

Double System Gene Expression Programming and It's Application in Function Finding Problems

Chaoxue Wang

School of Information and Control Engineering
Xi'an University of Architecture and Technology
Xi'an, China
wangchaoxue@163.com

Hui Dong

School of Information and Control Engineering
Xi'an University of Architecture and Technology
Xi'an, China
donghuiff@163.com

Kai Zhang

School of Information and Control Engineering
Xi'an University of Architecture and Technology
Xi'an, China
659798593@qq.com

Fangxiao Zhou

School of Information and Control Engineering
Xi'an University of Architecture and Technology
Xi'an, China
zhoufangxiao2002@163.com

Abstract—Gene Expression Programming (GEP) is a powerful evolutionary method for knowledge discovery and model learning, which simulates the process of natural evolution, and has the phenomena of low converging speed and readily being premature. After the birth of human society, the evolution processes of nature world and human show a rising acceleration characteristic under the artificial intervention. Inspired by this, this paper proposes a double system gene expression programming (DS-GEP), which consists of natural evolution system and artificial intervention system. The artificial intervention system includes individual intervention operation and population intervention operation. The individual intervention operation aims to repair the unfeasible genes in individuals with superior genes from gene pool. The population intervention operation uses extinction/restart strategy to form new population with high diversity after the evolution has fallen into stagnation. To validate the superiority of DS-GEP, DS-GEP and the standard GEP and an improved GEP in the relevant literatures are compared as regards enough function finding problems. The test results show that DS-GEP can overcome the stagnation and premature convergence phenomenon effectively during the evolutionary process, and promises competitive performance.

Keywords—double system gene expression programming; natural evolution system; artificial intervention system; function finding problems; gene pool

I. INTRODUCTION

GEP was proposed on the basis of genetic algorithm (GA) and genetic programming (GP) by Portugal scholar Candida in 2001[1, 2]. It adopts the dual architecture of genotype and phenotype, and retains the advantage of GA and GP, and possesses the characteristics that the algorithm flow is clear, realization is simple and precision is high, and especially it shows the excellent performance in complex function finding problems. Currently, GEP has been applied in physics, military science, demography,

dynamics, software engineering, water conservancy, mechanical engineering, medical science, meteorology, robot, etc. [3-12], but the problem of premature convergence still exists like other evolutionary algorithms.

Standard GEP simulates natural evolution process of living beings, which has serious randomness, blindness and spends a long period of time. Although the genetic operations of standard GEP could guarantee the probabilistic reachability of global optimal solution in theory, the phenomena of low converging speed and readily being premature still exists in practical application.

Humanity becomes the master of the nature world since human society emerged, and the natural evolution results which need a long geological ages of "survival of the fittest" in the past could be obtained within a comparative short period by using some artificial intervention ways. So, the evolution of human beings and other living beings is no longer a tortuous, long, passive natural evolution process, and becomes a semi natural evolution process characterized by acceleration, orientation, initiative under artificial intervention. Inspired by this, numerous researchers add artificial intervention methods to standard GEP. TANG Chang-jie et al. introduced three new modern transgene techniques, overlapped gene expression and backtracking evolution into GEP to interfere with the evolution process and got satisfied results [13]. JIANG Yue et al. used outbreeding strategy and dynamic fitness function strategy to interfere with population's evolution process, which can increase population diversity and improve performance of standard GEP [14]. JIA Li-yuan et al. proposed an immune gene expression programming by introducing the density and memory mechanism of immune algorithm, which can efficiently adjust the population diversity and prevent the loss of high quality solutions [15]. ZHANG Xuedong et al. introduced tabu search into genetic operations of GEP to improve the local searching ability, and then proposed gene expression programming based on parallel tabu

search (PTS-GEP) [16]. HU Jian-jun et al. adopted gene space balance strategy to interfere with the creation of initial population in standard GEP, so the diversity and fitness of initial population are excellent [17]. LI Tai-yong et al. proposed an adaptive population diversity tuning algorithm to maintain the population diversity and achieve high evolution efficiency in GEP [18].

