International Journal of Innovative Computing, Information and Control
Volume **20**, Number **1**, February **2024**

ICIC International ©2024 ISSN 1349-4198
pp. **297–311**

A CLUSTERING ALGORITHM COMBINING FUZZY C-MEANS AND ARTIFICIAL BEE COLONY ALGORITHM

Kan Ni

Graduate School of Science and Technology
Gunma University
1-5-1 Tenjincho, Kiryu 376-8515, Japan
t202d003@gunma-u.ac.jp

Received May 2023; revised September 2023

Abstract. *The current clustering algorithm has problems such as sensitivity to initial selection, poor global search ability, and low clustering efficiency, which not only affects its segmentation ability but also makes it difficult to meet the needs of practical applications. Therefore, in response to this issue, the Differential Evolution (DE) algorithm was used to improve the Artificial Bee Colony (ABC) algorithm, and in practical clustering applications, the improved algorithm was combined with Fuzzy C-Means (FCM) to obtain the Differential Evolution Artificial Bee Colony-Fuzzy Clustering (DEABC-FC) algorithm, which was experimentally analyzed. The experimental results showed that the mutation strategy results of DEABC-FC were around 60.574 and 1.541e02 on Iris and Glass, and around 1.228e-10 and 6.003e-09 on CMC and Wine datasets, both lower than those of Artificial Bee Colony-Fuzzy Clustering (ABC-FC) and FCM algorithms. In addition, the DEABC-FC algorithm has undergone 15 iterations on 4 datasets, and its clustering performance is relatively good, and it can effectively balance global and local search. At the same time, after changing the values of the mutation factor and crossover factor separately, it was found that increasing the value of the mutation factor improved the diversity of the population and the stability of the algorithm, but also had a certain impact on the convergence of the algorithm. The increase in the value of the crossover factor makes the DEABC-FC algorithm have faster convergence speed and better global optimization ability. Overall, DEABC-FC has stronger global optimization ability and convergence speed in clustering compared to ABC-FC and FCM methods, which can effectively offset the shortcomings of clustering algorithms and have important significance in practical clustering applications.*

Keywords: Data mining, Artificial bee colony, Convergence rate, Clustering algorithm, Mutation strategy

1. **Introduction.** The rapid development of computer information technology has led to the explosive growth of data information. After the hidden value of these data has been paid attention to, it is imminent to extract valuable information from a large amount of data [1]. As a highly intelligent data information extraction method, data mining has gradually attracted the attention of major enterprises. Among them, clustering is a relatively important branch in the data mining [2,3]. Numerous domestic and foreign scholars have studied it, for example, Lyu et al. [4] proposed a bidirectional clustering algorithm based on local density and applied it in the denoising; Xu et al. [5] used an average connection hierarchical clustering algorithm based on correlation similarity to measure the distribution of elements in coal; Qu and Wang [6] used Fuzzy Clustering (FC) algorithm to achieve image denoising and accurate segmentation of corpus callosum sagittal plane in diffusion tensor images. However, there is currently limited research on the combination

DOI: 10.24507/ijicic.20.01.297

of Swarm Intelligence (SI) algorithms and clustering; therefore, the improved Differential Evolution Artificial Bee Colony (DEABC) algorithm is obtained by combining the Artificial Bee Colony (ABC) algorithm with the Differential Evolution (DE) algorithm, and the Fuzzy C-Means (FCM) algorithm based on the improved ABC is obtained by combining it with the FCM algorithm in clustering application. The purpose is to solve the problems of traditional clustering algorithms that are sensitive to initial points, poor search ability and low adaptability, so as to promote the development of SI algorithms in the clustering. At the same time, it not only has the advantage of utilizing swarm intelligence algorithms to compensate for the shortcomings of traditional clustering algorithms, but also has the innovation of effectively combining swarm intelligence algorithms with clustering algorithms.

The remainder is divided into four parts. The first part is a summary and discussion of the current research on FCM algorithms both domestically and internationally. The second part is the study of clustering algorithms that integrate FCM and improved ABC algorithm. The third part is to verify the fusion algorithm. The fourth part is a summary of the entire article.

2. **Related Work.** In the era of dataization, data mining complies with the urgent needs of social development for data processing capabilities, and has broad application prospects. And clustering is an important branch of data mining. In the clustering algorithm, the FCM algorithm has been widely used because of its advantages of simple thinking and fast operation. Therefore, many scholars at home and abroad have carried out research on it. Huang et al. proposed a parallel FCM clustering algorithm based on the segmentation set, which improved the classification accuracy and efficiency of the FC algorithm [7]. The FCM algorithm was improved to build a database of relevant station types to achieve short-term passenger flow prediction for entry and exit [8]. Xu et al. proposed an optimization method for real-time response efficient segmentation of variable sets, thereby improving the real-time response of artillery structures [9]. Nie et al. proposed a new K-means clustering algorithm based on a detailed analysis of data mining to address the issue of maximizing the trajectory of objective functions in machine learning. This effectively optimized the trajectory maximization problem while improving the convergence speed of the algorithm [10]. Shunmuganathan et al. proposed an optimized SI algorithm for optimizing image segmentation by combining FCM with ABC algorithm, thereby improving the efficiency of image segmentation [11]. Ye et al. have effectively improved the accuracy of short-term wind power prediction by combining the FCM algorithm with the DE algorithm and the wavelength division method to address issues related to wind power prediction accuracy [12]. Wen et al. used the improved FCM algorithm to improve the effect of image segmentation [13]. Memon and Lee proposed a generalized improved FCM algorithm to enhance the image output strength [14]. Zheng et al. used the local FCM algorithm to classify the image, which improves the uniform segmentation ability of the image [15].

In addition, Deng et al. used the ABC algorithm to establish a stacked multi-objective portfolio model, which intuitively reflected the relationship between investor risk and return [16]. Forouzandeh et al. blurred the ABC algorithm. The Technology for Order Preference by Similarity to an Ideal Solution (TOPSIS) model was effectively combined to improve the accuracy of hotel recommendation for tourists [17]. Acquah et al. optimized the reliability model of the system, then ensuring the reliability of software quality in fields such as medicine and engineering [18]. Improvement of the ABC algorithm not only enhances the development of the algorithm, but also improves the global search ability of the algorithm [19]. Baesmat et al. used the ABC algorithm to improve the accuracy of short-term load forecasting, so as to effectively predict weather data [20]. Gao et al. improved the ABC algorithm with information learning, thereby improving the search ability of the algorithm [21]. Zabidi et al. integrated the binary into the ABC algorithm, thereby improving the convergence of the ABC algorithm [22]. Gu et al. used order optimization scheme to enhance its parameter estimation [23].

