Information Sciences 267 (2014) 1–15

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Further improvements in Feature-Weighted Fuzzy C-Means

Hong-Jie Xing [a,*], Ming-Hu Ha [b]

[a] Key Laboratory of Machine Learning and Computational Intelligence, College of Mathematics and Computer Science, Hebei University, Baoding 071002, China
[b] School of Science, Hebei University of Engineering, Handan 056038, China

ARTICLE INFO

Article history:
Received 4 July 2010
Received in revised form 19 November 2013
Accepted 18 January 2014
Available online 27 January 2014

Keywords:
Fuzzy C-Means
Feature-weight vector
Kernelized distance
Color image segmentation

ABSTRACT

In cluster analysis, certain features of a given data set may exhibit higher relevance than others. To address this issue, Feature-Weighted Fuzzy C-Means (FWFCM) approaches have emerged in recent years. However, there are certain deficiencies in the existing FWFCMs, e.g., the elements in a feature-weight vector cannot be adaptively adjusted during the training phase, and the update formulas of a feature-weight vector cannot be derived analytically. In this study, an Improved FWFCM (IFWFCM) is proposed to overcome these shortcomings. The IFWFCM_KD based on the kernelized distance is also proposed. Experimental results reported for five numerical data sets and the color images show that IFWFCM is superior to the existing FWFCMs. An interesting conclusion, that IFWFCM_KD might not improve the performance of IFWFCM, is also obtained by applying IFWFCM_KD to tackle the above-mentioned numerical data sets and color images.

1. Introduction

Clustering or cluster analysis is an important technology in many research areas, such as data mining, pattern recognition, and machine learning. Existing clustering methods can be roughly classified into five categories [13,24]:

- **Hierarchical methods**: There are two different types of hierarchical clustering approaches, namely, agglomerative hierarchical clustering and divisive hierarchical clustering.
- **Partitioning methods**: There are many partitioning techniques of cluster analysis. Among them, K-means, Fuzzy C-Means and self-organizing maps are the most typical approaches.
- **Density-based methods**: The density-based clustering approach uses the local density of the samples to determine the clusters. Therefore, they can detect non-convex and arbitrarily shaped clusters.
- **Grid-based methods**: Grid-based clustering methods partition a given data space into cells of a given grid and merge them to construct clusters
- **Model-based methods**: The model-based clustering approach utilizes model selection techniques to determine a model structure and uses maximum likelihood algorithms to estimate the parameters.

As a partitioning clustering method, Fuzzy C-Means (FCM) has been widely studied in many fields. FCM was first proposed by Dunn in 1974 [8] and then developed by Bezdek [1]. FCM is regarded as the most widely used in practice fuzzy clustering algorithm. It has been successfully used in applications such as remote sensing change detection [11], time series

* Corresponding author. Tel.: +86 15933987805.
E-mail address: hjxing@hbu.edu.cn (H.-J. Xing).

http://dx.doi.org/10.1016/j.ins.2014.01.033
0020-0255/

clustering [19], and color image segmentation [25]. In recent years, many improved versions of FCM have been proposed [5,4,26]. To eliminate the shortcoming caused by the random selection of the initial centers, Kim et al. proposed a novel initialization scheme for FCM in color image segmentation [16]. When kernel methods emerged [22], many kernel FCM (KFCM) methods were also designed. Graves and Pedrycz reviewed these KFCMs and compared them with FCM [12]. Through comparing the results of FCM and KFCM on many synthetic data sets and a number of benchmark data sets from the UCI machine learning repository [3], they draw the conclusion that the KFCMs can only produce marginal improvements over FCM.

For the above-mentioned FCM and its improved versions, it is assumed that all the features of the samples in a given data set make equal contribution when constructing the optimal clusters. However, for certain real-world data sets, some of the features can exhibit higher relevance in the clustering information than others. Thus, the features with higher relevance are more important to form the optimal clustering result than those with lower relevance. It is therefore desirable to revisit an FCM method in which different features possess different weights. Recently, several Feature-Weighted Fuzzy C-Means (FWFCM) approaches [28,14,10] have been proposed to address the above-mentioned problem. These variants exhibit two separate stages. At the first stage, the feature-weight vector is determined. Then, at the second stage, the FWFCM is trained by the samples with their features weighted by the obtained feature-weight vector. Unfortunately, the elements of the feature-weight vector are fixed in this second stage, which might not fully reflect their relevance in the clustering process. Therefore, much effort has been devoted to adjusting the feature-weight vector during the course of training the FCMs [9,23,29]. However, the approaches proposed in [9,23] both assign different feature weights for different features of the clusters rather than for different features of the entire data set. Although the method in [29] simultaneously updates the feature-weight vector and assigns different feature weights for the entire data set, there are still two issues that must be addressed.

- The update equation of the feature-weight vector cannot be obtained from the provided objective function. The reason is that the partial derivative of the given objective function with respect to feature-weight vector does not contain the feature-weight vector. Therefore, one cannot get the analytic expression of the update equation for the feature-weight vector.
- The experimental results might not be credible. For example, the sum of the elements in the feature-weight vector for the *Iris* data set is larger than one, which contradicts the constraint that the sum equals one.

To overcome the disadvantages of the existing FWFCMs, we propose an improved version referred to as Improved FWFCM (IFWFCM). In this version, the algorithmic framework and convergence properties of IFWFCM are recapitulated. The main contributions of the present study are as follows:

- In contrast to FWFCMs in [28,14], IFWFCM dynamically updates feature-weight vectors in its training phase rather than utilizing a fixed feature-weight vector. Because the feature-weight vector of the traditional FWFCMs remains fixed during the clustering procedure, the significance of certain features to the changing cluster information cannot be appropriately manifested. However, the elements in the updated feature-weight vector of IFWFCM can more accurately reflect the relevance of each feature when constructing clusters.
- According to the weighted distance provided in [29], we cannot analytically obtain the update equation of the feature-weight vector. We therefore reformulate the weighted distance measure to the form of its counterpart in [28], which indicates that the update equation of the feature-weight vector can be derived from the revised objective function. It should be mentioned here that the update equation of the feature-weight vector in this study is completely different from its counterpart in [29].
- Utilizing the kernelized distance measure, we generalize the proposed method to its kernelized version, i.e., IFWFCM based on the kernelized distance (IFWFCM_KD). Through an experimental study, we find that incorporating the kernelized distance measure into IFWFCM might not improve its performance; i.e., IFWFCM_KD cannot achieve more promising performance in comparison with IFWFCM.

The study is organized as follows. The traditional FCM is briefly reviewed in Section 2. The Improved FWFCM (IFWFCM) is expatiated in Section 3. IFWFCM based on the kernelized distance is described in Section 4. The results of the experiments are reported in Section 5. Some conclusions and suggestions for future work are given in Section 6. The theoretical derivations of the formulas in Section 3 are summarized in Appendix A.

2. Fuzzy C-Means

For a given data set $\mathbf{D} = \{\mathbf{x}_j\}_{j=1}^N$, with $\mathbf{x}_j = (x_{j1}, x_{j2}, \ldots, x_{jd}) \in \mathcal{R}^d$, the FCM clustering method minimizes the following objective function [2]

$$J(\mathbf{U}, \mathbf{V}; \mathbf{D}) = \sum_{i=1}^{C}\sum_{j=1}^{N}\mu_{ij}^m d_{ij}^2 = \sum_{i=1}^{C}\sum_{j=1}^{N}\mu_{ij}^m \|\mathbf{x}_j - \mathbf{v}_i\|^2, \tag{1}$$

where $\mathbf{U} = \left(\mu_{ij}\right)_{C\times N}$ is a fuzzy partition matrix in which its element μ_{ij} denotes the membership of the jth sample $\mathbf{x}_j$ belonging to the ith cluster, $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_C)^T = \left(v_{iq}\right)_{C\times d}$ is the center matrix that is composed of C cluster centers, $m > 1$ is the fuzzification exponent, and $\|\cdot\|$ is the Euclidean norm. It should be noted that the membership μ_{ij} should satisfy the constraints $\mu_{ij} \in [0,1], \sum_{i=1}^{C}\mu_{ij} = 1$ and $0 < \sum_{j=1}^{N}\mu_{ij} < N, i = 1,2,\ldots,C$ and $j = 1,2,\ldots,N$.

The update equations of μ_{ij} and $\mathbf{v}_i$ are as follows [2]

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{d_{ij}^2}{d_{kj}^2}\right)^{\frac{1}{m-1}}} \tag{2}$$

and

$$\mathbf{v}_i = \frac{\sum_{j=1}^{N}\mu_{ij}^m\mathbf{x}_j}{\sum_{j=1}^{N}\mu_{ij}^m}. \tag{3}$$

3. Improved Feature-Weighted Fuzzy C-Means

In this section, the proposed model, i.e., IFWFCM, is formulated. Moreover, the algorithm implementation for IFWFCM is explained. The convergence property and computational complexity are also analyzed.

