IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 26, 2019, accepted June 12, 2019, date of publication June 26, 2019, date of current version July 15, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2924957

A Kernel-Based Intuitionistic Fuzzy C-Means Clustering Using Improved Multi-Objective Immune Algorithm

WENKE ZANG, ZEHUA WANG, DONG JIANG, AND XIYU LIU
School of Business, Shandong Normal University, Jinan 250014, China
Corresponding author: Wenke Zang (wink@sdnu.edu.cn)

This work was supported in part by the National Science Foundation of China under Grant 61472231 and Grant 61402266.

ABSTRACT Clustering algorithms have attracted a lot of attentions recently in real-world applications. However, the traditional clustering algorithms still have plenty of defects which are not yet resolved. In this paper, a kernel-based intuitionistic fuzzy C-means clustering using improved multi-objective artificial immune algorithm (KIFCM-IMOIA) is proposed. In our algorithm, the kernel trick and the intuitionistic fuzzy entropy (IFE) are introduced into the objective functions, which improves the robustness to noises. In addition, an improved multi-objective optimization immune algorithm (IMOIA), which simultaneously optimizes the intra-cluster compactness and inter-cluster separation, is proposed to prevent the algorithm from falling into local optimum. The proposed IMOIA uses a novel active antibody selection strategy, a hybrid differential evolution strategy, and an adaptive mutation operator to maintain better distribution of the solutions with better convergence. Finally, we performed experiments using 14 UCI datasets and compared our algorithm with six clustering methods on three performance metrics. The experimental results show that our algorithm performs better than other algorithms.

INDEX TERMS Intuitionistic fuzzy C-means, kernel function, artificial immune algorithm, multi-objective optimization.

I. INTRODUCTION

As an unsupervised classification method, clustering algorithm is a research hotspot in recent decades and is widely used in pattern recognition [1], data mining [2], [3], image segmentation [4]–[6] and so on. Clustering methods are mainly divided into hard clustering [7]–[10] and soft clustering [11]–[14]. Among the soft clustering methods, fuzzy C-means (FCM) clustering algorithm [11] is the most flexible and widely used one. FCM proposes the membership degree that is used to indicate the extent to which each data point belongs to each cluster. In order to further improve the performance of the algorithm, the intuitionistic fuzzy C-means (IFCM) clustering algorithm was proposed in [12]. In addition to considering the membership degree of fuzzy sets, the introduction of non-membership degree and hesitation degree is an important extension of fuzzy set theory. Although the IFCM algorithm exhibits an absolute advantage over FCM, the performance of IFCM is still hampered by two noticeable problems, i.e., being prone to local optimum and sensitivity to noises.

In recent years, the application of multi-objective evolutionary algorithm (MOEA) in fuzzy clustering has become popular. The multi-objective clustering algorithms consider two or more clustering objectives, which prevents the algorithm from falling into local optimum and makes the clustering results more robust. In [15], the multi-objective clustering algorithm with automatic k-determination (MOCK) was proposed. This algorithm is built from original PESA-II [16]. The objectives used in MOCK are the overall deviation *Dev* and connectivity *Conn*. In [17], a multi-objective genetic algorithm with fuzzy C-means, denoted as FCM-NSGA, was presented. In FCM-NSGA, the non-dominated sorting genetic algorithm (NSGA-II) [18] is used to control the multi-objective optimization considering two objectives: the well-known FCM objective function J_{FCM} and the overlap-separation measure *OS*. The non-dominated sorting genetic algorithm using fuzzy membership chromosome

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

(NSGA-FMC) was proposed in [19]. In NSGA-FMC, two completely independent fuzzy objective functions, i.e., the fuzzy compactness π and separation S, are utilized. The algorithm is also built from the original NSGA-II [18].

Although the above FCM algorithms have achieved some promising results, they are sensitive to noises and outliers seeing that the Euclidean distance is used as the similarity measurement. Aiming at this problem, various improved methods have been proposed, in which the kernel-based clustering method has attracted extensive attention. In [20], a kernel-based fuzzy C-means (KFCM) algorithm was proposed. Experimental results show that the algorithm is robust to noises and outliers. In [21], a multiple kernel fuzzy C-means (MKFC) algorithm was proposed, which extends the FCM with a multiple kernel-learning setting. In [22], a novel multi-objective kernel clustering algorithm with automatic attribute weighting (MOKCW) was proposed. In MOKCW, two kernel-based objective functions J_c and F_s, which consider the compactness within the cluster and the separation between clusters respectively, are optimized by original NSGA-II [18].

Generally, conventional multi-objective clustering algorithms are built from the original MOEAs, such as PESA-II [16] and NSGA-II [18]. In order to effectively find the Pareto-optimal solutions in the solution space, many state-of-the-art MOEAs [23]–[26] have been proposed. Among them, artificial immune system (AIS) [27], an evolutionary algorithm based on the information processing mechanism of biological immune system, has been successfully used in multi-objective optimization problems (MOPs) and has a good application prospect. In recent years, many multi-objective immune algorithms (MOIAs) [26]–[29] have been proposed. For example, the non-dominated neighbor immune algorithm (NNIA) [26], a well-known multi-objective immune algorithm, uses non-dominated neighbor-based selection and proportional cloning method to enhance the local search ability in the less-crowded regions of the current front.

In this paper, by incorporating the intuitionistic fuzzy set and the kernel trick, two completely independent fuzzy objective functions J_{KIFCM} and S_{KIFCM}, which consider the intra-cluster compactness and the inter-cluster separation respectively, are proposed. This is done to improve the robustness to noises. Then, we present an improved multi-objective optimization immune algorithm as an underlying multi-objective optimization framework to prevent the algorithm from falling into local optimum. The IMOIA uses a center-based string encoding which is suitable for the clustering problem. In addition, a novel grid-based active antibody selection strategy, a hybrid differential evolution strategy and an adaptive mutation operator are presented to search through the solution space for optimal solutions. Finally, we compared our algorithm with a kernel-based intuitionistic fuzzy C-means clustering using NNIA (KIFCM-NNIA) and five state-of-the-art clustering algorithms by using 14 UCI datasets. Preliminary results show that our algorithm is superior to compared algorithms in terms of three clustering metrics, i.e., the clustering accuracy (*ACC*) [30], adjusted rand index (*ARI*) [31] and normalized mutual index (*NMI*) [32].

In the remainder of this paper, several basic concepts such as FCM, IFE and MOIA are presented in Section 2. Section 3 gives objective functions of the kernel-based intuitionistic fuzzy C-means problem. In Section 4, we present our algorithm. In Section 5, we present the experimental results. Section 6 concludes the paper.

II. RELATED CONCEPTS

In this section, we briefly review basic concepts of FCM [11], IFE [33] and MOIA [26].

A. FUZZY C-MEANS

Given a dataset containing n data samples, $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$, FCM algorithm outputs the membership degree u_{ik}, which represents the probability that the data sample $\mathbf{x}_k$ belongs to the *i-th* cluster. The membership degree can be obtained by finding the minimum value of the objective function J_{FCM}:

$$\begin{aligned} J_{FCM} &= \sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^m \|\mathbf{x}_k - \mathbf{v}_i\|^2 \\ s.t.\ &\sum_{i=1}^{c} u_{ik} = 1, \quad \forall k \\ &0 < \sum_{k=1}^{n} u_{ik} < n, \quad \forall i \end{aligned} \tag{1}$$

When the objective function J_{FCM} takes the minimum value, the corresponding membership u_{ik} and cluster center $\mathbf{v}_i$ can be calculated as:

$$\begin{cases} u_{ik} = \dfrac{1}{\sum_{j=1}^{c}\left[\|\mathbf{x}_k - \mathbf{v}_i\| / \|\mathbf{x}_k - \mathbf{v}_j\|\right]^{\frac{2}{m-1}}} \\ \mathbf{v}_i = \dfrac{\sum_{k=1}^{n} u_{ik}^m \mathbf{x}_k}{\sum_{k=1}^{n} u_{ik}^m} \end{cases} \tag{2}$$

where $V = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_c\}$, $\mathbf{v}_i (i = 1, 2, \ldots, c)$ represents the *i-th* cluster center, c represents the number of cluster centers, $\|\mathbf{x}_k - \mathbf{v}_i\|$ is the Euclidean distance between data point $\mathbf{x}_k$ and cluster center $\mathbf{v}_i$, m is a parameter used to determine the amount of fuzziness.

B. INTUITIONISTIC FUZZY ENTROPY

It is generally believed that the IFE gives the degree of blurring of a fuzzy set. In the intuitionistic fuzzy set $A = \{u_A(\mathbf{x}_k), \gamma_A(\mathbf{x}_k), \pi_A(\mathbf{x}_k) | \mathbf{x}_k \in X\}$, $u_A(\mathbf{x}_k)$, $\gamma_A(\mathbf{x}_k)$, and $\pi_A(\mathbf{x}_k)$ are the membership degree, non-membership degree,

and hesitation degree of $\mathbf{x}_k$ with respect to A, respectively. IFE can be defined as:

$$IFE(A) = \sum_{k=1}^{n} \pi_A(\mathbf{x}_k) e^{[1-\pi_A(\mathbf{x}_k)]} \tag{3}$$

where hesitation degree $\pi_A(\mathbf{x}_k)$ is expressed as:

$$\pi_A(\mathbf{x}_k) = 1 - u_A(\mathbf{x}_k) - \gamma_A(\mathbf{x}_k) \tag{4}$$

where the formula for non-membership degree $\gamma_A(\mathbf{x}_k)$ is defined as:

$$\gamma_A(\mathbf{x}_k) = (1 - u_A(\mathbf{x}_k)^\alpha)^{1/\alpha} \tag{5}$$

where the value of α is discussed in section 5. Notice that if A is a normal fuzzy set, then $IFE(A) = 0$, i.e., $\pi_A(\mathbf{x}_k) = 0, \forall \mathbf{x}_k$; if $u_A(\mathbf{x}_k) = \gamma_A(\mathbf{x}_k) = 0, \forall \mathbf{x}_k$, then $IFE(A) = n$; if the membership and non-membership of each element are reduced, then their sum is also reduced, the ambiguity is reduced, the hesitation degree is increased, and the *IFE* is increased.