From the research of domestic and foreign scholars, although FCM algorithm has advantages in computing speed, it also has problems such as lack of anti-noise, and the advantages of ABC algorithm can complement FCM algorithm, thus improving the overall performance of clustering algorithm. Therefore, the research of combining DE algorithm with FCM and ABC algorithm not only makes use of SI algorithm to make up for the disadvantages of traditional clustering algorithm, but also has the innovation of effectively combining SI algorithm with clustering algorithm. At the same time, the research on mutation and crossover factor in traditional algorithms is mostly empirical judgment. The research on using DEABC-FC algorithm to control these two parameters has theoretical basis in experiments. The experience summarized has also made its own contribution to the development of subsequent clustering algorithms.

3. **Clustering Algorithm Combining FCM and Improved ABC Algorithm.**

3.1. **An improved algorithm based on artificial peak swarm optimization and DE algorithm.**

To better mine information and improve the deficiencies of the traditional clustering algorithm in the division ability, the research combined ABC and DE algorithm to obtain an improved DEABC algorithm, and combined it with FCM to obtain the DEABC-FC algorithm, and practically apply it to clustering. The ABC algorithm is a group intelligence algorithm, which can simulate the honey-collecting behavior of bees. According to the nature of its work, the bee colony can be divided into collecting bees, observation bees and scout bees. It mainly studies the bees collecting honey, the division of labor, collaboration, and information interaction among individuals, enabling them to have optimal performance and adaptability [24]. Different bees share information by wagging their tails. In the searching for food, they cooperate with each other and change roles, so as to achieve high efficiency and energy saving. The calculation of this is shown in Equation (1).

$$fitness_i = \frac{1}{1 + f(z_i)} \tag{1}$$

In Equation (1), z_i represents the nectar source; f represents the objective function. Among them, the calculation expresses the quality of the nectar source by suitability, and its value indicates the pros and cons of the objective function. After understanding the principle of the nectar harvesting process, the ABC algorithm can be roughly estimated, as shown in Figure 1.

From Figure 1, the ABC algorithm first involves initializing it to obtain multiple initial honey sources; the second is to collect bees. According to Equation (3), in z_i, the vicinity of the nectar source generates a new nectar source, and changes it to the boundary value under the premise that the new nectar source exceeds the boundary, and compare the fitness value of the old and new nectar source. If the new nectar source is large, it needs to update it; otherwise it will remain unchanged. The value obtained by Equation (4) is used to select the nectar source. The last is the scout bee stage. If the nectar source is not updated after multiple disturbances, it is determined to be in a depleted state, so it returns to the bee collection stage for the next iteration until to the maximum number of times, thus ending the process. Among them, in the initialization stage, its initialization

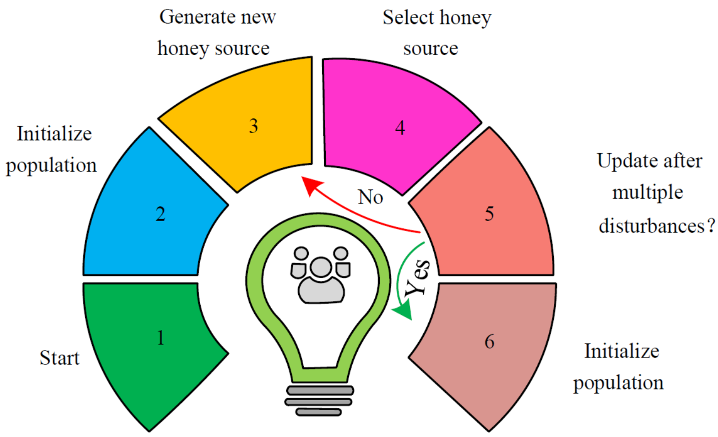


FIGURE 1. ABC algorithm flow diagram

is shown as Equation (2).

$$z_i = x_j^{\min} + rand(0,1)\left(x_j^{\max} - x_j^{\min}\right) \tag{2}$$

In Equation (2), j represents the dimension; $x_j^{\max}$ represents the upper bound of the j dimension, and $x_j^{\min}$ means the lower bound of the dimension j. In collecting bees, the new nectar source generation expression is shown in Equation (3).

$$v_{ij} = z_{ij} + \phi(z_{ij} - z_{lj}) \tag{3}$$

In Equation (3), z_{ij} represents honey source; ϕ represents a random number that controls the disturbance amplitude. In the observation bee stage, its probability calculation is shown in Equation (4).

$$P_i = \frac{fitness_i}{\sum_{j=1}^{SN} fitness_i} \tag{4}$$

In Equation (4), P_i denotes the probability of selecting nectar sources; SN refers to the number of initial nectar sources. The DEABC algorithm is a combination of the ABC algorithm and the DE algorithm. Therefore, after understanding the ABC algorithm, it needs to analyze the DE algorithm to understand the DEABC algorithm. It is worth noting that the DEABC algorithm is an improved ABC algorithm, so its process is roughly the same as the ABC algorithm. The DEABC algorithm combines the advantages of the two, improves the convergence speed of the algorithm, and can better balance the global and local search capabilities to prevent the algorithm from falling into the local optimal state.

The DE algorithm is a heuristic random search method based on the differences of individuals in the population, which can simulate the population evolution caused by the crossover, mutation and selection of genes. The DE algorithm randomly selects the best individual in the solution space. Through crossover, mutation and selection operations, the individuals of the group evolve from generation to generation until the optimal solution is obtained. Compared with other definite optimization methods, this method is less dependent on the object to be solved, and is the optimal method for solving the optimal solution of complex functions such as nonlinear, high-dimensional, non-differentiable [25], and its algorithm flow is shown in Figure 2.

From Figure 2, the specific process of the DE algorithm is to initialize it first; secondly, it is to evaluate the fitness value of each individual in the group; then, it is to determine whether the optimal solution conditions are met, and if so, the most optimal solution will be output. It is the optimal solution, if not, mutation operation will be performed, which

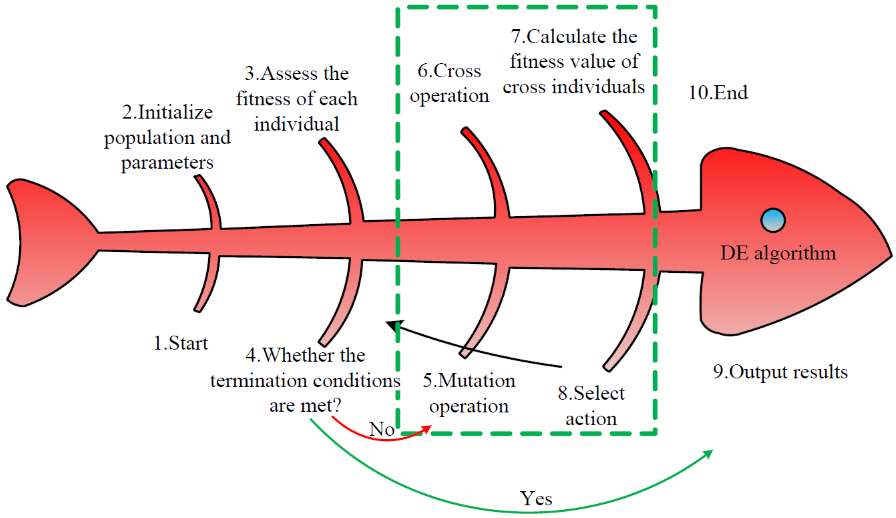


FIGURE 2. Flow diagram of DE algorithm

mutates in the three-generation population; then it recombines the mutated individual with the parent individual to generate a new individual; then it calculates the fitness value of the new individual. A judgment is needed to be made to determine whether it can promote the evolution of the group. If not, it will discard it, and it needs to re-execute the mutation, crossover and selection. If it can, it needs to terminate the process and output the optimal solution. Among them, its mutation is shown in Equation (5).

