***Research on Modified Artificial Bee Colony Clustering Algorithm***

Lilu Cao 1st Affiliation

Transportation and Management Department Dalian Maritime University

Dalian. R. China

e-mail: [crown\_0726@163.com](mailto:crown_0726@163.com)

Dashen Xue 2nd Affiliation

Transportation and Management Department Dalian Maritime University

Dalian. R. China

e-mail: [xds59@dlmu.edu.cn](mailto:xds59@dlmu.edu.cn)

*Abstract—*In order to overcome the disadvantages of the K-Means Clustering algorithm, such as the poor global search ability, being sensitive to initial cluster centric, as well as the vu-lnerable to trap in local optima and the slow convergence velocity in later period of the original Artificial Bee Colony (ABC) algorithm, a Modified ABC algorithm was proposed．Modified Artificial Bee Colony algorithm combined with K-means Clu-stering algorithm, named it as MABC-K-means algorithm, to establish Hybrid algorithm for solving framework. Through extensive testing, the MABC-K-means algorithm can improve cluster performance effectively. Finally, according to optimiz-ation solution strategy, instantiate Customer Relationship Mana-gement issue in the process of instantiating framework.

Keywords-cluster analysis; Modified Artificial Bee Colony algorithm; Customer Relationship Management

# Introduction

Main content of this paper consists of two aspects. On the one hand, Modified Artificial Bee Colony algorithm combined with K-means Clustering algorithm, get a hybrid solution strategy[1] and discuss the effectiveness of it. On the other hand, Study on the application of MABC-K-means algorithm in the Customer Relationship Management system. Achieve rapid analysis about business data of Customer Relationship Management system and find the valuable knowledge and rules for business decision making.

# Modified Artificial Bee Colony K-means algorithm

## Introduction of method

The process of Modified Artificial Bee Colony algorithm can be divided into three stages: Hire-bee stage, Follow-bee stage and Scout-bee stage.

Hire-bee stage, Hire-bee develop food source based their memory, and convey the information of food source to the Follow-bee. When a food source developed, Hire-bee will become Scout-bee and find new sources of food.

Follow-bee stage, Follow-bee wait for Hire-bee convey food source information in the hive, and then chooses the best food source from the information on the method of roulette.

Scout-bee stage, Scout-bee search for the new food source randomly in a certain range according to the internal motivation or external clues that possible. Once the Scout-bee finds new food sources, it will become Hire-bee, and constantly develop the food source until reaching the limit condition.

The work of Scout-bee can be regarded as a process of exploration, and the work of Hire-bee and Follow-bee as the development process throughout the whole search process.

In the Modified Artificial Bee Colony algorithm, the food source location represents the optimal solution of the optim-ization problem, and the food source income represents the solution's fitness of this problem. Firstly, the algorithm will randomly generate a plurality of food source location according to the population scale, each food source position represents a solution of D-dimensional in above problem, such as: .

(1)

(2)

indicates the i solution’s function value, and indicates the i solution fitness value. indicates the probability of Follow-bee select Scout-bee.

The update of food source location in algorithm can be expressed by the following formula:

(3)

, in this formula, represents the current optimal solution. The introduction of optimal solution[2] which can produce better feasible solution, thus can increase the algorithm's ability of development. Both and is random number, and the K must be different from I. is the i food source of the j elements. and are choice individual randomly from population, but the choice not the same.

If a solution is not update in long-term, then give the following solution and randomly generated a new solution by the Scout-bee. Its expression is as follows:

(4)

In this formula . The search performance of Modified Artificial Bee Colony algorithm depend on the balance of global exploration and local exploitation ability, while it depend on the selection of control parameters in a great degree.

The control parameters of this algorithm including, population size, value of "limit", ect. Among them, the value of "limit" shows that if a solution cannot be improved continuously after recycling "limit" times, and the food source of income is not the global optimal solution currently, then this solution into a local optimal solution, and then discard this solution.