3.1. Update equations

Suppose that we have a data set $\mathbf{D} = \{\mathbf{x}_j\}_{j=1}^N$ with $\mathbf{x}_j \in \mathcal{R}^d$. The fuzzy partition matrix $\mathbf{U} = \left(\mu_{ij}\right)_{C\times N}$ and the C cluster centers $\mathbf{V} = (v_{iq})_{C\times d}$ are both defined in Section 2. Thus, the objective function of IFWFCM to be minimized is given as follows

$$J(\mathbf{U},\mathbf{V},\mathbf{w};\mathbf{D}) = \sum_{i=1}^{C}\sum_{j=1}^{N}\mu_{ij}^m\left[d_{ij}^{(\mathbf{w})}\right]^2, \tag{4}$$

where $d_{ij}^{(\mathbf{w})} = \|\mathrm{diag}(\mathbf{w})(\mathbf{x}_j - \mathbf{v}_i)\|$ with $\mathbf{w} = (w_1, w_2, \ldots, w_d)^T$ is a feature-weight vector, and $\mathrm{diag}(\mathbf{w}) = \begin{pmatrix} w_1 & 0 & \ldots & 0 \\ 0 & w_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & w_d \end{pmatrix}$. Moreover, the elements μ_{ij} in the membership matrix $\mathbf{U}$ satisfy

$$\sum_{i=1}^{C}\mu_{ij} = 1, \tag{5}$$

while the elements w_q in the feature-weight vector $\mathbf{w}$ satisfy

$$\sum_{q=1}^{d}w_q = 1. \tag{6}$$

Utilizing the Lagrange multiplier technique, the aforementioned optimization problem can be converted to an unconstrained optimization problem that minimizes the following objective function

$$\widetilde{J}(\mathbf{U},\mathbf{V},\mathbf{w};\mathbf{D}) = \sum_{i=1}^{C}\sum_{j=1}^{N}\mu_{ij}^m[d_{ij}^{(\mathbf{w})}]^2 - \lambda\left(\sum_{i=1}^{C}\mu_{ij} - 1\right) - \beta\left(\sum_{q=1}^{d}w_q - 1\right), \tag{7}$$

where λ and β are Lagrange multipliers.

Therefore, problem (7) can be solved by finding the saddle point of the above Lagrange function and by taking the derivatives of the Lagrangian $\widetilde{J}(\mathbf{U},\mathbf{V},\mathbf{w};\mathbf{D})$ with respect to the parameters, i.e., μ_{ij}, $\mathbf{v}_i$, and w_k. Thereafter, we obtain the following update equations:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\left[d_{ij}^{(\mathbf{w})}\right]^2}{\left[d_{kj}^{(\mathbf{w})}\right]^2}\right)^{\frac{1}{m-1}}}, \tag{8}$$

$$\mathbf{v}_i = \frac{\sum_{j=1}^{N}\mu_{ij}^m\mathbf{x}_j}{\sum_{j=1}^{N}\mu_{ij}^m}, \tag{9}$$

and

$$w_q = \frac{1}{\sum_{l=1}^{d}\left[\frac{\sum_{i=1}^{C}\sum_{j=1}^{N}\mu_{ij}^m(x_{jq}-v_{iq})^2}{\sum_{i=1}^{C}\sum_{j=1}^{N}\mu_{ij}^m(x_{jl}-v_{il})^2}\right]} \tag{10}$$

Appendix A provides the detailed theoretical derivations of (8)–(10). For the update Eq. (9), there is an issue to be addressed.

- To obtain Eq. (9), the diagonal matrix $\mathrm{diag}(\mathbf{w})$ must be nonsingular, that is, $|\mathrm{diag}(\mathbf{w})| \neq 0$. Through the experiments, we observe that the elements w_q $(q = 1, 2, \ldots, d)$ in $\mathrm{diag}(\mathbf{w})$ are always nonzero. Hence, the update Eq. (9) holds.

3.2. IFWFCM algorithm

The overall procedure of the IFWFCM algorithm is summarized in Algorithm 1. It should be noted that during the initialization step in Algorithm 1, the values of the elements in the feature-weight vector $\mathbf{w}$ are initialized with the same value, i.e., $w_q = \frac{1}{d}$, with the indices $q = 1, 2, \ldots, d$. However, through experiments, we observe that the proposed approach is insensitive to the initialization of the feature weights. Although the initial feature weights can be different, the proposed method can obtain the same clustering results for a given data set.

Algorithm 1. The IFWFCM clustering algorithm

Input: Data set $\mathbf{D} = \{\mathbf{x}_j\}_{j=1}^N$

Output: Terminal fuzzy partition matrix $\mathbf{U}^{(t)}$, terminal center matrix $\mathbf{V}^{(t)}$, and terminal feature-weight vector $\mathbf{w}^{(t)}$

Initialization: Feature-weight vector $\mathbf{w}$, number of clusters C, fuzzification exponent m, termination tolerance ϵ, fuzzy partition matrix $\mathbf{U} = (\mu_{ij})_{C\times N}$ $(0 < \mu_{ij} < 1)$

repeat

for $t = 1, 2, \ldots$

Step 1: Compute the cluster centers $\mathbf{v}_i^{(t)} = \frac{\sum_{j=1}^N [\mu_{ij}^{(t-1)}]^m \mathbf{x}_j}{\sum_{j=1}^N [\mu_{ij}^{(t-1)}]^m}$.

Step 2: Calculate the distances $[d_{ij}^{(\mathbf{w})}]^2$ as:

$$\left[d_{ij}^{(\mathbf{w})}\right]^2 = (\mathbf{x}_j - \mathbf{v}_i)^T [\mathrm{diag}(\mathbf{w})]^2 (\mathbf{x}_j - \mathbf{v}_i), 1 \leqslant i \leqslant C, 1 \leqslant j \leqslant N.$$

Step 3: Update the fuzzy partition matrix:

$$\mu_{ij}^{(t)} = \frac{1}{\sum_{k=1}^C \left[\frac{d_{ij}^{(\mathbf{w})}}{d_{kj}^{(\mathbf{w})}}\right]^{\frac{2}{m-1}}}.$$

Step 4: Update the elements in the feature-weight vector:

$$w_q^{(t)} = \frac{1}{\sum_{l=1}^d \left(\frac{\sum_{i=1}^C \sum_{j=1}^N \left[\mu_{ij}^{(t)}\right]^m (x_{jq} - v_{iq})^2}{\sum_{i=1}^C \sum_{j=1}^N \left[\mu_{ij}^{(t)}\right]^m (x_{jl} - v_{il})^2}\right)}, 1 \leqslant q \leqslant d.$$

end for

until $\max_{ij}\left|\mu_{ij}^{(t)} - \mu_{ij}^{(t-1)}\right| < \epsilon$

3.3. Convergence property

Inspired by the convergence theorems and their proofs proposed by Bezdek in [1], we state the convergence property of IFWFCM in the following. Let $\mathbf{U} = (\mu_{ij})_{C\times N}, \mathbf{V} = (v_{iq})_{C\times d}$, and $\mathbf{w} = (w_1, w_2, \ldots, w_d)^T$.

Theorem 1. *Let* $\mathbf{V}$ *and* $\mathbf{w}$ *be fixed. Then,* $\mathbf{U}$ *is a local minimum of* $\widetilde{J}(\mathbf{U}, \mathbf{V}, \mathbf{w}; \mathbf{D})$ *of (7) if and only if* μ_{ij} *is given by (8).*

Proof. The necessity of the condition can be directly verified by the derivation of (8). To prove the sufficiency, we examine the second partial derivative of (7) with respect to the elements of $\mathbf{U}$. It can therefore be deduced that

$$h_{ts,ij}(\mathbf{U}) = \frac{\partial}{\partial \mu_{ts}}\left[\frac{\partial \widetilde{J}(\mathbf{U})}{\partial \mu_{ij}}\right] = \begin{cases} m(m-1)\mu_{ij}^{(m-2)}\left[d_{ij}^{(\mathbf{w})}\right]^2 & \text{if } t=i, s=j \\ 0 & \text{otherwise} \end{cases}, \tag{11}$$

where we substitute $\widetilde{J}(\mathbf{U},\mathbf{V},\mathbf{w};\mathbf{D})$ by $\widetilde{J}(\mathbf{U})$ for simplicity. Because $m > 1$ and $d_{ij}^{(\mathbf{w})} > 0$, we know that $m(m-1)\mu_{ij}^{(m-2)}\left[d_{ij}^{(\mathbf{w})}\right]^2 > 0$ and $H(\mathbf{U}) = (h_{ts,ij}(\mathbf{U}))_{CN\times CN}$ is a positive-definite diagonal matrix. Hence, (8) presents a sufficient condition for minimizing $\widetilde{J}(\mathbf{U})$. □

