# K-means Clustering Based on Improved Gravitational Search Algorithm (Postprint)

**Authors:** Wei Kangyuan, He Qing, Xu Qinshuai

**Date:** 2018-08-13T00:00:00+00:00

## Abstract

To address the problem that K-means algorithm clustering results are highly susceptible to initial cluster centers and prone to falling into local optimal solutions, a K-means clustering algorithm based on improved gravitational search is proposed. Firstly, an adaptive concept is introduced to control the decay factor of the gravitational coefficient, thereby enhancing the algorithm' s global exploration and local exploitation capabilities. Secondly, an immune clone selection mechanism is incorporated to enable the algorithm to effectively escape local optima, and the effectiveness and superiority of the improved gravitational search algorithm are verified through experiments on twelve benchmark test functions. Finally, by combining the improved gravitational search algorithm with K-means, a novel clustering algorithm named A2F-GSA-Kmeans is proposed, and experiments on six test datasets demonstrate that the algorithm achieves satisfactory clustering quality.

## Full Text

## Preamble

### Novel K-means clustering algorithm based on improved gravitational search algorithm

*Wei Kangyuana,b, He Qinga,b†, Xu Qinshuaia,b*

aCollege of Big Data & Information Engineering, bGuizhou Provincial Key Laboratory of Public Big Data, Guizhou University, Guiyang 550025, China

**Abstract:** In order to solve the problem that the clustering result of K-means algorithm gets affected by the initial cluster centers easily, this paper proposed a novel K-means clustering algorithm based on improved gravitational search algorithm. Firstly, it enhanced the global exploration and local exploitation capability of the algorithm with the introduction of adaptive concept to control the attenuation factor of gravitational constant. Then, by introducing immune clonal selection algorithm to make the algorithm jump out of the local optimum efficiently. The experimental results on twelve test functions prove the effectiveness and superiority of the improved GSA. Finally, by combining the improved GSA with K-means algorithm, this paper proposed a new clustering algorithm called A2F-GSA-Kmeans. The experimental results on six test datasets show that the algorithm has better clustering quality.

**Key words:** K-means clustering algorithm; gravitational search algorithm; attenuation factor of gravitational constant; immune clonal selection algorithm

## 0 Introduction

Clustering is an important method for data analysis that divides multiple abstract objects into multiple classes composed of similar objects according to corresponding criteria. It has been widely applied in many fields. Among clustering algorithms, K-means has the advantages of simplicity and efficiency when processing large amounts of data and has been widely used. However, its clustering results are extremely susceptible to the influence of initial cluster centers, leading to 陷入局部最优解错误! 未找到引用源。, and it requires users to specify the number of clusters. Different cluster numbers will yield different clustering results, directly affecting algorithm efficiency. Therefore, it is crucial that the algorithm itself can obtain the optimal number of clusters.

Swarm intelligence algorithms have been increasingly applied to clustering problems to compensate for the deficiencies of traditional clustering algorithms due to their powerful global search capabilities. For example, Yang Juqing et al. 错误! 未找到引用源。optimized the position and velocity update methods of the Bat Algorithm (BA) while introducing nonlinear inertia weight and limit threshold concepts to improve convergence performance. They combined the improved algorithm with K-means to propose a K-means algorithm based on improved BA, achieving good clustering results. Yu Zuojun et al. 错误! 未找到引用源。improved the search patterns of employed bees and onlooker bees in the Artificial Bee Colony algorithm by introducing arithmetic crossover operations, and combined it with K-means to propose a clustering algorithm that automatically finds the optimal number of clusters.

