A New Genetic Algorithm considering diversity of gene locus

Jun Liu [1, a], Jie Fu [2, b] and Xue Bai [2, c]

[1]School of Lanzhou University of Technology, Lanzhou 730050, China;

[2]School of Lanzhou University of Technology, Lanzhou 730050, China.

[a]lzhjliu@163.com, [b]jqd890222@163.com

Keywords: Genetic algorithm, prematurity, diversity of gene locus, mutation operator

Abstract. a new genetic algorithm is proposed in the paper. Different from other genetic algorithms, the diversity of gene locus is considered and a new mutation operator is proposed. The operator can suppress the survival rate of modes which tends to prematurity, and effectively ensure the diversity of populations. When the diversity of populations is abundant, the operator improves the survival rate of the dominant mode so that it accelerates the convergence of the algorithm. When the diversity of populations is barren, it can reduce the survival rate of dominant mode and increases the escaping rate from the local. The effectiveness of the operator as well as the convergence of the algorithm is proved in a theoretical way. Meanwhile, the validity of the new algorithm is testified by simulation experiments.

Introduction

Genetic algorithm (GA) is a kind of optimization algorithm which search for optimal solutions by simulating evolution of natural creatures and was put forward by an American professor, Holland, in 1975. Namely, it is also a kind of computational model which simulates Darwin's genetic choice and evolutionary selection [1]. The GA has been widely used in various fields, because it does not have continuous and differentiable conditions for optimization object function, and it has strong robustness and adaptability. In recent years, many scholars have done intensive studies on the proof of convergence problem of the algorithm, the premature phenomenon [2, 3], etc. Especially, many literatures have developed techniques to overcome the prematurity of the algorithm and have gotten improved performance of the algorithm finally [2, 4-8]. However, these researches usually focus on improving the technique by discussing the working mechanism of GA itself, so that the genetic individual or genetic parameters are treated as studying object and the impact incurred by the diversity of gene locus on the survival rate of dominant mode are usually overlooked, which bring disadvantages to the convergence efficiency while suppressing the premature of the algorithm. Aiming at handling the problem and different from other studies, a new GA in which an adaptive mutation operator concerning the diversity of gene locus is designed, is proposed in the paper. The operator can suppress the survival rate of modes which tends to prematurity, and effectively ensure the diversity of populations. When the diversity of populations is abundant, the new operator can improve the survival rate of the dominant mode so that it accelerates the convergence of the algorithm. When the diversity of populations is barren, it can reduce the survival rate of dominant mode and increases the rate of escaping from the local. The convergence of the algorithm is proved. Meanwhile, the validity and versatility of the new operator is testified by simulation experiments on a standard testing function model and a parameter optimization model of the support vector machine.

A new genetic algorithm Section Headings.

The diversity measurement of gene locus .

The key to improving the direction of mutation operator is to allocate suitable mutation probability for different gene locus. Therefore, the diversity measurement of gene locus should be established firstly.

Let P, N, p_m and $x_i, i = 1, 2, ..., N$ denote the population, population density, base mutation probability and a individual respectively. $x_{ij}, (j = 1, 2, ..., L)$ denote the $j-$th gene of the individual x_i. L denote the length of chromosome. Let $p_{mj}, (j = 1, 2, ..., L)$ denote the mutation probability of the $j-$th gene. Let S denote a arbitrary mode and its order is denoted by $o(S) = k$, then the gene locus determined by S is denoted by $l_i, (i = 1, 2, ... k)$. The measurement of the diversity of the $j-$th gene is defined as follows:

$$v_j = H(\frac{1}{N} \sum_{i=1}^{N} x_{ij}) \qquad (1)$$

Where $H(y) = 4y(1-y)$, y denotes a independent variable. Then the population diversity is defined as follows:

$$V = \frac{1}{L} \sum_{j=1}^{L} v_j \qquad (2)$$

Then the measurement of diversity of mode S is denoted by $V(S) = \frac{1}{k} \sum_{i=1}^{k} v_{l_i}$.

It is apparent that $v_j \in [0,1]$. When $v_j = 0$, the diversity of the $j-$th gene is lowest, which means the values of all the $j-$th genes are equal. When $v_j = 1$, the diversity of the $j-$th gene is highest, which means the $j-$th genes are evenly split to two groups with different values.

