Received January 27, 2020, accepted February 2, 2020, date of publication February 7, 2020, date of current version February 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2972333

An Enhanced Regularized k-Means Type Clustering Algorithm With Adaptive Weights

ZIHENG WU [1] **AND ZIXIANG WU** [2]

[1]School of Electrical and Information Engineering, Anhui University of Technology, Ma'anshan 243000, China
[2]Department of Thoracic Surgery, Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310000, China

Corresponding author: Ziheng Wu (wziheng@ahut.edu.cn)

This work was supported in part by the Natural Science Foundation of Anhui Province under Grant 1908085QF253, in part by the National Natural Science Foundation of China under Grant 61601004 and Grant 61602008, and in part by the Natural Science Foundation for Young Scientists of Anhui University of Technology under Grant QZ201812.

ABSTRACT K-means clustering algorithm is one of the most popular technique for clustering in machine learning, however, in the existing k-means clustering algorithm, the ability of the different features and the importance of the different data objects are treated equally; the discriminative ability of the different features and the importance of the different data objects cannot be differentiated effectively. In the light of this limitation, this paper put forward an enhanced regularized k-means type clustering algorithm with adaptive weights in which we introduced an adaptive feature weights matrix and an adaptive data weights vector into the objective function of the k-means clustering algorithm and we developed a new objective function with l^2-norm regularization to the weights of data objects and features, then we obtained the corresponding scientific updating iterative rules of the weights of the different features, the weights of the different data objects and the cluster centers theoretically. In order to evaluate the performance of the new algorithm put forward, extensive experiments were conducted. Experimental results have indicated that our proposed algorithm can improve the clustering performance significantly and are more effective with respects to three metrics: the successful clustering rate (*SCR*), normal mutual information (*NMI*) and RandIndex.

INDEX TERMS K-means clustering algorithm, machine learning, adaptive weights, l^2-norm regularization.

I. INTRODUCTION

Machine learning is an important branch of artificial intelligence, one main task of machine learning is to study and build a mathematical model which can learn from the set of input data. According to the different type of the set of input data, Machine learning models can be classified into supervised learning in which the set of input data contains desired output labels or unsupervised learning in which the set of input data contains no output labels. For some practical problems, it is not possible to get data labels [1], so it is obviously more suitable to apply unsupervised learning methods such as clustering for solving these problems. For unsupervised classification, clustering is a key technique in data analysis, the purpose of clustering is to partition a set of data objects into several groups in which the objects in a same group have higher similarities and the objects in different groups are far from each other according to a certain pre-defined similarity measure [2]. The clustering algorithm can be segmented into model-based method, density-based method, division-based method, mesh-base method and hierarchical-based method [3]. Clustering has been successfully employed to solve various real-life problems, with applications ranging from community discover [4], text organization [5], image segmentation [6] to bioinformatics [7] and so on.

K-means clustering algorithm is a hard clustering method which assumes that a data object either belongs to one cluster or does not belong to one cluster. Because of effectiveness and simplicity, k-means clustering algorithm has been one of the most widely used clustering algorithms for addressing various real-life problems [8]–[12].

Although the extensions of k-means clustering algorithm have improved the clustering quality, yet the existing k-means clustering algorithms cannot effectively distinguish

The associate editor coordinating the review of this manuscript and approving it for publication was Qilian Liang.

the discriminative ability of the different features and the importance of the different data objects at the same time, few studies have focused on the importance of the discriminative ability of the different features and the different data objects simultaneously. To this end, based on the previous research, in this paper, we proposed an enhanced regularized k-means type clustering algorithm with adaptive weights, in the new clustering algorithm, considering the different feature weights and data weights, an adaptive data weights vector and an adaptive feature weights matrix are introduced into the conventional k-means clustering algorithm, to encourage more data objects and features to participate in the identification of clusters, a new objective function with an l^2-norm regularization to the weights of data objects and features is constructed. By the proposed new objective function, the corresponding scientific updating iterative rules of the membership matrix, the weights of the different features, the weights of the different data objects and the cluster centers can be derived theoretically. In order to evaluate the performance of the new algorithm put forward, extensive experiments were conducted. Experimental results have corroborated that our proposed algorithm can improve the clustering performance significantly and are more effective with respects to three metrics: the successful clustering rate (*SCR*), normal mutual information (*NMI*) and RandIndex. The main contributions of this paper are as follows:

- We put forward an enhanced regularized k-means type clustering algorithm with adaptive weights. In the proposed algorithm, different from the existing k-means algorithms, the discriminative ability of the different features and the importance of the different data objects can be distinguished effectively at the same time.
- We added an l^2-norm regularization to the weights of data objects and features to encourage more data objects and features to participate in the identification of clusters.
- We developed a novel objective function for the regularized k-means type clustering algorithm with adaptive weights and give the scientific updating iterative rules by optimizing the corresponding objective function.

The rest of this paper is organized as follows: Section 2 introduces a brief background about k-means clustering algorithm. The details of our proposed algorithm are presented in Section 3. Experiments on different datasets are presented in Section 4. Finally, we conclude this paper in Section 5.

