



# Kernel intuitionistic fuzzy c-means and state transition algorithm for clustering problem

Xiaojun Zhou<sup>1,2</sup> · Rundong Zhang<sup>1</sup> · Xiangyue Wang<sup>1</sup> · Tingwen Huang<sup>3</sup> · Chunhua Yang<sup>1</sup>

Published online: 2 April 2020  
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

## Abstract

Clustering problems widely exist in machine learning, pattern recognition, image analysis and information sciences, etc. Although many clustering algorithms have been proposed, it is unpractical to find a clustering algorithm suitable for all types of datasets. Fuzzy c-means (FCM) is one of the most frequently-used fuzzy clustering algorithm for the reason that it is efficient, straightforward, and easy to implement. However, the traditional FCM taking Euclidean distance as similarity measurement can not distinguish the intersection between two clusters. Therefore, kernel function has been taken as similarity measurement to solve this issue. As a comprehensive partition criterion, intuitionistic fuzzy set which consider both membership degree and non-membership degree has been used to replace traditional fuzzy set to describe the natural attributes of objective phenomena more delicately. Thus, Kernel intuitionistic fuzzy c-means (KIFCM) has been proposed in this paper to settle clustering problem. Considering FCM is easily getting trapped in local optima due to its high sensitivity to initial centroid. State Transition Algorithm (STA) has been adopted in this study to obtain the initial centroid to enhance its stability. The proposed STA-KIFCM compared with some other clustering algorithms are implemented using five benchmark datasets. Experimental results not only show that the proposed method is efficient and can reveal encouraging results, but also indicate that the proposed method can achieve high accuracy.

**Keywords** Clustering problem · State transition algorithm · Kernel function · Intuitionistic fuzzy set · Fuzzy c-means

## 1 Introduction

Clustering problems exist in many areas including machine learning, pattern recognition, image analysis and information sciences, etc (Sipe 2001; Fouche and Langit 2011; Bastanlar and Ozuysal 2014). With the growing interest in automatically understanding, processing and summarizing data, many application domains have employed various clustering algorithms to identify patterns within a dataset. Clustering is the process of assigning data objects into a set of disjoint groups called clusters so that objects in each cluster are

more similar to each other than objects from different clusters. Clustering algorithms work by assigning objects to a group if they show a high level of similarity and by assigning objects to different groups if they are distinguished from each other. The common used clustering algorithms can be broadly classified as Hard and Fuzzy (Hathaway and Bezdek 1995). C-means is one of the most popular hard clustering algorithms which partition data objects into  $c$  clusters, where the number of clusters,  $c$ , is decided in advance according to application purposes. This model is inappropriate for real data sets where there are no definite boundaries between the clusters. Using the distance-based (DB) method can find the appropriate number of clusters in advance (Davies and Bouldin 1979). Fuzzy c-means (FCM) is one of the most popular unsupervised fuzzy clustering algorithm (Wang 1983; Zhang and Chen 2004). Since FCM is easily implemented and has obtained satisfactory results in many applications, it has become an important tool for clustering analysis (Zhou et al. 2014; Askari et al. 2017). Furthermore, due to the increasing complexity of the clustering environment and lack of knowledge about the domain of clustering prob-

Communicated by V. Loia.

**B** Rundong Zhang  
zrd1018@csu.edu.cn

- <sup>1</sup> School of Automation, Central South University, Changsha 410083, China
- <sup>2</sup> The Peng Cheng Laboratory, Shenzhen, Changsha 51800, China
- <sup>3</sup> Texas A&M University at Qatar, PO Box 23874, Doha, Qatar

lem in the process of distance measurement under an fuzzy environment, the kernel-based fuzzy c-means (KFCM) algorithm replaces the Euclidean distance metric which is widely used in previous literatures with a kernel metric. The kernel function is applied in order to achieve better mapping for nonlinear separable datasets. Previous studies have shown that the KFCM algorithm performs better than the FCM algorithm. FCM algorithm considers the degree of membership as studied in many literatures (Liu and Xu 2008; Zhang and Chen 2003; Atanassov and Rangasamy 1986). IFCM not only thinks about the degree of membership, but also considers the non-membership of the point in dataset. The IFCM algorithm can achieve better performance compared with the FCM algorithm (Chaira 2011). For example, the intuitionistic fuzzy c-means (IFCM) algorithm was applied in applications to medical images in Bezdek et al. (1984). It is more reasonable to replace fuzzy measurement with intuitionistic metric. Since kernel intuitionistic fuzzy c-means (KIFCM) is an extension of both KFCM and IFCM, this study replaces the Euclidean distance with kernel function and substitutes the fuzzy measurement with the intuitionistic fuzzy metric. KIFCM can describe the natural properties of clustering phenomena more delicately and comprehensively (Ghosh and Dubey 2013). In KIFCM, the goal is to minimize the sum of distances between data points and their respective parent cluster centroid, taking into account the similarity between elements and corresponding clustering center in each cluster. KIFCM has been widely studied in many literatures because of the nice recognition performance and robustness (Papa-georgiou and Iakovidis 2013). Although KIFCM algorithm appears good performance in fuzzy clustering problems, but there still exists some revealed deficiency in motivating the proposal of alternative approaches for fuzzy clustering. Such as initialization with randomly generated centers, high probability of getting trapped into local minima and low accuracy. Therefore, this study intends to improve the KIFCM algorithm by employing global search optimization algorithm.

Many global optimization algorithms such as genetic algorithm (GA) and particle swarm optimization (PSO) have been successfully employed in many clustering applications in previous literatures (Whitley 1994; Hosseinabadi et al. 2019; Kuo et al. 2016; Filho et al. 2015; Deng et al. 2019). They aim to solve optimization problems without being trapped into local minima. The state transition algorithm is a novel intelligent global search algorithm for many optimization problems due to its versatility and simplicity (Li et al. 2012; Zhou et al. 2012; Huang et al. 2018; Zhou et al. 2018; Huang et al. 2018; Zhou et al. 2018). In this paper, STA and KIFCM are combined together to improve the accuracy and obtain better clustering centers. Five benchmark datasets are implemented to compare the proposed method with other methods such as FCM, IFCM, KFCM and KIFCM. The experimental results indicate that the proposed hybrid method achieve better per-

formance and realize higher accuracy than other methods. The novelty and contributions of this paper are listed below:

- (i) This study applies the intuitionistic fuzzy set and kernel function together to settle clustering problems. Using the intuitionistic fuzzy set can partition clustering data points reasonably. Taking the kernel function as the distance metric can distinguish the overlapping clustering data points accurately.
- (ii) This study designs STA to avoid getting trapped into local optima. STA has good performance both in wide exploration and deep exploration which is the trade-off between local and global search.
- (iii) Combining STA with the promising KIFCM algorithm can increase accuracy obviously. Accuracy is the most important index in evaluating the performance of clustering algorithm in clustering problems. This study firstly intends to employ STA and KIFCM to manage clustering problems.

The remainder of this study is organized as follows. Section 2 presents the basic theory of FCM algorithm. Section 3 presents the hybrid method. The experimental results are presented in Sect. 4. Finally, Sect. 5 concludes this study.