Theorem 2. *Let* $\mathbf{U}$ *and* $\mathbf{w}$ *be fixed. Then,* $\mathbf{V}$ *is a local minimum of* $\widetilde{J}(\mathbf{U},\mathbf{V},\mathbf{w};\mathbf{D})$ *of (7) if and only if the vectors* $\mathbf{v}_i$ *in* $\mathbf{V}$ *are given by (9).*

Proof. Because $\mathbf{U}$ and $\mathbf{w}$ are both fixed, the proof of Theorem 2 is the same as that of Proposition 2 in [1]. □

Theorem 3. *Let* $\mathbf{U}$ *and* $\mathbf{V}$ *be fixed. Then,* $\mathbf{w}$ *is a local minimum of* $\widetilde{J}(\mathbf{U},\mathbf{V},\mathbf{w};\mathbf{D})$ *of (7) if and only if* w_q *is given by (10).*

Proof. The necessity condition can be proved by the derivation of (10). To verify the sufficiency, we examine the second partial derivative of (7) with respect to the elements of $\mathbf{w}$. It can be deduced that

$$h_{p,q}(\mathbf{w}) = \frac{\partial}{\partial w_p}\left[\frac{\partial \widetilde{J}(\mathbf{w})}{\partial w_q}\right] = \begin{cases} 2\sum_{i=1}^{C}\sum_{j=1}^{N}\mu_{ij}^{m}(x_{jq}-v_{iq})^2 & \text{if } p=q \\ 0 & \text{otherwise} \end{cases}, \tag{12}$$

where we substitute $\widetilde{J}(\mathbf{U},\mathbf{V},\mathbf{w};\mathbf{D})$ by $\widetilde{J}(\mathbf{w})$ for simplicity. Because $\sum_{i=1}^{C}\sum_{j=1}^{N}\mu_{ij}^{m}(x_{jq}-v_{iq})^2 > 0$, we know that $H(\mathbf{w}) = (h_{p,q}(\mathbf{w}))_{d\times d}$ is a positive-definite diagonal matrix. Thus, (10) is also a sufficient condition for minimizing $\widetilde{J}(\mathbf{w})$. □

According to the theoretical result obtained by Bezdek [1], we have $\widetilde{J}(\mathbf{U}^{(t+1)},\mathbf{V}^{(t+1)},\mathbf{w}^{(t+1)};\mathbf{D}) \leqslant \widetilde{J}(\mathbf{U}^{(t)},\mathbf{V}^{(t)},\mathbf{w}^{(t)};\mathbf{D})$ based on Theorems 1–3. Therefore, the proposed IFWFCM converges after a certain number of iterations.

3.4. Computational complexity

Considering the development of the IFWFCM, the complexity is $O((2T-1)NCd)$, where T is the total number of iterations, N is the number of samples, C is the number of clusters, and d is the number of features. In comparison, the computational complexities of the FWFCM proposed by Wang et al. and the FWFCM proposed by Hung et al. are $O((B+TC)Nd)$ and $O((NT_1+TC)Nd)$, respectively. B is the bootstrap replication number in [14], while T_1 is the number of iterations that are used for learning the feature weights in [28]. Although the computational complexity of IFWFCM is more expensive than that of the traditional FCM, i.e., $O(TNCd)$, it is still suitable for dealing with large data sets. This topic will be examined in Section 5.

4. IFWFCM based on the kernelized distance

In this section, IFWFCM based on the kernelized distance (IFWFCM_KD) is introduced. The objective function of IFWFCM_KD to be minimized is as follows

$$J_{KD}(\mathbf{U},\mathbf{V},\mathbf{w};\mathbf{D}) = \sum_{i=1}^{C}\sum_{j=1}^{N}\mu_{ij}^{m}(d_{ij\mathbf{w}}^{\Phi})^2, \tag{13}$$

where $d_{ij\mathbf{w}}^{\Phi}$ is the weighted kernelized distance, which is defined as

$$d_{ij\mathbf{w}}^{\Phi} = \|\Phi(\text{diag}(\mathbf{w})\mathbf{x}_j) - \Phi(\text{diag}(\mathbf{w})\mathbf{v}_i)\|. \tag{14}$$

According to the kernel trick [21], the explicit expression of the nonlinear mapping $\Phi(\cdot)$ in (14) need not be computed. The nonlinear mapping can be performed implicitly by employing a nonlinear kernel function. Moreover, the kernel function can be evaluated on samples in the original input space. Thus, the weighted kernelized distance $d_{ij\mathbf{w}}^{\Phi}$ can be calculated by

$$d_{ij\mathbf{w}}^{\Phi} = \sqrt{K(\tilde{\mathbf{x}}_j,\tilde{\mathbf{x}}_j) - 2K(\tilde{\mathbf{x}}_j,\tilde{\mathbf{v}}_i) + K(\tilde{\mathbf{v}}_i,\tilde{\mathbf{v}}_i)}, \tag{15}$$

where $\tilde{\mathbf{x}}_j = \text{diag}(\mathbf{w})\mathbf{x}_j$ and $\tilde{\mathbf{v}}_i = \text{diag}(\mathbf{w})\mathbf{v}_i$. One can refer to [27] for the choice of kernel functions.

Taking the constraints (5) and (6) into consideration, we obtain the objective function with Lagrange terms as

$$\widetilde{J}_{KD}(\mathbf{U},\mathbf{V},\mathbf{w};\mathbf{D}) = \sum_{i=1}^{C}\sum_{j=1}^{N}\mu_{ij}^{m}\left(d_{ij\mathbf{w}}^{\Phi}\right)^{2} - \lambda\left(\sum_{i=1}^{C}\mu_{ij} - 1\right) - \beta\left(\sum_{q=1}^{d}w_q - 1\right). \tag{16}$$

In the paper, the Gaussian kernel function $K(\mathbf{x},\mathbf{y}) = \exp\left\{-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}\right\}$ is utilized for calculating the kernelized distance. Hence, the following update equations for μ_{ij}, $\mathbf{v}_i$ and w_q can be obtained.

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{C}\left[\frac{\left(d_{ij\mathbf{w}}^{\Phi}\right)^2}{\left(d_{kj\mathbf{w}}^{\Phi}\right)^2}\right]^{\frac{1}{m-1}}}, \tag{17}$$

$$\mathbf{v}_i = \frac{\sum_{j=1}^{N}\mu_{ij}^{m}K(\mathrm{diag}(\mathbf{w})\mathbf{x}_j,\mathrm{diag}(\mathbf{w})\mathbf{v}_i)\mathbf{x}_j}{\sum_{j=1}^{N}\mu_{ij}^{m}K(\mathrm{diag}(\mathbf{w})\mathbf{x}_j,\mathrm{diag}(\mathbf{w})\mathbf{v}_i)}, \tag{18}$$

and

$$w_q = \frac{1}{\sum_{l=1}^{d}\left[\frac{\sum_{i=1}^{C}\sum_{j=1}^{N}\mu_{ij}^{m}K(\mathrm{diag}(\mathbf{w})\mathbf{x}_j,\mathrm{diag}(\mathbf{w})\mathbf{v}_i)(x_{jq}-v_{iq})^2}{\sum_{i=1}^{C}\sum_{j=1}^{N}\mu_{ij}^{m}K(\mathrm{diag}(\mathbf{w})\mathbf{x}_j,\mathrm{diag}(\mathbf{w})\mathbf{v}_i)(x_{jl}-v_{il})^2}\right]}. \tag{19}$$

The proof of convergence for IFWFCM_KD can be obtained similarly from that for IFWFCM in Section 3.

5. Experimental results

To achieve better clustering performance, the settings of the parameters in the FWFCM proposed by Wang et al. are as follows: the learning rate $\eta = 0.8$, the maximum number of iterations $N_I = 500$, the number of points uniformly selected from the interval [0, 1] to determine the parameter β is set at $N_P = 1000$, and the termination threshold $\delta = 10^{-10}$. For the FWFCM proposed by Hung et al., the bootstrap replication number B = 200, which is the same as the value used in [14].

