Hindawi
Mathematical Problems in Engineering
Volume 2018, Article ID 8724084, 8 pages
https://doi.org/10.1155/2018/8724084

Research Article

Research on Clustering Method of Improved Glowworm Algorithm Based on Good-Point Set

Yaping Li,[1,2,3] **Zhiwei Ni**,[1,2] **Feifei Jin**,[1,2] **Jingming Li**,[1,2] **and Fenggang Li**[1,2]

[1]*School of Management, Hefei University of Technology, Hefei, Anhui 230009, China*
[2]*Key Laboratory of Process Optimization and Intelligent Decision-Making, Ministry of Education, Hefei, Anhui 230009, China*
[3]*Anhui Economic Management Institute, Hefei, Anhui 230059, China*

Correspondence should be addressed to Zhiwei Ni; zhiwein@163.com

Received 26 October 2017; Accepted 8 January 2018; Published 5 March 2018

Academic Editor: Mohammed Nouari

As an important data analysis method in data mining, clustering analysis has been researched extensively and in depth. Aiming at the limitation of K-means clustering algorithm that it is sensitive to the distribution of initial clustering center, Glowworm Swarm Optimization (GSO) Algorithm is introduced to solve clustering problems. Firstly, this paper introduces the basic ideas of GSO algorithm, K-means algorithm, and good-point set and analyzes the feasibility of combining them for clustering optimization. Next, it designs a clustering method of improved GSO algorithm based on good-point set which combines GSO algorithm and classical K-means algorithm together, searches data object space, and provides initial clustering centers for K-means algorithm by means of improved GSO algorithm and thus obtains better clustering results. Major improvement of GSO algorithm is to optimize the initial distribution of glowworm swarm by introducing the theory and method of good-point set. Finally, the new clustering algorithm is applied to UCI data sets of different categories and numbers for clustering test. The advantages of the improved clustering algorithm in terms of sum of squared errors (SSE), clustering accuracy, and robustness are explained through comparison and analysis.

1. Introduction

As an unsupervised data analysis method, clustering analysis is widely applied in such fields as data mining, pattern recognition, machine learning, and artificial intelligence [1]. Different from classification, clustering algorithm realizes categorization by gathering data objects through certain similarity metric and clustering criterion without any prior knowledge. As a branch of statistics, clustering analysis has been studied extensively. Clustering method can be mainly classified into division method, hierarchy method, and density-based method. The K-means algorithm proposed by James Macqueen is a typical clustering algorithm based on division [2]. However, the clustering result of K-means algorithm is greatly affected by initial clustering center point and is very sensitive to outliers. Literature [3] optimizes the K-means algorithm by integrating the coding, crossing, and aberrance thoughts of genetic algorithm (GA) with the local optimizing ability of K-means clustering algorithm and proposes the K-means clustering algorithm based on GA. Hierarchy-based clustering methods mainly include CURE algorithm [4] and Chameleon algorithm [5], of which one cluster is represented by multiple points in CURE algorithm, making the processing of nonspherical data sets better. Representative algorithms of density-based clustering methods include DBSCAN algorithm [6], which is able to effectively identify class cluster of any shape, but is very sensitive to the setting of artificial parameters (e.g., radius). Rodriguez and Laio put forward a new density-based density peaks clustering (DPC) algorithm [7] in 2014. In this algorithm, density peaks (i.e., clustering centers) are selected manually through "decision diagram" first, and then, residual data points are allocated to each clustering center on this basis to obtain corresponding clustering result. It is noteworthy that, in recent years, some scholars have started introducing the heuristic swarm optimization algorithm into clustering analysis of different fields and improving clustering effect by virtue of the global searching ability of swarm optimization algorithm. A clustering analysis method combining PSO and K-means is proposed in literature [8] through the global

searching ability of particle swarm algorithm. In addition, Cuckoo algorithm, artificial bee colony algorithm, artificial fish swarm algorithm, and so forth [9–12] are also started to be introduced in the research of clustering algorithm.

The GSO algorithm [13] proposed by Krishnanand and Ghose is a new swarm intelligence optimization algorithm, which is more efficient in solving multimodal problems compared with traditional swarm intelligence optimization algorithms [14]. Aljarah and Ludwig put forward a new clustering based GSO algorithm in 2013. In this algorithm, the GSO algorithm is adjusted to solve the data clustering problem to locate multiple optimal centroids [15]. An new approach for cluster analysis based on GSO algorithm and K-means has been proposed by Onan and Korukoglu [16]. Due to the multimodal nature of multimedia data, Pushpalatha and Ananthanarayana proposed the GSO algorithm based Multimedia Document Clustering (GSOMDC) algorithm to group the multimedia documents into topics in 2015 [17]. A fuzzy clustering algorithm based on GSO algorithm (GSO-KFCM) is proposed by Cheng and Bao in 2017. In this algorithm, the GSO algorithm obtains the optimal solution as the initial clustering center of the kernelized fuzzy mean clustering algorithm [18].

