Neurocomputing 335 (2019) 59–71

Contents lists available at ScienceDirect

Neurocomputing

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

A study on multi-kernel intuitionistic fuzzy C-means clustering with multiple attributes

Shan Zeng[a], Zhiyong Wang[b], Rui Huang[c,*], Ling Chen[a], David Feng[b,d]

[a] College of Mathematics and Computer Science, Wuhan Polytechnic University, Wuhan, Hubei 430023, PR China
[b] School of Information Technology and The BMIT research group, The University of Sydney, NSW 2006, Australia
[c] School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, Guangdong 518172, PR China
[d] The Med-X Research Institute, Shanghai Jiao Tong University, Shanghai 200240, PR China

ARTICLE INFO

Article history:
Received 17 September 2018
Revised 5 December 2018
Accepted 20 January 2019
Available online 23 January 2019

Communicated by Dr. Sato-Ilic Mika

Keywords:
Fuzzy C-means clustering
Intuitionistic fuzzy sets
Multi-kernel model

ABSTRACT

Fuzzy C-means (FCM) clustering has been widely applied in various data-driven applications. While traditional FCM clustering algorithms handle uncertainty with type-2 Fuzzy Sets (T2 FSs), the recently-proposed intuitionistic fuzzy sets (IFSs) have shown advantages for describing vague and uncertain data by taking both membership degree and non-membership degree into account. However, intuitionistic fuzzy C-means (IFCM) algorithms generally do not take the importance of individual attributes and the structure of the data into account, when multi-modal and imbalanced features are involved in an application. Therefore, in this paper, we propose to address this issue of IFCM with multi-kernel mapping. First, different types of features are grouped. Second, a composite kernel is constructed to map each attribute group into an individual kernel space and to linearly combine these kernels with optimal weights. Comprehensive experiments have been conducted on a wide range of datasets, such as machine learning repository (UCI) dataset, fabric dataset, hyperspectral imaging classification and MRI brain images segmentation to demonstrate the superior performance of the proposed clustering algorithm in comparison with the state-of-the-art in the field.

1. Introduction

In the era of big data, data analytics has become increasingly important in almost every domain which will gain benefits from data-driven solutions. Clustering is one of the useful tools for explorative data analysis, which aims to discover the structure among given data. Fuzzy C-means [1] is one of the most popular clustering algorithms, which overcomes the hard partition problem of K-means [2] with soft partition. That is, a given sample can belong to different clusters with different membership degrees. It has been widely used in many fields such as pattern recognition [3], image processing [4,5], and data mining [6].

Since Bezdek's masterwork, many studies have been focused on improving the performance of the conventional FCM algorithm from different perspectives, such as FDCM-SSR [7], GIFP-FCM [8], FCM_S [9], FCM_S1 and FCM_S2 [10], EnFCM [11], FGFCM [12], and Sparse FCM [13]. Furthermore, how to handle the uncertainty problem of fuzzy clustering has received increasing research interests in the past decade. Type-2 Fuzzy Sets (T2 FSs) as one of the representatives have demonstrated their potential in dealing with various uncertainties [14]. For general T2 FSs, their constrained variants, namely the interval-valued fuzzy sets, are typically used in fuzzy clustering due to the simple calculation, leading to the interval-valued fuzzy C-means clustering algorithm [15,16]. Linda and Manic presented a novel uncertain fuzzy clustering algorithm, namely general type-2 fuzzy C-means (GT2_FCM) [17], using the α-planes representation theorem. By introducing the multiple-kernel function and the interval-valued fuzzy sets, Nguyen et al. [18] proposed a multiple kernel interval-valued FCM algorithm.

Recently, intuitionistic fuzzy sets (IFSs) introduced by Atanassov and Rangasamy [19], have shown advantages for describing vague and uncertain data by taking both the membership degree and non-membership degree into account. By introducing IFSs into the conventional FCM algorithm, intuitionistic fuzzy C-means (IFCM) algorithms are able to achieve higher accuracy, better robustness and faster convergence than the traditional FCM. Iakovidis et al. [20] proposed an intuitionistic fuzzy clustering algorithm for computer vision, which adopted a similarity measure over IFSs. Xu and Wu [21] proposed an intuitionistic FCM algorithm which adopted the intuitionistic fuzzy distance instead of fuzzy distance. Combined the concepts of IFSs with the PFCM algorithm, Le et al. [22] proposed a novel clustering algorithm for the applications

* Corresponding author.
E-mail address: ruihuang@cuhk.edu.cn (R. Huang).

https://doi.org/10.1016/j.neucom.2019.01.042
0925-2312/

of geo-demographic analysis. Considering the hesitation degree, Chaira [23] proposed a new IFCM algorithm for medical image analysis, and the performance was better than the conventional FCM and type-2 FCM algorithms. To improve Chaira's IFCM algorithm, Lin [24] proposed IFCM with kernel functions (EKIFCM) to take the advantages of kernels. Verma et al. [25] proposed a novel intuitionistic FCM algorithm for brain image segmentation, which utilizes an intuitionistic fuzzy factor to handle noise and uncertainty information without any parameter training. However, they generally do not take the importance of individual features and the structure of the data into account, when multi-modal and imbalanced features are involved in an application.

Kernel methods [26] are an effectual technique to integrate multi-modal data (such as geometric, textural, images, graphs and vectors), and an effective way for non-linear analysis and modeling of feature spaces. Thus, fuzzy clustering with kernel methods has been widely studied [27–29]. For example, Pedrycz et al. proposed a kernel-based FCM (KFCM) algorithm by projecting the original features into a Hilbert space with a much higher dimension. As a result, the structure of the data can be better discovered with clear separation in a new feature space. Additionally, by considering that all the features are not equally relevant to all the prototypes, a novel weighted fuzzy kernel-clustering algorithm (WKFCM) [30] was proposed by integrating weighted mean into KFCM. As in practical applications data are often acquired from multiple sources, multiple kernels instead of a single kernel can deliver more flexibility on kernel selections and data representation. Huang et al. [31] proposed a multiple kernel FCM (MKFCM) algorithm which utilizes composite kernels instead of a single fixed kernel. Chen et al. [32] adopted the FCM algorithm with multiple kernels for image segmentation, in which different attributes by different kernels are used to produce a linearly composited kernel. By adaptively adjusting the kernel weights, FCM algorithms with multiple kernels are shown to be more robust to ineffective kernels and redundant attributes [33,34]. In addition, rather than using multiple kernels in a clustering task, the ensemble clustering technique aims to combine multiple base clusterings to achieve better and more robust clustering results [35–37].

While intuitionistic FCM and the kernel models have separately demonstrated great potential for clustering complex multi-modal data, there have been few studies jointly taking both advantages into account. Taking image segmentation as an example, we can define different kernels for different features, e.g., one kernel function for the intensity information and another kernel function for the texture information respectively, and apply a combined kernel in intuitionistic FCM to obtain better image segmentation results. Since the intuitionistic fuzzy sets (IFSs) have shown advantages for describing vague and uncertain data by taking both membership degree and non-membership degree into account, the multiple kernel FCM (MKFCM) handle uncertainty with IFSs are able to achieve: (1) higher clustering accuracy; (2) better robustness to noise; (3) faster convergence; and (4) more applicability for complex data sets. Therefore, we propose a novel multi-kernel intuitionistic fuzzy C-means (MKIFCM) clustering algorithm to extend the intuitionistic FCM algorithm with linearly combined kernels. The proposed MKIFCM will possess the advantages of both intuitionistic fuzzy sets and multiple kernel functions in practical clustering problems and provide robust clustering for data acquired from multiple heterogeneous or homogeneous sources.

The remainder of this study is organized as follows. In Section 2, we introduce Atanassov's intuitionistic fuzzy sets and a novel intuitionistic FCM algorithm. We then explain our proposed multi-kernel intuitionistic FCM algorithm, namely MKIFCM, in Section 3. Section 4 presents comprehensive experimental results to demonstrate the effectiveness of our proposed algorithm and its superior performance in comparison with the state-of-the-art algorithms in the field. Finally, Section 5 concludes the paper.

2. Intuitionistic FCM clustering algorithm

As an extension to traditional fuzzy sets, an intuitionistic fuzzy set A in a finite set E can be denoted as

$$A = \{\langle x, u_A(x), \nu_A(x)\rangle | x \in E\} \tag{1}$$

where the functions $u_A(x)$, $\nu_A(x)$: $E \rightarrow [0, 1]$ are the fuzzy membership degree and the non-membership degree of a sample $x \in E$, respectively. For all $x \in E$, the following holds:

$$0 \le u_A(x) + \nu_A(x) \le 1 \tag{2}$$

For all $x \in E$, when $u_A(x) + \nu_A(x) = 1$, the intuitionistic fuzzy set (IFS) is simplified into a traditional fuzzy set.

IFS considers another uncertainty parameter, and the hesitation degree was introducted for parametric fuzzy complement. The hesitation degree $\pi_A(x)$ of a sample $x \in E$ in A is defined as follows:

$$\pi_A(x) = 1 - u_A(x) - \nu_A(x), 0 \le \pi_A(x) \le 1 \tag{3}$$

and Yager's intuitionistic fuzzy complement function is defined as:

$$N(u(x)) = g^{-1}(g(1) - g(u(x))) \tag{4}$$

where $g(\cdot)$ is an increasing function and g: $[0, 1] \rightarrow [0, 1]$.