5.1. Synthetic data sets

To test the effectiveness of the proposed method, two synthetic data sets are generated. All of the samples in the two data sets are generated from the different Gaussian distributions. For convenience, the two synthetic data sets are referred to as *Dataset1* and *Dataset2*, respectively. There are 200 two-dimensional samples in *Dataset1*, with 50 samples in each cluster. The means μ_i $(i = 1, 2, 3, 4)$ of the samples for the four clusters are $(2,2)^T, (2,4)^T, (4,2)^T$, and $(4,4)^T$, while the covariance matrices $\Sigma_i = \begin{pmatrix} 0.25 & 0 \\ 0 & 0.25 \end{pmatrix}$. For *Dataset2*, there are 100 two-dimensional samples in each of the three clusters. The means μ_i $(i = 1, 2, 3)$ of the samples for the three clusters are $(2,2)^T, (2,4)^T$, and $(2,6)^T$, while $\Sigma_i = \begin{pmatrix} 2 & 0 \\ 0 & 0.2 \end{pmatrix}$. Fig. 1 shows the two synthetic data sets.

It can be observed from Fig. 1(a) that the contributions of the two features, i.e., F1 and F2, are approximately the same when constructing the clusters. Nevertheless, it can be deduced from Fig. 1(b) that the contribution of the second feature F2 is larger than that of the first feature F1 when generating the clusters.

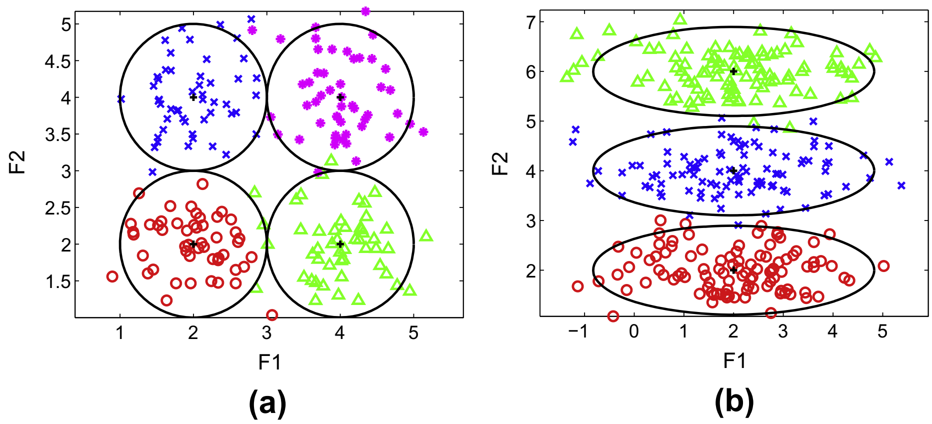

Fig. 1. The two synthetic data sets with their samples obeying Gaussian distributions: (a) *Dataset1*; (b) *Dataset2*.

For FCM, FWFCM proposed by Wang et al. [28], FWFCM proposed by Hung et al. [14] and the proposed method, the fuzzification exponent m is set to be 1.5, and the termination tolerance ϵ is 10^{-4}. Therefore, the feature weights learned by the latter three methods using *Dataset1* are illustrated in Fig. 2(a), while Fig. 2(b) is for *Dataset2*. From Fig. 2(a), we can find that all of the methods except for FWFCM proposed by Wang et al. can assign approximately equal feature weights for F1 and F2 of *Dataset1*. For *Dataset2*, it can be observed from Fig. 2(b) that both FWFCM proposed by Wang et al. and the proposed method can learn appropriate feature weights because they both assign a larger value for F2 and a smaller value for F1. However, FWFCM proposed by Hung et al. cannot learn appropriate feature weights for *Dataset2* because it assigns a larger value for F1 than for F2.

The error rates of the four approaches reported for the two synthetic data sets are summarized in Table 1. Moreover, the running time of the four methods on the two data sets is also included in the table. Fig. 3 demonstrates the performances of the four methods on *Dataset2*. It can be observed from Table 1 that all of the four approaches have the same error rate on *Dataset1*. Moreover, the proposed method achieves more promising performance on *Dataset2* in comparison with the other three approaches. It can also be observed from the results on *Dataset2* in Table 1 and Fig. 3 that the samples scaled by an appropriate feature-weight vector can improve the performance of FCM, while an inappropriate vector can deteriorate the performance of FCM.

Furthermore, it is shown in Table 1 that the running time of the proposed method is even lower than for FCM on *Dataset1*. Compared to FWFCM proposed by Wang et al., the proposed method comes with lower running time for the two synthetic data sets. Moreover, the proposed method requires the same running time but exhibits a superior performance in comparison with FWFCM proposed by Hung et al. on *Dataset2*.

5.2. Benchmark data sets

In the following experiments, the three benchmark data sets are utilized. It should be noted here that the three data sets are all originally designed for supervised learning rather than for unsupervised learning. In the following experiment, the provided class labels of the three data sets are used to evaluate the performance of the aforementioned clustering approaches.

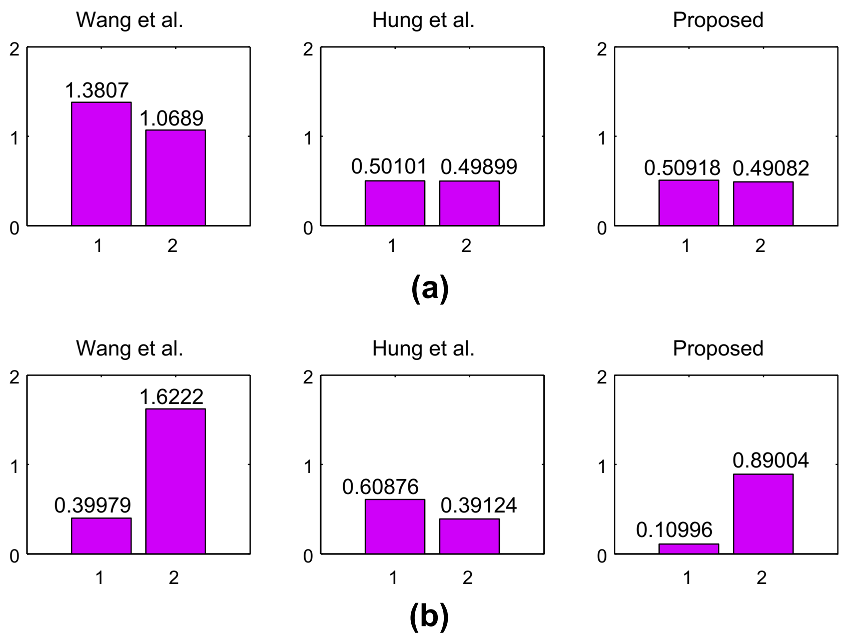

Fig. 2. Feature weights determined by using three approaches on the two synthetic data sets: (a) *Dataset1*; (b) *Dataset2*.

Table 1
Error rates and running time of the four clustering methods reported on the two data sets.

Data set	FCM		Wang et al.		Hung et al.		Our method	
	Error (%)	Time (s)	Error (%)	Time (s)	Error (%)	Time (s)	Error (%)	Time (s)
Dataset1	4	0.09	4	74.14	4	0.02	4	0.02
Dataset2	16	0.05	2	165.80	30	0.05	1.67	0.05

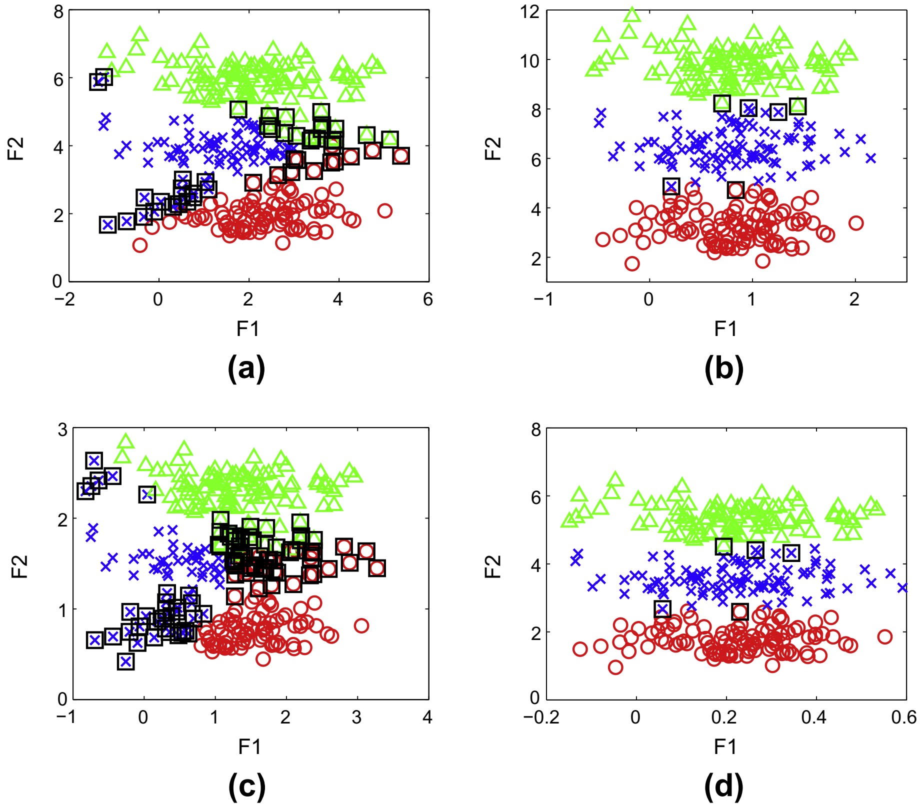

Fig. 3. The clustering results and the error rates of the four different methods upon *Dataset2*: (a) FCM (*Err* = 16%); (b) Wang et al. (*Err* = 2%); (c) Hung et al. (*Err* = 30%); and (d) proposed (*Err* = 1.67%)

The description of the three benchmark data sets is as follows.