C. MULTI-OBJECTIVE IMMUNE ALGORITHM

MOPs are aimed at optimizing multiple, possibly conflicting objectives, simultaneously. The general definition of a MOP [34] is shown below:

$$\begin{cases} \min F(\mathbf{s}) = (f_1(\mathbf{s}), f_2(\mathbf{s}), \ldots, f_k(\mathbf{s})) \\ subject\ to \\ \mathbf{s} = (s_1, s_2, \ldots, s_m) \in \Omega \end{cases} \tag{6}$$

where $\mathbf{s} = (s_1, s_2, \ldots, s_m)$ is an m-dimensional candidate solution, Ω is the decision space, $F(\mathbf{s})$ is an objective vector, and $f_i(\mathbf{s})$ is the *i-th* objective function. Considering two candidate solutions $\mathbf{s}_A \in \Omega$ and $\mathbf{s}_B \in \Omega$, it is said that $\mathbf{s}_A$ dominates $\mathbf{s}_B$, i.e., $\mathbf{s}_A \succ \mathbf{s}_B$, if and only if

$$\begin{aligned} &\forall i = 1, 2, \ldots, m \quad f_i(\mathbf{s}_A) \leq f_i(\mathbf{s}_B) \\ &and\ \exists j = 1, 2, \ldots, m \quad f_j(\mathbf{s}_A) < f_j(\mathbf{s}_B) \end{aligned} \tag{7}$$

In MOP, we say a solution is a Pareto-optimal solution when it is not dominated by any other. The set of non-dominated solutions is called as Pareto front. The goal of MOEA is to find a set of Pareto-optimal solutions that approximate the true Pareto front.

MOIA is a new bionic algorithm based on the principles and processes of biological immune system. NNIA is the most representative multi-objective immune algorithm. In NNIA, antigens refer to the multi-objective problems and the corresponding constraints. The potential solution of MOP is regarded as an antibody, e.g., the candidate solution $\mathbf{s} = (s_1, s_2, \ldots, s_m)$ in Eq.(6). Thus, a no-dominated solution is regarded as a dominant antibody. Only partial less-crowded non-dominated individuals are called active antibodies. An antibody population is composed by a set of antibodies and a dominant population is the set of dominant antibodies.

In the biological immune system, when external antigen is detected by the biological immune system, the B-cell will eliminate invaders by the reactive procedures immediately, for example, clonal selection and affinity maturation by hyper-mutation. This is similar with the multi-objective immune algorithm. In NNIA, the antibodies with better affinities will be selected to reproduce by cloning. After that, the evolutionary operations are applied on each antibody in the clone population to realize the affinity maturation process. Then antibodies with higher affinity will retain as memory cells to maintain the population diversity. We show the population evolution process of NNIA in Figure 1, where A_t is active population which composed of less-crowded non-dominated antibodies, C_t is the clone population, C_t' is an evolved population of C_t, D_t is dominant population.

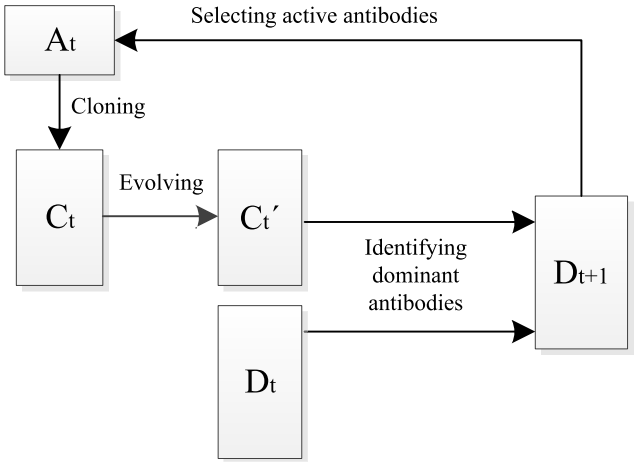

FIGURE 1. Population evolution process in the *t-th* generation.

III. OBJECTIVE FUNCTIONS OF MULTI-OBJECTIVE CLUSTERING

The performance of the multi-objective clustering algorithm depends critically on the clustering objectives. In this paper, we choose the FCM objective function J_{FCM} [11] and fuzzy separation S_{FCM} [24] as the two clustering objectives. The definition of J_{FCM} is given in Eq.(1). S_{FCM} is expressed as

$$S_{FCM} = \sum_{q=1}^{c} \sum_{p=1, p\neq q}^{c} (u_{qp})^m \left\| \mathbf{v}_q - \mathbf{v}_p \right\|^2 \tag{8}$$

where $\mathbf{v}_q$ and $\mathbf{v}_p$ represent the *q-th* and *p-th* cluster center respectively, $\left\| \mathbf{v}_q - \mathbf{v}_p \right\|$ is the Euclidean distance between two cluster centers $\mathbf{v}_q$ and $\mathbf{v}_p$, the membership between two cluster centers u_{qp} is calculated as:

$$u_{qp} = \frac{1}{\sum_{k=1, k\neq p}^{c} \left[\left\| \mathbf{v}_p - \mathbf{v}_q \right\| / \left\| \mathbf{v}_p - \mathbf{v}_k \right\| \right]^{\frac{2}{m-1}}} \tag{9}$$

A. INTUITIONISTIC FUZZY C-MEANS CLUSTERING

In IFCM [12], the data points $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ are classified into c homogeneous groups or clusters represented as $F = \{F_1, F_2, \ldots, F_c\}$. $u_{Fi}(\mathbf{x}_k)$, $\gamma_{Fi}(\mathbf{x}_k)$ and $\pi_{Fi}(\mathbf{x}_k)$ are the membership degree, non-membership degree, and hesitation

degree for $\mathbf{x}_k$ in *i-th* group F_i, respectively. For simplicity, $u_{Fi}(\mathbf{x}_k)$, $\gamma_{Fi}(\mathbf{x}_k)$, and $\pi_{Fi}(\mathbf{x}_k)$ are denoted as u_{ik}, γ_{ik} and π_{ik}, respectively. The objective function of IFCM proposed in [12] is represented as follows:

$$J_{IFCM} = \sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik}^*)^m \|\mathbf{x}_k - \mathbf{v}_i\|^2 + \sum_{i=1}^{c}\pi_i^* e^{1-\pi_i^*} \quad (10)$$

where $u_{ik}^* = u_{ik} + \pi_{ik}$ represents the intuitionistic fuzzy membership of the *k-th* data point to the *i-th* cluster center, π_i^* is defined by:

$$\pi_i^* = \frac{1}{n}\sum_{k=1}^{n}\pi_{ik} \quad (11)$$

where the definition of hesitation degree π_{ik} is consistent with Eq.(4):

Then, the second clustering objective S_{IFCM} is calculated as:

$$S_{IFCM} = \sum_{p=1}^{c}\sum_{q=1,q\neq p}^{c}(u_{qp}^*)^m \|\mathbf{v}_q - \mathbf{v}_p\|^2 + \sum_{p=1}^{c}\pi_p^* e^{1-\pi_p^*} \quad (12)$$

where π_p^* is calculated as:

$$\pi_p^* = \frac{\sum_{q=1,q\neq p}^{c}\pi_{qp}}{c-1} \quad (13)$$

B. GAUSSIAN RADIAL BASIS FUNCTION

The IFCM using the Euclidean distance is sensitive to noises and outliers. An effective way to solve this problem is to use kernel method to project data into higher dimensional space. In this paper, Gaussian radial basis function (GRBF) [35] is used to improve the objective function of IFCM.

$$K(x,y) = \exp\left(\frac{-\|x-y\|^2}{\sigma}\right) \quad (14)$$

where $\|x-y\|$ is the Euclidean distance between two data points x and y, σ is the bandwidth parameter. However, choosing the appropriate bandwidth value can be very difficult. The method of bandwidth selection in this study will be detailed described in the next section.

In KFCM [20], a nonlinear map is defined as $\Phi : x \to \Phi(x) \in F$, where $x \in X$. X is the data space. F is the transformed feature space with higher dimension. In other word, a nonlinear map $\Phi(\cdot)$ is defined to map the points in the original space into the high-dimensional feature space. The Euclidean distance $\|x-y\|^2$, which measures the similarity between data points x and y, is replaced by the Euclidean distance $\|\Phi(x)-\Phi(y)\|^2$ between mapped points $\Phi(x)$ and $\Phi(y)$ in the transformed feature space.

$$\begin{aligned}\|\Phi(x)-\Phi(y)\|^2 &= (\Phi(x)-\Phi(y))^T(\Phi(x)-\Phi(y))\\ &= K(x,x)+K(y,y)-2K(x,y)\\ &= 2(1-K(x,y))\\ &= 2\left(1-\exp\left(\frac{-\|x-y\|^2}{\sigma}\right)\right)\end{aligned} \quad (15)$$

C. OBJECTIVE FUNCTION OF PROPOSED ALGORITHM

According to the definition given in the previous section, we reconstruct the clustering objectives represented by Eq.(10) and Eq.(12). The multi-objective clustering problem can be described as:

$$\min F(\mathbf{v}) = (f_1(\mathbf{v}), f_2(\mathbf{v})) \quad (16)$$

The first clustering objective which represents the intra-cluster compactness is calculated as

$$f_1(\mathbf{v}) = J_{KIFCM} = 2\sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik}^*)^m\left(1-\exp\left(\frac{-\|\mathbf{x}_k-\mathbf{v}_i\|^2}{\sigma_1}\right)\right) + \sum_{i=1}^{c}\pi_i^* e^{1-\pi_i^*} \quad (17)$$

where $u_{ik}^* = u_{ik} + \pi_{ik}$ is the kernel-based intuitionistic fuzzy membership and u_{ik} is calculated as:

$$u_{ik} = \frac{1}{\sum_{j=1}^{c}\left[(1-K(\mathbf{x}_k,\mathbf{v}_i))/(1-K(\mathbf{x}_k,\mathbf{v}_j))\right]^{\frac{1}{m-1}}} \quad (18)$$

