

Adaptation Ethnic Group Evolution Algorithm

Hao Chen

School of Computer Science
and Engineering, Xi'an
University of Technology
Xi'an 710048, China
chenhao_xaut@yahoo.com.
cn

Du-Wu Cui

School of Computer Science
and Engineering, Xi'an
University of Technology
Xi'an 710048, China
cuidw@xaut.edu.cn

Xue Li

International Business
School of Shaanxi Normal
University,
Xi'an 710062, China
Snow_lx2008@163.com

Zhan-Min Wang

School of Computer Science
and Engineering, Xi'an
University of Technology
Xi'an 710048, China
zhanmin@xaut.edu.cn

Abstract—Enlightened by the analysis method of modern anthropology for sociogroup, the ethnic group evolution algorithm (EGEA)*, whose population is formed of several ethnic groups, is proposed. The ethnic group is a kind of clustering form for individuals, and represents the possible evolution trends, which are implicit in population. In EGEA, ethnic groups are used as the competition and generation unit of individuals, which improve the parallelism and diversity of population evolution. In this paper, an improving EGEA, adaptation ethnic group evolution algorithm (AEGEA), is developed. In AEGEA, a series of adaptation control strategies are used to adjust evolution parameters dynamically so as to make the searching process more steady and efficient. The simulations tests of numerical optimization show AEGEA can restrain premature convergence phenomenon effectively during the evolutionary process while increasing the convergence speed greatly.

Keywords—genetic algorithm, ethnic group evolution algorithm, adaptation ethnic group evolution mechanism

I. INTRODUCTION

As we known, the genetic algorithm (GA) pertains to searching algorithms with an iteration of generation-and-test. Two operators—crossover and mutation—give each individual the chance of optimization and ensure the evolutionary tendency with the selection mechanism of survival of the fittest. This simulation mechanism for biological evolution sets up a basic searching framework, but it takes on some limitation also. Because the crossover operator and mutation operator search for new individuals in search space randomly and semi-blindly, which lead to the double property of search process [1]: it not only makes the algorithm converge also let it prematurely converge to the non-optimal area. Vose and Liepins [2] believe the genetic operators are not enough to lead population get to the optimum fitness static schema in coding space. So, making use of the idea and method in natural science to improve GA has become an important research trend.

In modern anthropology research, the ethnic group [3] has been used as the organization unit of sociogroup to observe the transformation and development of society. This is a kind of research method for analyzing complicated huge group.

* Supported by the National Natural Science Foundation of China under Grant No. 60743009 and 60501006; the Natural Science Foundation of Shaanxi Province of China under Grant No. 2006F-43.

Classifying a large group into some simple units is in favor of discovering the possible evolutionary trends and adjusting control strategies timely. Taking this research method into GA, a new evolution mode, named as ethnic group evolution algorithm (EGEA) [4], is developed.

The canonical EGEA use static parameters. So, it is short of dynamical adjustment ability. In this paper, a series of adaptation control strategies are taken into EGEA, and the adaptation ethnic group evolution algorithm (AEGEA) is developed. To validate AEGEA, simulation tests of numerical optimization are performed. The results show AEGEA is available to the numerical optimization problems.

II. ETHNIC GROUP EVOLUTION MECHANISM

A. Definitions of ethnic group evolution

1) Macrogamete

In population evolution process, the stability of individuals is in favor of the inheritance and accumulation of genetic information, but the evolution of individuals depends on the variation of genes. If the balance between stability and variability of individual is broken, the efficiency of population evolution will be disturbed heavily. A new idea for harmonizing the inheritance with exploration of genetic information is enlightened by sexual reproduction of higher creature. Female gametes have larger size and less number, who are more stable and used to carry energy and information. On the contrary, male gametes have less size and larger number, who are more active. Depending on the cooperation between female gametes and male gametes, the sexual reproduction ensures the stability of heredity while improves the number of mutation. Imitated by this mechanism, we make use of a role classification process to select representative individuals from population, and named it as “macrogamete” M . The macrogamete will keep more stable and be used to carry genetic information. Other individuals will have higher probability to change themselves and be used to explore new information.

