of parameters. Finally, we employ the values of parameters corresponding to the best result.

### B. Artificial Datasets

The running results of VWKFC on five artificial datasets (i.e., Data-a, Data-b, Data-c, Data-d and Data-e) are shown as Fig. 2.

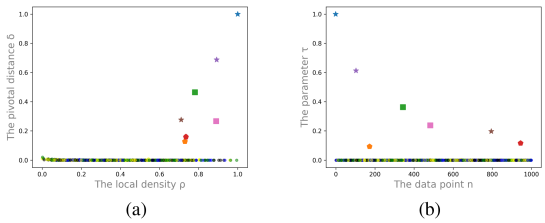


Fig. 3. (a) $\rho - \delta$ decision graph of Data-b dataset. (b) $\tau$ distribution diagram of Data-b dataset.

From Fig. 2, we can see that data points belonging to different clusters are marked with different colors and shaped after the operation of VWKFC. We can find that the VWKFC algorithm performs well in these datasets. Among them, the black dots in the rendering diagram represent the location of the cluster centers obtained by VWKFC. It can be found that the location of the cluster centers obtained by VWKFC is accurate. Especially in the Data-e dataset, 12 different cluster centers are accurately found in the case of severe class overlap.

To verify the effectiveness of the proposed KHDI algorithm (as the initialization method), Fig. 3(a) is the $\rho - \delta$ decision graph of Data-b processed by KHDI. As can be seen from Fig. 3(a) the local density is evenly distributed within the interval of $[0, 1]$. Among them, the pivotal distance $\delta$ of most data points is at a level less than 0.05, and these data points are represented by a dot "·" in the figure. By observing Fig. 3(a), there exist 7 data points that are different from other data points. They have a pivotal distance $\delta$ and a large local density $\rho$ that are significantly higher than other data points. The seven points are marked by the different symbols. Correspondingly, Fig. 3(b) is the parameter $\tau$ distribution diagram of Data-b. It is observed that $\tau$ for these 7 data points in Fig. 3(b) is evidently larger than most data points. The value of $\tau$ of most data points is almost close to 0. According to the idea of KHDI, these 7 data points correspond to the seven initial cluster centers for Data-b. One can well perceive from Fig. 3 that KHDI can distinguish potential cluster centers from common data points and effectively discover initial cluster centers.

In Table III, we offer the value of the evaluation indexes obtained by comparative algorithms after running on Data-a and Data-b. The results of FCM, KFCM and RLM are the worst. The effect of V-FCM is better due to the introduction of viewpoints. Then, EWFCM, FRFCM, DVPFCM and H-FCM-PSO also

Authorized licensed use limited to: Chulalongkorn University provided by UniNet. Downloaded on January 17,2024 at 08:32:14 UTC from IEEE Xplore. Restrictions apply.