The hesitation degree π_{ik} is defined as:

$$\pi_{ik} = 1 - u_{ik} - (1-u_{ik}^{\alpha})^{1/\alpha} \quad (19)$$

The second clustering objective which represents the inter-cluster separation is as follows:

$$\begin{aligned}f_2(\mathbf{v}) &= \frac{1}{S_{KIFCM}}\\ &= \frac{1}{2\sum_{q=1}^{c}\sum_{p=1,p\neq q}^{c}(u_{qp}^*)^m\left(1-\exp\left(\frac{-\|\mathbf{v}_q-\mathbf{v}_p\|^2}{\sigma_2}\right)\right)+\sum_{q=1}^{c}\pi_q^* e^{1-\pi_q^*}}\end{aligned} \quad (20)$$

where $u_{qp}^* = u_{qp} + \pi_{qp}$ and u_{qp} is calculated as:

$$u_{qp} = \frac{1}{\sum_{j=1,j\neq q}^{c}\left[(1-K(\mathbf{v}_q,\mathbf{v}_p))/(1-K(\mathbf{v}_q,\mathbf{v}_j))\right]^{\frac{1}{m-1}}} \quad (21)$$

In this paper, the calculation of bandwidth σ depends on the distance variance of all samples in the given dataset. The bandwidth parameter σ_1 in Eq.(17) is expressed as:

$$\sigma_1 = \left(\frac{1}{n-1}\sum_{k=1}^{n}(d_k-\bar{d})^2\right)^{\frac{1}{2}} \quad (22)$$

where d_k is the Euclidean distance between data point $\mathbf{x}_k$ and $\bar{\mathbf{x}}$, $\bar{\mathbf{x}}$ represents the mean value of all data points, i.e., $\bar{\mathbf{x}} = (\sum_{k=1}^{n}\mathbf{x}_k)/n$, $\bar{d}$ represents the mean value of d_k, i.e., $\bar{d} = (\sum_{k=1}^{n}d_k)/n$.

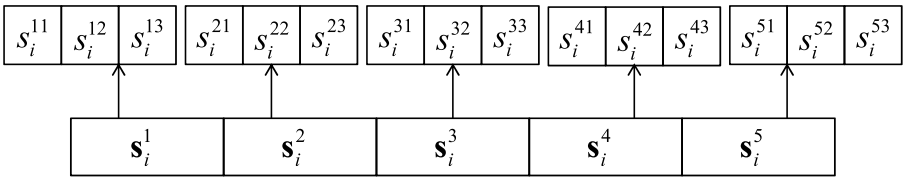

FIGURE 2. An antibody with 5 cluster centers in three dimensions.

The bandwidth parameter σ_2 in the second objective function is calculated by

$$\sigma_2 = \left(\frac{1}{c-1} \sum_{q=1}^{c} (d_q - \bar{d})^2 \right)^{\frac{1}{2}} \tag{23}$$

where d_q is the Euclidean distance between cluster center $\mathbf{v}_q$ and $\bar{\mathbf{v}}$, $\bar{\mathbf{v}}$ represents the mean value of all cluster centers. $\bar{d}$ represents the mean value of d_q.

Notice that our multi-objective clustering algorithm optimizes two different clustering criteria, i.e., J_{KIFCM} and S_{KIFCM}. J_{KIFCM} represents the intra-cluster compactness, the smaller its value, the better the clustering result. On the contrary, S_{KIFCM} is inter-cluster distance which needs to be maximized.

IV. A KERNEL-BASED INTUITIONISTIC FUZZY C-MEANS CLUSTERING USING IMPROVED MULTI-OBJECTIVE IMMUNE ALGORITHM

In this section, we present the basic idea and the general framework of the proposed algorithm KIFCM-IMOIA.

A. INITIALIZATION OF ANTIBODY POPULATION

In this paper, the antigen refers to the clustering objective, i.e., $\min F(\mathbf{v}) = (f_1(\mathbf{v}), f_2(\mathbf{v}))$. An antibody refers to a candidate solution of the clustering problem, i.e., a set of clustering centers. In our algorithm, antibodies are encoded using real numbers. $P_t = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_N\}$ represents the antibody population of current generation, where N denotes the size of population, $\mathbf{s}_i$ denotes the *i-th* antibody. Specifically, $\mathbf{s}_i = \{\mathbf{s}_i^1, \mathbf{s}_i^2, \ldots, \mathbf{s}_i^c\}$ represents a solution to the clustering problem, i.e., c cluster centers. One cluster center $\mathbf{s}_i^j$ is an m- dimensional vector that can be expressed as $\mathbf{s}_1^j = \{s_1^{j1}, s_1^{j2}, \ldots, s_1^{jm}\}$. Then, an antibody can be represented by $\mathbf{s}_i = \{s_i^1, s_i^2, \ldots, s_i^{c\times m}\}$. Figure 2 shows an example of an antibody $\mathbf{s}_i$ with 5 cluster centers in three dimensions.

B. GRID-BASED SELECTION OF ACTIVE ANTIBODIES

In NNIA, the selection of active antibodies is based on the crowding distance [28]. However, there are shortcomings of this method. Considering Figure 3, there are 8 antibodies at the Pareto front, and they are labeled by $\mathbf{s}_1$ through $\mathbf{s}_8$. The corresponding crowding distance values of 8 antibodies are Inf, 1.0323, 0.6552, 0.5744, 0.1689, 0.4429, 0.8193, Inf. Supposing that five antibodies need to be selected to participate in subsequent evolution, antibodies $\mathbf{s}_4$ through $\mathbf{s}_6$ with lower crowding distances will be discarded. However, this will result in skewed uniformity of the remaining dominant antibodies in the Pareto front.

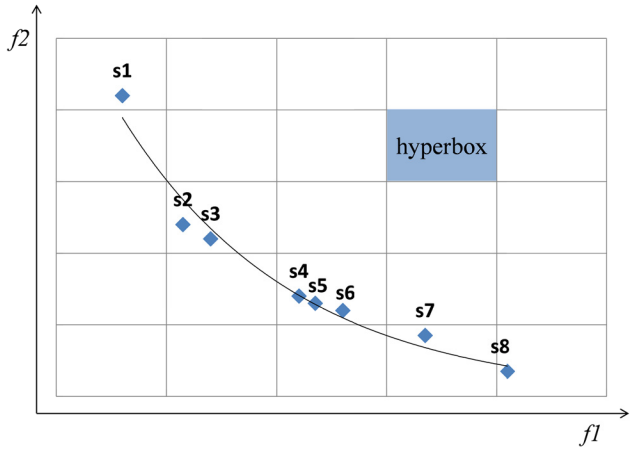

FIGURE 3. Grid-based selection of active antibodies.

To overcome this problem, we apply grid-based selection method. Assuming that the 2-dimensional objective space should be divided into K hyper-boxes, the width wid_j of a hyper-box of the *j-th* objective is calculated as:

$$wid_j = \frac{ub_j - lb_j}{\sqrt{K}} \tag{24}$$

where ub_j and lb_j are the upper and lower boundaries of grid in the *j-th* objective respectively. ub_j and lb_j are calculated as:

$$\begin{cases} lb_j = \min(f_j) - \dfrac{\max(f_j) - \min(f_j)}{2 \times \sqrt{K}} \\ ub_j = \max(f_j) + \dfrac{\max(f_j) - \min(f_j)}{2 \times \sqrt{K}} \end{cases} \tag{25}$$

where $\max(f_j)$ and $\min(f_j)$ are the maximum and minimum value of the *j-th* objective in current population, respectively.

To show the details of grid-based selection method, its pseudo-code is provided in Algorithm 1.

As shown in Figure 3, the 2-dimensional objective space is divided into 25 hyper-boxes. Then, the grid-based selection method is used to select active antibodies from dominant population. According Algorithm 1, two solutions in $\mathbf{s}_4$, $\mathbf{s}_5$ and $\mathbf{s}_6$ are selected randomly to be deleted, and then one of solutions $\mathbf{s}_2$ and $\mathbf{s}_3$ will be deleted. Thus, five solutions, i.e., $\mathbf{s}_1$, $\mathbf{s}_2(\mathbf{s}_3)$, $\mathbf{s}_4(\mathbf{s}_5/\mathbf{s}_6)$, $\mathbf{s}_7$, $\mathbf{s}_8$, will be selected as active antibodies. However, the five active antibodies selected by NNIA

Algorithm 1 Grid-Based Selection of Active Antibodies

Input: K: the number of hyper-boxes
D_t: the dominant population
N: the maximum size of active population
Output: the active population A_t

1. Divide the objective space into K hyper-boxes based on Eq.(24) and Eq.(25);
2. n_k = the size of population in each hyper-box;
while $|D_t| > N$ **do**
3. Select one hyper-box with the highest n_k;
4. Delete an antibody from the selected hyper-box, randomly;
5. Update D_t;
end while
6. $A_t = D_t$

are $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_7, \mathbf{s}_8$. Thus, grid-based selection method of active antibodies helps to maintain better uniformity of the solutions.

C. PROPORTIONAL CLONING

Given a set of active antibodies $S = (\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_N)$. Each antibody $\mathbf{s}_i$ in S needs to be reproduced with q_i times. First, the variant crowding distance (VCD) [24] of each active antibody is calculated. Then, the active antibody with higher VCD has a larger q_i for enhancing local search around the active antibody. The value of q_i is calculated by

$$q_i = \left\lceil n_C \times \frac{VCD(\mathbf{s}_i)}{\sum_{j=1}^{N} VCD(\mathbf{s}_j)} \right\rceil \tag{26}$$

where n_C donates expectant size of the clone population. It should be noted that the VCD of boundary solutions are infinity. Thus, when calculating the values of q_i for the boundary solutions, the values of VCD are twice the maximum value of all solutions except the boundary solutions.

D. HYBRID DIFFERENTIAL EVOLUTION STRATEGY

Differential evolution (DE) [28] is a powerful stochastic search method, which has been widely used in multi-objective evolutionary algorithms. In this paper, we propose a hybrid differential evolution strategy consisting of two well-known DE strategies [36], i.e., rand/1/bin and best/1/bin, denoted by DE1 and DE2 respectively. DE1 and DE2 have different advantages: DE1 is conducive to maintaining the diversity of population, and DE2 is beneficial to accelerating the convergence of algorithm.