When $g(x) = x^{\alpha}$, Eq. (4) can be rewritten as:

$$N(u(x)) = (1 - u^{\alpha}(x))^{1/\alpha}, \alpha > 0 \text{ where } N(1) = 0, N(0) = 1, \tag{5}$$

From Yager's intuitionistic fuzzy complement $N(u(x))$, the non-membership values and the hesitation degree are calculated as:

$$\nu_A(x) = (1 - u_A^{\alpha}(x))^{1/\alpha} \tag{6}$$

$$\pi_A(x) = 1 - u_A(x) - (1 - u_A^{\alpha}(x))^{1/\alpha} \tag{7}$$

Thus, IFS becomes:

$$A = \left\{\left\langle x, u_A(x), (1 - u_A^{\alpha}(x))^{1/\alpha}\right\rangle | x \in E\right\} \tag{8}$$

By introducing the intuitionistic fuzzy entropy (IFE) in the traditional fuzzy clustering algorithm, the objective function of the intuitionistic fuzzy clustering algorithm can be written as [23–25]:

$$J_{IFCM}(U, V) = \sum_{i=1}^{C}\sum_{j=1}^{N} {u_{ij}^*}^m d(x_j, \nu_i)^2 + \sum_{i=1}^{C} \pi_i^* e^{1-\pi_i^*} \tag{9}$$

where C is the number of clusters; ν_i is the ith cluster center and V denotes the prototype matrix; $U \in \mathbb{R}^{N\times C}$ is a fuzzy partition matrix whose elements are u_{ij} which denotes the fuzzy membership degree of the jth sample belonging to the ith cluster; $m > 1$ is the fuzzification degree; $\pi_i^* = \frac{1}{N}\sum_{j=1}^{N}\pi_{ij}$ and π_{ij} is the hesitation degree calculated by Eq. (7); u_{ij}^* is the intuitionistic fuzzy membership of the jth sample belonging to the ith cluster and is defined based on the hesitation degree as follows:

$$u_{ij}^* = u_{ij} + \pi_{ij} \tag{10}$$

Similarly, the cluster prototypes can be updated with the parametric fuzzy complement. By substituting Eq. (10) in the conventional FCM algorithm, the cluster prototypes are updated as follows:

$$\nu_i^* = \frac{\sum_{j=1}^{N} u_{ij}^* x_j}{\sum_{j=1}^{N} u_{ij}^*} \tag{11}$$

However, existing intuitionistic fuzzy clustering algorithms often neglect the role of the original fuzzy membership degree when introducing the intuitionistic fuzzy membership degree. The ideal intuitionistic fuzzy clustering objective function should consider not only the difference between the classification samples and the cluster centers, but also the role of the fuzzy membership degree, the non-membership degree and the hesitation degree. Therefore, the fuzzy membership degree and the non-membership degree are combined in the objective function to develop an improved intuitionistic fuzzy clustering algorithm. Moreover, similar to the algorithmsin [23–25], we also introduce the IFE with hesitation degree to represent the uncertainty of each class in the cluster. The objective function of the improved intuitionistic FCM algorithm is then defined as:

$$J_{IIFCM}(U,V)=\sum_{i=1}^{C}\sum_{j=1}^{N}u_{ij}^{m}d_{ij}^{2}+\sum_{i=1}^{C}\sum_{j=1}^{N}(1-\gamma_{ij})^{m}d_{ij}^{2}+\sum_{i=1}^{C}\left(\frac{1}{N}\sum_{j=1}^{N}\pi_{ij}\right)e^{1-\frac{1}{N}\sum_{j=1}^{N}\pi_{ij}}$$
$$s.t.\ u_{ij}\in[0,1],\ \sum_{i=1}^{C}u_{ij}=1,\ u_{ij}+\gamma_{ij}+\pi_{ij}=1,\ 1\le i\le C,\ 1\le j\le N \tag{12}$$

where γ_{ij} is the non-membership values of the jth data belonging to the ith cluster, and π_{ij} is hesitation degree.

Joining two equations $u_{ij}+\gamma_{ij}+\pi_{ij}=1$ and $u_{ij}^{*}=u_{ij}+\pi_{ij}$, we can obtain that: $u_{ij}^{*}=1-\gamma_{ij}$. Then, substituting u_{ij}^{*} and $\pi_i^{*}=\frac{1}{N}\sum_{j=1}^{N}\pi_{ij}$ into Eq. (12), we have,

$$J_{IIFCM}(W,U,V)=\sum_{i=1}^{C}\sum_{j=1}^{N}(u_{ij}^{m}+(u_{ij}^{*})^{m})d_{ij}^{2}+\sum_{i=1}^{C}\pi_i^{*}e^{1-\pi_i^{*}}$$
$$s.t.\ u_{ij}\in[0,1],\ \sum_{i=1}^{C}u_{ij}=1,\ 1\le i\le C,\ 1\le j\le N \tag{13}$$

From Eqs. (6) and (7), we can see that the more different the membership value and the non-membership value are, the lower the hesitation value is. The closer the membership value and the non-membership value are, the higher the hesitation value is. In this case, actual membership value of the jth data belonging to the ith cluster should be within the interval $[\,u_{ij},u_{ij}^{*}]$. It is worth noting from Eq. (13) that our method inherits the advantages of intuitionistic fuzzy clustering and type-2 fuzzy clustering for handling uncertainty.

3. Multi-kernel intuitionistic FCM clustering algorithm

In this paper, we propose a novel multi-kernel intuitionistic fuzzy C-means (MKIFCM) clustering algorithm to extend the intuitionistic FCM algorithm with linearly combined kernels and MKFCM with IFSs. The flowchart is shown in Fig. 1.

In this section, we introduce our proposed MKIFCM algorithm from two aspects: formulating the objective clustering function and optimizing the clustering process.

3.1. Objective clustering function

Set ϕ is a nonlinear mapping function, ϕ: $x\to\phi(x)$, where $x\in\mathbb{R}^p$ is a sample in the original feature space and $\phi(x)\in\mathbb{R}^q$ denotes the mapped high feature space, $p\ll q$. By referring to Eq. (13), the objective function of the kernel-based intuitionistic FCM algorithm is defined as below:

$$J_{KIFCM}(W,U,V)=\sum_{i=1}^{C}\sum_{j=1}^{N}(u_{ij}^{m}+(u_{ij}^{*})^{m})\left\|\phi(x_j)-\phi(v_i)\right\|^{2}+\sum_{i=1}^{C}\pi_i^{*}e^{1-\pi_i^{*}}$$
$$s.t.\ u_{ij}\in[0,1],\ \sum_{i=1}^{C}u_{ij}=1,\ 1\le i\le C,\ 1\le j\le N \tag{14}$$

With kernel function $\kappa(x,y)=\phi(x)^{T}\cdot\phi(y)$, we can have,

$$\begin{aligned}\left\|\phi(x_j)-\phi(v_i)\right\|^{2}&=\phi(x_j)^{T}\phi(x_j)+\phi(v_i)^{T}\phi(v_i)-2\phi(x_j)^{T}\phi(v_i)\\&=\kappa(x_j,x_j)+\kappa(v_i,v_i)-2\kappa(x_j,v_i)\end{aligned}\tag{15}$$

Kernel-based intuitionistic FCM is an effectual method for nonlinear analysis and modeling. However, each kernel function has its own unique characteristics and each feature group vector has its specific structures. To effectively represent complex structures of each feature group vector, our method adopts a linearly combined kernels as similarity measurement function in this paper. Firstly, a special kernel mapping is assigned to each feature group vector as:

$$\begin{aligned}D_{ij,k}^{(\phi_k)}&=\left\|\phi_k(x_{jk})-\phi_k(v_{ik})\right\|^{2}=\phi_k(x_{jk})^{T}\phi_k(x_{jk})+\phi_k(v_{ik})^{T}\phi_k(v_{ik})\\&\quad-2\phi_k(x_{jk})^{T}\phi_k(v_{ik})\\&=\kappa^{(k)}(x_{jk},x_{jk})+\kappa^{(k)}(v_{ik},v_{ik})-2\kappa^{(k)}(x_{jk},v_{ik})\end{aligned}\tag{16}$$

where $D_{ij,k}^{(\phi_k)}$ denotes the distance measurement of the jth sample belonging to the ith cluster for the kth feature group vector with the ϕ_k mapping function, $\kappa^{(k)}$ denotes the kernel function of the kth feature group vector. Secondly, the similarity measurement is defined as the following Eq. (17) to take the advantage of all the features:

$$\begin{aligned}D_{ij}^{2}=D(x_j,v_i)^{2}&=\sum_{k=1}^{K}\omega_{ik}^{\beta}D_{ij,k}^{(\phi_k)}=\sum_{k=1}^{K}\omega_{ik}^{\beta}\left\|\phi_k(x_{jk})-\phi_k(v_{ik})\right\|^{2}\\&=\sum_{k=1}^{K}\omega_{ik}^{\beta}(\kappa^{(k)}(x_{jk},x_{jk})+\kappa^{(k)}(v_{ik},v_{ik})-2\kappa^{(k)}(x_{jk},v_{ik}))\end{aligned}\tag{17}$$

where ω_{ik} is the weight of the kth feature groups vector respective to the ith cluster, $\omega_{ik}\in[0,1]$, $\sum_{k=1}^{K}\omega_{ik}=1$, ; $\beta>1$ is the coefficient as the fuzzification degree m.

Based on the similarity measurement function (Eq. (17)), the objective function of MKIFCM is rewritten as:

$$\begin{aligned}J(W,U,V)&=\sum_{i=1}^{C}\sum_{j=1}^{N}(u_{ij}^{m}+(u_{ij}^{*})^{m})\sum_{k=1}^{K}\omega_{ik}^{\beta}\left\|\phi_k(x_{jk})-\phi_k(v_{ik})\right\|^{2}+\sum_{i=1}^{C}\pi_i^{*}e^{1-\pi_i^{*}}\\&=\sum_{i=1}^{C}\sum_{j=1}^{N}(u_{ij}^{m}+(u_{ij}^{*})^{m})\sum_{k=1}^{K}\omega_{ik}^{\beta}(\kappa^{(k)}(x_{jk},x_{jk})+\kappa^{(k)}(v_{ik},v_{ik})\\&\quad-2\kappa^{(k)}(x_{jk},v_{ik}))+\sum_{i=1}^{C}\pi_i^{*}e^{1-\pi_i^{*}}\end{aligned}$$
$$s.t.\ u_{ij}\in[0,1],\ \omega_k\in[0,1]\ \text{ and }\ \sum_{k=1}^{K}\omega_k=1,$$
$$\sum_{i=1}^{C}u_{ij}=1,\ 1\le j\le N,\ 1\le k\le K \tag{18}$$

If we utilize the Gaussian kernel function, $\kappa(x,x)=1$. Thus, Eq. (18) can be simplified as below:

$$J(W,U,V)=\sum_{i=1}^{C}\sum_{j=1}^{N}(u_{ij}^{m}/2+(u_{ij}^{*})^{m}/2)\sum_{k=1}^{K}\omega_{ik}^{\beta}(2-2\kappa^{(k)}(x_{jk},v_{ik}))$$

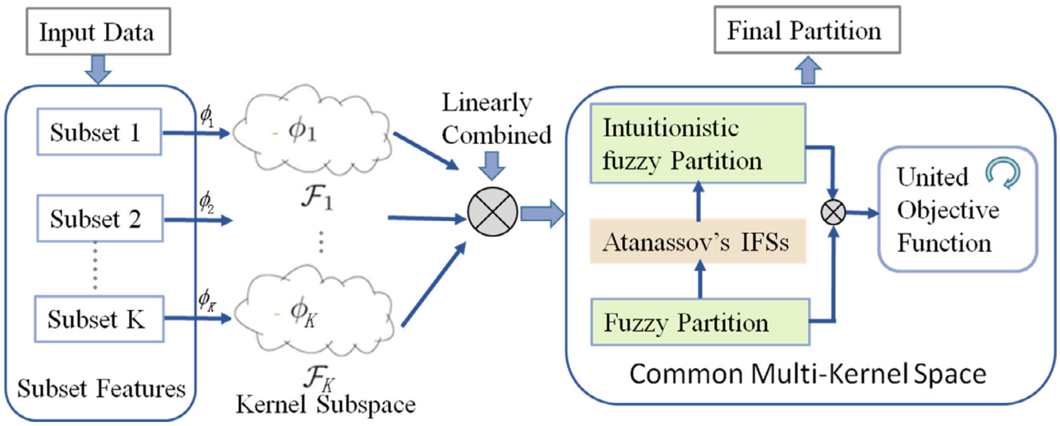

Fig. 1. Schematic framework of our proposed model.

$$
\begin{aligned}
&+\sum_{i=1}^{C}\pi_i^* e^{1-\pi_i^*}\\
&=\sum_{i=1}^{C}\sum_{j=1}^{N}(u_{ij}^m+(u_{ij}^*)^m)\sum_{k=1}^{K}\omega_{ik}^{\beta}(1-\kappa^{(k)}(x_{jk},v_{ik}))\\
&+\sum_{i=1}^{C}\pi_i^* e^{1-\pi_i^*}
\end{aligned}
\tag{19}
$$

where $\kappa^{(k)}(x_{jk},v_{ik})=\exp(-\|x_{jk}-v_{ik}\|^2/\sigma_k^2)$.