## 2 Theoretical basis of FCM algorithm

The basic theoretical knowledge necessary for this study are presented in this section. This section includes two parts. The first part includes distance-based (DB) c-means algorithm used to determine the number of clusters. Another part includes KIFCM algorithm which combines the intuitionistic fuzzy sets and kernel function into the conventional FCM.

### 2.1 Distance-based c-means algorithm

The conventional clustering algorithm partitions set of  $n$  objects  $X = \{x_1, x_2, \dots, x_n\}$  in  $R^d$  dimensional space into  $c$  ( $2 \leq c \leq n$ ) fuzzy clusters with clustering centers  $C = \{c_1, c_2, \dots, c_c\}$ . Determining the number of clusters is a significant preliminary study before applying the proposed hybrid method to get the best clustering centers. The DB c-means algorithm aims to divide the dataset into several clusters and calculate the DB index for each cluster. The best number of clusters can be obtained by taking the ratio of intra-cluster distance to inter-cluster distances as metric. If a dataset can be divided into  $k$  clusters, the DB value can be calculated as follows:

$$DB(k) = \frac{1}{k} \sum_{i=1}^k D_i \quad (1)$$

where  $D_i$  is defined as below:

$$D_i = \max_{i \neq j} \left\{ \frac{\bar{d}_i + \bar{d}_j}{d_{ij}} \right\} \tag{2}$$

here,  $d_{ij}$  represents the inter-distance which is the distance between centroid  $i$  and  $j$ , but  $\bar{d}_i$  represents the intra-distance which is the average distance between each point in cluster  $i$  to the centroid of cluster  $i$ .

Figure 1 shows the result of evaluating the DB value corresponding to 2–10 clusters. It applies c-means algorithm which begins with a random initial centroid to cluster the Old Faithful Geysers (OFG) dataset. At each trial, the results might be different. But the error of evaluation can be reduced when 20 repetitions are conducted. The result shows that the smallest DB value is obtained at 3.

### 2.2 Intuitionistic and kernel FCM algorithm

In this section, the related background necessary for this study will be introduced. This includes the intuitionistic fuzzy logic and kernel FCM algorithm.

#### 2.2.1 Intuitionistic fuzzy logic

Fuzzy set theory, proposed by Zadeh, has been successfully applied in various fields (Zadeh 1999). The theory states that the degree of membership of an element to a fuzzy set is a single value between 0 and 1. But in Atanassov’s intuitionistic fuzzy set, both the degree of membership  $\mu(x)$  and the degree of non-membership  $\nu(x)$  are considered (Zhao et al. 2010). The intuitionistic fuzzy logic can describe the natural attributes of objective phenomena more delicately and comprehensively. In the intuitionistic fuzzy set, an object  $x \in A$  consists of the degree of membership  $\mu_A(x)$  and the degree of non-membership  $\nu_A(x)$ , where  $\mu(x) \in [0, 1]$ ,  $\nu(x) \in [0, 1]$  and  $0 \leq \mu_A(x) + \nu_A(x) \leq 1$ . If  $\mu_A(x) = 1 - \nu_A(x)$ ,  $\forall x \in A$ , then  $A$  degrades into a fuzzy set. In intuitionistic fuzzy set, there is a hesitation degree  $\pi_A(x)$ , defined as below:

$$\pi_A(x) = 1 - \mu_A(x) - \nu_A(x) \tag{3}$$

Obviously,  $\pi_A(x) \in [0, 1]$ . Therefore, the degree of membership in the IFCM algorithm is calculated as below:

$$\mu_{ij}^* = \mu_{ij} + \pi_{ij} \tag{4}$$

In this paper, Yager’s intuitionistic fuzzy complement is taken as the intuitionistic fuzzy complement which can be written as below:

$$\pi_{ij} = 1 - \mu_{ij} - \left(1 - \mu_{ij}^\alpha\right)^{1/\alpha}, \quad -1 < \alpha < 0 \tag{5}$$

where  $\pi_{ij}$  is Yager’s hesitation degree in the IFCM algorithm. According to Yager’s intuitionistic fuzzy complement,  $\nu_{ij} = (1 - \mu_{ij}^\alpha)^{1/\alpha}$  and  $\pi_{ij} = 1 - \mu_{ij} - \nu_{ij}$ , we can obtain the Eq. (5). So  $\mu_{ij}^*$  can be obtained as below:

$$\mu_{ij}^* = 1 - \left(1 - \mu_{ij}^\alpha\right)^{1/\alpha}, \quad -1 < \alpha < 0 \tag{6}$$

where  $\alpha$  is a parameter in the Yager-generating function.  $\alpha$  is set to  $-1/2$  in this paper, because the test shows a better result when  $\alpha$  equals to  $-1/2$ .  $\mu_{ij}^*$  is called the Yager’s intuitionistic membership degree in the IFCM algorithm.

#### 2.2.2 Kernel FCM algorithm

Differing from conventional clustering algorithm, the fuzzy clustering of objects is described by a fuzzy matrix  $\mu$  with  $n$  rows and  $c$  columns in which  $n$  is the number of data objects and  $c$  is the number of clusters. The element  $\mu_{ij}$  which is the element in  $i$ th row and  $j$ th column in  $\mu$  matrix indicates the degree of membership of the  $i$ th object belonging to the  $j$ th cluster. The FCM aims to minimize the sum of distances between data points and their respective parent cluster centroid as defined as below:

$$\begin{aligned} \min J &= \sum_{i=1}^n \sum_{j=1}^c [\mu_{ij}]^m \times d_{ij}(x_i, c_j) \\ &\sum_{j=1}^c \mu_{ij} = 1, \quad \forall i = 1, 2, \dots, n \\ \text{s.t. } &0 \leq \mu_{ij} \leq 1, \quad \forall i = 1, 2, \dots, n; \\ &\forall j = 1, 2, \dots, c \end{aligned}$$

where the fuzzy parameter  $m$  ( $m > 1$ ) determines the amount of fuzziness of the resulting clusters.  $\mu_{ij}$  and  $c_j$  can be calculated by Eqs. (7) and (8), respectively.  $d_{ij}(x_i, c_j) = \|x_i - c_j\|$  is the Euclidian distance from object  $x_i$  to the clustering center  $c_j$ .

