Vol. 11 No. 8 PROGRESS IN NATURAL SCIENCE August 2001

A new non-monotone fitness scaling for genetic algorithm *

LI Minqiang (李敏强)** and KOU Jisong (寇纪淞)

Institute of Systems Engineering, Tianjin University, Tianjin 300072, China

Received November 13, 2000; revised December 22, 2000

Abstract The properties of selection operators in the genetic algorithm (GA) are studied in detail. It is indicated that the selection of operations is significant for both improving the general fitness of a population and leading to the schema deceptiveness. The stochastic searching characteristics of GA are compared with those of heuristic methods. The influence of selection operators on the GA's exploration and exploitation is discussed, and the performance of selection operators is evaluated with the premature convergence of the GA taken as an example based on One-Max function. In order to overcome the schema deceptiveness of the GA, a new type of fitness scaling, non-monotone scaling, is advanced to enhance the evolutionary ability of a population. The effectiveness of the new scaling method is tested by a trap function and a needle-in-haystack (NiH) function.

Keywords: genetic algorithms, selection operator, non-monotone fitness scaling.

The genetic algorithm (GA) is a new type of global optimization method by simulating the process of the natural selection and the structure of heredity[1~3]. First, the feasible solution space of the original optimization problems is transformed into a finite discrete space of binary-coded strings (called chromosomes), in which each string (called individual) is corresponding to a feasible solution for the optimization problem. Then the GA carries out a global searching, with a population of individuals, in the binary-coded space by adopting three operators: selection, crossover and mutation. The selection operation mimics the competition of natural organisms in choosing individuals from the current population to forming a candidates' pool, in which the fitter ones with larger fitness hold a higher survival probability, and their hereditary information is inherited by the individuals in the next generation of the population. The crossover and mutation operations are used to exercise changes in candidates by simulating the natural sexual hybridization and variation of organisms, in order to form a new population with more fitter individuals. In the process of evolution, the fitness function constitutes a feedback index of the environment to the population of individuals.

The GA is characterized with a global searching ability, implicit parallelism, robustness and scalability. Its power in solving complex optimization problems comes from the evolution process of a population of feasible solutions in the form of binary-coded representation by using simple genetic operators, where the selection operator is the crucial one for controlling GA's searching performance.

With the selection operations, the GA can implement the process of the natural selection of fittest individuals, and ensure that fine genetic materials, called schema existing among binary-coded

* Project supported by the National Natural Science Foundation of China (Grant No. 69974026).

** Corresponding author, E-mail: mqli@tju.edu.cn

strings, are transferred to new individuals. In solving optimization problems, it is the selection operator that controls the searching direction and trajectory of the GA, so the improper design and parameter settings may lead the GA away from converging to the global optima[3~5]. Therefore, this paper will analyze the relation of the selection operation and the GA searching capability, and the phenomena of population premature and schema deceptiveness. Then, we put forward a new method of selection using non-monotone fitness scaling, which is applied to typical optimization problems.

1 Selection operation and GA performance

In fact, the population-based evolution of the GA for solving optimization problems is a typical stochastic searching process, in which the GA should carry out both the exploration in the definite space for global optimality, and the exploitation in the current optima neighborhood for local optimality, which we call, respectively, the reforming and refining capabilities of the GA. With the exploration, the GA aims to overcome the local optima trap and schema deceptiveness, so as to make global searching on the total binary-coded space. With the exploitation, the GA would refine the current optima in a small neighborhood of the definite space, so as to maintain the evolutionary ability of the population for climbing toward the local optima by inheriting fine schema in individuals or genes in chromosomes. It is hard to endow the GA with the two capabilities simultaneously at any generation in its evolution process.

Since the selection operation guides the searching direction of the GA, it is also the most important factor affecting the searching capabilities of reforming and refining. Let us take as an example the one-dimensional optimization function: $\max f(x)$ $(x \in [a, b])$. The searching sphere of stochastic-search algorithms at any time is the total definition space $[a, b]$, and that of the traditional local-search algorithms is a small neighborhood containing the current best solution x_0 with a radius of σ $(\sigma \ll b - a)$. In comparison (see Fig. 1), the GA exercises a search in the probability distribution on a much larger neighborhood around solution x_0, and the size of neighborhood and the probability distribution is controlled by the selection operation.

