International Journal of Smart Home
Vol. 9, No. 3 (2015), pp. 205-214
http://dx.doi.org/10.14257/ijsh.2015.9.3.19

# Research on Improved Firefly Optimization Algorithm Based on Cooperative for Clustering

Zhao Hongwei[1],Tian Liwei*[1] and Wang Dongzheng[2]

*1.School of Information Engineering, Shenyang University*
*2.Electronic Information and Engineering College, Dalian University of Technology*
*zhw30@163.com,tianliwei@163.com*

### *Abstract*

*This paper built a optimization model and proposed an improved firefly optimization algorithm called CFA, which is based on firefly Cooperative. The main idea of CFA is to extend the single population FA to the interacting multi-swarms by cooperative Models. In this work, firstly, CFA algorithm is used for optimizing six widely-used benchmark functions and the comparative results produced by, firefly optimization algorithm(FA) are studied. Secondly, CFA algorithm used in data mining, clustering analysis on several typical data sets. The performance of typical data clustering results showed that the biological heuristic algorithm based on clustering analysis algorithm with the existing success of FA compared to faster convergence, and the clustering of higher quality.*

***Keywords:*** *Firefly, Particle Swarm Optimization, SwarmIntelligence, Data clustering*

## 1. Introduction

Nature-inspired meta-heuristic algorithms are becoming powerful in solving modern global optimization problems especially for the NP-hard optimization. In recent years, many Nature-inspired meta-heuristic algorithms have been proposed: such as Ant Colony Optimization (ACO), Particle Swarm Algorithm (PSO), Firefly Algorithm (FA), *etc.* In recent years, firefly algorithms as a rich source of potential engineering applications and computational model have attracted more and more attentions. A few models have been developed to model the firefly behaviors and been applied for solving practical problems [1].

The Firefly Algorithm (FA) [2] is a new swarm intelligence algorithm that has been used for solving some optimization problems. This algorithm is based on the behavior of social fireflies. In social firefly colonies, each individual seems to have its own agenda and yet the group as a whole appears to be highly organized. Algorithms based on nature have been demonstrated to show effectiveness and efficiency, which has been applied successfully to some engineering problems, such as constrained optimization problems, neural networks and clustering [3].

In this study, A novel Firefly algorithm (CFA) is designed. As a generalized neighborhood search algorithm, CFA uses swarm intelligence of biosphere to solve optimization problems, by means of heuristic search strategy, whose capacity of tracking changes rapidly gives algorithm the ability of global optimization, because of the characteristics of global convergence itself, and the initial value can be set as fixed or random allowing parameters to be set in a wider scope. CFA has strong adaptability and parallelism; many behavior combinations can be selected due to its good flexibility, and it can get better optimization performance which genetic algorithm and particle swarm optimization do not possess. This artificial intelligence model, based on biological behavior, is different from the classical

ISSN: 1975-4094 IJSH

pattern. Firstly design a single entity perception, behavioral mechanisms, and then place a group of entities in the environment so that they can solve the problems in environment interaction, however making the best reaction under the stimulation of the environment is the basic idea of CFA.

The Firefly Algorithm (FA), which a population-based numerical optimization algorithm in the literature [4-5]. FA has been applied successfully to clustering. However, experimentation with complex and multimodal benchmark functions reveal that the FA algorithm possesses a poor convergence behavior compared to other SI algorithms and its performance also heavily decreases with the growth of the search space dimensionality. Tao, X. M (2010) introduced the K-means algorithm [6] to speed up the iteration, but the performance was unstable because of many random processes in FA which affected the practical application of the method. Xie, J. Y., (2010) obtained clusters automatically for the amount of K and applied them to arbitrary shape of data, better parallelism, but the quality of ultimately clustering quality was affected by the number and the size of grids which led to some limitations [7].