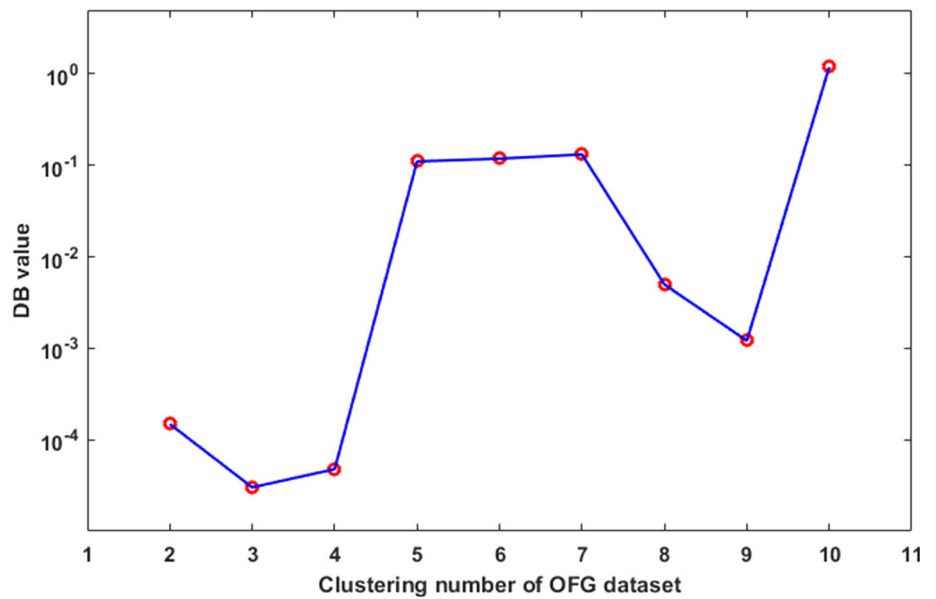
$$\mu_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{ik}}\right)^{\frac{2}{m-1}}} \tag{7}$$

$$c_j = \frac{\sum_{i=1}^n [\mu_{ij}]^m x_i}{\sum_{i=1}^n [\mu_{ij}]^m} \tag{8}$$

Therefore, in the traditional FCM algorithm, the membership degree and the clustering center can be obtained by Eqs. (7) and (8), respectively.

In the previous FCM algorithm, the distance between two independent data points  $x$  and  $y$  is calculated using Euclidean distance  $d(x, y) = \|x - y\|$ . This is because the Euclidean distance is the simplest and most used distance measurement. In this paper, the kernel function works as the distance measurement to substitute the Euclidean distance

**Fig. 1** DB value for different clustering numbers with OFG dataset



measurement (Fan et al. 2009). Defining a nonlinear map as  $\Phi : x \rightarrow \Phi(x) \in F$ , where  $x \in X$ .  $X$  represents the data space, and  $F$  represents the transformed feature space with higher or even infinite dimension. The bidimensional kernel function frequently used are listed as follows:

- Gaussian function

$$K(x, y) = \exp\left(-\|x - y\|^2/\sigma^2\right), \tag{9}$$

- Radial basis (RBF) functions

$$K(x, y) = \exp\left(-\sum_i |x_i^a - y_i^a|^b/\sigma^2\right) \tag{10}$$

$(0 < b \leq 2)$

- Hyper tangent function

$$K(x, y) = 1 - \tanh\left(-\|x - y\|^2/\sigma^2\right), \tag{11}$$

where  $x$  and  $y$  represent two independent data points,  $\sigma$  represents the standard deviation which can be set in advance. Obviously, RBF function reduces into Gaussian function with  $a = 1$  and  $b = 2$ . Taking the kernel function as metric has advantage in distinguishing overlapping data points than conventional Euclidean distance measurement.

Further improvement is achieved by adding a kernel function and intuitionistic fuzzy set into the conventional FCM. The kernel intuitionistic fuzzy c-means (KIFCM) algorithm maps the data point  $x_i$  and centroid  $c_j$  in the kernel space as  $\varnothing(x_i)$  and  $\varnothing(c_j)$ , respectively. The sum of distances between data points and their respective parent cluster centroid that

the KIFCM aims to minimize is defined as follows:

$$\begin{aligned} \min J &= \sum_{i=1}^n \sum_{j=1}^c [\mu_{ij}]^m \times \|\varnothing(x_i), \varnothing(c_j)\|^2 \\ \text{s.t. } &\sum_{j=1}^c \mu_{ij} = 1, \quad \forall i = 1, 2, \dots, n \\ &0 \leq \mu_{ij} \leq 1, \quad \forall i = 1, 2, \dots, n; \\ &\quad \forall j = 1, 2, \dots, c \end{aligned}$$

where  $\|\varnothing(x_i), \varnothing(c_j)\|^2$  is the distance between  $\varnothing(x_i)$  and  $\varnothing(c_j)$  measured by kernel function, calculated by using Eq. (12):

$$\|\varnothing(x_i), \varnothing(c_j)\|^2 = K(x_i, x_i) + K(c_j, c_j) - 2K(x_i, c_j) \tag{12}$$

where  $K(x, y)$  is the kernel function. This paper applies the Hyper tangent functions as the kernel function. The Hyper tangent function is defined in Eq. (11). So, the kernel metric which is taken as the distance measurement can be defined as Eq. (13):

$$d_{ij}(x_i, c_j)^2 = \|\varnothing(x_i) - \varnothing(c_j)\|^2 = 2(1 - K(x_i, c_j)) \tag{13}$$

The formulations used to calculate the degree of membership  $\mu_{ij}$  and clustering center  $c_j$  are defined as Eqs. (14) and (15), respectively.

$$\mu_{ij} = \frac{(1/(1 - K(x_i, c_j)))^{\frac{1}{m-1}}}{\sum_{k=1}^c (1/(1 - (K(x_i, c_k))))^{\frac{1}{m-1}}} \tag{14}$$

$$c_j = \frac{\sum_{i=1}^n [\mu_{ij}]^m K(x_i, c_j) x_i}{\sum_{i=1}^n [\mu_{ij}]^m K(x_i, c_j)} \tag{15}$$

### 2.2.3 Kernel intuitionistic FCM algorithm

This study substitutes the Hyper tangent function for the Euclidean distance measurement as the distance measurement. Based on the intuitionistic fuzzy set, the kernel intuitionistic FCM algorithm which evolves the conventional fuzzy logic into the intuitionistic fuzzy set can be obtained (Lin 2014). The membership  $\mu_{ij}^*$  can be calculated according to Eq. (16). The new clustering center  $c_j^*$  can be calculated by Eq. (17).

$$\mu_{ij} = \frac{(1/(1 - K(x_i, c_j)))^{\frac{1}{m-1}}}{\sum_{k=1}^c (1/(1 - (K(x_i, c_k))))^{\frac{1}{m-1}}}$$

$$\mu_{ij}^* = 1 - \left(1 - \mu_{ij}^{-1/2}\right)^{-2} \tag{16}$$

$$c_j^* = \frac{\sum_{i=1}^n [\mu_{ij}^*]^m K(x_i, c_j) x_i}{\sum_{i=1}^n [\mu_{ij}^*]^m K(x_i, c_j)} \tag{17}$$

The intuitionistic clustering center can be obtained by Eq. (17) in IFCM algorithm. The KIFCM algorithm can be obtained by combining the kernel function and intuitionistic fuzzy set into FCM algorithm. The KIFCM algorithm can be stated in **Algorithm 1**.

---

**Algorithm 1.** KIFCM.

---

**Step 1:** Set the parameter  $m$  and initialize the membership  $\mu_{ij} (i = 1, 2, \dots, n; j = 1, 2, \dots, c)$  matrix.