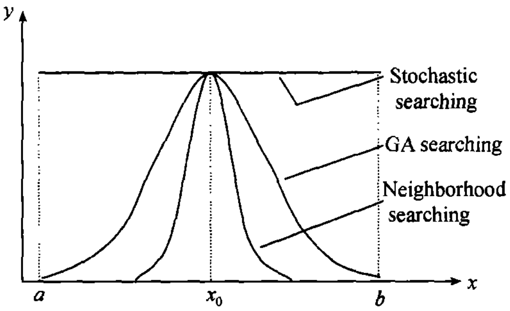

Fig. 1 Comparison of three searching methods.

With specified parameter settings, the GA displays different searching performances. In the initial generations, the population is in its largest diversity, and the GA can get the information of all the definite space, or can get to any point in the binary-coded space if errors of the random sampling are not taken into consideration. In the direction of the selection processing (the selection pressure is larger than 0), the diversity of the population goes down monotonously along with the evolution process, the speed of which is decided by the selection pressure, and the larger the pressure, the quicker the descending speed. In the interim phase of generations, the searching activities of the GA are gradually constrained in a smaller and smaller subspace of definition, and its exploration process is then changed into exploitative searching.

If the selection operation is not designed properly, the exploration process would terminate rather early, and the GA can hardly find the global optimal neighborhood. On the other hand, if it is too late to switch to the exploitation process, the GA would have no time to do refining for the current best solution. So the selection operator has been, up to now, an important research topic of GA[5,6].

2 Selection operation and population premature

In the practical applications of GA as a searching method for solving optimization problems, it is often found that the population does not converge to the global optima, which is especially true for a multi-modal function optimization. In the binary-code space, there still exists a marked Hamming distance between the best strings in the current population and the global optimal string, but the solutions represented by the current best strings can hardly be improved, which is called the premature convergence of the population[7,8].

The premature convergence of the population is characterized by a remarkable similarity of individuals, low diversity of the population, and lack of effective alleles (the genes contained in the global optima string) in individual strings[7,8]. With the three genetic operators, the GA cannot combine lower order schemata into higher order ones. This phenomenon is related to the properties of the optimization function, population size, schema deceptiveness, etc., and the top factor is the selection operation. The diversity of the population drops down quickly with a high selection pressure. On the other hand, when the selection pressure is low, the schema competition is weakened. The above two cases will results in a great decrease of the probability of the higher order schemata recombination, and the evolution of the population will stagnate at interim states.

Now, One-Max optimization function is taken as an instance to show the relation between the selection operation and the premature convergence of the GA. The One-Max optimization is a special binary bit unitation function[5], which is the sum of "1" alleles in the bit string. It is a GA-easy optimization problem, and takes the general form

$$f(\boldsymbol{a}) = \sum_{l=1}^{L} a_l, \tag{1}$$

where $\boldsymbol{a} = (a_1, a_2, \cdots, a_L)$, $a_l \in \{0,1\}$ $(l = 1,2,\cdots,L)$, with its global optimum string being $\boldsymbol{a}^* = (1,1,\cdots,1)$, and the optimal function value $f(\boldsymbol{a}^*) = L$.

Let us compare the results of the GA in solving One-Max optimization function under different selection pressures. The parameters and genetic strategies setting of the GA include the string length $L = 120$, population size $n = 120$; the ranking selection and elitist retaining with $\eta^+ = \{1.1, 1.6, 2.0\}$, two-points crossover with probability $p_c = 0.6$, bit mutation with probability $p_m = 1/(L \times 50) = 0.000167$. The maximum number of evolution generations is prescribed as 300. The computation results are shown in Figure 2.

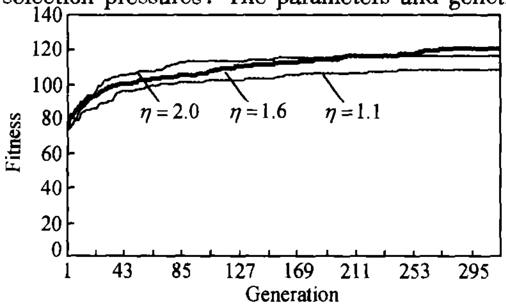

Fig. 2 Performance of GA in solving One-Max function.

Figure 2 shows that each of low or high selection pressures affects the searching capabilities of the GA even for GA-easy problems. Moreover, considering the improper design of other genetic parameters, the premature convergence of the GA is very common in solving complicated optimization problems with fitness functions featured with multi-modality (or multi-peaks), non-linearity, non-additivity, etc., which will result in a series schema deceptiveness in the population evolution. Those kinds of optimization problems are called GA-hard ones.