As an important research direction of data mining, clustering algorithm is a suitable means of classifying data for different patterns based on the different characteristics of different objects [8]. But the traditional clustering has greater ability of local search, for it is very sensitive to the initial cluster centers and easily falls into local optimum. If outliers are randomly selected as the initial centers, the whole quality of classification will decline. AF is less sensitive to initial values, even if to global optimization, which has bad convergence and slower iteration rate in late period [9]. Aiming at the advantages and disadvantages of both algorithms, this paper presents a global optimization idea to improve clustering algorithm based on CFA, the result of which on a small data set shows that the improved algorithm obtains clear classifications and better performance [10].

The performance of the CFA algorithm on clustering is compared with the results of other nature inspired techniques FA and Particle Swarm Intelligence (PSO) algorithm on the UCI database [11]. The CFA, FA and PSO algorithms are in the same class of population-based, SI optimization techniques. Hence, we compare the performance of the CFA algorithm with FA and PSO algorithms.

## 2. Standard AF algorithm

### 2.1. Basic Firefly Algorithm

Firefly Algorithm is a new evolutionary algorithm and its principle is to use the luminescence properties of fireflies, searching for other brighter individual in the designated area and get closer to it, in order to achieve its location optimization. The information of each firefly including position, light intensity and attractiveness between neighboring fireflies, and this information cause the location update of fireflies. Variation in luminescent can be analytically expressed by the following Gaussian form [12]:

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

where $I$ is the new light intensity, $I_0$ is the maximum fluorescence brightness value of each firefly, $\gamma$ is the light absorption coefficient and $R_j$ is the distance between any two fireflies $i$ and $j$.

The attractiveness to the luminescent can be analytically represented as

$$\beta\ (\gamma) = \beta_0 \times e^{-\gamma R_j^2} \qquad (2)$$

where $\beta_0$ is the maximum attractiveness and $\beta_0 \in [0,1]$.

The movement of a firefly is attracted to another more attractive (brighter) firefly is determined by

$$x_i(t+1) = x_i(t) + \beta_0 \times e^{-\gamma R_j} \times (x_j(t) - x_i(t)) + \alpha \times (rand - 1/2) \qquad (3)$$

$x_i(t)$ and $x_j(t)$ is the position of firefly $i$ and $j$ at generation $t$, $\alpha$ is the random parameter and $rand$ is a random number that uniformly distributed in [0 1].

Firefly Algorithm is using firefly optimization process to simulate the search process of the optimal solution. Each firefly represents a viable solution to the problem that randomly distributed in the solution space , the corresponding fluorescent brightness evaluate whether the corresponding feasible solution is good , the position of fireflies is constantly updating under the influence of the relative attractiveness to find the optimal solution. In the basic Firefly Algorithm search model, if brighter individual is found in the neighborhood, then the searching firefly will move towards the brighter one with a certain step, but the searching firefly have not consider the influence of the historical best position of group .Suppose a firefly individual find a brighter firefly individual in the neighborhood and move towards it ,but its moving route of the individual deviate from the direction of optimum position and will inevitably reduce the convergence speed and easily to run into local optimum. Therefore, a Firefly Algorithm with the influence of the historical best position of group is proposed when considering the influence of the historical best position of group [13, 14].

### 2.2.The FA Algorithm Steps

In what follows we briefly outline the original FA algorithm steps:

**[Step1]** $m$ fireflies are Randomly placed within the search range, maximum attractiveness is $\beta_0$, the light absorption is $\gamma$ ,the randomization parameter is $\alpha$ ,the maximum number of iterations is $T$ , the position of fireflies is random distributed.

**[Step2]** Calculate the fluorescence brightness of fireflies. Calculate the objective function values of Firefly Algorithm that use the improved Firefly Algorithm as the largest individual fluorescence brightness value $I_0$ .

**[Step3]** Update the position of firefly. When the firefly $i$ is no only attracted by a brighter firefly $j$ but also influenced by the historical best position of group, the position formula is updating as function(4).The brightest fireflies will update their position as the following function:

$$x_{best}(t+1) = x_{best}(t) + \alpha \times (rand - 1/2) \qquad (4)$$

Where $x_{best}(t)$ is the global optimal position at generation $t$.