**Step 2:** Calculate the clustering center  $c_j^* (j = 1, 2, \dots, c)$  according to Eq. (17).

**Step 3:** Calculate the distance measurement  $\|\phi(x_i), \phi(c_j)^2\| (i = 1, 2, \dots, n; j = 1, 2, \dots, c)$  using Eqs. (13) and (11).

**Step 4:** Update and normalize  $\mu_{ij}^* (i = 1, 2, \dots, n; j = 1, 2, \dots, c)$  according to Eq. (16).

**Step 5:** Go to **Step 2** if KIFCM terminating condition is not met.

---

When the relative change in the centroid values becomes small, the KIFCM algorithm is terminated.

### 3 Hybrid method with STA and KIFCM

The hybrid method proposed in this paper are included in this section. The first part of this section is STA which is a global search optimization algorithm. The second part of this section is KIFCM which shows a good performance in

clustering problems. KIFCM algorithm which was proposed firstly in this paper is presented in this section.

### 3.1 State transition algorithm

STA is a novel intelligent global optimization algorithm proposed by Huang et al. (2019), inspired by the understanding of temporal and spatial correlation. STA has been widely used to solve many global search problems (Zhou et al. 2018, 2019; Huang et al. 2019; Zhou et al. 2019a, b; Han et al. 2018). Its heuristic ideas enable itself to search large and discontinuous search spaces without getting trapped into the local optima. Based on the current state  $x_k$ , the unified form of generation of a new state  $x_{k+1}$  in STA can be described as follows:

$$\begin{cases} x_{k+1} = A_k x_k + B_k u_k \\ y_{k+1} = f(x_{k+1}) \end{cases} \tag{18}$$

where  $x_k = [x_{k1}, x_{k2}, \dots, x_{kn}]^T$  stands for a  $n$ -dimensional state, corresponding to a solution of an optimization problem;  $u_k$  is a function of  $x_k$  and historical states.  $A_k$  and  $B_k$  are state transition matrices, which are usually some state transformation operators;  $f(\cdot)$  is the objective function of fitness function, and  $y_{k+1}$  is the function value at  $x_{k+1}$ . Using state space transformation for reference, four special state transition operators are designed to generate continuous solutions for an optimization problem. The four special state transition operators can be described as below:

– *Rotation transformation*

$$x_{k+1} = x_k + \alpha \frac{1}{n \|x_k\|_2} R_t x_k, \tag{19}$$

where  $\alpha$  is a positive constant, called the rotation factor;  $R_r \in \mathbb{R}^{n \times n}$ , is a random matrix with its entries being uniformly distributed random variables defined on the interval  $[-1, 1]$ , and  $\|\cdot\|_2$  is the 2-norm of a vector. This rotation transformation has the function of searching in a hypersphere with the maximal radius  $\alpha$ .

– *Translation transformation*

$$x_{k+1} = x_k + \beta R_t \frac{x_k - x_{k-1}}{\|x_k - x_{k-1}\|_2}, \tag{20}$$

where  $\beta$  is a positive constant, called the transition factor;  $R_t \in \mathbb{R}$  is a uniformly distributed random variable defined on the interval  $[0, 1]$ . The translation transformation has the function of searching along a line from  $x_{k-1}$  to  $x_k$  at the starting point  $x_k$  with the maximum length  $\beta$ .

– *Expansion transformation*

$$x_{k+1} = x_k + \gamma R_e x_k, \tag{21}$$



where  $\gamma$  is a positive constant, called the expansion factor;  $R_e \in \mathbb{R}^{n \times n}$  is a random diagonal matrix with its entries obeying the Gaussian distribution. The expansion transformation has the function of expanding the entries in  $x_k$  to the range of  $[-\infty, +\infty]$ , searching in the whole space.

– *Axesion transformation*

$$x_{k+1} = x_k + \delta R_a x_k, \quad (22)$$

where  $\delta$  is a positive constant, called the axesion factor;  $R_a \in \mathbb{R}^{n \times n}$  is a random diagonal matrix with its entries obeying the Gaussian distribution and only one random position having nonzero value. The axesion transformation is aiming to search along the axes, strengthening single dimensional search.

With the state transformation operators, sampling technique and update strategy, the procedures of the state transition algorithm can be described in **Algorithm 2**.

---

#### Algorithm 2. STA.

---

**Step 1:** Set parameters of STA including  $\alpha$  ( $\alpha_{min}, \alpha_{max}$ ),  $\beta, \gamma, \delta$ , SE and maximum iterative count.

**Step 2:** Initialize  $Best$  and  $fBest$ .

**Step 3:** Generate SE samples by Expansion transformation on the basis of the current  $Best$  and update the current  $Best$  by strategy of greed.

**Step 4:** Generate SE samples by Rotation transformation on the basis of the current  $Best$  and update the current  $Best$  by strategy of greed.

**Step 5:** Generate SE samples by Axesion transformation on the basis of the current  $Best$  and update the current  $Best$  by strategy of greed.

**Step 6:** Go to **Step 3**, if the terminating condition is not met.

---

When the current  $Best$  needs to be abandoned, Translation transformation can be taken to update the  $Best$ . That is the update strategy of greed. Generally, the maximum iteration reached is taken as the terminating condition. In **Algorithm 2**, SE is called search enforcement, representing the times of transformation by a certain operator. The changeable rotation factor can not only speed up the process of finding optimal solution, but also increase the accuracy of the solution. For better understanding of more detailed steps of STA, the STA toolbox can be downloaded via the following link: <https://www.mathworks.com/matlabcentral/fileexchange/52498-state-transition-algorithm>

---

#### Algorithm 3. STA-KIFCM.

---

**Step 1** Set parameters of STA including  $\alpha, \beta, \gamma, \delta$ , SE and maximum iterative count, input and regularize dataset of KIFCM.

**Step 2** Initialize  $Best$  and  $fBest$ .

**Step 3** STA

**Step 3.1** Calculate  $d$  and  $\mu^*$  using Eqs. (13) and (16), respectively.

**Step 3.2** Calculate  $\mu^*$  by Eq. (16) and normalize it.

**Step 3.3** Update  $Best$  by four operators and  $fBest$  by rule of greed.

**Step 3.4** If STA terminating condition is not met, go to **Step 3.3**.

**Step 4** KIFCM

**Step 4.1** Calculate  $d$  and  $\mu^*$  using Eqs. (13) and (16), respectively.

**Step 4.2** Calculate  $\mu^*$  by Eq. (16) and normalize it.

**Step 4.3** Update the  $Best$  using Eq. (17) and  $fBest$ .

**Step 4.4** If KIFCM terminating condition is not met, go to **Step 4.1**.

**Step 5** Go to **Step 3**, if STA-KIFCM terminating condition is not met.

---

### 3.2 STA-KIFCM algorithm

Good performance of the KIFCM algorithm has been shown by many literatures (Pena et al. 1999). However, the results of KIFCM is not stable since KIFCM is highly dependent on the initial clustering centers. The searching ability of the KIFCM algorithm is a local search procedure. Previously, many studies focus on designing a novel cluster initialization (Babu and Murty 1993). STA has good performance in generating heuristic and stochastic initial clustering centers (Zhou et al. 2016). There exist four operators which can generate diverse candidate clustering centers in STA. After that, KIFCM can obtain the clustering centers rapidly and reasonably. So, the hybrid method can generate the optimal clustering centers accurately. The hybrid method with STA and KIFCM can not only avoid getting trapped into the local optima, but also improve the degree of accuracy efficiently. The procedures of the hybrid method can be stated in **Algorithm 3**.

