Novel Clonal Selection Algorithm Improving Selection Operator

Tao Gong[1, 2, 3], Long Li[1], Changsheng Guo[1] and Xiaolei Gong[1]

[1] *College of Information S. & T., Donghua University, Shanghai 201620, China*

[2] *Engineering Research Center of Digitized Textile & Fashion Technology, Ministry of Education, Donghua University, Shanghai 201620, China*

[3] *Department of Computer Science, Purdue University, West Lafayette 47907, USA*
E-mail: taogong@dhu.edu.cn

Abstract

To optimize the search of the clonal selection algorithm for the optimal solution, the selection operator is improved and a novel clonal selection algorithm is designed. The selection operator is compared with such selection operators as Roulette Wheel Selection operator, Tournament Selection operator, and Sampling Selection operator. The selection operator is improved to increase the accuracy and efficiency for finding the optimal solutions. The experimental results show that the novel clonal selection algorithm performs better than the traditional clonal selection algorithm due to the adjustment of the selection operator.

Keywords: *Clonal selection; roulette wheel selection; tournament selection; sampling selection; optimal solutions.*

1. Introduction

As we know, there are some analogous points between immune algorithm (GA) and clonal selection algorithm (CSA), though they are substantially different. In recent decades, some researchers have attempted to integrate the two algorithms as new hybrid algorithms, which will perform better than each single algorithm. For example, Licheng Jiao et al. added immune methods into the GA and designed an immune genetic algorithm [1], which mainly includes two steps, i.e. vaccination and immune selection. Antariksha Bhaduri used GAIN, a hybrid mathematic algorithm based on the GA and Artificial Immune Network, for optimizing university time table scheduling problem [2]. Selection operator is an essential of genetic algorithm, which determines the performances such as convergence, feasibility and efficiency of the whole algorithm. Apparently, it occupies position among the series of steps researchers concern mostly of immune algorithm. In this paper, selection operators usually employed in genetic algorithm and based on fitness are transplanted to the CSA after adjusted and improved to lead the CSA to perform more astringently and multifariously. Firstly, the clonal selection algorithm and the importance of selection operators are discussed. Then, some usual genetic selection operators are introduced and improved edition provided and embedded into the CSA. And in the end, simulation experiments suggest that the improved selection operators can be used the CSA.

2. Clonal Selection Algorithm and Selection Operator

Both the genetic algorithm and the clonal selection algorithm are originated from some biological mechanisms in nature. Leandro Nunes de Castro proposed the clonal selection

algorithm in 2000 [4]. From then on, more and more researchers are interested in the CSA, and thousands of papers are published.

The clonal selection algorithm mainly constitutes of selection operation, clonal operation, mutation operation, memory operation and several other steps. From Figure 1, basic flow of the clonal selection algorithm is presented. Fitness calculation means that fitness between antigen and every antibody of the population is calculated to measure similarity/affinity during antigen-antibody interactions. It can be described by a function mathematically. For example, in optimization problem the objective function value of the related antibody is commonly used, or in handwriting character recognition an equation contributes to the similarity between feature matrices, normalized, of binary images [5]. Selection operation will save the antibodies of which fitness supers and others would be discarded. After selection operation, clonal operation makes antibody selected is cloned based on the size of the fitness, which is called clonal operation. In the mutation operation, mutation ratio is proportional to the value of fitness. Greater antibody fitness determines smaller the mutation rate, and vice versa. In order to improve the local search ability, the ratio tends to a decreasing trend according to the number of iterations.

The algorithm involves two selection operations: the first happens till the cloning and selection operation and the second happens till the excellent antibody is memorized. Thus, selection operation has great significance for the algorithm.

3. Design of Selection Operators

In the clonal selection algorithm, the selection operation is always based on the fitness of antigen and antibody. So the following discusses Genetic Related selection operators that use the information concerning the genetic relatedness of individuals are proposed [6].