**[Step4]**Recalculate the fluorescence brightness value $I_0$ by using the distance measure function $r_{tr}(S)$ after updating the location and searching the local area for the strongest fluorescence brightness individual, updating the optimal solution when the target value is improved ,otherwise unchanged.

**[Step5]** When reach the maximum iteration number $T$ , record the optimal solution, otherwise repeat step (c), (d), (e) and start the next search. The optimal solution is also the global optimum value $H_{\max}$ and the global optimum image threshold is the corresponding threshold value $(s,t)$ at the position $x_{best}(t)$.

## 3. The Cooperative Firefly Algorithm(CFA)

In order to overcome the premature convergence of classical FA, Cooperative optimization model is incorporated into FA to construct an improved FA in this paper. In order to improve the balance between the exploration and exploitation in CFA we propose a modification of Eq. (5) used in traditional FA.

The aim of the modification is to restore balance between exploration and exploitation affording increased probability of escaping basin of attraction of local optima. In the proposed CFA ,the firefly $i$ find a brighter firefly $j$ when iterative search use Firefly Algorithm ,then $i$ move towards $j$ with a certain step, but the direction of movement will deflect under the influence of the historical best position of group. The direction that $i$ towards $j$ synthesize with the direction that $i$ towards the historical best position of group （$x_{best}$） is the deflect direction, in this way each search is affected by better solutions thereby improving the convergence rate. The principle of Firefly Algorithm with the influence of the historical best position of group .Suppose any firefly $i$ in the searching range is attracted by a brighter firefly $j$ and influenced by the historical best position of group, then the original direction of movement will change and move towards the optimal direction, thus speeding up the convergence rate.

$$\begin{aligned} & x_i(t+1) = x_i(t) + \beta_0 e^{-\gamma R_j^2} \\ & \times (x_j(t) - x_i(t)) + \beta_1 e^{-\gamma R_{\text{best}}^2} \\ & \times (x_{best}(t) - x_i(t)) + \lambda(rand - 1/2) \end{aligned} \tag{5}$$

In the proposed CFA, Eq. (4) of traditional FA based on constants values of $\alpha$ and is modified by Eq. (5) Using new variables, $\lambda$ and $\beta_1$. In this case, the fireflies are adjusted by:

Schematic diagram of Firefly Algorithm with the influence of the historical best position of group. The movement of a firefly is attracted to another more brighter firefly with the influence of the historical best position of group is determined by where is the updating position of firefly, $x_i(t)$ is the initial position of firefly which play an important role in balancing the global searching and the local searching, $\beta_0 \times e^{-\gamma R_j^2} \times (x_j(t) - x_i(t))$ represents the position of fireflies update under the attraction between fireflies, $\beta_1 e^{-\gamma R_{best}^2} \times (x_{best}(t) - x_i(t))$ represent the updating position of fireflies under the influence of the historical best position of group, $\lambda \times (rand - 1/2)$ is the random parameter that can avoid the result falling into local optimum.

The pseudo-code of CFA is described in Table 1. The CFA can improve the algorithm convergence rate significantly.

**Table 1. The CFA Algorithm**

| ***Algorithm：　The Improved FA algorithm*** |
|---|
| Input: |
| Create an initial population of fireflies n within |
| d-dimensional search space $x_{ik}$, i = 1, 2, . . . , n and |
| k = 1, 2, . . . , d |
| Evaluate the fitness of the population f ($x_{ik}$) which is directly |
| proportional to light intensity $I_{ik}$ |
| Algorithm' s parameter— β 0, γ |
| Output: |
| Obtained minimum location: $x_i$ |
| min |
| begin |
| repeat |
| for i = 1 to n |
| for j = 1 to n |
| if ($I_j < I_i$) |
| Move firefly i toward j in |
| d-dimension using Eq. (4) |
| end if |
| Attractiveness varies with distance r via |
| exp[−r₂] |
| Evaluate new solutions and update light |
| intensity using Eq. (5) |
| end for j |
| end for i |
| Rank the fireflies and find the current best |
| until stop condition true |
| end |

# 4. Experimental Result

## 4.1 Benchmark Functions

Ten well-known benchmark functions are used in the test. These functions contain three uni-modal functions, four functions and three rotated functions.