Assuming that each antibody in clone population is represented by $\mathbf{s}_i = (s_i^1, s_i^2, \ldots, s_i^{cm})$, the mutant vector $\mathbf{v}_i = (v_i^1, v_i^2, \ldots, v_i^{cm})$ corresponding to $\mathbf{s}_i$ is generated as

$$\mathbf{v}_i = \begin{cases} \mathbf{s}_{r1} + F \times (\mathbf{s}_{r2} - \mathbf{s}_{r3}), & if\ rand_i \le 0.5 \\ \mathbf{s}_{best} + F \times (\mathbf{s}_{r2} - \mathbf{s}_{r3}), & otherwise \end{cases} \tag{27}$$

where $\mathbf{s}_{r1}$, $\mathbf{s}_{r2}$ and $\mathbf{s}_{r3}$ are randomly selected from the dominant population, $\mathbf{s}_{best}$ is randomly selected from active population, $rand_i$ is a random number over interval [0, 1], F is control parameter and its value is 0.5 in this paper.

Then a trial vector $\mathbf{y}_i = (y_i^1, y_i^2, \ldots, y_i^{cm})$ will be obtained from its parents $\mathbf{s}_i$ and $\mathbf{v}_i$ by using the following crossover rule:

$$y_i^j = \begin{cases} v_i^j, & if\ rand_i < 0.5\ or\ j = I_i \\ s_i^j, & otherwise \end{cases} \tag{28}$$

where s_i^j, v_i^j and y_i^j represent the j-th variable of the clone antibody $\mathbf{s}_i$, mutant vector $\mathbf{v}_i$ and trial vector $\mathbf{y}_i$, respectively, $rand_i$ is a random number over interval [0, 1], I_i is a random integer in [1, cm].

E. ADAPTIVE MUTATION OPERATOR

After the differential evolution, adaptive mutation operator is used to enhance the population diversity. For a trial vector $\mathbf{y}_i = (y_i^1, y_i^2, \ldots, y_i^{cm})$, adaptive mutation is defined as follows:

$$z_i^j = \begin{cases} y_i^j + \sigma_j \times (ub_j - lb_j), & if\ rand_i < p_t \\ y_i^j, & otherwise \end{cases} \tag{29}$$

where z_i^j is the j-th variable of the antibody after mutation, $rand_i$ is a random number over interval [0, 1], ub_j and lb_j are the upper and lower limits of the j-th variable of all antibodies in current population, respectively. σ_j is defined by

$$\sigma_j = \begin{cases} (2 \times r_i)^{\frac{1}{\eta+1}} - 1, & if\ r_i < 0.5 \\ 1 - (2 - 2 \times r_i)^{\frac{1}{\eta+1}}, & otherwise \end{cases} \tag{30}$$

where r_i is a random number in [0, 1], η is the mutation distribution parameter and its value is 20 in this paper.

In addition, the mutation probability p_t in the t-th generation population is defined as follows:

$$p_t = \frac{1 + \lambda \times (1 - \frac{t}{T})}{m} \tag{31}$$

where m is the number of decision variables, λ is pre-defined parameter in [0, 1] and its value is 0.5 in this paper, T is the maximum generation.

F. SELECTION OF OPTIMAL SOLUTION

In the last generation of the IMOIA, the approximate Pareto-optimal set A_T is reported. As shown in Figure 4, all the solutions in this front are considered to be equally important. But in practical problems, we must choose a single solution from the Pareto set.

In this paper, we use a semi-supervised method proposed in [17] to select the optimal solution in the obtained Pareto set. Assuming that 10% of the class labels for the entire data set are known, we refer to the data set consisting of data samples with known class labels as the test set. For each solution in Figure 4, the clustering labels of test set are assigned based on membership matrixes.

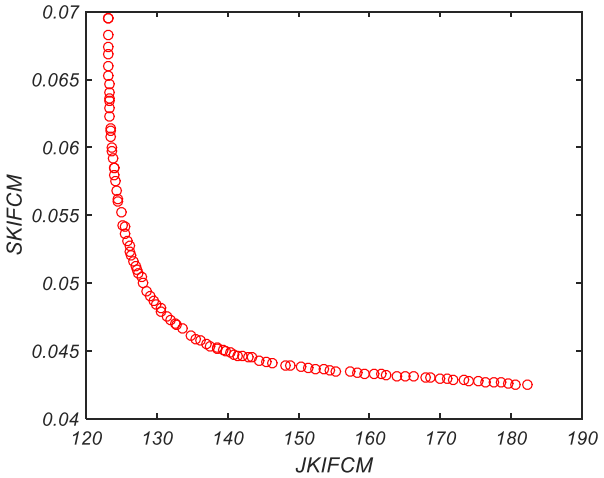

FIGURE 4. A set of solutions obtained in the last generation of KIFCM-IMOIA.

Let G and L represent the clustering label set obtained by our algorithm and the true classification label set, respectively. Then, we select the best solution with the minimum Minkowski score (MS) [17].

$$MS(G, L) = \sqrt{\frac{n_{01} + n_{10}}{n_{11} + n_{10}}} \tag{32}$$

where n_{11} represents the number of pairs of points that are in the same cluster in G and L, n_{01} and n_{10} are the number of pairs in the same cluster only in L and G, respectively.

G. ALGORITHM PROCESS

To show the details of KIFCM-IMOIA, its pseudo-code is provided in Algorithm 2.

KIFCM-IMOIA starts in step 1 by creating an initial antibody population randomly. Then, our algorithm goes into its main loop until the number of iterations reaches the maximum generation T. In step 2, we compute the corresponding objective functions of each antibody in current population according to Eq.(17) and Eq.(20). In step 3, we select non-dominated solutions from current population. In step 4, we select active antibodies from dominant population. If the size of dominant population exceeds the maximum, partial antibodies are selected from dominant population as active antibodies; otherwise, dominant population is the active population. Then, the proportional cloning is used to generate clone population with N antibodies. In step 6, differential evolution strategy and adaptive mutation operator are applied on cloned antibodies to form a new population. Then, the new population and dominant population are combined as next-generation population. The iteration continues until the number of generations reaches maximum. At the end of IMOIA, the approximate Pareto-optimal set A_T is reported. Then, we choose the optimal clustering centers according to our needs in the approximate Pareto-optimal set. In this paper, we select the final solution based on a semi-supervised method as detailed in section 4.6. At the end, the optimal solution, i.e., the set of cluster centers, is reported.

V. EXPERIMENTS AND ANALYSIS

As mentioned before, our algorithm uses improved multi-objective immune algorithm to obtain the optimal solutions.

Algorithm 2 KIFCM-IMOIA

Parameters: T: maximum generation
N : maximum size of population
Input: $X = \{x_k\}_{k=1}^{n}$: the dataset
c: the number of clusters
Output: the set of cluster centers $V = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_c\}$

1. Randomly create an initial antibody population $P_0 = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_N\}$ as detailed in section 4.1;
   while $t < T$ **do**
2. Calculate the objective functions of each antibody in P_t according to Eq.(17) and Eq.(20);
3. Select non-dominated antibodies from P_t to form D_t as detailed in [26];
4. Select active antibodies as following steps:
   If $|D_t| > N$
   $A_t =$ select N active antibodies from D_t by Algorithm 1;
   else
   $A_t = D_t$
   end if
5. Applying proportional cloning on A_t to form C_t as detailed in section 4.3;
6. Apply the hybrid differential evolution strategy and adaptive mutation operator on C_t to form C_t';
7. $P_{t+1} = D_t \cup C_t'$;
   end while
8. Select the optimal solution from A_T as detailed in section 4.6;
9. Decode the optimal solution into the set of cluster centers $V = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_c\}$;
10. **Return** the set of cluster centers $V = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_c\}$.

TABLE 1. UCI data sets used in experiment.

Dataset	Instances	Features	Number of clusters
Hayes-Roth (D1)	160	5	3
Haberman's Survival Data (D2)	306	3	2
Seeds (D3)	210	7	3
Glass Identification (D4)	214	10	6
Ecoli (D5)	336	8	8
Balance Scale (D6)	625	4	3
Optical Recognition of Handwritten Digits (1/ 2/ 7/ 9) (D7)	718	64	4
Contraceptive Method Choice (D8)	1473	9	3
Letter Recognition (A/ B) (D9)	1555	16	2
Letter Recognition (C/ D) (D10)	1541	16	2
Shuttle (2/3/4/5/6/7) (D11)	4799	9	6
Skin Segmentation (portion) (D12)	5000	3	2
Occupancy Detection (D13)	9752	5	2
Electrical Grid Stability Simulated Data (D14)	10000	14	2