To overcome the premature convergence of the GA, the large mutation or adaptive mutation is adopted usually, but with only a limited improvement[7]. Whitley[8] and Kuo[9] et al. suggested that new fitness scaling methods might help to tackle these GA-hard problems.

3 Non-monotone fitness scaling

Fitness scaling is a necessary step for both maximum and minimum optimization problems in improving the searching performance of the GA. As to the numerical mapping from the objective function to the fitness function, we have the general form T: $g(x) \rightarrow f(x)$, $x \in R$ (assuming $f(x) > 0$). Since the selection operation of individuals is based on their fitness, and the selection is proportional to the fitness in the form of roulette wheeling, the transformation will affect the selection pressure of the population. It is necessary for mapping to keep the order of individuals unchanged. That is for x_1, $x_2 \in R$, if $g(x_1) > g(x_2)$, then $f(x_1) > f(x_2)$, and if $g(x_1) = g(x_2)$, then $f(x_1) = f(x_2)$.

According to Neo-Darwinism[1,2,9], there exist three types of evolution selection for organisms: the stabilizing selection, directional selection and disruptive selection. The normal selection operation in the GA is a kind of directional selection, which leads the population towards the states of higher fitness. The stabilizing selection is a randomized operation in fact, and has no direct relation with the fitness of individuals, so that the property of the population is kept constant. The disruptive selection reverses the order of ranking of individuals in the population, and the fit ones are chosen with smaller probability, and total quality of the population drops down, or significant changes may happen in the evolution process of population.

The directional selection of organisms in nature is not casual, and it can only happen under specified conditions, such as a large enough population size, a suitable environment, the immunity of livings. Viewed from the total evolution process, the populations of organisms in nature evolve indeed from simple to complex, but degeneration does take place at some phases or locally. Similarly, the selection rule of survival of the fittest can also be used along with other kinds of selection operations in the GA.

In applying the GA to optimization problems, the directed selection ensures the successive evolution of the population, but also results in premature convergence and schema deceptiveness. To overcome this difficulty, we must design new selection strategies and operators. Through a large number of experiments of the GA in solving complicated optimization problems, and with due consideration of the work of others[3,9], we have formulated the no-monotone fitness scaling logic.

Definition 1. Given any optimization function opt$g(x)$ ($x \in R$, $g(x) \geqslant 0$), its fitness function is $f(x)$, and for $x_1, x_2 \in R$, there exists $x_0 \in R$, $x_1 < x_0 < x_2$. If the original fitness values satisfy $f(x_1) > f(x_0) > f(x_2)$, then through the new scaling operation, we have $f'(x_1) \geqslant f'(x_0) \leqslant f'(x_2)$. This type of scaling is called no-monotone fitness scaling method.

It can be seen that the non-monotone fitness scaling is a kind of disruptive selection, and it does not meet the constant requirement of the individual order. In this case, the most and least fit individuals would have higher survival probabilities, which means, on the other hand, that the interim individuals contribute much less to the next generation of the population. Here two types of basic no-monotone fitness scaling methods are given.

3.1 Averaged fitness scaling

Suppose that the population is $P = \{\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n\}$ with size n, and the fitness values of individuals are $f(\boldsymbol{a}_1), f(\boldsymbol{a}_2), \cdots, f(\boldsymbol{a}_n)$ respectively. Then the averaged fitness scaling is defined as

$$f'(\boldsymbol{a}_j) = \left| f(\boldsymbol{a}_j) - \bar{f} \right|, \quad \boldsymbol{a}_j \in P, \tag{2}$$

where $\bar{f} = \frac{\sum_{i=1}^{n} f_i}{n}$ is the averaged fitness of the population.

For any $\boldsymbol{a}_j \in P$, the larger $f'(\boldsymbol{a}_j)$ means the higher survival probability. In other words, those individuals with their fitness values far away from the averaged one will have more copies to be reproduced into the next generation or for crossover and mutation. Those individuals with their fitness close to the averaged fitness are given less chance to be selected and will have less copies to be reproduced.