$$v_i = x_{r1} + F(x_{r2} - x_{r3}) \tag{5}$$

In Equation (5), v_i indicates the variation individual; x_{r1}, x_{r2}, x_{r3} represent three different individuals in the parent group; F severs as the variation factor, which is an F factor used to control the interference of the x_{r2}, x_{r3} vector to the individual. The degree of disturbance to the individual by the score vector is generated by the linear combination of x_{r1}. The higher the value, the greater the variation of the variant v_i, thus determining that the population is more diverse. The crossover is shown in Equation (6).

$$u_{i,j} = \begin{cases} v_{i,j} & rand \leq CR \\ x_{i,j} & rand \geq CR \end{cases} \tag{6}$$

In Equation (6), $u_{i,j}$ means a new individual; *rand* expresses a random number with a value between 0 and 1; CR refers to a crossover factor, whose value is between $[0, 1]$. The two jointly regulate the influence of the variant on the new individual. The selection is shown in Equation (7).

$$x_i^{k+1} = \begin{cases} u_i & f(u_i) < f\left(x_i^k\right) \\ x_i^k & f(u_i) > f\left(x_i^k\right) \end{cases} \tag{7}$$

In Equation (7), k is the algebra of the population; $f(x)$ denotes the objective function defined for practical problems, and its value can judge the individual fitness.

3.2. **DEABC-FC algorithm combining FCM and improved ABC algorithm.** The DE algorithm has greatly expanded the interference processing method of the ABC algorithm, so that the convergence speed and optimization effect have been better improved. Its practical application still needs to be explored, so the study applies it to the clustering. As one of the clustering algorithms, the FCM algorithm has the advantages of wide application range, simple algorithm and fast running speed, but it also has its limitations, such as the sensitivity to the initial point, which will lead to unstable results and large deviations. The search performance is poor, which cannot meet the clustering requirements, etc. [26]. Therefore, the FCM algorithm is effectively combined with the improved DEABC algorithm, and the advantages and disadvantages of the two are combined to complement each other, and the DEABC-FC algorithm is obtained. The FCM algorithm combines hard clustering and fuzzy membership matrix to perform fuzzy division, and the satisfied expression of this matrix is shown in Equation (8).

$$\sum_{j=1}^{k} u_{ij} = 1 \tag{8}$$

In Equation (8), u_{ij} means the numerical value in the matrix U. There are various methods to measure the clustering effect, and the distance measurement is generally used. The study uses the Euclidean distance to measure the similarity of sample points, and the objective function of the FCM algorithm is shown in Equation (9).

$$J = \sum_{j}^{n} \sum_{j=1}^{k} u_{ij}^{m} \left\| x_i - v_j \right\| \tag{9}$$

In Formula (9), m denotes a weighted index, which is used to control the impact of membership on the clustering; x_i represents a data object with dimension D, and v_j means the j cluster center. On this basis, Equation (8) can be derived according to Lagrange multiplier, and the function obtained after derivation is shown in Equation (10).

$$\begin{cases} v_j = \dfrac{\sum_{i=1}^{n} u_{ij}^{m} x_i}{\sum_{i=1}^{n} u_{ij}^{m}} \\ u_{ij} = \dfrac{1}{\sum_{k=1}^{c} \left(\frac{\|x_i - v_j\|^2}{\|x_i - v_k\|^2} \right)^{\frac{2}{m-1}}} \end{cases} \tag{10}$$

In Equation (10), m stands for fuzzy index; k means the number of partitions; c indicates their total number. The DEABC-FC algorithm integrates the FCM algorithm, so the main idea of the DEABC-FC algorithm is that the FCM algorithm can accelerate the clustering. In the DEABC-FC algorithm, the initialized honey source is composed of $k * d$, the center of the dimensional cluster, which represents a fuzzy cluster. The specific process of the DEABC-FC algorithm is shown in Figure 3.

From Figure 3, the DEABC-FC algorithm firstly initializes the nectar source, and divides it according to the FCM, to calculate its fitness value according to Equation (1); secondly, it determines whether it obtains the best nectar source, if there is, it needs to output the nectar source and end the process, if not, it will proceed to the next cycle iteration; then in the bee-picking stage, according to the optimized disturbance equation, it generates a new nectar source. It also performs an FCM division on the new nectar source by using the new clustering center, calculates the fitness value according to Equation (1), and selects the nectar source with the largest fitness after obtaining the fitness value; then it is judged whether it has been updated after multiple disturbances, if not, it should give up; if there is, it will generate a new nectar source and calculate its fitness; finally it returns to the previous in the judgment stage. The process can be ended until the optimal nectar source is output. Among them, the expression of the initialized nectar source Z is shown in Equation (11).

$$Z = [z_1, z_2, \ldots, z_{SN}], \quad z_i = [z_{i1}, z_{i2}, \ldots, z_{im}] \tag{11}$$

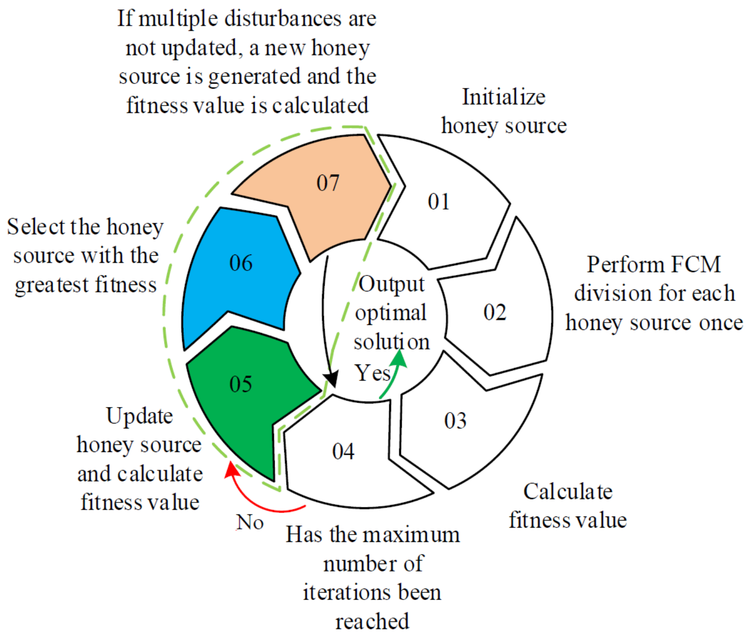


FIGURE 3. Specific flow of DEABC-FC algorithm

In Equation (11), SN indicates the honey source element sequence number; m represents a cluster center, and its calculation is shown in Equation (12).