2) Ethnic group

The ethnic group E is a mapping of organization structure for macrogamete, and it is generated by a certain clustering rule. Based on this definition, the ethnic group takes on some characteristics: (1) a clustering rule means using a special view to analyze population, and different view create different ethnic

group; (2) the ethnic group is just a structure mapping, which store the clustering index for macrogamete. The clustering process doesn't take individuals out of population to generate some new subpopulations; (3) the structure of ethnic group is changing with the iteration of population.

3) Evaluation criterions of EGEA

The population of EGEA consists of an n -tuple of binary strings A of length L . With m dimension, $A = (a_{1,1} a_{1,2} \dots a_{1,l_1} \dots a_{m,1} a_{m,2} \dots a_{m,l_m})$, $w = (1, 2, \dots, m)$, $k = (1, 2, \dots, L_w)$, $\sum_{w=1}^m l_w = L$. F_i is the fitness of A_i and the fittest individual is written as A_{best} . \bar{F} and F_{max} is the average fitness and best fitness of current population.

a) The fitness contrast ratio

The fitness contrast ratio of A_i is

$$\bar{R}_i = (F_{max} - F_i) / (F_{max} - \bar{F}) \quad (1)$$

If $F_i > \bar{F}$, then $0 \leq \bar{R}_i < 1$; else $\bar{R}_i \geq 1$.

b) The code difference ratio

The code difference ratio between A_i and A_j is

$$D(A_i, A_j) = D_{ij} = \delta_{ij} / \sum_{w=1}^m \left(\sum_{k=1}^{L_w} \eta_{w,k} \right) \quad (2)$$

where $\delta_{ij} = \sum_{w=1}^m \left[\sum_{k=1}^{L_w} (b_{w,k} \times \eta_{w,k}) \right]$, $b_{w,k} = \begin{cases} 1, & a_{i,w,k} \neq a_{j,w,k} \\ 0, & a_{i,w,k} = a_{j,w,k} \end{cases}$, and

the weight of gene is $\eta_{w,k} = l_w - k_w + 1$.

In practical operation, the code difference ratio between A_i and A_{best} is mainly care about, so $D(A_i, A_{best}) = D_i$ is the working formula.

c) The race exponent

The race exponent of A_i is

$$Q(A_i, A_{best}) = Q_i = D_i^{\rho \bar{R}} \quad (3)$$

where $\rho \in \begin{cases} (0,1) & , \bar{R} < 1 \\ [1,2) & , \bar{R} \geq 1 \end{cases}$. Let the race exponent of A_{best} is 1,

and if $D_i = 0$ let $Q_i = 0$, then $Q \in [0,1]$.

In essence, the race exponent is a kind of fitness scaling, which integrates fitness contrast ratio with individual code difference ratio and develops from the idea of keeping individuals balance between fitness growth and population diversity preservation.

d) The population diversity measurement

The average code difference ratio of population in the k generation is

$$\gamma_k = \frac{1}{n-1} \sum_{i=1}^{n-1} \left(\frac{1}{n-i} \sum_{j=i+1}^n D_{ij} \right) \quad (4)$$

The $\gamma \in (0,1)$ is more close to 1 means the population is more diverse. In this paper, it is used to be the measurement of population diversity.

B. Canonical ethnic group evolution algorithm

The core idea of ethnic group evolution is to make use of clustering method to build up ordered organization form, the ethnic group, in random population, and use ethnic groups as the intermediate unit between population and individual to control the generation and competition process among individuals. The figure 1 is the flowchart of EGEA.

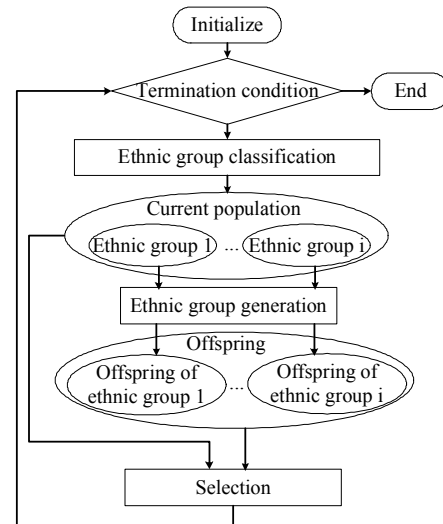


Figure 1. Flowchart of EGEA

From figure 1 we can see the first step of EGEA is to create ethnic groups in population. A classical view of clustering is the similarity of binary code among individuals. Macrogamete is a group of representative individuals in population. Performing individuals clustering in it can reduce computation and improve the accuracy of clustering. So, a classical ethnic group form is generated from macrogamete according to the code difference ratio of them. This clustering process is also an analysis process of population structure. Each of ethnic groups represents one evolution tendency within population. In this clustering process, the macrogamete factor $\lambda \in (0,1)$ and ethnic group radius $\theta \in (0,1)$ are two important parameters. The λ is used to set up the acceptance value of race exponent for choosing macrogamete. The θ is used to set up the radius among ethnic groups. So, we can make use of λ and θ to control the structure of ethnic group.

