

Novel Fuzzy Clustering Algorithm Based on Fireflies

Li Dan^{1,a}, Luo Ke^{1,b}, Sun Zhen^{2,c}

¹ Institute of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha 410114, China

² Institute of Computer and Communication Engineering, Zhengzhou Institute of Light Industry, Zhengzhou, 450002, China

^alid2014happy@126.com, ^bluok@csust.edu.cn, ^c zzsunzhen@126.com

Keywords: Fuzzy c-means; Firefly; Chaos theory

Abstract. Aiming at the existence of fuzzy C-means algorithm was sensitive to the initial clustering center and its shortcoming of easily plunged into local optimum, this paper proposed a novel fuzzy clustering algorithm based on fireflies. The algorithm employed the chaos initialization individuals as the initial population. Then it utilized the improved fireflies as the accurately clustering center and received a new clustering center as the initial clustering center of fuzzy C-means. Thus it can overcome the fuzzy C-means' sensitivity to the initial clustering center and solve the deficiency of easily falling into local optimum. Simulation experiment results based on UCI standard data sets show that the algorithm can avoid falling into local optimum and precocious, it also gets better performance and results compared with other algorithms.

Introduction

Clustering is a process of dividing collection of physical or abstract objects into a certain amount of classes or clusters. It makes the clusters in the similar degree is high, low degree of similarity between clusters. Clustering analysis has been widely used in many fields, such as data mining, biology, statistics, image processing, pattern recognition and business intelligence, etc [1].

Fuzzy C-means algorithm has the characteristics of simple calculation and speediness, it has been widely applied. However, there are still existing problems like falling into local optimal easily and the algorithm is sensitive to the initial clustering center, etc. Therefore, in order to overcome these shortcomings of the fuzzy C-means algorithm, many scholars combine with the fuzzy C-means algorithm and particle swarm optimization (pso) [2], artificial bee colony algorithm [3], genetic algorithm, ant colony algorithm and artificial fish algorithm [4].

In 2008, Yang Xinshe proposed firefly algorithm [5,6]. The firefly algorithm is a novel and intelligent optimization algorithm, which mimic natural biological fireflies glow and the biological characteristics of mutual attraction. The algorithm has a good search and optimization performance. It dropped some biological significances of the firefly, only using their luminescence properties to search its sight partners, through comparing the fluorescence intensity and the degree of attraction to determine a certain direction and, realizing the optimized position in the process of iteration. Algorithm has been applied to many areas such as image processing [7], pipeline scheduling [8], path planning.

Firefly algorithm has strong global search ability and convergence speed, less algorithm parameters, and it is suitable for parallel processing. This paper combined firefly algorithm with good local searching ability of the fuzzy C-means, took the advantages of both algorithms in order to get better results. Therefore, this paper proposed a novel fuzzy clustering algorithm based on fireflies, using logistic map produced by chaos variable initialization population, and modifying parameters of fireflies' location updating formula, utilizing the characteristics of the chaos optimization algorithm in order to overcome premature convergence. In addition, it can improve the population diversity and searching ergodicity by chaos theory. What's more, it avoids trapping into local optimum and improving the ability of global optimization.

Propaedeutics

Fuzzy C-means algorithm. To a limited data set on given feature space $X = \{x_1, x_2, \dots, x_n\}$, X is divided into c ($2 \leq c \leq n$) class, Assuming a class of c for clustering center $V = \{v_1, v_2, \dots, v_c\}$, u_{ij} is sample of the membership degree matrix, $u_{ij} \in [0,1]$, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, c$; m is the fuzzy weighted index, this paper sets $m \in [1,3]$, the objective function of FCM algorithm can be formulated as:

$$J_f(U, V) = \sum_{i=1}^n \sum_{j=1}^c u_{ij}^m \|x_i - v_j\|^2. \quad (1)$$

Formula (1) is subjected to the $u_{ij} \in [0,1]$, $J_f(U, V)$ achieves minimum [9]. According to the Lagrange multiplication, derivative of formula (1) under the constraints of $u_{ij} \in [0,1]$ can be as:

$$v_j = \frac{\sum_{i=1}^n u_{ij}^m x_i}{\sum_{i=1}^n u_{ij}^m}; u_{ij} = 1 / \sum_{k=1}^c \left(\|x_i - v_j\|^2 / \|x_i - v_k\|^2 \right)^{2/m-1}. \quad (2)$$