$$m = k * d \tag{12}$$

In Equation (12), d stands for the upper limit of the dimension j. In addition, in the bee-picking stage, the optimized disturbances are shown in Equations (13)-(15).

$$DEABC/rand/1 : s_i = z_{r1} + F(z_{r2} - z_{r3}) \tag{13}$$

In Equation (13), s_i denotes the vector generated by the DEABC algorithm according to the optimization strategy for mutation.

$$DEABC/best/1 : s_i = z_{best} + F(z_{r1} - z_{r2}) \tag{14}$$

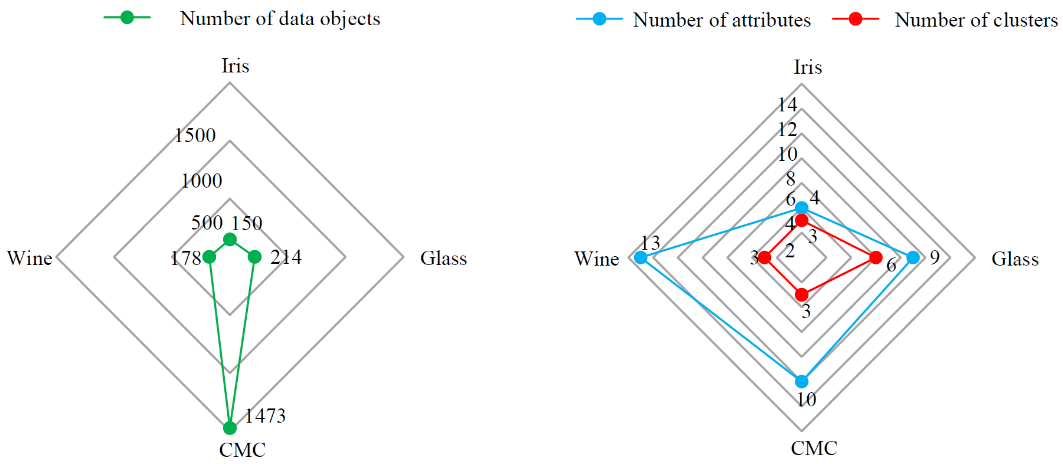
In Equation (14), z_{best} represents the current global optimal solution; $r1$ and $r2$ belong to $[1, SN]$.

$$DEABC/current\text{-}to\text{-}best/1 : s_i = z_i + F(z_{best} - z_i) + F(z_{r1} - z_{r2}) \tag{15}$$

The information used by the three optimized strategies is different. Among them, Equation (13) is to use the existing honey resources and has a high interference efficiency; Equation (14) is to make full use of the current best value of honey source information, so that the algorithm has a faster convergence speed, but its diversity is low, and it is easy to fall into a local optimum; Equation (15) is to make full use of the existing and optimal nectar sources. Its own local search balances the overall and local search capabilities. The advantages and disadvantages are that when using the interference mode of Equation (14), more parent information can be used to obtain a better mutation vector, but it will increase the amount of calculation, thereby reducing the operation efficiency.

The DEABC-FC algorithm introduces the concept of mutation and crossover in differential evolution, improves the traditional random generation method. It also makes it obtain better optimization direction and higher quality candidate solutions, and improves the robustness and convergence rate of the algorithm. Although the concept of differential evolution has some complexity in each iteration, the method is much faster to solve than the required time and space complexity.

4. **Application and Performance Analysis of DEABC-FC Algorithm.** After understanding the idea and process of DEABC-FC algorithm, its specific application and performance can be analyzed. To compare and analyze the clustering performance of DEABC-FC and other clustering methods, the study used a large number of experimental data in the UCI machine learning database to conduct experiments. Before the experiment, the dataset needed to be standardized. A standard test dataset was represented by Iris, Glass, CMC, and Wine. Its content is shown in Figure 4.



(a) Number of data objects in four datasets

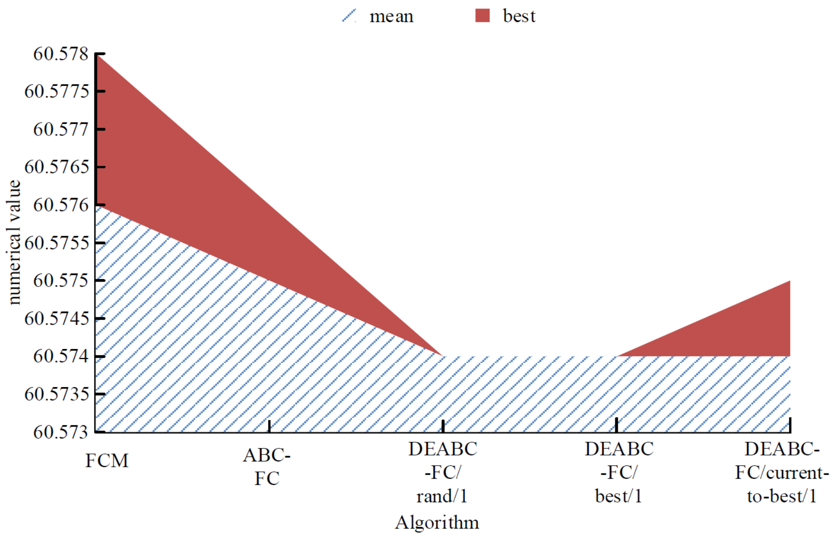
(b) The attributes of the four datasets and the number of clusters

FIGURE 4. Specific contents of the four standard datasets

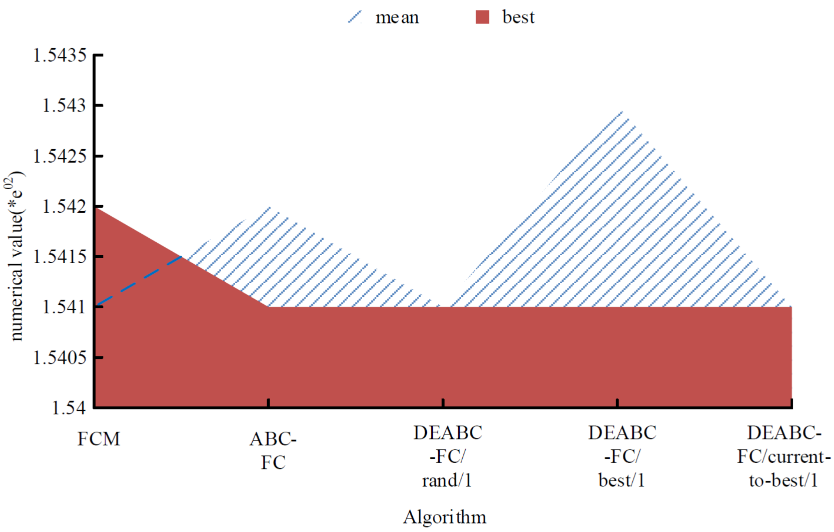
The four standardized datasets in Figure 4 include the Iris, Glass, CMC, and Wine datasets. Figure 4(a) represents the number of data objects and Figure 4(b) indicates the number of attributes and clusters. In addition, a total of 150, 214, 1473, and 178 samples were collected in the Iris, Glass, CMC, and Wine datasets, respectively.