EGEA makes use of the ethnic group as the generation unit and allots the offspring quantity for each ethnic group according to its competitive capacity. The generation process

of each ethnic group includes matrimony, crossover and mutation three steps. The matrimony is a kind of selection operation, which based on race exponent of individuals and center about macrogamete of each ethnic group to select spouse. Then, the suited crossover and mutation operator are chosen to generate offspring. Ethnic group generation mechanism ensures the generation probability of representative individuals. Meanwhile, the competition for existence among ethnic groups improves the searching efficiency of population.

The next population is selected from current population and offspring of ethnic groups, and a new ethnic group evolution process will begin until the iteration process has finished.

III. ADAPTION ETHNIC GROUP EVOLUTION ALGORITHM (AEGEA)

In order to make the searching process of EGEA have self-regulation ability, a series of adaptation control strategies are taken into EGEA.

1) Creating ethnic group

Step 1. Evaluating the race exponent for all individuals.

Step 2. Selecting macrogamete: the macrogamete factor λ is decided by following formula

$$\lambda = (\bar{Q} + \zeta) / 2 \quad (5)$$

where \bar{Q} is the average race exponent of current population and $\zeta = 0.6$. If $Q_i > \lambda$, then $A_i \in M$.

Step 3. Building up ethnic group: the first step is choosing A_{center} , which are the core individual of each ethnic group. Ranking macrogamete by race exponent, and let $A_{best} \in A_{center}$. Searching the rank, if $A_i \notin A_{center}$ and $\forall A_{center}$ has $D(A_i, A_{center}) > \theta$, then $A_i \in A_{center}$. The ethnic group radius factor θ is decided by following formula

$$\theta = (\gamma + \varepsilon) / 2 \quad (6)$$

where $\varepsilon = 0.4$, and γ is the diversity of last population.

Then, allocating the remaining macrogamete to each ethnic group. For all $A_i \in M$ and $A_i \notin A_{center}$, if $A_{center,j}$ is the center individual of E_j , and $D(A_i, A_{center,j})$ is the least code difference ratio for all A_{center} , then $A_i \in E_j$.

Step 4. Allotting the generation scale: the generation scale of E_k is defined as

$$S_k = n \frac{Q_{c,k}}{\sum Q_c} \quad (7)$$

where $Q_{c,k}$ is the race exponent of $A_{center,k}$ and $\sum Q_c$ is the sum race exponent of all ethnic group center individuals.

2) Ethnic group generation

Step 1. Selecting parents: based on race exponent, a macrogamete M_i from current ethnic group, and an individual A_j from whole population are selected by roulette selection.

Step 2. Evaluating the mating result: the attraction degree is decided by following formula

$$\sigma_{ij} = D_{ij}(Q_i + Q_j) \quad (8)$$

if $\sigma_{ij} > \text{Random}(0,1)$, then mate succeed; else return step 1.

Step 3. Choosing crossover operator to produce offspring.

Step 4. The mutation probability of offspring is defined as

$$P_m = \begin{cases} \text{random}(0.001, 0.01) & , \max(Q_i, Q_j) = 1 \\ 1 - \max(Q_i, Q_j) & , \text{otherwise} \end{cases} \quad (9)$$

Step 5. If the offspring count of current ethnic group is smaller than the generation scale, then return step 1; else start a new generation process for next ethnic group.