3.1. Roulette Wheel Selection

Roulette Wheel Selection [7], one of the earliest selection operators, so simple and practical that widely adapted, is raised by John Holland, the creator of genetic algorithm. To a great extent, it is on the basis of the ratio, fitness account for in the population and stochastic which takes more than the former. The principle can be described as:

$$P_i = \frac{F_i}{\sum_j^N F_j} \tag{1}$$

Here, N is the total number of population, i is the label of the ith individual, F_i is the fitness of the ith individual, P_i is the stay ratio of the ith individual.

It is apparent from the equation (1) that the greater the fitness of individual, the ratio left will be bigger. And the equation selects only one individual once, so iterated to select multiple demanded. In the case of allowing individual select repeatedly, one of the computing process is to use P_1, P_1+P_2, …, $P_1+P_2+\cdots+P_i$, …, $P_1+P_2+\cdots+P_{N-1}$ to divide [0, 1] into N intervals after calculating $\{P_i\}$; then to produce a stochastic number p between [0, 1]. When $P_1+P_2+\cdots+P_{i-1} < p \le P_1+P_2+\cdots+P_i$ is given, the ith individual will be selected.

3.2. Tournament Selection

Most tournament selection [8] is stochastic, divides the population into several subpopulation, just as tournament has several stadiums. Fitness comparison processes among individuals in the subpopulation. Elect the most high-fitness individual or some

most high-fitness individuals, preferred to be stayed. Though there is strong stochastic in the selection operation, still it can ensure high-fitness individuals selected in a great probability and the lowest fitness individual condemned. Provided the number of population is *N*, tournament selection can be: (1) divide the population into m subpopulations, so every subpopulation has *N*/*m* (rounded, commonly set to be 2); (2) compare and sort fitness in each subpopulation separately, select the top *n*.

3.3. Sampling Selection

There are two types of sampling selection [9]: one is in a determined way, and the other performs stochastically. Determined sampling selection guarantees high-fitness individuals preserved and low discarded. Here is:

(1) According to fitness, calculate number of offspring:

$$n_i = N \cdot \frac{F_i}{\sum_j^N F_j} \tag{2}$$

(2) Round down n_i, so achieve the number of ith individual offspring $n_i^{'}$; then set the offspring individuals, whose sum is $N^{'} = N - (n_1^{'} + \cdots + n_i^{'} + \cdots + n_N^{'})$.

(3) According to the size of decimal part of n_i, sort the population, preserve the former $N^{'}$ individuals.

The other, Remainder Stochastic Sampling with Replacement, the improved edition of the deterministic sampling selection, defers from the above on the third step. In this selection operator, the third step is replaced by the other stochastic selection, for example, Roulette Wheel Selection, and is determined by decimal part of n_i.

Based on the above three selection operators, the clonal selection algorithm (CSA) is designed as below.

```
Start algorithm
  Init antibody population and antigen
  FOR:
    Fitness calculation
    First selection
    Clone operation
    Mutation operation
    Fitness calculation
    Second selection
    Remember operation
    Mutation ratio adjustment
    IF: Converge judgment == Yes
        Go to End of the algorithm
    ELSE
        Add new population
        Go to the FOR beginning
    END
  END
End of the algorithm
```

Figure 1. Clonal Selection Algorithm based on the Selection Operators

4. Improvement of Clonal Selection Algorithm

In general, the selection operation of genetic algorithm have large randomness, it is obviously not suitable for clone selection algorithm if employed directly. In the progress of selection operation of roulette wheel selection and tournament selection, stochastic operation ranges the entire population commonly, which means great randomness. As shown in Figure 2, distribution of fitness between random antibodies produced by computer and antigen, it can be revealed that in random selection, those low-fitness antibodies that will be selected to selection operation have a great impact to offspring.