TABLE V
RUNNING RESULTS OF COMPARISON ALGORITHMS ON UCI DATASETS

| | | FCM | KFCM | RLM | V-FCM | EWFCM | FRFCM | DVPFCM | H-FCM-PSO | VWKFC |
|---|---|---|---|---|---|---|---|---|---|---|
| Iris | ACC(+) | 0.597 | 0.7289 | 0.8065 | 0.8667 | 0.8933 | 0.9133 | 0.9225 | 0.9333 | **0.9467** |
| | CH(+) | 9.7035 | 10.2406 | 10.906 | 10.1723 | 11.6567 | 11.8354 | 12.295 | 12.3027 | **12.3074** |
| | NMI(+) | 0.2498 | 0.3315 | 0.6832 | 0.7269 | 0.7433 | 0.7705 | 0.7833 | 0.8322 | **0.8366** |
| | EARI(+) | 0.6932 | 0.7521 | 0.7562 | 0.8487 | 0.8524 | 0.8838 | 0.9118 | 0.9253 | **0.9299** |
| | XB(-) | 0.7407 | 0.4083 | 0.6410 | 0.398 | 0.391 | 0.206 | 0.3345 | 0.2025 | **0.199** |
| Wine | ACC(+) | 0.6551 | 0.727 | 0.7541 | 0.7876 | 0.8034 | 0.809 | 0.8596 | 0.8708 | **0.8764** |
| | CH(+) | 2.0802 | 2.1887 | 2.9635 | 4.1096 | 4.6689 | 7.3996 | 8.2208 | 9.5626 | **10.374** |
| | NMI(+) | 0.1769 | 0.281 | 0.3712 | 0.4342 | 0.4771 | 0.5494 | 0.6709 | 0.6711 | **0.6714** |
| | EARI(+) | 0.1588 | 0.2593 | 0.2873 | 0.3029 | 0.4049 | 0.574 | 0.7325 | 0.7867 | **0.8066** |
| | XB(-) | 3.219 | 2.3283 | 2.4835 | 2.2456 | 2.1449 | 1.6047 | 1.4138 | 1.0297 | **0.8497** |
| Seeds | ACC(+) | 0.7952 | 0.7983 | 0.8295 | 0.8348 | 0.8524 | 0.8905 | 0.8952 | 0.8995 | **0.9** |
| | CH(+) | 8.0245 | 10.0106 | 10.8873 | 13.431 | 14.1835 | 15.7293 | 15.7699 | 15.8007 | **15.8565** |
| | NMI(+) | 0.4799 | 0.4825 | 0.5332 | 0.5905 | 0.6337 | 0.6654 | 0.6684 | 0.6729 | **0.6976** |
| | EARI(+) | 0.5319 | 0.5471 | 0.6122 | 0.7503 | 0.8039 | 0.8273 | 0.8359 | 0.8494 | **0.8585** |
| | XB(-) | 0.7752 | 0.6853 | 0.6450 | 0.5699 | 0.3962 | 0.3919 | 0.3517 | 0.2876 | **0.1205** |
| Spect | ACC(+) | 0.5723 | 0.7352 | 0.7667 | 0.7753 | 0.794 | 0.7945 | 0.8427 | 0.8315 | **0.8543** |
| | CH(+) | 16.5504 | 18.7896 | 18.998 | 19.3829 | 21.9245 | 31.8734 | 33.2145 | 33.2275 | **33.2571** |
| | NMI(+) | 0.1129 | 0.2174 | 0.2197 | 0.2405 | 0.2518 | 0.2695 | 0.2742 | 0.2286 | **0.3209** |
| | EARI(+) | 0.2636 | 0.5774 | 0.5447 | 0.6348 | 0.8657 | 0.8879 | 0.8935 | 0.8942 | **0.8943** |
| | XB(-) | 2.529 | 1.8233 | 2.3126 | 1.3781 | 1.31 | 0.7732 | 0.7489 | 0.7190 | **0.6983** |
| Breast_cancer | ACC(+) | 0.7699 | 0.8297 | 0.8552 | 0.8897 | 0.8944 | 0.9156 | 0.9315 | 0.9315 | **0.9332** |
| | CH(+) | 40.8972 | 42.2908 | 66.2103 | 78.7936 | 80.3375 | 88.4362 | 92.565 | 94.0327 | **96.7998** |
| | NMI(+) | 0.2931 | 0.3615 | 0.4763 | 0.5293 | 0.5319 | 0.5728 | 0.626 | 0.6300 | **0.646** |
| | EARI(+) | 0.7896 | 0.7939 | 0.8012 | 0.8134 | 0.8575 | 0.8579 | 0.8735 | 0.8754 | **0.8876** |
| | XB(-) | 0.9735 | 0.839 | 0.6780 | 0.598 | 0.3747 | 0.3546 | 0.3174 | 0.2647 | **0.2647** |
| Letter_AB | ACC(+) | 0.8329 | 0.8533 | 0.8795 | 0.8853 | 0.8887 | 0.9119 | 0.9235 | 0.9286 | **0.935** |
| | CH(+) | 129.0012 | 129.8611 | 130.6947 | 134.0196 | 134.2925 | 140.1665 | 159.9534 | 164.9831 | **176.2127** |
| | NMI(+) | 0.3905 | 0.4045 | 0.478 | 0.4799 | 0.5021 | 0.5731 | 0.6424 | 0.6439 | **0.6449** |
| | EARI(+) | 0.8074 | 0.8122 | 0.7909 | 0.8135 | 0.8177 | 0.8224 | 0.8568 | 0.8582 | **0.8647** |
| | XB(-) | 1.5822 | 1.1709 | 1.0738 | 0.9006 | 0.7316 | 0.6327 | 0.6024 | 0.5633 | **0.432** |
| Wifi localization | ACC(+) | 0.5389 | 0.547 | 0.5824 | 0.6295 | 0.676 | 0.7595 | 0.924 | 0.9260 | **0.948** |
| | CH(+) | 56.0916 | 61.5782 | 63.5621 | 65.7539 | 66.3763 | 66.7259 | 68.835 | 69.4152 | **70.5556** |
| | NMI(+) | 0.4371 | 0.5406 | 0.5903 | 0.6049 | 0.6771 | 0.6815 | 0.7928 | 0.8068 | **0.8356** |
| | EARI(+) | 0.5208 | 0.6113 | 0.6645 | 0.6696 | 0.6722 | 0.7296 | 0.8599 | 0.8943 | **0.9018** |
| | XB(-) | 0.6325 | 0.5665 | 0.5696 | 0.5333 | 0.469 | 0.4269 | 0.303 | 0.3007 | **0.0906** |
| Statlog | ACC(+) | 0.7565 | 0.7614 | 0.7842 | 0.8026 | 0.8202 | 0.8511 | 0.8963 | 0.9378 | **0.9521** |
| | CH(+) | 78.2974 | 82.5009 | 89.5367 | 97.1245 | 109.7552 | 113.0569 | 234.1459 | 244.7384 | **442.7055** |
| | NMI(+) | 0.1846 | 0.2065 | 0.2288 | 0.2436 | 0.2497 | 0.3436 | 0.4019 | 0.6104 | **0.7418** |
| | EARI(+) | 0.2688 | 0.2856 | 0.3110 | 0.4138 | 0.4346 | 0.6818 | 0.8625 | 0.8849 | **0.9077** |
| | XB(-) | 1.6579 | 1.3423 | 1.3037 | 1.2869 | 0.6826 | 0.456 | 0.4016 | 0.2714 | **0.1303** |

obtain nice values. The VWKFC algorithm achieves best results in terms of all evaluation indexes.

Why is the VWKFC algorithm the best here? In fact, VWKFC achieves the initialization of the cluster centers through KHDI, which avoids the algorithm falling into the local optimal value due to improper initialization of the cluster centers. In the process of clustering, the selection of features in VWKFC is realized through feature weights, making the clustering result more accurate. The influence of noise and outliers on clustering effect is reduced by introducing high-density viewpoint obtained by KHDI.

### C. UCI Datasets

Furthermore, we evaluate the performance of VWKFC on 8 UCI datasets. Table IV shows the specific details of these UCI datasets, including the number of data samples, the number of attributes, and the reference classes. Table V shows the index values obtained by each clustering algorithm running on different UCI datasets.

From Table V, FCM and KFCM show a more modest effect. As a non FCM clustering method, RLM performs better than FCM and KFCM. By introducing viewpoints, V-FCM gets better results than FCM, KFCM and RLM. Then, EWFCM and FRFCM obtain more excellent performance. In virtue of using RLM-based initialization strategy and viewpoint idea, DVPFCM obtains more perfect performance. In addition, by using the optimization mechanism of PSO, the H-FCM-PSO algorithm also obtains good results. But VWKFC is still the best one.