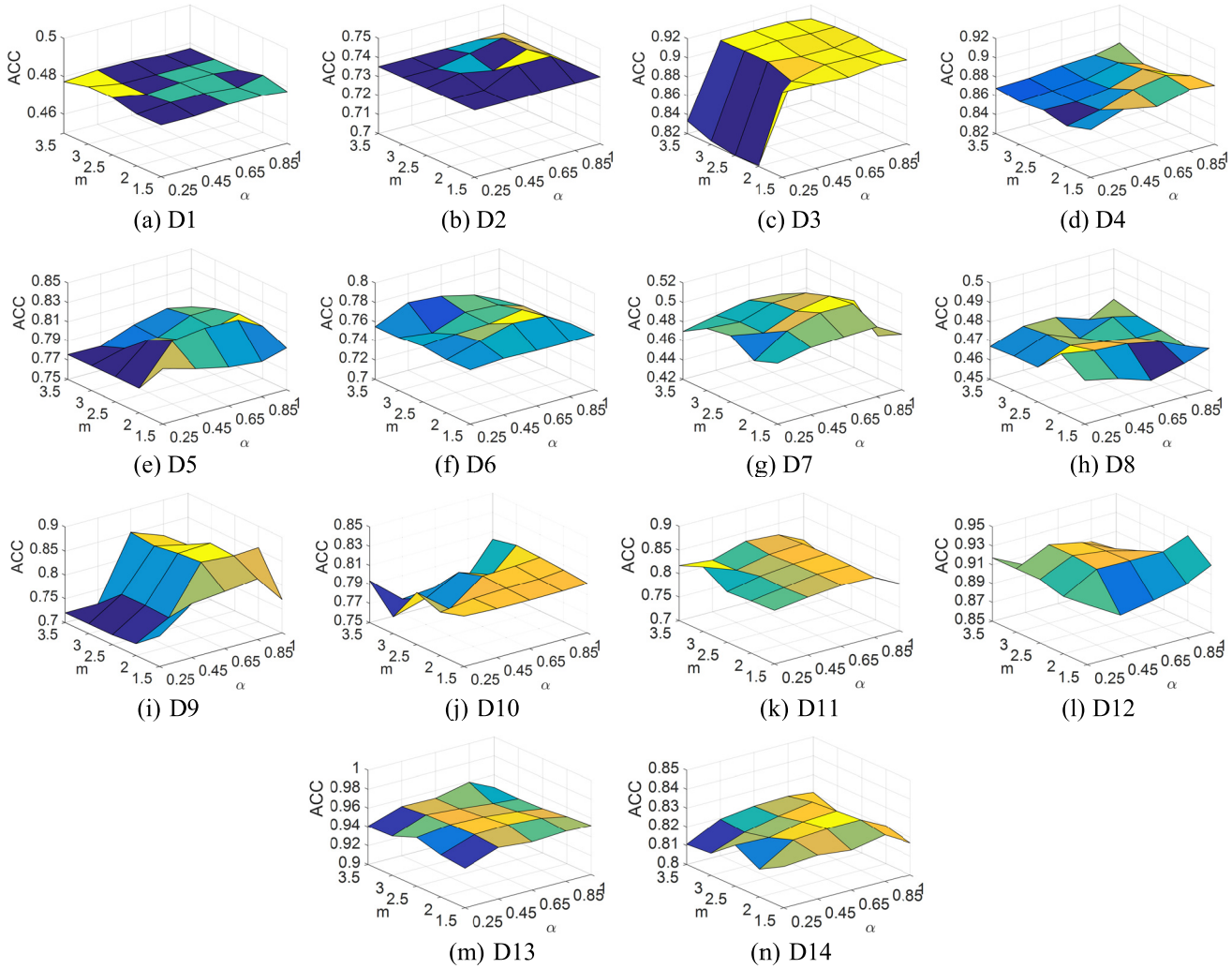

FIGURE 5. *ACC* against the parameters *m* and α on UCI data sets.

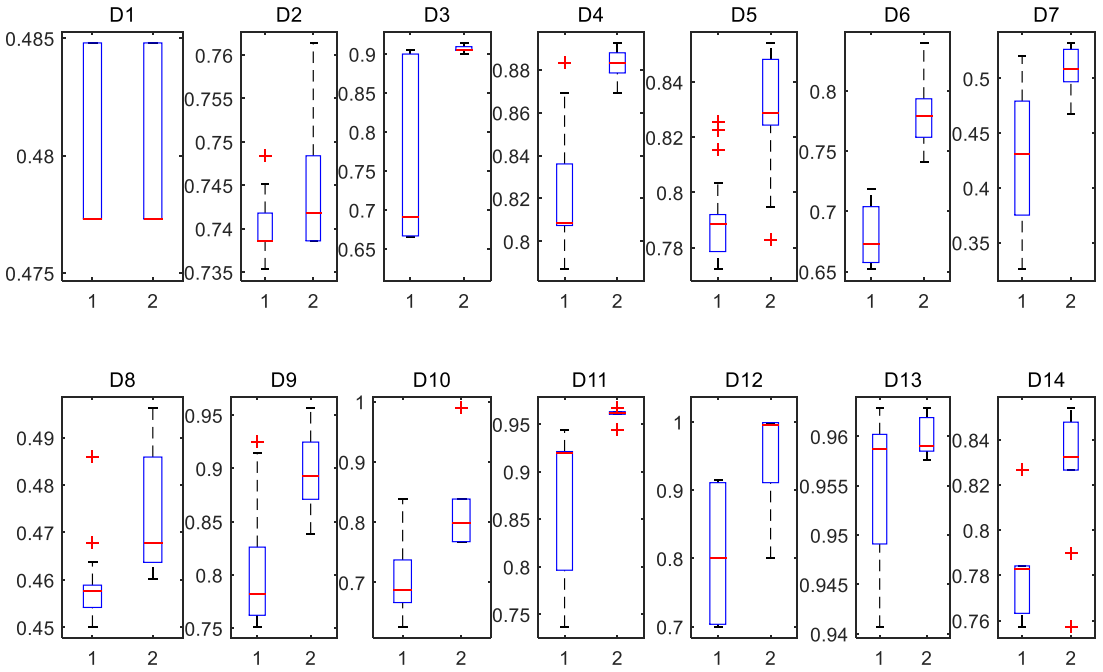

FIGURE 6. Box plots of *ACC* obtained by two algorithms on 14 UCI datasets.

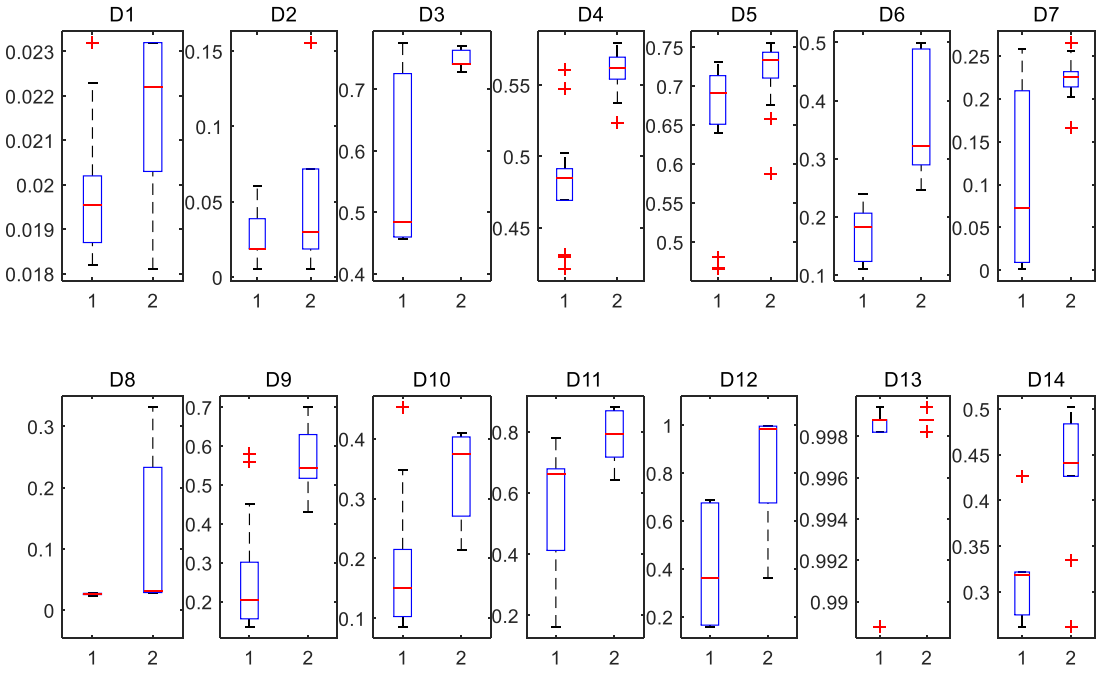

FIGURE 7. Box plots of *ARI* obtained by two algorithms on 14 UCI datasets.

To verify the impact of IMOIA, we compared our algorithm with KIFCM-NNIA. In addition, in order to confirm the performance of our algorithm, we compared our algorithm with five famous clustering algorithms, i.e. AP [9], DBSCAN [10], FCM [11], KFCM [20] and IFCM [12]. Fourteen UCI real datasets are used to compare the performance of our algorithm and comparison algorithms. The details of the data sets are shown in the Table 1.

A. EVALUATION ON METRICS

In this paper, three clustering metrics are used to measure the performance of our algorithm: *ACC* [30], *ARI* [31] and *NMI* [32].

The *ACC* is calculated as:

$$ACC = \frac{\sum_{k=1}^{n} \delta(l_k, map(g_k))}{n} \tag{33}$$

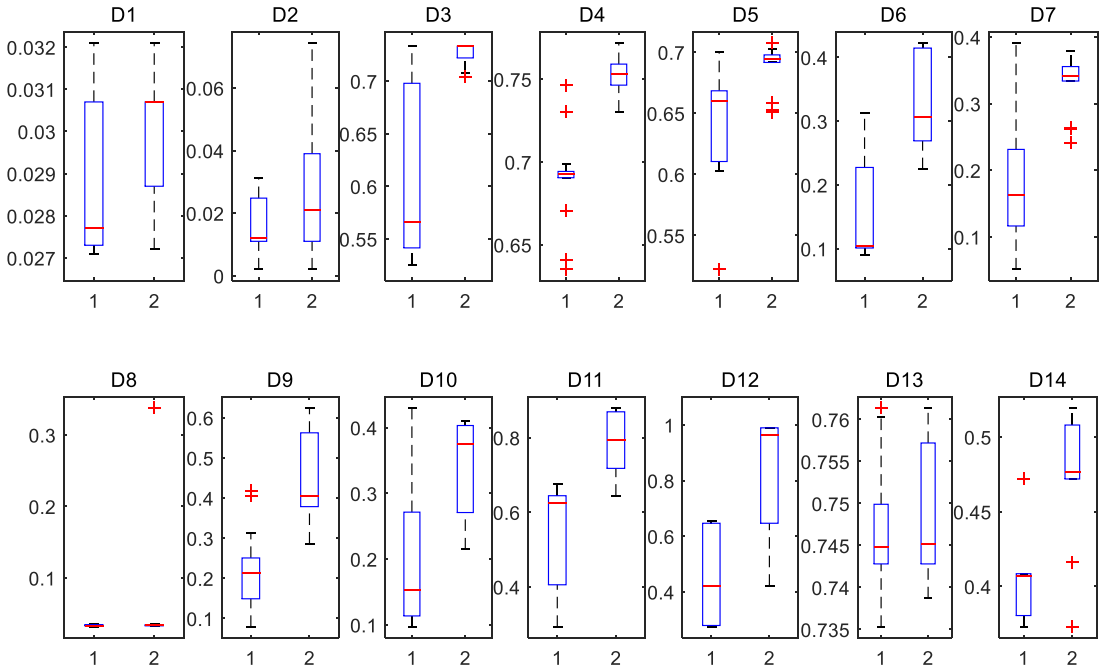

FIGURE 8. Box plots of *NMI* obtained by two algorithms on 14 UCI datasets.

where n represents the number of data samples, g_k and l_k represent the cluster label obtained by our algorithm and the true classification label of the data sample $\mathbf{x}_k$, respectively, $map(\cdot)$ is a mapping function that maps cluster labels obtained by our algorithm to the true classification labels. When $l_k = map(g_k)$, the function value of $\delta(l_k = map(g_k))$ is 1, otherwise it is 0.