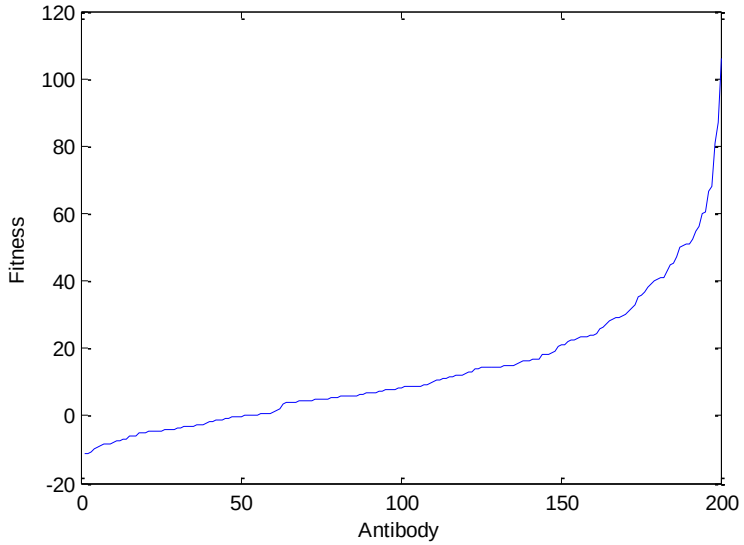

Figure 2. Distribution of Fitness between Random Antibodies and Antigen

It is these considerations that the improved operation can be that the selection range in antibodies with higher fitness, not only ensuring less low-fitness impact, but also bringing certain randomness. One of the solutions is to provide a parameter r_1, which is used to control the individual ratio of antibodies join in selection operation. Another parameter r_2 means the selection ratio in roulette wheel selection and the number of sub-groups in tournament selection.

The randomness of sampling selection mainly related with the decimal part of the number of offspring, the integer part directly represents the degree of fitness, larger fitness then more probability selected. Yet, fundamentally, the selection probability of an individual has a great relationship with the distribution of fitness. If the distribution of affinity in population is uniform, In other words, the variance is small, difference between selection probability of individuals would be small and randomness larger; If the distribution variance larger, high-fitness individuals would more likely be selected, or definitely. In order to adjust or fix the randomness of sampling selection, the proportion accounted for the fitness of the total population is divided into parameter S_n parts. And the length of a part replaces the unit '1' to calculate the numbers of individuals. Adjust equation (1) to the following formula:

$$n_i = N \cdot \frac{F_i}{\sum_j^N F_j} \cdot \frac{1}{\frac{Integer(\max(F) - \min(F))}{S_n}} \tag{3}$$

5. Experiments and Analysis

To validate the feasibleness of applying the three strategies in the clonal selection algorithm, function 1,which has several peaks ,one of them is the highest and one of them is close to the highest, is employed to do the test. Function 1 has two regions. One is called Highest Region which is a locate region of the highest peak, and the other called Sub-highest Region, also a locate region but of the sub highest peak. Because the stochastic element added into clonal selection algorithm should be proper, just replace the first selection in the algorithm to the top three strategies. Some parameters are set as below:

Population Number 200
Code Method 2-bit
Mutation Ratio 0.01
Generation 100
Number of the 2nd Selection 10

Normal Selection parameter: ***Number of the 1st Selection*** : 10.

Roulette Wheel Selection parameter: r_1 : 0.2, r_2 : 0.5

Tournament Selection parameter: r_1 : 0.2, r_2 : 0.5

Stochastic or Determined Sampling Selection parameter: s_n : 5

The region parameter of Highest Region is set as the rectangle [9.1 8.8 10 10], and the parameter of Sub-highest Region is set as the rectangle [9 3.4 10 6.7].

The fitness function is calculated as below:

$$f(x,y) = (0.5 + \frac{\sin(\sqrt{x^2+y^2})^2 - 0.5}{1+0.001\cdot(x^2+y^2)})\cdot x^2 + x + y \quad (4)$$

100 tests were processed applying the above strategies. In every test, we got a best fitness. Table 1 shows the statistics data on the best fitness value, the worst fitness value, number of individuals located in the Highest Region, number of individuals located in the Sub-highest Region, the mean value of fitness and the variance of fitness. From the result data, we can confirm that the new clonal selection algorithms of the three selection operators are better than that of the traditional selection operator.