For example, in the Iris dataset, the ACC value of VWKFC reaches 0.9467. On datasets with higher feature dimensions, such as Breast_cancer, Spect and Statlog, the ACC value of VWKFC reaches 0.9332, 0.8543 and 0.9521. This indicates that VWKFC has certain advantages over previous algorithms in processing datasets with more attribute characteristics. Compared with the traditional feature weight clustering algorithms (e.g., EWFCM and FRFCM), the VWKFC algorithm initializes the cluster center and introduces the viewpoint through the KHDI algorithm to avoid that the objective function falls into the

Authorized licensed use limited to: Chulalongkorn University provided by UniNet. Downloaded on January 17,2024 at 08:32:14 UTC from IEEE Xplore. Restrictions apply.

TABLE VI
PERCENTAGE IMPROVEMENT IN ACC OF VWKFC COMPARED WITH RELATED ALGORITHMS

| | FCM | KFCM | RLM | V-FCM | EWFCM | FRFCM | DVPFCM | H-FCM-PSO |
|---|---|---|---|---|---|---|---|---|
| Data-a | 0.0806 | 0.0454 | 0.0274 | 0.0137 | 0.0101 | 0.0067 | 0.0033 | 0.0033 |
| Data-b | 0.5477 | 0.4354 | 0.1563 | 0.1149 | 0.0399 | 0.002 | 0.0007 | 0.0010 |
| Iris | 0.5857 | 0.2988 | 0.1738 | 0.0923 | 0.0597 | 0.0365 | 0.0262 | 0.0142 |
| Wine | 0.3378 | 0.2055 | 0.1621 | 0.1127 | 0.0908 | 0.0833 | 0.0195 | 0.0063 |
| Seeds | 0.1317 | 0.1273 | 0.0849 | 0.0781 | 0.0558 | 0.0106 | 0.0053 | 0.0006 |
| spect | 0.4923 | 0.1619 | 0.1142 | 0.1018 | 0.0759 | 0.0752 | 0.0137 | 0.0759 |
| Breast_cancer | 0.2121 | 0.1247 | 0.0912 | 0.0488 | 0.0433 | 0.0192 | 0.0018 | 0.0018 |
| Letter_AB | 0.1225 | 0.0957 | 0.0631 | 0.0561 | 0.052 | 0.0253 | 0.0124 | 0.0068 |
| Wifi localization | 0.7591 | 0.733 | 0.5824 | 0.5059 | 0.4023 | 0.2481 | 0.0259 | 0.0232 |
| Statlog | 0.2585 | 0.2504 | 0.2141 | 0.1862 | 0.1608 | 0.1186 | 0.0622 | 0.0150 |
| The average | 0.3528 | 0.2478 | 0.1669 | 0.131 | 0.099 | 0.0625 | 0.0171 | 0.0148 |

TABLE VII
AVERAGE NUMBER OF ITERATIONS OF THE ALGORITHMS RUNNING ON DIFFERENT DATASETS

| | FCM | KFCM | V-FCM | EWFCM | FRFCM | DVPFCM | H-FCM-PSO | VWKFC |
|---|---|---|---|---|---|---|---|---|
| Data-a | 13(3) | 43(8) | 20(6) | 31(7) | 6(2) | 20(5) | 20(4) | **6(1)** |
| Data-b | 31(5) | 52(8) | 29(4) | 20(3) | **4(1)** | 32(7) | 32(6) | 9(2) |
| Iris | 17(3) | 29(5) | 18(4) | 37(6) | 6(2) | 152(8) | 52(7) | **5(1)** |
| Wine | 25(5) | 41(6) | 23(4) | 67(8) | **2(1)** | 22(3) | 48(7) | 9(2) |
| Seeds | 19(3) | 29(6) | 26(4) | 111(8) | **5(1)** | 28(5) | 46(7) | 8(2) |
| spect | 15(4) | 32(7) | 5(3) | 37(8) | 5(2) | 32(6) | 27(5) | **4(1)** |
| Breast_cancer | 15(4) | 37(7) | 7(2) | 85(8) | **2(1)** | 19(5) | 25(6) | 14(3) |
| Letter_AB | 22(5) | 15(3) | 9(2) | 74(8) | **3(1)** | 16(4) | 28(6) | 39(7) |
| Wifi localization | 41(7) | 61(8) | 36(5) | 39(6) | **3(1)** | 15(3) | 30(4) | 11(2) |
| Statlog | 16(4) | 25(6) | 16(3) | 49(7) | 7(2) | 57(8) | 24(5) | **4(1)** |
| The average | 21.4(4) | 36.4(6) | 18.9(3) | 55.0(8) | **4.3(1)** | 39.3(7) | 33.2(5) | 10.9(2) |

local minimum point due to improper initialization of cluster centers. Compared with V-FCM and DVPFCM, the VWKFC algorithm can obtain better clustering effect when processing high-dimensional data by assigning weights to features.

Table VI shows the percentage improvement of the VWKFC algorithm in ACC compared with the comparison algorithms in each dataset, including artificial and UCI datasets. From Table VI, we get 1.48% improvement of the proposed VWKFC in ACC compared with H-FCM-PSO, and 1.71% improvement than DVPFCM, and 6.25% improvement than FRFCM, and 9.90% improvement than EWFCM, and 13.10% improvement than V-FCM, and finally 16.69% improvement than RLM.