The Gravitational Search Algorithm (GSA) 错误! 未找到引用源。is a novel swarm intelligence optimization algorithm proposed by Professor Esmat Rashedi et al. in 2009. The algorithm searches for global optimal solutions by simulating universal gravitation in physics. Research has shown that when optimizing benchmark test functions, the classical GSA algorithm' s optimization accuracy and convergence speed are significantly superior to Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) 错误! 未找到引用源。. However, similar to other metaheuristic algorithms, the classical GSA algorithm also suffers from premature convergence and easy 陷入局部极值等缺陷. Based on this, scholars have implemented many effective improved GSA algorithms in recent years. For instance, Liu et al. 错误! 未找到引用源。utilized chaotic mapping to optimize particle position initialization in GSA and introduced an adaptive decreasing inertia weight coefficient into the position update formula, proposing the AC-GSA algorithm, which achieved good results in both classical benchmark function tests and optimizing least squares support vector machine hyperparameters. Sun et al. 错误! 未找到引用源。improved GSA' s Kbest and velocity update methods based on particle heterogeneity, using individual and global optimal values to propose the LIGSA algorithm. This allowed particles to learn from K nearest neighbors to fully explore the search space and effectively prevent premature convergence, while global optimal value guidance accelerated convergence. Mirjalili et al. 错误! 未找到引用源。improved the gravitational coefficient G using chaotic mapping based on its importance for balancing global exploration and local exploitation capabilities, verifying its effectiveness in jumping out of local optima to achieve higher optimization precision. However, although existing research has improved the optimization effect of classical GSA, mostly focusing on combining PSO to improve GSA' s velocity and position update methods 错误! 未找到引用源。8~ 错误! 未找到引用源。, effective balancing of exploration and exploitation capabilities and solving premature convergence problems still require in-depth research.

In summary, this paper addresses the problems of GSA easily 陷入局部最优 and premature convergence. Firstly, the gravitational coefficient in GSA is improved to enhance global exploration and local exploitation capabilities while maintaining high precision and convergence speed. Then, an immune clonal selection mechanism is introduced to improve premature convergence. Finally, the improved GSA algorithm is applied to K-means clustering to optimize cluster centers and obtain the best number of clusters by adjusting K values using clustering evaluation functions.

## 1.1 K-means Algorithm

K-means is a partition-based clustering algorithm that first randomly selects K data points from the sample space as cluster centers, then calculates the Euclidean distance between other data points and these centers to partition the data. The algorithm flow is shown in Figure 1 [Figure 1: see original paper].

For Euclidean distance data, this paper uses Mean Square Error (MSE) as the clustering objective function, where smaller MSE values indicate better clustering effects. It is defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{k}\sum_{y\in C_i}\|y - z_i\|^2$$

where $z_i$ represents cluster centers.

This paper uses the silhouette coefficient 错误! 未找到引用源。to evaluate clustering quality under different numbers of clusters to find the optimal cluster number. The silhouette coefficient for each cluster is expressed as:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

where: $a(i)$ represents the average distance between sample $i$ and other samples in the same cluster; $b(i)$ represents the minimum average distance between sample $i$ and all samples in other clusters.

For the entire dataset, the average silhouette index evaluates clustering effectiveness as follows:

$$sil = \frac{1}{N} \sum_{i=1}^{N} s(i)$$

where $N$ represents the sample size of the dataset. Moreover, $-1 \leq sil \leq 1$. If $sil$ is close to 1, the clustering quality is good.

## 1.2 Classical GSA Algorithm

The Gravitational Search Algorithm (GSA) treats all particles in space as mass objects moving without resistance according to Newton' s second law, where objects with larger mass occupy better positions. Through mutual gravitational forces between objects, the optimal solution is found. The classical GSA algorithm flow is shown in Figure 2 [Figure 2: see original paper].