On this basis, ABC-FC algorithm based on ABC was introduced, and it was compared with the FCM algorithm and the DEABC-FC algorithm for 1000 times on four data, respectively. Independent tests were carried out, and the optimal value parameters, average value, standard deviation, average number of iterations, and average operation time were counted, and were represented by *best*, *mean*, *std*, *iter* and *time*, respectively. In addition, for DEABC-FC alone, the study tested its three mutation strategies and the value of its mutation and crossover factor for many times, so that the algorithm could obtain the optimal clustering result. At the same time, two groups of experiments were carried out, denoted by A and B. Group A was used to record the optimal solution, stability and convergence rate of clustering. Group B was used to describe the convergence of clustering. The experimental results on Iris and Glass are shown in Figure 5.

From Figure 5, in these 1000 clustering experiments, the values of the three DEABC-FC mutation strategies on the two datasets *best* were approximately 60.574 and 1.541e02, which were significantly lower than the ABC-FC algorithm and FCM algorithm. The clustering result was smaller, which showed that the DEABC algorithm had better global optimization performance and could better classify. From the *mean* value point of view, the values of the ABC-FC algorithm were 60.575 and 1.542e02, respectively, which were lower than 60.576 and 1.541e02 of the FCM algorithm, indicating that the ABC effectively improved the FCM algorithm. While for DEABC-FC, compared with the ABC-FC algorithm, it has been further improved, so its overall optimization ability was better. In addition, the experimental results on the datasets CMC and Wine are shown in Table 1.



(a) Mean and best values of several algorithms on dataset Iris



(b) Mean and best values of several algorithms on dataset Glass

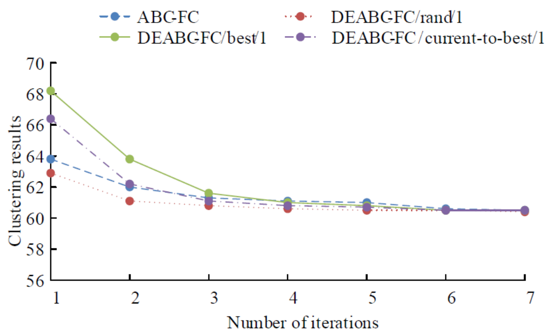
FIGURE 5. Comparison of clustering performance of several algorithms on datasets Iris and Glass

Table 1 shows the comparison of the results of several algorithms on the datasets CMC and Wine. From the values in the table, the values of the three mutation strategy algorithms of DEABC-FC on the CMC and Wine datasets were significantly lower than those of ABC-FC and FCM. Therefore, the performance and global optimization ability of DEABC-FC could be greatly improved through appropriate parameter selection. In addition, the DEABC-FC could effectively improve the convergence speed of the algorithm because it introduced the idea of crossover mutation in differential evolution. Therefore, when the parameters were properly selected, the number of computations of DEABC-FC was significantly reduced, and the convergence speed was greatly accelerated.

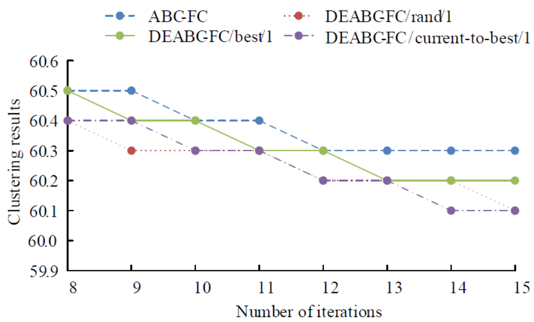
Group B experiments compared the clustering performance of the three mutation strategies of the ABC-FC and the DEABC-FC on four datasets. The experimental record results were also divided into two groups, which combine the four datasets in pairs. The experimental results on Iris and Glass are shown in Figure 6.

TABLE 1. Comparison of clustering performance of several algorithms on datasets CMC and Wine

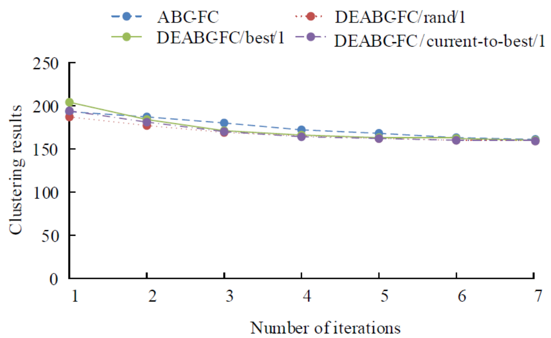
CMC					
–	FCM	ABC-FC	DEABC-FC/rand/1	DEABC-FC/best/1	DEABC-FC/ current-to-best/1
mean	1.814e04	1.813e04	1.812e04	1.812e04	1.812e04
std	1.550e-08	2.568e-07	1.246e-10	1.228e-10	1.265e-10
best	1.814e04	1.813e04	1.812e04	1.812e04	1.812e04
time	0.054	0.813	0.203	0.310	0.301
iter	39.079	39.894	9.905	14.803	14.272
Wine					
mean	1.796e06	1.795e06	1.794e06	1.794e06	1.794e06
std	1.796e-07	1.218e-06	6.052e-09	6.003e-09	5.725e-09
best	1.796e06	1.795e06	1.794e06	1.794e06	1.794e06
time	0.017	0.309	0.108	0.121	0.122
iter	55.859	63.105	17.054	21.053	29.351



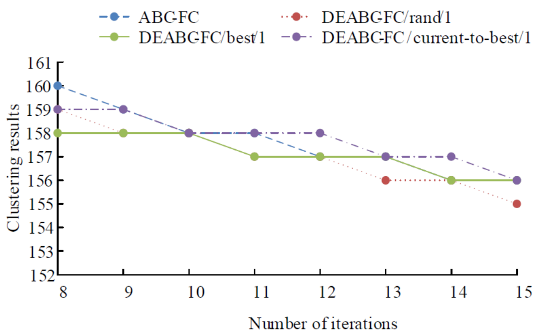
(a) Clustering results with 1-7 iterations on Iris