On the basis of predecessors, this paper proposes a double system GEP (DS-GEP) by systematically simulating the methods and technologies of human intervention natural evolution process, which consists of natural evolution system and artificial intervention system. In DS-GEP, the artificial system aims at accelerating the convergence speed, prevent evolution stagnation and premature convergence by intervening with the natural evolution process of standard GEP. Control experiments about function finding problems suggest that DS-GEP has competitive performance not only in convergence speed but also in solution quality.

II. INTRODUCTION OF GEP

Standard GEP algorithm could be defined as a nine-meta group: $GEP = \{C, E, P_0, M, \varphi, \Gamma, \Phi, \Pi, T\}$, where C is the coding means; E is the fitness function; P_0 is the initial population; M is the size of population; φ is the selection operator; Γ is the crossover operator; Φ is the point mutation operator; Π is the string mutation operator; T is the termination condition. In GEP, individual is also called chromosome, which is formed by gene and linked by the link operator. The gene is a linear symbol string which is composed of head and tail. The head involves the functions from function set and the variables from the terminator set, but the tail merely contains the variables from the terminator set. The basic steps of standard GEP are as follows [1]:

- (1) Inputting relevant parameters, creating the initial population;
- (2) Computing the fitness of each individual;
- (3) If the termination condition is not met, go on the next step, otherwise, terminate the algorithm;
- (4) Retaining the best individual;
- (5) Selecting operation;
- (6) Point mutating operation;
- (7) String mutating operation (IS transposition, RIS transposition, Gene transposition);
- (8) Crossover operation (1-point recombination, 2-point recombination, Gene recombination);
- (9) Go to (2).

III. DOUBLE SYSTEM GENE EXPRESSION PROGRAMMING

This paper adds artificial intervention system, which consists of individual intervention and population intervention, to standard GEP and makes up the double system GEP (DS-GEP) that includes natural evolution system and artificial intervention system. The flowchart of DS-GEP is shown as Fig .1.

A. Fitness function

In statistics, the method to assess the relevance degree between two groups of data usually uses the correlation

coefficient. In this paper the fitness function is devised as: $fitness = R^2 = 1 - SSE / SST$ [19], where

$$SSE = \sum_{j=1}^m (y_j - \hat{y}_j)^2 \quad (1)$$

$$SST = \sum_{j=1}^m (y_j - \bar{y})^2 \quad (2)$$

where, y_j is the observation data; \hat{y}_j is the forecast data which is computed with formula and observation data; \bar{y} is the mean of y ; SSE is the residual sum of squares; SST is the total sum of squares of deviations; m is the size of data.

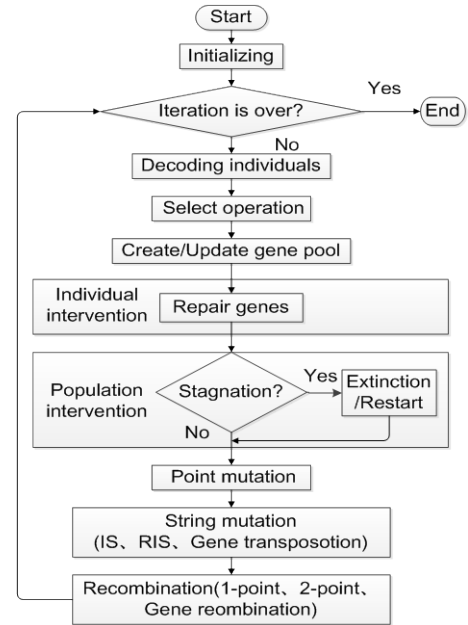


Figure 1. Flowchart of DS-GEP

B. Natural evolution system

The process of natural evolution is completed by standard GEP whose genetic operators mainly include selection operator, crossover operator, point mutation operator and string mutation operator.

The selection operator is the tournament method with elitist strategy. The crossover operation includes single point recombination, 2-point recombination and gene recombination. The point mutation includes the single-point and multi-point mutation; the string mutation includes IS transposition, RIS transposition and Gene transposition [1].

C. Artificial intervention operation

1) *Creating and updating gene pool*: The gene pool is created with the eugenic individuals from initial population: import the genes belonging to the first j eugenic individuals according to the fitness (j is the size of gene pool) into gene pool. The updating of gene pool starts from the second generation: if the genes from the best individual in current generation aren't in gene pool, then import them

into gene pool to replace the genes whose fitness is the lowest.