The classical GSA algorithm is described as follows: Assume a population composed of $N$ particles $X_i = (x_i^1, x_i^2, \cdots, x_i^D), i = 1, 2, \cdots, N$ in a $D$-dimensional search space, where $x_i^d$ represents the position of particle $i$ in dimension $d$. During the $t$-th iteration, the inertial mass $M_i(t)$ of particle $i$ is updated based on its fitness value:

$$M_i(t) = \frac{m_i(t)}{\sum_{j=1}^{N} m_j(t)}$$

$$m_i(t) = \begin{cases} \frac{f_i(t) - \text{worst}(t)}{\text{best}(t) - \text{worst}(t)} & \text{if best}(t) \neq \text{worst}(t) \\ 1 & \text{otherwise} \end{cases}$$

where $f_i(t)$ represents the fitness value of particle $i$ at iteration $t$. For minimization problems, the best and worst fitness values are defined as:

$$\text{best}(t) = \min_{j \in \{1, 2, \cdots, N\}} f_j(t)$$

$$\text{worst}(t) = \max_{j \in \{1,2,\cdots,N\}} f_j(t)$$

Conversely, they can be used for maximization problems.

When performing the $t$-th iteration, the mutual force between particle $i$ and particle $j$ in dimension $k$ is defined as:

$$F_{ij}^k(t) = G(t) \times \frac{M_i(t) \times M_j(t)}{R_{ij}(t) + \varepsilon} \times (x_j^k(t) - x_i^k(t))$$

where: $M_i(t)$ represents the inertial mass of acting particle $i$; $M_j(t)$ represents the inertial mass of acted particle $j$; $\varepsilon$ is a constant; $G(t)$ represents the gravitational constant at iteration $t$; $R_{ij}(t)$ represents the distance between particles $i$ and $j$ (generally Euclidean distance), calculated as:

$$R_{ij}(t) = \|X_i(t), X_j(t)\|_2$$

In GSA, the gravitational constant $G(t)$ is defined as:

$$G(t) = G_0 \times e^{-\alpha \times \frac{t}{T}}$$

where: $G_0$ represents the universal gravitational constant at the initial moment; $\alpha$ represents the attenuation factor of the gravitational constant, generally a constant; $T$ represents the maximum number of iterations.

In GSA, let the total force acting on particle $i$ in dimension $k$ at iteration $t$ be:

$$F_i^k(t) = \sum_{j \in \text{Kbest}, j \neq i} \text{rand}_j \times F_{ij}^k(t)$$

where $\text{rand}_j$ is a random number in $[0, 1]$; $F_{ij}^k(t)$ is defined as in equation (6); Kbest is the set of particles that exert gravitational force on others, initially containing $N$ particles and gradually decreasing to 1, defined as:

$$\text{Kbest} = \text{final\_per} + (100 - \text{final\_per}) \times \left(1 - \frac{t}{T}\right)$$

where final_per represents the percentage of particles that exert force on others.

According to Newton' s second law, the acceleration of particle $i$ in dimension $k$ at iteration $t$ is defined as:

$$a_i^k(t) = \frac{F_i^k(t)}{M_i(t)}$$

In each iteration of GSA, particles update their velocity $v_i^k$ and position $x_i^k$ as follows:

$$v_i^k(t+1) = \text{rand}_i \times v_i^k(t) + a_i^k(t)$$

$$x_i^k(t+1) = x_i^k(t) + v_i^k(t+1)$$

where $\text{rand}_i$ is a random number in $[0, 1]$.

## 2 Improved Gravitational Search Algorithm

To address the problems of classical GSA easily 陷入局部最优解 and premature convergence, this paper proposes an improved gravitational search algorithm based on adaptive attenuation factor of gravitational coefficient and immune clonal selection mechanism, called A2F-GSA (adaptive attenuation factor based gravitational search algorithm).

### 2.1 Adaptive Nonlinear Attenuation of Gravitational Coefficient

In GSA, the gravitational constant $G$ is crucial for finding optimal solutions, as shown in equation (7), where the main parameters are $G_0$ and $\alpha$, typically constant values. Research shows that when parameter $G_0$ is set to 100, the algorithm achieves optimal performance 错误! 未找到引用源。. For parameter $\alpha$, adjusting its value reveals that: in the early iterations, a smaller $\alpha$ value ensures larger step sizes for particles, improving global exploration capability; in the middle and later iterations, a larger $\alpha$ value accelerates convergence and enhances local exploitation capability 错误! 未找到引用源。.