Suppose that G and L respectively represent the clustering label set obtained by our algorithm and the true classification label set. The *ARI* is defined by

$$ARI = \frac{RI - E(RI)}{\max(RI) - E(RI)} \tag{34}$$

where *E(RI)* represents the expected value of *RI*. *RI* is the rand index which is defined as:

$$RI = \frac{a+b}{C_2^n} \tag{35}$$

where a represents the number of pairs of points that are in the same cluster in G and L, and b represents the number of pairs of points in different clusters in G and L, n represents the number of samples in the data set, C_2^n indicates the number of pairs that can be composed in the dataset. The range of *ARI* is $[-1, 1]$. The larger the value, the closer the clustering result is to the true classification.

The *NMI* is defined by

$$NMI = \frac{\sum_{i=1}^{c}\sum_{j=1}^{c} n_{ij} \log\left(\frac{n \cdot n_{ij}}{n_i \cdot n_j'}\right)}{\sqrt{\left(\sum_{i=1}^{c} n_i \log\left(\frac{n_i}{n}\right)\right)\left(\sum_{j=1}^{c} n_j' \log\left(\frac{n_j'}{n}\right)\right)}} \tag{36}$$

where c represents the number of clusters, n represents the number of samples in the data set, n_{ij} represents the number of data samples belonging to the *i-th* cluster in the label set obtained by the algorithm and belonging to the *j-th* cluster in the true label set, simultaneously. n_i is the number of data samples belonging to the *i-th* cluster center obtained by the algorithm, and n_j' is the number of data samples belonging to the *j-th* cluster in the real case. *NMI* effectively measures the statistical information between the clustering result distribution obtained by the algorithm and the actual classification label distribution. The range of *NMI* values is [0, 1]. Generally speaking, the larger the *NMI* value, the better the clustering result.

B. ANALYSIS OF PARAMETERS OF KIFCM-IMOIA

In this section, we evaluated the effects of parameters m and α. In FCM [11], the value of m is 2. In IFCM [12], the values of m and α are 2 and 0.85, respectively. In this section, m and α are tested in sets {1.5, 2, 2.5, 3, 3.5} and {0.25, 0.45, 0.65, 0.85, 1}, respectively. Our algorithm is executed 10 times for each UCI dataset to investigate m and α. Figure 5 shows the curved surfaces of *ACC* with the variations of m and α on 14 UCI data sets.

As shown in Figure 5, the values of m and α have no obvious effects on *ACC* values of D1, D2, D4, D6, D7, D8 and D11. For D3, D5, D9, D12, D13 and D14, the values of *ACC* decrease obviously when the value of α is less than 0.5. For D10, the value of *ACC* decreases obviously when the value of m is greater than 2. In the following experiments, the values of m and α are assigned to 2 and 0.65, respectively.

TABLE 2. The average values and their deviations of clustering results of six algorithms on 14 UCI datasets.

Dataset	Algorithm	ACC	ARI	NMI
Hayes-Roth	AP	0.4167	4.96E-05	0.0094
	DBSCAN	0.3864	0	4.31E-16
	FCM	0.4015 (0.0000)	0.0067 (0.0000)	0.0173 (0.0000)
	KFCM	0.4303 (5.68E-05)	0.0179 (7.91E-06)	0.0321 (6.45E-06)
	IFCM	0.4015 (0.0000)	0.0085 (0.0000)	**0.036 (0.0000)**
	KIFCM-IMOIA	**0.4796 (1.24E-05)**	**0.0217 (2.48E-06)**	0.0301 (1.73E-06)
Haberman's Survival Data	AP	0.7353	0.0272	0.0037
	DBSCAN	0.7353	0.1421	0.0096
	FCM	0.7549 (0.0000)	0.1571 (0.0000)	0.0682 (0.0000)
	KFCM	**0.7553 (0.0000)**	**0.1656 (0.0000)**	**0.0698 (0.0000)**
	IFCM	0.7516 (0.0000)	0.1556 (0.0000)	0.0654 (0.0000)
	KIFCM-IMOIA	0.7439 (5.90E-05)	0.0502 (0.0022)	0.0266 (0.0005)
seeds	AP	0.6429	0.4413	0.489
	DBSCAN	0.85	0.4584	0.4706
	FCM	0.8095 (0.0000)	0.5125 (0.0000)	0.4801 (0.0000)
	KFCM	0.8143 (0.0000)	0.5223 (0.0000)	0.4902 (0.0000)
	IFCM	0.7905 (0.0000)	0.4715 (0.0000)	0.4475 (0.0000)
	KIFCM-IMOIA	**0.9075 (1.66E-05)**	**0.7499 (0.0002)**	**0.7268 (0.0001)**
Glass Identification	AP	0.5047	0.2523	0.3654
	DBSCAN	0.3551	0	3.06E-16
	FCM	0.6075 (0.0000)	0.2222 (0.0000)	0.3689 (0.0000)
	KFCM	0.5411 (0.0031)	0.1751 (0.0032)	0.3284 (0.0024)
	IFCM	0.5888 (0.0000)	0.2121 (0.0000)	0.3582 (0.0000)
	KIFCM-IMOIA	**0.8826 (3.87E-05)**	**0.5568 (0.0003)**	**0.7522 (0.0001)**
Ecoli	AP	0.6726	0.5083	0.483
	DBSCAN	0.4435	0.0381	0.1183
	FCM	0.7988 (0.0002)	0.3541 (4.56E-05)	0.5513 (1.24E-05)
	KFCM	0.8126 (0.0005)	0.3598 (0.0004)	0.5604 (6.21E-05)
	IFCM	0.7827 (0.0000)	0.3824 (0.0000)	0.5272 (0.0000)
	KIFCM-IMOIA	**0.8319 (0.0003)**	**0.7163 (0.0024)**	**0.6897 (0.0003)**
Balance Scale	AP	0.7616	0.0873	0.2273
	DBSCAN	0.4608	0	5.05E-16
	FCM	0.6368 (0.0008)	0.1206 (0.0014)	0.1004 (0.0008)
	KFCM	0.7005 (0.0036)	0.2188 (0.0076)	0.1989 (0.0067)
	IFCM	0.6307 (0.0028)	0.1103 (0.0054)	0.0911 (0.0032)
	KIFCM-IMOIA	**0.7816 (0.0009)**	**0.3562 (0.0094)**	**0.3226 (0.0049)**
Optical Recognition of Handwritten Digits (1/2/7/9)	AP	**0.5089**	0.2169	0.2663
	DBSCAN	0.3195	0	3.37E-16
	FCM	0.3195 (0.0000)	0 (0.0000)	3.37E-16 (0.0000)
	KFCM	0.3195 (0.0000)	0 (0.0000)	3.37E-16 (0.0000)
	IFCM	0.3195 (0.0000)	0 (0.0000)	3.37E-16 (0.0000)
	KIFCM-IMOIA	0.5083 (0.0004)	**0.2247 (0.0008)**	**0.3364 (0.0013)**
Contraceptive Method Choice	AP	0.4345	0.0224	0.0201
	DBSCAN	0.427	0	4.33E-16
	FCM	0.4369 (3.52E-07)	0.0149 (2.33E-08)	0.0245 (5.01E-08)
	KFCM	0.4328 (2.46E-05)	0.0046 (0.0002)	0.0180 (6.91E-06)
	IFCM	0.4412 (7.68E-05)	0.0131 (0.0002)	0.0269 (3.58E-05)
	KIFCM-IMOIA	**0.4725 (0.0002)**	**0.0993 (0.0142)**	**0.0642 (0.0086)**
Letter Recognition (A\B)	AP	0.7204	-0.0141	0.0132
	DBSCAN	0.7204	0	7.08E-11
	FCM	0.7204 (0.0000)	-0.008 (0.0000)	3.14E-04 (0.0000)

TABLE 2. *(Continued.)* **The average values and their deviations of clustering results of six algorithms on 14 UCI datasets.**

	KFCM	0.7204 (0.0000)	-0.0011 (0.0000)	0.0029 (0.0000)
	IFCM	0.8495 (0.0000)	0.4807 (0.0000)	0.4355 (0.0000)
	KIFCM-IMOIA	**0.8971 (0.0013)**	**0.5694 (0.0058)**	**0.4509 (0.0101)**
Letter Recognition (C\D)	AP	0.6768	0.1108	0.1789
	DBSCAN	0.5455	0	6.70E-16
	FCM	0.5859 (0.0000)	0.0196 (0.0000)	0.0183 (0.0000)
	KFCM	0.5960 (0.0000)	0.0271 (0.0000)	0.0232 (0.0000)
	IFCM	0.6061 (0.0000)	0.0353 (0.0000)	0.0304 (0.0000)
	KIFCM-IMOIA	**0.8296 (0.0059)**	**0.3623 (0.0038)**	**0.3436 (0.0047)**
Shuttle (2/3/4/5/6/7)	AP	0.7037	0.2115	0.2608
	DBSCAN	0.5122	0.0358	0.0441
	FCM	0.9554 (0.0000)	0.3300 (0.0000)	0.4986 (0.0000)
	KFCM	0.8641 (2.94E-05)	0.1262 (0.0021)	0.2257 (0.0022)
	IFCM	0.9539 (0.0000)	0.3281 (0.0000)	0.4942 (0.0000)
	KIFCM-IMOIA	**0.9598 (5.54E-05)**	**0.8281 (0.0125)**	**0.7833 (0.0087)**
Skin Segmentation (portion)	AP	0.516	3.30E-04	0.0133
	DBSCAN	0.8712	0.0228	0.1998
	FCM	0.5008 (0.0000)	-1.91E-04 (0.0000)	1.93E-06 (0.0000)
	KFCM	0.5048 (0.0000)	-1.01E-04 (0.0000)	6.99E-05 (0.0000)
	IFCM	0.5002 (0.0000)	-1.93E-04 (0.0000)	1.21E-07 (0.0000)
	KIFCM-IMOIA	**0.9432 (0.0048)**	**0.8041 (0.0517)**	**0.7966 (0.04557)**
Occupancy Detection	AP	0.8327	0.6682	0.2052
	DBSCAN	0.9519	0.0966	0.3734
	FCM	0.9477 (0.0000)	0.8966 (0.0000)	0.63 (0.0000)
	KFCM	0.9501 (6.60E-05)	0.9128 (0.0017)	0.6492 (0.0025)
	IFCM	0.9482 (0.0000)	0.9006 (0.0000)	0.6334 (0.0000)
	KIFCM-IMOIA	**0.9598 (3.67E-06)**	**0.9989 (1.08E-07)**	**0.7486 (6.31E-05)**
Electrical Grid Stability Simulated Data	AP	0.7091	0.0147	0.0512
	DBSCAN	0.6380	0	7.05E-16
	FCM	0.6380 (0.0000)	0.0345 (0.0010)	0.0282 (0.0007)
	KFCM	0.6395 (9.81E-06)	0.0367 (0.0013)	0.0299 (0.0009)
	IFCM	0.6380 (0.0000)	0.0256 (0.0008)	0.0212 (0.0005)
	KIFCM-IMOIA	**0.8276 (0.0007)**	**0.4316 (0.0045)**	**0.4750 (0.0017)**