This paper introduces GSO algorithm into clustering analysis, regards each glowworm as a feasible solution in clustering center of data object space, searches data object space through the optimization process of glowworm, and solves clustering center by obtaining multiple extreme points. In this way, it combines GSO algorithm and K-means algorithm together, provides initial clustering centers for K-means algorithm by means of GSO algorithm, solves the problem that K-means algorithm is sensitive to initial clustering centers, and thus obtains better clustering effects. Meanwhile, considering the effect of the initial distribution of glowworm swarm on clustering center search, this paper optimizes the initial glowworm swarm distribution in GSO algorithm by introducing the theory and method of good-point set [19, 20], which improves the global searching performance of GSO algorithm. The research in this paper mainly includes 3 parts. Section 2 gives explanations on relevant algorithms and theories, which puts forward the optimization idea for clustering analysis-oriented GSO algorithm. Section 3 introduces improved GSO algorithm based on good-point set, combines improved GSO algorithm with K-means algorithm together, and designs the algorithm framework and implementation steps for new clustering method (GSOK_GP algorithm). Section 4 selects UCI data sets of different categories and numbers for clustering experiment and analysis for the GSOK_GP algorithm designed in this paper.

2. Description of Relevant Algorithms

2.1. K-Means Clustering Algorithm

2.1.1. Basic Ideas of K-Means Clustering Algorithm. Basic ideas of K-means clustering algorithm: select k data points at random in the data objects to be clustered to act as initial clustering center points, and allocate other data points to corresponding clustering center points based on their similarity with such initial clustering center points. After one round of allocation, recalculate the clustering centers of each category based on the clustering result of the round, and then, allocate residual data points to obtain the clustering result of the new round. Repeat this process for given times or until the convergence of data center points.

2.1.2. Steps of K-Means Clustering Algorithm

(1) Problem Description. $X = \{x_1, x_2, \ldots, x_N\}$ represents a given data object, where x_i represents data vector point. Divide X into several disjoint clusters $C = \{C_1, C_2, \ldots, C_k\}$, where $X = \bigcup_{i=1}^{k} C_i$, $C_i \cap C_j = \emptyset$, $C_i \neq \emptyset$.

(2) Related Definitions

Definition 1. Euclidean distance between data points

$$\operatorname{dis}\left(x_i, x_j\right) = \sqrt[2]{\left(x_i - x_j\right)\left(x_i - x_j\right)^T}. \tag{1}$$

Definition 2. SSE of clustering results

$$\mathrm{SSE} = \sum_{i=1}^{k} \sum_{x_i \in z_j} \operatorname{dis}\left(x_i - z_j\right), \tag{2}$$

where z_j is the cluster center of x_i. SSE is taken as an important indicator for evaluating clustering result in general.

(3) Implementation Steps of K-Means Algorithm

Step 1. Randomly select k samples $z(1) = \{z_1^1, z_2^1, \ldots, z_k^1, \}$ as initial clustering centers.

Step 2. Allocate other data points in data object X to existing clustering center z_i^1 as per given principles (e.g., shortest Euclidean distance).

Step 3. Recalculate clustering center $z(2) = \{z_1^2, z_2^2, \ldots, z_k^2\}$ and $z_i^2 = (1/m)\sum_{j=1}^{m} x_j$, as per clustering result, where x_j is the data point allocated to clustering center point z_i^1.

Step 4. If $z_i^1 \neq z_i^2$, that is, the new clustering center is different from the original one, turn to Step 2 for iteration again, until the convergence of clustering center points or reaching maximum iterations.

It can be learnt from the steps above that initial clustering centers have significant effect on the clustering result and operating efficiency of K-means clustering algorithm and may lead to premature local optimum of K-means clustering algorithm, which causes clustering results with large difference in turn.

2.2. Main Ideas and Steps of GSO Algorithm. In GSO algorithm, each glowworm is deemed as a feasible solution of target problem in space. Glowworms gather towards high brightness glowworm through mutual attraction and location movement, and multiple extreme points are found out in the

solution space of a target problem. In this way, the problem is solved. Its main ideas can be described as follows.