3) Simulated annealing rank selection

In general, the fitness expresses the absolute value of individual. Ranking current population and its offspring according to it. The position numbers of individuals express its relative value. Selecting next population by evaluating individuals synthetically, which can form more reasonable population. Based on this idea, the simulated annealing rank selection operator is developed. In this selection operator, race exponent is used to measure the absolute value of individual, and a simulated annealing parameter is used to measure its relative value. The selection probability of individual in k position of rank is defined as

$$P_s(k) = \sigma Q_k + (1 - \sigma) \hbar_k \quad (10)$$

where $\hbar(k) = \exp(-\gamma/T_k)$ is the simulated annealing parameter of individual in k position of rank, γ is the diversity parameter of current population and $T_k = T_0 \beta^k$ is the annealing temperature. In generally, $T_0 = 2n$, β is a decimal fraction which is less than 1, and weighted value $\sigma \in (0,1)$ is used to control the affection proportion between race exponent and simulated annealing parameter.

This selection process can effectively control the survival pressure and keep population structure more reasonably. Since the population diversity affects the distribution of race exponent. Therefore, a relationship between σ and γ can be built up. In this paper, a simple formula is used

$$\sigma = \gamma \quad (11)$$

4) Adjacent domain searching

In order to improve the precision of solution, an adjutant searching operator based on numeric coding is used to search the adjacent domain of best individual, which is get in ethnic group evolution process based on binary coding.

Step 1. The start point, searched in the m-dimensional space based on binary coding, can be represented as $A_{best} = (x_1, \dots, x_m)$, and F^0 is its fitness. The searching adjacent domain of x_i is $[o_i - \eta_i, o_i + \eta_i]$, where o_i is the i-th dimension real number of A_{best} , and $\eta_i = \frac{|\bar{x}_i - \underline{x}_i|}{(2^{l_i+1} - 2)}$, \bar{x}_i and \underline{x}_i is the upper limit and lower limit of x_i , l_i is the binary coding length of x_i . If we set up the searching step number is $\alpha = 1, 2, \dots$, then the initial searching step width of x_i is $g = \eta_i / \alpha$.

Step 2. For each o_i , move one unit step at increase and decrease trend to generate tow new points. Comparing the fitness among these tow points and start point, then choosing the trend with best fitness as the searching trend of i-th dimension value.

Step 3. For each searching steps, generate a new points, and the variable of each dimension is

$$o_i^{k+1} = o_i^k + hg^k d_i^k \quad (12)$$

where $d_i^k = -1, 0, 1$ is the searching trend of x_i , and the searching step $h = 1, \dots, \alpha$. Evaluating the fitness of these α new points and finding next start point $A_{best}^{k+1} = (o_1^{k+1}, \dots, o_m^{k+1})$ with best fitness F^{k+1} .

Step 4. If $F^{k+1} < F^k$, then stop search and output the final searching result A_{best}^k ; else come back step 2 to search interval $[o_i^{k+1} - g^{k+1}, o_i^{k+1} + g^{k+1}]$ with new start point $A_{best}^{k+1} = (o_1^{k+1}, \dots, o_m^{k+1})$, and the new searching step width is $g^{k+1} = g^k / \alpha$.

IV. SIMUATIONS

In order to test the performance of AEGEA, six benchmark functions have been used

Generalized Schwefel's Problem 2.26:

$$f_1(x) = \sum_{i=1}^m (-x_i \sin \sqrt{|x_i|}), S = [-500, 500]^m;$$

Generalized Rastrigin's Function:

$$f_2(x) = \sum_{i=1}^m [x_i^2 - 10 \cos(2\pi x_i) + 10], S = [-5.12, 5.12]^m;$$

Ackley's Function:

$$f_3(x) = -20 \exp \left(-0.2 \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2} \right) - \exp \left(\frac{1}{n} \sum_{i=1}^n \cos(2\pi x_i) \right) + 20 + e,$$

$$S = [-32, 32]^m;$$

Generalized Griewank Function:

$$f_4(x) = \frac{1}{4000} \sum_{i=1}^m x_i^2 - \prod_{i=1}^m \cos \left(\frac{x_i}{\sqrt{i}} \right) + 1, S = [-600, 600]^m;$$

Sphere Model:

$$f_5(x) = \sum_{i=1}^m x_i^2, S = [-100, 100]^m;$$

Schwefel's Problem 2.22:

$$f_6(x) = \sum_{i=1}^m |x_i| + \prod_{i=1}^m |x_i|, S = [-10, 10]^m$$

A. Comparison between EGEA and AEGEA on functions with 30 dimensions

Both EGEA and AEGEA is to run 200 generations for each function, and use the same code length for each function. The code length need to ensure the deviation of variable lower than 10^{-3} for each function. The population sizes of them both are $n=100$. The λ and θ of EGEA are 0.6 and 0.4. Furthermore, the EGEA use linear rank selection operator and adjacent domain searching operator.