3.2. Clustering optimization

To minimize the objective function Eq. (19), we need to simultaneously search for the optimal partition matrix U, cluster prototypes V, and weight vector W.

3.2.1. Partition matrix optimization

The distance between sample x_j and cluster center v_i in the multiple kernel space is defined as $D_{ij}^2=\sum_{k=1}^{K}\omega_{ik}^{\beta}{D_{ij,k}^{(\phi_k)}}^2=\sum_{k=1}^{K}\omega_{ik}^{\beta}\|\phi_k(x_{jk})-\phi_k(v_{ik})\|^2$. When V and W are fixed, the distances can be obtained. Considering the value of u_{ij}^* is obtained with $u_{ij}^*=u_{ij}+\pi_{ij}$ in the IFCM, we only need to calculate u_{ij} in Eq. (19) and then obtain u_{ij}^* through u_{ij}. By using the constraint $\sum_{i=1}^{C}u_{ij}=1$ and the Lagrangian optimization, the minimization of $J(U, V, W)$ in Eq. (19) is equivalent to the optimization of u_{ij} as below:

$$
\begin{aligned}
J(u_{ij},\alpha_j)=&\sum_{i=1}^{C}\sum_{j=1}^{N}(u_{ij}^m+(u_{ij}^*)^m)D_{ij}^2-\sum_{j=1}^{N}\alpha_j\left(\sum_{i=1}^{C}u_{ij}-1\right)\\
&+\sum_{i=1}^{C}\pi_i^* e^{1-\pi_i^*}
\end{aligned}
\tag{20}
$$

Therefore, the optimal value of u_{ij} can be obtained by setting $\partial J(u_{ij},\alpha_j)/\partial u_{ij}=0$ and $\partial J(u_{ij},\alpha_j)/\partial \alpha_j=0$. Thus, we have the following equations

$$
\begin{cases}
\frac{\partial J(u_{ij},\alpha_j)}{\partial u_{ij}}=mu_{ij}^{m-1}D_{ij}^2-\alpha_j=0\\
\frac{\partial J(u_{ij},\alpha_j)}{\partial \alpha_j}=\sum_{i=1}^{C}u_{ij}-1=0
\end{cases}
\tag{21}
$$

Then u_{ij} is $u_{ij}=\{{}^{\alpha_j}/{}_{mD_{ij}^2}\}^{{}^{1}/m-1}$. Under the constraint $\sum_{i=1}^{C}u_{ij}=1,\alpha_j$ can be eliminated and the closed-form representation of u_{ij} can be obtained as:

$$
u_{ij}=1\Big/\sum_{h=1}^{c}\left[\frac{D_{ij}^2}{D_{hj}^2}\right]^{1/(m-1)}
\tag{22}
$$

Substitute Eq. (17) into Eq. (22), we obtain the iterative solution for u_{ij} as:

$$
u_{ij}=1\Big/\sum_{h=1}^{c}\left[\frac{\sum_{k=1}^{K}\omega_{ik}^{\beta}(1-\kappa^{(k)}(x_{jk},v_{ik}))}{\sum_{k=1}^{K}\omega_{hk}^{\beta}(1-\kappa^{(k)}(x_{jk},v_{hk}))}\right]^{1/(m-1)}
\tag{23}
$$

From Eqs. (6) and (7), we have,

$$
\begin{aligned}
\gamma_{ij}&=(1-u_{ij}^{\alpha})^{1/\alpha}\\
&=\left(1-1\Big/\left(\sum_{h=1}^{c}\left[\frac{\sum_{k=1}^{K}\omega_{ik}^{\beta}(1-\kappa^{(k)}(x_{jk},v_{ik}))}{\sum_{k=1}^{K}\omega_{hk}^{\beta}(1-\kappa^{(k)}(x_{jk},v_{hk}))}\right]^{1/(m-1)}\right)^{\alpha}\right)^{\frac{1}{\alpha}}
\end{aligned}
\tag{24}
$$

$$
\begin{aligned}
&u_{ij}^*\\
&=1-\left(1-1\Big/\left(\sum_{h=1}^{c}\left[\frac{\sum_{k=1}^{K}\omega_{ik}^{\beta}(1-\kappa^{(k)}(x_{jk},v_{ik}))}{\sum_{k=1}^{K}\omega_{hk}^{\beta}(1-\kappa^{(k)}(x_{jk},v_{hk}))}\right]^{1/(m-1)}\right)^{\alpha}\right)^{\frac{1}{\alpha}}
\end{aligned}
\tag{25}
$$

3.2.2. Cluster prototype optimization

For the kth feature groups vector, suppose that the fuzzy memberships $U=[u_{ij}]_{C\times N}$ and weights $W=[\omega_{ik}]_{C\times K}$ are fixed, the optimal prototypes can be derived. By substituting Eq. (25) in the traditional FCM method, the updated cluster prototypes are written as:

$$
v_{ik}=\frac{\sum_{j=1}^{N}(u_{ij}^m+u_{ij}^*)\kappa^{(k)}(x_{jk},v_{ik})\cdot x_{jk}}{\sum_{j=1}^{N}(u_{ij}^m+u_{ij}^*)\kappa^{(k)}(x_{jk},v_{ik})}
\tag{26}
$$

These cluster prototypes can be directly computed when Gaussian kernel functions are adopted, which is different from the case that cluster prototypes are in the implicit kernel-induced feature space using other kernel function.

3.2.3. Weight optimization

For the kth feature groups vector, suppose that U and D_{ij}^2 are fixed, the minimum of $J(U, V, W)$ in Eq. (19) can be solved with La-

Algorithm 1 Multi-kernel Intuitionistic fuzzy C-means clustering (MKIFCM).

Input: Given a set of N data points $X=\{x_i\}_{i=1}^{N}$in P dimension space, a base set of kernel functions $\{\kappa_k\}_{k=1}^{K}$, the exponent β, the number of clusters C, the fuzzy index m, the termination criterion ζ and T, and initialized partition matrix $U^{(0)}=\{u_{ij}\}_{i,j=1}^{C,N}$ and weights $\{\omega_{ik}\}_{i,k=1}^{C,K}$.
Output: The partition matrix $U=\{u_{ij}\}_{i,j=1}^{C,N}$ and the weights $\{\omega_{ik}\}_{i,k=1}^{C,K}$for each class.
1. Select suitable Gaussian kernel functions for each feature groups vector; and calculate the initial clustering prototypes by Eq. (11).
2. **Procedure** MKIFCM (Data X, Number C, combined-kernel $\{\kappa_{com}^{(k)}\}_{k=1}^{K}$)
3. **Repeat**
4. Calculate $\gamma_{ij}=(1-u_{ij}^{\alpha})^{1/\alpha}$ and $u_{ij}^{*}=1-\gamma_{ij}$ by Eqs. (24) and (25)
5. Calculate distances D_{ij}^2 by Eq. (17)
6. Update weights $\{w_k\}_{k=1}^{K}$by Eq. (32)
7. Update partition matrix $U=\{u_{ij}\}_{i,j=1}^{C,N}$ by Eq. (23)
8. **End for**
9. **Until** $|J^{(t)}-J^{(t-1)}|\le\zeta$ or the number of iterations $t>T$
10. **Return** $U=\{u_{ij}\}_{i,j=1}^{C,N}$ and $\{w_k\}_{k=1}^{K}$
11. **End procedure**

grangian optimization by solving the following optimization problem on ω_{ik}:

$$J(\omega_{ik},\eta_i)=\sum_{i=1}^{C}\sum_{j=1}^{N}\sum_{k=1}^{K}(u_{ij}^m+(u_{ij}^{*})^m)\omega_{ik}^{\beta}(1-\kappa^{(k)}(x_{jk},v_{ik}))+\sum_{i=1}^{C}\pi_i^{*}e^{1-\pi_i^{*}}-\sum_{i=1}^{C}\eta_i\left(\sum_{k=1}^{K}\omega_{ik}-1\right) \tag{27}$$

Then the optimal value of w_{ik} can be obtained by setting $\partial J(w_{ik},\eta_i)/\partial w_{ik}=0$ and $\partial J(w_{ik},\eta_i)/\partial\eta_i=0$. As a result, we have the following equations:

$$\frac{\partial J(\omega_{ik},\eta_i)}{\partial\omega_{ik}}=\beta\sum_{j=1}^{N}(u_{ij}^m+(u_{ij}^{*})^m)\omega_{ik}^{\beta-1}(1-\kappa^{(k)}(x_{jk},v_{ik}))-\eta_i=0 \tag{28}$$

$$\frac{\partial J(\omega_{ik},\eta_i)}{\partial\eta_i}=\sum_{k=1}^{K}\omega_{ik}-1=0 \tag{29}$$

Then ω_{ik} is obtained as below:

$$\omega_{ik}=\left\{\eta_i\Big/\beta\sum_{j=1}^{N}(u_{ij}^m+(u_{ij}^{*})^m)(1-\kappa^{(k)}(x_{jk},v_{ik}))\right\}^{1/\beta-1} \tag{30}$$

With the constraint $\sum_{k=1}^{K}\omega_{ik}=1$, we have

$$\eta_i^{1/\beta-1}=\frac{1}{\sum_{k=1}^{K}\left(\frac{1}{\beta\sum_{j=1}^{N}(u_{ij}^m+(u_{ij}^{*})^m)(1-\kappa^{(k)}(x_{jk},v_{ik}))}\right)^{1/\beta-1}} \tag{31}$$

Finally, η_i can be eliminated and the closed-form expresentation of the optimal weight for the kth feature groups vector as

$$\omega_{ik}=\frac{1}{\sum_{t=1}^{K}\left(\frac{\sum_{j=1}^{N}(u_{ij}^m+(u_{ij}^{*})^m)(1-\kappa^{(k)}(x_{jk},v_{ik}))}{\sum_{j=1}^{N}(u_{ij}^m+(u_{ij}^{*})^m)(1-\kappa^{(t)}(x_{jt},v_{it}))}\right)^{1/\beta-1}} \tag{32}$$

Our proposed clustering method is summarized in Algorithm 1.