II. RELATED WORKS

Although the k-means clustering algorithm has a good performance in clustering detection, there still some shortcomings in k-means clustering algorithm, many extensions of k-means clustering algorithm have been proposed. The conventional k-means clustering algorithm is sensitive to the initial clustering centers and the number of clusters needs to be tuned manually. To solve the problem of determining the best initial seeds and most appropriate clustering number, an improved k-means algorithm based on density canopy was put forward [13]. In order to overcome the drawback that the clustering algorithms only consider within-cluster information, by using between-cluster information to add a new term in the distance measurement, an enhanced soft subspace clustering algorithm was proposed [14]. A k-means++ algorithm selecting the initial clustering centers by maximizing the distances among them was presented [15]. Since the conventional k-means clustering algorithm needs much computational cost, a novel variant of k-means which aims to make a better trade-off between efficiency and clustering quality was proposed [16]. An improved k-means algorithm with an acceleration mechanism for producing the new cluster centers was presented [17]. The conventional k-means clustering algorithm and lots of its variations are easy to get stuck at local optima, a global k-means clustering algorithm in which one cluster center is added dynamically at a time and each data object is used as a candidate for the kth cluster center was put forward [18], [19]. In addition, to meet the needs of the demands of clustering datasets in different applications, several variations of k-means clustering algorithm were put forward [20]–[24]. In the basic k-means algorithm, during the clustering process, all the data objects and all the features have equivalent effect [25]. In the clustering process, in order to differentiate the discriminative ability of the different features, lots of weighting feature methods were put forward [5], [26]–[30]. A kernel-based multi-objective clustering algorithm with automatic attribute weighting was put forward [31]. A k-means type clustering algorithm using a new fashion for features weighting with an l^2-norm regularization was proposed [32].

III. AN ENHANCED REGULARIZED K-MEANS TYPE CLUSTERING ALGORITHM WITH ADAPTIVE WEIGHTS

In the traditional k-means clustering algorithm, the ability of the different features and the importance of the different data objects are treated equally, the importance of the different data objects and the discriminative ability of the different features cannot be differentiated. In [32], considering the weights of features, a weighting k-means clustering algorithm is presented; however, the importance of the different data objects cannot be distinguished. In order to differentiate the importance of the different data objects and the discriminative ability of the different features effectively at the same time, we put forward an enhanced regularized k-means type clustering algorithm with adaptive weights(*dfkmeans*-l^2) in which an adaptive data weights vector W and an adaptive feature weights matrix R are introduced considering the different data weights and feature weights simultaneously.

Given a dataset $X = \{x_k\}_{k=1}^{n}$, n is the number of the data objects in the dataset, $x_k = [x_{k1}, x_{k2}, \cdots, x_{km}]$, m is the number of features; U is the membership matrix, $u_{ik} = 1$ indicates that the kth data object is assigned to the ith cluster, otherwise, if the kth data object is not assigned to the ith

cluster, $u_{ik} = 0$; W is an adaptive data weights vector $W = [w_1, \cdots, w_k \cdots, w_n]$, w_k show the weight of data object x_k; R is an adaptive feature weights matrix

$$R = \begin{bmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & \vdots & \vdots \\ r_{c1} & \cdots & r_{cm} \end{bmatrix}$$

in which the *ith* row shows the weight vector of all m features in cluster i, $1 \leq i \leq c$, c is the number of clusters;

$$V = \begin{bmatrix} v_{11} & \cdots & v_{1m} \\ \vdots & \vdots & \vdots \\ v_{c1} & \cdots & v_{cm} \end{bmatrix}$$

is the matrix of the cluster centers in which the *ith* row shows the center of cluster i. The objective function of the proposed k-means clustering algorithm can be formulated as:

$$P_{dfkmeans-l^2}(X, U, V, W, R) = \sum_{k=1}^{n} w_k^2 \sum_{i=1}^{c} u_{ik} \sum_{j=1}^{m} r_{ij}^2 (x_{kj} - v_{ij})^2 + \frac{1}{2}\alpha \sum_{k=1}^{n} w_k^2 + \frac{1}{2}\beta \sum_{i=1}^{c}\sum_{j=1}^{m} r_{ij}^2 \tag{1}$$

Subjected to the constraint:

$$\sum_{j=1}^{m} r_{ij} = 1 \quad \sum_{i=1}^{c} u_{ik} = 1 \quad \prod_{k=1}^{n} w_k = 1$$
$$r_{ij} \in [0,1] \quad u_{ik} \in [0,1], \ 1 \leq i \leq c, \ 1 \leq k \leq n \tag{2}$$

In Eq.(1), we add an l^2-norm regularization to the weights of data objects and features to encourage more data objects and features to participate in the identification of clusters. α is a hyper-parameter to harmonize the weighting scatter of data objects and the certainty of each data object and control the regularization of the weights of each data object in the identification of a cluster; the larger α is, the difference between the values of the weights of each data object will be smaller, the smaller α is, the difference between the values of the weights of each data object will be larger; β is a hyper-parameter to harmonize the weighting scatter of features and the certainty of each feature and control the regularization of the weights of each feature in the identification of a cluster; the larger β is, the difference between the values of the weights of each feature in the identification of a cluster will be smaller, the smaller β is, the difference between the values of the weights of each feature in the identification of a cluster will be larger.

IV. THE PROPOSED K-MEANS CLUSTERING ALGORITHM

In this subsection, we minimize the objective function of Eq.(1) to get the iterative rules of the proposed k-means clustering algorithm. According to the objective function, we can get the corresponding Lagrangian function as follows:

$$\vartheta(U, V, W, R) = \sum_{k=1}^{n} w_k^2 \sum_{i=1}^{c} u_{ik} \sum_{j=1}^{m} r_{ij}^2 (x_{kj} - v_{ij})^2 + \frac{1}{2}\alpha \sum_{k=1}^{n} w_k^2 + \frac{1}{2}\beta \sum_{i=1}^{c}\sum_{j=1}^{m} r_{ij}^2 + \phi_1(\prod_{k=1}^{n} w_k - 1) + \phi_2(\sum_{j=1}^{m} r_{ij} - 1) \tag{3}$$

To get the iterative rules, we adopt the common approach by optimizing one variable of U, V, W and R and fixing the other variables in the objective function, so we can get the corresponding iterative rules by minimizing the Lagrangian function of Eq.(3).