2) *Individual intervention operation*: some unfeasible individuals could be produced by population initialization or genetic operations, which contain some unfeasible inferior genes, such as the genes including log function whose argument is less than or equal to zero, or the genes which include the information that a number is divided by zero. The traditional method is throwing these individuals out of population, as may lose the superiority genes in individuals. This paper adopts repairing mean that replaces the inferior genes in unfeasible solutions with superior genes in gene pool to change unfeasible individuals into better feasible ones. This method not only could diffuse the high quality genes from the gene pool, but also could retain the feasible and superior genes in unfeasible solutions.

3) *Population intervention operation*: The Standard GEP as the natural evolution process, due to the reduction of diversity, will emerge the phenomenon of stagnation and premature with the evolution. This paper uses extinction/restart strategy to restrain this phenomenon and make population continue healthy evolution by improving population diversity.

The detailed method is as follows: when the best individual of each generation doesn't improve for continuous n (n is an algorithm parameter) generations, the evolution will be thought into stagnation and premature, and then m (m is an algorithm parameter) individuals that are randomly produced will be imported into population to replace the m worst individuals to form new population with high diversity to go on with the evolution and search process.

IV. CONTROL EXPERIMENTS

Experiment 1: In order to test the validity of DS-GEP, DS-GEP is compared with standard GEP about F_1 [2], which is shown as (3):

$$F_1 : 5a_n^4 + 4a_n^3 + 3a_n^2 + 2a_n + 1 \quad (3)$$

In *Experiment 1*, 10 groups of training data are produced by F_1 ($a_n=1, 2, 3, 4, 5, 6, 7, 8, 9, 10$). The algorithm parameters are shown as Tab. 1, and the test results are shown as Tab. 2. It can be seen from Tab. 2 that the optimizing rate of DS-GEP is obviously higher than standard GEP.

TABLE I. THE PARAMETERS OF ALGORITHMS

Option	Parameter
Times of running	50
Max evolution generation	100
Size of population	40
Function set	+-* /
Terminator set	a
Link operator	+
The length of head	4
Number of gene	6

Ration of crossover(1 point and 2 point)	0.2
Ration of point mutation	0.31
Ration of recombination	0.1
Ration of string mutation(IS and RIS)	0.2
Length of IS element	{1,2,3,4,5}
Length of RIS element	{1,2,3,4,5}

TABLE II. TEST RESULTS OF EXPERIMENT 1

Option	DS-GEP	GEP
Times of experiment	50	50
Times of hit	41	35
Optimizing ratio	82%	70%

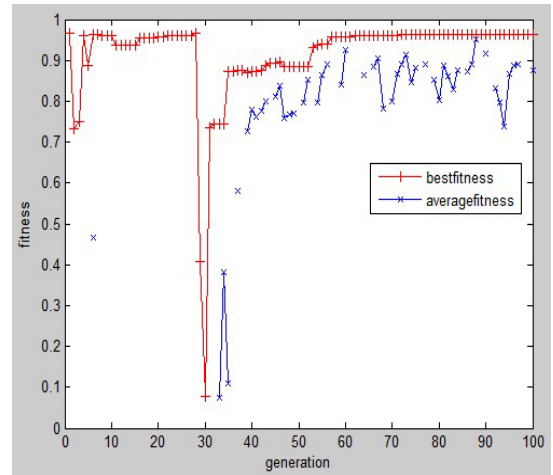


Figure 2. Evolution curve of standard GEP in experiment 1

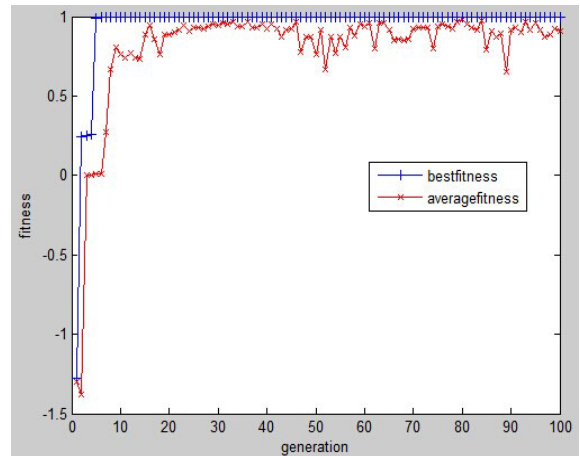


Figure 3. Evolution curve of DS-GEP in experiment 1

The evolution curve of DS-GEP and standard GEP are respectively shown as Fig. 2, and Fig. 3. First of all, it can be found from Fig.2 that the average fitness curve is interrupted in some generations. This indicates that in these generations there are infeasible solutions, whose fitnesses are set to negative infinity, and so the average fitnesses are negative infinity and can't be displayed. Secondly, it can be found from the best fitness curve of Fig.2 that stagnation repeatedly appears in the evolution

