



A Kernel Fuzzy C-means Clustering Algorithm Based on Firefly Algorithm

Chunying Cheng^(✉) and Chunhua Bao

College of Computer Science and Technology,
Inner Mongolia University for Nationalities, Tongliao 028043, China
chengchunying_80@163.com

Abstract. In this paper, the KFCM algorithm selects the Gaussian kernel function, maps the data into the high-dimensional feature space for clustering, and uses the optimal clustering center of the firefly algorithm as the initial value of KFCM, and then processes it through KFCM. Based on class analysis, a kernel fuzzy C-means clustering algorithm based on firefly algorithm (FA-KFCM) is proposed. Numerical experiments results show that FA-KFCM is superior to other algorithms in clustering accuracy and time efficiency.

Keywords: Firefly algorithm · Clustering · Kernel fuzzy C-means clustering · The classical UCI data set

1 Introduction

The kernel fuzzy C-means clustering algorithm introduces the idea of kernel in the traditional FCM algorithm, solves the problem of linear indivisibility, and realizes effective clustering of various data structures. Currently, it has been widely applied in many fields, but there are still many problems, such as sensitivity to the initial clustering center, long calculation time, diversity of kernel functions and parameter selection. [1] Therefore, we can use the kernel method to improve the accuracy of clustering.

A new swarm intelligence algorithm, the firefly algorithm (FA) [2], was proposed by Yang in 2008. FA is an interdisciplinary research achievement using swarm intelligence and stochastic algorithms, and it is strong and fast. The firefly algorithm has become an increasingly important tool of Swarm Intelligence that has been applied in almost all areas of optimization, as well as engineering practice.

FA has been used in several fields, for example, for solving minimizing the makespan for the permutation flow shop scheduling problem [3], solving the task graph scheduling problem [4] and solving the Job shop scheduling problem [5]. This study applies The FA-KFCM algorithm overcomes the FCM algorithm trapped local optimum and being sensitive to initial value effectively, and enhances the capacity of local search of firefly algorithm. And the validity of the algorithm is verified by Iris, Cmc, Wine and Zoo data in the public dataset.

2 FA-KFCM Algorithm

In this section, we first describe the firefly algorithm and kernel fuzzy C-means. Then, we combine the firefly algorithm kernel fuzzy C-means with mechanism obtain our FA-KFCM algorithm.

2.1 FA Algorithm

Firefly algorithm is based on a physical formula of light intensity I hat decreases with the increase of the square of the distance r^2 . However, as the distance from the light source increases, the light absorption causes that light becomes weaker and weaker. These phenomena can be associated with the objective function to be optimized.

- (1) All fireflies are unisex.
- (2) Their attractiveness is proportional to their light intensity.
- (3) The light intensity of a firefly is affected or determined by the landscape of the fitness function.

In the standard firefly algorithm, the light intensity I of a firefly representing the solution s is proportional to the value of fitness function $I(x) \propto f(x)$, whilst the light intensity $I(r)$ varies according to the following equation: [6]

$$I(r) = I_0 \times e^{-\gamma r^2} \tag{1}$$

The attractiveness β_0 of fireflies is proportional to their light intensities $I(r)$. The attraction β can be described by Eq. (1)

$$\beta(r) = \beta_0 \times e^{-\gamma r^2} \tag{2}$$

Where β_0 is the attractiveness at $r = 0$. The light intensity I and attractiveness β are in some way synonymous.

The distance between any two fireflies x_i and x_j in the basic firefly algorithm is calculated from the Euclidean distance:

$$r_{ij} = \|x_i - x_j\| = \sqrt{\sum_{k=1}^n (x_{i,k} - x_{j,k})^2} \tag{3}$$

Firefly i is moved by the attraction of its bright firefly j , The position of the movement can be expressed as:

$$x_i = x_i + \beta_0 \times e^{-\gamma r_{ij}^2} \times (x_j - x_i) + \alpha(rand - 0.5) \tag{4}$$

Where x_i, x_j are the positions of fireflies i and j in the solution space; is the step factor, which is the constant on the interval $[0, 1]$; $rand$ is the random factor, which obeys the interval $[0, 1]$ Evenly distributed.

2.2 KFCM Algorithm

The kernel function clustering method has a much better performance than the classical clustering algorithm. The kernel clustering method performs better clustering of samples by non-linear mapping of samples of the input space.