(b) Clustering results with 8-15 iterations on Iris



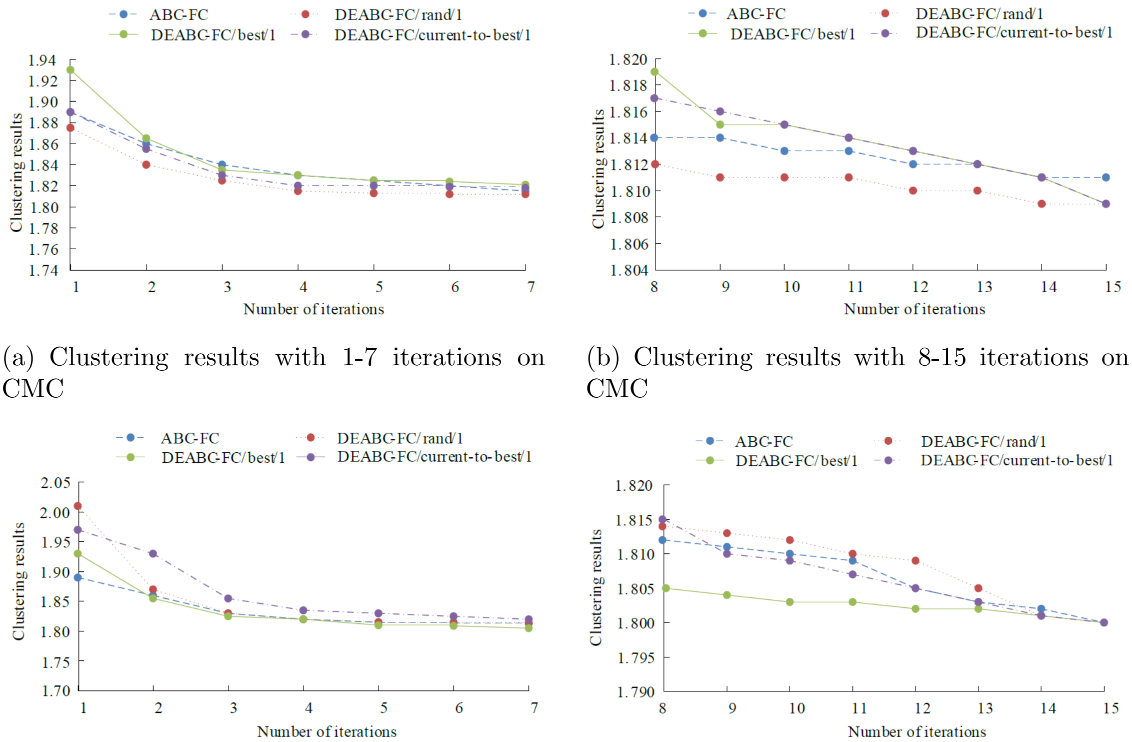
(c) Clustering results with 1-7 iterations on Glass



(d) Clustering results with 8-15 iterations on Glass

FIGURE 6. Clustering results on Iris and Glass

From Figure 6, the three mutation strategies of DEABC-FC in the Iris dataset, the highest were 62.9, 68.2, 66.4 under 15 iterations, and the lowest were 60.1, 60.2, 60.1, respectively, while the ABC-FC had the highest 63.8, the lowest was 60.3; in the Glass dataset, the highest three mutation strategies were 187, 204, 194, and the lowest were 155, 156, 156, respectively, while the ABC-FC had the highest 193, and the lowest was 156. From the experimental results of the two datasets, it was obvious that the DEABC-FC



(a) Clustering results with 1-7 iterations on CMC

(b) Clustering results with 8-15 iterations on CMC

(c) Clustering results with 1-7 iterations on Wine

(d) Clustering results with 8-15 iterations on Wine

FIGURE 7. Clustering results on CMC and Wine

had a faster decline rate than the ABC-FC, indicating that the DEABC-FC could achieve better convergence and ensure that the perturbation moves towards a more favorable direction. In addition, its results on the CMC and Wine datasets are shown in Figure 7.

From Figure 7, the three mutation strategies of DEABC-FC in the CMC dataset had the highest values of 1.875, 1.93, and 1.89 under 15 iterations, and the lowest values of 1.809, 1.809, and 1.809, respectively, while the ABC-FC was the highest of 1.89, the lowest was 1.811; in the Glass dataset, the highest of the three mutation strategies were 2.01, 1.93, and 1.97, and the lowest was 1.8, 1.8, and 1.8, respectively, while the highest of the ABC-FC algorithm was 1.89, and the lowest was 1.8. Likewise, the DEABC-FC performed significantly better than the ABC-FC algorithm in the two datasets, indicating that the improved DEABC-FC could better grasp various information in the population, thereby speeding up the convergence, and a good balance was achieved between global and local search. Therefore, it had a good effect on improving the performance of the clustering algorithm.

In the two groups of experiments, the study found that the clustering effect and convergence rate of the DEABC-FC were affected by variation factors F and cross-over factors CR, so the study used the control variable method to change the value of F and control other related parameters to remain unchanged. At the same time, it set the number of bee colonies to 10, the maximum number of iterations to 300, and the update limit to 30 times. Similarly, the clustering results on the four datasets were divided into two groups. Group C was the experiment on the Iris dataset, and group D was the experiment on the CMC dataset. This experiment mainly focused on three algorithms: FCM, ABC-FC and DEABC-FC/rand/1. The experimental results of group C are shown in Table 2.

Table 2. Clustering results on the dataset Iris

Iris								
–	DEABC-FC/rand/1						ABC-FC	FCM
F	0.1			0.5			–	–
CR	0.1	0.5	0.9	0.1	0.5	0.9	–	–
mean	60.583	60.581	60.579	60.589	60.579	60.579	60.580	60.582
std	8.769e-04	1.612e-07	3.561e-13	2.750e-04	3.725e-13	3.641e-13	2.751e-08	1.345e-07
best	60.582	60.582	60.580	60.583	60.581	60.581	60.583	60.589
time	63.81	16.49	9.11	76.68	17.20	7.69	26.95	23.21

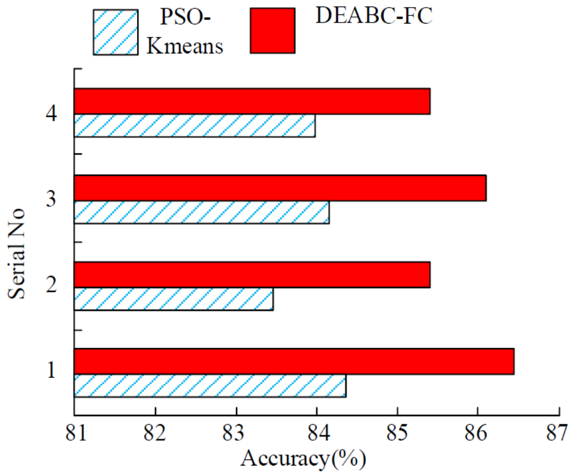
Table 3. Clustering results on the dataset CMC