3.2 Adjustable fitness scaling

Given population $P = \{\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n\}$, calculate $f_{\min} = \min\{f(\boldsymbol{a}_j),\ j = 1, 2, \cdots, n\}$, and $f_{\max} = \max\{f(\boldsymbol{a}_j),\ j = 1, 2, \cdots, n\}$. Adjustable fitness scaling is defined as

$$f'(\boldsymbol{a}_j) = \left| f(\boldsymbol{a}_j) - \bar{f}' \right|, \quad \boldsymbol{a}_j \in P, \tag{3}$$

where $\bar{f}' = f_{\min} + \alpha(f_{\max} - f_{\min})$, $\alpha \in [0, 1]$. If $\alpha = 0$, the adjustable fitness scaling becomes the traditional scaling algorithm with the least fit individual not selected. When $\alpha = 1$, it is the disruptive selection with the greatest degeneration. By choosing a proper value of α, we can regulate the pressure and disruptive degree of the selection operation.

Besides, we need to consider the requirements of the GA in different phases of evolution, so that the exploration and exploitation of searching can be integrated effectively. Let T be the maximum number of generations for evolution, the no-monotone fitness scaling is adopted within t generations, where $t \leqslant \beta T$, $0 \leqslant \beta \leqslant 1$, and usually we have $\beta = 2/3$. When $t \leqslant \beta T$, the GA carries out exploration, and makes global searching on the total definition space. After $t > \beta T$, the GA turns to the directed selection, and starts to exploitative searching.

4 Numerical examples

The practical optimization models in automatic control, machine learning and management of economics and enterprises are all of complicated ones, and they can hardly be solved with traditional analytic methods and searching algorithms, which are usually stuck in local optima. Two examples are designed to be solved by the GA with no-monotone fitness scaling.

4.1 Trap function

Trap function is a special bit unitation function[2], and takes the form

$$f(u) = \begin{cases} \frac{a}{z}(z - u), & \text{if } u \leqslant z; \\ \frac{b}{L - z}(u - z), & \text{otherwise}, \end{cases} \tag{4}$$

where a and b are constants, and $a < b$. The global optimal string is "111…111", and its deceptive attractor is "000…000", and $u(111\cdots111) = L$, $f(L) = b$, $u(000\cdots000) = 0$, $f(0) = a$. For $u < z$, $f(u)$ is a monotone-falling function, and when $u \geqslant z$, $f(u)$ is a monotone-rising function.

The parameters of the GA are set as: the string length $L = 30$, population size $n = 100$, two-points crossover with $p_c = 0.6$, mutation with $p_m = 1/L = 0.0333$.

(i) Assuming $a = 80$, $b = 100$, we analyse the relation between the searching capability of the GA and $z = \text{int}(L/2) + k$, where the proportional selection and elitist retainment are used, and the maximum number of generations of evolution is prescribed as 200. The calculation is repeated 120 times for each case, the statistical results are listed in Table 1, where G_a, P_a are the number of the averaged generations and the probability converging to a respectively, and G_b, P_b are those converging to b.

Table 1 Relation between GA searching capability and parameter z

Averaged generation and probability	$z > \text{int}(L/2) = 15$					
	$z = 15 + 1$	$z = 15 + 3$	$z = 15 + 5$	$z = 15 + 7$	$z = 15 + 9$	$z = 15 + 12$
G_a	31.25	29.73	30.06	32.08	33.48	35.83
P_a	0.3333	0.9167	1	1	1	1
G_b	27.50	34.00	–	–	–	–
P_b	0.6667	0.0833	0	0	0	0
Averaged generation and probability	$z \leqslant \text{int}(L/2) = 15$					
	$z = 15 - 9$	$z = 15 - 7$	$z = 15 - 5$	$z = 15 - 3$	$z = 15 - 1$	$z = 15$
G_a	–	–	–	–	–	32.50
P_a	0	0	0	0	0	0.0556
G_b	35.35	32.91	31.86	28.08	26.83	24.03
P_b	1	1	1	1	1	0.9444

" – " denotes that the GA cannot converge to a or b in the maximum number of generations.

For $z > \text{int}(L/2) = 15$, the schema deceptiveness of trap function gets stronger along with the increasing of z, and the population finally converges to $a = 80$. When $z \leqslant \text{int}(L/2) = 15$, the schema deceptiveness is weakened as z decreases, and $b = 100$ becomes the converged point of GA population.

The competition of the same order of schema is intensified around $z = \text{int}(L/2) = 15$, and if the schema containing effective alleles or the alleles of deceptive attractor become dominant, the population would converge quickly to the global optima or to the deceptive attractor.