The first function is Sphere function whose global minimum value is 0 at (0, 0, … , 0). Initialization range for the function is [−5.12, 5.12]. It is a unimodal function with non-separable variables.

$$f_1(x)=\sum_{i=1}^{n} x_i^2 \qquad x\in[-5.12,5.12]^D \tag{6}$$

The second function is Rosenbrock function whose global minimum value is 0 at (1, 1, ⋯ , 1). Initialization range for the function is [−15, 15]. It is a unimodal function with non-separable variables. Its global optimum is inside a long, narrow, parabolic shaped flat valley. So it is difficult to converge to the global optimum.

$$f_2(x)=\sum_{i=1}^{n} 100\times(x_{i+1}-x_i^2)^2+(1-x_i)^2 \quad x\in[-3,3]^D \tag{7}$$

The fourth function is Rastrigin function whose global minimum value is 0 at (0, 0, ⋯ , 0). Initialization range for the function is [−15, 15]. It is a multimodal function with separable variables.

$$f_4(x)=\sum_{i=1}^{D}\left(x_i^2-10\cos(2\pi x_i)+10\right) \tag{8}$$

The third function is Quadric function whose global minimum value is 0 at (0, 0, ⋯ , 0). Initialization range for the function is [-10, 10]. It is a unimodal function with non-separable variables.

$$f_3(x)=\sum_{i=1}^{D}\left(\sum_{j=1}^{i}x_j\right)^2 \tag{9}$$

### 4.2 Results for the 10-D Problems

For most cases in our implementation, we can take $\beta_0$ = 1, α in[0,1], ɤ=1, and ƛ=1:5. In addition. In the 10 dimensions should be determined by the actual scales of the problem of interest. The parameter g now characterizes the variation of the attractiveness, and its value is crucially important in determining the speed of the convergence and how the FA algorithm behaves. Table 2 lists the experimental results for each algorithm on four functions.

From the results, we observe that CFA achieved better results on all test problems than the original FA. As we can see in Figure 1, under the influence of the serial heterogeneous cooperative approach: the firefly starts exploring the search space at every phase. So, the fireflies slow down near the optima to pursue the more and more precise solutions.

**Table 2. Comparison among PSO, FA, CFA and GA on 10-D Problems**

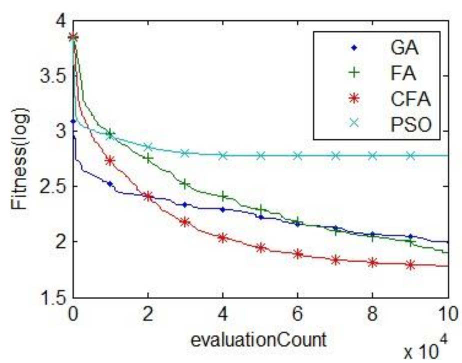
| 10D | | PSO | CFA | FA | GA |
|---|---|---|---|---|---|
| Sphere | Best | 6.06E+00 | **1.48E-07** | 1.00E+00 | 3.11E-04 |
| | Worst | 3.05E+01 | **1.48E-07** | 1.00E+00 | 3.45E-02 |
| | Mean | 1.96E+01 | **1.48E-07** | 2.10E+00 | 1.18E-02 |
| | Std | 5.41E+00 | **1.48E-07** | 1.00E+00 | 9.40E-03 |
| Rosenbrock | Best | 8.91E+00 | **1.48E-07** | 1.96E-02 | 7.81E+00 |
| | Worst | 6.22E+01 | **4.07E-07** | 4.07E+00 | 9.81E+00 |
| | Mean | 1.21E+01 | **2.96E-07** | 7.29E-01 | 8.66E+00 |
| | Std | 9.50E+00 | **1.95E-07** | 1.50E+00 | 5.56E-01 |
| Rastrigin | Best | 9.95E+00 | **1.48E-07** | 2.98E+00 | 3.30E+00 |
| | Worst | 5.37E+01 | **9.75E-02** | 2.59E+01 | 7.01E+00 |
| | Mean | 3.64E+01 | **1.11E-02** | 1.08E+01 | 4.95E+00 |
| | Std | 1.06E+01 | **2.57E-02** | 4.60E+00 | 1.15E+00 |
| Quadric | Best | 3.56E+01 | **1.39E-01** | 2.71E+01 | 1.68E-01 |
| | Worst | 1.32E+02 | **4.40E-01** | 7.97E+01 | 6.32E-01 |
| | Mean | 9.98E+01 | **2.70E-01** | 6.30E+01 | 3.44E-01 |
| | Std | 2.48E+01 | **8.46E-02** | 1.23E+01 | 1.54E-01 |