Besides, we calculate the average number of iterations of each algorithm on different datasets, which is shown in Table VII. Since there is no iterative process in the running process of the RLM algorithm, we do not compare it with RLM here. We sort the average number of iterations of different algorithms on the same dataset in ascending order, and the numbers in parentheses represent the positions that have been sorted. As can be seen, the average numbers of iterations of VWKFC on Data-a, Iris, Spect, Statlog are the lowest among all algorithms, and the average numbers of iterations of VWKFC on Data-b, Wine, Seeds and Wifi localization are the second among all algorithms. It is worth noting that, on the dataset of Letter_AB, the average number of iterations of VWKFC is 39, which is greater than the number of iterations of all algorithms except EWFCM. This indicates that VWKFC may be not perfect in processing dataset of character recognition. From the perspective of the average number of iterations, FRFCM is the best while VWKFC is the second best. Experiments manifest that VWKFC can converge after fewer iterations than other algorithms. This is mainly due to the introduction of more accurate initial cluster centers and viewpoint by KHDI, which greatly accelerates the convergence speed.

In Table VIII, we calculate the average running time of each algorithm on different datasets. Here we have sorted the average running time of these algorithms in ascending order and marked the sorted position in parentheses. As can be seen, compared with some algorithms, VWKFC needs more running time although it runs fewer iterations. This is because the VWKFC algorithm in addition needs to calculate the weight matrix. However, the extra cost in running time is worth due to the resulting benefit of VWKFC.

### D. ORL Face Dataset

We exhibit the results on the ORL face dataset. To verify the effect of VWKFC in face clustering, we use a total of 200 face images of the last 20 people in ORL face dataset. The evaluation indexes ACC and CHI are used to evaluate the clustering effect. Table IX shows the results of each algorithm on ORL face dataset. Among them, the index ACC of VWKFC on ORL face dataset is 0.7032, and the index CH is 10.6749. Experiments show that the VWKFC algorithm can achieve the best clustering effect on the ORL face dataset.

### E. High-Dimensional Datasets

We show the running results of comparison algorithms on two high-dimensional datasets, including Libras (with 90 dimensions and 360 samples) and D256 [34] (with 256 dimensions and 1024 samples). From Table X, we find that DVPFCM gets good result and VWKFC obtains the best performance. This shows that the effect of the VWKFC algorithm is also ideal for high-dimensional datasets.

### F. Datasets With Various Shapes

Table XI show the results on three datasets with various shapes, which include Pathbased, Jain, and Spiral. From Table XI, VWKFC and RLM obtain the best result from five evaluation indexes. In detail, Path-based and Jain have relatively dense data in the same cluster, so distance-based clustering algorithm can get a better effect. As for the Spiral dataset, data is spiral and the data distribution shows the characteristic shape of dataset ontology, so density-based clustering method (e.g. RLM) is suitable, meanwhile VWKFC also works fine.

### G. Determining the Number of Clusters

Here we can provide useful information on how to determine the number of clusters. From the KHDI algorithm, the value of $\tau$ of most data points is almost close to 0, but there exist some special points with high $\tau$ values. The number of these special points gives the suggestion of clustering number. For example, in Fig. 3(b), there exist 7 special points with high $\tau$ values, which implies that the clustering number may be 7 (noting that the ground-truth clustering number is 7).

Authorized licensed use limited to: Chulalongkorn University provided by UniNet. Downloaded on January 17,2024 at 08:32:14 UTC from IEEE Xplore. Restrictions apply.

TABLE VIII
AVERAGE RUNNING TIME OF EACH ALGORITHM ON DIFFERENT DATASETS

| | FCM | KFCM | V-FCM | EWFCM | FRFCM | DVPFCM | H-FCM-PSO | VWKFC |
|---|---|---|---|---|---|---|---|---|
| Data-a | **0.0510(1)** | 0.2340(7) | 0.1130(3) | 0.0540(2) | 0.4130(8) | 0.2260(6) | 0.1749(5) | 0.1580(4) |
| Data-b | 3.1330(4) | 6.8050(7) | 4.9050(6) | 17.6750(8) | **0.1290(1)** | 0.6544(2) | 4.6653(5) | 2.8550(3) |
| Iris | 0.0520(2) | 0.1400(5) | 0.1100(4) | 0.2060(7) | **0.0300(1)** | 4.1430(8) | 0.2251(6) | 0.0900(3) |
| Wine | 0.1460(3) | 0.2900(4) | 0.1300(2) | 1.9080(7) | **0.0500(1)** | 5.0130(8) | 0.3955(6) | 0.3800(5) |
| Seeds | 0.0950(2) | 0.2200(4) | 0.2300(5) | 10.4880(8) | **0.0700(1)** | 5.8260(7) | 0.3725(6) | 0.2100(3) |
| spect | **0.0250(1)** | 0.1300(2) | 0.1500(4) | 63.1930(8) | 0.1300(3) | 0.3810(6) | 0.2660(5) | 0.4990(7) |
| Breast_cancer | **0.1490(1)** | 0.1800(2) | 0.6380(5) | 22.3940(8) | 0.5270(3) | 1.3410(6) | 0.5373(4) | 1.7550(7) |
| Letter_AB | **0.4170(1)** | 0.9960(3) | 4.2370(5) | 12.8140(7) | 0.4390(2) | 8.7850(6) | 1.4623(4) | 15.9260(8) |
| Wifi localization | 2.6730(2) | 6.0380(4) | 9.4950(5) | 19.9190(8) | **0.2600(1)** | 19.8210(7) | 5.3542(3) | 11.0970(6) |
| Statlog | **0.5640(1)** | 1.2860(2) | 9.3530(5) | 72.2470(8) | 2.5410(4) | 53.4610(7) | 1.9877(3) | 17.9110(6) |
| The average | 0.7305(2) | 1.6319(4) | 2.9361(5) | 22.0898(8) | **0.4589(1)** | 9.9651(7) | 1.5441(3) | 5.0881(6) |