A new GA considering diversity of gene locus. Two mutation operators a and b are introduced to the GA.

Operator a: reverse the value of the $j-$th genes of the individual $x_i, i = 1, 2, ..., N$ by the mutation probability $p_{mj} = \beta v_j p_m + \gamma p_m$, where β and γ are two self-defined parameters.

Operator b: reverse the value of the $j-$th genes of the individual $x_i, i = 1, 2, ..., N$ by the mutation probability $p_{mj} = \beta(1 - v_j) p_m + \gamma p_m$.

Then the following is the new GA: (1) Initial populations. (2)Crossover operation and selection operation are carried out on populations.(3)Count the current population diversity V, if $V \leq \theta$, the mutation operator is b, if $V > \theta$,the mutation operator is a (θ is a threshold value). (4)Mutation operation is carried out on every gene by utilizing the selected mutation operator. (5) Whether the termination condition of algorithm is satisfied, if not, return to step (2), otherwise output analysis result as the optimal solution for the current population. (6)End the algorithm. Different from other GA, the diversity of gene locus is considered in the GA. We improve the direction of mutation operation by allocating suitable mutation probability to different gene locus. Note that the selection operator and crossover operator here are traditional ones, and we want to see what there are any differences when the diversity of gene locus is considered in the GA. By the way, the strategy of elite individual to be preserved is adopted here.

Analysis

Theorem 1: if the mutation operator a is adopted, for the superiority mode S, the following exists: $p\{S \in M[P(t)] \| S \in P(t)\} \geq 1 - \{1 - [1 - 4\beta p_m \frac{n_s}{N}(1 - \frac{n_s}{N}) - rp_m]^k\}^{n_s}$ .Where $n_s \in (\frac{N}{2}, N)$ denote the quantity of the superiority mode S in the population $P(t)$. M denotes mutation operation.

Theorem 2: the new GA is of global convergence.

Proof: let MC denote the crossover and mutation operation. The individual x_i' is accessible from x_i by crossover and mutation operation if $p\{MC(x_i) = x_i'\} > 0$ exists[10]. We firstly give a Lemma here:

Lemma [2,5]: If an evolutionary algorithm conforms to the following two conditions, then the algorithm will converge to the global optimization solution with probability one:

Condition 1: for any two points x_i and x_i' in the feasible domain, x_i' is accessible from x_i, i.e., $p\{MC(x_i) = x_i'\} > 0$.

Condition 2: the population sequence of $P(1), P(2), \cdots, P(t)$ is monotonic under condition that one of individuals in $P(t+1)$ is not inferior to any individuals in $P(t)$ for $\forall t$.

Numerical experiments

Firstly, the following standard test function is adopted to test the performance of the new operator and the GA:

$$f(x, y) = 0.5 - \frac{\sin^2(x^2 + y^2)^{0.5} - 0.5}{1 + 0.001(x^2 + y^2)}, \quad x, y \in [-10, 10].$$

The global optimal solution of the above function is $f(0,0) = 1$. The test function is as shown as in Fig.1.
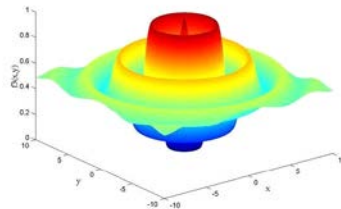

Fig.1

It is apparent that the test function has many local optimal solutions as shown in Fig.2., meanwhile, the global optimal solution is surrounded by the local optimal solutions. So it is easy to get into prematurity. Let $N = 50$, $P_m = 0.001$, $\beta = 3$, $\gamma = 0.1$. The probability of the crossover operation is $P_c = 0.72$. The algorithm will stop after 120 generations. The numerical comparisons of basic GA(B_GA), adaptive GA(A_GA) and the new GA(N_GA) in the paper are shown in Table 1. Note that $P_m = 0.01$ for the later stage of A_GA.