### 4.3 Results for the 20-D Problems

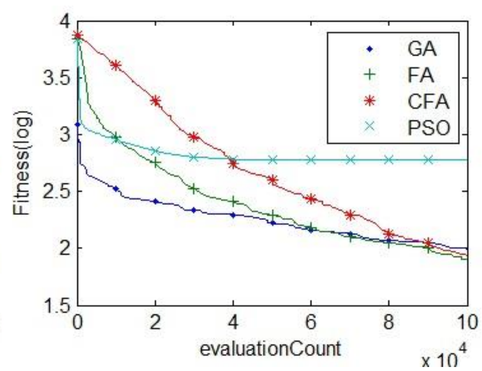
This experiment conducted on 20-D problems to compare the proposed CFA with the original FA, from the results, we observe that CFA achieved better results on all test problems than the original FA, GA and PSO. This experiment runs 30 times respectively for each algorithm on each benchmark function. Table 3 lists the experimental results for each algorithm on functions Sphere, Rosenbrock, Rastrigrin and Quadric. Figure 1 shows the search progresses of the average values found by all algorithms over 30 runs for four functions. From Figure 1, the CFA algorithm surpasses all other algorithms on four functions.

**Table 3. Results for All Algorithms on Benchmark Functions**

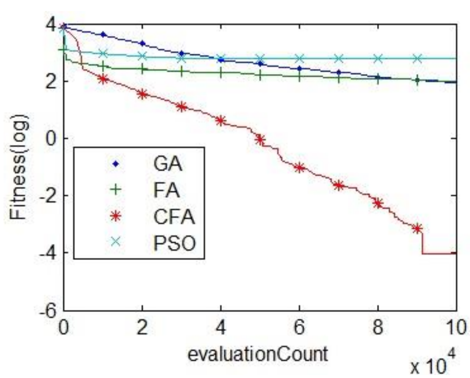
| 20D | | PSO | FA | CFA | GA |
|---|---|---|---|---|---|
| Sphere | Best | 7.8172e-021 | 3.2041e-021 | **1.6297e-024** | 8.84E-01 |
| | Worst | 2.8808e-015 | 6.2272e-012 | **2.2354e-017** | 3.46E-01 |
| | Mean | 1.5536e-016 | 2.3795e-013 | **6.0619e-019** | 3.89E-01 |
| | Std | 5.3120e-016 | 1.0571e-012 | **3.5620e-018** | 1.15E+00 |
| Rosenbrock | Best | 1.6760e-001 | 4.3780e-001 | **6.1030e-002** | 6.56E+02 |
| | Worst | 8.9007e+002 | 3.5381e+003 | **1.5974e+002** | 4.56E+02 |
| | Mean | 5.1215e+001 | 3.2367e+001 | **2.0250e+001** | 2.01E+02 |
| | Std | 1.4576e+002 | 5.7988e+001 | **3.9885+001** | 2.43E+03 |
| Rastrigrin | Best | 8.8337e-010 | 1.2299e+000 | **2.9849e+000** | 6.81E+01 |
| | Worst | 7.9647e+000 | 1.9900e+001 | **4.0793e+001** | 1.79E+01 |
| | Mean | 3.9028e+000 | 8.5567e+000 | **1.4632e+001** | 3.17E+01 |
| | Std | 2.7131e+001 | 5.6043e+000 | **1.0297e+001** | 1.16E+02 |
| Quadric | Best | 3.9600e-002 | 2.1600e-002 | **2.7010e-002** | 1.88E+01 |
| | Worst | 2.6810e-001 | 1.5520e-001 | **2.0680e-001** | 1.42E+00 |
| | Mean | 8.3612e-002 | 9.8000e-002 | **9.0800e-002** | 1.41E+01 |
| | Std | 4.6112e-002 | 4.1500e-002 | **3.7100e-002** | 2.01E+01 |