After run 50 trials, the experimental statistic results between the EGEA and the AEGEA of f_1 - f_6 with 30 dimensions are shown in Table 1. As can be see, AEGEA find the exact global optimum of f_2, f_3 and f_6 in all trials, but EGEA just find the exact global optimum of f_6 . For all the six functions, both the mean function value and the mean number of function evaluations of AEGEA are much better than those of EGEA. These results show the adaptation mechanism effectively improve the searching capacity of EGEA.

B. Comparison between AEGEA and four outstanding algorithms on functions with 30 dimensions

OGA/Q [5], HTGA [6], MAGA [7], and StGA [8] are four methods obtain good performance on numerical optimization problem. In [5], the termination criterion of OGA/Q was the quality of the solution couldn't be further improved in successive 50 generations after 1000 generations. In [6], HTGA would be stopped as soon as the smallest function value of the fitness member of the population is less than or equal to the mean function value given by the OGA/Q. In [7], the termination criterion of MAGA is to run 150 generations for each function. In [8], StGA make use of generation number to be the termination criterion and most generation number of six functions is less than 200 generations. Since the termination criteria of these four algorithms are different, to make a fair comparison, we let the computational cost of AEGEA be less than these four algorithms, and compare the qualities of their final solutions at the given computational cost. Therefore, the termination criterion of AEGEA is to run 200 generations for each function, and the parameters are same as experiment A.

From table 2 we can see the solution of AEGEA for f_2 and f_6 is much better than other algorithms. For f_1 , only StGA find the exact global optimum. For f_3, f_4 and f_5 , the solution of MAGA is much better than other algorithms, and the solution of AEGEA are close to HTGA and StGA and much better than OGA/Q. To summarize, the results show AEGEA is competent for the numerical optimization problems.

TABLE I. COMPARISON OF EXPERIMENTAL RESULTS BETWEEN EGEA AND AEGEA

Function		f_1	f_2	f_3	f_4	f_5	f_6	
EGEA	Generations	Maximum	199	199	199	199	199	119
		Minimum	176	65	124	105	95	39
		Mean	186	156	157	125	136	87
		Standard deviations	14.3	54.2	38.6	28.1	36.1	24.9
	Function value	Maximum	-12563.7140	2.4	0.83	0.28	1.3	0
		Minimum	-12569.4698	0	0	0	0	0
		Mean	-12568.2564	6.8e-2	0.16	0.06	0.45	0
		Standard deviations	8.7e-1	5.6e-1	8.3e-2	9.2e-3	6.7e-2	0
	Function evaluations	Maximum	47895	25689	35692	41965	26582	15987
		Minimum	28954	14562	25468	20548	13687	6829
		Mean	36478	18745	30258	31452	20687	9654
		Standard deviations	6547	3562	4893	6895	4165	2983
AEGEA	Generations	Maximum	199	189	173	199	188	93
		Minimum	119	58	59	80	43	25
		Mean	178	90	77	176	118	47
		Standard deviations	27.2	47.6	25.1	35.9	44.1	19.3
	Function value	Maximum	-12568.6798	0	0	1.4e-2	0.29	0
		Minimum	-12569.4866	0	0	0	0	0
		Mean	-12569.2367	0	0	2.5e-4	6.8e-2	0
		Standard deviations	1.5e-1	0	0	8.7e-3	3.9e-4	0
	Function evaluations	Maximum	37579	16795	29001	36819	23181	14836
		Minimum	21790	9005	15614	21787	11324	4822
		Mean	28786	11114	21409	25858	20391	9009
		Standard deviations	5947	2546	3958	4827	3238	2869

