Clustering Analysis Based on Chaos Immune Algorith

[1]Fan Cui, [2]Zhen Guo, [2]Le Xu

1China University of Mining and Technology, Beijing State Key Laboratory of Coal Resources and Safe Mining, Beijing 100083, China
E-mail: cuifan@cumtb.edu.cn
2China University of Mining and Technology, Beijing College of Geoscience and Surveying Engineering, Beijing 100083, China

Corresponding author is Fan Cui

Abstract
In Chaos Immune Algorithm algorithm, the ergodic property of chaos phenomenon is used to optimize the initial population, so it can accelerate the convergence of Immune Algorithms. Chaotic systems are sensitive to initial condition system parameters. Through the clone selection operator, antibody circulation and supplement, Clone operator and excellent individual chaotic disturbance, local optimums were avoided, so the global optimization was obtained. As for this issue, chaotic immune algorithm, ALECO-2, BPMA and BP algorithm were applied to test the algorithm performance in two group experiments. Theory and experiment showed that the Chaos Immune Algorithm can get global optimum clustering center, and greatly improve the amplitude of operation.
Keywords: CHAOS, IMMUNE ALGORITHM, CLUSTER CLASSIFICATION

1. Introduction

Clustering Analysis is non-supervisory pattern recognition(Cheung, YM., 2003; Kanade, M. et al, 2007). Clustering is to group the data, and each group of data thus generated is termed as a cluster, and each data in a cluster is named as an object. The purpose of clustering is to make the objects from the same cluster resemble each other in characteristic as much as possible, while objects from different clusters differentiate each other in characteristic as much as possible. The task of clustering is to divide an unmarked mode into some subclasses according to certain rules. It is required that analogous samples be classified into the same group while non-analogous samples into different group, therefore, it is termed as non-supervisory classification. Currently, different methods of clustering have been applied into many fields like data mining, pattern recognition, image processing, Laser Radar target detection, and remote sensing technique (Hung, CC. et al, 2011; Tseng, VS. et al, 2005; Farnaz, F. et al, 2013; Cuncun, W., 2013).

Overview of cluster analysis as follows:

The clustering can be simply described as follows: classify n vectors $X_p\left(p=1,2,\cdots,n\right)$ into c categories $G_j\left(j=1,2,\cdots,c\right)$, and the clustering center of each category should be c_j. This pattern consists of the following steps.

(1) Select appropriately the initial center $G_1^0, G_2^0, \cdots, G_c^0$ of c categories.

(2) In k iterations, if we adjust any sample vector X_p, into any category of c categories, as for all $i \neq j, i, j \in [1,c]$, if $\left\|X_p - G_j^k\right\| < \left\|X_p - G_i^k\right\|$, then $X_p \in S_j^k$, of which, S_j^k is the category with G_j^k as the center.

(3) Recalculate the new center G_j^{k+1}: $G_j^{k+1} = \frac{1}{N_j}\sum_{X_p \in S_j^k} X_p$ of category S_j^k that we got from step 2, N_j in the formula is the number of samples in category S_j^k, the condition for ending the iteration is $i \neq j, j=1,2,\cdots,c$ $G_j^k = G_j^{k+1}$.

As is shown, the clustering is an issue of optimization of grouping in nature. And the selection of the initial center influences the result of clustering greatly. Optimization of grouping can be solved by immune algorithm, and chaotic optimization arithmetic operators are introduced into immune algorithm. The ergodic property of chaos phenomenon is used to optimize the quality of the initial population and improve the efficiency of calculation.

2. Research Method

The core of Chaos Immune Algorithm is based on the biologic immune mechanism reflected in mathematics the embodiment of biological immune mechanism in mathematics. Combining with random choice and determinacy Artificial Immune Algorithm is a heuristic random searching algorithm which has the ability to develop. The chaotic immune algorithm is indicated as in Figure 1.