Firefly algorithm. The firefly algorithm's critical factor to realize optimization is light intensity and attractiveness. The location of the firefly is the better, the greater its brightness is, that is to say, a firefly's attractiveness is proportional to the light intensity, the brightness and attractiveness decreases with the increase of distance. In the process of optimization the related definition can be stated as

$$I = I_0 e^{-\gamma r}. \quad (3)$$

Formula(3)is the relative fluorescence intensity. Where, I_0 is the strongest fluorescence intensity at $r = 0$, γ is absorption coefficient for media, r_{ij} is the distance between firefly i and j .

$$\beta = \beta_0 e^{-\gamma r_{ij}^2}. \quad (4)$$

Formula(4)is the attractiveness .Where, β_0 is the biggest attractiveness.

$$x_i = x_i + \beta_0 e^{-\gamma r_{ij}^2} (x_j - x_i) + \alpha [rand - 1/2]. \quad (5)$$

Formula(5)is the movement of firefly j attracted another less attractive firefly i . Where, x_i is the position of the firefly i , x_j is the position of the firefly j . α is step length coefficient in initialization time, $rand$ is a random number generator uniformly distributed in $[0,1]$ [5,6].

Chaos theory. The process of solving the optimization problem is that, chaotic sequences from chaotic space are mapped to the solution space, thus the optimization process is mapped to the traversal of chaotic orbit search process. So that it can jump out of local extreme value point and avoid precocity and then obtain the global optimal solution.

Logistic mapping is a class of chaotic system. It is simple but has widely used. It can be stated as

$$l_{k+1} = \mu l_k (1 - l_k), l_k \in (0,1). \quad (6)$$

Where, l_k is real value sequence, $k = 0, 1, 2, 3, \dots$; μ is parameter ($\mu \in [0,4]$), When the chaotic mapping in the mapping state, sequence generated by logistic map has characteristic of chaotic sequences in a certain initial conditions of l_0 [10].

Proposed IFAFCM approach

Although the firefly algorithm has characteristics of the strong global search ability and convergence speed, less algorithm parameters and it is suitable for parallel processing. The algorithm has the shortcomings of precociousness and slow convergence speed in the later. So this paper introduces chaos theory on the basis of firefly algorithm, chaos optimization can overcome the shortcomings of the premature convergence. Firstly, Chaos theory was used to optimize the initial population. Secondly, it modifies parameters of the firefly's location updating formula. It traverses the feasible region so that it improves the global search ability and then gets the global optimal.

Chaos initialization. The algorithm applies the logistic map which is most classic and widely used. The processes of producing chaos initialization are introduced as follows:

a. Firstly, generated a d dimensional vector $l_1 = [l_{1,1}, l_{1,2}, \dots, l_{1,d}]$ randomly, and $l_{1,d} \in [0,1]$.

- b. Put l_i as chaos initialization vector to iterate, chaotic variables which are obtained by formula (6) mapping to the solution space from chaotic space.
- c. Calculate the objective function value and sort it, obtain better individual.

Chaos initialization is retaining the initialization algorithm of random nature, and using the chaos characteristics improved the searching ergodicity and the diversity of population.

Modify the firefly's position updating formula parameters.By using logistic mapping system improvement after update position of fireflies is given by (7):

$$x_i = x_i + \beta_0 e^{-\gamma(t)r_i^2} + \alpha(t)[rand - 1/2] \quad (7)$$

Where, $\alpha(t) = [1 - \alpha(t-1)]\gamma(t-1)\mu_2$, $\gamma(t) = [1 - \gamma(t-1)]\gamma(t-1)\mu_1$; μ_1 and μ_2 are chaos theory parameters, $\mu_1 = \mu_2 = 4$.

After each iteration, logistic mapping system produces three random parameters values which belong to [0,1]. A chaotic variable that logistic chaotic map generated has the characteristics of aperiodicity and ergodicity. Therefore, it can traverse the total value of the feasible region, so as to avoid falling into local optimum and improve the ability of global optimization.

Algorithm description.1) Fireflies coding

Assumes that the data dimension is d , dividing the class cluster into c , each firefly is composed of c clustering center vector. Therefore, the location of the firefly is $L_i(C_{i1}, C_{i2}, \dots, C_{ij}, \dots, C_{ik})$, C_{ij} is the i fireflies center of the coordinates of the vector which is the j clustering center. Each cluster center is d dimensional vector, so the location of the firefly is $c \times d$ dimensional vector.