Iris data set: This data set [3] consists of 150 four-dimensional samples. It comes with three classes of equal size (50).

Crude_oil data set: This data set is a five-dimensional data set [15] with 56 samples and three classes. There are five features for each sample. Moreover, there are 7 samples in the first class, 11 samples in the second class, and 38 samples in the third class.

Bupa data set: This data set [3] consists of 345 samples with six features. It is divided into two classes. There are 145 samples in the first class, and the remaining 200 samples are in the second class.

For each data set, the number of clusters C for each clustering method is set to be the same as the corresponding number of classes (which is provided with the dataset). Moreover, the fuzzification exponent m is set to be 1.5, while the termination tolerance ϵ is still 10^{-4}. The feature weights, error rates and running time obtained by the three different methods, i.e., the FWFCM proposed by Wang et al., the FWFCM proposed by Hung et al. and the proposed method on the *Iris* data set, are summarized in Table 2. The results for the different methods on the *Crude_oil* and *Bupa* data sets are included in Tables 3 and 4, respectively. Moreover, the error rates and running time of the traditional FCM are also provided in Tables 2–4.

From Tables 2–4, it can be easily observed that the proposed method achieves superior performance on all of the three data sets in comparison with the other three methods. For the *Iris* data set, the two FWFCMs and IFWFCM methods outperform the traditional FCM approach because their error rates are all less than that of FCM. In comparison with the two FWFCM methods, the proposed approach is the most efficient because the learning time is the shortest. For the *Crude_oil* and *Bupa* data sets, although they have a longer computing time, the proposed method achieves better performance in comparison

Table 2
Clustering results obtained for *Iris* data set.

Method	Feature-weight vector	Error rates (%)	Time (s)
FCM	–	12	0.02
Wang et al.	(0.0005, 0.0000, 1.9829, 0.1355)	6	46.39
Hung et al.	(0.1020, 0.1022, 0.3377, 0.4580)	4	0.05
Our method	(0.1194, 0.1134, 0.4346, 0.3327)	3.33	0.02

Table 3
Clustering results obtained for *Crude_oil* data set.

Method	Feature-weight vector	Error rates (%)	Time (s)
FCM	–	37.5	0.02
Wang et al.	(0.0542, 1.6552, 0.8471, 0.1593, 0.0163)	37.5	8.97
Hung et al.	(0.1570, 0.1732, 0.3638, 0.1064, 0.1996)	39.29	0.05
Our method	(0.1134, 0.1007, 0.1023, 0.6228, 0.0608)	30.36	0.03

Table 4
Clustering results obtained for *Bupa* data set.

Method	Feature-weight vector	Error rates (%)	Time (s)
FCM	–	48.41	0.02
Wang et al.	(0.6240, 0.5724, 0.4655, 0.7079, 1.5158, 0.8796)	47.25	296.45
Hung et al.	(0.5365, 0.0458, 0.5099, 0.4653, 0.4813, 0.0341)	49.86	0.11
Our method	(0.1536, 0.0803, 0.2337, 0.2361, 0.2107, 0.0856)	45.8	0.14

with the traditional FCM method. However, the performances of the method proposed by Hung et al. are both worse than those of FCM. Although the method proposed by Wang et al. can improve the performances of FCM on the *Crude_oil* and *Bupa*, the computational complexity is prohibitively demanding.

To get an intuitive view, the clustering results of the above four approaches on the *Iris* data set are depicted in Fig. 4. In the figure, the third and fourth features of the *Iris* data set are chosen to demonstrate the results. The reason is that, for the three feature-weighted methods, i.e., the method proposed by Wang et al., the approach proposed by Hung et al. and our proposed method, the learned weights for these two features are all larger than those of the first and second features.

5.3. Color image segmentation

In this experiment, the proposed method is compared with the traditional FCM and FWFCM proposed by Hung et al. in color image segmentation. As stated in Section 3.4, the computational requirement for the FWFCM proposed by Wang et al. is

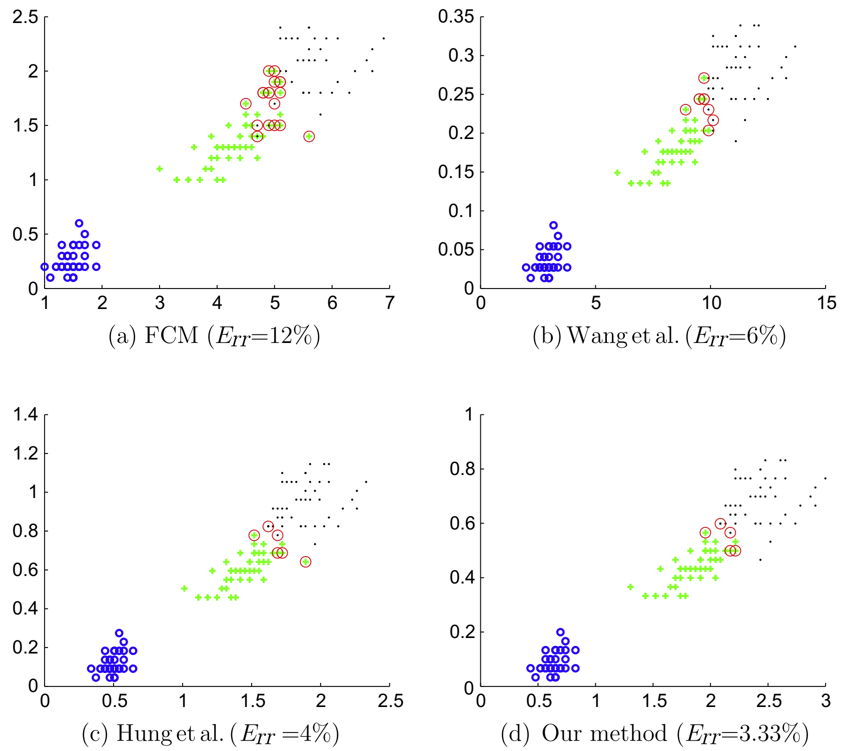

Fig. 4. The clustering results and the error rates of the four different approaches on the third and fourth features of the *Iris* data set. The samples enclosed by the red circle are those misclassified. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5
Number of clusters for all the methods upon the ten different images in the Berkeley segmentation dataset.

Image	*Plane*	*Buffalo*	*Horse1*	*Horse2*	*Sun flower*
Number of clusters	3	2	3	3	5
Image	*Leopard*	*Hawk*	*Bird*	*Night*	*Pyramid*
Number of clusters	3	2	5	3	3

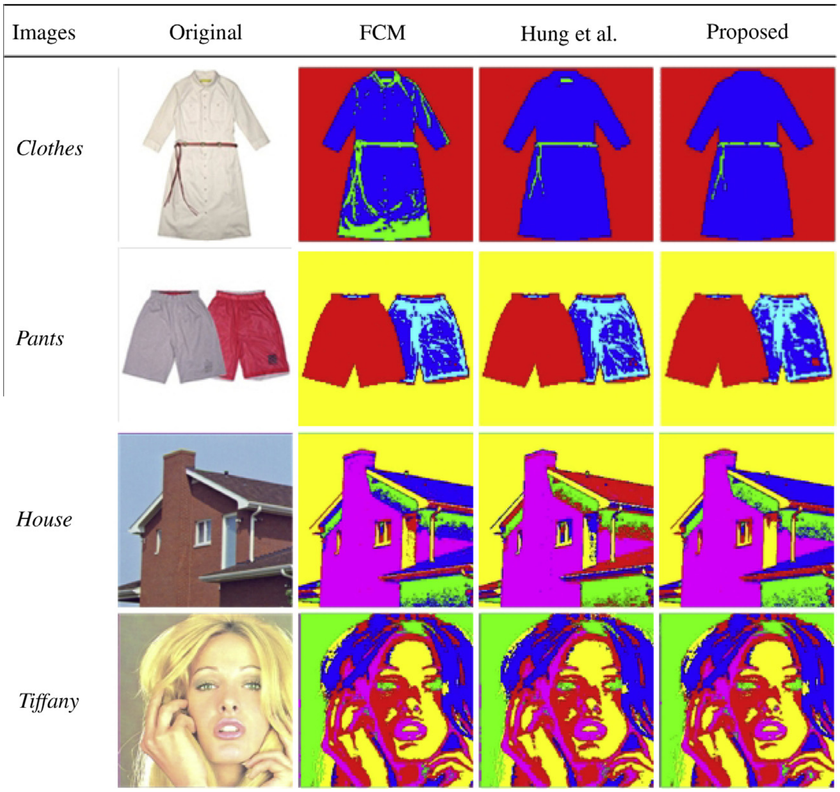

Fig. 5. Four original images from [16] and the USC-SIPI image database together with their segmentation results by the three methods. For each image group, from left to right: Original image, result of FCM, result of FWFCM proposed by Hung et al., and result of the proposed method.