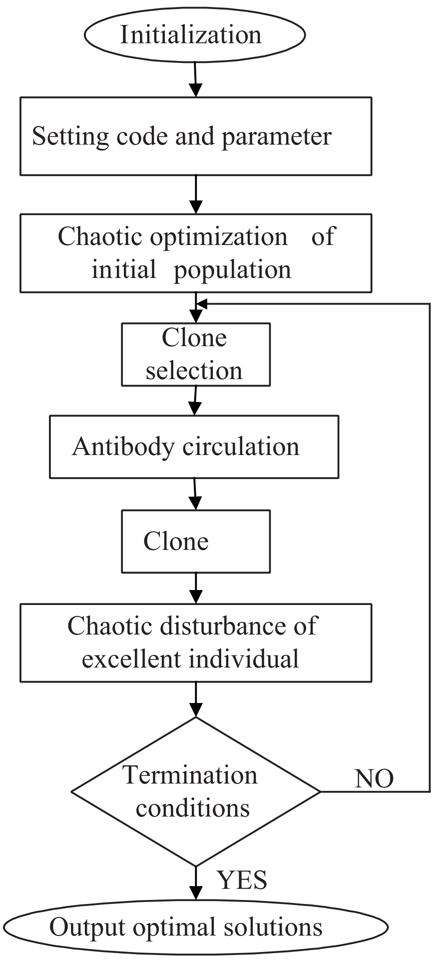

Figure 1. The process of Chaos Immune Algorithm

Major operators of Chaotic Immune Algorithm:

2.1. Chaotic Optimization of the Initial Population

Ag was defined initial antigen and Ab were defined initial antibody populations. M represented the scale of antibody populations.

$X_{\grave{u}\grave{u}\ddot{u}}=\{x_1, x_2, \cdots, x, \cdots, x\}$ represented an initial antibody, $m=1,2,\cdots,M$ $i=1,2,\cdots,n$. n was the dimension of variable X_m. Similar with genetic algorithm, x_{mi} was called allele, $x_{mi}\in[a_{mi}, b_{mi}]$. Antibody bit string was divided into l segments and every segment length was n_i, so, $n=\sum_{i=1}^{l} n_i$ represented total length of antibody gene segment. $Y(0)=(y_{0,1}, y_{0,2}, \cdots, y_{0,i})$ ($i=1,2,\cdots,n$) was defined initial antibody populations center .

After k iterations, the euclidean distance reciprocal between individual and antibody populations center was defined affinity function:

$$\frac{1}{aff(X_m)}=\left(\sum_{i=1}^{n}\left|x_{mi}-y_{ki}\right|^2\right)^{1/2} \quad (1)$$

$$m=1,2,\cdots,M \quad i=1,2,\cdots,n$$

Main operators as follows:

1) Chaos optimization of initial population

The initial Ab were optimized through chaos operator. According to equation

2) a chaotic-type initial value γ_{mi}^0 was produced.

$$\gamma_{mi}^0=(x_{mi}-a_{mi})/(b_{mi}-a_{mi}) \quad (2)$$

$$m=1,2,\cdots,M \quad i=1,2,\cdots,n$$

The logistic map was used to generate chaotic variable γ_i^k.

$$\gamma_i^{k+1}=\mu\gamma_i^k\left(1-\gamma_i^k\right) \quad (3)$$

$$i=1,2,\cdots,n \quad k=0,1,2,\cdots$$

Where μ was the control parameter, after determined the value of μ, with the arbitrary initial value $\gamma_i^0\in(0,1)$ (except 0.25, 0.5, 0.75 the fixed point of equation(3)), an assured time series $\gamma_i^1, \gamma_i^2, \cdots\gamma_i^k$ can be iterated (Alatas, B. et al, 2009).