A. THE ITERATIVE RULES OF U

By fixing $V = \overset{*}{V}$, $W = \overset{*}{W}$ and $R = \overset{*}{R}$, the updating iterative rules of the membership u_{ik} is:

$$u_{ik} = \begin{cases} 1, & if \sum_{j=1}^{m} r_{ij}^2 (x_{kj} - v_{ij})^2 < \sum_{j=1}^{m} r_{pj}^2 (x_{kj} - v_{pj})^2, \\ & \quad p \neq i, \ 1 \leq p \leq c \\ 0, & otherwise \end{cases} \tag{4}$$

B. THE ITERATIVE RULES OF W

We can obtain the updating iterative rules of the data weight w_k by setting the derivatives of the Lagrangian function to zero with respect to w_k:

$$\frac{\partial \vartheta_{DwfwFcm}(\overset{*}{U}, \overset{*}{V}, W, \overset{*}{R})}{\partial w_k} = 2w_k \sum_{i=1}^{c} u_{ik} \sum_{j=1}^{m} r_{ij}^2 (x_{kj} - v_{ij})^2 + \alpha w_k + \phi_1 \prod_{p=1, p\neq k}^{n} w_p = 0$$

$$\phi_1 = \frac{-[2w_k \sum_{i=1}^{c} u_{ik} \sum_{j=1}^{m} r_{ij}^2 (x_{kj} - v_{ij})^2 + \alpha w_k]}{\prod_{p=1, p\neq k}^{n} w_p}$$

$$w_k = \frac{-\phi_1 \prod_{p=1, p\neq k}^{n} w_p}{[2 \sum_{i=1}^{c} u_{ik}^2 \sum_{j=1}^{m} r_{ij}^2 (x_{kj} - v_{ij})^2 + \alpha]} = \left[\frac{-\phi_1}{[2 \sum_{i=1}^{c} u_{ik} \sum_{j=1}^{m} r_{ij}^2 (x_{kj} - v_{ij})^2 + \alpha]} \right]^{1/2}$$

Because $\prod_{p=1}^{n} w_p = 1$

We can get:

$$\phi_1 = \left[\prod_{p=1}^{n} [-2\sum_{i=1}^{c} u_{ip} \sum_{j=1}^{m} r_{ij}^2 (x_{pj} - v_{ij})^2 - \alpha] \right]^{1/n}$$

$$w_k = \left[\frac{\left[\prod_{p=1}^{n} [2\sum_{i=1}^{c} u_{ip} \sum_{j=1}^{m} r_{ij}^2 (x_{pj} - v_{ij})^2 + a] \right]^{1/n}}{2\sum_{i=1}^{c} u_{ik} \sum_{j=1}^{m} r_{ij}^2 (x_{kj} - v_{ij})^2 + a} \right]^{1/2} \quad (5)$$

C. THE ITERATIVE RULES OF R

We can obtain the updating iterative rules of the feature weight r_{ij} by setting the derivatives of the Lagrangian function to zero with respect to r_{ij}:

$$\frac{\partial \vartheta(\overset{*}{U}, \overset{*}{V}, \overset{*}{W}, R)}{\partial r_{ij}} = 2\sum_{k=1}^{n} w_k^2 u_{ik} r_{ij} (x_{kj} - v_{ij})^2 + \phi_2 + \beta r_{ij} = 0$$

$$r_{ij} = -\frac{\phi_2}{2\sum_{k=1}^{n} w_k^2 u_{ik} (x_{kj} - v_{ij})^2 + \beta}$$

Because $\sum_{p=1}^{m} r_{ip} = 1$ [see continuation in (6), as shown at the bottom of this page.]

D. THE ITERATIVE RULES OF V

We can obtain the updating iterative rules of the cluster centers v_{ij} by setting the derivatives of the Lagrangian function to zero with respect to v_{ij}:

$$v_{ij} = \frac{\sum_{k=1}^{n} w_k^2 u_{ik} x_{kj}}{\sum_{k=1}^{n} w_k^2 u_{ik}} \quad (7)$$

The overall procedures of our proposed k-means algorithm can be described as follows.

TABLE 1. The overall procedures of the proposed k-means algorithm.

The procedures of the proposed algorithm
1: **Input:** dataset $X = \{x_k\}_{k=1}^{n}$, c , α , β
2: **Output:** W, U, R, V
3:**Initialize:** Randomly initialize the cluster centers $V^0 = \begin{bmatrix} v_{11} & \cdots & v_{1m} \\ \vdots & \vdots & \vdots \\ v_{c1} & \cdots & v_{cm} \end{bmatrix}$,the adaptive data weights vector $W^0 = [w_1, \cdots, w_k \cdots, w_n]$ and the adaptive feature weights matrix $R^0 = \begin{bmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & \vdots & \vdots \\ r_{c1} & \cdots & r_{cm} \end{bmatrix}$
4: **repeat**
5: **update:** the membership matrix U with Eq.(4), the data weights W with Eq.(5), the feature weights R with Eq.(6), the centroids V with Eq.(7) respectively.
6: **until convergence**

The stop criterion of all related algorithms is that the number of iterations exceeds the maximum number of iterations or for two consecutive iterations, the change in the objective function is less than the set threshold δ, in this paper, the maximum number of iterations and δ are set to 100 and 0.0001 respectively.