The initial cluster center of K-Means clustering algorithm is selected randomly cause local optimal solution rather than the global optimal solution. According to this situation, the Modified Artificial Bee Colony algorithm combined with K-Means algorithm in this paper, a new algorithm was proposed, called MABC-K-means algorithm. In this hybrid algorithm, generate the initial cluster center by Modified Artificial Bee Colony algorithm and calculate the average value of cluster. Each object in the collection of samples as the nearest cluster again, then solving the average of sample object of each cluster. The algorithm start convergence until the criteria function do not change, finally get the number of clusters.

The clustering procedure of Modified Artificial Bee Colony K-means algorithm is as follows:

Step one: Parameters initialization. The specific parameters including the number of clusters(FoodNumber, Limit), the maximum times of iterations (MaxCycle), the initial times of cycles (cycle, GlobalMin), the maximum times of cycles to K-means algorithm (MCNKM) and the initial times of cycles (cycleKM);

Step two: Create the initial solution set (i=1,2,... , FoodNumber) and calculate one solution's fitness;

Step three: The Scout-bee searches in the neighborhood to produce the new solution , calculate the solution's fitness;

Step four: Determine the fitness of is better than . If the current solution is better than the original solution, vi instead of , otherwise, remains unchanged;

Step five: Calculate the probability values for each solution ;

Step six: the Follow-bee select the corresponding cluster according to the probability, and research in the neighborhood. Then calculate the new fitness of cluster points, selecting the food source according to the step four;

Step seven: If the current yield is smaller than the optimal solution, then the current income replace the original income;

Step eight: The unprocessed data after Limit times iterations will replace with a new random data;

Step nine: Determine whether satisfie the termination condition, if the Modified Artificial Bee Colony algorithm reached the maximum times of cycles, then output the best cluster centers. Otherwise, go to step three;

Step ten: The results of Modified Artificial Bee Colony algorithm as the center points of K-means algorithm, then this algorithm initialization;

Step Eleven: Calculate the fitness of K-means algorithm in this center points, and repeat step ten;

Step twelve: Assign data to the largest cluster of fitness.

The steps of algorithm flow chart as shown in figure 1: 

1. Artificial bee colony algorithm flowchart

## Data model validation

To test the MABC algorithm's performance in the optimization of function, using 5 testing functions in this paper. The test result will compare with the standard Differential Evolution algorithm(DEA)[3], standard Genetic algorithm(GA) and standard Artificial Bee Colony algorithm(ABC)[4]. The number of population of MABC algorithm, standard Differential Evolution algorithm, standard Genetic algorithm and standard Artificial Bee Colony algorithm is 125(that is, SN=62), the value of Limit is SN\*D.

In running the test process of algorithm, each test process will run 30 times, although time-consuming, but

the results of running will more meaningful and more precise.

### The comparison of MABC algorithm with ABC, DEA and GA

In this part, MABC algorithm will compare with the standard Differential Evolution algorithm, standard Genetic algorithm and standard Artificial Bee Colony algorithm. Each algorithm compared separately in 10 dimensional, 20 dimensional and 30 dimensional. The function's mean and variance is the parameters of comparison for the per-formance of the algorithm, test results as shown in table 1.