Step 1. Initialize glowworm swarm $Z = \{z_1, z_2, \ldots, z_n\}$. Glowworm number n in swarm, step s, fluorescein initial value l_0, fluorescein volatilization rate ρ, domain change rate β, decision domain initial value γ_0, domain threshold $\gamma_{\max}$, and other parameters related need to be initialized and assigned in the initialization.

Step 2. Calculate glowworm fitness based on objective function. Calculate the fitness $f(z_i)$ of each glowworm z_i at its location based on specific objective function $y = \max(f(z))$.

Step 3. Calculate the moving direction and step of glowworm. Each glowworm z_i searches for glowworms z_j with higher fluorescein value l_i within its own decision radius r_i, and determine the next moving direction and step based on fluorescein value and distance.

Step 4. Update glowworm locations. Update the location of each glowworm z_i based on determined moving direction and step.

Step 5. Update the decision domain radius of glowworm.

Step 6. Judge whether the algorithm has converged or reached the maximum iterations (itmax) and determine whether to enter the next round of iteration.

It can be learnt from the steps above that algorithm execution efficiency can be improved and premature local optimum of algorithm can be avoided by optimizing the initial distribution of glowworm swarm.

2.3. Basic Theory of Good-Point Set. Basic definition and structure of good-point set are as follows:

(1) Assume G_s is a unit cube in S-dimensional Euclidean space, which is expressed as

$$x \in G_s,$$
$$x = (x_1, x_2, \ldots, x_s), \quad \text{where, } 0 \le x_i \le 1,\ i = 1, 2, \ldots, s. \tag{3}$$

(2) Assume $P_n(k)$ is a point set with the number of n in G_s, which is expressed as

$$P_n(k) = \left\{\left(x_1^{(n)}(k), x_2^{(n)}(k), \ldots, x_s^{(n)}(k)\right)\right\}, \quad \text{where, } 0 \le k \le n,\ 0 \le x_i^{(n)}(k) \le 1,\ i = 1, 2, \ldots, s. \tag{4}$$

(3) Assume $r = (r_1, r_2, \ldots, r_s)$ is a given point in G_s and $N_n(r) = N_n(r_1, r_2, \ldots, r_s)$ is the number of points not satisfying the inequality below in point set $P_n(k)$.

$$0 \le x_i^{(n)}(k) \le r_i, \quad \text{where, } i = 1, 2, \ldots, s. \tag{5}$$

$\varphi(n) = \sup|N_n(r)/n - |r||$, where $r \in G_s, |r| = r_1 r_2 \cdots r_s$, and $\varphi(n)$ is known as the deviation of point set $P_n(k)$.

(4) Assume $\varphi(n)$ is the deviation of $P_n(k) = \{(x_1^{(n)} * k, x_2^{(n)} * k, \ldots, x_s^{(n)} * k),\ k = 1, 2, \ldots, n\}$ and meets the requirements below:

$\varphi(n) = C(r, \varepsilon) n^{-1+\varepsilon}$, where $C(r, \varepsilon)$ is a constant related to r and ε, $\varepsilon > 0$.

$P_n(k)$ is known as a good-point set and r a good point.

It has been proved by applicable theorems that, with respect to approximate integration, the order of deviation $\varphi(n)$ is only relevant to n and irrelevant to the space dimensions of the sample. Therefore, good-point set can provide better support for the calculation in high-dimensional spaces [20]. Meanwhile, as for a point set object whose distribution is unknown, the deviation $\varphi(n)$ of n points $P_n = \{x_1, x_2, \ldots, x_n\}$ obtained by virtue of good-point set is significantly superior to n points obtained by random method. Therefore, a better initial distribution scheme can be provided for the swarm distribution in swarm intelligence algorithm based on this feature of good-point set.

3. Design of GSOK_GP Algorithm

This paper proposes an improved GSO algorithm based on good-point set to solve clustering problems on the basis of analysis of relevant algorithms above and characteristics of clustering problems. Its main ideas can be described as firstly, optimize the initial distribution of glowworm swarm through good-point set, so as to optimize GSO algorithm. Secondly, optimize the initial clustering centers in clustering data objects, and obtain characteristics of multiple extreme points and a clustering center point set with optimized GSO algorithm. Thirdly, select k extreme points as the initial clustering center of K-means algorithm in the clustering center point set as per maximum distance principle. Fourthly, execute the K-means algorithm with initial clustering center to figure out the clustering result. The algorithm framework is shown as Figure 1. Where $t \le$ itmax means the iterations are no greater than maximum iterations, flag $> k$ means the number of extreme points is greater than the number k of initial clustering centers required.