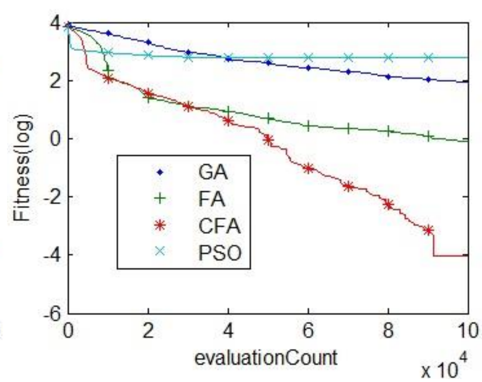

(a) Sphere-10D



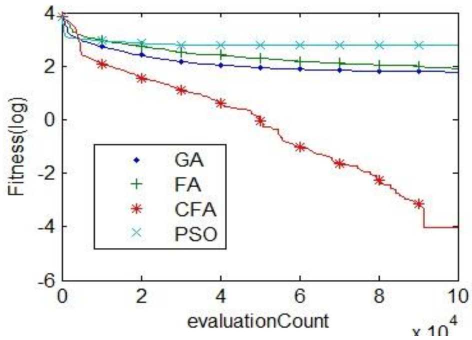

(b) Sphere-20D



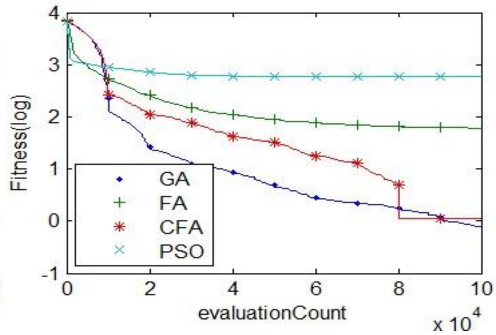

(c) Rosenbrock-10D



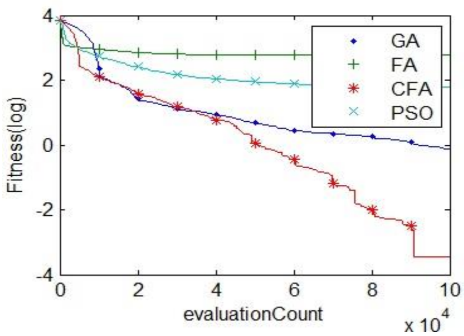

(d) Rosenbrock-20D



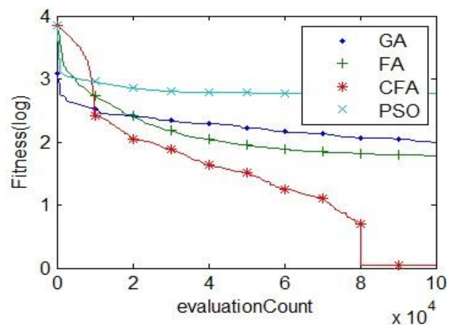

(e) Rastrigin-10D




(f) Rastrigin-20D




(g) Quadric-10D




(h) Quadric-20D

**Figure 1. Convergence Results of GA , FA, CFA and PSO on 10D and 20-D benchmark functions. (a) Sphere-10; (b) Sphere-20; (c) Rosenbrock-10; (d) Rosenbrock-20 (e) Rastrigin-10; (f) Rastrigin-20; (g) Quadric-10; (h) Quadric-20**

## 5. Data Clustering Experimental Results

To analyze the performance of the proposed CFA approach for clustering algorithm, the results of PSO FA and GA with different data sets have been compared in this paper, which are selected from the UCI machine learning repository.