TABLE II. COMPARISON OF EXPERIMENTAL STATISTIC RESULTS AMONG AEGEA, OGA/Q, HTGA, MAGA AND StGA

Function		f_1	f_2	f_3	f_4	f_5	f_6
AEGEA	MNFE	28786	11114	21409	25858	20391	9009
	MFV (Std)	-12569.2367 (1.5e-1)	0 (0)	0 (0)	2.5e-4 (8.7e-3)	6.8e-2 (3.9e-4)	0 (0)
OGA/Q	MNFE	302166	224710	112421	134000	112559	112612
	MFV (Std)	-12569.4537 (6.447e-4)	0 (0)	4.44e-16 (3.9e-16)	0 (0)	0 (0)	0 (0)
MAGA	MNFE	10862	11427	9656	9777	9502	9591
	MFV (Std)	-12569.4866 (7.121e-12)	0 (0)	4.44e-16 (0)	0 (0)	0 (0)	0 (0)
HTGA	MNFE	163468	16267	16632	20999	20844	14285
	MFV (Std)	-12569.46 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
StGA	MNFE	1500	28500	10000	52500	30000	17600
	MFV (Std)	-12569.5 (0)	4.42e-13 (1.1e-13)	3.52e-8 (3.5e-9)	2.44e-17 (4.5e-17)	2.45e-15 (5.2e-16)	2.03e-7 (2.95e-8)

MNFE: Mean Number of Function Evaluation MFV: Mean Function Value Std: Standard deviations

C. Analysis of ethnic group evolution process

The figure 2 shows the variation of function value, population diversity and ethnic group number of AEGEA with iteration for 6 functions with 100 dimensions. As can be see, the searching speed of AEGEA is very quickly and most of function value curve is smooth and continuous that means AEGEA can search problem space with high availability and successfully avoid the evolution process to be at a standstill. The curve of population diversity show the ethnic groups make population keep a steady convergence trend and avoid population diversity to decline hastily. The curve of ethnic group number show the variation of ethnic group structure

takes on certain regularity. It changes from expansion to assimilation. In initial phase of evolution, it's likely to exist many evolution trends in the population at same time, which make ethnic groups take on an expanding process. The expansion of ethnic group make population parallel search several possible evolution trends and avoid missing the correct evolution trend. When population is close to the global optimum, ethnic groups begin to overlap and assimilate each other that make individuals get the global optimum from different way. This process help EGEA to improve the accuracy and efficiency of population convergence while restrains premature convergence greatly.