4. Experimental results and discussions

We conducted comprehensive experiments to evaluate the performance of our proposed clustering algorithm. In particular, our algorithm was compared with five representative fuzzy clustering algorithms: Conventional Fuzzy C-means Clustering (FCM) [1], Intuitionistic FCM (IFCM) [23], Weighted Kernel FCM (WKFCM) [30], Multiple-kernel FCM (MKFCM) [31] and Intuitionistic FCM based on kernel (EKIFCM) [24]. The comparison was conducted on various datasets in the following five experiments: (1) six datasets from the UCI machine learning repository [38], (2) Brodatz texture images [39], (3) the fabric dataset [40], (4) the Indian Pines hyperspectral image database [41] and (5) MRI brain images segmentation [42].

4.1. Evaluation metrics

To evaluate the performance of the clustering methods for comparison, we adopt seven performance metrics: classification rate (CR) [29], the best CR, classification coefficient [43], classification entropy [43], Davies–Bouldin [44], Dun [44], and Iterations Counts (IC).

Classification rate (CR) has been widely used to measure the performance of clustering algorithms. It is defined as [29]:

$$CR=\frac{\max\{\sum_{i=1}^{N}\delta(l_i,map(c_i))\}}{N} \tag{33}$$

where N is the total number of samples, l_i and c_i are the original label and the clustering result of the ith sample, respectively, $map(\cdot)$ denotes a permutation function, and $\delta(a,b)$ is the number of samples that are in both clusters a and b. Higher CR indicates better clustering performance.

Classification coefficient U_{pc} and classification entropy U_{pe} [43] are to evaluate the classification precision of a partition matrix. They are defined as:

$$U_{pc}=\frac{\sum_{j=1}^{N}\sum_{i=1}^{c}\mu_{ij}}{N} \tag{34}$$

$$U_{pe}=\frac{-\sum_{j=1}^{N}\sum_{i=1}^{c}\left(\mu_{ij}logu_{ij}\right)}{N} \tag{35}$$

The larger the classification coefficient or the smaller the classification entropy is, the better the partition result is.

The DB (Davies–Bouldin) [44] indicator is the ratio between intra-class compactness and inter-class dispersibility. It is defined as:

$$DB=\frac{1}{c}\sum_{i=1}^{c}\max_{i\ne j}\left(\frac{\sigma_i+\sigma_j}{d(p_i,\ p_j)}\right) \tag{36}$$

where σ_i represents the average distance from the cluster prototype p_i to all the samples in the ith class, $d(p_i,p_j)$ is the distance from the cluster prototype p_i to the cluster prototype p_j.

D (Dunn) [44] is the ratio between the shortest inter-class distance and the longest intra-class distance, and defined as:

$$D=\min_{1\le i\le c}\left\{\min_{1\le i\le c,i\ne j}\left[\frac{d(p_i,p_j)}{\max_{1\le k\le n}\Delta_{ik}}\right]\right\} \tag{37}$$

Table 1
Description of the six UCI datasets and features.

Dataset	Feature Group	Attributes in each feature group	Dimension	Cluster	Size	
Glass Identification dataset (GID)	Feature 1	Refractive index	1	9	6	214
	Feature 2	Sodium, Magnesium, Aluminum, Silicon, Potassium, Calcium, Barium, Iron	8			
IRIS	Each attribute as one feature	Sepal length, Sepal width, Petal length, Petal width	4		3	150
Wine	Each attribute as one feature	Alcohol, Malic acid, Ash, Alcalinity of ash, Magnesium, Total phenols, Flavanoids, Nonflavanoid phenols, Proanthocyanins, Color intensity, Hue, OD280/OD315 of diluted wines, Proline	13		3	178
Wisconsin Prognostic Breast Cancer (WPBC)	Each attribute as one feature	Clump thickness, Uniformity of cell size, Uniformity of cell shape, Marginal adhesion, Single epithelial cell size, Bare nuclei, Bland chromatin, Normal nucleoli, Mitoses	9		2	699
Multiple features (MF)	Mfeat-fou	76 Fourier coefficients of the character shapes	76	649	10	2000
	Mfeat-fac	216 profile correlations	216			
	Mfeat-kar	64 Karhunen–Love correlations	64			
	Mfeat-pix	240 pixel averages in 2×3 windows	240			
	Mfeat-zer	47 Zernike moments	47			
	Mfeat-mor	6 morphological variables	6			
Image segmentation (IS)	Shape	9 features for the shape information of 7 images	9	19	7	2310
	RGB	10 features for the RGB values of 7 images	10			

where $\max_{1\le k\le n} \Delta_{ik}$ represents the maximum distance from the cluster prototype to the samples in its class, and $d(p_i + p_j)$ is the distance from the cluster prototype p_i to the cluster prototype p_j.

The smaller the DB value or the larger the D value, indicating that the higher compactness within the same cluster while the greater separation between the different classes, and the better the clustering quality.

Iterations Counts (IC) represents the number of iterations. Under the same termination criterion, the fewer the iterations for clustering, the less the time required in average for the algorithm.

The validity index based on ratio of within-class compactness and between-class separation is robust to data dimension and noise. However, these metrics are not suitable for image data. The validity index based on membership degree can be applied to datasets with low dimension, as membership degree has a decreasing trend with increasing number of classes. When samples have label information, CR is the best evaluation metric. Thus, CR, U_{pc}, U_{pe}, DB, D and IC are adopted in Experiment 1, and only CR is adopted in other experiments. In all examples, the average results of ten runs of each method are used to evaluate the classification performance.

In this paper, we use a Gaussian kernel for each feature groups vector. For the kth feature groups vector, to choose scale parameter σ_k, let the minimal value allowed for the Gaussian kernel over the data set be δ. We then obtain the corresponding σ_k as

$$\sigma_k = \min_{i,j}(-\|x_{ik} - x_{jk}\|^2 / \log(\delta)) \tag{38}$$

4.2. Experiment 1: UCI machine learning datasets

This experiment was conducted on six datasets of the UCI repository [38] which has been widely used in machine learning related areas, including the Glass Identification dataset (GID), IRIS, Wine, Wisconsin Prognostic Breast Cancer (WPBC), Multiple Features (MF) and Image Segmentation (IS) dataset. The Glass Identification database dataset has a larger number of samples, a lower standard deviation of features, and a larger number of classes. The IRIS dataset is a typical dataset in classification domain. The Wine dataset has a larger number of features. The Wisconsin Prognostic Breast Cancer is medical data and the types of features are numerical and binary. The Multiple Features dataset has the largest number of classes and attributes. The Image Segmentation dataset has the largest number of samples. Table 1 shows the details of the feature group vectors.

As shown in Table 2, our MKIFCM outperforms all the other algorithms in terms of CR, except achieving comparable results to FCM on the WPBC dataset. Note that WPBC has simple struction, which may not be suitable for demonstrating the superiority of our algorithm. Therefore, our method is more suitable for real and complicated datasets (e.g., MF) with a large number of attributes, classes and instances. It should be noted that the value of fuzzification degree m has a great influence on the performance of various fuzzy clustering algorithms. In this paper, the value range of the fuzzification degree m is set to (1, 5], and the value of fuzzification degree m when each algorithm achieves optimal performance is shown in Table 2.

For the GID dataset, although there is no significant difference between MKIFCM and other three algorithms (i.e., FCM, IFCM and EKIFCM) in terms of average CR, the maximum CR of our method outperforms these three algorithms. Furthermore, MKIFCM has more superior performance than the WKFCM and MKFCM. For the IRIS dataset, MKIFCM outperforms all the other algorithms, and MKFCM and EKIFCM are superior to the other three algorithms in term of average CR and the maximum CR. For the MF dataset, there are significant differences in the data values for each dimension. The mean variance of some important features is discarded due to small values. Thus, the dataset is normalized to reduce the impact of each dimension. MKIFCM achieves the best CR 92.96% among all the algorithms. A similar pattern can be observed in the Wine dataset. For the IS dataset, MKIFCM and EKIFCM demonstrate stronger stability than the other algorithms (Fig. 1).

The WPBC dataset is medical data which contain both numerical and binary values. We can see that the conventional clustering methods (FCM and IFCM) can achieve a better CR than the kernel-based fuzzy clustering methods. We believe the reasons could be: first, the data structure of this dataset is relatively simple and Euclidean distance was well qualified; second, we treated each attribute as a feature group, which may lead to multiple parameter selection for building the kernel function. Moreover, by combining multiple kernels, Zadeh's fuzzy set and intuitionistic fuzzy set mechanism, our algorithm is able to obtain

Table 2
Comparison with different algorithms on the UCI repository.