By studying the characteristics of exponential functions and considering the impact of parameter $\alpha$ on algorithm performance, this paper proposes an attenuation factor that adaptively varies based on iteration number, defined as:

$$\alpha(t) = \gamma \times e^{\eta \times \frac{t}{T}}$$

where: $t$ represents the current iteration number; $T$ represents the maximum iteration number; $\gamma$ and $\eta$ are function parameters. By selecting appropriate parameters to control the variation range of the $\alpha$ function, we set $\gamma = 100$ and $\eta = 0.1$. The adaptive $\alpha$ generates larger gravitational coefficient $G$ in early iterations to more effectively enhance global exploration capability, and smaller $G$ in later iterations to effectively improve local exploitation capability.

### 2.2 Introduction of Immune Clonal Selection Mechanism

Analysis of classical GSA shows that as the algorithm iterates, particles gradually gather around those with better fitness values, causing particle distribution to shrink and diversity to decrease, making the algorithm prone to 陷入局部最优. Inspired by the DQABCI algorithm proposed in 错误! 未找到引用源。that combines artificial bee colony and Clonal Selection Algorithm (CSA), and the MAPCPSOI algorithm in 错误! 未找到引用源。based on PSO and CSA, this paper introduces an immune clonal selection mechanism into classical GSA. By combining GSA' s optimization capability with CSA' s characteristics of selection, replication, mutation, and reselection, the algorithm gains the ability to jump out of local optima and improves premature convergence.

The Clonal Selection Algorithm (CSA) 错误! 未找到引用源。, proposed by De Castro et al. in 2000, is an optimization algorithm based on the micro-evolutionary process within artificial immune systems. It achieves a balance between global exploration and local exploitation by constructing memory units 错误! 未找到引用源。, evolving from a single optimal individual to a population optimal solution set, expanding the search area while demonstrating immune system diversity 错误! 未找到引用源。.

CSA mainly includes three stages: clonal replication, clonal mutation, and clonal selection. The clonal replication stage expands population size and search space; the clonal mutation stage increases population diversity to construct a new population; the clonal selection stage selects antibodies with high fitness (affinity) from the new population for the next generation, compressing the population while moving it toward better solutions.

The specific steps of the immune clonal selection mechanism introduced in this paper are as follows:

**(1) Particle Affinity Calculation**

In the improved GSA algorithm, particles are treated as antibodies. Therefore, the fitness function of particles is defined as the affinity function:

$$a_i(t) = \frac{f_i(t)}{\sum_{i=1}^{N} f_i(t)}$$

where $f_i(t)$ represents the fitness value of particle $i$ at iteration $t$.

**(2) Clonal Replication**

Select the particle with the highest affinity value in the current population for position information cloning replication. The number of copies is $M$, where we select $M = 10$.

**(3) Clonal Mutation**

Perform position mutation on the newly generated particles from clonal replication. This paper introduces mutation operators based on Gaussian and Cauchy distributions, whose probability density functions are described as:

$$C(x) = \frac{1}{\pi} \times \frac{\gamma}{\gamma^2 + (x - x_0)^2}$$

$$N(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Analysis shows that in the early iterations, the algorithm should maintain larger mutation step sizes to expand the global search range; in later iterations, as the population gradually converges to the global optimum, mutation step sizes should gradually decrease to facilitate algorithm convergence. The Cauchy distribution-based mutation operator has a larger mutation scale compared to Gaussian mutation, making it suitable for early evolution; the Gaussian distribution-based mutation operator is suitable for subsequent iterations, enabling faster convergence. Therefore, inspired by 错误! 未找到引用源。' s hybrid mutation approach for GSA, this paper proposes an improved mutation strategy that adaptively adjusts with iteration number:

$$x' = x + \lambda \times \left[\left(1 - \frac{t}{T}\right) \times C(0,1) + \frac{t}{T} \times N(0,1)\right]$$

where: $x$ represents the position information of the cloned particle; $x'$ represents the new position after mutation; $\lambda$ is a regulation coefficient, set to $\lambda = 0.1$ in this paper; $t$ represents the current iteration number; $T$ represents the maximum iteration number; $N(0,1)$ and $C(0,1)$ represent random numbers following Gaussian and Cauchy distributions, respectively.