Several stopping rules can be used. The terminate condition in the **Step 3.4** is the maximum number of iterations reached. The **Step 4.4** is terminated when the relative change in the degree of membership becomes small. The **Step 5** stop when the objective function cannot be minimized anymore. Figure 2 shows the total flowchart of the STA-KIFCM.

## 4 Experiment results and discussion

This section shows experiments which use five benchmarks datasets to evaluate the performances of fuzzy clustering methods. In this section, the comparison algorithms will be introduced in Sect. 4.1. Then, the experimental setup, results and discussion are included in the rest of this section.

### 4.1 Comparison algorithms

In this paper, three evolutionary algorithms are adopted to generate the initial centroid for KIFCM. Two typical evolutionary algorithms named GA and PSO are used to compare with STA to overcome a drawback of KIFCM algorithm which is highly dependent on initial centroid. All the parameters of these algorithms are used as the recommended settings on their corresponding reference (Kuo et al. 2018). The STA-KIFCM proposed in this paper has been detailedly described in Fig. 2. In fact, the function of GA and PSO are similar to STA in generating the initial centroid for KIFCM. The main differences between these evolutionary algorithms are the way of generating candidate solutions and updating optimal solution strategy. The detail descriptions of differences between GA-KIFCM, PSO-KIFCM and STA-KIFCM are given as follows.

- (1) GA-KIFCM: the way of generating candidate solution is based on crossover operator and mutation operator. The strategy of updating optimal solution is the rule of greedy (Whitley 1994; Hosseinabadi et al. 2019; Kuo et al. 2018).
- (2) PSO-KIFCM: the vector and position updating formula are adopted to generate candidate solution. The updating optimal solution can be summarized as learning from the best particle (Filho et al. 2015; Deng et al. 2019; Kuo et al. 2018).

In addition, the framework of GA-KIFCM and PSO-KIFCM are similar to STA-KIFCM which has been illustrated in Fig. 2. Compared to STA-KIFCM, GA and PSO are adopted to take the place of STA in GA-KIFCM and PSO-KIFCM, respectively.

### 4.2 Experimental setup

Table 3 shows the parameters of PSO-KIFCM and GA-KIFCM methods which have been set in Kuo et al. (2018). The parameters are set according to the rule that the number of evaluating target function are uniform. The preparation for the experiment are outlined below:

**Table 1** Datasets characteristics

Datasets parameter	Iris	Wine	Tae	Seed	OFG
Number of clusters	3	3	3	6	3
Number of features	4	13	5	9	2
Number of instances	150	178	151	214	272

#### – Parameters of STA-KIFCM settings

According to previous literatures, this study sets the weighting exponent  $m$  as 2 (Pedrycz and Rai 2008). The standard deviation  $\sigma$  in kernel function is chosen as  $\sqrt{2}$ . The parameters  $\beta$ ,  $\gamma$ ,  $\delta$  in STA are chosen as 1 and the parameter  $\alpha$  is decreased linearly from 1 to  $e^{-4}$  over the course of iterations to facilitate global exploration in the early stages and exploitation in the latter stages. The search enforcement (SE) in STA can be chosen as 30, which means that the number of candidate solutions in every iteration is 30.

#### – Datasets description

This part adopts five benchmark clustering datasets which can be downloaded from the UCI machine learning repository, with the exception of the Old Faithful Geyser (OFG) dataset, to evaluate the performance of the proposed algorithm. So many previous studies have taken these datasets into various fields including biology, oncology, education, medicine, chemistry and agriculture (Michielssen et al. 1992). Table 1 gives a brief overview of these datasets characteristics.

- *Iris dataset* was introduced by Fisher (Fisher 1936) consisting of three species of the Iris flowers (Virginica, Versicolour and Setosa) includes a total of 150 instances with 4 features (Length and width of the sepals and petals).
- *Wine dataset* consisting of wine derived from three cultivars grown in the same region includes a total of 178 instances with 13 features (alcohol, malic acid, ash, alkalinity of ash, magnesium, total phenols, flavanoids, nonflavanoid phenols, proanthocyanins, color intensity, hue, OD280/OD315 of diluted wines and proline).
- *Tae dataset* consists of evaluations of teaching performance over three regular semesters and two summer semesters of 151 teaching assistant (TA) assignments at the Statistics Department of the University of Wisconsin-Madison. The scores were divided into three roughly equal-sized categories (“low”, “medium”, and “high”) to form the class variable.
- *Seed dataset* consisting of kernels belonging to three different varieties of wheat: Kama, Rosa and Cana-