1. the results of improved ABC algorithm to compared with the GA, EA and ABC1's

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Equation | Algorithm | GABC |  | EABC |  | ABC1 |  | MABC1 |  |
|  | Dimensions | Mean value | Variance | Mean value | Variance | Mean value | Variance | Mean value | Variance |
| Griewank | 10 | 0.050 | 0.029523 | 0.222366 | 0.0781 | 0.00087 | 0.002535 | 0 | 0 |
|  | 20 | 1.0139 | 0.026966 | 0.59036 | 0.2030 | 2.01e-08 | 6.76e-08 | 0 | 0 |
|  | 30 | 1.234 | 0.11045 | 0.8211 | 0.1394 | 2.87e-09 | 8.45e-10 | 0 | 0 |
| Rastrigin | 10 | 1.392 | 0.76319 | 0.43404 | 0.2551 | 0 | 0 | 0 | 0 |
|  | 20 | 6.031 | 1.4537 | 1.8135 | 0.2551 | 1.45e-08 | 5．06e-08 | 0 | 0 |
|  | 30 | 10.43 | 2.6386 | 3.0527 | 0.9985 | 0.033874 | 0.181557 | 0 | 0 |
| Rosenbrock | 10 | 46.31 | 33.8217 | 25.303 | 29.7964 | 0.034072 | 0.045553 | 0.0040 | 0.0043 |
|  | 20 | 103.9 | 29.505 | 72.452 | 27.3441 | 0.13614 | 0.132013 | 0.0030 | 0.0039 |
|  | 30 | 166.2 | 59.5102 | 98.407 | 35.5791 | 0.219626 | 0.152742 | 0.0014 | 0.0024 |
| Ackley | 10 | 0.5926 | 0.22482 | 0.19209 | 0.1951 | 7.8e-11 | 1.16e-09 | 5.151e-015 | 1.7161e-015 |
|  | 20 | 0.924 | 0.22599 | 0.32321 | 0.097353 | 1.6e-11 | 1.9e-11 | 2.044e-013 | 2.3554e-014 |
|  | 30 | 1.098 | 0.24956 | 0.3771 | 0.098762 | 3e-12 | 5e-12 | 1.866e-011 | 7.6374e-012 |
| Schwefel | 10 | 1.951 | 1.3044 | 0.32037 | 1.6185 | 1.27e-09 | 4e-12 | 1.273e-004 | 2.8571e-020 |
|  | 20 | 7.285 | 2.9971 | 1.4984 | 0.84612 | 19.83971 | 45.12342 | 2.555e-004 | 6.3246e-010 |
|  | 30 | 13.53 | 4.9534 | 3.272 | 1.6185 | 146.8568 | 82.3144 | 3.82e-004 | 2.7330e-007 |

This section uses 5 standard test functions, they are Griewank, Rastrigin, Rosenbrock, Ackley, Schwefel. The test consists of two processes:

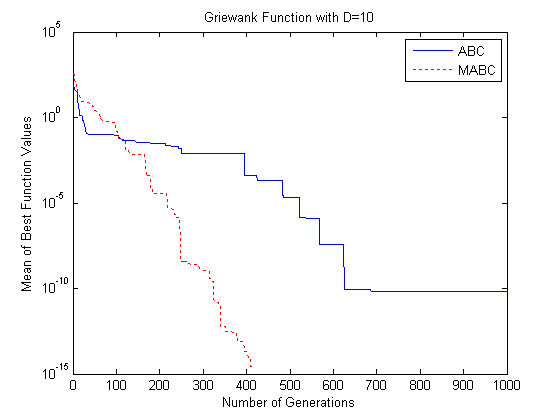
The function test separate in iteration of 500 times, 750 times and 1000 times, the three iteration corresponded are 10 dimension, the 20 dimension and 30 dimension. The function evaluation number of three dimension of each test function corresponded are 1.5104, 2.5104 and 3.0104;

In 1000 times, 1500 times and 2000 times iterative testing, related settings same as the first procedure, the times of evaluations of each test function respectively is 3.0104, 5.0104, 6.0104.

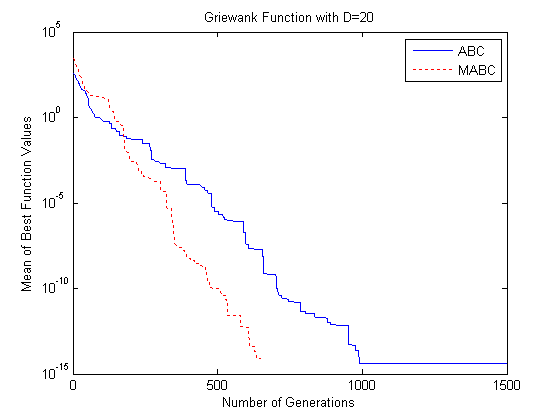
The results show that MABC has better performance in most of the functions, notably the precision is higher in Griewank function and Rastrigin function.