**(4) Clonal Selection**

The mutated particles and original particles form a new population denoted as $P'$. For the new population $P'$, calculate the affinity values based on particle position information, select the best individuals to enter the next iteration, achieving population compression while ensuring solution quality.

## 3 K-means Algorithm Based on Improved GSA

To address the problem of K-means algorithm being susceptible to initial cluster centers and 陷入局部最优解, this paper uses the improved gravitational search algorithm to optimize cluster centers and obtains different clustering results by adjusting the number of clusters, proposing a K-means algorithm based on improved GSA called A2F-GSA-Kmeans.

The gravitational search algorithm performs optimization randomly without being affected by the initial solution. Therefore, this paper first proposes two improvement strategies to overcome GSA's tendency to 陷入局部最优解 and premature convergence. Then, the improved GSA with strong global search capability is combined with K-means to optimize cluster centers, proposing the new clustering algorithm A2F-GSA-Kmeans. Finally, clustering evaluation functions are used to evaluate different clustering results to obtain the best number of clusters.

The specific steps of the algorithm are as follows:

a) Initialize parameters including population size $N$, maximum iteration number $T$, randomly initialize particle positions $X_i$, gravitational constant $G_0$, gravitational coefficient attenuation factor $\alpha$ parameters, and optimal particle replication number $M$.

b) Set the initial number of clusters $k = 2$, with $k$ ranging in $[2, k_{\max}]$, where $k_{\max} \leq \sqrt{n}$.

c) Perform clustering on the dataset with the current specified number of clusters, complete individual position initialization, calculate fitness values, select the current optimal solution Gbest, and compute the corresponding clustering validity index.

d) Update particle inertial masses $M_i(t)$ according to equations (2) and (3).

e) Update the gravitational constant $G(t)$ according to equations (7) and (14).

f) Calculate acceleration $a_i^k(t)$ according to equation (11).

g) Update particle velocities and positions according to equations (12) and (13).

h) Clone replicate $M$ current optimal particles, perform mutation operations according to equation (21), and select the best individual from the mutated new population to enter the next iteration.

i) Check if the maximum iteration number is reached. If yes, proceed to step j); otherwise, return to step c).

j) Increment $k = k + 1$. If $k > k_{\max}$, proceed to step k); otherwise, return to step c).

k) Find the optimal number of clusters by comparing the *sil* index values.

## 4 Experimental Results and Analysis

This paper uses MATLAB R2015b development environment for simulation experiments running on a Windows 7 operating system computer to verify the effectiveness of the improved algorithm.

### 4.1 Performance Analysis of Improved GSA Algorithm

To verify the performance of the proposed A2F-GSA algorithm, relevant algorithm parameters are set as shown in Table 1 . Twelve benchmark test functions (shown in Table 2 ) are introduced for simulation experiments, including: continuous unimodal functions (F1~F4), multimodal high-dimensional functions (F5~F8), and multimodal low-dimensional functions (F9~F12).

**Table 1 Algorithm Parameter Settings**

| Parameter | GSA | GG-GSA | A2F-GSA |
|---|---|---|---|
| Population size | 50 | 50 | 50 |
| Maximum iterations | 1000 | 1000 | 1000 |
| Initial gravitational constant | 100 | 100 | 100 |
| Gravitational coefficient attenuation factor | 20 | 20 | Adaptive |

Table 3 shows the mean, minimum, and standard deviation values obtained from 30 runs of GSA 错误! 未找到引用源。, GG-GSA 错误! 未找到引用源。, and A2F-GSA algorithms on the benchmark test functions. Figure 3 [Figure 3: see original paper] shows the convergence process comparison between GSA and A2F-GSA algorithms when optimizing benchmark test functions at dimension 30. Figure 4 [Figure 4: see original paper] shows the convergence process comparison between GSA and A2F-GSA algorithms when optimizing benchmark test functions under mixed dimensions.