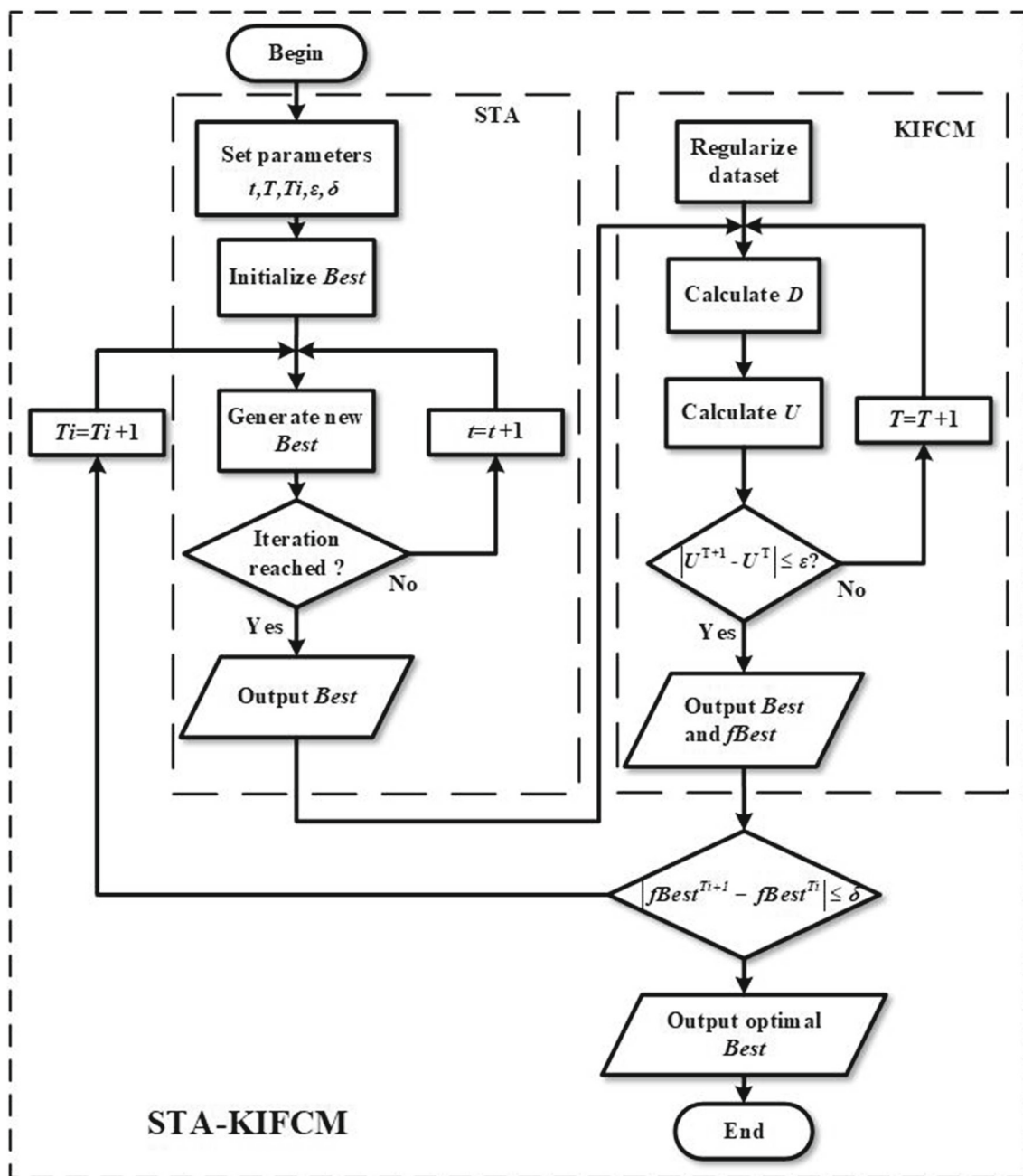


Fig. 2 The flowchart of the STA-KIFCM method

dian, 70 elements each, randomly selected for the experiment includes seven geometric parameters of wheat kernels (area, perimeter, compactness, length of kernel, asymmetry coefficient, length of kernel grooves).

- *Old Faithful Geyser (OFG) Dataset* describe the waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA (Hutchinson et al. 1997). 272 observations on 2 variables (erup-

tions numeric eruption time in mins, waiting numeric waiting time to next eruption) are contained in the data frame.

#### – Performance indices

The performance indices are used to evaluate the effectiveness of fuzzy clustering method shown in Table 2. The performance indices are listed as follows:



**Table 2** The comparison of results obtained by each algorithm

Methods	Datasets	Inter-distance	T-consuming	Accuracy (%)
FCM	Iris	2.6910	0.0182	93
	Wine	18.6902	0.0201	90
	Tae	19.4265	0.0257	92
	Seed	2.0096	0.1164	94
	OFG	1.3687	0.0411	64
IFCM	Iris	2.8641	0.0185	93
	Wine	27.3673	0.0258	94
	Tae	31.1523	0.0254	85
	Seed	2.6412	0.2078	95
	OFG	1.6098	0.0376	59
KFCM	Iris	2.2539	1.4142	95
	Wine	16.6493	1.4142	99
	Tae	16.3404	1.4142	92
	Seed	1.5472	1.4142	90
	OFG	1.2241	1.4142	61
KIFCM	Iris	2.9788	1.4142	95
	Wine	26.1836	1.4142	98
	Tae	29.0528	1.4142	86
	Seed	2.6453	1.4142	85
	OFG	1.6765	1.4142	60
PSO-KIFCM	Iris	0.0100	19.0313	96
	Wine	0.0843	26.5734	99
	Tae	0.0099	21.7344	91
	Seed	0.1306	87.4375	90
	OFG	0.0055	38.5625	90
GA-KIFCM	Iris	0.0294	22.1875	97
	Wine	0.0703	28.0469	98
	Tae	0.1034	22.5938	92
	Seed	0.1402	93.4844	96
	OFG	0.1013	39.5156	87
STA-KIFCM	Iris	0.0100	15.5313	99
	Wine	0.0346	20.1833	99
	Tae	0.0110	16.7344	95
	Seed	0.0452	75.0313	97
	OFG	0.0055	36.0625	95

**Table 3** Parameter settings for PSO-KIFCM and GA-KIFCM methods

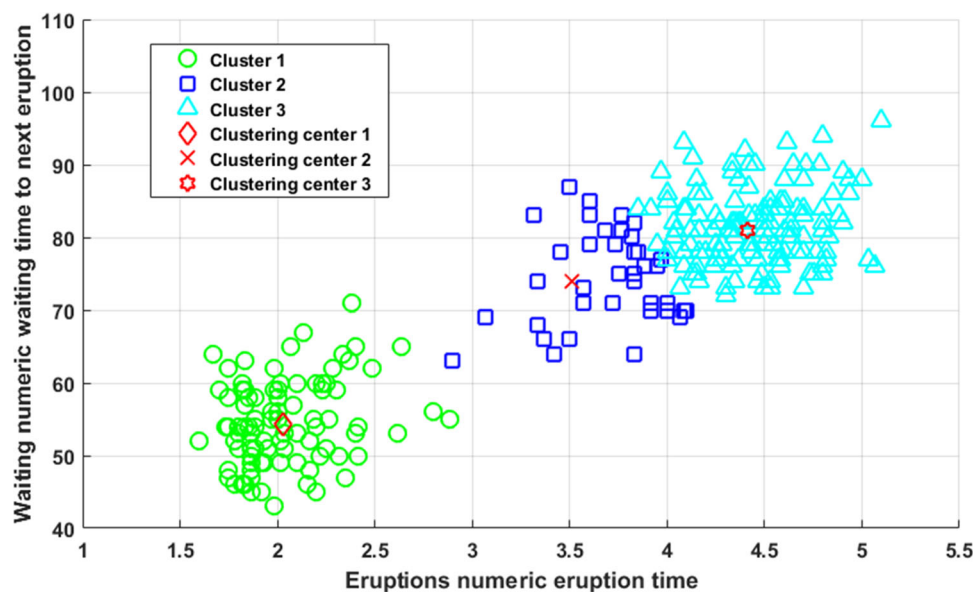
Methods	Factors	Parameters
PSO-KIFCM	Maximum number of iteration	40
	Number of particles	50
	$c_1$	0.5
	$c_2$	0.5
GA-KIFCM	Maximum number of iteration	40
	Number of chromosomes	50
	Crossover rate	0.85
	Mutation rate	0.001

(a) *Inter-distance*. The sum of distances between data points and their respective parent cluster centroid, smaller value of which are desirable and indicate greater compactness of clustering. The inter-distance can be formulated as follows:

$$d = \sum_{i=1}^k \sum_{x_j \in S_i} d_{ji}(x_j, c_i)^2 \tag{23}$$

where  $k$  represents the number of clustering centers,  $S_i$  represents the  $i$ th cluster and  $x_j$  represents the  $j$ th point belonging to cluster  $S_i$ .  $d_{ji}(x_j, c_i)^2$  represents the dis-

**Fig. 3** The clustering result of OFG dataset



tance between point  $x_j$  and  $c_i$ . In FCM and IFCM, the Euclidean distance  $d_{ji}(x_j, c_i)^2 = \|x_j - c_i\|^2$  is taking as the distance measurement. But the kernel function  $d_{ji}(x_j, c_i)^2 = \|\phi(x_j), \phi(c_i)\|^2$  is the distance metric in KFCM, KIFCM GA-KIFCM PSO-KIFCM and STA-KIFCM.