Figure 2 to 4 represents that the convergence curves of convergence performance of MABC in all test functions after compared with the standard Artificial Bee Colony algorithm[5].

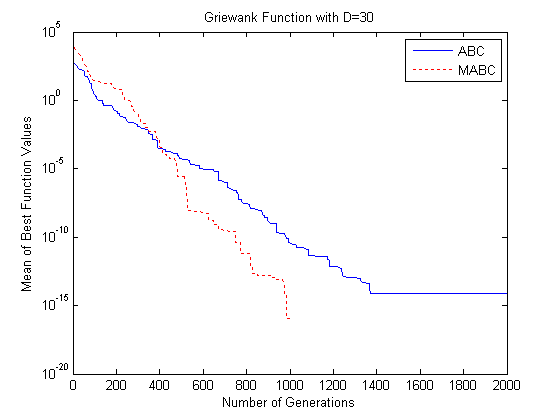
The above Figure shows, compared with the ABC algorithm, Modified Artificial Bee Colony algorithm has obvious advantages no matter in convergence or different dimensions. Otherwise, it more effective than the ABC algorithm.



1. MABC and ABC on Griewank function convergence curve (D = 10)



1. MABC and ABC on Griewank function convergence curve (D = 20)



1. MABC and ABC on Griewank function convergence curve (D = 30)

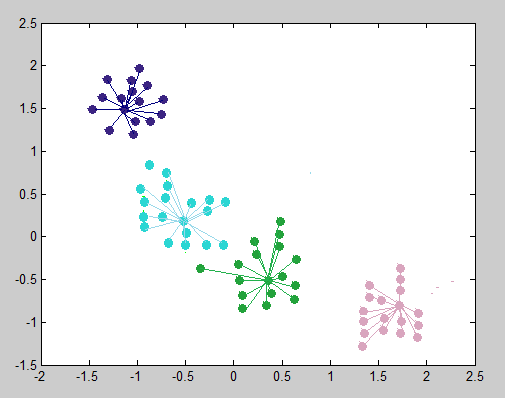
## The application about clustering of Modified Artificial Bee Colony algorithm in Customer Relationship Management

When the algorithm is used to customer clustering, We will use the customers and their orders of a e-commerce platform in the first quarter as basic data. And choose 100 per as the sample in these basic data. In order to validate the feasibility of algorithm preliminary, This paper only extracts simple DataSet for clustering analysis. Cluster variables of experiment is the customer's average times of consumption and average amount production in the e-commerce platform in the first quarter. The data after the early screening and weight, select the MABC-K-means algorithm to customer clustering analysis. And the MATLAB was used to realize the clustering results intuitively.

When use the MABC-K-means algorithm to customer clustering analysis, the results as follows: table 2 is the result of customer clustering, and the legend is shown in Figure 5.

1. Results of clients cluster analysis

|  |  |  |  |
| --- | --- | --- | --- |
| Cluster categories | Cluster centers | A single cluster number | Clustering time |
| 1 | (0.3430，-0.4339) | 21 | 0.488289 |
| 2 | (-1.1789，1.5174) | 18 |
| 3 | (-0.6679，0.1425) | 28 |
| 4 | (1.8242，-0.9121) | 31 |



1. Results of clients cluster analysis

Analysis the running results of the above algorithm can be seen that, MABC-K-means algorithm can take the customer that should be together in a class better, and remove the Hot Pixel in the algorithm data, so enhance the effection of clustering. According to customer clustering results based on MABC-K-means algorithm and analysis it with the original data, we can see that the consumption characteristics of different customer types as shown in table 3.