in order $O((NT_1 + TC)Nd)$. This is prohibitive for segmenting the color images in the following experiments. Therefore, the results of the FWFCM proposed by Wang et al. cannot be demonstrated in this subsection. The two images, i.e., *Clothes* and *Pants* from [16] together with the two images, i.e., *House* and *Tiffany* from the USC-SIPI image database,[1] are utilized. Moreover, the ten natural color images from the famous public Berkeley segmentation dataset[2] are also used to test the efficiency of the proposed method. Similar to the settings of the parameters in [14], the fuzzification exponent m for all of the methods is set to be 2, and the termination tolerance ϵ is 10^{-4}. The number of clusters C is 3 and 4 for *Clothes* and *Pants*, respectively, while is 5 for both *House* and *Tiffany*. For the ten images taken from the Berkeley segmentation dataset, the fuzzification exponent m and the termination tolerance ϵ are also set to be 2 and 10^{-4}, respectively. The numbers of clusters C for all of the approaches on the ten images are reported in Table 5.

To make all the methods more efficient to deal with color image segmentation, the CIELAB color space rather than the RGB color space is chosen to describe the pixels in a given image. There are many published studies that address this issue. One can refer to [16,14,17] for detailed explanations.

The segmentation results of FCM, FWFCM proposed by Hung et al. and the proposed method on *Clothes*, *Pants*, *House* and *Tiffany* are summarized in Fig. 5. The sizes of *Clothes* and *Pants* are both 128×128 pixels, while the size is 256×256 pixels for *House* and *Tiffany*. For the *Clothes* image, it is shown in Fig. 5 that FCM incorrectly classifies the belt and coat because the whole belt region and certain parts of the coat region are all green in color. In comparison with the segmentation result of FCM, the results of FWFCM proposed by Hung et al. and our method are more close to human perceptions. However, there

[1] http://sipi.usc.edu/database/database.php?volume=misc.
[2] http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/BSDS300/html/dataset/images/color.

are still misclassification regions for both methods, which include the neckline region obtained by FWFCM proposed by Hung et al. and the belt region achieved by the proposed method. For the *Pants* image, the segmentation results of the three methods are exactly the same except that the rectangular mark on the red pants in the original image are incorrectly classified in part by the former two methods. With respect to *House* and *Tiffany*, the segmentation results of the three different approaches are nearly the same from the visual perspective.

The segmentation results of the three approaches on the ten images from the Berkeley segmentation dataset are shown in Fig. 6. The sizes of the original images in Fig. 6 are all 481 $\times$ 321 pixels. From the viewpoint of human perception, most of the segmentation results are approximately the same. However, for *Sun flower*, the segmentation result of the proposed method can segment the region of the pistil correctly, while the other two methods cannot. For most of the images, especially *Plane*,

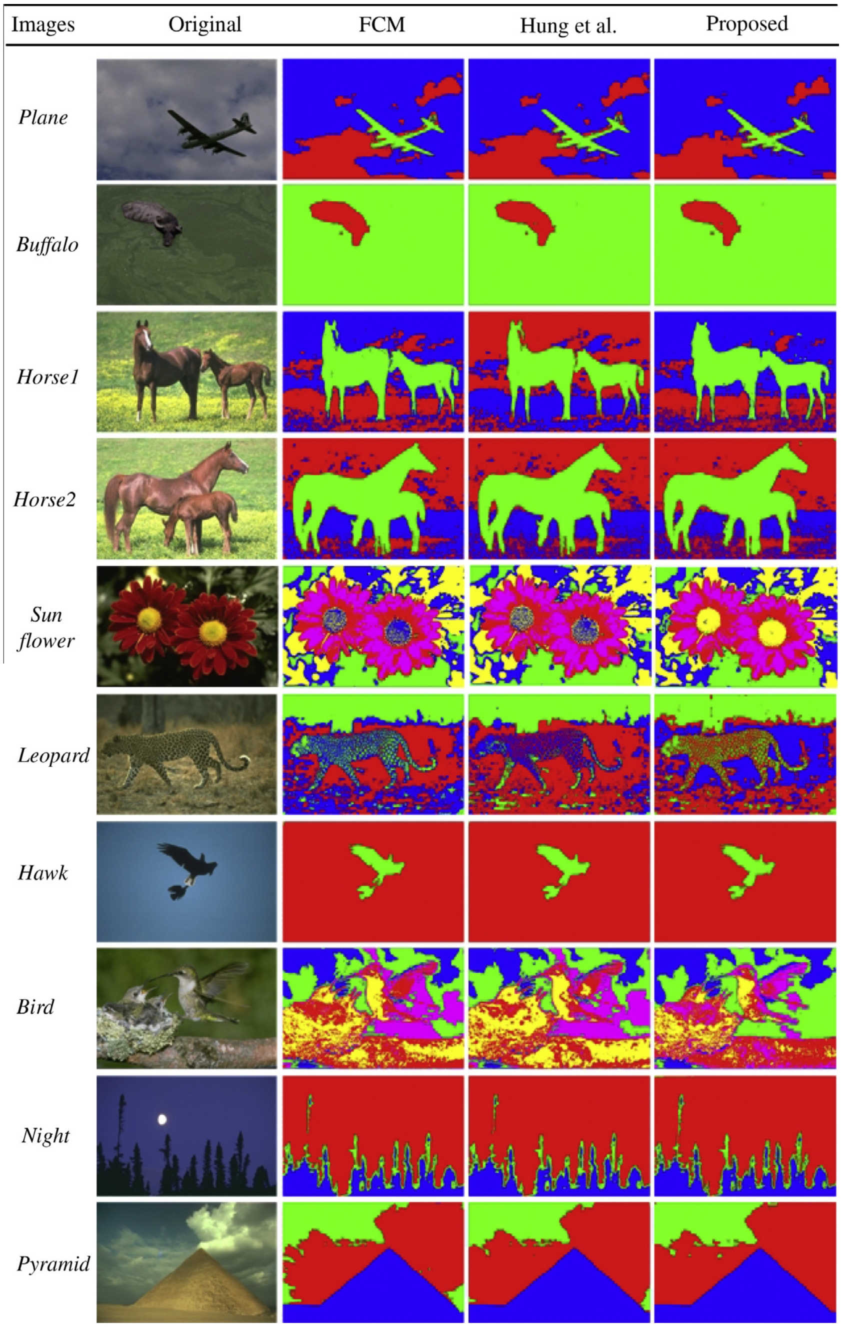

Fig. 6. Ten original images from the Berkeley segmentation dataset and their segmented results by the three methods. For each image group, from left to right: Original image, result of FCM, result of FWFCM proposed by Hung et al., and result of the proposed method.

Horse2 and *Bird*, the segmentation results of the proposed method on the region of the background are better than those of the other two approaches.

To quantitatively evaluate the segmentation results, the evaluation function proposed by Liu and Yang [18] is utilized. The evaluation function $F(I)$ is given below

$$F(I) = \frac{1}{1000(N \times M)} \sqrt{R} \sum_{j=1}^{R} \frac{e_j^2}{\sqrt{A_i}}, \tag{20}$$

where I is an image to be segmented, $N \times M$ is the total pixels in I, R is the number of regions in the segmented image, A_i is the number of pixels of the ith region, and e_i is the color error of the ith region. It should be noted here that the smaller the value of $F(I)$, the better the segmentation result is. Table 6 shows a quantitative evaluation of the segmented results. Moreover, the running times of the three different approaches used for the 14 images are also included in Table 6.