The algorithm base on CFA algorithms is used for data clustering on Iris data sets, which is able to provide the same partition of the data points in all runs. Iris data is thus selected from the UCI machine learning repository, cluttering result of which sets by FA and the CFA clustering algorithm is presented in Figure 2. From the result Figure 2, for all real data sets, the basic clustering algorithm with CFA outperforms the other methods.

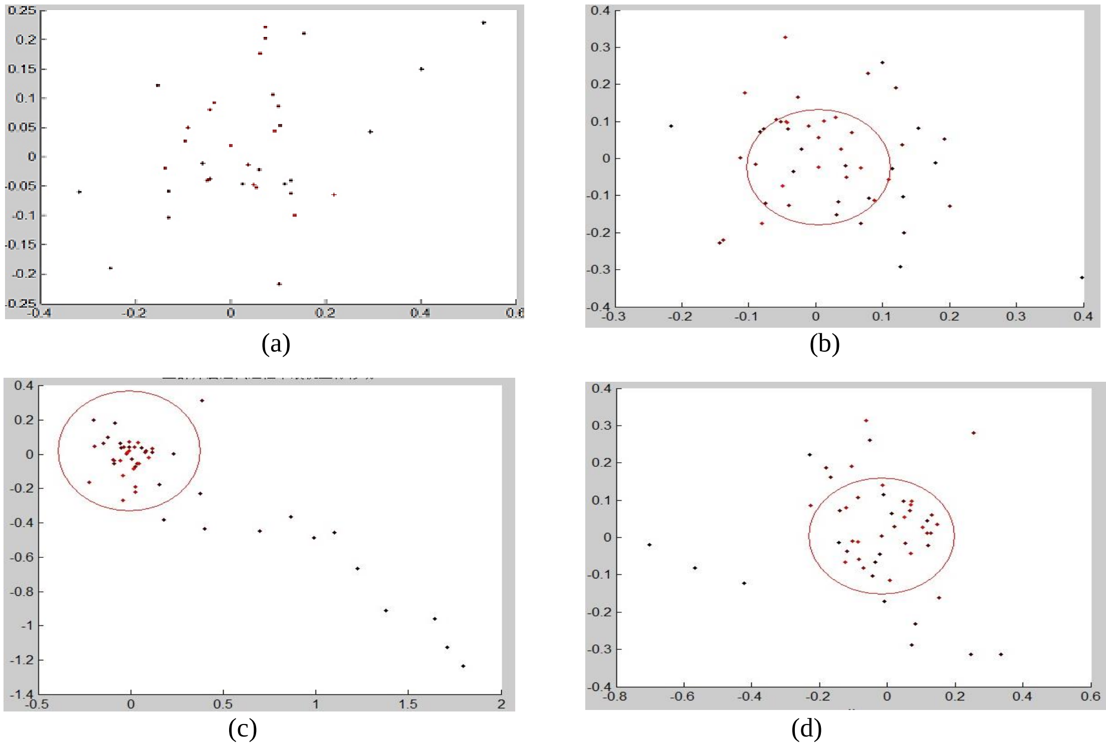



**Figure 2. The Data Distribution of Iris Data Sets and the Clustering Result by CFA and Hybrid Algorithm. (a) Iris Distribution (b) Iris Clustering Result Base on PSO Algorithm (c) Iris Clustering Result Base on FA Algorithm (d) Iris Clustering Result Base on CFA Algorithm**

## 6. Conclusion

In This paper investigates a new nature inspired algorithm—the FA is used for clustering and evaluating its performance. Firstly, based on the cooperative approaches, a novel Firefly (FA) algorithm is presented, namely Cooperative Firefly(CFA). Secondly, in order to demonstrate the performance of the CFA algorithm, we compared it with those of FA, PSO,GA optimization algorithms on several benchmark functions. The FA algorithm is compared with PSO, FA and GA as all these methods are in the same class of population-based, nature inspired optimization techniques. Comparison of experimental results show, the clustering algorithm based on CFA makes similar data gather obviously,and it distinguishes samples precisely while also improving the cluster quality and obtaining better centers with clear division which represents reducing computation amount .