process, and especially the search falls into local optima at the 67th generation and can't jump out.

It can be seen from Fig. 3 that DS-GEP finds the global optimal solution at the 10th generation and keep it to the last generation; at the same time, the average fitness curve undulate periodically and isn't interrupted, which shows that the individual intervention operation can repair infeasible solutions to feasible ones and the periodical population intervention operation can improve population diversity effectively.

Experiment 2: In order to test the progressiveness of DS-GEP, an improved GEP (MDC-GEP) [20] is chosen to compare with DS-GEP. Test functions are the same as the functions in [20] so that we could compare objectively. They are shown as (4)-(7). The algorithm parameters are shown as Tab. 1, and the test results of *Experiment 2* are shown in Tab. 3.

$$F_2 : \cos^2(x^2), x \in [0, 3] \quad (4)$$

$$F_6 : \cos(10^x), x \in [0, 2] \quad (5)$$

$$F_7 : \frac{3x^2 + 5x + 1}{5y^2 + 3}, x, y \in [0, 5] \quad (6)$$

$$F_{10} : \frac{\sin x_1 + \cos x_2}{\sqrt{e^{x_3}}} + \tan(x_4 - x_5), x_i \in [0, 2\pi], \quad i = 1, 2, \dots, 5. \quad (7)$$

TABLE III. STATISTICS TABLE OF EXPERIMENT 2

Function	Algorithm	Max fitness	Min fitness	Average fitness
F_2	MDC-GEP	0.9991	0.7865	0.9371
	DS-GEP	0.9937	0.8134	0.9114
F_6	MDC-GEP	0.9954	0.8237	0.9465
	DS-GEP	0.9873	0.8568	0.9372
F_7	MDC-GEP	0.9462	0.8133	0.8956
	DS-GEP	0.9548	0.8475	0.9158
F_{10}	MDC-GEP	0.9771	0.8954	0.9520
	DS-GEP	0.9853	0.9157	0.9647

Tab. 3 shows that the performance of DE-GEP and MDC-GEP is roughly the same for the simple functions F_2 and F_6 . However, as to the complex functions F_7 and F_{10} , DE-GEP shows better performance than MDC-GEP.

V. CONCLUSION

Enlightened by human intervening with natural evolution process, a double system GEP (DS-GEP) is presented by systematically introducing an artificial intervention system into the original natural evolution process of standard GEP. In artificial intervention system, the individual intervention operation can repair the

unfeasible genes in individuals with superior genes from the gene pool which is created and updated by the eminent individuals from population to improve the quality of solution, and the population intervention operation uses extinction/restart strategy to form new population with high diversity to prevent the algorithm premature after the evolution has fallen into stagnation. The control experiments about function finding problems show that DS-GEP can overcome the stagnation and premature convergence phenomenon effectively during the evolutionary process, and promise competitive performance not only in the convergence speed but also in the quality of solution.

ACKNOWLEDGMENT

Support from the National Natural Science Foundation of China (NO. 31170393), the Nature Science Foundation of Shaanxi Province (NO. 2012JM8023), and the Nature Science Special Foundation of Department of Education of Shaanxi Province (NO. 12JK0726) are gratefully acknowledged.