E. CONVERGENCE AND COMPLEXITY ANALYSIS

The objective function values against the iterations with different initial centroids is shown in figure 1, from figure 1, we can conclude that with the increment of the iterations, the value of the objective function is decreasing, the algorithm is able to converge after a few of iterations. There are only two computational updating steps in the traditional k-means algorithm: updating the centroids V and membership matrix U. There are four computational updating steps in our proposed algorithm: updating the membership matrix U, the data weights W, the feature weights F and the centroids V. The computational updating cost is $O(tcmn)$, n is the number of data objects, c is the number of clusters, t is the number of iterations, m is the number of features.

$$\phi_2 = \frac{\prod_{p=1}^{m} \left[-\sum_{k=1}^{n} 2w_k^2 u_{ik} (x_{kp} - v_{ip})^2 + \beta \right]}{\sum_{q=1}^{m} \left(\left[\prod_{p=1}^{m} \left[\sum_{k=1}^{n} 2w_k^2 u_{ik} (x_{kp} - v_{ip})^2 + \beta \right] \right] \Big/ [2w_k^2 u_{ik} (x_{kq} - v_{iq})^2 + \beta] \right)}$$

$$r_{ij} = \frac{\prod_{p=1}^{m} \left[\sum_{k=1}^{n} w_k^2 u_{ik} (x_{kp} - v_{ip})^2 + \beta \right] \Big/ (w_k^2 u_{ik} (x_{kj} - v_{ij})^2 + \beta)}{\sum_{q=1}^{m} \left(\left[\prod_{p=1}^{m} \left[\sum_{k=1}^{n} w_k^2 u_{ik} (x_{kp} - v_{ip})^2 + \beta \right] \right] \Big/ [w_k^2 u_{ik} (x_{kq} - v_{iq})^2 + \beta] \right)} \quad (6)$$

TABLE 2. The weights of the different features in the different cluster on the *Breasttissue* dataset by *dfkmeans-l^2* when $\beta = 0.2$.

Feature / Cluster	1	2	3	4	5	6	7	8	9
1	0.1198	0.0829	0.0614	0.1245	0.1301	0.1146	0.1254	0.1231	0.1182
2	0.1225	0.0781	0.0289	0.1299	0.1338	0.1218	0.1299	0.1272	0.1249
3	0.0436	0.1000	0.0673	0.1197	0.1946	0.1575	0.1588	0.1133	0.0452
4	0.1213	0.2389	0.0469	0.0742	0.2036	0.0965	0.0478	0.0535	0.1172
5	0.1470	0.0290	0.0150	0.1183	0.1759	0.1124	0.1394	0.1048	0.1581
6	0.1644	0.1774	0.1762	0.0787	0.0578	0.0617	0.079	0.0764	0.1283

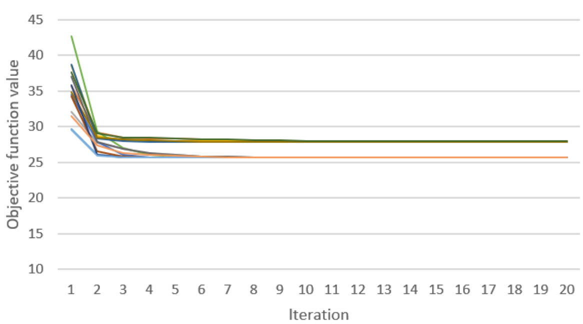


FIGURE 1. The objective function values against the iterations with different initial centroids.

V. EXPERIMENTS AND RESULTS

In this section, to evaluate the performance of the proposed algorithm, we conduct experiments using *UCI* datasets as the test object and do the related experimental analysis under the environment of Matlab. *Haberman*, *Breasttissue*, *IRIS*, *Banknote*, *Sonar* and *Wdbc* datasets of *UCI* datasets are used. In *Haberman* dataset, the number of categories is 2, the number of attributes is 3 and the sample size is 306; In *Breasttissue* dataset, the number of categories is 6, the number of attributes is 9 and the sample size is 106; In *IRIS* dataset, the number of categories is 3, the number of attributes is 4 and the sample size is 150; In *Banknote* dataset, the number of categories is 2, the number of attributes is 5 and the sample size is 1372; In *Sonar* dataset, the number of categories is 2, the number of attributes is 60 and the sample size is 208; In *Wdbc* dataset, the number of categories is 2, the number of attributes is 30 and the sample size is 569.

In this paper, all the datasets are dealt with normalization, we compare the clustering results produced by the traditional k-means clustering algorithm, the k-means clustering algorithm considering data weights(*dkmeans-l^2*), the k-means clustering algorithm considering feature weights (*fkmeans-l^2*), the entropy weighting k-means(*EWkmeans*) [32] and the proposed k-means algorithm about three performance metrics including Rand index(*RI*), the successful classification rate (*SCR*) and normalized mutual information(*NMI*) [33]. *RI* represents the percentage of the pairs of the data objects which are classified correctly in the clustering results. *SCR* stands for the accuracy of clustering which represents the percentage of the data objects which are classified correctly in the clustering results. *NMI* is a normalization of the mutual information (*MI*) score used to measure the clustering quality between 0 (no mutual information) and 1 (perfect clustering). The objective functions of *dkmeans-l^2* and *fkmeans-l^2* are shown as follows:

$$P_{dkmeans-l^2}(X,U,V,W) = \sum_{k=1}^{n} w_k^2 \sum_{i=1}^{c} u_{ik} \sum_{j=1}^{m} (x_{kj} - v_{ij})^2 + \frac{1}{2}\alpha \sum_{k=1}^{n} w_k^2$$

$$P_{fkmeans-l^2}(X,U,V,R) = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik} \sum_{j=1}^{m} r_{ij}^2 (x_{kj} - v_{ij})^2 + \frac{1}{2}\beta \sum_{i=1}^{c} \sum_{j=1}^{m} r_{ij}^2$$

The weights of the different features in the different cluster on the *Breasttissue* dataset by the proposed k-means algorithm (*dfkmeans-l^2*) when $\beta = 0.2$ are shown in table 2, the weights of different data objects by the proposed k-means algorithm (*dfkmeans-l^2*) on the *Breasttissue* dataset when $\alpha = 0.03$ are shown in figure 2.Through table 2 and figure 2 we can see that *dfkmeans-l^2* can distinguish the importance of the different data objects and the discriminative ability of the different features. The weights of different data