3.1. Initial Swarm Optimization Based on Good-Point Set. Optimization of initial distribution of glowworm swarm is to represent the characteristics of solution space more scientifically utilizing glowworm swarm in essence. Randomly generated glowworm swarm cannot cover all conditions of solution space in most cases. Therefore, uniform distribution of glowworm swarm in solution space is an effective strategy. More uniform distribution of swarm can be realized with the theory and method of good-point set above.

Assume the initial glowworm swarm number is n; select n points in s-dimensional space to act as glowworm locations. Select the good-point set $P_n(k) = \{(x_1^{(n)} * k, x_2^{(n)} * k, \ldots, x_s^{(n)} * k),\ k = 1, 2, \ldots, n\}$ composed of n good points in s-dimensional space with good-point set theory. There are mainly three methods:

(1) Square root sequence method: $r_k = \{\sqrt{p_k},\ 1 \le k \le s\}$, where p_k are different primes.

(2) Cyclotomic field method: $r_k = \{2\cos(2\pi k/p),\ 1 \le k \le s\}$, where p is the smallest prime meeting $(p-3)/2 \ge s$.

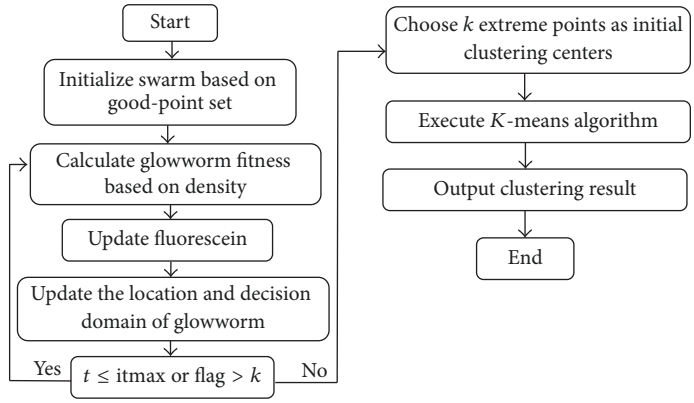

Figure 1: Flow of GSOK_GP algorithm.

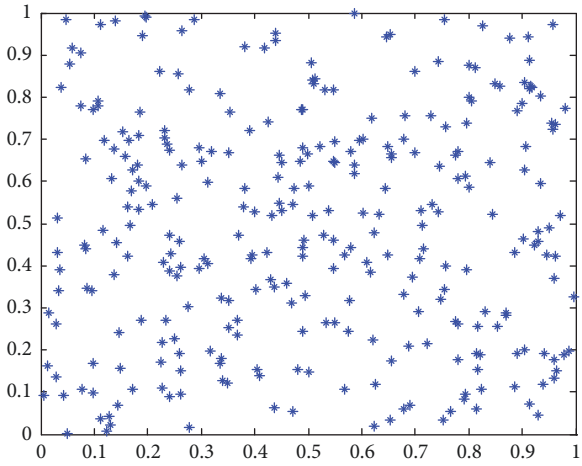

Figure 2: Randomly distributed n data points (glowworms).

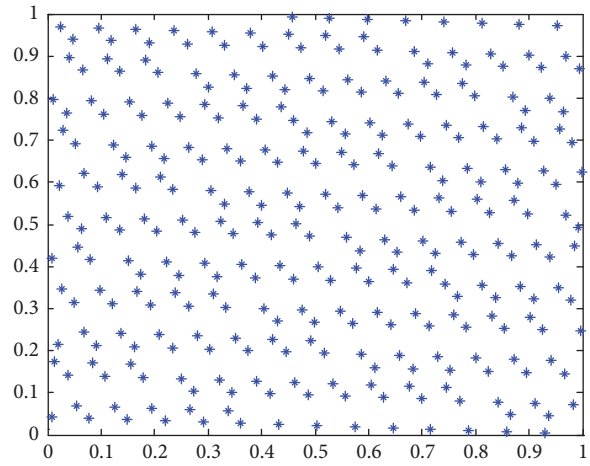

Figure 3: n data points (glowworms) distributed in good-point set exponential sequence method.

(3) Exponential sequence method: $r_k = \{e^k,\ 1 \leq k \leq s\}$.

Assuming $s = 2$ and $n = 300$, construct good-point set (i.e., initial glowworm swarm distribution) with exponential sequence method. Figures 1 and 2 show the data points (glowworms) distribution under random condition and when applying exponential sequence method, respectively.