TABLE IX
RUNNING RESULTS OF THE VWKFC ALGORITHM ON THE ORL FACE DATASET

| | FCM | KFCM | RLM | V-FCM | EWFCM | FRFCM | DVPFCM | H-FCM-PSO | VWKFC |
|---|---|---|---|---|---|---|---|---|---|
| ACC(+) | 0.23 | 0.3 | 0.3759 | 0.45 | 0.5212 | 0.6305 | 0.6822 | 0.5700 | **0.7032** |
| CH(+) | 1.0168 | 3.039 | 3.1288 | 3.342 | 4.7389 | 7.4563 | 7.5977 | 7.8624 | **10.6749** |

TABLE X
RUNNING RESULTS OF COMPARISON ALGORITHMS ON TWO HIGH-DIMENSIONAL DATASETS

| | | FCM | KFCM | RLM | V-FCM | EWFCM | FRFCM | DVPFCM | H-FCM-PSO | VWKFC |
|---|---|---|---|---|---|---|---|---|---|---|
| Libras | ACC(+) | 0.3500 | 0.3639 | 0.3667 | 0.3750 | 0.3917 | 0.4139 | 0.4278 | 0.4250 | **0.4306** |
| | CH(+) | 0.7197 | 0.8205 | 0.8971 | 0.8386 | 0.9255 | 0.9315 | 1.0157 | 1.0181 | **1.0611** |
| | NMI(+) | 0.4072 | 0.4697 | 0.4720 | 0.4766 | 0.5034 | 0.5259 | 0.5493 | 0.5356 | **0.5582** |
| | EARI(+) | 0.3903 | 0.5114 | 0.5135 | 0.5205 | 0.5347 | 0.5934 | 0.6092 | 0.6201 | **0.6404** |
| | XB(-) | 2.7475 | 1.9227 | 1.9376 | 1.8947 | 1.6962 | 0.7396 | 0.3064 | 0.2590 | **0.2458** |
| Dim256 | ACC(+) | 0.8125 | 0.8281 | 0.8496 | 0.8750 | 0.9131 | 0.8994 | 0.9189 | 0.9209 | **0.9229** |
| | CH(+) | 2.3081 | 2.7514 | 2.8341 | 2.8415 | 2.8657 | 2.8571 | 2.8661 | 2.8865 | **2.8876** |
| | NMI(+) | 0.8660 | 0.8755 | 0.9048 | 0.9121 | 0.9467 | 0.9291 | 0.9477 | 0.9503 | **0.9679** |
| | EARI(+) | 0.8849 | 0.9186 | 0.9199 | 0.9325 | 0.9566 | 0.9698 | 0.9617 | 0.9753 | **0.9837** |
| | XB(-) | 122.1710 | 83.3129 | 68.4021 | 57.6211 | 45.1444 | 42.5325 | 33.2638 | 27.9157 | **22.7201** |

TABLE XI
RUNNING RESULTS OF COMPARISON ALGORITHMS ON DATASETS WITH VARIOUS SHAPES

| | | FCM | KFCM | RLM | V-FCM | EWFCM | FRFCM | DVPFCM | H-FCM-PSO | VWKFC |
|---|---|---|---|---|---|---|---|---|---|---|
| Pathbased | ACC(+) | 0.6567 | 0.6867 | 0.7333 | 0.7067 | 0.7100 | 0.7133 | 0.7267 | 0.7433 | **0.7467** |
| | CH(+) | 12.9497 | 13.1794 | 22.4901 | 20.6823 | 20.8275 | 22.1214 | 23.0783 | 23.2108 | **23.4285** |
| | NMI(+) | 0.3833 | 0.4580 | 0.5404 | 0.4749 | 0.4938 | 0.5135 | 0.5362 | 0.5493 | **0.5512** |
| | EARI(+) | 0.6558 | 0.6872 | 0.6318 | 0.7545 | 0.7709 | 0.7900 | 0.8140 | 0.8166 | **0.8240** |
| | XB(-) | 0.6827 | 0.4655 | 0.4897 | 0.3834 | 0.3566 | 0.2749 | 0.2429 | 0.2425 | **0.2387** |
| Jain | ACC(+) | 0.7668 | 0.8150 | 0.8740 | 0.8204 | 0.8633 | 0.8686 | 0.8794 | 0.8740 | **0.8847** |
| | CH(+) | 19.2806 | 25.8889 | 58.8733 | 29.3787 | 31.5121 | 48.6993 | 51.0734 | 55.5433 | **61.2501** |
| | NMI(+) | 0.3474 | 0.3524 | 0.3888 | 0.4398 | 0.5177 | 0.5221 | 0.5435 | 0.5325 | **0.5547** |
| | EARI(+) | 0.5595 | 0.6287 | 0.6495 | 0.7237 | 0.8155 | 0.8225 | 0.8355 | 0.8347 | **0.8446** |
| | XB(-) | 0.8460 | 0.5788 | 0.5771 | 0.5481 | 0.5331 | 0.3030 | 0.2806 | 0.2755 | **0.2519** |
| Spiral | ACC(+) | 0.3462 | 0.3494 | **1.0000** | 0.3526 | 0.3590 | 0.3654 | 0.3622 | 0.3622 | 0.3654 |
| | CH(+) | 14.3450 | 16.5469 | 18.0545 | 18.3157 | 21.5884 | 21.0739 | 21.0144 | 22.1455 | **22.7596** |
| | NMI(+) | 0.0016 | 0.0019 | **1.0000** | 0.0023 | 0.0027 | 0.0030 | 0.0032 | 0.0032 | 0.0032 |
| | EARI(+) | 0.5227 | 0.5840 | **1.0000** | 0.6981 | 0.7473 | 0.8128 | 0.8031 | 0.8073 | 0.8131 |
| | XB(-) | 0.9251 | 0.5568 | 0.3744 | 0.5147 | 0.3914 | 0.3191 | 0.3487 | 0.3171 | **0.2852** |