2) Algorithm steps

Step 1: Initialize the algorithm's parameters, the individual which is produced by chaos initialization is the initial population;

Step 2: Formula (1) is used to calculate each individual target, as its maximum brightness

Step 3: The relative brightness of l and attractiveness of β between each firefly are calculated respectively by the formula (3) and formula (4);the firefly mobile direction shall be determined according to the size of l . Using the formula (7) to update the location of the fireflies, disturb in the best position of fireflies randomly.

Step 4: According to the location of the fireflies after the location of updating, recalculate the brightness of the fireflies.

Step 5: To rank all fireflies according to brightness of the fireflies, determine the current optimal position;

Step 6: Whether algorithm has reached termination conditions, if algorithm has reached, the output of the optimal partition clustering, otherwise return Step 3;

Step 7: The new clustering center is regarded as the initial clustering center;

Step 8: According to the fuzzy C-means algorithm to calculate the new clustering center;

Step 9: Repeating Step 8. It can stop until reach convergence law or maximum number of iteration.

Experiment results analysis

The simulation environment: operating system: Windows 7; compile software: Microsoft Visual C++6.0, Matlab 2012, Inter Core i5 CPU, memory: 4GB.

To verify the feasibility and effectiveness of the algorithm in this paper, the result of algorithm in this paper compares to the standard of FCM algorithm and the other algorithm based on the literature [2] (Hereinafter referred to as MPSO) on the Glass and the Iris standard data set. Experiment involves the characteristics of several of data sets as shown in table 1.

Table 1 Characteristics of data sets in the experiment

Data set	Number of samples	Number of attributes	Number of categories
Iris	450	4	3
Glass	214	9	6

The algorithm analyzed respectively with FCM algorithm, MPSO algorithm and proposed algorithm on the IRIS and GLASS data set in this paper. Algorithms of fuzzy weighted index is $m = 2$. Parameter of the MPSO algorithm is set as: Size scale particle groups $N = 20$, $\omega\phi_1$ linearly decreases from 0.9 to 0.4, $\eta_1 = \eta_2 = 0.5$. IFAFCM algorithm parameter is set as: population size $N = 20$, the largest number of iterations $\max T = 500$, step length coefficient $\alpha = 0.5$, the biggest attraction $\beta_0 = 1$, light absorption coefficient $\gamma = 1$. The experimental results are shown in table 2.

Table 2 Comparison of accuracy of clustering and number of iterations for various clustering algorithms

Data set	Algorithm	Accuracy of clustering (%)	Number of iterations
Glass	FCM	25.69	62
	MPSO	73.41	96
	IFAFCM	80.56	55
Iris	FCM	89.14	30
	MPSO	92.00	57
	IFAFCM	93.78	26

By comparing the experimental results in table 2, accuracy of clustering algorithm is superior to other algorithms on the Glass and the Iris data set in this paper. In addition, accuracy of clustering algorithm is higher on the Glass data set. Through the comparison of the number of iterations, it shows that the algorithm can quickly find the optimal solution, the accuracy and rate of convergence of the algorithm are proven, it improve the clustering effect significantly.

With the changing of the number of iterations increases, the objective function value of MPSO, FCM and IFAFCM algorithms on the Glass and the Iris data set is shown in Figure 1 and Figure 2.

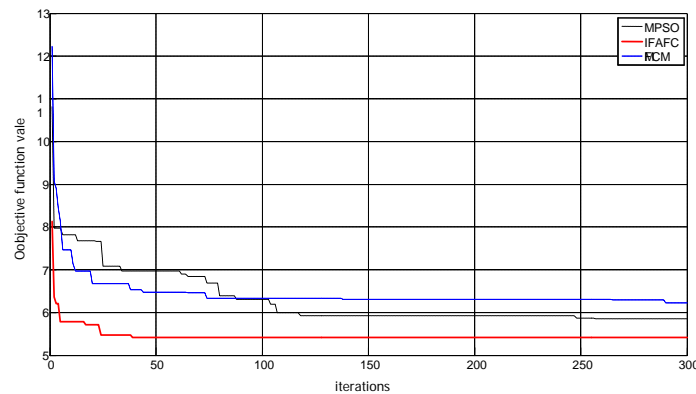


Figure 1 Comparison of curves of objective function for Glass data set using three algorithms