The comparison between Figures 2 and 3 indicates that the data point distribution in exponential sequence method is more uniform, which is able to cover the solution space better. In the meantime, the structure of its good-point set is more stable; that is, the distribution effect is consistent when n is unchanged. Therefore, a better initial glowworm distribution can be obtained by applying good-point set in initial glowworm swarm distribution.

3.2. Flow of GSOK_GP Algorithm. Glowworm individuals are deemed as the feasible solutions of a clustering center point when combining improved glowworm algorithm with K-means algorithm to solve clustering problems. In view of the characteristic that clustering center points are surrounded by data points of data objects, the density of clustering center points is represented by an extreme value of various data point densities within one domain. Therefore, take the density of glowworm individuals in data object set as their fitness, and obtain a superior initial clustering center point set through optimizing of density extreme value by glowworms. The main algorithm flow is as follows.

Step 1. Initialize with the glowworm swarm based on good-point set. As for the data set $X = \{x_1, x_2, \ldots, x_N\}$ needing to be clustered, initialize and assign glowworm number n in swarm, initial location of glowworm, step s, fluorescein initial value l_0, fluorescein volatilization rate ρ, domain change rate β, decision domain initial value γ_0, domain threshold $\gamma_{\max}$, and other parameters related in the Euclidean space where X is limited.

$$\begin{aligned} X &= \{x_1, x_2, \ldots, x_N\}, \\ x_i &= (x_{i1}, x_{i2}, \ldots, x_{is},) \\ Z &= \{z_1, z_2, \ldots, z_n\}, \\ z_i &= (z_{i1}, z_{i2}, \ldots, z_{is},) \\ z_{ij} &\in [\min(x_i), \max(x_i)]. \end{aligned} \tag{6}$$

Step 2. Calculate glowworm fitness, namely, the number $f(z_i)$ of data points in data set $X = \{x_1, x_2, \ldots, x_N\}$ in the domain where glowworm z_i distance is γ_i.

$$\begin{aligned} \text{Dset } X(z_i) &= \left\{x_j \mid \text{dis}\left(x_j, z_i\right) < \gamma_i\right\}. \\ f(z_i) &= \text{count}\left(\text{Dset } (z_i)\right) \end{aligned} \tag{7}$$

Step 3. Update glowworm fluorescein. $l_i(t)$ represents the fluorescein value of glowworm t in round z_i of iteration.

$$l_i(t+1) = (1-\rho) l_i(t) + \lambda f\left(x_i(t)\right). \tag{8}$$

Step 4. Determine moving direction. Glowworm z_i searches the glowworm with higher fluorescein in decision domain and selects the glowworm z_j with higher fluorescein through roulette approach, which acts as the moving direction of the next step. Dset $Z(z_i)$ represents the glowworm set in the domain, Dset $L(z_i)$ represents the glowworm set with higher fluorescein in the domain, and $P(\text{Dset } L(z_i))$ represents the probability of each glowworm to be selected. Choose the glowworm z_j with the maximum probability to act as the moving direction of glowworm z_i.

$$\begin{aligned} \text{Dset } Z(z_i) &= \left\{z_j \mid \text{dis}\left(z_j, z_i\right) < \gamma_i,\ \ j \neq i\right\} \\ \text{Dset } L(z_i) &= \left\{z_j \mid \text{dis}\left(z_j, z_i\right) < \gamma_i, l_j > l_i,\ \ j \neq i,\right\} \\ P\left(\text{Dset } L(z_i)\right) &= \{p_1, p_2, \ldots, p_m\} \\ P\left(z_j\right) &= \max\left(p_1, p_2, \ldots, p_m\right). \end{aligned} \tag{9}$$

Step 5. Update location. Glowworm z_i moves by the step s towards the direction of glowworm z_j to complete location update of all glowworms.

$$z_i = z_i + \frac{z_j - z_i}{\left\|z_j - z_i\right\|} \times s. \tag{10}$$

Step 6. Update decision domain. $\gamma_i(t)$ represents the decision radius of glowworm z_i in round t iteration, n_S represents the threshold of glowworm number in the domain, and n_t represents the glowworm number within the decision radius.

$$\gamma_i(t+1) = \gamma_i(t) + \beta\left(n_S - n_t\right). \tag{11}$$

Step 7. Judge the termination condition of glowworm search and enter iteration of the next round.

Step 8. Glowworm algorithm terminates, and k extreme points are output to act as the initial clustering center points for K-means algorithm.

Step 9. Execute K-means algorithm and output clustering result.