C. RESULTS FROM KIFCM-IMOIA AND KIFCM-NNIA

To verify the impact of IMOIA, we compared our algorithm with KIFCM-NNIA on 14 UCI datasets. In each algorithm, the maximum generation T is 100 and the maximum size of population N is 100.

Figure 6–Figure 8 show the box plots of two algorithms for three metrics, i.e., *ACC*, *ARI* and *NMI*. In each plot, the left box represents the result of KIFCM-NNIA and the right box represents the result of KIFCM-IMOIA; the median value of the metrics is indicated by the red line at the center of the box; the top line of the box is the position of the third quartile and the bottom line is the position of the first quartile; the upper and lower limits of the whiskers represent the maximum and minimum values of the metric, respectively; the outlier is represent by the red symbol "+". Note that a higher and more compact box indicates better clustering results.

As shown in Figure 6, the boxes of the two algorithms are consistent on D1. In the other 13 data sets, the performance of our algorithm is obviously better than that of KIFCM-NNIA. Especially for D6, the minimum *ACC* of our algorithm is higher than the maximum *ACC* of KIFCM-NNIA.

As shown in Figure 7 and Figure 8, for D1, D3 and D10, the maximum *ARI* obtained by KIFCM-NNIA is higher than that obtained by KIFCM-IMOIA. For D1, D7 and D10, the maximum *NMI* obtained by KIFCM-NNIA is higher than that obtained by KIFCM-IMOIA. However, the boxes obtained by our algorithm are lower. For other data sets, KIFCM-IMOIA performs better. The reason for this phenomenon is that IMOIA uses a novel active antibody selection strategy, a hybrid differential evolution strategy and an adaptive mutation operator to maintain a better distribution of solutions with better convergence.

D. COMPARISON OF RESULTS WITH OTHER CLUSTERING ALGORITHMS

In this experiment, we compared our algorithm with five famous clustering algorithms based on three

TABLE 3. T-test results on 14 UCI datasets.

Data set	Std. Err.			T value			95% Conf. Intvl			Two-tailed P		
	ACC	ARI	NMI	ACC	ARI	NMI	ACC	ARI	NMI	ACC	ARI	NMI
D1	0.002	0.001	0.001	32.443	6.453	-3.772	[0.0462, 0.0523]	[0.0026, 0.0050]	[-0.0030, -0.0009]	0.023	0.009	0.004
D2	0.001	0.009	0.004	-8.125	-13.471	-10.854	[-0.0142, -0.0859]	[-0.1326, -0.0983]	[-0.0512, -0.0352]	0.000	0.000	0.000
D3	0.001	0.003	0.002	131.616	93.645	102.345	[0.0965, 0.0995]	[0.2322, 0.2426]	[0.2349, 0.2446]	0.000	0.000	0.000
D4	0.001	0.003	0.002	242.145	108.037	199.966	[0.2728, 0.2774]	[0.3283, 0.3410]	[0.3794, 0.3872]	0.000	0.000	0.000
D5	0.005	0.010	0.004	3.699	36.918	36.718	[0.0089, 0.0299]	[0.3371, 0.3758]	[0.1223, 0.1364]	0.000	0.000	0.000
D6	0.005	0.017	0.013	3.753	15.196	7.461	[0.0091, 0.0309]	[0.2327, 0.3051]	[0.0692, 0.1215]	0.000	0.000	0.000
D7	0.003	0.005	0.007	-0.179	1.522	10.652	[-0.0076, 0.0064]	[-0.0027, 0.0183]	[0.0567, 0.0836]	0.859	0.139	0.000
D8	0.003	0.022	0.017	11.252	3.939	2.194	[0.0258, 0.0370]	[0.0415, 0.1310]	[0.0025, 0.0720]	0.000	0.001	0.036
D9	0.007	0.014	0.018	7.336	6.389	0.839	[0.0344, 0.0609]	[0.0603, 0.1171]	[-0.0222, 0.0530]	0.000	0.000	0.408
D10	0.005	0.011	0.013	25.269	22.304	13.103	[0.1133, 0.1332]	[0.2284, 0.2745]	[0.1390, 0.1904]	0.000	0.000	0.000
D11	0.001	0.020	0.017	3.251	24.451	16.746	[0.0016, 0.0072]	[0.4565, 0.5398]	[0.2499, 0.3194]	0.003	0.000	0.000
D12	0.013	0.042	0.039	5.701	18.812	15.300	[0.0462, 0.0978]	[0.6964, 0.8662]	[0.5171, 0.6766]	0.000	0.000	0.000
D13	0.002	0.008	0.009	6.406	11.337	10.762	[0.0067, 0.0129]	[0.0705, 0.1015]	[0.0806, 0.1183]	0.000	0.000	0.000
D14	0.005	0.012	0.008	24.484	33.880	56.025	[0.1086, 0.1284]	[0.3917, 0.4421]	[0.4083, 0.4393]	0.000	0.000	0.000

clustering metrics. To make the comparison fare, the maximum number of iterations for each compared algorithm is 100. In this paper, each algorithm is executed 30 times for each dataset. The evaluation metrics are obtained from the output at the end of each run. Table 2 shows the average values and their deviations of clustering results of six algorithms on 14 UCI datasets.

As shown in Table 2, the clustering results of KIFCM-IMOIA are better than other algorithms on 11 UCI datasets. For Hayes-Roth, our algorithm performs slightly worse than IFCM and KFCM on *NMI*. For Haberman's Survival Data, our algorithm performs worse than KFCM on three clustering metrics. For Optical Recognition of Handwritten Digits, our algorithm performs slightly worse than AP on *ACC*. However, for other data sets, our algorithm performs significantly better than other algorithms. Especially for seeds, Glass Identification, Letter Recognition (C\D), Skin Segmentation and Electrical Grid Stability Simulated Data, the improvement of our algorithm is obvious. As can be seen from Table 2, the improvement rates of *ACC* are about 10%, the *ARI* improvement rates are approximately 20%, and the improvement rates of *NMI* are over 10% on these five data sets.

E. T-TEST RESULTS

In this section, we first select the best method from five well-known clustering algorithms according to Table 2. Then, we compared the numerical results of KIFCM-IMOIA and the selected best method using the t-test. In our experiment, the sample size is set to 30. Table 3 shows the t-test results, including the standard error of the difference, the T value, the 95% confidence interval and the two-tailed P value.

By performing the independent-samples T test on the results of KIFCM-IMOIA and the selected best method, it can be inferred whether there are significant differences between the two algorithms. In Table 3, the positive T value indicates that our algorithm performs better than other algorithms. If the two-tailed P value is less than 0.05, it indicates that there is significant difference between the result of our algorithm and that of the compared algorithm. If the 95% confidence

TABLE 4. Computational complexity of the seven algorithms.

Algorithm	AP	DBSCAN	FCM	KFCM	IFCM	KIFCM-NNIA	KIFCM-IMOIA
Time complexity	$O(Tn^3)$	$O(nlog(n))$	$O(Tcn)$	$O(Tcn)$	$O(Tcn)$	$O(TNcn)$	$O(TNcn)$

interval is to the right of zero, then our algorithm performs significantly better than the comparison algorithms.

As shown in Table 3, the T values are negative of three clustering metrics on D2 which means that our algorithm is performing poorly on this dataset. For D1, the T value is negative of *NMI* and the two-tailed P value is less than 0.05 which means that our algorithm performs slightly worse than the compared algorithms on *NMI*. For D7, the value in the 95% confidence interval is on the left side of zero and the two-tailed P value is greater than 0.05, which means that our algorithm performs worse than the compared algorithms, but the difference is not obvious. For D9, the T value is positive of *NMI* and the two-tailed P value is less than 0.05, which means that our algorithm performs better than the compared algorithms, but the difference is not obvious. For other datasets, our proposed algorithm can achieve significant clustering performance. All two-tailed P values are less than 0.05, so the numerical differences of KIFCM-IMOIA and the best method among other five compared algorithms are statistically significant. In other words, the performance of our algorithm is significantly improved compared with other algorithms.

F. COMPUTATIONAL COMPLEXITY

In this subsection, we will discuss the time complexity of seven algorithms. Assuming that N is maximum size of population, n is the size of dataset, c is the number of cluster centers. The steps to calculate the time complexity of KIFCM-IMOIA are as follows:

1) In the initialization step of antibody population, the time complexity is $O(N)$.
2) The time complexity of computing objective functions of antibodies in population is $O(Ncn)$.
3) In the step of identifying non-dominated antibodies in the population, the time complexity is $O((N + N_D)^2)$, where N_D is the number of non-dominated antibodies in previous generation population.
4) In the step of selecting active antibodies from current population, the time complexity is $O(N + N_D)$.
5) The time complexity of cloning, differential evolution, and mutation is $O(N)$.