(ii) Assume $z = \text{int}(L/2) + 5 = 20$, $a = 60$ or $a = 80$, $b = 100$, and the averaged and adjustable fitness scaling are adopted respectively with $\alpha = 0.67$ and $\beta = 2/3$. The maximum number of generations of the population evolution is 200, the other parameters are set as above. The computation is repeated 120 times for each case, and the statistical results are listed in Table 2.

Table 2 Comparison of GA with and without no-monotone fitness scaling

Averaged generation and probability	Directed selection		Averaged fitness scaling		Adjustable fitness scaling	
	$a = 80$	$a = 60$	$a = 80$	$a = 60$	$a = 80$	$a = 60$
G_a	31.25	46.70	43.74	29.83	120.44	141.33
P_a	1	1	0.9500	0.7667	0.6500	0.0500
G_b	–	–	29.00	22.75	140.76	144.87
P_b	0	0	0.0500	0.2333	0.3500	0.9500

Computation precision: $\leqslant 0.00001$.

Table 2 shows that the averaged fitness scaling made a little improvement over the directed selection with $a = 80$, and displayed some influence with $a = 60$, which means that it has a certain degree of capability for overcoming the schema deceptiveness. When the adjustable fitness scaling was applied, significant improvements were made with $a = 80$, and its capability of overcoming the schema deceptiveness reached nearly 100% with $a = 60$.

4.2 Needle-in-a-haystack problem

As to needle-in-a-haystack problems (NiH), the global optima are surrounded by the worst solutions, which blocks the way of the schema recombination from low order to high order. In these cases, it is very difficult for the GA to move from local optima to the neighborhood of global optima in randomized searching. To analyze the searching process of the GA, a typical NiH function is designed as

$$\max f(x, y) = \left(\frac{a}{b + (x^2 + y^2)}\right)^2 + (x^2 + y^2)^2, \quad x, y \in [-5.12, 5.12], \tag{5}$$

where $a = 3.0$, $b = 0.05$, $\max f(0,0) \approx 3600$, and four local optimal points with the same function value (2748.78) are $(-5.12, 5.12)$, $(-5.12, -5.12)$, $(5.12, 5.12)$, $(5.12, -5.12)$. With the changing of the control parameter $\{a, b\}$, this function will be featured with different degrees of schema deceptiveness, and the above four local optima points are the deceptive attractors.

The parameters of the GA are set as: $L = 40$ (x, y are all coded with 20 binary bits), $n = \{20, 40, 60, 80, 100, 120\}$, $p_c = 0.6$, $p_m = 1/L = 0.025$. The proportionate selection and the elitist retainment are used, and the maximum number of generations of the population evolution is prescribed as 300. The computation is repeated 120 times for each case, and the statistical results converging to the global optima are listed in Table 3.

Table 3 Performance of GA for NiH with different population sizes

Averaged generation and probability	$n = 20$	$n = 40$	$n = 60$	$n = 80$	$n = 100$	$n = 120$
G_0	88.60	83.43	81.21	82.59	83.37	81.43
P_0	0.2083	0.5250	0.5750	0.6083	0.6417	0.6667

With the population size fixed at 60, the searching trajectories of GA under schema deception and after overcoming schema deception are shown in Fig. 3, and the fitness growing of the current best individual in the population along with the number of generations of evolution is shown in Figure 4.

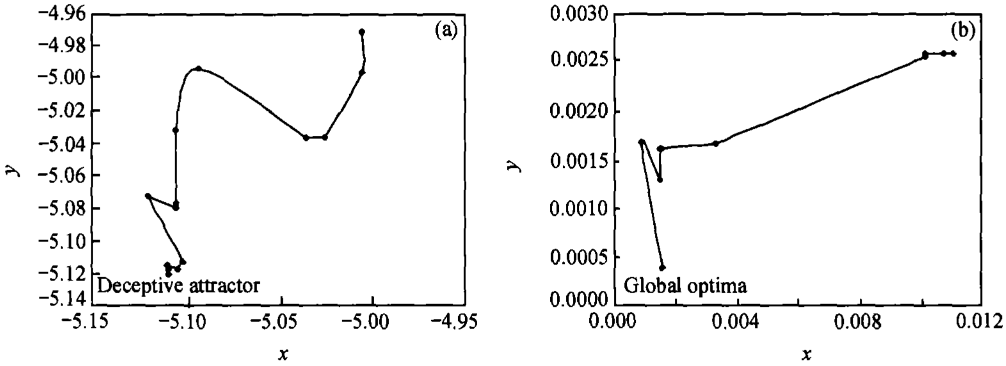

Fig. 3 Searching trajectory of GA. (a) With schema deception; (b) after overcoming schema deception.