3.3. Key Strategies in GSOK_GP Algorithm

3.3.1. Density-Based Fitness Function. Cluster center is a glowworm data point surrounded by adjacent points of low local density in GSOK_GP algorithm; therefore, cluster center can be interpreted as a local optimal point on fitness.

$$\begin{aligned} f(z_i) &= \text{count}\left(\text{Dset}\left(z_i\right)\right) \\ \text{count}\left(\text{Dset}\left(z_i\right)\right) &= \sum_{z_j \in \text{Dset}(z_i)} d_{ij} \\ d_{ij} &= \begin{cases} 1, & \text{dis}\left(x_j, z_i\right) < \gamma_i \\ 0, & \text{dis}\left(x_j, z_i\right) \geq \gamma_i. \end{cases} \end{aligned} \tag{12}$$

3.3.2. Weighted Euclidean Distance. Since there is large difference in value range of the data object in different dimensions, partial attributes with a large value range may have greater effect on the Euclidean distance between data points if only Euclidean distance is applied, which will cause greater effect on the clustering result. Therefore, calculation of Euclidean distance needs to be adjusted through different weights allocation in the process of initial clustering center search by the glowworm if assuming each dimension of the data object has the same effect on the clustering result without prior knowledge.

Assumption 3. Value range of data object X in each dimension is expressed as follows:

$$U = \left\{[a_1, b_1], [a_1, b_1], \ldots, [a_n, b_n]\right\}. \tag{13}$$

Set $d_i = b_i - a_i$, $D = (d_1, d_2, \ldots, d_n,)$.

$W = (w_1, w_2, \ldots, w_n,)$ represents the weight to be assigned to different dimensions:

$$w_i = \frac{1/d_i}{\sum 1/d_i}. \tag{14}$$

Improved Euclidean distance calculation method is redefined in this way.

$$\text{dis}\left(x_i, x_j\right) = \sqrt[2]{\left(x_i - x_j\right)\left(x_i - x_j\right) W^T}. \tag{15}$$

It should be noted that adjustment for Euclidean distance calculation method is only applied in the process of searching initial clustering center in GSO algorithm, and general Euclidean distance calculation approach needs to be adopted in algorithm evaluation, so as to compare and analyze with other algorithms.

3.3.3. Selection of Extreme Point. A relatively large distance between cluster centers is necessary in clustering algorithm. Therefore, select k centers in multiple cluster centers to constitute the initial clustering centers of K-means algorithm; that is, selecting k extreme points in extreme point set $J = \{J_1, J_2, \ldots, J_p\}$ to act as the initial clustering centers of K-means algorithm needs to follow distance maximization principle. When $p > k$, the basic steps for selecting extreme points are as follows:

(1) Firstly, select the glowworm with the highest fitness to act as the first clustering center point.

Table 1: Selection of experimental data set.

Data set	Number of dimensions	Number of categories	Number of samples
Iris	4	3	150
Glass	9	6	214

(2) Secondly, calculate the distances from other clustering center points to the first clustering center point, and select the one with the largest distance to act as the second clustering center point.

(3) Repeat step (2) to calculate the sum of the distances from other clustering center points to clustering centers selected, and select the one with the largest distance to act as the next clustering center point until k clustering center points are obtained.

4. Experiment and Analysis

4.1. Experimental Environment. Matlab is employed to compile GSOK_GP algorithm and two UCI data sets shown in Table 1 are selected to test its effectiveness in this paper. Design parameters of GSO algorithm referring to relevant literatures, and select relevant parameters of M-GSO algorithm as follows based on actual clustering problems: $N = 50$, $\rho = 0.4$, $\lambda = 0.6$, $\beta = 0.08$, $s = 1$, $l_0 = 5$, and $n_s = 6$, with maximum iterations: 100.

SSE, clustering accuracy, and robustness are used to evaluate clustering effect of algorithm in this paper. SSE employs the sum of the Euclidean distances from all data objects to their cluster center points. The calculation approach is as follows:

SSE $= \sum_{j=1}^{k} \sum_{x_i \in z_j} \sqrt[2]{(x_i - z_j)(x_i - z_j)^T}$, where z_j is the cluster center point of x_i.

The clustering accuracy proposed by Gan et al. is taken as one of the clustering effect evaluation standards in this paper [21]. Clustering accuracy refers to the proportion of accurately classified samples to total samples. The definition of clustering accuracy AC is as follows:

$$\mathrm{AC} = \frac{\sum_{i=1}^{k} a_i}{n}, \tag{16}$$

where k represents the number of categories of data sets, n represents the total number of samples in the data set, a_i represents the number of samples accurately classified into Category i.