The values of $F(I)$ produced by FCM lower than those of the proposed method on all of the images except for *House*, *Tiffany*, *Buffalo*, and *Night* refer to Table 6. Similarly, the values of $F(I)$ in FCM are also lower than those of FWFCM proposed by Hung et al. for all of the images except for *Plane*, *Buffalo*, *Sun flower*, and *Hawk*. In comparison with FWFCM proposed by Hung et al., the proposed method achieves smaller values of $F(I)$ on seven images and larger values on the remaining seven images. Moreover, the running times of FCM are shorter than those of the proposed method on all of the images except for *Horse2*, *Leopard*, and *Pyramid*. The elapsed times of FCM are also shorter than those for FWFCM proposed by Hung et al. for all of the images except for *Horse2*. Compared to FWFCM proposed by Hung et al., the proposed method comes with lower computing overhead on all of the images except for *Pants*, *House*, and *Tiffany*.

Therefore, the results in Table 6 indicate that FCM is superior to the other two methods in the color image segmentation. Nevertheless, as stated before, FCM cannot precisely segment some of the regions within certain images, for example, the rectangular mark of *Pants* in Fig. 5 and the pistil region of *Sun flower* together with the background regions of *Plane*, *Horse2* and *Bird* in Fig. 6. In contrast, the proposed method can successfully address the aforementioned problems. Moreover, considering the values of the evaluation function and the elapsed time, we can find, from Table 6, that the proposed method is more promising than FWFCM proposed by Hung et al. for segmenting color images.

5.4. Experimental results for IFWFCM_KD

To determine the impact of kernelization on the proposed IFWFCM, we apply the IFWFCM_KD method introduced in Section 4 to study all of the numerical data sets and color images used in Sections 5.1, 5.2 and 5.3. It should be mentioned here that the performance of IFWFCM_KD largely depends on the choice of the width parameter σ in the Gaussian kernel function. In this study, the optimal parameter σ is chosen from the set of values $\{10^{-6}, 10^{-5}, \ldots, 10^5, 10^6\}$ by considering the performance of IFWFCM_KD. Through experiments, we find that the performance of IFWFCM_KD deteriorates when the value of σ is chosen to be less than 0.1 or greater than 10^6. Moreover, the optimal clustering performance of IFWFCM_KD can be obtained and it remain fixed when the value of σ is selected to be greater than 1 and less than 10^5. Therefore, in the following experiments, the width parameter σ in the Gaussian kernel function is set to be 10 for all of the numerical data sets and color images.

The results of IFWFCM_KD upon the synthetic and benchmark data sets are summarized in Table 7. To make the results more readable, the results of the proposed IFWFCM method are also included in the table. It should be noted here that all of the results of IFWFCM are taken directly from Sections 5.1 and 5.2. From Table 7, we can easily find that the error rates of

Table 6
Values of $F(I)$ and computing time of the three different methods reported for the 14 selected images.

Image	FCM		Hung et al.		Proposed method	
	$F(I)$	Time (s)	$F(I)$	Time (s)	$F(I)$	Time (s)
Clothes	0.2933	0.50	0.3367	1.06	0.3343	0.69
Pants	0.4043	0.31	0.5758	0.86	0.4906	1.52
House	6.4719	8.30	6.5074	9.95	6.3281	10.36
Tiffany	3.4583	5.11	3.5724	7.48	3.4238	11.17
Plane	3.7435	4.52	3.7372	18.41	3.9908	9.06
Buffalo	4.8305	1.83	4.8292	11.19	4.8249	3.16
Horse1	29.529	2.84	29.7882	13.45	31.603	6.58
Horse2	21.94	25.47	22.0214	15.5	22.394	6.97
Sun flower	74.113	6.69	73.9109	19.22	86.438	16.16
Leopard	13.442	9.45	14.1762	26.52	13.6787	6.98
Hawk	4.1866	1.88	4.181	13.41	4.3814	2.48
Bird	16.679	16.02	17.4773	30.47	18.236	24.38
Night	4.1134	4.03	4.1167	15.91	3.9883	6.09
Pyramid	18.188	4.50	18.6582	15.56	19.017	4.36

Table 7
Error rates and running time of IFWFCM and its kernelized version reported for the synthetic and benchmark data sets.

Data set	IFWFCM		IFWFCM_KD	
	Error rate (%)	Time (s)	Error rate (%)	Time (s)
Dataset1	4	0.02	4	0.17
Dataset2	1.67	0.05	1.67	0.3
Iris	3.33	0.02	3.33	0.08
Crude_oil	30.36	0.03	30.36	0.06
Bupa	45.8	0.14	45.80	0.23

IFWFCM_KD are all the same as their counterparts in IFWFCM. However, the running times of IFWFCM_KD are all longer than those of IFWFCM.

The segmentation results of IFWFCM_KD for the four original color images in Fig. 5 are summarized in Fig. 7, and their corresponding running times are included in Table 8. Comparing the sub-figures in Fig. 7 with their counterparts in Fig. 5, we find that they are all the same. However, it is shown in Table 8 that the running times of IFWFCM_KD are much longer than those of IFWFCM. It should be noted that the values of $F(I)$ in IFWFCM_KD on the segmented results are not

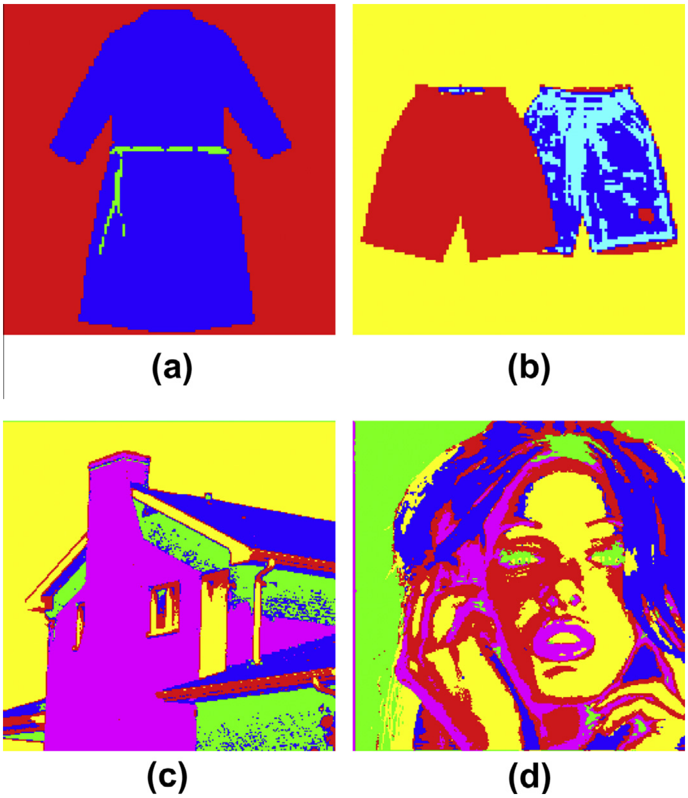

Fig. 7. The segmentation results of IFWFCM_KD upon the four images from [16] and the USC-SIPI image database: (a) *Clothes*, (b) *Pants*, (c) *House* and (d) *Tiffany*.

Table 8
Running time of IFWFCM and its kernelized version upon the four different images (s).

Image	IFWFCM	IFWFCM_KD
Clothes	0.69	40.73
Pants	1.52	64.61
House	10.73	1186.14
Tiffany	12.48	1199.77
Average	6.36	622.81

included in Table 8 only because they are the same as those of IFWFCM. Moreover, the segmentation results of IFWFCM_KD on the original images in Fig. 6 are also not demonstrated here because they are the same as those of IFWFCM.

Therefore, in comparison with IFWFCM, IFWFCM_KD produces the same performances on the synthetic and benchmark data sets and the same segmentation results on the color images. However, the computational complexity of IFWFCM_KD is more expensive than that of IFWFCM.

6. Conclusions

Many FWFCMs have been proposed in the field of machine learning to make the traditional FCM address the samples using weighted features. The proposed approach, i.e., IFWFCM, reformulates the objective function and the update equations from the existing FWFCMs. The algorithm implementation, convergence property and computational complexity of IFWFCM are also included. Furthermore, we extend the proposed IFWFCM to its kernelized version, i.e., IFWFCM_KD. Experimental results show that IFWFCM is superior to the existing FWFCMs. Moreover, through experiments, we observe that IFWFCM_KD might not improve the performance of IFWFCM on certain data sets and color images.

To make the proposed IFWFCM more efficient, there are two tasks for future investigation. First, similar to the traditional FCM, the number of clusters for the proposed method must be predefined. To overcome this shortcoming, the model selection [20] and cluster validity function-based strategies [30,7] will be considered. Second, it is well known that incorporating spatial constraints into the FCM objective function can enhance the robustness of the original clustering algorithm to noise and outliers [6]. We therefore will investigate this issue for our proposed IFWFCM to make it more robust with respect to segmenting color images that are corrupted by noise.