To discover the issue of determine the number of clusters, we employ some commonly-used clustering validity indexes including the XB index, the Wu-and-Li index (WLI) [35] together with the imbalanced index (IMI) [36]. Now, supposing that we do not know the clustering number of an artificial dataset Data-b and a UCI dataset Aggregation. We employ the clustering number corresponding to the extreme values of these cluster validity indexes as the ideal number of clusters. Table XII shows the optimal number of clusters obtained by the VWKFC algorithm on the strength of these indexes. We observe that the results of XB, WLI and IMI show that the ideal clustering number of the Data-b and Aggregation datasets is all 7. So the VWKFC algorithm can recognize how many clusters are in the Data-b and Aggregation datasets. From such point of view, it can help find the correct clustering number.

Authorized licensed use limited to: Chulalongkorn University provided by UniNet. Downloaded on January 17,2024 at 08:32:14 UTC from IEEE Xplore. Restrictions apply.

TABLE XII
THREE CLUSTER VALIDITY INDEXES FOR THE TWO DATASETS USING VWKFC

| | | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Data-b | XB(-) | 0.3404 | 0.5767 | 0.7809 | 0.5242 | 0.3564 | **0.1046** | 0.7289 | 0.7048 | 0.6300 | 0.6728 | 0.6844 | 0.6655 | 0.6154 | 0.6904 |
| | WLI(-) | 0.0667 | 0.0828 | 0.0540 | 0.0840 | 0.0477 | **0.0446** | 0.0622 | 0.0540 | 0.0567 | 0.0538 | 0.0462 | 0.0461 | 0.0472 | 0.0589 |
| | IMI(-) | 0.2024 | 0.2025 | 0.1777 | 0.1420 | 0.1651 | **0.1110** | 0.1439 | 0.1224 | 0.1584 | 0.1479 | 0.1275 | 0.1358 | 0.1311 | 0.1417 |
| Aggregation | XB(-) | 0.5743 | 0.7348 | 0.6829 | 0.6656 | 0.5816 | **0.5497** | 0.5957 | 0.5665 | 0.7603 | 0.7383 | 0.7669 | 0.7440 | 0.7007 | 0.7104 |
| | WLI(-) | **0.1618** | 0.4807 | 0.4239 | 0.4175 | 0.4054 | 0.3952 | 0.4063 | 0.4335 | 0.4881 | 0.4666 | 0.5129 | 0.5146 | 0.5710 | 0.5094 |
| | IMI(-) | 0.2820 | 0.3335 | 0.2762 | 0.2548 | 0.2505 | **0.2381** | 0.2590 | 0.2715 | 0.2636 | 0.3209 | 0.2556 | 0.3753 | 0.4721 | 0.3659 |

## IV. CONCLUSION

In this study, we propose the density viewpoint-based weighted kernel fuzzy clustering (VWKFC) algorithm. The main work and contributions are as follows:

i) We establish the kernel-based hypersphere density initialization (KHDI) algorithm. We use the kernel distance instead of the Euclidean distance to calculate local density of data points. A more accurate initialization of the cluster centers is provided before the algorithm iteration. It prevents the algorithm from falling into a local minimum due to improper initialization of the cluster centers.

ii) We put forward a novel computing method for density viewpoints. A new density radius is given in KHDI, and then the data point with the highest local density obtained by KHDI is regarded as the viewpoint, and the structure of the whole data is observed on this basis. By this way, the influence of noise and outliers on the cluster center during iteration is reduced to a certain extent.

iii) The concept of the weight information granule is proposed. In order to obtain good results when the algorithm processed datasets with high feature dimensions and clustering irrelevant feature attributes, we introduce feature weight matrix. Different weights are assigned to different features to reduce the influence of unrelated features in clustering process. At the same time, we assign a sample weight to each data point, resulting in that the influence of noise and outliers to the clustering can be controlled to be small.

iv) Based on the kernel function, viewpoints obtained from the KHDI algorithm, weight information granules as well as a maximum entropy regularization item, we propose a new fuzzy clustering algorithm called the VWKFC algorithm. And its convergence theorem is proved.

v) Experimental results verify that VWKFC is superior over the comparison algorithms in both artificial datasets and UCI datasets with regard to five evaluation indexes. The number of iterations required by the VWKFC algorithm is generally lower than that of the comparison algorithms. This implies that the proposed algorithm is able to achieve convergence at a faster speed. Moreover, main parameters in VWKFC are analyzed. We find that VWKFC can provide useful information on how to determine the number of clusters, which employ three acknowledged clustering validity indexes. Finally, the proposed algorithm is applied to ORL face dataset, high-dimensional datasets, datasets with various shapes, and obtains best clustering result.

In the future work, we will further explore the application of the VWKFC algorithm to more practical fields which include text clustering, data mining, image processing and others. We hope that the proposed VWKFC algorithm can play a vital role in different fields. In addition, we are going to combine fuzzy clustering with fuzzy reasoning [37], [38].

## REFERENCES


[1] T. Campbell, B. Kulis, and J. How, "Dynamic clustering algorithms via small-variance analysis of markov chain mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 6, pp. 1338–1352, Jun. 2019.

[2] N. Gupta, S. Ari, and N. Panigrahi, "Change detection in landsat images using unsupervised learning and RBF-based clustering," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 5, no. 2, pp. 284–297, Apr. 2021.