In addition, the robustness indicators proposed in literature [22] are used to identify the algorithm stability in this paper. The algorithm robustness in this paper is calculated with the mean square error of results of multiple experiments as per the calculation formula below:

$$R = \frac{\mathrm{AC}^* - \mathrm{AC}'}{\mathrm{AC}^*} \times 100\%, \tag{17}$$

where AC^* is the optimal value of clustering accuracy and AC' is the average value of clustering accuracy obtained by operating the algorithm multiple times. The smaller the R is, the higher the algorithm robustness will be.

Table 2: Experimental results of iris data set.

Algorithm	Average value of SSEs	Average value of ACs (%)	Standard deviation value of ACs
K-means	102.57	83.95	0.0451
PSOK	99.61	87.39	0.0420
GSOK_GP	97.32	89.33	0

Table 3: Experimental results of glass data set.

Algorithm	Average value of SSEs	Average value of ACs (%)	Standard deviation value of ACs
K-means	241.03	51.70	0.0157
PSOK	233.23	52.20	0.0108
GSOK_GP	225.08	53.50	0.0090

Table 4: Comparison of robustness (R) of each algorithm (%).

Algorithm	Iris	Glass
K-means	10.7	5.73
PSOK	8.07	3.65
GSOK_GP	0	2.97

4.2. Experimental Results and Analysis. The data of executing GSOK_GP algorithm 20 times for Iris and Glass data sets, respectively, and independently is shown in Tables 2, 3, and 4. The data of executing K-means algorithm and PSOK algorithm 20 times is cited from literature [9].

There are 150 sample objects in Iris data set, each of which has 4 attributes, which can be classified into 3 categories in total. The experimental results of Iris data set are shown in Table 2.

There are 214 data sets in Glass data set; each object has 9 attributes, which can be classified into 6 categories in total. The experimental results of Glass data set are shown in Table 3.

It can be learnt from Tables 2 and 3 that GSOK_GP algorithm is superior to traditional k-means algorithm and PSOK algorithm on SSE and average accuracy.

Calculation results based on comparing the robustness of traditional k-means algorithm, PSOK algorithm, and GSOK_GP algorithm are shown in Table 4.

Table 4 indicates that the operation results of 20 independent operations of GSOK_GP algorithm for Iris data set are consistent, which proves significant stability. And the fluctuation in the operation results of 20 independent operations for Glass data set is obviously smaller than that of k-means algorithm and PSOK algorithm. Therefore, GSOK_GP algorithm has better robustness in the experiments.

5. Conclusion

Traditional K-means clustering algorithm is widely used due to its simple principle and high execution efficiency. However, K-means algorithm relies on initial clustering centers, which leads to large difference in the clustering result, low accuracy, and lack of stability of traditional K-means algorithm. In this

paper, the initial clustering centers in K-means algorithm are optimized with improved glowworm algorithm based on good-point set, and the clustering effect is improved.

The GSOK_GP algorithm proposed in this paper is mainly applied to solving data object clustering problems under unsupervised learning conditions. The difference between the GSOK_GP algorithm and traditional clustering methods is that it combines GSO algorithm and K-means algorithm together to improve the clustering effect. In particular, as for the effect of initial clustering centers on clustering results, this paper provides more scientific descriptions for data object space by introducing the theory and method of good-point set and obtains superior initial clustering center points with the searching ability of GSO algorithm. Through comparison and analysis, GSOK_GP algorithm is proved to have better clustering effect and stability.

In addition, the adverse effect of computing efficiency of GSOK_GP algorithm for glowworm density in case of large data object has also been noticed, which means that the convergence of GSOK_GP algorithm needs to be improved further, so as to apply it better when addressing clustering problems under large data volume.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The work was supported by National Natural Science Foundation of China (nos. 91546108, 71271071, 71490725, and 71521001), fund of Provincial Excellent Young Talents of Colleges and Universities of Anhui Province (no. 2013SQRW115ZD), fund of Support Program for Young Talents of Colleges and Universities of Anhui Province, fund of Natural Science of Colleges and Universities of Anhui Province (no. KJ2016A162), fund of Social Science Planning Project of Anhui Province (no. AHSKYG2017D136), and fund of Scientific Research Team of Anhui Economic Management Institute (no. YJKT1417T01).

References

[1] J. Guang, M. Liu, and D. Zhang, "Spectral clustering algorithm based on effective distance," *Journal of Frontiers of Computer Science and Technology*, vol. 11, no. 11, pp. 1365–1372, 2014.

[2] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.

[3] B. Lu and F. Ju, "An optimized genetic K-means clustering algorithm," in *Proceedings of the 2012 International Conference on Computer Science and Information Processing, CSIP 2012*, pp. 1296–1299, China, August 2012.