**Table 2 Benchmark Test Functions**

| Function | Name | Search Range |
|---|---|---|
| F1 | Schwefel 1.2 | [-100,100] |
| F2 | Schwefel 2.22 | [-10,10] |
| F3 | Schwefel 2.21 | [-100,100] |
| F4 | Step | [-100,100] |
| F5 | Ackley | [-32,32] |
| F6 | Griewank | [-600,600] |
| F7 | Penalized | [-50,50] |
| F8 | Generalized Penalized | [-50,50] |
| F9 | Shekel Goldstein-Pri | [-65.536,65.536] |
| F10 | Hartman | [-2,2] |
| F11 | Shekel' s Family | [0,1] |
| F12 | ··· | [0,10] |

**Table 3 Comparison of A2F-GSA with Other Algorithms on Benchmark Functions**

| Function | Metric | GSA 错误! 未找到引用源。 | GG-GSA 错误! 未找到引用源。 | A2F-GSA |
|---|---|---|---|---|
| F1 | Mean | 2.244E-17 | 5.417E-23 | 7.032E-32 |
| | Min | 1.130E-17 | 2.896E-25 | 5.257E-32 |
| | Std.Dev | 6.862E-18 | 2.201E-22 | 3.017E-32 |
| F2 | Mean | 2.289E-08 | 1.260E-10 | 1.442E-15 |
| | Min | 1.802E-08 | 3.510E-12 | 1.282E-15 |
| | Std.Dev | 2.658E-09 | 2.337E-10 | 1.643E-16 |
| … | … | … | … | … |
| F12 | Mean | -3.863E+00 | -3.863E+00 | -3.863E+00 |
| | Min | -3.863E+00 | -3.863E+00 | -3.863E+00 |
| | Std.Dev | 2.710E-15 | 2.710E-15 | 2.220E-16 |

As shown in Table 3, for the 12 benchmark test functions, A2F-GSA demonstrates significantly better convergence accuracy for functions F1-F3 compared to other algorithms. For the step function F4, both classical GSA and its improved variants can obtain the theoretical optimal solution, and Figures 3(a)-(c) show that convergence speed is significantly improved compared to classical GSA. This indicates that the proposed adaptive gravitational coefficient attenuation method can effectively balance global exploration and local exploitation capabilities, thereby improving solution effectiveness and efficiency.

For multimodal functions with multiple local extrema, Table 3 and Figures 3(d)-(f) show that under 30-dimensional test conditions, A2F-GSA' s optimization accuracy and speed for functions F5-F8 are significantly superior to other algorithms, demonstrating that introducing the immune clonal selection mechanism enables the algorithm to effectively jump out of local optima and improve premature convergence. Under mixed dimensions, functions F9-F12 have fewer local extrema compared to high-dimensional multimodal functions. Table 3 and Figures 4(a)-(b) show that A2F-GSA achieves theoretical optimal solutions for functions F11, F12, and F13 with lower standard deviations than other algorithms, although its optimization performance on function F9 is slightly inferior to GG-GSA, it still shows significant improvement over classical GSA.

In summary, for the 12 benchmark test functions selected, the proposed A2F-GSA algorithm achieves optimal solution accuracy, convergence speed, and robustness.

### 4.2 Performance Analysis of Improved Clustering Algorithm

To verify the effectiveness of the proposed K-means algorithm based on improved GSA in finding the optimal number of clusters, this paper conducts experiments on six datasets from the public UCI 错误! 未找到引用源。repository: Normal07, Cancer, Iris, Wine, Glass, and Abalone. The feature distributions of these datasets are shown in Table 4 .