Table 1. Experimental Results of Different Selection Operators in the CSA

Strategy	Experimental results					
	Best solution	***Worst solution***	***Right Region***	***Wrong Regoin***	***Mean Fitness***	***Variance***
Traditional selection operator	111.5988	107.2897	63	37	109.2938	1.2448
Roulette selection operator	**111.6650**	**109.0063**	**29**	**8**	**111.0783**	**1.0591**
Tournament selection operator	**111.6650**	**108.0164**	**67**	**14**	**110.7421**	**1.3136**
Det Sampling selection operator	**111.6650**	**106.8363**	**55**	**5**	**111.2906**	**0.8050**
Sto Sampling selection operator	**111.6650**	**106.9761**	**48**	**23**	**110.3836**	**1.9359**

6. Conclusion

This paper applied some selection operators into the clonal selection algorithm to improve the performances such as the best solution, the worst solution and the fitness. Experimental simulations show that the new clonal selection algorithms of the Roulette selection operator, the Tournament selection operator and the Sampling selection operator perform better than that of the traditional selection operator. This improvement

will result in better security optimization such as learning of unknown viruses of some computer systems and networks.

Acknowledgements

The work was supported in part by grants from Natural Science Foundation of Shanghai (08ZR1400400), the Shanghai Educational Development Foundation (2007CG42), the National Natural Science Foundation of China (60874113), NSF-0242840, and NSF-0219110.

References

[1] O. Sukwong, H. S. Kim and J. C. Hoe, "Commercial antivirus software effectiveness: an empirical study", IEEE Computer, 44(3): 63-70 **(2011)**.

[2] L. Jiao and L. Wang, "A Novel Genetic Algorithm Based on Immunity", IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 30(5): 552-561 **(2000)**.

[3] A. Bhaduri, "University Time Table Scheduling using Genetic Artificial Immune Network", In: Proceedings of 2009 International Conference on Advances in Recent Technologies in Communication and Computing, **(2009)** pp. 289-292.

[4] L. N. De Castro and F. J. Von Zuben, "The Clonal Selection Algorithm with Engineering Applications", In: Proceedings of GECCO 2000, **(2000)** pp. 36-37.

[5] U. Garain, M. P. Chakraborty and D. Dasgupta, "Recognition of Handwritten Indic Script Using Clonal Selection Algorithm", In: Proceedings of the 5th international conference on Artificial Immune Systems, **(2006)** pp. 256-266.

[6] A. Gog, D. Dumitrescu and B. Hirsbrunner, "New Selection Operators based on Genetic Relatedness for Evolutionary Algorithms", Evolutionary Computation. CEC 2007, **(2007)** pp. 4610-4614.

[7] D. B. Fogel, "An Introduction to Simulated Evolutionary Optimization", IEEE Transactions on Neural Networks, 5(1): 3-14 **(1994)**

[8] T. Blickle and L. Thiele, "A Mathematical Analysis of Tournament Selection", In: L. Eshelman, ed., Genetic Algorithms Proceedings of the 6th International Conference (ICGA1995), **(1995)** pp. 9-16.

[9] P. Yang and J. H. Zheng, "Comparison and research over genetic selection operators", Computer Engineering and Applications, 43(15): 59- 62 **(2007)**.

Authors

Tao Gong

Dr. Gong received the MS degree in Pattern Recognition and Intelligent Systems and PhD degree in Computer Science from the Central South University respectively in 2003 and 2007.He is an associate professor of computer sciences at Donghua University, China, and he was a visiting scholar at Department of Computer Science and CERIAS, Purdue University, USA. He is the General Editors-in-Chief of the first leading journal Immune Computation in its field, and an editorial board member of some international journals such as Journal of Computers in Mathematics and Science Teaching, International Journal of Security and Its Applications, and International Journal of Multimedia and Ubiquitous Engineering. He is a Technical Committee Chair of ISEEIP 2012, and a Publicity Chair of ISA 2012. He is also a program committee member of some international conferences such as IEEE ICNC 2011, IEEE BMEI 2011, WMSE 2011, ICARIS 2012, IST 2012, ISA 2012 and SIS2013 etc. He is a Life Member of Sigma Xi, The Scientific Research Society, a Vice-Chair of IEEE Computer Society Task Force on Artificial Immune Systems, and Chen Guang Scholar of Shanghai.