Therefore, the worst time complexity of one generation for KIFCM-IMOIA can be simplified as $O(Ncn)$. Assuming that T is the maximum number of generations, the time complexity becomes $O(TNcn)$.

Table 4 summarizes the computational complexity of the seven algorithms. The time complexity of KIFCM-IMOIA is worse than that of FCM, KFCM and IFCM, but they are in the same order of magnitude. The time complexity of AP and DBSCAN are $O(Tn^3)$ and $O(nlog(n))$, respectively, which are not as good as KIFCM-IMOIA. The time complexity of KIFCM-NNIA is $O(TNcn)$, which is same with KIFCM-IMOIA.

VI. CONCLUTION

In order to prevent the algorithm from falling into local optimum and improve robustness to noises, we present KIFCM-IMOIA. Our algorithm combines kernel method and multi-objective immune optimization algorithm with IFCM. The kernel method projects data into higher dimensional spaces, which improves robustness to noises. In addition, the multi-objective immune optimization algorithm considers the separation between clusters and the compactness within clusters simultaneously. This operation helps the proposed algorithm to avoid falling into local optimum. First, to verify the impact of IMOIA, we compared our algorithm with KIFCM-NNIA. Experimental results show that IMOIA can maintain a better distribution of solutions with better convergence. Then, extensive experiments were performed to compare the performance of our algorithm with five famous clustering algorithms, i.e. AP [9], DBSCAN [10], FCM [11], KFCM [20] and IFCM [12], on 14 UCI data sets. Experimental results show that KIFCM-IMOIA performs better than other algorithms on three clustering metrics, including the clustering accuracy, adjusted rand index and normalized mutual index.

REFERENCES

[1] M. R. N. Kalhori and M. H. F. Zarandi, "Interval type-2 credibilistic clustering for pattern recognition," *Pattern Recognit.*, vol. 48, pp. 3652–3672, Nov. 2015.
[2] X. Huang, Y. Ye, L. Xiong, R. Y. K. Lau, N. Jiang, and S. Wang, "Time series k-means: A new k-means type smooth subspace clustering for time series data," *Inf. Sci.*, vol. 367, pp. 1–13, Nov. 2016.
[3] I. Saha and U. Maulik, "Incremental learning based multiobjective fuzzy clustering for categorical data," *Inf. Sci.*, vol. 267, pp. 35–57, May 2014.
[4] H. Verma, R. K. Agrawal, and A. Sharan, "An improved intuitionistic fuzzy C-means clustering algorithm incorporating local information for brain image segmentation," *Appl. Soft Comput.*, vol. 46, pp. 543–557, Sep. 2016.
[5] M. Zhang, L. Jiao, W. Ma, J. Ma, and M. Gong, "Multi-objective evolutionary fuzzy clustering for image segmentation with MOEA/D," *Appl. Soft Comput.*, vol. 48, pp. 621–637, Nov. 2016.
[6] F. Zhao, Z. Zeng, H. Q. Liu, and J. L. Fan, "A Kriging-assisted reference vector guided multi-objective evolutionary fuzzy clustering algorithm for image segmentation," *IEEE Access*, vol. 7, pp. 21465–21481, 2019.
[7] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, Jun. 2014.
[8] M. R. P. Ferreira and F. de A. T. de Carvalho, "Kernel-based hard clustering methods in the feature space with automatic variable weighting," *Pattern Recognit.*, vol. 47, pp. 3082–3095, Sep. 2014.
[9] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, Feb. 2007.
[10] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2nd Int. Conf. Knowl. Discovery Data Mining*, 1996, pp. 226–231.

[11] J. C. Bezdek, W. Full, and R. Ehrlich, "FCM: The fuzzy c-means clustering algorithm," *Comput. Geosci.*, vol. 10, nos. 2–3, pp. 191–203, 1984.
[12] T. Chaira, "A novel intuitionistic fuzzy C means clustering algorithm and its application to medical images," *Appl. Soft Comput.*, vol. 11, no. 2, pp. 1711–1717, 2011.
[13] Y. Wang, X. Duan, X. Liu, C. Wang, and Z. Li, "A spectral clustering method with semantic interpretation based on axiomatic fuzzy set theory," *Appl. Soft Comput.*, vol. 64, pp. 59–74, Mar. 2018.
[14] X. Liu, X. Yong, and H. Lin, "An improved spectral clustering algorithm based on local neighbors in kernel space," *Comput. Sci. Inf. Syst.*, vol. 8, pp. 1143–1157, Oct. 2011.
[15] J. Handl and J. Knowles, "An evolutionary approach to multiobjective clustering," *IEEE Trans. Evol. Comput.*, vol. 11, no. 1, pp. 56–76, Feb. 2007.
[16] D. W. Corne, N. R. Jerram, J. D. Knowles, and M. J. Oates, "PESA-II: Region-based selection in evolutionary multiobjective optimization," in *Proc. 3rd Annu. Conf. Genetic Evol. Comput.*, 2001, pp. 283–290.
[17] S. Wikaisuksakul, "A multi-objective genetic algorithm with fuzzy *c*-means for automatic data clustering," *Appl. Soft Comput.*, vol. 24, pp. 679–691, Nov. 2014.
[18] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.
[19] C.-L. Yang, R. J. Kuo, C.-H. Chien, and N. T. P. Nguyen, "Non-dominated sorting genetic algorithm using fuzzy membership chromosome for categorical data clustering," *Appl. Soft Comput.*, vol. 30, pp. 113–122, May 2015.
[20] D.-Q. Zhang and S.-C. Chen, "Clustering incomplete data using kernel-based fuzzy C-means algorithm," *Neural Process. Lett.*, vol. 18, pp. 155–162, Dec. 2003.
[21] H. C. Huang, Y. Y. Chuang, and C. S. Chen, "Multiple kernel fuzzy clustering," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 1, pp. 120–134, Feb. 2012.
[22] Z. Zhou and S. Zhu, "Kernel-based multiobjective clustering algorithm with automatic attribute weighting," *Soft Comput.*, vol. 22, pp. 3685–3709, Jun. 2018.
[23] C. Han, L. Wang, Z. Zhang, J. Xie, and Z. Xing, "A multi-objective genetic algorithm based on fitting and interpolation," *IEEE Access*, vol. 6, pp. 22920–22929, 2018.
[24] W. Zang, W. Zhang, Z. Wang, D. Jiang, X. Liu, and M. Sun, "A novel double-strand DNA genetic algorithm for multi-objective optimization," *IEEE Access*, vol. 7, pp. 18821–18839, 2019.
[25] R. E. Haber, G. Beruvides, R. Quiza, and A. Hernandez, "A simple multi-objective optimization based on the cross-entropy method," *IEEE Access*, vol. 5, pp. 22272–22281, 2017.
[26] M. Gong, L. Jiao, H. Du, and L. Bo, "Multiobjective immune algorithm with nondominated neighbor-based selection," *Evol. Comput.*, vol. 16, no. 2, pp. 225–255, 2008.
[27] Q. Lin, J. Chen, Z.-H. Zhan, W.-N. Chen, C. A. C. Coello, Y. Yin, C.-M. Lin, and J. Zhang, "A hybrid evolutionary immune algorithm for multiobjective optimization problems," *IEEE Trans. Evol. Comput.*, vol. 20, no. 5, pp. 711–729, Oct. 2016.
[28] Y. Qi, Z. Hou, M. Yin, H. Sun, and J. Huang, "An immune multi-objective optimization algorithm with differential evolution inspired recombination," *Appl. Soft Comput.*, vol. 29, pp. 395–410, Apr. 2015.
[29] Z. Zhang, X. Wang, and J. Lu, "Multi-objective immune genetic algorithm solving nonlinear interval-valued programming," *Eng. Appl. Artif. Intell.*, vol. 67, pp. 235–245, Jan. 2018.
[30] R. Shang, W. Zhang, F. Li, L. Jiao, and R. Stolkin, "Multi-objective artificial immune algorithm for fuzzy clustering based on multiple kernels," in *Proc. IEEE Symp. Ser. Comput. Intell.*, Nov./Dec. 2017, pp. 1–8.
[31] N. X. Vinh, J. Epps, and J. Bailey, "Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance," *J. Mach. Learn. Res.*, vol. 11, pp. 2837–2854, Jan. 2010.
[32] Z. He, X. Xu, and S. Deng, "*k*-ANMI: A mutual information based clustering algorithm for categorical data," *Inf. Fusion*, vol. 9, pp. 223–233, Apr. 2008.
[33] E. Szmidt and J. Kacprzyk, "Entropy for intuitionistic fuzzy sets," *Fuzzy Sets Syst.*, vol. 118, no. 3, pp. 467–477, 2001.
[34] J. D. Schaffer, "Multiple objective optimization with vector evaluated genetic algorithms," in *Proc. 1st Int. Conf. Genetic Algorithms*, 1985, pp. 93–100.
[35] J.-J. Guo and P. B. Luh, "Selecting input factors for clusters of Gaussian radial basis function networks to improve market clearing price prediction," *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 665–672, May 2003.
[36] S. Das and P. N. Suganthan, "Differential evolution: A survey of the state-of-the-art," *IEEE Trans. Evol. Comput.*, vol. 15, no. 1, pp. 4–31, Feb. 2011.

WENKE ZANG received the M.S. and Ph.D. degrees from Shandong Normal University, China, in 2005 and 2018, respectively, where he is currently an Associate Professor. His research interests include machine learning, data mining, and service science.

ZEHUA WANG is currently pursuing the master's degree with the School of Business, Shandong Normal University, China. Her research interests include artificial intelligence, genetic algorithm, data mining, and machine learning.

DONG JIANG is currently pursuing the master's degree with the School of Business, Shandong Normal University, China. Her research interests include machine learning, genetic algorithm, data mining, and artificial intelligence.

XIYU LIU received the Ph.D. degree in mathematical sciences from Shandong University, in 1990. He is currently a Professor, the Doctorial Supervisor, and the Dean of the School of Management Science and Engineering, Shandong Normal University, China. His currently research interests include membrane computing and data mining.