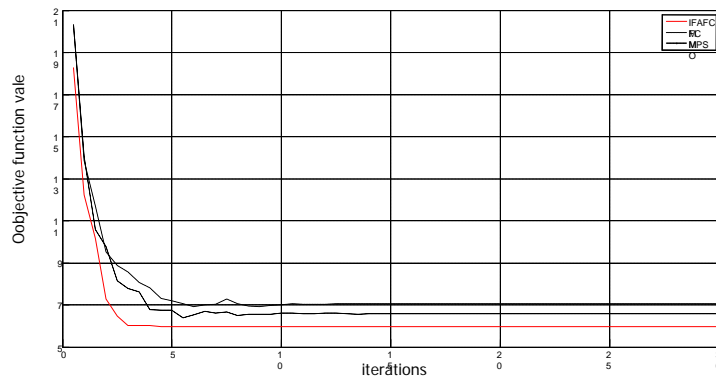


Figure 2 Comparison of curves of objective function for Iris data set using three algorithms

Figure 1 and Figure 2 shows: The early uses of chaos initialization fireflies as initial population in each data set in this paper, the objective function values is very small at the beginning of the iteration, and reach the local optimum quickly. With the increase of the number of iterations, chaotic state integrated into the fireflies position of updating parameter, thus jumped out of local optimum then achieve the global optimal finally. Algorithm is superior to other algorithms on the number of iterations and the convergence speed; it shows the stability and robustness of this algorithm.

Summary

This paper proposed a novel fuzzy clustering algorithm based on fireflies. The algorithm made up for the FCM algorithm's easily falling into local optimum, and solved the FCM algorithm's sensitive to the initial clustering center of the problem. The algorithm utilized the logistic map chaotic variables as firefly's population initialization in order to speed up the convergence. Chaotic strategy was introduced, this algorithm made the firefly algorithm parameters transforms randomly in the feasible region, thus it can jump out of local optimum. Ultimately, it improved the global search ability and achieved better clustering effect and higher accuracy.

Acknowledgements

This work was supported by National Natural Science Foundation of China under Grant NO.11171095, 71371065; Supported by Natural Science Hengyang Mutual Fund of Hunan Province under Grant NO.10JJ8008; Supported by the Science and Technology Planning Project of Hunan Province of China under Grant NO.2013SK3146.

References

- [1] Han J W, Kan B, Pei J Z. Data mining concepts and techniques 3rd. Beijing: China Machine Press. 288-292,2012.
- [2] Pu P B, Wang G, Liu T A; Research of improved fuzzy C-means algorithm based on particle swarm optimization. Computer Engineering and Design ,Vol. 29,no.16,p.4277-4279,2008.
- [3] Luo K, Li L, Zhou B X. A honey-bee mating optimization clustering algorithm. Acta Electronica Sinica, vol. 42, no.12, p.2435-2441, 2014.
- [4] Xiao L C, Ying Z, Jun T S, et al. Method of image Segmentation based on fuzzy C-means clustering Algorithm and artificial fish swarm algorithm. ICISS, p.254-257.2010
- [5] Yang X S. Nature-inspired metaheuristic algorithms. [S.l.]: Luniver Press, p.83-96, 2008.
- [6] Yang X S. Firefly algorithms for multimodal optimization: Stochastic Algorithms. Foundations and Applications. Berlin: Springer Verlag, p.169-178, 2009.
- [7] Horng M H, Jiang T W. Multilevel image thresholding selection based on the firefly algorithm. UIC/ATC, IEEE, p.58-63, 2010.
- [8] Student U G. A multi-agent brokering approach and jumper firefly algorithm for job scheduling in cloud computing. ICICA, no.21, p.52-58, 2014.
- [9] Bezdek J C, Ehrlich R, Full W. FCM: The fuzzy c-means clustering algorithm. Computers & Geosciences, Vol.10, no.2, p.191-203, 1984.
- [10] Gandomi A H, Yang X S, Talatahari S, et al. Firefly algorithm with chaos. Communications in Nonlinear Science and Numerical Simulation, Vol.18, no.1, p.89-98, 2013.