- (b) *Time-consuming*. Time complexity is commonly used to evaluate the rapidity of algorithm. The less time consuming with high validity is desirable.
- (c) *Accuracy*. Each data point in the tested datasets are labeled according to their membership value for each cluster. The final cluster label for each data point is defined as the cluster which has the highest membership value. The same data point may be labeled differently under each repetition. The accuracy of algorithm labeling is the ratio of the correct labels to total labels.

### 4.3 Result and discussion

The results of the experiment with five benchmark datasets are shown in this part. The result of proposed STA-KIFCM method compared with other methods have been shown in Table 2. In order to reduce the error and contingency, each algorithm is executed 100 times. The comparison results indicate that low accuracy has been achieved by FCM, IFCM and KFCM algorithms with OFG dataset. The reason is so many points in OFG dataset are overlapped and close with each other. Original points in OFG dataset have been shown in Fig. 3. Although FCM and IFCM algorithms cost less time, they behave lower accuracy than other algorithms. KFCM and KIFCM algorithm cost almost the same time. Because they are not very different in programming. The formulation of calculating the degree of membership is the only

difference exists in these two algorithms. High accuracy and short Inter-distance can be obtained by STA-KIFCM methods with less time than PSO-KIFCM, GA-KIFCM. Figure 4 displays the comparison result about accuracy. Although the performance of PSO-KIFCM and GA-KIFCM are similar to STA-KIFCM under Iris, Wine and Tae datasets, STA-KIFCM has the highest accuracy in these three datasets. Moreover, not only does STA-KIFCM perform the highest accuracy but also the differences between STA-KIFCM and GA-KIFCM, PSO-KIFCM are obvious under Seed and OFG datasets. So, we can conclude that STA-KIFCM behaves a better performance than PSO-KIFCM and GA-KIFCM under testing all datasets. In obtaining the optimal initial clustering centroid for KIFCM, STA achieves the highest accuracy with the lowest time-consuming under five datasets. It shows that STA has excellent performance than GA and PSO in speedability and optimality when solving the optimization problem in clustering. Figure 3 obtained by the STA-KIFCM method displays the clustering result of OFG dataset. OFG dataset can be clustered into 3 clusters which has been verified in Sect. 2.1. So many data points in cluster 2 are close with points in cluster 3. It is hard to partition the data in the intersection between Clusters 2 and 3 by Euclidean distance measurement. And the intersection between Clusters 2 and 3 can not be easily distinguished by vision. But STA-KIFCM method can classify these points graphically and find the accurate clustering centers exactly. It is easy to distinguish the edge between cluster 2 and cluster 3.

## 5 Conclusions

In this paper, we integrate KIFCM with STA to overcome the shortcomings of the conventional FCM. Using the novel

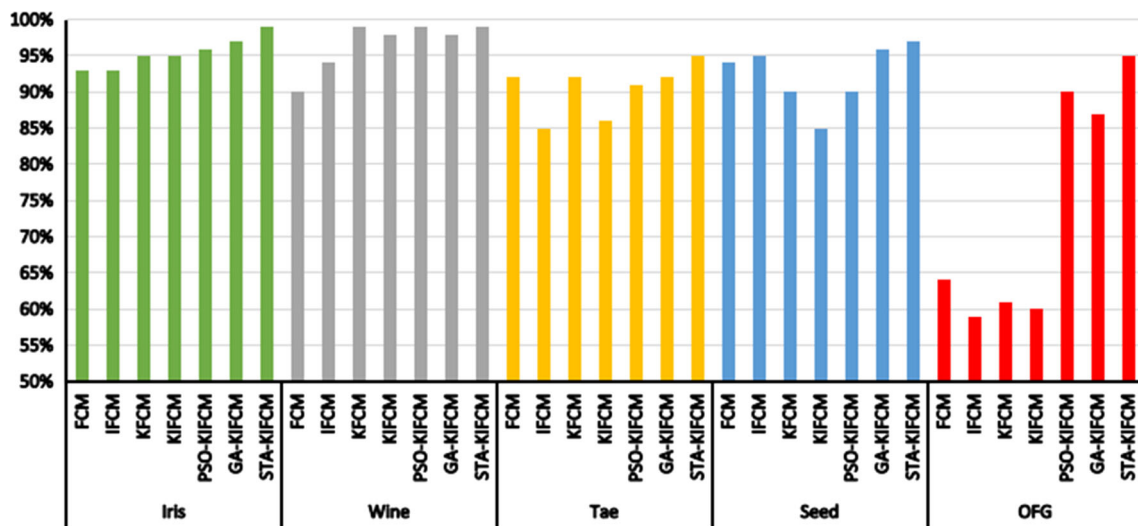


Fig. 4 The comparison of accuracy for different methods

intelligent global search algorithm STA to provide a better initial centroid can improve the results of KIFCM. Determining the parameters of STA can control the trade-off between local and global search. If the data points are distributed in a large area, wide exploration is required. On the other hand, if the data points are located within a relatively small area, deep exploitation is needed. STA can prevent the KIFCM algorithms from getting trapped into local optima because it aims to find the optimal initialized cluster centroid for KIFCM algorithms. The proposed STA-KIFCM method are evaluated using five well-known clustering datasets and compare with PSO-KIFCM and GA-KIFCM. Experimental results over five well-known datasets show that the STA-KIFCM method is efficient and can reveal very encouraging results in term of quality of clustering centers found because of the four operators in STA. The proposed method in this paper have intrinsic advantages, clusters are more internally homogeneous and more diverse from each other, and provide better partitioning of testing datasets. According to these results, further study can focus on three main perspectives. The first is that reducing the time complexity of the hybrid method should be further considered. The second is applying the proposed clustering method to solve real-world problems in industrial process. The third is replacing the kernel function measurement with different similarity measurements to study the influence of similarity measurement on KIFCM clustering results.

### Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** This article does not contain any studies with human participants performed by any of the authors.

**Funding** This work was funded by the National Natural Science Foundation of China (Grant Nos. 61860206014, 61873285), the Innovation-Driven Plan in Central South University (Grant No. 2018CX012), the 111 Project (Grant No. B17048) and Hunan Provincial Natural Science Foundation of China (Grant No. 2018JJ3683) and the Fundamental Research Funds for the Central Universities of Central South University (Grant No. 2019zzts577).