If the sets $Y = \{y_j | j = 1, 2, \dots, M\}$, the objective function of fuzzy C-means is as shown in formula (5)

$$J_b(U, V) = \sum_{j=1}^M \sum_{i=1}^C u_{ji}^m \|y_j - v_i\|^2 \tag{5}$$

Where u_{ji} is

$$\forall j, \sum_{i=1}^C u_{ji} = 1; \forall j, i, \mu_{ji} \in [0, 1]; \forall i, \sum_{j=1}^M \mu_{ji} > 0 \tag{6}$$

Using Lagrange multiplier method:

$$\bar{J}(U, V, \delta) = J_b(U, V) + \sum_{j=1}^M \delta_j (\sum_{i=1}^C \mu_{ji} - 1) \tag{7}$$

Introducing nonlinear mapping here $\varphi : y \rightarrow \varphi(y)$

$$\|\varphi(y_j) - \varphi(v_i)\| = KFCM(x_j, x_j) + KFCM(v_i, v_i) - 2KFCM(y_j, v_i) \tag{8}$$

The objective function of KFCM is:

$$J_\varphi = \sum_{j=1}^M J_j = \sum_{j=1}^M \sum_{i=1}^C \mu_{ji}^b \|\varphi(y_j) - \varphi(v_i)\|^2 \tag{9}$$

Here the kernel function selects the Gaussian function as:

$$KFCM(y, x) = \exp[-(y - x)^2 / \sigma^2] \tag{10}$$

Substituting Eq. (10) into Eq. (8), then Eq. (9) is:

$$J_\varphi = 2 \sum_{j=1}^M \sum_{i=1}^C \mu_{ji}^b [1 - KFCM(y_j, v_i)] \tag{11}$$

For J_φ , the partial derivatives of u and v are separately obtained as partial derivatives

$$v_j = \frac{\sum_{j=1}^M \mu_{ji}^b KFCM(y_j, v_i) y_j}{\sum_{j=1}^M \mu_{ji}^b KFCM(y_j, v_i)} \tag{12}$$

$$\mu_{ji} = \frac{(1 - KFCM(x_j, v_i))^{-1/(b-1)}}{\sum_{i=1}^C (1 - KFCM(x_j, v_i))^{-1/(b-1)}} \tag{13}$$

Kernel fuzzy C-means first calculates the kernel function by the formula (10), and updates the membership matrix according to the formulas (12) and (13) until the optimal cluster center is found, and then the final clustering result is obtained.

2.3 The Proposed FA-KFCM Algorithm

In the FA-KFCM algorithm, the position of each firefly represents a clustering center, expressed in terms of vector $V = \{v_1, v_2, \dots, v_C\}$, where v_i is the i th cluster center. The light intensity of the firefly is determined by the objective function of the fuzzy clustering. According to the characteristics of the firefly algorithm, the light intensity function of the firefly can be defined as:

$$I(V) = \frac{1}{1 + J_\varphi(U, V)} \tag{14}$$

The specific algorithm steps of the KFCM algorithm based on the FA are:

- Step 1: Initialize the light absorption coefficient γ , randomized parameter α , Maximum number of iterations T_{max} , Maximum attraction β_0 , C and b .
- Step 2: Initialize the position of the firefly V_1, V_2, \dots, V_N .
- Step 3: Calculate the $KFCM(y_j, v_i)$.
- Step 4: For each firefly, calculate the U according to Eq. (12), calculate the light intensity of each firefly $I(V_j)$ according to Eq. (14).
- Step 5: Compare the light intensity of fireflies. If $I(V_i) > I(V_j)$, it means that firefly j is in a good position, attract fireflies i to move to itself, and calculate the attraction and update position according to formula (2) and formula (4).
- Step 6: According to (12), update the membership matrix of the fireflies.
- Step 7: According to (14), recalculate the light intensity of the fireflies.
- Step 8: Repeats steps (5) to (7), until the finds out the number of iterations is found.
- Step 9: Output result.

2.4 Experimental Evaluation

This article uses MATLAB7.0 as a tool for programming under the Windows 8 operating system. The experiments used Iris, Cmc, Wine and Zoo public dataset in the UCI dataset to test the clustering accuracy and time efficiency of the FA-KFCM

algorithm proposed in this paper, and compares it with the references [7] and [8]. The results are shown in Table 1.

Table 1. Data experiment sample dataset.

Dataset	Class	Dimension	Number of samples
IRIS	3	4	180
CMC	3	9	1800
WINE	3	13	150
ZOO	3	8	110

In this paper, the number of clusters is 3, fuzzy index $b = 2$, maximum iteration number $T_{max} = 100$, light absorption coefficient $\gamma = 0.9$, Maximum attraction $\beta_0 = 1$, randomized parameter $\alpha = 0.1$. After running the current algorithm 50 times, the results are shown in Table 2.

Table 2. Comparison of clustering results of three algorithms on dataset.