FIGURE 2. **The weights of different data objects on *Breasttissue* dataset by *dfkmeans-l^2* when $\alpha = 0.03$.**

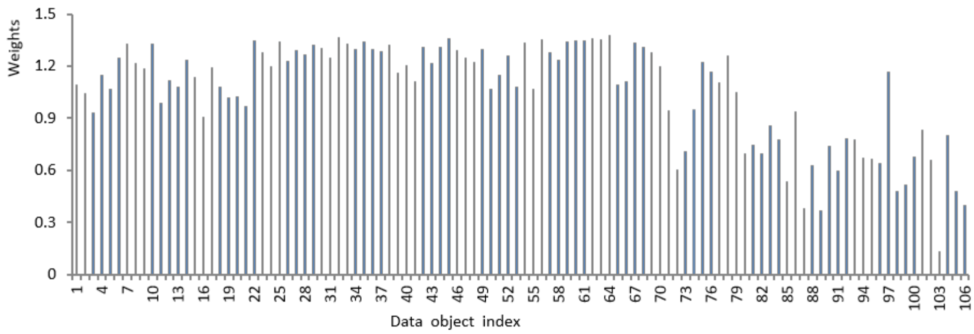


FIGURE 3. **The weights of different data objects on *Breasttissue* dataset by *dfkmeans-l^2* when $\alpha = 0.0001$.**

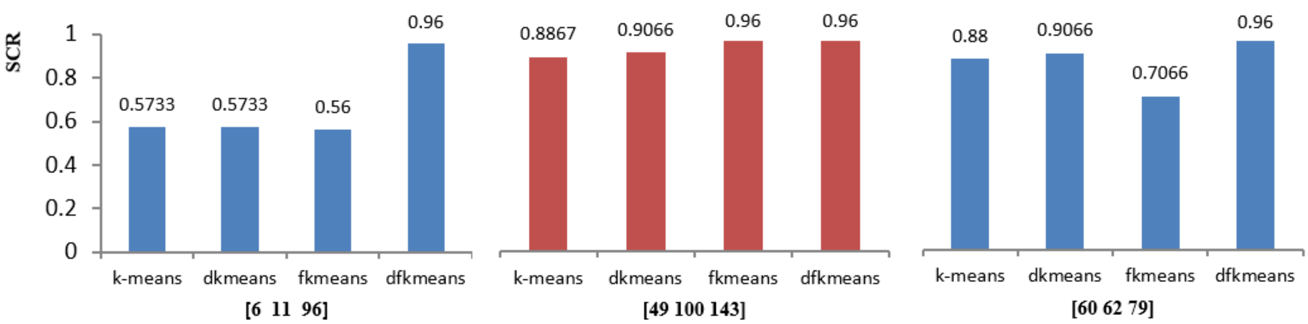
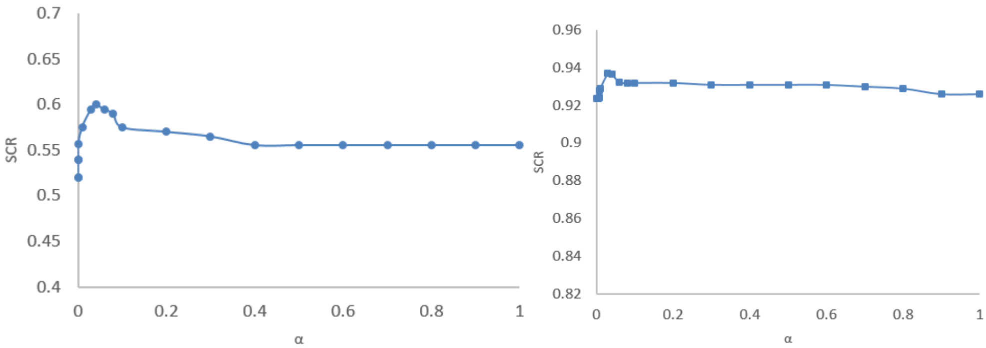


FIGURE 4. **The comparison of SCR with the different initial clustering centers.**

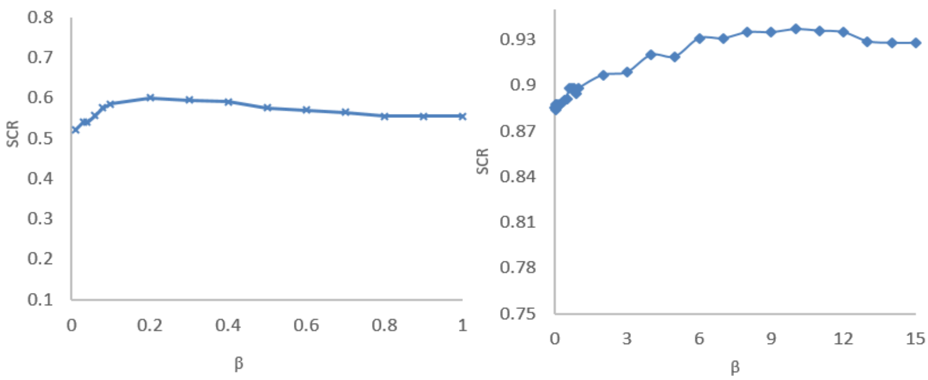
objects by *dfkmeans-l^2* on the *Breasttissue* dataset when $\alpha = 0.0001$ are shown in figure 3, in compared with figure 2, we can see that the weights of different data objects by *dfkmeans-l^2* when $\alpha = 0.0001$ are much sparser than those by *dfkmeans-l^2* when $\alpha = 0.03$. When $\alpha = 0.0001$, the values of the weights of partial data objects will be much larger than the other weights, which could result in the phenomenon that the clustering result are dominated by the partial data objects with larger weights. So in the presented algorithm, we can regulate the value of the hyper-parameter α to make sure that we can not only obtain weights to represent the different importance of each data object, but also stimulate more data objects to contribute to the process of clustering.