REFERENCES

- [1] C. Ferreira, "Gene Expression Programming: A New Adaptive Algorithm for Solving Problems," *Complex System*, vol. 13, pp. 87-129, February 2001.
- [2] C. Ferreira, "Function finding and the creation of numerical constants in gene expression programming," 7th Online World Conference on Soft Computing in Industrial Applications, August 2002.
- [3] L. Teodorescu, "Gene Expression Programming Approach to Event Selection in High Energy Physics," *IEEE Transactions on Nuclear Science*, vol. 53, pp. 2221-2227, August 2006.
- [4] LUO Songtao, "The War Function Mining Based on GEP and its Application in Military," Chengdu, Sichuan University, October 2006.
- [5] LIU Mengwei, LI Xia, LIU Tao, "A Gene Expression Programming Algorithm for Population Prediction Problems," *Acta Scientiarum Naturalium Universitatis Sunyatseni*, vol. 49, pp. 115-120, June 2010.
- [6] Nor Azazi Zakaria, H. Md. Azamathulla, Chun Kiat Chang, Aminuddin Ab. Ghani, "Gene expression programming for total bed material load estimation—a case study," *Science of the Total Environment*, vol. 408, pp. 5078-5085, June 2010.
- [7] Zhang Yongqiang, Xiao Jing, "A Software Reliability Modeling Method Based on Gene Expression Programming," *Applied Mathematics & Information Sciences*, vol. 6, pp. 125-132, March 2012.
- [8] H. Md Azamathulla, "Gene-expression programming to predict scour at a bridge abutment," *Journal of Hydroinformatics*, vol. 34, pp. 324-331, June 2012.
- [9] Yang Yang, Li Xinyu, Gao Liang, Shao Xinyu, "Modeling and impact factors analyzing of energy consumption in CNC face milling using GRASP gene expression programming," *The International Journal of Advanced Manufacturing Technology(online)*, pp. 1-17, May 2013.
- [10] LONG Wei, ZHANG Xiao-dong, WANG Hao, et al, "Prediction of Enhancement Effect of Nitroimidazoles on Irradiation by Gene Expression Programming," *Chemical Research in Chinese Universities*, vol. 29, pp. 519-525, March 2013.
- [11] Seydou Traore, Aytac Guven, "New algebraic formulations of evapotranspiration extracted from gene-expression programming in the tropical seasonally dry regions of West Africa," *Irrigation Science*, vol. 31 pp. 1-10, March 2013.
- [12] Chih-Hung Wu, I-Sheng Lin, Ming-Liang Wei, and Tain-Yu Cheng, "Target Position Estimation by Genetic Expression Programming for Mobile Robots With Vision Sensors," *IEEE Transactions on Instrumentation and measurement*, vol. 62, pp. 3218-3230, December 2013.

- [13] TANG Changjie, PENG Jing, ZHANG Huan, and ZHONG Yixiao, "Three new techniques for knowledge discover by gene expression programming— transgene, overlapped gene expression and backtracking evolution," *Journal of Computer Applications*, vol.25, pp. 1978-1981, September 2005.
- [14] JIANG Yue, TANG Chang-jie, ZHENG Ming-xiu, et al, "Outbreeding Strategy with Dynamic Fitness in Gene Expression Programming," *Journal of sichuan university (engineering science edition)*, vol. 39, pp. 121-126, March 2007.
- [15] JIA Li-yuan, LI Lei, Xi Sheng-feng, "Application of Immune Algorithm in Gene Expression Programming," *Computer Simulation*, vol. 25, pp. 189-191, March 2008.
- [16] ZHANG Xue-dong, RAO Yuan, YUAN Chang-an, ZHAO Chua-xin, "Improving model accuracy using Gene Expression Programming and Tabu Search," *Computer Engineering and Applications*, vol. 45, pp.35-38, February 2009.
- [17] HU Jian-jun, WU Xiao-yun, "Superior Population Producing Strategy in Gene Expression Programming," *Journal of Chinese Computer Systems*, vol. 8, pp. 1660-1662, August 2009.
- [18] LI Tai-yong, TANG Chang-jie, WU Jiang, et al, "Adaptive Population Diversity Tuning Algorithm for Gene Expression Programming," *Journal of University of Electronic Science and Technology of China*, vol. 39, pp. 279-283, March 2010.
- [19] ZUO Jie, "Core technology research of Gene Expression Programming," Chengdu, Sichuan University, October 2004.
- [20] XUAN Shi-bin, LIU Yi-Guang, "GEP Evolution Algorithm Based on Control of Mixed Diversity Degree," *Pattern Recognition and Artificial Intelligence*, vol. 25, pp. 186-194, April 2012.