Algorithm	Average number of iterations				The maximum number of iterations				Mean fitness value			
	Iris	Cmc	Wine	Zoo	Iris	Cmc	Wine	Zoo	Iris	Cmc	Wine	Zoo
Ref [7]	100	100	100	100	32	37	41	36	89.4	5545	16487	3531
Ref [8]	100	100	100	100	26	29	34	22	86.6	5378	15799	3269
FA-KFCM	100	100	100	100	21	20	25	19	85.1	5217	15541	3120

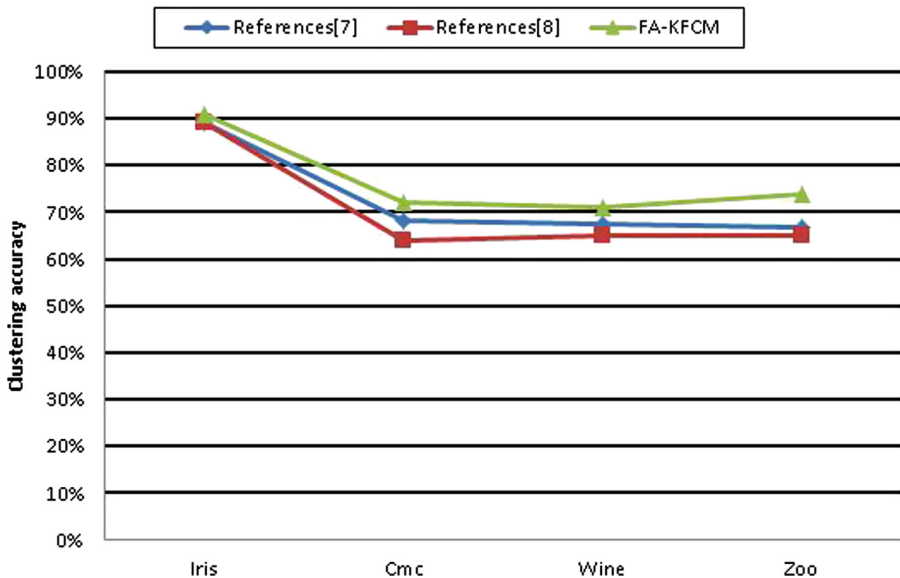


Fig. 1. Comparison of clustering precision of three algorithms.

From the comparison results of Table 2, the results obtained by FA-KFCM are obviously better than those of other algorithms and the clustering quality of FA-KFCM is better. For four datasets, the Average number of iterations and the maximum number of iterations obtained by the algorithm in this paper are all the minimum of several clustering algorithms, which further shows that the clustering quality of the algorithm is better.

It could be learned from Fig. 1 that when clustering four data samples, Iris, Cmc, Wine and Zoo, the average clustering accuracy of the FA-KFCM algorithm is better than that of the other two algorithms.

3 Conclusion

In this paper, the FA algorithm and KFCM algorithm are analyzed in detail, and the FA-KFCM algorithm is proposed. The algorithm uses the optimal clustering center of the firefly algorithm as the initial value of KFCM, and then processes the clustering analysis through KFCM. In order to verify the validity of the new algorithm, the author used Iris, Cmc, Wine and Zoo data for numerical experiments and compared with the literature algorithm. The experimental results show that the proposed algorithm has better performance and good clustering results.

Acknowledgments. This work was supported by Inner Mongolia University for Nationalities Funds of China under Grant No. NMDYB18008.

References

1. Kim, D.W., Lee, D.: Evaluation of the performance of clustering algorithms in kernel-Induced feature space. *Pattern Recogn.* **38**, 607–611 (2005)
2. Yang, X.S.: Firefly algorithm. In: *Nature-Inspired Metaheuristic Algorithms*, pp. 79–90 (2008)
3. Sayadi, M., Ramezani, R., GhaariNasab, N.: A discrete firefly metaheuristic with local search for makespan minimization in permutation flow shop scheduling problems. *Int. J. Ind. Eng. Comput.* **1**(1), 1–10 (2010)
4. Hnig, U.: A firefly algorithm-based approach for scheduling task graphs in homogeneous systems. In: *Informatics*, pp. 24–33 (2010)
5. Khadwilard, A., Chansombat, S., Thepphakorn, T.: Application of firefly algorithm and its parameter setting for job shop scheduling. In: *1st Symposium on Hands-On Research and Development*, pp. 1–10 (2011)
6. Yang, X.S.: Optimization and metaheuristic algorithms in engineering. In: *Metaheuristic in Water Geotechnical and Transport Engineering*, pp. 1–23. Elsevier (2013)
7. Zhongwei, G., Fuyuan, X.: Clustering algorithm based on adaptive chaotic particle swarm optimization. *Comput. Eng. Des.* **36**(6), 1510–1585 (2015)
8. Kuo, R.J., Syu, Y.J., Zhenyao, C.: Integration of particle swarm optimization and genetic algorithm for dynamic clustering. *Inf. Sci.* **195**(15), 124–140 (2012)