[3] W. Huang, S. K. Oh, and W. Pedrycz, "Fuzzy wavelet polynomial neural networks: Analysis and design," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 5, pp. 1329–1341, Oct. 2017.

[4] X. T. Li and K. C. Wong, "Evolutionary multiobjective clustering and its applications to patient stratification," *IEEE Trans. Cybern.*, vol. 49, no. 5, pp. 1680–1693, May 2019.

[5] Y. H. Shen, W. Pedrycz, and X. Wang, "Clustering homogeneous granular data: Formation and evaluation," *IEEE Trans. Cybern.*, vol. 49, no. 4, pp. 1391–1402, Apr. 2019.

[6] P. Jha, A. Tiwari, N. Bharill, M. Ratnaparkhe, M. Mounika, and N. Nagendra, "A novel scalable kernelized fuzzy clustering algorithms based on in-memory computation for handling Big Data," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 5, no. 6, pp. 908–919, Dec. 2021.

[7] E. Zhu, Y. Zhang, P. Wen, and F. Liu, "Fast and stable clustering analysis based on grid-mapping k-means algorithm and new clustering validity index," *Neurocomputing*, vol. 363, pp. 149–170, Oct. 2019.

[8] Y. M. Tang, F. J. Ren, and W. Pedrycz, "Fuzzy C-means clustering through SSIM and patch for image segmentation," *Appl. Soft Comput.*, vol. 87, pp. 1–16, Feb. 2020.

[9] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, Jun. 2014.

[10] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. San Francisco, CA, USA: Academic, 1981.

[11] W. Pedrycz, A. Amato, V. D. Lecce, and V. Piuri, "Fuzzy clustering with partial supervision in organization and classification of digital images," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 4, pp. 1008–1026, Aug. 2008.

[12] K. Xu, W. Pedrycz, Z. Li, and W. Nie, "Constructing a virtual space for enhancing the classification performance of fuzzy clustering," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 9, pp. 1779–1792, Sep. 2019.

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

[14] N. R. Pal, K. Pal, and J. C. Bezdek, "A possibilistic fuzzy c-means clustering algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 4, pp. 517–530, Aug. 2005.

[15] S. C. Chen and D. Q. Zhang, "Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure," *IEEE Trans. Syst. Man, Cybern. B., Cybern.*, vol. 34, no. 4, pp. 1907–1916, Aug. 2004.

[16] L. Chen, C. L. Chen, and M. Lu, "A multiple-kernel fuzzy c-means algorithm for image segmentation," *IEEE Trans. Syst. Man Cybern. Part B. Cybern.*, vol. 41, no. 5, pp. 1263–1274, Oct. 2011.

[17] Y. Ding and X. Fu, "Kernel-based fuzzy c-means clustering algorithm based on genetic algorithm," *Neurocomputing*, vol. 188, pp. 233–238, May 2016.

[18] J. Zhou, L. Chen, C. L. P. Chen, Y. Zhang, and H. X. Li, "Fuzzy clustering with the entropy of attribute weights," *Neurocomputing*, vol. 198, pp. 125–134, Jul. 2016.


Authorized licensed use limited to: Chulalongkorn University provided by UniNet. Downloaded on January 17,2024 at 08:32:14 UTC from IEEE Xplore. Restrictions apply.

[19] M. S. Yang and Y. Nataliani, "A feature-reduction fuzzy clustering algorithm based on feature-weighted entropy," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 2, pp. 817–835, Apr. 2018.
[20] H. Verma, D. Verma, and P. K. Tiwari, "A population based hybrid FCM-PSO algorithm for clustering analysis and segmentation of brain image," *Expert Syst. Appl.*, vol. 167, Apr. 2021, Art. no. 114121.
[21] W. Pedrycz, V. Loia, and S. Senatore, "Fuzzy clustering with viewpoints," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 2, pp. 274–284, Apr. 2010.
[22] Y. M. Tang, X. H. Hu, W. Pedrycz, and X. C. Song, "Possibilistic fuzzy clustering with high-density viewpoint," *Neurocomputing*, vol. 329, pp. 407–423, Feb. 2019.
[23] W. I. Zangwill, *Nonlinear Programming: A Unified Approach*. Englewood Cliffs, NJ, USA, Prentice-Hall, 1969.
[24] F. Werner and Y. N. Sotskov, *Math. of Econ. and Bus.*, London, U.K.: Routledge, 2006.
[25] N. Hindman, "Basically bounded sets and a generalized Heine-Borel theorem," *Am Math. Mon.*, vol. 80, no. 5, pp. 549–552, 1973.
[26] A. Asuncion and D. J. Newman, "UCI machine learning repository," School of Information and Computer Science, University of California, 2007. [online]. Available: http://archive.ics.uci.edu/ml/index.php
[27] D. Graves and W. Pedrycz, "Kernel-based fuzzy clustering and fuzzy clustering: A comparative experimental study," *Fuzzy Sets Syst.*, vol. 161, no. 4, pp. 522–543, Feb. 2010.
[28] A. Strehl and J. Ghosh, "Cluster ensembles - a. knowledge reuse framework for combining multiple partitions," *J. Mach. Learn. Res.*, vol. 3, pp. 583–617, Dec. 2002.
[29] T. Calinski and J. Harabasz, "A dendrite method for cluster analysis," *Commun. Stat.*, vol. 3, no. 1, pp. 1–27, Sep. 1972.
[30] R. J. G. B. Campello, "A fuzzy extension of the rand index and other related indexes for clustering and classification assessment," *Pattern Recognit. Lett.*, vol. 28, no. 7, pp. 833–841, May 2007.
[31] H. C. Huang, Y. Y. Chuang, and C. S. Chen, "Multiple kernel fuzzy clustering," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 1, pp. 120–134, Feb. 2012.
[32] L. Hubert and P. Arabie, "Comparing partitions," *J. Classification.*, vol. 2, no. 1, pp. 193–218, Dec. 1985.
[33] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 8, pp. 841–847, Aug. 1991.
[34] P. Fränti, O. Virmajoki, and V. Hautamäki, "Fast agglomerative clustering using a k-nearest neighbor graph," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1875–1881, Nov. 2006.
[35] C. H. Wu, C. S. Ouyang, L. W. Chen, and L. W. Lu, "A new fuzzy clustering validity index with a median factor for centroid-based clustering," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 3, pp. 701–718, Jun. 2015.
[36] Y. Liu, Y. Jiang, T. Hou, and F. Liu, "A new robust fuzzy clustering validity index for imbalanced data sets," *Inform. Sci.*, vol. 547, pp. 579–591, Feb. 2021.
[37] Y. M. Tang, W. Pedrycz, and F. J. Ren, "Granular symmetric implicational method," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 6, no. 3, pp. 710–723, Jun. 2022.
[38] Y. M. Tang and W. Pedrycz, "Oscillation-bound estimation of perturbations under Bandler-Kohout subproduct," *IEEE Trans. Cybern.*, vol. 52, no. 7, pp. 6269–6282, Jul. 2022.