We have conducted experiments of the value of hyper-parameter α and β for its effects on the clustering results. We have compared the clustering performance (*SCR, RI,NMI*) with the increments of α and β with 0.002, 0.02, 0.2,1 when the values of α and β are in the range of (0,0.01], (0.01,0.1], (0.1,1] and (1,15] respectively. Generally, the best clustering performance can be derived when the range of α and β are (0.01,0.5] and (0.1,10] respectively depending on different datasets. The larger α is, the difference between the values of the weights of each data object will be smaller, the values of the weights of each data object will tend to be similar; the smaller α is, the difference between the values of the weights of each data object will be larger. The larger β is, the difference between the values of the weights of each feature in the identification of a cluster will be smaller, the values of the weights of each feature will tend to be similar; the smaller β is, the difference between the values



(a) (b)

FIGURE 5. **The comparison of the averaged SCR about setting different values of α: (a) *Sonar* dataset with $\beta = 0.2$; (b) *Wdbc* dataset with $\beta = 10$.**



(a) (b)

FIGURE 6. **The comparison of the averaged SCR about setting different values of β with $\alpha = 0.03$: (a) *Sonar* dataset; (b) *Wdbc* dataset.**

of the weights of each feature in the identification of a cluster will be larger. Take the *Sonar and Wdbc* datasets as example, after 100 runs, the averaged *SCR* derived by *dfkmeans*-l^2 about setting different values of α is shown in figure 5; the averaged *SCR* derived by *dfkmeans*-l^2 about setting different values of β is shown in figure 6. In this paper, we study the clustering results with α set to 0.03, β set to 0.2 in *Haberman*, *Breasttissue, IRIS*, *Banknote* and *Sonar* datasets and β set to 10 in *Wdbc* dataset.

Take the *IRIS* dataset as example, if we fix the initial clustering centers, the comparison of *SCR* when the four algorithms are convergent with the different initial clustering centers is shown in figure 4.From figure 4, we can see that when the number of the original data in *IRIS* dataset selected as the initial clustering centers is [6 11 96], the SCR of the proposed *dfkmeans*-l^2 is 0.96,which is much higher than the other three algorithms; when the number of the original data selected as the initial clustering centers is [49 100 143], the *SCR* of *dfkmeans*-l^2 is 0.96 which is equal to *fkmeans*-l^2 and higher than 0.9066 of *dkmeans*-l^2 and 0.8867 of k-means; when the number of the original data selected as the initial clustering centers is [60 62 79], the *SCR* of *dfkmeans*-l^2 is 0.96 which is higher than 0.7066 of *fkmeans*-l^2, 0.9066 of *dkmeans*-l^2 and 0.88 of k-means algorithm, so we can conclude that the proposed algorithm is more robust and has better stability than the traditional k-means, *dkmeans*-l^2 and *fkmeans*-l^2 algorithm.

The clustering results are averaged over twenty independent experiments to mitigate the effect of initialization. The comparison about three performance metrics and running time on the different datasets among the different algorithms is shown in table 3.

From table 3, we can see that among different algorithms, the running time is consistent with the complexity analysis in section 4.5, the factor resulting in the differences of the running time is the different number of iterations; the averaged three performance metrics: rand index (*RI*), normalized mutual information (*NMI*) and the successful classification rate (*SCR*) of the proposed algorithm (*dfkmeans*-l^2) are all much higher than the other four algorithms: the traditional k-means, *dkmeans*-l^2, *fkmeans*-l^2 and *EWkmeans* algorithm; although the standard deviation by the *dfkmeans*-l^2 are not always smaller than the other algorithms, we can see that the value by the *dfkmeans*-l^2 with the averaged *SCR* subtracting

TABLE 3. **The comparison of the *SCR, RI, NMI* and running time index on different datasets among different algorithms (The standard deviation in bracket).**