Glass								
–	DEABC-FC/rand/1						ABC-FC	FCM
F	0.1			0.5			–	–
CR	0.1	0.5	0.9	0.1	0.5	0.9	–	–
mean	1.822e04	1.819e04	1.819e04	1.823e04	1.822e04	1.822e04	1.820e04	1.821e04
std	9.311e-04	5.790e-09	1.239e-10	0.003	2.311e-05	1.251e-10	2.572e-07	1.551e-08
best	1.819e04	1.818e04	1.818e04	1.8120e04	1.818e04	1.817e04	1.820e04	1.821e04
time	71.91	21.01	13.25	83.95	21.11	9.88	40.00	39.10

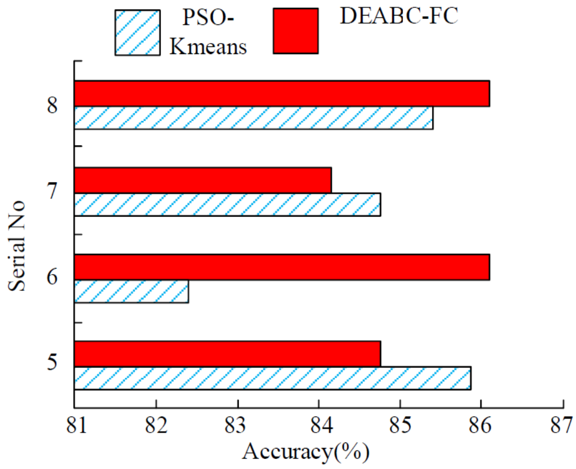
From Table 2, when $F = 0.1$, $CR = 0.1/0.5/0.9$, the *best* values of DEABC-FC/rand/1 were 60.582, 60.582 and 60.580, respectively, which were lower than 60.583 of ABC-FC and 60.589 of FCM. At that time, $F = 0.5$, the *best* values of DEABC-FC/rand/1 were 60.583, 60.581 and 60.581, which were also lower than the other two algorithms. The results showed that the increase of the value F would make the diversity of the group, reduce the risk of premature maturity of the algorithm, and improve the stability of the algorithm, but it would reduce the convergence of the algorithm in some aspects. Therefore, the value of F must be considered population diversity. In addition, the experimental results of group D are shown in Table 3.

From Table 3, with the increase of the CR value, *mean* also increased and *best* decreased at the same time. In addition, although the increase of the CR value led to a decrease in the local development ability of the mutation vector, it also improved the convergence speed and global search ability of the DEABC-FC. Overall, the changes in F and CR values could cope with different datasets, and their results were indeed impressive for the DEABC-FC algorithm. They could significantly enhance the convergence speed and global search ability of the DEABC-FC algorithm, and their transformation could simplify the optimization steps. On this basis, to further verify the effectiveness of DEABC-FC, the research introduced the PSO-Kmeans clustering algorithm in the previous literature to compare the accuracy. The results are shown in Figure 8.

From Figure 8(a), the average accuracy of PSO-Kmeans in the first four experiments was 84.01%, while the average accuracy of DEABC-FC was 85.85%, which was significantly higher than the comparison algorithm. From Figure 8(b), the average accuracy of PSO-Kmeans in the last four experiments was 84.67%, while the average accuracy of DEABC-FC was 85.28%, which was also higher than the comparison algorithm. In general, the standard deviation of DEABC-FC was 0.0074, which was significantly lower than 0.0103 of the comparison algorithm, and had higher algorithm accuracy. To verify the practical application effect of the DEABC-FC algorithm, its application in actual power grid data analysis was studied and compared with the power grid data analysis method



(a) Comparison results of algorithm accuracy in the first four experiments



(b) Comparison results of algorithm accuracy in the last four experiments

FIGURE 8. Comparison results of accuracy of two algorithms

TABLE 4. Comparison results of different methods in actual power grid data analysis

–	32 kB	64 kB	128 kB	256 kB	512 kB	1024 kB
A	91.5%	90.0%	89.5%	87.5%	85.0%	78.5%
B	91.5%	90.5%	90.9%	87.1%	84.8%	76.9%
DEABC-FC	97.5%	96.5%	94.9%	94.9%	92.5%	89.1%

(A) using entropy weight grey model and the data analysis method (B) using Bayesian network. The results are shown in Table 4.

From Table 4, the accuracy of the DEABC-FC algorithm was 97.5%, 96.5%, 94.9%, 94.9%, 92.5%, and 89.1%, respectively, for data sizes of 32 kB, 64 kB, 128 kB, 256 kB, 512 kB, and 1024 kB, which were higher than the comparison method. Overall, the DEABC-FC algorithm had effectiveness and practicality in the analysis of actual power grid data.

5. **Conclusion.** To better explore and improve the defects of traditional clustering methods in segmentation ability, the study improved the ABC algorithm by introducing the DE algorithm. To effectively combine, the DEABC-FC algorithm was obtained, and its application performance was analyzed. The experimental results showed that the mutation strategy of the DEABC-FC algorithm was approximately 60.574 and 1.541e02 on the Iris and Glass datasets, and approximately 1.228e-10 and 6.003e-09 on the CMC and Wine datasets. The numerical values *std* and *best* were lower than those of the ABC-FC and FCM algorithms, showing that the algorithm was more stable and had better performance. In addition, on the four datasets, the clustering results of the DEABC in 15 iterations showed a faster decline, achieving a better balance between global and local search. At the same time, the study kept other influencing factors unchanged, and after changing the value of F alone, it was found that the increase of F could improve the diversity of the group and the stability of the algorithm, but it would also affect the convergence of the algorithm. The increase of the CR value would make the DEABC-FC have higher convergence speed and better global search performance, thus making the optimization effect more stable. In addition, in the comparison of the introduced algorithms, the average accuracy of DEABC-FC was higher than that of the comparison

algorithm. In practical applications, the DEABC-FC algorithm had the highest analysis accuracy of 97.5%, which was much higher than the comparison method. In general, in clustering, the DEABC-FC had better global search ability and convergence speed than the ABC-FC and the FCM, which can better neutralize the shortcomings of the clustering algorithm. It is worth noting that the research proposes that the parameter adaptability of the DEABC-FC is not good in practical applications, which reduces the practicality of the algorithm, so follow-up improvements are needed.

REFERENCES

[1] Y. Wu, S. Shi, J. Liu and H. Zhu, ADL-RAID: Energy-saving data layout for dynamic loads, *International Journal of Computer Systems Science & Engineering*, vol.33, no.5, pp.309-316, 2018.