Title	Method	Parameters	CR	The best CR	U_{pc}	U_{pe}	DB	D	IC
GID ($c=6$)	FCM	$m=1.08$	0.5339 ± 0.0038	0.5421	**0.9702**	**0.0190**	0.6022	0.2822	36
	IFCM	$m=1.08$, $\alpha=2$	0.5383 ± 0.0072	0.5511	0.9673	0.0282	**0.0038**	**10.70**	**21**
	WKFCM	$m=2$	0.3598 ± 0.0049	0.4579	0.7006	0.2376	0.2470	0.0096	68
	MKFCM	$m=2.85$	0.3794 ± 0.0028	0.4439	0.6950	0.2475	0.4315	0.0089	50
	EKIFCM	$m=2$,$\alpha=2$	0.5514 ± 0	0.5514	0.9680	0.0273	0.0039	10.68	69
	MKIFCM	$m=2.85$, $\alpha=1.25$	**0.5654** ± 0.0005	**0.5657**	0.0841	1.7906	0.0260	0.0000	**21**
Iris ($c=3$)	FCM	$m=1.08$	0.8933 ± 0	0.8933	**0.9830**	**0.0112**	0.0058	13.64	14
	IFCM	$m=2$, $\alpha=2$	0.8810 ± 0.0072	0.9000	0.5812	0.3872	0.0001	97.75	47
	WKFCM	$m=2$	0.8657 ± 0.0103	0.9267	0.4310	0.6080	**0.0000**	**148.7**	**8**
	MKFCM	$m=2$	0.9343 ± 0.0030	0.9467	0.6612	0.3070	0.0003	28.06	84
	EKIFCM	$m=2$, $\alpha=2$	0.9600 ± 0	0.9600	0.6459	0.3215	0.0011	4.824	76
	MKIFCM	$m=2$, $\alpha=0.85$	**0.9667** ± 0.0008	**0.9669**	0.6099	0.3588	0.0018	3.978	50
Wine ($c=3$)	FCM	$m=1.08$	0.6899 ± 0.0014	0.7022	**0.9842**	**0.0101**	0.0059	11.84	**16**
	IFCM	$m=1.08$, $\alpha=2$	0.7022 ± 0	0.7022	0.9828	0.0114	0.0059	12.10	32
	WKFCM	$m=2$	0.6517 ± 0	0.6517	0.9658	0.0238	0.1224	0.2470	42
	MKFCM	$m=1.08$	0.6573 ± 0	0.6573	0.9653	0.0246	0.1338	0.2482	43
	EKIFCM	$m=2.7$, $\alpha=14$	0.7581 ± 0	0.7581	0.3593	0.7661	**0.0000**	**600.7**	154
	MKIFCM	$m=1.5$, $\alpha=0.85$	**0.9270** ± 0	**0.9270**	0.3547	0.7354	0.0016	3.056	61
WPBC ($c=2$)	FCM	$m=2$	**0.9528** ± 0	**0.9528**	0.6439	0.2670	0.0002	16.03	**12**
	IFCM	$m=2$, $\alpha=2$	0.9485 ± 0	0.9485	0.6636	0.2515	**0.0001**	25.02	25
	WKFCM	$m=2$	0.7897 ± 0	0.7897	0.8696	0.0752	**0.0001**	**31.99**	91
	MKFCM	$m=1.08$	0.7854 ± 0	0.7854	**0.9892**	**0.0079**	0.1580	0.4675	53
	EKIFCM	$m=4.5$, $\alpha=1$	0.9071 ± 0	0.9071	0.6510	0.2633	**0.0001**	25.11	13
	MKIFCM	$m=2$, $\alpha=0.85$	0.9328 ± 0	0.9328	0.5247	0.3937	0.0002	0.1870	142
MF ($c=10$)	FCM	$m=1.08$	0.8020 ± 0.0280	0.8930	0.8328	0.1564	0.0006	0.9190	54
	IFCM	$m=1.08$, $\alpha=2$	0.7911 ± 0.0311	0.8755	0.8385	0.1555	**0.0005**	**1.0900**	135
	WKFCM	$m=1.08$	0.8035 ± 0.0263	0.8855	0.5984	0.4948	0.0012	0.3940	300
	MKFCM	$m=1.28$	0.8481 ± 0.0055	0.9036	0.4784	0.6842	0.0013	0.2940	300
	EKIFCM	$m=1.28$, $\alpha=2$	0.8970 ± 0	0.8970	0.8310	0.1654	**0.0005**	**1.0900**	**28**
	MKIFCM	$m=1.28$, $\alpha=2$	**0.9295** ± 0.0004	**0.9296**	0.5878	0.5065	0.0007	0.8420	40
IS ($c=7$)	FCM	$m=1.08$	0.6370 ± 0.0228	0.6667	**0.9384**	**0.0483**	0.0009	7.6600	58
	IFCM	$m=1.08$, $\alpha=2$	0.6483 ± 0.0142	0.6645	0.9369	0.0504	0.0009	8.6000	37
	WKFCM	$m=1.08$	0.6503 ± 0.0074	0.6645	0.9366	0.0506	0.0009	8.5700	**12**
	MKFCM	$m=1.08$	0.6359 ± 0.0043	0.6455	0.8523	0.1471	**0**	11.200	300
	EKIFCM	$m=1.85$, $\alpha=2$	0.6571 ± 0	0.6571	0.3839	0.9678	0.0001	2.6400	124
	MKIFCM	$m=1.28$, $\alpha=9$	**0.6693** ± 0.0011	**0.6697**	0.4363	0.8900	**0**	**13.900**	99

Table 3
Description of the Brodatz texture image dataset.

Dataset	Feature groups	Feature groups vector generation by Gabor filter	Dimension	Cluster	Size
Brodatz texture	1	10 features generated with five orientations and two scales starting from 0.4	10	7	4096
	2	15 features generated with five orientations and three scales starting from 0.5	15		
	3	30 features generated with six orientations and five scales starting from 0.6	30		
	4	40 features generated with eight orientations and five scales starting from 0.25	40		

Table 4
The CR of all algorithms in texture images.

Texture	FCM	IFCM	WKFCM	MKFCM	EKIFCM	MKIFCM
CR	0.4541 ± 0.0035	0.6809 ± 0.0045	0.5566 ± 0.0125	0.6894 ± 0.0017	0.6924 ± 0	**0.8057** ± 0.0011
The best CR	0.4558	0.6868	0.6832	0.6898	0.6924	**0.8063**

better clustering performance on CR. It strongly indicates that our algorithm gains benefits from the advantages of both MKFCM and IFCM for clustering complicated datasets.

4.3. Experiment 2: Brodatz texture images

In this experiment, the Brodatz texture dataset was used for comparing FCM, IFCM, WKFCM, MKFCM, EKIFCM and the proposed MKIFCM algorithms. The Brodatz texture dataset [39] consists of seven texture images of size 128 × 128. To extract different texture features, we utilized the Gabor filter [45] in this experiment. The detailed information of each grouped feature is listed in Table 3.

As shown in Table 4, the CR of MKIFCM reaches 80.57%, which is the best compared to all the algorithms that generally have CR below 70%. The clustering results of a sample image with different algorithms are shown in Fig. 2.

For image datasets, different methods can be used to extract features, and the dimension of features is higher than the previous experiments. The updating mechanism of FCM would easily be influenced by a distance measurement which directly adopts Euclidean distance, thus, the classification results is the worst. Although also adopting Euclidean distance, IFCM introduces hesitation degrees to maximize the inter-class differences and obtains better clustering performance. WKFCM adopts single kernel function, which is not able to handle complex datasets well. Al-

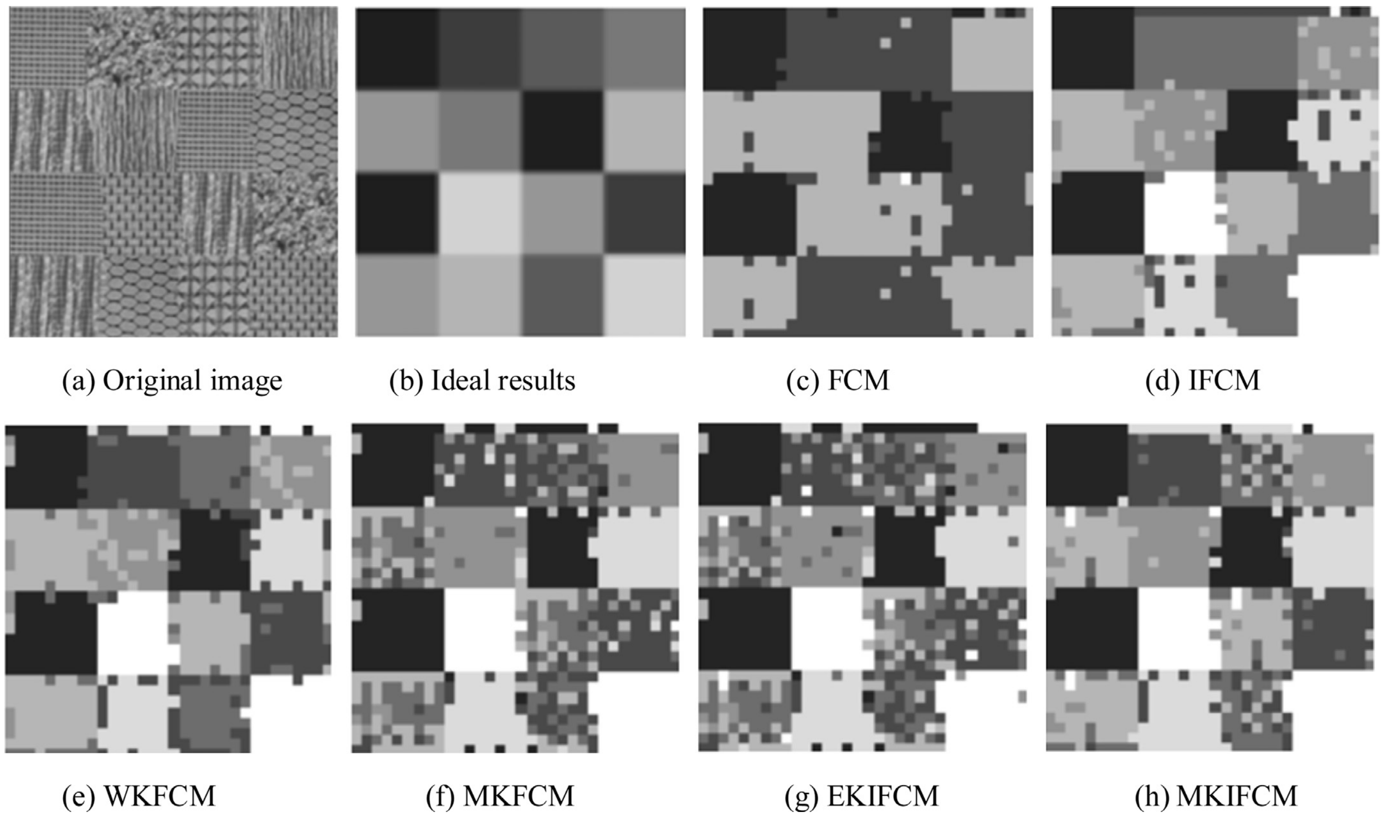

Fig. 2. The clustering results of a sample image with different algorithms on the Brodatz texture dataset.

though EKIFCM adopts single kernel function, EKIFCM is able to achieve better performance than WKFCM because of the intuitionistic fuzzy set and the GA mechanism. MKFCM maps each feature group into each kernel space and linearly combines these kernels to build a composite kernel, and can handle the clustering problem of complex datasets well. Thus, MKFCM obtains better clustering performance than FCM and WKFCM. Moreover, MKIFCM inherits the excellent performance of Atanassov's intuitionistic fuzzy sets and multi-kernel method which can better handle uncertainty and complex clustering problems than a single method such as FCM, IFCM, WKFCM, MKFCM and EKIFCM. Similar results are also observed in Experiments 3 and 4.

4.4. Experiment 3: fabric defect imaging dataset

This experiment was conducted on fabric defect images with plain weave and pattern. Those images were obtained from the TILDA fabric image library of the University of Freiburg [40]. In the experiment, the fabric defect images were normalized to a size of 256×256 as input. Each test image was evenly divided into 16×16 image blocks. The features of fabric defect images were extracted with Gabor filters (48 features generation with eight orientation and six scales starting from 0.6). As fabric defect is generally in the direction of the latitude and longitude lines, we treated the each orientation's feature as a Feature group. The existing visual saliency maps of plain weave and patterned fabric defect images were compared with the experimental renderings of different algorithms.