		k-means	*dkmeans-l^2*	*fkmeans-l^2*	*EWkmeans*	*dfkmeans-l^2*
SCR	*Haberman*	0.5163(± 0.0089)	0.5184(± 0.0062)	0.6900(± 0.084)	0.5816(± 0.1101)	**0.7461(± 0.0032)**
	Breasttissue	0.4921(± 0.0328)	0.5161(± 0.0306)	0.5416(± 0.0572)	0.5136(± 0.0367)	**0.5880(± 0.0332)**
	IRIS	0.8322(± 0.1239)	0.8694(± 0.8728)	0.8861(± 0.1454)	0.9011 (± 0.0162)	**0.9421(± 0.0249)**
	Banknote	0.5759(± 0.0037)	0.5764(± 0.0024)	0.6491(± 0.1031)	0.6377(± 0.1268)	**0.6806(± 0.1151)**
	Sonar	0.5504(± 0.0105)	0.5552(± 0.0025)	0.5628(± 0.0131)	0.5423(± 0.0211)	**0.5929(± 0.0187)**
	Wdbc	0.9279(± 0.0000)	0.9297(± 0.0000)	0.9296(± 0.0107)	0.9277(± 0.0130)	**0.9369(± 0.0055)**
RI	*Haberman*	0.4990(± 0.0005)	0.4991(± 0.0003)	0.5838(± 0.053)	0.5330(± 0.051)	**0.6199(± 0.0031)**
	Breasttissue	0.7818(± 0.0270)	0.7979(± 0.0147)	0.8043(± 0.0224)	0.7932(± 0.021)	**0.8179(± 0.0138)**
	IRIS	0.8521(± 0.0621)	0.8728(± 0.0499)	0.9011(± 0.0842)	0.8874(± 0.0160)	**0.9365(± 0.0268)**
	Banknote	0.5112(± 0.0011)	0.5113(± 0.0007)	0.5636(± 0.0776)	0.5671(± 0.0913)	**0.5975(± 0.08419)**
	Sonar	0.5029(± 0.0020)	0.5037(± 0.0005)	0.5058(± 0.0034)	0.5040(± 0.004)	**0.5156(± 0.0073)**
	Wdbc	0.8660(± 0.0000)	0.8691(± 0.0000)	0.8692(± 0.0184)	0.8660(± 0.0220)	**0.8817(± 0.0096)**
NMI	*Haberman*	0.0010(± 0.0006)	0.0009(± 0.0003)	0.0459(± 0.0334)	0.0141(± 0.026)	**0.0652(± 0.0039)**
	Breasttissue	0.5089(± 0.0134)	0.5124(± 0.0270)	0.5227(± 0.0291)	0.5251(± 0.021)	**0.5346(± 0.0299)**
	IRIS	0.7182(± 0.0601)	0.7535(± 0.0521)	0.8143(± 0.0936)	0.7593(± 0.0219)	**0.8522(± 0.0286)**
	Banknote	0.0167(± 0.0011)	0.0186(± 0.0012)	0.1013(± 0.1210)	0.1044(± 0.1441)	**0.1408(± 0.1361)**
	Sonar	0.0075(± 0.0040)	0.0108(± 0.0009)	0.0358(± 0.0168)	0.0218(± 0.020)	**0.0729(± 0.0227)**
	Wdbc	0.6232(± 0.0000)	0.6215(± 0.0000)	0.6478(± 0.1485)	0.6316(± 0.0448)	**0.6698(± 0.0246)**
Running time(second)	*Haberman*	0.1233(± 0.0093)	1.5451(± 0.2174)	0.6450(± 0.0448)	0.7003(± 0.1569)	1.2560(± 0.1933)
	Breasttissue	0.2524(± 0.0448)	1.2887(± 0.3138)	0.9598(± 0.0155)	0.8601(± 0.1837)	1.304(± 0.0293)
	IRIS	0.3413(± 0.0970)	0.7535(± 0.0521)	0.4483(± 0.1123)	0.4888(± 0.0963)	0.6375(± 0.1343)
	Banknote	11.105(± 1.7791)	20.223(± 1.4470)	11.933(± 1.0250)	11.383(± 1.8996)	10.770(± 1.5350)
	Sonar	1.8618(± 0.2627)	2.8125(± 0.4685)	3.6365(± 0.4408)	1.9945(± 0.0930)	3.8960(± 0.6757)
	Wdbc	2.949(± 0.5506)	3.8739(± 1.1872)	3.7295(± 1.2509)	4.0103(± 1.395)	4.0676(± 0.6689)

the standard deviation is still larger than the averaged *SCR* by the other three algorithms, which indicates that the algorithm presented in this paper has a better clustering quality.

VI. CONCLUSION

This paper put forward an enhanced regularized k-means type clustering algorithm with adaptive weights in which we introduced an adaptive feature weights matrix and an adaptive data weights vector into the objective function of the traditional k-means clustering algorithm and we developed a new objective function with l^2-norm regularization to the weights of data objects and features. Experimental results show that the novel algorithm proposed in this paper has better and promising performance with respects to three evaluation metrics: Rand index (*RI*), the successful classification rate (*SCR*), normalized mutual information (*NMI*).

ACKNOWLEDGMENT

(Ziheng Wu and Zixiang Wu contributed equally to this work.)

REFERENCES

[1] F. Breve, L. Zhao, M. Quiles, W. Pedrycz, and J. Liu, "Particle competition and cooperation in networks for semi-supervised learning," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 9, pp. 1686–1698, Sep. 2012.

[2] J. Z. Huang, M. K. Ng, H. Rong, and Z. Li, "Automated variable weighting in k-means type clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 657–668, May 2005.

[3] J.-G. Sun, "Clustering algorithms research," *J. Softw.*, vol. 19, no. 1, pp. 48–61, Jun. 2008.

[4] J. Liu, C. Aggarwal, and J. Han, "On integrating network and community discovery," in *Proc. 8th ACM Int. Conf. Web Search Data Mining (WSDM)*, 2015, pp. 117–126.

[5] L. Jing, M. K. Ng, and J. Z. Huang, "An entropy weighting k-means algorithm for subspace clustering of high-dimensional sparse data," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 8, pp. 1026–1041, Aug. 2007.

[6] G. Liu, Y. Zhang, and A. Wang, "Incorporating adaptive local information into fuzzy clustering for image segmentation," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3990–4000, Nov. 2015.

[7] G. Moreno-Hagelsieb, Z. Wang, S. Walsh, and A. Elsherbiny, "Phylogenomic clustering for selecting non-redundant genomes for comparative genomics," *Bioinformatics*, vol. 29, no. 7, pp. 947–949, Apr. 2013.

[8] J. Wang, S.-T. Wang, and Z.-H. Deng, "A novel text clustering algorithm based on feature weighting distance and soft subspace learning," *Chin. J. Comput.*, vol. 35, no. 8, p. 1655, 2012.

[9] J. Xie, S. Jiang, W. Xie, and X. Gao, "An efficient global K-means clustering algorithm," *J. Comput.*, vol. 6, no. 2, pp. 271–279,2011.

[10] J. Zhang, Y. Yang, J. Yang, and Z. Zhang, "Algorithm for initialization of K-means clustering center based on optimized-division," *J. Syst. Simul.*, vol. 21, no. 9, pp. 2586–2590, 2009.