Chaos optimization algorithm is mapping the chaos space to the solution space, using the inherent properties of chaotic variables to fulfill the overall search. According to equation (4), map the chaos variables from chaos space to the solution space.

$$x_{mi}=a_{mi}+(b_{mi}-a_{mi})\gamma_{mi}^k \quad (4)$$

$$m=1,2,\cdots,M \quad i=1,2,\cdots,n$$

Set x_{mi}^* as the optimal solution at the current phase of coarse-grained search, aff^* is the optimal objective function value for current phase. After each iteration, the individual affinity function $aff(k)$ was calculated, iteration termination condition was $aff(k)\le aff^*$. The antibody population after optimization was denoted by Ab'.

2.2. Clone Selection Operator

After the combination of antibody with antigen, antigen can be destroyed through a series of reactions which are based on antibody concentration. Antibody concentration $d(X_m)$ was defined by the following expression (Riccardo, C., 2003).

$$d(X_m) = \left[\frac{1}{M} \sum_{q=1}^{M} \frac{1}{1 + \sqrt{\sum_{i=1}^{n} (x_{mi} - x_{qi})}} \right]^{\alpha \cdot \left(1 - \frac{g}{G}\right)} \tag{5}$$

Where α was system parameter to adjust algorithm convergence speed, g was current evolution generation and G was the maximum evolution generation.

Affinity function $aff(X_m)$ was adjusted through the following expression. After the adjustment, $aff(X_m)$ was denoted by $\tilde{aff}(X_m)$.

$$\tilde{aff}(X_m) = aff(X_m) / d(X_m) \tag{6}$$

These antibodies which have bigger individual affinity value and the lower concentration can be promoted, on the contrary, those antibodies with smaller individual affinity value and the higher concentration can be inhibited, thus this process ensured the diversity of antibody group, so as to escape from local optimums.

2.3. Antibody Circulation and Supplement

The antibody circulation and supplement mechanism of biological immune system was simulated in order to ensure the diversity of antibody group and realize global search. Each time before antibody group Ab' were cloned, M_s highest affinity antibodies in a random generation antibody group Ab_r with population size of M_r replaced M_s lowest affinity antibodies in Ab'.

2.4. Clone Operator

According to $\tilde{aff}$, antibodies in Ab' were ordered by sort descending. Take top m_c antibodies to be cloned, the antibody group cloned was denoted by Ab_c. As the following equation, the population size of Ab_c can be calculated.

$$M_c = \sum_{j=1}^{m_c} round\left(\frac{\beta M}{j}\right) \tag{7}$$

Where M_c was the population size of Ab_c. β was proliferation coefficient to control antibody group size owing to its influence of algorithm iteration and calculating time. j was the sequence number of antibody by sort descending. $round(\cdot)$ was rounding operation.

2.5. Variation Operator

Set $\bar{X} = (\bar{x}_1, \cdots, \bar{x}_{i-1}, \bar{x}_i, \cdots, \bar{x}_n)$ as a parent entity, according to equation (8) to mutate, then after the Variation, the offspring individual was $\tilde{X}_m$, the antibody group mutated was denoted by Ab_m.

$$\tilde{X}_m = \bar{X}_m + \eta N(0,1) e^{-\hat{aff}} \tag{8}$$

After $(0,1)$ standardization, aff of $\bar{X}_m$ was denoted by $\hat{aff}$. $N(0,1)$ is normal random function with mean value $\mu = 0$ and variance $\sigma = 1$. The proportionality constant η can control attenuation of negative exponential function. According to equation (8), the bigger the antibody affinity value, the smaller variation, that was beneficial to maintain the stability of the local optimal solution.

2.6. Excellent Individual Chaotic Disturbance

After the clone selection, clone and variation, we get the current optimal solution is $X^* = (x_1^*, \cdots, x_{i-1}^*, x_i^*, \cdots, x_n^*)$, according to equation (9), make chaotic disturbances to x_i^*.

$x_i(k') = x_i^* + \varphi u_{k,i}$ §ùè ⋯ n (9)

In the equation, $x_i(k')$ is the chaotic variable of a smaller range of ergodicity relative to equation(9), φ is the adjustment coefficient related to the iteration number k', in this paper set $\varphi = 1 - \left(\frac{k'-1}{k'}\right)^n$. Continue to carry out iterative search with $x_i(k')$, the termination condition is ùùì(') ≤ *.