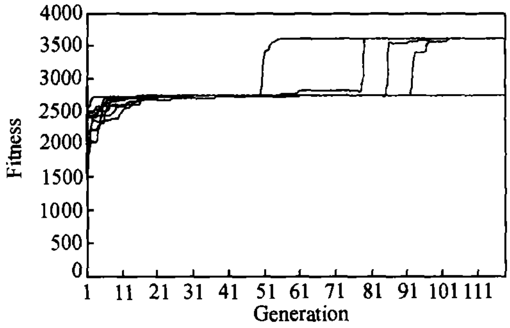

Fig. 4 Performance of GA for NiH.

The schema deception led the searching process towards deceptive attactors, and then the GA was trapped in local optimal points (located randomly at one of the four points). When the schema deception was overcome by genetic operators, the GA shifted its search to the neighborhood of the global optima, and finally converged to the global optimal point. It can be seen that the global optimal convergence of the GA becomes a stochastic event under the circumstance of hard schema deceptiveness.

Now, we apply the averaged and adjustable fitness scaling to the GA in solving the NiH problem, where $\alpha = 0.33$, $\beta = 2/3$, and the other parameters are set as above. The computation is repeated 120 times for each case, and the statistical results converging to the global optima are listed in Table 4.

Table 4 Performance of GA for NiH with and without no-monotone fitness scaling

Averaged generation and probability	Directed selection		Averaged fitness scaling		Adjustable fitness scaling	
	$n = 60$	$n = 100$	$n = 60$	$n = 100$	$n = 60$	$n = 100$
G_0	81.21	83.37	84.15	74.15	81.61	78.22
P_0	0.5750	0.6417	0.9083	0.9833	0.8889	0.9167

Computation precision $\leqslant 0.1$.

Table 4 shows that for the NiH optimization problem, the GA can achieve a probability of 90% for finding the global optima by using the averaged and adjustable fitness scaling. When the schema deceptiveness of NiH function is increased, the global convergence of the GA with the directed selection drops down. On the other hand, the GA with no-monotone fitness scaling keeps a marked performance.

5 Conclusion

In this paper, the property of the selection operation has been discussed in detail, and a new type of no-monotone fitness scaling has been advanced to deal with the premature convergence of the population and the schema deceptiveness, and its effectiveness has been shown in two numerical experiments. Furthermore, no-monotone fitness scaling should be studied further about its form and parameters by taking into consideration different sizes of population, population initializing, and the random error of three genetic operators.

References

1 Richard, K. B. et al. Foundation of genetic algorithms, 4. San Francisco: Morgan Kaufmann Publishers, Inc., 1997.

2 Mitchell, M. An Introduction to Genetic Algorithms, Cambridge-Massachusetts: The MIT Press, 1996.

3 Goldberg, D. E. et al. Optimizing global-local search hybrids. In: Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-99), Banzhaf, W. et al. eds. San Francisco: Morgan Kaufmann Publishers, 1999: 26.

4 Yun, W. et al. The analysis of global convergence and computational efficiency for genetic algorithm. Control Theory and Applications (in Chinese), 1996, 13(4): 455.

5 Goldberg, D. E. et al. A comparative analysis of selection schemes used in genetic algorithms. In: Foundations of Genetic Algorithms, Rawlins, G. J. E. ed., San Mateo: Morgan Kaufmann Publishers, 1991: 69.

6 Miller, B. et al. Genetic algorithms, selection schemes, and the varying effects of noise. Evolutionary Computation, 1996, 4(2): 25.

7 XUE, Z. et al. The characteristics and prevention of population premature of GA. Sciences in China (in Chinese), Ser. E, 1996, 26(4): 363.

8 Whitley, D. Fundamental principles of deception in genetic search. In: Foundations of Genetic Algorithms, Rawlins, G. J. E. ed., San Mateo: Morgan Kaufmann Publishers, 1991: 221.

9 Kuo, T. et al. A genetic algorithm with disruptive selection. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 1996, 26(2): 65.