[2] W. Wei, X. Zhang and L. Yang, Full-cycle state evaluation of S700K switch machine based on residual network and fuzzy clustering, *International Journal of Innovative Computing, Information and Control*, vol.18, no.4, pp.1203-1216, 2022.

[3] A. Buczak and E. Guven, A survey of data mining and machine learning methods for cyber security intrusion detection, *IEEE Communications Surveys & Tutorials*, vol.18, no.2, pp.1153-1176, 2017.

[4] B. Lyu, W. Wu and Z. Hu, A novel bidirectional clustering algorithm based on local density, *Scientific Reports*, vol.11, no.1, pp.1-10, 2021.

[5] N. Xu, R. B. Finkelman, S. Dai, C. Xu and M. Peng, Average linkage hierarchical clustering algorithm for determining the relationships between elements in coal, *ACS Omega*, vol.6, no.9, pp.6206-6217, 2021.

[6] Y. Qu and Y. Wang, Segmentation of corpus callosum based on tensor fuzzy clustering algorithm, *Journal of X-Ray Science and Technology*, vol.29, no.5, pp.1-14, 2021.

[7] D. Huang, J. Wang and Z. Wen, Parallel FCM clustering algorithm of fuzzy number based on cut set, *Journal of Computational Methods in Sciences and Engineering*, vol.21, no.3, pp.1-9, 2021.

[8] T. W. Lu, E. Yao, S. Liu and W. Zhou, Short-time forecast of entrance and exit passenger flow for new line of urban rail transit during growth period, *Tiedao Xuebao/Journal of the China Railway Society*, vol.42, no.5, pp.19-28, 2020.

[9] F. Xu, G. Yang and L. Wang, Artillery structural dynamic responses optimization based on Stackelberg game method, *Journal of Low Frequency Noise, Vibration and Active Control*, vol.41, no.1, pp.140-159, 2022.

[10] F. Nie, Z. Li, R. Wang and X. Li, An effective and efficient algorithm for K-means clustering with new formula, *IEEE Transactions on Knowledge and Data Engineering*, vol.35, no.4, pp.3433-3443, 2023.

[11] N. Shunmuganathan, V. Sheshathri and R. Sankarasubramanian, A hybrid image segmentation method using firefly and artificial bee colony algorithms for color images, *Journal of Algebraic Statistics*, vol.13, no.3, pp.4235-4243, 2022.

[12] L. Ye, B. Dai, M. Pei, P. Lu, J. Zhao, M. Chen and B. Wang, Combined approach for short-term wind power forecasting based on wave division and Seq2Seq model using deep learning, *IEEE Transactions on Industry Applications*, vol.58, no.2, pp.2586-2596, 2022.

[13] J. Wen, T. Yang, Y. Shou, Y. Tang and W. W. Hu, Improved evidential fuzzy c-means method, *Journal of Systems Engineering and Electronics*, vol.29, no.1, pp.187-195, 2018.

[14] K. H. Memon and D. H. Lee, Generalised fuzzy c-means clustering algorithm with local information, *Fuzzy Sets & Systems*, vol.11, no.1, pp.1-12, 2018.

[15] J. Zheng, D. Zhang, K. Huang and Y. Sun, An adaptive image segmentation method based on the fuzzy c-means with spatial information, *IET Image Processing*, vol.12, no.5, pp.785-792, 2018.

[16] X. Deng, X. He and C. Huang, A new fuzzy random multi-objective portfolio model with different entropy measures using fuzzy programming based on artificial bee colony algorithm, *Engineering Computations*, vol.39, no.2, pp.627-649, 2022.

[17] S. Forouzandeh, K. Berahmand, E. Nasiri and M. Rostami, A hotel recommender system for tourists using the artificial bee colony algorithm and fuzzy TOPSIS model: A case study of Tripadvisor, *International Journal of Information Technology and Decision Making*, vol.20, no.1, pp.399-429, 2021.

[18] S. Acquah, Z. Li and A. Krampah-Nkoom, Enhanced algorithm of artificial bee colony (ABC) to optimize models of system reliability, *International Journal of Advanced Research*, vol.9, no.1, pp.835-866, 2021.

[19] B. Zhao, T. W. Sung and X. Zhang, A quasi-affine transformation artificial bee colony algorithm for global optimization, *Journal of Intelligent and Fuzzy Systems*, vol.40, no.4, pp.1-18, 2021.
[20] K. H. Baesmat, I. Masodipour and H. Samet, Improving the performance of short-term load forecast using a hybrid artificial neural network and artificial bee colony algorithm, *Canadian Journal of Electrical and Computer Engineering*, vol.44, no.3, pp.275-282, 2021.
[21] W. F. Gao, L. L. Huang, S. Y. Liu and C. Dai, Artificial bee colony algorithm based on information learning, *IEEE Transactions on Cybernetics*, vol.45, no.12, pp.2827-2839, 2017.
[22] A. Zabidi, N. M. Tahir, I. Mohd and Z. I. Rizman, The performance of binary artificial bee colony (BABC) in structure selection of polynomial NARX and NARMAX models, *International Journal on Advanced Science Engineering and Information Technology*, vol.7, no.2, pp.373-379, 2017.
[23] W. Gu, Y. Yu and H. Wei, Artificial bee colony algorithm-based parameter estimation of fractional-order chaotic system with time delay, *IEEE/CAA Journal of Automatica Sinica*, vol.4, no.1, pp.107-113, 2017.
[24] I. B. Jingle and J. J. A. Celin, Mining optimized positive and negative association rule using advance ABC algorithm, *Journal of Theoretical and Applied Information Technology*, vol.95, no.24, pp.6846-6855, 2017.
[25] Y. X. Su and C. Rui, Multi-objective particle swarm-differential evolution algorithm, *Neural Computing & Applications*, vol.28, no.2, pp.1-12, 2017.
[26] J. Yang, Y. S. Ke and M. Z. Wang, An adaptive clustering segmentation algorithm based on FCM, *Turkish Journal of Electrical Engineering and Computer Sciences*, vol.25, no.6, pp.4533-4544, 2017.

Author Biography

Kan Ni received the B.Sc. degree in Electrical Engineering and Automation from Nanjing University of Posts and Telecommunications, China, 2015; the M.Sc. degree in Electronic and Information Engineering · Mathematical Sciences from Gunma University, Japan, 2020; the M.Sc. degree in Agricultural Electrification and Automation from Yangzhou University, China, 2020.

Mr. Ni is currently a full-time Ph.D. candidate at the Electronic and Information Engineering · Mathematical Sciences, Gunma University, Japan. His research interests include image processing, fuzzy reasoning, artificial intelligence, and neuro-fuzzy learning algorithms for system identification. He has published over 10 papers in journals and conferences.