3. Results and Analysis

As for this issue, chaotic immune algorithm, ALECO-2, BPMA and BP algorithm were applied to test the algorithm performance in two group experiments.

3.1. Tset 1

Let's input a 4 that is in normal distribution into 2 to carry out the simulated experiment, there are 20 samples, the samples' vectors of average value is $[-1.5 \quad -1.5 \quad -1.5 \quad -1.5]$ and $[2.2 \quad 2.2 \quad 2.2 \quad 2.2]$, the sample's covariance matrix is $\begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$ and $\begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix}$, there is overlap among the samples (Leandro, DC. et al, 2002; CCarlotta, D. et al, 2007).

As for this issue, chaotic immune algorithm, ALECO-2, BPMA and BP algorithm were applied. On the basis of large number of simulated operations, we found out the optimum studying parameter for each algorithm. With the optimum parameter, the comparative curve of convergence of each algorithm is shown in Figure 2. Table 1 if the average value and standard balance of the times of iteration when converge after operating for 10 times at random. CIA is tested in another group with 20 samples for each patter, the rate of correct identification is 95.5%.

Table 1. The statistics of convergence of different algorthm

Algorithm	CIA	ALECO-2	BPMA	BP
Average Iteration Times	24.1	58.2	60.14	128.2
Standard Balance of Iteration Times	10.17	22.27	20.03	47.45

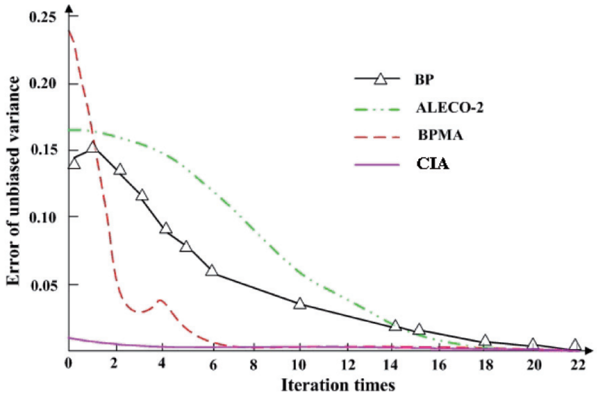

Figure 2. Comparative Curve of Convergence

3.2. Test 2

The marketing theory usually divides the market competitor into 4 types: (1) market leader; (2) market challenger; (3) market follower; (4) market filler. After the repeated research and comprehensive analysis of the 4 types of market competitors five characteristic factors can be extracted: 1) x_1 is the relative market share; 2) x_2 is the level of price changes for the enterprise; (3) x_3 is the capability for enterprise in new product development; 4) x_4 is the capability for enterprise distribution channels and physical distribution; 5) x_5 is the capability for a comprehensive marketing. Various characteristic factors and the corresponding grade of the actual meaning and values can be further divided, as shown in Table 2.

Selecting 280 of the garment industry manufacturers, the samples are randomly divided into the training sample set (140 samples) and test sample set (140 samples), four characteristic factor values are calculated in each sample. ALECO-2, BPMA and BP algorithm were applied to do a classified comparative test. The results are shown in Table 3.

4. Conclusions

Focusing on data clustering, this article tries to combine chaotic operator with immune algorithm. This chaotic immune algorithm integrates the advantages of both chaotic operator and clone operator. The outputs of theory and simulation both indicate that the chaotic immune algorithm can ensure the global optimum clustering and improve the amplitude of operation greatly.