**Table 4 Dataset Feature Description**

| Dataset | Samples | Features | Classes |
|---|---|---|---|
| Normal07 | 300 | 2 | 3 |
| Cancer | 699 | 9 | 2 |
| Iris | 150 | 4 | 3 |
| Wine | 178 | 13 | 3 |
| Glass | 214 | 9 | 6 |
| Abalone | 4177 | 8 | 3 |

Algorithm parameters use the basic settings of A2F-GSA algorithm, with the search range for number of clusters $k \in [2, k_{\max}]$. In experiments, the proposed A2F-GSA-Kmeans algorithm and classical GSA algorithm are each run 20 times on the selected test datasets. The number of times correct clustering results are obtained in the 20 runs is recorded, and clustering accuracy is calculated as:

$$\text{Clustering Accuracy}(\%) = \frac{\text{Number of correct clustering results}}{\text{Total experimental runs}} \times 100\%$$

The results are compared with the K-means algorithm based on improved artificial bee colony algorithm in 错误! 未找到引用源。, as shown in Table 5 .

**Table 5 Algorithm Accuracy Comparison**

| Dataset | K-means | Improved-Kmeans 错误! 未找到引用源。 | A2F-GSA-Kmeans |
|---|---|---|---|
| Normal07 | 100% | 100% | 100% |
| Cancer | 85% | 90% | 95% |
| Iris | 88% | 92% | 93% |
| Wine | 82% | 85% | 87% |
| Glass | 65% | 72% | 78% |
| Abalone | 45% | 48% | 50% |

Table 5 shows that for low-dimensional and clearly separated Normal07 data, all runs achieve correct clustering. For less clearly separated Iris and Wine data, A2F-GSA-Kmeans achieves slightly better accuracy than classical GSA clustering. For complex datasets like Glass and high-dimensional Cancer, A2F-GSA-Kmeans demonstrates significantly better clustering accuracy than other algorithms, verifying the effectiveness of the proposed improvements. However, for the noisy Abalone dataset, none achieve satisfactory clustering accuracy, requiring further research.

## 5 Conclusion

This paper first proposes an adaptive gravitational coefficient attenuation factor function to replace constant values, enabling the gravitational coefficient $G$ to change nonlinearly from large to small, effectively improving global exploration and local exploitation capabilities. Simultaneously, introducing the immune clonal selection mechanism into GSA enables the algorithm to effectively jump out of local optima and improve premature convergence. Simulation results on 12 benchmark test functions verify that the proposed A2F-GSA algorithm achieves better optimization performance than other algorithms. Then, combining A2F-GSA with K-means, a new A2F-GSA-Kmeans clustering algorithm is proposed. Experiments on six UCI test datasets demonstrate that the proposed A2F-GSA-Kmeans algorithm achieves significantly improved clustering quality compared to K-means algorithms based on classical GSA and improved bee colony algorithms.

## References

[1] Chen T W, Ikeda M. Design and implementation of low-power hardware architecture with single-cycle divider for on-line clustering algorithm [J]. IEEE Trans on Circuits & Systems I Regular Papers, 2013, 60 (8): 2111-2119.

[2] 杨菊蜻, 张达敏. 基于改进 BA 算法的 K-means 聚类 [J]. 计算机应用研究, 2018, 35 (05): 1454-1457. (Yang Juqing, Zhang Damin. K-means clustering algorithm based on improved BA algorithm [J]. Application Research of Computers, 2018, 35 (05): 1454-1457.)

[3] 于佐军, 秦欢. 基于改进蜂群算法的 K-means 算法 [J]. 控制与决策, 2018, 33 (1): 181-185. (Yu Zuojun, Qin Huan. K-means algorithm based on improved artificial bee colony algorithm [J]. Control and Decision, 2018, 33 (1): 181-185.)

[4] Rashedi E, Nezamabadi-Pour H, Saryazdi S. GSA: a gravitational search algorithm [J]. Information Sciences. 2009, 179 (13): 2232-2248.

[5] Liu Chao, Niu Peifeng, Li Guoqiang, et al. A hybrid heat rate forecasting model using optimized LSSVM based on improved GSA [J]. Neural Processing Letters, 2017, 45 (1): 299-318.