As shown in Fig. 3, FCM, IFCM and WKFCM completely missed the fabric defects and the performance of MKFCM is better than that of EKIFCM. This indicates that multiple kernel mechanism is better than a single kernel mechanism for multi-view data. Moreover, as shown in Fig. 3, the fabric defects identified by our algorithm were accurately segmented, and the defect areas were marked effectively.

4.5. Experiment 4: Indian Pines hyperspectral image

This experiment was conducted on the hyperspectral image acquired by AVIRIS imaging spectrometer from the northwest district of Indian Pines in June 1992 [41]. The total number of bands used is 220 except the 1st, 33rd, 97th and 161st bands for testing. The whole scene with spatial resolution 145×145 pixels contains 16 classes of objects of interest of which the size is between 20 and 2468 pixels. Note that only 4 classes of objects numbered 2, 5, 10, and 11 were selected in this experiment due to the number of instances is too large. Fig. 4(a) and (b) shows the pseudo color image of bands 24, 12 and 5, and the ground-truth map, respectively. Each different color represents a class in Fig. 4(b).

As shown in Table 5, the CRs of FCM and IFCM are both 37.49%. This indicates that there is no significant difference between FCM and IFCM for clustering high-dimensional data. By introducing single kernel, the CRs of WKFCM and EKIFCM are better than those of FCM and IFCM. Moreover, the fuzzy clustering algorithms with multiple-kernel, MKFCM and MKIFCM perform clearly better than the other algorithms. It is observed that the maximum CR of our proposed algorithm is the best with 61.23%. Fig. 5 shows the clustering results of six representative algorithms including FCM, IFCM, WKFCM, MKFCM, EKIFCM and MKIFCM.

4.6. Experiment 5: MRI brain images segmention

In this experiment, MRI brain images from http://brainweb.bic.mni.mcgill.ca/brainweb. The data used in the experiment is a real

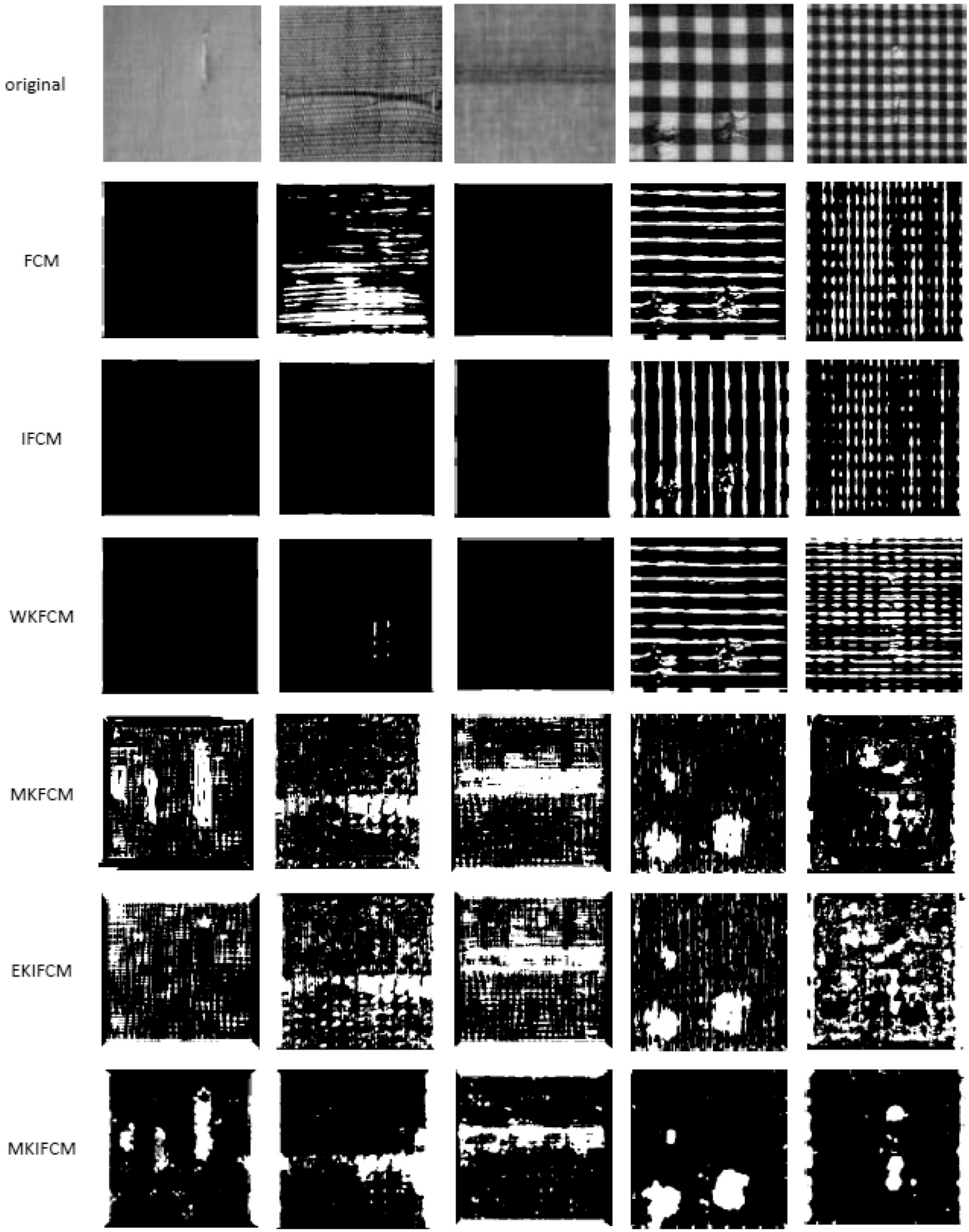

Fig. 3. Comparison with visual saliency maps on fabric defect images.

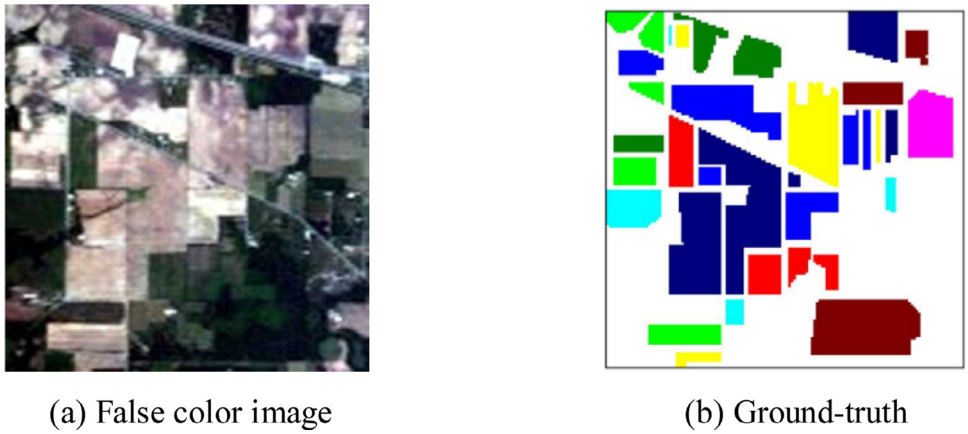

Fig. 4. Illustration of the Indian Pines hyperspectral image. (a) Pseudo-color image of bands 24, 12, and 5. (b) The ground-truth map of the objects of interest.

Table 5
The CR of different algorithms.

	FCM	IFCM	WKFCM	MKFCM	EKIFCM	MKIFCM
CR	0.3749 ± 0.0012	0.3749 ± 0.0009	0.4535 ± 0.0058	0.5844 ± 0.0015	0.4805 ± 0	**0.6123** ± 0.0007
The best CR	0.3751	0.3753	0.4736	0.5853	0.4805	**0.6128**

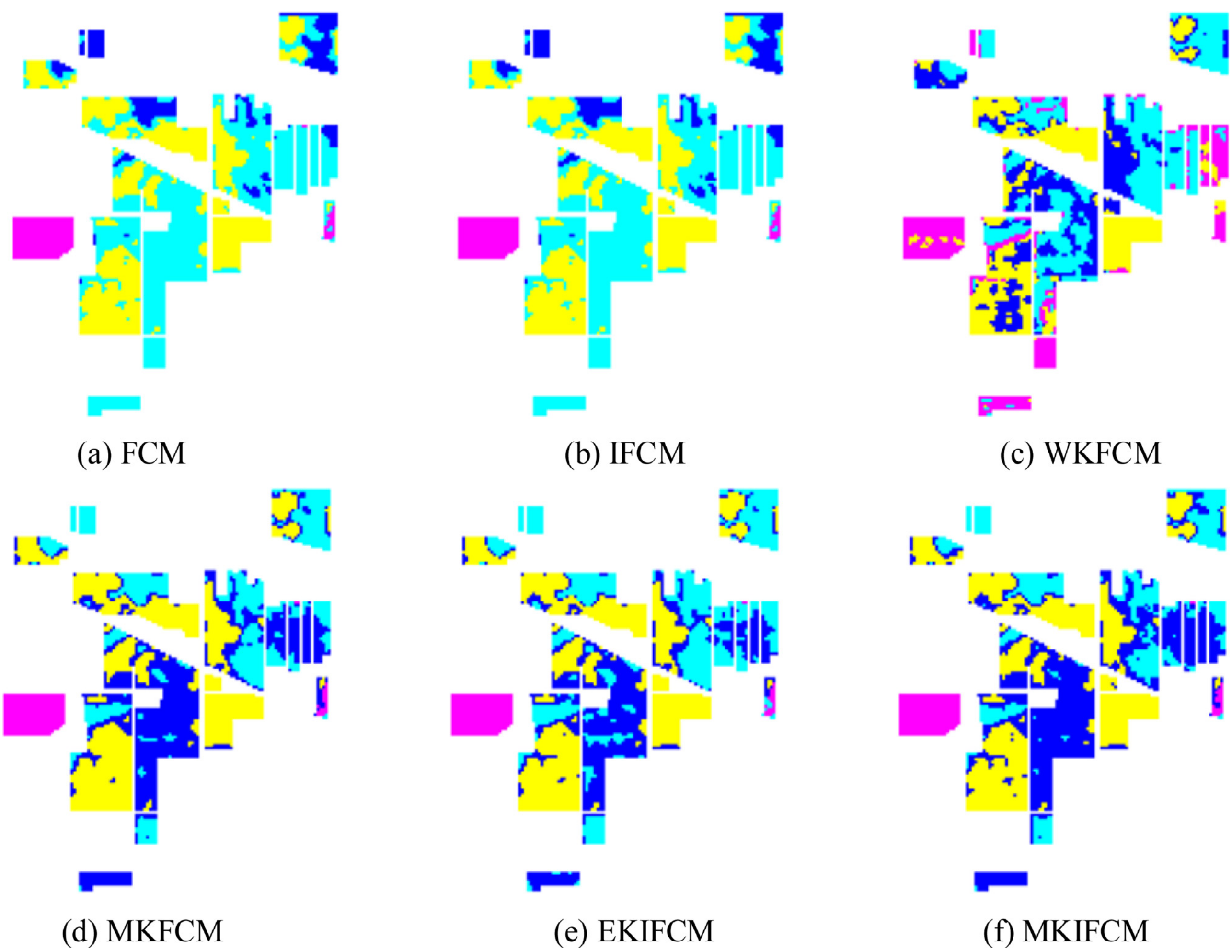

Fig. 5. Clustering results of the hyperspectral image with different algorithms.