Combining with the analysis on the table following, we can come to the following conclusions: the number of first type customers in this table is the least, the degree is highest and have generous income. They often branch in Bejing, Shanghai or Guangzhou and other developed cities. In addition, the age of them are between of 25 to 35; Second types of customers than other customers in number, but less than the fourth customers. Most of them distributed in two or three line city, have general income and general education. Their age around 30; The third types of customers from first-tier cities, have higher incomes and higher levels of education and other characteristics. Although consumption of their is less, but the amount of consumption is very high; The fourth types of customers have the largest number. From their point of view, the degree of this kind of customer are relatively low and have less income. Their age distribution is not uniform. The average consumption times and the amount of consumption average are less than the above;

1. Consumption characteristics of different customer types

|  |  |
| --- | --- |
| Customer types | Customer characteristics |
| Category 1 | The average consumption times is 6.807  The average amount of consumption is 968.461 |
| Category 2 | The average consumption times is 13.19  The average amount of consumption is 521.47 |
| Category 3 | The average consumption times is 3.67  The average amount of consumption is 1897.51 |
| Category 4 | The average consumption times is 4.2  The average amount of consumption is 253.67 |

1. Customer category information table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Customer types | City | Educational background | Age | Income |
| 1 | First-line | High | 25-35 | High |
| 2 | Two or three-line | General | 30 | General |
| 3 | First-line | High | 35-45 | High |
| 4 | Uneven | Low | Uneven | Low |

1. Percentage of clients and value

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Customer types | 1% | 2% | 3% | 4% |
| Number percentage | 8% | 26% | 16% | 41% |
| Percentage value | 23% | 15% | 49% | 17% |
| Customer level | Important | VIP | SVIP | Generic |

Through the analysis we can see that although the number of first class customers at least, but they brought a higher income for the enterprise. These customers can be considered to have potential value. Enterprise should pay more attention to these customers; The number of second customer no more and no less, and did not bring much revenue for the enterprise also, so the enterprise should put some resources to develop this kind of customers; The number of third is not the most, but has brought a greater part of profit for the enterprise, this kind of customer is truly valuable to the enterprise, so enterprise should concentrate superior resources to maintaining these customers; The number of fourth is the largest , but income they brings to the enterprise is relatively small, they belong to the ordinary user. Enterprise need not spend too much energy to this kind of customers.

## Conclusion

The advantages that Regulating the global optimization and local optimization of Modified Artificial Bee Colony algorithm combined with advantages that fast convergence speed of K-means means algorithm, improves the robustness of the two algorithm. Through a lot of experimental analysis. The algorithm not only overcomes the disadvantages that poor stability of traditional K-means algorithm, and the clustering effect is improved obviously, also, it proves the rationality and validity of the combination of these two methods.

Through the Modeling of problem domain, realization of rules and Study of methods in this problem, Modified Artificial Bee Colony algorithm engine is implemented. Through quantitative description of the problem classifier and the study of the algorithm recommended method, instantiate the framework for solving this kind of problem. Through a large number of experimental analysis and comparison, proved that the algorithm is feasible and reasonable.

##### References

1. Yiwen Zhong,Juan Lin,Jing Ning,Xiaoyu Lin. Hybrid Artificial Bee Colony Algorithm with Chemotaxis Behavior of Bacterial Foraging Optimization Algorithm[C].Seventh International Conference on Natural Computation,2011,2:1171-1174.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
2. G. Zhu and S. Kwong, "Gbest-guided artificial bee colony algorithm for numerical function optimization," Applied Mathematics and Computation, vol. 217, pp. 3166-3173, 2010.
3. K. V. Price, R. M. Storn, and J. A. Lampinen, Differential Evolution: A Practical Approach to Global Optimization. Berlin, Germany: Springer-Verlag, 2005.
4. Dervis Karaboga, Bahriye Basturk. “Artificial Bee Colony (ABC) Optimization Algorithm for solving constrained optimization problems,” in Foundation of Fuzzy Logic and Soft Computing, 12thInternational Fuzzy Systems Association, World Congress, IFSA 2007,P.Melin, Mexico: Springer, Lecture Notes in Artificial Intelligence Vol. 4529, June 2007, pp. 789-798.
5. S.N.Omkar, J.Senthilnath, Rahul Khandelwal, G.Narayana, S.Gopalakrishnan.Artificial Bee Colony (ABC) for multi-objective design optimization of compositestructures [J]. Applied Soft Computing, 2011, 11:489-499.