Acknowledgments

The authors would like to thank Prof. Dae-Won Kim for having provided us with the color images. This work is partly supported by the National Natural Science Foundation of China (Nos. 60903089; 61073121; 61075051; and 61170040) and the Natural Science Foundation of Hebei Province (Nos. F2013201060; F2012201023; and F2012402037). In addition, the authors would like to express their special thanks to the referees for their valuable comments and suggestions for improving this study.

Appendix A

In this appendix, the detailed derivation for obtaining the update Eqs. (8)–(10) is provided. Setting the partial derivatives of the Lagrangian function (7) with respect to μ_{ij} and β to zero, we can obtain

$$\mu_{ij} = \left(\frac{\lambda}{m\left[d_{ij}^{(\mathbf{w})}\right]^2} \right)^{\frac{1}{m-1}} \tag{A.1}$$

and

$$\sum_{i=1}^{C} \mu_{ij} = 1 \tag{A.2}$$

Therefore, $1 = \sum_{k=1}^{C} \mu_{kj} = \left(\frac{\lambda}{m}\right)^{\frac{1}{m-1}} \sum_{k=1}^{C} \left(\frac{1}{\left[d_{kj}^{(\mathbf{w})}\right]^2} \right)^{\frac{1}{m-1}}$, which yields $\left(\frac{\lambda}{m}\right)^{\frac{1}{m-1}} = \frac{1}{\sum_{k=1}^{C} \left(\frac{1}{\left[d_{kj}^{(\mathbf{w})}\right]^2} \right)^{\frac{1}{m-1}}}$. Substituting into Eq. (A.1), yields the update Eq. (8), i.e.,

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\left[d_{ij}^{(\mathbf{w})}\right]^2}{\left[d_{kj}^{(\mathbf{w})}\right]^2} \right)^{\frac{1}{m-1}}}. \tag{A.3}$$

Setting the partial derivative of the Lagrangian function (7) with respect to $\mathbf{v}_i$ to zero yields the update Eq. (8),

$$\mathbf{v}_i = \frac{\sum_{j=1}^{N} \mu_{ij} \mathbf{x}_j}{\sum_{j=1}^{N} \mu_{ij}}. \tag{A.4}$$

It should be noted that the above equation holds under the condition $|\mathrm{diag}(\mathbf{w})| \neq 0$.

Setting the partial derivatives of the Lagrangian function (7) with respect to w_k and β to zero, we can obtain

$$w_q = \frac{\beta}{2\sum_{i=1}^{C}\sum_{j=1}^{N}\mu_{ij}^{m}(x_{jq} - v_{iq})^2} \tag{A.5}$$

and

$$\sum_{q=1}^{d} w_q = 1 \tag{A.6}$$

Hence, it can be deduced from $1 = \sum_{l=1}^{d} w_l = \beta \sum_{l=1}^{d} \frac{1}{2\sum_{i=1}^{C}\sum_{j=1}^{N}\mu_{ij}^{m}(x_{jl}-v_{il})^2}$ that $\beta = \frac{1}{\sum_{l=1}^{d}\left[\frac{1}{2\sum_{i=1}^{C}\sum_{j=1}^{N}\mu_{ij}^{m}(x_{jl}-v_{il})^2}\right]}$. Substituting into Eq. (A.5), yields the update Eq. (A.5),

$$w_q = \frac{1}{\sum_{l=1}^{d}\left[\frac{\sum_{i=1}^{C}\sum_{j=1}^{N}\mu_{ij}^{m}(x_{jq}-v_{iq})^2}{\sum_{i=1}^{C}\sum_{j=1}^{N}\mu_{ij}^{m}(x_{jl}-v_{il})^2}\right]} \tag{A.7}$$

References

[1] J.C. Bezdek, A convergence theorem for the fuzzy ISODATA clustering algorithms, IEEE Trans. Patt. Anal. Mach. Intell. 2 (1) (1980) 1–8.
[2] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithm, Plenum Press, New York, 1981.
[3] C.L. Blake, C.J. Merz, UCI Repository of Machine Learning Databases, Department of Information and Computer Science, University of California, Irvine, CA, USA, 1998 <http://www.ics.uci.edu/mlearn/MLRepository.html>.
[4] C. Borgelt, Accelerating fuzzy clustering, Inform. Sci. 179 (2009) 3985–3997.
[5] B.K. Brouwer, A method of relational fuzzy clustering based on producing feature vectors using FastMap, Inform. Sci. 179 (2009) 3561–3582.
[6] S. Chen, D. Zhang, Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure, IEEE Trans. Syst., Man, Cybernet.-Part B: Cybernet. 34 (2004) 1907–1916.
[7] S. Das, S. Sil, Kernel-induced fuzzy clustering of image pixels with an improved differential evolution algorithm, Inform. Sci. 180 (2010) 1237–1256.
[8] J.C. Dunn, Some recent investigations of a new fuzzy partition algorithm and its application to pattern classification problems, J. Cybernet. 4 (1974) 1–15.
[9] H. Frigui, O. Nasraoui, Unsupervised learning of prototypes and attribute weights, Patt. Recog. 37 (2004) 567–581.
[10] H. Fu, A.M. Elmisery, A new feature weighted fuzzy c-means clustering algorithm, in: Proceedings of IADIS European Conference on Data Mining, Algarve, Portugal, 2009, pp. 11–18.
[11] A. Ghosh, N.S. Mishra, S. Ghosh, Fuzzy clustering algorithms for unsupervised change detection in remote sensing images, Inform. Sci. 181 (2011) 699–715.
[12] D. Graves, W. Pedrycz, Kernel-based fuzzy clustering and fuzzy clustering: a comparative experimental study, Fuzzy Sets Syst. 161 (2010) 522–543.
[13] J.W. Han, M. Kamber, Data Mining: Concept and Techniques, Morgan Kaufmann, San Mateo, 2000.
[14] W.L. Hung, M.S. Yang, D.H. Chen, Bootstrapping approach to feature-weight selection in fuzzy c-means algorithms with an application in color image segmentation, Patt. Recog. Lett. 29 (2004) 1317–1325.
[15] R.A. Johnson, D.W. Wichern, Applied Multivariate Statistical Analysis, Prentice-Hall, London, 1982.
[16] D.W. Kim, K.H. Lee, D. Lee, A novel initialization scheme for the fuzzy c-means algorithm for color clustering, Patt. Recog. Lett. 25 (2004) 227–237.
[17] C.P. Lim, W.S. Ooi, An empirical analysis of colour image segmentation using fuzzy c-means clustering, Int. J. Knowl. Eng. Soft Data Parad. 2 (1) (2010) 97–106.
[18] J. Liu, Y.H. Yang, Multiresolution color image segmentation technique, IEEE Trans. Patt. Anal. Mach. Intell. 16 (1994) 689–700.
[19] E.A. Maharaj, P. D'Urso, Fuzzy clustering of time series in the frequency domain, Inform. Sci. 181 (2011) 1187–1211.
[20] D. Pelleg, A. Moore, X-means: extending k-means with efficient estimation of the number of clusters, in: Proceedings of the 17th International Conference on Machine Learning, 2000, pp. 727–734.
[21] B. Schölkopf, The kernel trick for distances, in: Advances in Neural Information Processing Systems, 2001, pp. 301–307.
[22] J. Shawe-Taylor, N. Cristianini, Kernel Methods for Pattern Analysis, Cambridge University Press, Cambridge, 2004.
[23] H. Shen, J. Yang, S. Wang, X. Liu, Attribute weighted mercer kernel based fuzzy clustering algorithm for general non-spherical datasets, Soft Comput. 10 (2006) 1061–1073.
[24] K.P. Soman, S. Diwakar, A. Ajay, Insight into Data Mining: Theory and Practice, Prentice-Hall of India, New Delhi, 2006.
[25] K.S. Tan, N.A. Mat Isa, Color image segmentation using histogram thresholding – fuzzy C-means hybrid approach, Patt. Recog. 41 (1) (2011) 1–15.
[26] D.M. Tsai, C.C. Lin, Fuzzy c-means based clustering for linearly and nonlinearly separable data, Patt. Recog. 44 (2011) 1750–1760.
[27] V. Vapnik, Statistical Learning Theory, Wiley, 1998.
[28] X.Z. Wang, Y.D. Wang, L.J. Wang, Improving fuzzy c-means clustering based on feature-weight learning, Patt. Recog. Lett. 25 (2004) 1123–1132.
[29] Y. Yue, D. Zeng, L. Hong, Improving fuzzy c-means clustering by a novel feature-weight learning, in: Proceedings of 2008 IEEE Pacific–Asia Workshop on Computational Intelligence and Industrial Application, 2008, pp. 173–177.
[30] Y. Zhang, W. Wang, X. Zhang, Y. Li, A cluster validity index for fuzzy clustering, Inform. Sci. 178 (2008) 1205–1218.