Table 6
The DC and SA of different algorithms.

			FCM	IFCM	WKFCM	MKFCM	EKIFCM	MKIFCM
1%Noise 0%INU	DC	CSF	0.9530	0.9533	0.9514	0.9714	0.9671	**0.9732**
		GW	0.9640	0.9671	0.9675	0.9788	0.9751	**0.9800**
		WM	0.9868	0.9882	0.9883	0.9922	0.9909	**0.9925**
	SA		0.9745	0.9762	0.9763	0.9848	0.9823	**0.9857**
1%Noise 20%INU	DC	CSF	0.9475	0.9477	0.9458	0.9660	0.9613	**0.9697**
		GW	0.9522	0.9535	0.9515	0.9648	0.9617	**0.9671**
		WM	0.9798	0.9799	0.9778	0.9837	0.9828	**0.9841**
	SA		0.9659	0.9660	0.9643	0.9747	0.9725	**0.9762**
1%Noise 40%INU	DC	CSF	0.9366	0.9395	0.9381	0.9559	0.9457	**0.9596**
		GW	0.9250	0.9280	0.9251	0.9383	0.9332	**0.9420**
		WM	0.9630	0.9635	0.9592	0.9667	0.9661	**0.9681**
	SA		0.9460	0.9466	0.9439	0.9550	0.9515	**0.9575**
3%Noise 0%INU	DC	CSF	0.9499	0.9501	0.9494	0.9600	0.9569	**0.9620**
		GW	0.9555	0.9582	0.9591	0.9607	0.9592	**0.9628**
		WM	0.9819	0.9828	0.9826	0.9831	0.9831	**0.9833**
	SA		0.9686	0.9699	0.9703	0.9720	0.9711	**0.9733**
3%Noise 20%INU	DC	CSF	0.9460	0.9473	0.9468	0.9597	0.9537	**0.9614**
		GW	0.9480	0.9493	0.9489	0.9541	0.9515	**0.9549**
		WM	0.9772	0.9774	0.9760	0.9778	0.9777	**0.9788**
	SA		0.9629	0.9631	0.9626	0.9670	0.9653	**0.9675**
3%Noise 40%INU	DC	CSF	0.9308	0.9318	0.9312	0.9467	0.9374	**0.9484**
		GW	0.9221	0.9222	0.9190	0.9295	0.9249	**0.9304**
		WM	0.9583	0.9595	0.9565	0.9627	0.9623	**0.9625**
	SA		0.9435	0.9438	0.9394	0.9486	0.9456	**0.9490**

brain image with a T1 weight and a Slice thickness of 1 mm. The noise levels (calculated relative to the brightest tissue) increased from 1% to 3% and the intensity non-uniformity (INU) is 0%, 20%, 40% ("RF"), respectively. The cerebrospinal (CSF), gray matter (GM), and white matter (WM) sections were selected in the z-axis direction of the slice data. The segmentation results are quantitatively evaluated by Dice coefficient and segmentation accuracy. The formula for the Dice coefficient is:

$$DC(R,T) = \frac{2|R \cap T|}{|R| + |T|} \tag{39}$$

where R represents the clustering result and T represents the actual label.

The formula for segmentation accuracy is

$$SA(R,T) = \sum_{i=1}^{c} \frac{R_i \cap T_i}{\sum_{i=1}^{c} T_i} \tag{40}$$

As shown in Table 6, the DC and SA of MKIFCM are the best compared to all the other algorithms. The segmentation results of a sample image usingdifferent algorithms are shown in Fig. 6. From Table 6 and Fig. 6, we can see that MKFCM's performance is second only to MKIFCM, and is better than other algorithms. This shows that the introduction of multiple-kernel function in the FCM algorithm framework for similarity measurement can improve the performance of the algorithm significantly. Table 6 and Fig. 6 show that IFCM outperforms FCM, MKIFCM outperforms MKFCM, and EKIFCM outperforms WKFCM. This indicates that the introduction of intuitionistic fuzzy sets in the fuzzy clustering algorithm can improve the performance of the algorithm. For MRI brain images, when the sample pixels belong to the boundary region, the hesitation values for such sample pixels to each cluster may be high. Hence, the hesitation value will improve the actual membership value of a sample to different prototypes and has almost equal membership value to each prototype. In addition, the performance of EKIFCM is better than IFCM and WKFCM, indicating that the choice of parameters has a direct impact on the results.

4.7. Discussion

As shown in the above experimental results, FCM and IFCM are not able to achieve good performances on all the datasets except WPBC. The reason is that the similarity measure of FCM and IFCM adopts Euclidean distance which does not measure complex data structures well. WKFCM, EKIFCM and MKFCM utilize the kernel method to obtain better performance than conventional algorithms. From analyzing the experimental results, the MKFCM obtains better performance than WKFCM because the effective multiple-kernel method which can better handle clustering problems in more complex datasets than a single-kernel method. The EKIFCM obtains more accurate classification rates and stable performance because of the kernel function, intuitionistic fuzzy set, and GA mechanism. However, the GA mechanism in EKIFCM requires more processing time than the other methods.

Overall, MKIFCM can provide more stable and better performance in terms of classification rate. The superior performance of MKIFCM can be attributed to two factors. First, the different types of features are grouped, and a composite kernel is constructed to map each feature group into an individual kernel space and linearly combine these kernels with optimal weights. This ensures that MKIFCM can perform better with complex datasets than other methods. Second, the hesitation degree that arises while defining the membership function as a novel uncertainty is introduced in MKIFCM, which can precisely determine the membership function. This is because that the Yager class of IFSs (hesitation degrees) can enhance the membership to amplify the distinction of clusters, and therefore can maximize the good samples in the class and yield better classification rates.

Furthermore, we also studied the sensitivity of each algorithm to initialization, and the result shows that the final performance of EKIFCM and MKIFCM are not easily influenced by initialization.

5. Conclusions

In this paper, we present a novel multi-kernel intuitionistic fuzzy C-means clustering algorithm, namely MKIFCM. The MKIFCM

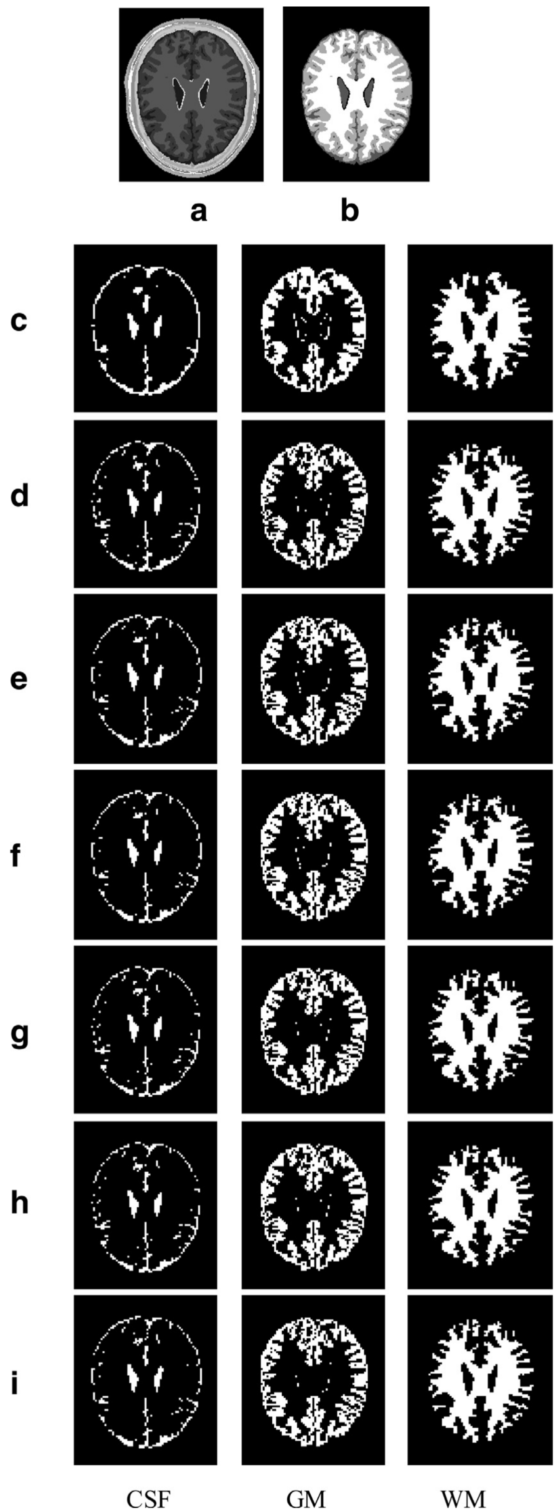

Fig. 6. (a) Original image with INU=0 and 1% noise; (b) image composed of CSF, GM and WM; (c)–(i) the clustering results of FCM, IFCM, WKFCM,MKFCM, EKIFCM and MKIFCM, respectively.

algorithm constructs a linear composite kernel to map each feature group into a new kernel space, and the optimal weights for linearly combining these kernels are automatically adjusted in the clustering procedure. Effective kernels or important features can contribute more to the MKIFCM clustering process and thus make the performance improved. Our comprehensive experiments over a diverse set of datasets clearly demonstrate that our proposed MKIFCM algorithm is able to achieve accurate and robust clustering in the presence of noise, and fast convergence. In the future, we will aim to take the prior knowledge of data structure into consideration for choosing kernel functions.

Acknowledgments

The authors are grateful to anonymous reviewers for their valuable suggestion and comments that have improved this paper. This research was partially supported by NSFC-CAAC (Grant No. U1833119), Hubei natural science foundation (Grant No. 2017CFB500), Wuhan science and technology foundation (Grant No. 2018020401011299), Robotic Discipline Development Fund (2016-1418) from Shenzhen Gov, China.