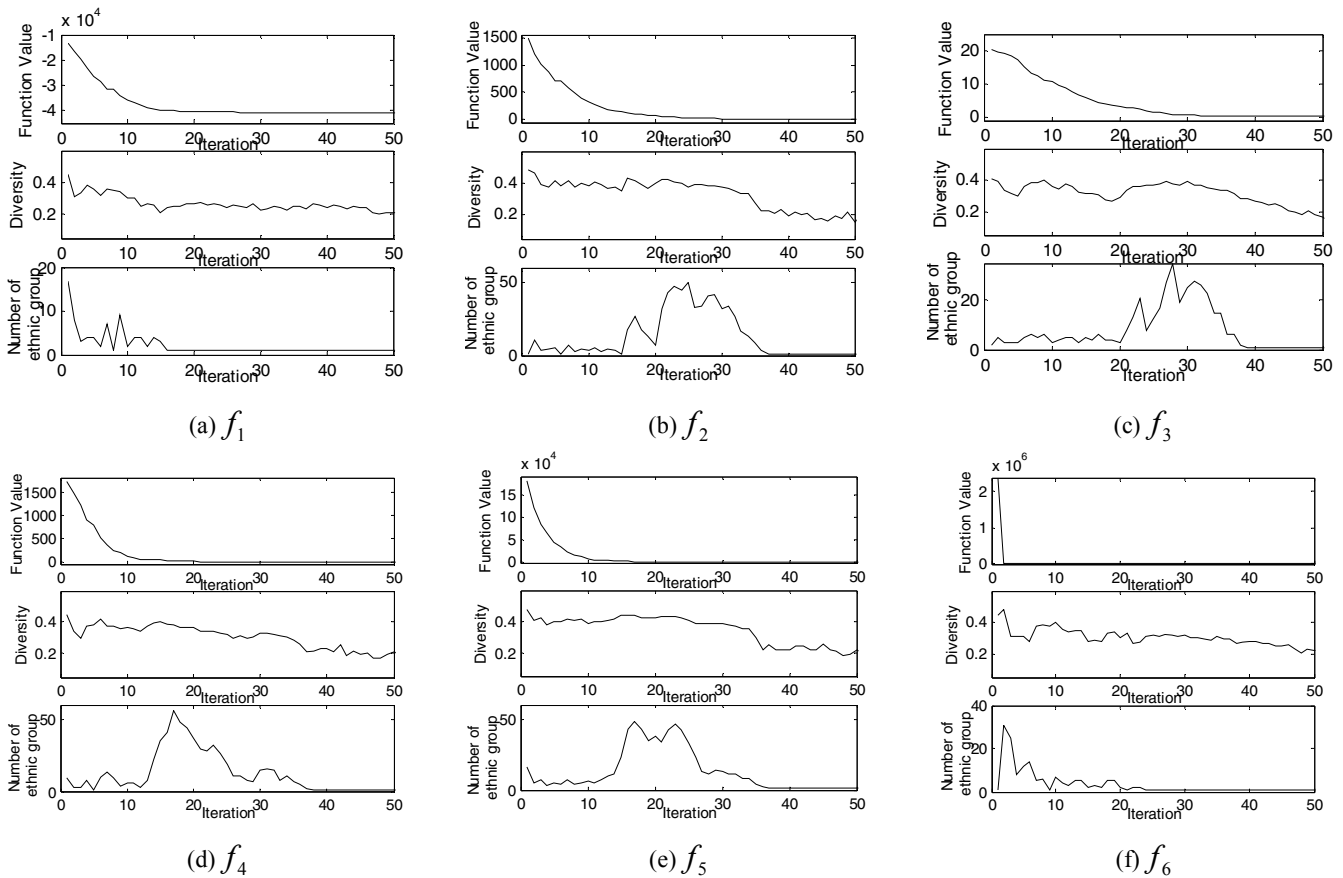


Figure 2. The variation of function value, population diversity and number of ethnic group of AEGEA with iteration for 6 functions with 100 dimensions

V. CONCLUSION

Based on ethnic group evolution mechanism and a series of adaptation control strategies, an improving ethnic group evolution algorithm AEGEA, is developed. The simulations for numerical optimization show that AEGEA is not only feasible but also valid and is helpful in avoiding the premature convergence phenomenon while increasing the convergence speed greatly. Meanwhile, The ethnic group evolution mechanism can not only keep the validity and diversity of population evolution but also provide a general framework for mining the evolutionary experience and knowledge from population, and make use of it to guide and accelerate the evolutionary process of ethnic group. In the future, we will continue with this research work.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their valuable comments and suggestions.

REFERENCES

[1] Xu Zong-ben, Gao Yong, "Precaution and analysis on characteristic of genetic algorithm", Science in China (series E), 1996, 26(4): 364-375.

[2] Vose, M.D., and Liepins, G.E. "Schema Disruption". In Proceeding of Fourth International Conference on Genetic Algorithms (ICGA 4), Belew, F. and Brooker, L. (ed.), San Mateo, CA: Morgan Kaufmann Publisher, Inc., 237-242, 1991.

[3] Max Weber. "The ethnic group" T. Parsons, Theory of Society.

[4] Hao Chen, Du-Wu Cui, Xue Li, and Zhan-Min Wang, "The Harmonious Evolution of Ethnic Group Algorithm", in Haikou China, Third International Conference on Natural Computation (ICNC2007), vol IV, IEEE Computer Society, 380-384, 2007.

[5] Y. W. Leung and Y. Wang. "An orthogonal genetic algorithm with quantization for global numerical optimization". IEEE Transactions on Evolutionary Computation, 2001, 5(1): 41-53.

[6] Jinn-Tsong Tsai, Tung-Kuan Liu, and Jyh-Horng Chou. "Hybrid Taguchi-Genetic Algorithm for Global Numerical Optimization". IEEE Transactions on Evolutionary Computation, 2004, 8(4): 365-377.

[7] W. C. Zhong, J. Liu, M. Z. Xue and L. C. Jiao. "A multiagent genetic algorithm for global numerical optimization". IEEE Transactions on Systems, Man and Cybernetics -Part B: Cybernetics, 2004, 34(2): 1128-1141.

[8] Zhenguo Tu and Yong Lu. "A Robust Stochastic Genetic Algorithm (StGA) for Global Numerical Optimization". IEEE Transactions on Evolutionary Computation, 2004, 8(5): 456-470.