Table 1 experimental results

	B_GA	A_GA	N_GA
Optimum	0.9697	0.9912	0.9967
Times of prematurity	37	8	3

It can be drawn that the more accurate solution is gotten by the new GA and the new GA gets the least prematurity times. Fig.2, Fig. 3 and Fig.4 show the distribution of individuals gotten by B_GA, A_GA and N_GA respectively. It is clear to show that that the new operators of the paper perform well.
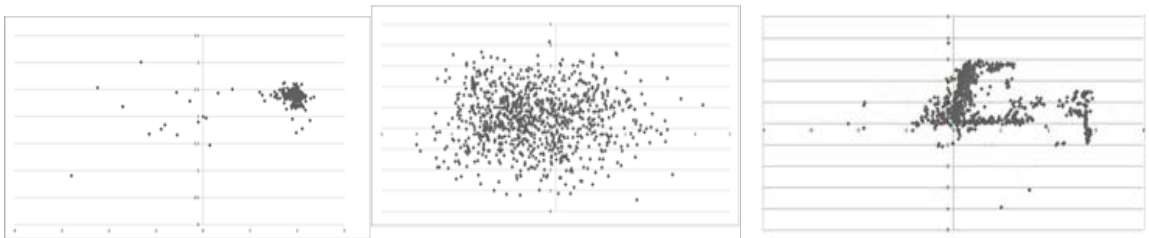

Fig. 2 Two or more references   Fig.3 Individuals gotten by A_GA Fig.4 Individuals gotten by N_GA

Fig.2 shows that the B_GA falls into local optimum and almost all of the individuals are in a small area, which means the B_GA gets into prematurity shortly so as to do harm to the performance of the algorithm. In Fig. 3, although the A_GA compared with the B_GA gets a larger searching area and there is no obvious tendency of prematurity, the distribution of the individuals is almost random. That is to say, the evolutionary information is not utilized so that the searching direction of the algorithm is fully random, which is also disadvantageous to improving the performance of the algorithm. The Fig. 4 shows that the searching of the N_GA covers a large area and the search areas are continuous from a local area to anther nearby. So the N_GA can keep a large search scope, meanwhile, especially strengthen the searching directivity by fully utilized the genetic information. This balance the searching area and convergence efficiency of the algorithm, namely, the N_GA avoids the prematurity of algorithm and has considerable convergence efficiency.

Conclusions

The diversity of gene locus is considered in the paper and a GA, in which a new mutation operator is adopted, is proposed. When the diversity of populations is abundant, the operator improves the survival rate of the dominant mode so that it accelerates the convergence of the algorithm. When the diversity of populations is barren, it can reduce the survival rate of dominant mode and increases the escaping rate from the local. The theoretical analysis and numerical experiments show that the algorithm perform well. Combining with other techniques, the mutation operator can be widely used to improve the performance of other genetic algorithms.

References

[1] Holland,J.H.Adaption in natural and artificial systems[M].The University of Michigan Press,1975.

 [2] ZHUANG J,WANG S A.Study on Self-adjusting Genetic Algorithm[J]. Journal of System Simulation. 2003,15(2):281-286

[3] CHEN G L et al. Genetic Algorithm and Its Application[M].Beijing:The People's Posts and Telecommunication Press,1996

[4] XU K, LEI Y C,ZHU H Q. Research on Adaptive Genetic Algorithm with Tabu Search[J]. Industrial Control Computer. 2011,24(3):55-57

[5] ZHANG L, LU Y L, YANG G Z et al. Parallel Quantum Genetic Algorithm Using Sphere Multi-regionDivision[J]. 2011,33(5):1035-1041

 [6] YE Y L,SAN Y. Hybrid Optimization algorithm and its performance[J]. Journal of Jilin University(Engineering And Technology Edition).2009,39(1):131-136)

[7] ZHANG J Z,JIANG T. Improved adaptive genetic algorithm[J]. Computer Engineering And Applications. 2010,46(11):53-55

[8] ZHANG L,ZHANG B. Research on the Mechanism of Genetic Algorithms[J].Journal of Software.2000,11(7):945-952

[9] ZHANG W X,LIANG Y. Foundations of Mathematics for the Genetic Algorithm[M]. Xi'an : Xi'an Jiaotong Unversity Press,2001:54-79

[10] Back T. Evolutionary algorithms in theory and practice[M]. New York: Oxford University Press，1996:21-28

[11] V Vapnik. The nature of statistical learning theory[M].New York:Springer-Verlag,200.

[12] WANG Z C, ZHANG H X. Application and research on regulation of parameters in support vector regression[J].Computer Engineering andDesign.2011,32,(8):2821-2824

[13] A D Sanchez. Advanced support vector machines and kernel methods[J]．Neuro Computering，2003,55(1)：5-20．