[6] Sun Genyun, Zhang Aizhu, Wang Zhenjie, et al. Locally informed gravitational search algorithm [J]. Knowledge-Based Systems, 2016, 104 (C): 134-144.

[7] Mirjalili S, Gandomi A H. Chaotic gravitational constants for the gravitational search algorithm [J]. Applied Soft Computing, 2017, 53: 227-238.

[8] Mirjalili S, Lewis A. Adaptive gbest-guided gravitational search algorithm [J]. Neural Computing & Applications, 2014, 25 (7-8): 1569-1584.

[9] Darzi S, Kiong T S, Islam M T, et al. A memory-based gravitational search algorithm for enhancing minimum variance distortionless response beamforming [J]. Applied Soft Computing, 2016, 47 (C): 103-118.

[10] Vijay K B, K. V. Arya, An effective gbest-guided gravitational search algorithm for real-parameter optimization and its application in training of feedforward neural networks [J]. Knowledge-Based Systems, 2018, 143: 236-257.

[11] 朱连江, 马炳先, 赵学泉. 基于轮廓系数的聚类有效性分析 [J]. 计算机应用, 2010, 30 (S2): 139-141+198. (Zhu Lianjiang, Ma Bingxian, Zhao Xuequan. Clustering validity analysis based on silhouette coefficient [J]. Journal of Computer Applications, 2010, 30 (S2): 139-141+198.)

[12] 范炜锋. 万有引力搜索算法的分析与改进 [D]. 广州: 广东工业大学, 2014. (Fang Weifeng. Analysis and improvement of gravitational search algorithm [D]. Guangzhou: Guangdong University of Technology, 2014.)

[13] 蒋建国, 谭雅, 董立明, 等. 改进的万有引力搜索算法在边坡稳定分析中的应用 [J]. 岩土工程学报, 2016, 38 (3): 419-425. (Jiang Jianguo, Tan Ya, Dong Liming, et al. Application of modified gravitational search algorithm in slope stability analysis [J]. Chinese Journal of Geotechnical Engineering, 2016, 38 (3): 419-425.)

[14] 赵辉, 李牧东, 翁兴伟. 分布式人工蜂群免疫算法求解函数优化问题 [J]. 控制与决策, 2015, 30 (7): 1181-1188. (Zhao Hui, Li Mudong, Weng Xingwei. Distributed artificial bee colony immune algorithm for the problems of function optimization [J]. Control and Decision, 2015, 30 (7): 1181-1188.)

[15] 吴建辉, 章兢, 李仁发, 等. 多子种群微粒群免疫算法及其在函数优化中应用 [J]. 计算机研究与发展, 2012, 49 (9): 1883-1898. (Wu Jianhui, Zhang Jing, Li Renfa, et al. A multi-subpopulation PSO immune algorithm and its application on function optimization [J]. Journal of Computer Research and Development, 2012, 49 (9): 1883-1898.)

[16] DeCastro L N, Zuben V. Learning and optimization using the clonal seletion. Issue on Artificial Immune System (AIS) . 2002, 6 (3): 239-351.

[17] 舒万能, 丁立新. 克隆选择算法的优化和品质因数 [J]. 软件学报, 2016, 27 (11): 2763-2776. (Shu Wanneng, Ding Lixin. Optimization and quality factor of clonal selection algorithm [J]. Journal of Software, 2016, 27 (11): 2763-2776.)

[18] Mohammadi M, Raahemi B, Akbari A, et al. Improving linear discriminant analysis with artificial immune system-based evolutionary algorithms [J].

Information Sciences, 2012, 189 (7): 219-232.

[19] Zhang Nan, Li Chaoshun, Li Ruhai, et al. A mixed-strategy based gravitational search algorithm for parameter identification of hydraulic turbine governing system [J]. Knowledge-Based Systems, 2016, 109: 236-247.

[20] University of California, Irvine. UCI machine learning repository [DB/OL]. [2013-06-19]. http://archive.ics.uci.edu/ml/datasets.html