[11] S.-B. Zhou, Z.-Y. Xu, and X.-Q. Tang, "Method for determining optimal number of clusters in K-means clustering algorithm," *J. Comput. Appl.*, vol. 30, no. 8, pp. 1995–1998, Sep. 2010.

[12] X.-F. Lei, "An efficient clustering algorithm based on local optimality of K-means: An efficient clustering algorithm based on local optimality of k-means," *J. Softw.*, vol. 19, no. 7, pp. 1683–1692, Oct. 2008.

[13] G. Zhang, C. Zhang, and H. Zhang, "Improved K-means algorithm based on density Canopy," *Knowl.-Based Syst.*, vol. 145, pp. 289–297, Apr. 2018.

[14] Z. Deng, K.-S. Choi, F.-L. Chung, and S. Wang, "Enhanced soft subspace clustering integrating within-cluster and between-cluster information," *Pattern Recognit.*, vol. 43, no. 3, pp. 767–781, Mar. 2010.

[15] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding," in *Proc. 18th Annu. ACM-SIAM Symp. Discrete Algorithms, Soc. Ind. Appl. Math.*, 2007, pp. 1027–1035.

[16] W.-L. Zhao, C.-H. Deng, and C.-W. Ngo, "K-means: A revisit," *Neurocomputing*, vol. 291, pp. 195–206, May 2018.

[17] L. Bai, J. Liang, C. Sui, and C. Dang, "Fast global K-means clustering based on local geometrical information," *Inf. Sci.*, vol. 245, pp. 168–180, Oct. 2013.

[18] A. M. Bagirov, "Modified global -means algorithm for minimum sum-of-squares clustering problems," *Pattern Recognit.*, vol. 41, no. 10, pp. 3192–3199, Oct. 2008.

[19] A. Likas, N. Vlassis, and J. J. Verbeek, "The global K-means clustering algorithm," *Pattern Recognit.*, vol. 36, no. 2, pp. 451–461, Feb. 2003.

[20] Y. Ding, Y. Zhao, X. Shen, M. Musuvathi, and T. Mytkowicz, "Yinyang K-means: A drop-in replacement of the classic K-means with consistent speedup," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, pp. 579–587.

[21] I. Dhillon and D. Modha, "Concept decompositions for large sparse text data using clustering," *Mach. learn.*, vol. 42, no. 1, pp. 143–175, 2001.

[22] P. Bradley, U. Fayyad, and C. Reina, "Scaling clustering algorithms to large databases," in *Proc. 4th Int. Conf. Knowl. Disc. Data Mining*, 1998, pp. 9–15.

[23] J. Drake and G. Hamerly, "Accelerated K-means with adaptive distance bounds," in *Proc. 5th NIPS Workshop Optim. Mach. Learn.*, 2012, pp. 2–5.

[24] W. Sun, J. Wang, and Y. Fang, "Regularized K-means clustering of high-dimensional data and its asymptotic consistency," *Electron. J. Statist.*, vol. 6, no. 6, pp. 148–167, 2012.

[25] S. Z. Selim and M. A. Ismail, "K-means-type algorithms: A generalized convergence theorem and characterization of local optimality," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 1, pp. 81–87, Jan. 1984.

[26] X. Huang, Y. Ye, and H. Zhang, "Extensions of K-means-type algorithms: A new clustering framework by integrating intracluster compactness and intercluster separation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 8, pp. 1433–1446, Aug. 2014.

[27] X. Chen, Y. Ye, X. Xu, and J. Z. Huang, "A feature group weighting method for subspace clustering of high-dimensional data," *Pattern Recognit.*, vol. 45, no. 1, pp. 434–446, Jan. 2012.

[28] X. Li, G. Xu, S. Ren, and M. Yao, "Reliability-based regularized weighted soft k-means algorithm for subspace clustering," *J. NanJing Univ., Natural Sci.*, vol. 53, no. 3, pp. 525–536, 2017.

[29] X. Huang, Y. Ye, H. Guo, Y. Cai, H. Zhang, and Y. Li, "DSKmeans: A new K-means-type approach to discriminative subspace clustering," *Knowl.-Based Syst.*, vol. 70, pp. 293–300, Nov. 2014.

[30] A. Ahmad and L. Dey, "A K-means type clustering algorithm for subspace clustering of mixed numeric and categorical datasets," *Pattern Recognit. Lett.*, vol. 32, no. 7, pp. 1062–1069, May 2011.

[31] Z. Zhou and S. Zhu, "Kernel-based multiobjective clustering algorithm with automatic attribute weighting," *Soft Comput.*, vol. 22, no. 11, pp. 3685–3709, Jun. 2018.

[32] X. Huang, X. Yang, J. Zhao, L. Xiong, and Y. Ye, "A new weighting K-means type clustering framework with an l2 -norm regularization," *Knowl.-Based Syst.*, vol. 151, pp. 165–179, Jul. 2018.

[33] T. M. S. Filho, B. A. Pimentel, R. M. C. R. Souza, and A. L. I. Oliveira, "Hybrid methods for fuzzy clustering based on fuzzy c-means and improved particle swarm optimization," *Expert Syst. Appl.*, vol. 42, nos. 17–18, pp. 6315–6328, Oct. 2015.

ZIHENG WU was born in 1988. He received the Ph.D. degree from the University of Science and Technology of China, in 2018. He is currently a Lecturer with the Anhui University of Technology. His research interests include machine learning, artificial intelligence, grey systems theory, medical informatics, science engineering, and intelligent control.

ZIXIANG WU received the Ph.D. degree from Zhejiang University, in 2019. He is currently with the Second Affiliated Hospital, Zhejiang University School of Medicine. His current research interests include thoracic surgery, artificial intelligence, and medical informatics.