## Acknowledgements

The authors are supported financially by International cooperation project (Project No.S2012ZR0191) and the Natural Science Foundation of Liaoning Province (Project No.2013020011) and the Social Science Foundation of Liaoning Province (Project No.L14ASH001) and this work is supported in part by the International S&T Cooperation Program of China (ISTCP) under Grant 2011DFA91810-5 and Program for New Century Excellent Talents in University of Ministry of Education of China under Grant NCET-12-1012.

## References

[1] M. A. Badamchizadeh, A. Nikdel and M. Kouzehgar, “Comparison of Genetic Algorithm and Particle Swarm Optimization for Data Fusion Method Based on Kalman Filter”, International Journal of Artificial Intelligence, vol. 5, no. 10, **(2010)**.

[2] E. Falkenauer, Editor, “Genetic Algorithms and Grouping Problems”, New York: Wiley, Chichester, **(1998)**.

[3] Y. Kao and K. Cheng, “An ACO-based clustering algorithm”, Edited M. Dorigo, Springer, Berlin, LNCS, vol. 4150, **(2006),** pp. 340–347.

[4] R. Younsi and W. Wang, “A new artificial immune system algorithm for clustering”, Edited Z. R. Yang, Springer, Berlin, LNCS, vol. 3177, **(2004),** pp. 58–64.

[5] I. Falco and A. D. Cioppa and E. Tarantino, “Facing classification problems with particle swarm optimization”, Applied Soft Computing, vol. 7 , no. 3, **(2007)**.

[6] X. D. He and L. D. Qu, “Firefly Clustering Algorithm. Application Research of Computers, vol. 26, no. 10, **(2009)**.

[7] D. Karaboga and C. Ozturk, “A novel clustering approach: Articial Bee Colony (ABC) algorithm”, Applied Soft Computing, vol. 11, no. 1, **(2011)**.

[8] H. Li and Q. F. Zhang, “Multi objective optimization problems with complicated Pareto sets”, IEEE Transactions on Evolutionary Computation, vol. 13, no. 2, **(2009)**.

[9] B. Liu and Y. Q. Zhou, “A Hybrid Clustering Algorithm Based on Firefly Algorithm”, Journal of Computer Engineering and Applications, vol. 44, no. 18, **(2008).**

[10] I. E. Evangelou, D. G. Hadjimitsis, A. Lazakidou and C. Clayton, Editors, “Data mining and knowledge discovery in complex image data using artificial neural networks”, in Proceedings of the Workshop on Complex Reasoning an Geographical Data, **(2001)** December 1-3, Paphos, Cyprus.

[11] M. S. Kamel and S. Z. Selim, “New algorithms for solving the fuzzy clustering problem”, Pattern Recognition, vol. 27, no. 3, **(1994)**.

[12] T. Sato and M. Hagiwara, Editors, “Bee system: finding solution by a concentrated search; Systems, man, and cybernetics”, IEEE international conference on computational cybernetics and simulation, **(1997)** October 12-15, Orlando, FL, USA.

[13] X. M. Tao, J. Xu and L. B. Yang, “An Improved Hybrid Algorithm Based on Particle Swarm Optimization and K-means Algorithm”, Journal of Electronics &Information Technology, vol. 32, no. 1, **(2010)**.

[14] S. Suresh, N. Sundararajan and P. Saratchandran, “A sequential multi-category classifier using radial basis function networks”, Neuro computing, vol. 71, no. 7, **(2008).**

## Authors

**Zhao Hongwei,** male, M.Sc. ,associate professor, Shenyang University, supervisor for postgraduates, School of Information Engineering, His main research areas are Swarm intelligence theory and Optimization decision method.

**Tian Liwei**, male, ph. D, professor, Shenyang University, supervisor for postgraduates, School of Information Engineering, His main research areas are Computational Biology and Swarm intelligence theory.