Table 2. Featured Factors of Enterprise Competitive Position

	Featured factor	Grade	Real meaning	Corresponding value
x_1	Relative market share		The comparison of market share in the largest market competitors	0.1-10
x_2	The level of price change	1	High	9-10
		2	Medium	8-8.9
		3	General	6-7.9
		4	Weak	0-5.9
x_3	capability of new product development	1	High	9-10
		2	Medium	8-8.9
		3	General	6-7.9
		4	Weak	0-5.9
x_4	Capability of saling channels and physical distribution	1	High	9-10
		2	Medium	8-8.9
		3	General	6-7.9
		4	Weak	0-5.9
x_5	Capability of comprehensive marketing	1	High	9-10
		2	Medium	8-8.9
		3	General	6-7.9
		4	Weak	0-5.9

Table 3. Accuracy of different algorithms

Algorithm		Market leader	Market challenger	Market follower
BP	Iterations	90	2000	2200
	Correct rate	76.10%	72.38%	71.77%
BPMA	Iterations	82	1900	2100
	Correct rate	80.30%	80.42%	81.75%
ALECO-2	Iterations	77	1800	2040
	Correct rate	86.10%	82.38%	81.77%
CIA	Iterations	70	1788	2000
	Correct rate	87.10%	89.08%	88.98%
Algorithm		**Market filler**	**Average correct rate**	
BP	Iterations	200	200	
	Correct rate	73.34%	73.34%	
BPMA	Iterations	194	194	
	Correct rate	83.55%	83.55%	
ALECO-2	Iterations	183	183	
	Correct rate	83.34%	83.34%	
CIA	Iterations	175	175	
	Correct rate	88.75%	88.75%	

Acknowledgements

This work is supported by “Key Programe of National Natural Sience Fund” — “The Early Warning of Mine Roof Disaster”(51234005).

References

1. Alatas, B., Akin, E., Ozer, A.B., 2009. Chaos embedded particle swarmoptimization algorithms. Chaos Solitons Fractals. 40(4): 1715-1734.
2. CCarlotta, D., Dimitrios, G., Sheng, M., 2007. Locally adaptive metrics for clustering high dimensional data. Data Mining and Knowledge Discover. 14(1): 63-97.
3. Cheung, Y.M., 2003. k-Means: A new generalized k-means clustering algorithm. Pattern Recognition Letters. 24(15): 2883-2893.
4. Cuncun, W., 2013. Intelligent search technology combining semantic grid and Clustering. TELKOMNIKA Indonesian Journal of Electrical Engineering. 11(8): 4803-4809.
5. Farnaz, F., Ebrahim, P., Shahram, T., 2013. Retinal identification based on density clustering and fuzzy logic and connecting this information to electronic health record. International Journal of Electrical and Computer Engineering. 3(3): 359-365.
6. Hung, CC., Kulkarni, S., Kuo, B.C., 2011. A new weighted fuzzy c-means clustering algorithm for remotely sensed image classification. IEEE Journal of Selected Topics in Signal Processing. 5(3): 543-553.
7. Kanade, M., Hall, O., 2007. Fuzzy ants and clustering. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans. 37(5): 758-769.
8. Leandro, D.C., Fernando, V.Z., 2002. Learning and optimization using the clonal selection principle. IEEE Transactions on Evolutionary Computation. 6(3): 239-251.
9. Riccardo, C., 2003. Chaotic sequences to improve the performance of evolutionary algorithms. IEEE Transactions on Evolutionary Computation. 7(3):289-304.
10. Tseng, V.S., Kao, C.P., 2005. Efficiently mining gene expression data via a novel parameterless clustering method. IEEE/ACM Trans. on Computational Biology and Bioinformatics. 2(4): 355-365.

Optimization of Microgrid Dispatch by Using Co-evolutionary Method

[1]Chen Juan, [1]Huang Yuan-sheng, [2]Lu Bin

1Department of Economic Management, North China Electric Power University, Baoding, China
2Department of Computer Science & Technology, North China Electric Power University, Baoding, China

Corresponding author is Chen Juan

Abstract

The drawbacks of high cost, high pollution and high energy consumption of traditional power system become increasingly apparent, so the energy crisis and energy conservation is an urgent requirement to promote new energy generation and renewable energy generation such distributed generations to become a useful supplement for