[4] Y.-J. Zhou, C. Xu, and J.-G. Li, "Unsupervised anomaly detection method based on improved CURE clustering algorithm," *Tongxin Xuebao/Journal on Communication*, vol. 31, no. 7, pp. 18–32, 2010.

[5] H. Wilcox, R. C. Nichol, G.-B. Zhao, D. Bacon, K. Koyama, and A. K. Romer, "Simulation tests of galaxy cluster constraints on chameleon gravity," *Monthly Notices of the Royal Astronomical Society*, vol. 462, no. 1, Article ID stw1617, pp. 715–725, 2016.

[6] Y. Jing, G. Jiawei, L. Jiye et al., "An improved DBSCAN clustering algorithm based on data field," *Journal of Frontiers of Computer Science and Technology*, vol. 6, no. 10, pp. 903–911, 2012.

[7] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.

[8] Z. Pei, X. Hua, and J. Han, "The clustering algorithm based on particle swarm optimization algorithm," in *Proceedings of the International Conference on Intelligent Computation Technology and Automation, ICICTA 2008*, pp. 148–151, chn, October 2008.

[9] T. Hassanzadeh and M. R. Meybodi, "A new hybrid approach for data clustering using firefly algorithm and K-means," in *Proceedings of the 16th CSI International Symposium on Artificial Intelligence and Signal Processing, AISP 2012*, pp. 7–11, Iran, May 2012.

[10] Y. Huihua, W. Ke, L. Lingqiao, W. Wen, and H. Shengtao, "K-means clustering algorithm based on adaptive cuckoo search and its application," *Journal of Computer Applications*, vol. 36, no. 8, pp. 2066–2070, 2016.

[11] D. Yu-ting, W. Song, and M. Wei, "Artificial colony clustering algorithm based on global information," *Microelectronics & Computer*, vol. 34, no. 2, pp. 20–24, 2017.

[12] H.-t. Yu, M.-J. Jia, and H.-q. Wang, "Clustering algorithm based on artificial fish swarm," *Computer Science*, vol. 39, no. 12, pp. 60–64, 2012.

[13] K. N. Krishnanand and D. Ghose, "Glowworm swarm optimization for simultaneous capture of multiple local optima of multimodal functions," *Swarm Intelligence*, vol. 3, no. 2, pp. 87–124, 2009.

[14] N. Zainal, A. M. Zain, and N. H. M. Radzi, "Glowworm swarm optimization (GSO) for optimization of machining parameters," *Journal of Intelligent Manufacturing*, vol. 27, no. 4, pp. 797–804, 2016.

[15] I. Aljarah and S. A. Ludwig, "A new clustering approach based on glowworm swarm optimization," in *Proceedings of the 2013 IEEE Congress on Evolutionary Computation, CEC 2013*, pp. 2642–2649, Mexico, June 2013.

[16] A. Onan and S. Korukoglu, "Improving performance of glowworm swarm optimization algorithm for cluster analysis using K-means," in *Proceedings of the International Symposium on Computing in Science Engineering*, vol. 10, pp. 291–297, 2013.

[17] K. Pushpalatha and V. S. Ananthanarayana, "A New glowworm swarm optimization based clustering algorithm for multimedia documents," in *Proceedings of the 17th IEEE International Symposium on Multimedia, ISM 2015*, pp. 262–265, USA, December 2015.

[18] C. Cheng and C. Bao, "A kernelized fuzzy C-means clustering algorithm based on glowworm swarm optimization algorithm," in *Proceedings of the 9th International Conference on Computer and Automation Engineering, ICCAE 2017*, pp. 78–82, Australia, February 2017.

[19] L. G. Hua and Y. Wang, *Applications of Number Theory to Numerical Analysis*, Springer, Berlin, Germany, 1981.

[20] Y. Chen, X. Liang, and Y. Huang, "Improved quantum particle swarm optimization based on good-point set," *Zhongnan Daxue Xuebao (Ziran Kexue Ban)/Journal of Central South University (Science and Technology)*, vol. 44, no. 4, pp. 1409–1414, 2013.

[21] G. Gan, J. Wu, and Z. Yang, "A genetic fuzzy k-Modes algorithm for clustering categorical data," *Expert Systems with Applications*, vol. 36, no. 2, pp. 1615–1620, 2009.

[22] P. Xiaoying, C. Xuejing, L. Angru, and Z. Pu, "Firefly partition clustering algorithm based on self-adaptive step," *Computer Applied Research*, vol. 34, no. 12, pp. 3576–3579, 2017.