**Yiming Tang** (Member, IEEE) received the Ph.D. degree from the Hefei University of Technology, Hefei, China, in 2011. He is currently an Associate Professor with the Hefei University of Technology and is also a Visiting Professor with the University of Alberta, Edmonton, AB, Canada. He has authored or coauthored more than 80 academic papers, such as the IEEE TRANSACTIONS ON CYBERNETICS and the IEEE TRANSACTIONS ON FUZZY SYSTEMS. His research interests include clustering, fuzzy reasoning, fuzzy system, granular computing, image processing, and affective computing. He is an Associate Editor for the *Information Sciences*, the IEEE ACCESS and the *Journal of Intelligent and Fuzzy Systems*, a Member of Editorial Board of the *PLOS ONE*, a Guest Editor of the *Symmetry*, and a Senior Member of China Computer Federation (CCF), and also a Senior Member of the Chinese Association for Artificial Intelligence (CAAI). He is a Professional Committee of Machine Learning of CAAI, a Professional Committee of Granular Computing and Knowledge Discovery of CAAI, an Executive Committee of Cooperative Computing of CCF, and a Professional Committee of Non-classical Logic and Computing of Chinese Society of Logic. He is a reviewer of more than 40 journals and conferences. He was the recipient of the IEEE ACCESS Outstanding Associate Editor of 2021.

**Zhifu Pan** received the bachelor's degree in 2019 from Hubei University, Wuhan, China. He is currently working toward the master's degree. His research interests fuzzy clustering, granular computing, and image processing.

**Witold Pedrycz** (Life Fellow, IEEE) received the M.Sc., Ph.D., and D.Sc. degrees from the Silesian University of Technology, Gliwice, Poland, in 1977, 1980, and 1984, respectively. He is currently a Professor and the Canada Research Chair of computational intelligence with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. He is also with the Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland, where he was elected as a Foreign Member in 2009. He also holds an appointment as Special Professor with the School of Computer Science, University of Nottingham, Nottingham, U.K. He is the author of 14 research monographs that cover various aspects of computational intelligence and software engineering. His research interests include computational intelligence, fuzzy modeling and granular computing, knowledge discovery and data mining, fuzzy control, and software engineering. He is an Editor-in-Chief of the *Information Sciences*. He is also an Associate Editor for the IEEE TRANSACTIONS ON FUZZY SYSTEMS and the IEEE TRANSACTIONS ON SYSTEMS, *Man*, and *Cybernetics and Systems*. He has edited a number of volumes. He was a Member of numerous program committees of IEEE conferences in the area of fuzzy sets and *neurocomputing*. He was the recipient of the prestigious Norbert Wiener Award from the IEEE SYSTEMS, *Man*, and *Cybernetics Council* in 2007, the IEEE Canada Computer Engineering Medal in 2008 and the Cajastur Prize for Soft Computing in 2009 from the European Centre for Soft Computing for pioneering and multifaceted contributions to granular computing. He is a Fellow of the Royal Society of Canada.

**Fuji Ren** (Senior Member, IEEE) received the Ph.D. degree in 1991 from the Faculty of Engineering, Hokkaido University, Sapporo, Japan. From 1991 to 1994, he was with CSK as a Chief Researcher. In 1994, he joined the Faculty of Information Sciences, Hiroshima City University, Hiroshima, Japan, as an Associate Professor. Since 2001, he has been a Professor with the Faculty of Engineering, Tokushima University, Tokushima, Japan. He is a Chair Professor with the University of Electronic Science and Technology of China, Chengdu, China, from 2022. His research interests include natural language processing, artificial intelligence, affective computing, and emotional robot. He is the Academician of the Engineering Academy of Japan and EU Academy of Sciences. He is a Senior Member of IEEE, Editor-in-Chief of the *International Journal of Advanced Intelligence*, a Vice President of CAAI, and a Fellow of the Japan Federation of Engineering Societies, Fellow of IEICE, Fellow of CAAI. He is the President of International Advanced Information Institute, Japan.

**Xiaocheng Song** received the bachelor's degree in 2015 from Huaibei Normal University, Huaibei, China. He is curently working toward the master's degree. His research interests include fuzzy clustering and image processing.

Authorized licensed use limited to: Chulalongkorn University provided by UniNet. Downloaded on January 17,2024 at 08:32:14 UTC from IEEE Xplore. Restrictions apply.