References

[1] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, 1981.
[2] A.K. Jain, Data Clustering: 50 Years Beyond K-means, Springer, Berlin Heidelberg, 2008.
[3] A. Baraldi, P. Blonda, A survey of fuzzy clustering algorithms for pattern recognition I, IEEE Trans. Syst. Man. Cybern. 29 (6) (1999) 778–785.
[4] M. Yambal, H. Gupta, Image segmenatation using fuzzy C means clustering: a survey, Int. J. Adv. Res. Comput. Commun. Eng. (2013) 2927–2929.
[5] M. Gong, L. Su, M. Jia, W. Chen, Fuzzy clustering with a modified MRF energy function for change detection in synthetic aperture radar images, IEEE Trans. Fuzzy Syst. 22 (1) (2014) 98–109.
[6] T.M. Nguyen, Q.M.J. Wu, Online feature selection based on fuzzy clustering and its applications, IEEE Trans. Fuzzy Syst. 24 (6) (2016) 1294–1306.
[7] J. Gu, L. Jiao, S. Yang, F. Liu, Fuzzy double C-means clustering based on sparse self-representation, IEEE Trans. Fuzzy Syst. 99 (2017) 1-1.
[8] L. Zhu, F.L. Chung, S. Wang, Generalized fuzzy C-means clustering algorithm with improved fuzzy partitions, IEEE Trans. Syst Man Cybern. B 39 (3) (2009) 578–591.
[9] M.N. Ahmed, S.M. Yamany, N. Mohamed, A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data, IEEE Trans. Med. Imag. 21 (3) (2002) 193–199.
[10] S. Chen, D. Zhang, Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure, IEEE Trans. Syst. Man. Cybern. B 34 (4) (2004) 1907–1916.
[11] L. Szilagyi, Z. Benyo, S.M. Szilagyi, MR brain image segmentation using an enhanced fuzzy C-means algorithm, in: Proceeding of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS2003), 2003, pp. 724–726.
[12] W. Cai, S. Chen, D. Zhang, Fast and robust fuzzy C-means clustering algorithms incorporating local information for image segmentation, Pattern Recogn. 40 (3) (2007) 825–838.
[13] X.Y. Chang, Q.N. Wang, Y.W. Liu, Y. Wang, Sparse regularization in fuzzy C-means for high-dimensional data clustering, IEEE Trans. Cybern. 47 (9) (2017) 2616–2627.
[14] B.Q. Hu, C.Y. Wang, On Type-2 Fuzzy Relations and Interval-Valued Type-2 Fuzzy Sets, Elsevier North-Holland, Inc, 2014.
[15] C. Hwang, C.H. Rhee, Uncertain fuzzy clustering: interval type-2 fuzzy approach to C-means, IEEE Trans. Fuzzy Syst. 15 (1) (2007) 107–120.
[16] Z. Ji, Y. Xia, Q. Sun, Interval-valued possibilistic fuzzy C-means clustering algorithm, Fuzzy Sets Syst. 253 (3) (2014) 138–156.
[17] O. Linda, M. Manic, General type-2 fuzzy C-means algorithm for uncertain fuzzy clustering, IEEE Trans. Fuzzy Syst. 20 (5) (2012) 883–897.
[18] D.D. Nguyen, T.N. Long, T.P. Long, Towards hybrid clustering approach to data classification: multiple kernels based interval-valued fuzzy C-means algorithms, Fuzzy Sets Syst. 279 (2015) 17–39.
[19] K.T. Atanassov, P. Rangasamy, Intuitionistic fuzzy sets, Fuzzy Sets Syst. 20 (1) (1986) 87–96.
[20] D.K. Iakovidis, N. Pelekis, E. Kotsifakos, Intuitionistic fuzzy clustering with applications in computer vision, in: Proceeding of the international Conference on Advanced Concepts for Intelligent Vision Systems (ACIVS2008), 2008, pp. 764–774.
[21] Z. Xu, J. Wu, Intuitionistic fuzzy C-means clustering algorithms, J. Syst. Eng. Electron. 21 (4) (2010) 580–590.
[22] H.S. Le, B.C. Cuong, P.L. Lanzi, N.T. Thong, A novel intuitionistic fuzzy clustering method for geo-demographic analysis, Expert Syst. Appl. 39 (10) (2012) 9848–9859.
[23] T. Chaira, A novel intuitionistic fuzzy C means clustering algorithm and its application to medical images, Appl. Soft Comput. 11 (2) (2011) 1711–1717.
[24] K.P. Lin, A novel evolutionary kernel intuitionistic fuzzy c-means clustering algorithm, IEEE Trans. Fuzzy Syst 22 (5) (2014) 1074–1087.
[25] H. Verma, R.K. Agrawal, A. Sharan, An improved intuitionistic fuzzy c-means clustering algorithm incorporating local information for brain image segmentation, Appl. Soft Comput. 46 (3) (2016) 543–557.

[26] J. Shawetaylor, N. Cristianini, Kernel methods for pattern analysis: ranking, clustering and data visualisation, J. Am. Stat. Assoc. 101 (476) (2004) 1730-1730.
[27] M.S. Yang, H.S. Tsai, A Gaussian kernel-based fuzzy c-means algorithm with a spatial bias correction, Pattern Recogn. Lett. 29 (12) (2008) 1713–1725.
[28] D. Graves, W. Pedrycz, Performance of kernel-based fuzzy clustering, Electronic. Lett. 43 (25) (2007) 1445–1446.
[29] D. Graves, W. Pedrycz, Kernel-based Fuzzy Clustering and Fuzzy Clustering: A Comparative Experimental Study, 161, Elsevier North-Holland, Inc, 2010, pp. 522–543.
[30] H. Shen, J. Yang, S. Wang, Attribute weighted mercer kernel based fuzzy clustering algorithm for general non-spherical datasets, Soft. Comput. 10 (11) (2006) 1061–1073.
[31] H.C. Huang, Y.Y. Chuang, C.S. Chen, Multiple kernel fuzzy clustering, IEEE Trans. Fuzzy Syst. 20 (1) (2012) 120–134.
[32] L. Chen, C.L.P. Chen, M. Lu, A multiple-kernel fuzzy c-means algorithm for image segmentation, IEEE Trans. Syst . Man. Cybern. B 41 (5) (2011) 1263–1274.
[33] N. Baili, H. Frigui, Fuzzy clustering with multiple kernels, in: Proceeding of the 2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE2011), 2011, pp. 490–496.
[34] S. Zeng, X.Y. Wang, H. Cui, C.J Zheng, D. Feng, A unified collaborative multi-kernel fuzzy clustering for multiview data, IEEE Trans. Fuzzy Syst. 26 (3) (2018) 1671–1687.
[35] D. Huang, C.D. Wang, J.H. Lai, Locally weighted ensemble clustering, IEEE Trans. Cybern. 48 (5) (2018) 1460–1473.
[36] I. Khan, J.Z. Huang, K. Ivanov, Incremental density-based ensemble clustering over evolving data streams, Neurocomputing 191 (2016) 34–43.
[37] D. Huang, J.H. Lai, C.D. Wang, Ensemble clustering using factor graph, Pattern Recogn. 50 (2016) 131–142.
[38] A. Asuncion, D.J. Newman, UCI machine learning repository irvine, 2007.
[39] T. Randen, in: Brodatz Textures [EB/OL], in: http://www.ux.uis.no/~tranden/brodatz.html, 2007.
[40] TILDA. Textile defect image database[DB]. University of Freiburg. Germany. http://lmb.informatik.unifreiburg.de/research/dfgtexture/tilda.
[41] S. Zeng, Z.Y. Wang, C.J. Gao., Z. Kang, D. Feng, Hyperspectral Image Classification with Global-Local Discriminant Analysis and Spatial-spectral Context, IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 11 (12) (2018) 5005–5018.
[42] MRI brain images, Available: http://brainweb.bic.mni.mcgill.ca/brainweb.
[43] M.K. Pakhira, S. Bandyopadhyay, U. Maulik, A study of some fuzzy cluster validity indices, genetic clustering and application to pixel classification, Fuzzy Sets Syst. 155 (2) (2005) 191–214.
[44] U. Maulik, S. Bandyopadhyay, Performance evaluation of some clustering algorithms and validity indices, IEEE Trans. Pattern Anal. Mach. Intell. 24 (12) (2002) 1650–1654.
[45] D.A. Clausi, H. Deng, Design-based texture feature fusion using Gabor filters and co-occurrence probabilities, IEEE Trans. Image Process. 14 (7) (2005) 925–936.

Shan Zeng received the B.E. degree in mechanical engineering in 2003 and the M.E. degree in mechatronics engineering in 2009 from Wuhan Polytechnic University. He received the Ph.D. degree in pattern recognition and intelligent systems in 2012 from Huazhong University of Science and Technology. In 2003, he joined the College of Mathematics and Computer Science, Wuhan Polytechnic University, Wuhan, China, where he is currently an Associate Professor. During 2015–2016, he was a visiting scholar at Macau University and The University of Sydney. His research interests include pattern recognition, machine learning, image processing, and hyperspectral imaging with their applications in nondestructive testing of food quality.

Zhiyong Wang received his B. Eng. and M. Eng. Degrees in electronic engineering from South China University of Technology, Guangzhou, China, and his Ph.D. degree from Hong Kong Polytechnic University, Hong Kong. He is currently an Associate Professor and Associate Director of the Multimedia Laboratory with the School of Information Technologies, The University of Sydney, Australia. His research interests focus on multimedia computing, including multimedia information processing, retrieval and management, and pattern recognition.

Rui Huang received his B.S. degree in computer software from Peking University in 1999, M.E. degree in pattern recognition and intelligent systems from Chinese Academy of Sciences in 2002, and Ph.D. degree from Rutgers University in 2008, specializing in computer vision. He is currently a faculty member at The Chinese University of Hong Kong, Shenzhen. His research experience includes subspace analysis for face recognition, deformable models and probabilistic graphical models for image and video analysis. His current research interests include machine learning methods for intelligent video surveillance, computer vision for robotics.

Ling Chen received the B.S. degree in electronic information engineering from Wuhan University of Science and Technology, Wuhan, China in 2016. She is currently pursuing the M.S. degree with mechanical and electronic engineering in Wuhan Polytechnic University, Wuhan, China. Her research interests include image processing, hyperspectral image analysis, computer vision, machine learning.

(David) Dagan Feng is currently Director (Research), Institute of Biomedical Engineering & Technology, and Academic Director, USYD-SJTU Joint Research Alliance. He has been Associate Dean, Faculty of Science, University of Sydney. He has been the Chair Professor, Advisory Professor, Guest Professor, Adjunct Professor or Chief Scientist in different world-known universities and institutes. He is the Founder and Director of the Biomedical & Multimedia Information Technology Research Group at the University of Sydney. He has served as Chairs or Editors of different committees and key journals in the area. Professor Feng has been elected as Fellow of ACS (Australia), HKIE (Hong Kong), IET (UK), IEEE (USA), and Australian Academy of Technological Sciences and Engineering. He received his Ph.D. in Computer Science from the University of California, Los Angeles in 1988.