**Informed consent** Informed consent was obtained from all individual participants included in the study.

### References

- Askari S, Montazerin N, Zarandi MH et al (2017) Generalized entropy based possibilistic fuzzy C-Means for clustering noisy data and its convergence proof. *Neurocomputing* 219:186–202
- Atanassov KT, Ranganamy P (1986) Intuitionistic fuzzy sets. *Fuzzy Sets Syst* 20(1):87–96
- Babu GP, Murty MN (1993) A near-optimal initial seed value selection in K-means algorithm using a genetic algorithm. *Pattern Recognit Lett* 14(10):763–769
- Bastanlar Y, Ozuyusal M (2014) Introduction to machine learning. *Methods Mol Biol* 1107:105–128
- Bezdek JC, Ehrlich R, Full WE et al (1984) FCM: the fuzzy c-means clustering algorithm. *Comput Geosci* 10:191–203
- Chaira T (2011) A novel intuitionistic fuzzy C means clustering algorithm and its application to medical images. *Appl Soft Comput* 11(2):1711–1717
- Davies DL, Bouldin DW (1979) A cluster separation measure. *IEEE Trans Pattern Analysis and Mach Intell* 1(2):224–227
- Deng W, Yao R, Zhao H et al (2019) A novel intelligent diagnosis method using optimal LS-SVM with improved PSO algorithm. *Soft Comput* 23(7):2445–2462
- Fan J, Han M, Wang J et al (2009) Single point iterative weighted fuzzy C-means clustering algorithm for remote sensing image segmentation. *Pattern Recognit* 42(11):2527–2540

- Filho TD, Pimentel BA, De Souza RM et al (2015) Hybrid methods for fuzzy clustering based on fuzzy c-means and improved particle swarm optimization. *Expert Syst Appl* 42(17):6315–6328
- Fisher RA (1936) The use of multiple measurements in taxonomic problems. *Ann Human Genet* 7(2):179–188
- Fouche G, Langit L (2011) Introduction to data mining. *Scalable Comput Pract Exp* 9(4):369–402
- Ghosh S, Dubey SK (2013) Comparative analysis of K-means and fuzzy C-means algorithms. *Int J Adv Comput Sci Appl* 4:4
- Han J, Yang C, Zhou X et al (2018) A two-stage state transition algorithm for constrained engineering optimization problems. *Int J Control Autom Syst* 16(2):522–534
- Hathaway RJ, Bezdek JC (1995) Optimization of clustering criteria by reformulation. *IEEE Trans Fuzzy Syst* 3(2):241–245
- Hosseinabadi AAR, Vahidi J, Saemi B et al (2019) Extended genetic algorithm for solving open-shop scheduling problem. *Soft Comput* 23(13):5099–5116
- Huang Z, Yang C, Zhou X et al (2018) A novel cognitively inspired state transition algorithm for solving the linear bi-level programming problem. *Cogn Comput*. 1:1–11
- Huang Z, Yang C, Zhou X et al (2018) A hybrid feature selection method based on binary state transition algorithm and relief. *IEEE J Biomed Health Inf* 23(5):1888–1898
- Huang M, Zhou X, Huang T et al (2019) Dynamic optimization based on state transition algorithm for copper removal process. *Neural Comput Appl* 31(7):2827–2839
- Hutchinson RA, Westphal JA, Kieffer SW (1997) In situ observations of old faithful geyser. *Geology* 25(10):875–878
- Kuo RJ, Mei CH, Zulvia FE et al (2016) An application of a metaheuristic algorithm-based clustering ensemble method to APP customer segmentation. *Neurocomputing* 205:116–129
- Kuo RJ, Lin TC, Zulvia FE et al (2018) A hybrid metaheuristic and kernel intuitionistic fuzzy c-means algorithm for cluster analysis. *Appl Soft Comput* 67:299–308
- Li C, Zhou J, Kou P et al (2012) A novel chaotic particle swarm optimization based fuzzy clustering algorithm. *Neurocomputing* 83:98–109
- Lin K (2014) A novel evolutionary kernel intuitionistic fuzzy C-means clustering algorithm. *IEEE Trans Fuzzy Syst* 22(5):1074–1087
- Liu J, Xu M (2008) Kernelized fuzzy attribute C-means clustering algorithm. *Fuzzy Sets Syst* 159(18):2428–2445
- Michielssen E, Ranjithan S, Mitra R (1992) Optimal multilayer filter design using real coded genetic algorithms. *IEE Proc J (Optoelectronics)* 139(6):413–420
- Papageorgiou EI, Iakovidis DK (2013) Intuitionistic fuzzy cognitive maps. *IEEE Trans Fuzzy Syst* 21(2):342–354
- Pedrycz W, Rai P (2008) Collaborative clustering with the use of fuzzy C-means and its quantification. *Fuzzy Sets Syst* 159(18):2399–2427
- Pena JM, Lozano JA, Larranaga P et al (1999) An empirical comparison of four initialization methods for the K-means algorithm. *Pattern Recognit Lett* 20(10):1027–1040
- Sipe MA (2001) Statistical pattern recognition. *Pattern Anal Appl* 4(1):71–72
- Wang P (1983) Pattern recognition with fuzzy objective function algorithms (James C. Bezdek). *Siam Rev* 25(3):442–442
- Whitley D (1994) A genetic algorithm tutorial. *Stat Comput* 4(2):65–85
- Zadeh LA (1999) Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets Syst* 100(1):9–34
- Zhang DQ, Chen SC (2003) Clustering incomplete data using kernel-based fuzzy C-means algorithm. *Neural Process Lett* 18(3):155–162
- Zhang DQ, Chen SC (2004) A novel kernelized fuzzy C-means algorithm with application in medical image segmentation. *Artif Intell Med* 32(1):37–50
- Zhao H, Xu Z, Ni M et al (2010) Generalized aggregation operators for intuitionistic fuzzy sets. *Int J Intell Syst* 25(1):1–30
- Zhou X, Yang C, Gui W et al (2012) State transition algorithm. *J Ind Manag Optim* 8(4):1039–1056
- Zhou KL, Chao F, Yang SL (2014) Fuzziness parameter selection in fuzzy c-means: the perspective of cluster validation. *Sci China Inf Sci* 57(11):1–8
- Zhou X, Gao DY, Yang C et al (2016) Discrete state transition algorithm for unconstrained integer optimization problems. *Neurocomputing* 173:864–874
- Zhou X, Shi P, Lim CC et al (2018) A dynamic state transition algorithm with application to sensor network localization. *Neurocomputing* 273:237–250
- Zhou X, Yang C, Gui W (2018) A statistical study on parameter selection of operators in continuous state transition algorithm. *IEEE Trans Cybern* 49(10):3722–3730
- Zhou X, Zhou J, Yang C et al (2018) Set-point tracking and multi-objective optimization-based PID control for the goethite process. *IEEE Access* 6:36683–36698
- Zhou X, Huang M, Huang T et al (2019) Dynamic optimization for copper removal process with continuous production constraints. *IEEE Trans Ind Inf*. <https://doi.org/10.1109/TII.2019.2943500>
- Zhou X, Yang K, Xie Y et al (2019a) A novel modularity-based discrete state transition algorithm for community detection in networks. *Neurocomputing* 334:89–99
- Zhou X, Long J, Xu C et al (2019) An external archive-based constrained state transition algorithm for optimal power dispatch. *Complexity*. <https://doi.org/10.1155/2019/4727168>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.