International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.9, No.11, (2016), pp.309-318
http://dx.doi.org/10.14257/ijsip.2016.9.11.28

Research and Application of GEP Algorithm Based on Cloud Model

Zhang Rui[1,2],Hou Shasha[1] and Gao Hui[2]

[1]School of Automation, Harbin University of Science and Technology, Harbin, Heilongjiang, China 150080
[2]Chengdu Donglu Traffic Science and Technology Co., Ltd. Chengdu, Sichuan, China 610031
zr_gh@sina.com

Abstract

Aiming at the traditional GEP algorithm adopted fixed rate of mutation and crossover rate in the process of evolution, and ignored the dynamic change of individual fitness, which leaded to the presence of premature convergence and local optimization problem. By using the cloud adaptive strategy and cloud cross strategy of cloud model, a genetic algorithm based on cloud model (Cloud Model Gene Expression Programming, CMGEP) was proposed. The algorithm adjusted the mutation rate and crossover rate in evolution through the cloud adaptation strategy according to the change of dynamic， and timely calculated population similarity to achieve cloud cross to increase the diversity of population and jump out of the premature convergence. It was applied to the field of railway engineering and its results were compared with those obtained by traditional GEP Algorithm and CMGEP Algorithm. Experiments show that the algorithm can improve the adaptability and the prediction accuracy, it has better convergence.

Keywords: *gene expression programming; cloud model; similarity; Subgrade Settlement prediction*

1. Introduction

In the field of railway engineering, subgrade settlement prediction [1] is a critical technology during the process of high speed railway subgrade construction. It has a direct relationship with the safe operation of high-speed railway and service life, so the subgrade settlement prediction is a problem that railway departments focus on . With the widespread concern, an endless stream of new Algorithm, the prediction method which commonly used are curve fitting, gray algorithms, artificial neural networks and genetic algorithms, *etc.*, but the methods that are described above are some flaws in practical applications. For example, the curve fitting and gray algorithm is only suitable for prediction of a small amount of data, neural networks and genetic algorithms can easily fall into the local optimal problem. Gene expression programming [2-3] (Gene expression programming, abbreviation GEP) combines the advantages of genetic algorithm and genetic programming, with the ability to use simple coding and solve complex problems, especially suitable for application in function prediction. Compared with the traditional forecasting methods, GEP algorithm has higher prediction accuracy, but also its own shortcomings in local convergence area.

At present there are many methods for local convergence of GEP algorithm. Li Taiyong [4] proposed a weighted adaptive population diversity methods which is able to improve the efficiency of evolution. Li Qu [5] proposed an adaptive multi performance of GEP algorithm to improve the convergence rate .Li Kangshun [6] proposed a non-coding region inserted through the operator GEP algorithm overcomes local optimization

ISSN: 2005-4254 IJSIP

problem; Zhang Kai [7], who proposed the combination of the SFLA and GEP to solve the problem of premature convergence. Experimental show that these methods can improve the performance of GEP Algorithm, but its own mutation rate and cross rate still can not be adjusted flexibly according to the status of the evolution of the population, remain low population diversity and premature convergence problem.

In this paper, we use strong adaptive and adjust capacity of the cloud model [9], introducing the cloud adaptation strategy and the cloud cross strategy into the traditional GEP algorithm, then CM-GEP algorithm is proposed. In the evolutionary process of GEP Algorithm, according to the dynamic change of population fitness to adjust individual mutation rate and crossover rate; while taking advantage of the similarity to judge the variety of the populations in evolution. Cloud cross when the similarity is greater than the threshold, introduce outstanding individual and improve the diversity of population, so that the GEP Algorithm as soon as possible to jump out of the premature and local optimum, effectively improve the speed of convergence.

2. Basic Concept

Definition 1:Design a functions set of population F={+,-,*,/,^,L,E,Q,S,C,T}, terminals set T={ t,?}, where ^ said power index, L said natural logarithm, E represents index, Q represents the square root, S represents a sine function, C represents the cosine function, T represents tangent function, t denotes times,? denotes Random constant.

(1) the CM-GEP model is five tuple, CM-GEP=(N,A_N,F,T,L), where N is the population size, A_N is the number of genes, and L is the connection function "+" .

(2) CM-GEP fitness function, set n is the sample size, Cij is the prediction result according to the corresponding expression of the i-th individual that take advantage of variable data of the j-th sample requirements, Tj is the actual function value of the j-th sample ,so the individual's fitness function as Formula (1).

$$f_i = 1 \Big/ \left(\sum_{i=1}^{n} \left(C_{ij} - T_j \right)^2 / n \right) + 1 \tag{1}$$

the fitness function is an assessment of the environmental adaptability of the individual in the population, which controls the direction of the evolutionary algorithm, and plays an important role in the GEP algorithm.

Definition 2:Set G for the population, f_i for individual fitness, E_f , En_f and He_f respectively for fitness expectation, fitness entropy and fitness hyper entropy, then:

Set the common set F={ f_i }, $\forall f_i \in R$, R for the real set, T for the language subset on the domain F, the existence of $C_T\left(f\right): F \rightarrow [0,1], \forall f_i \in F$, $f_i \rightarrow C_T\left(f_i\right)$,and f_i to the determination of C to meet the Equation (2), said the distribution of C on the F called the fitness of cloud, denoted as $C\left(E_f, En_f, He_f\right)$,each f_i is called a cloud $Drop\left(f_i, C_T\left(f_i\right)\right)$, where exp is the exponent function. The distribution of fitness cloud is shown in Figure 1.

$$u_i = exp\left(-\frac{1}{2} \left(\left(f_i - E_f \right) \big/ En'_f \right)^2 \right) \tag{2}$$

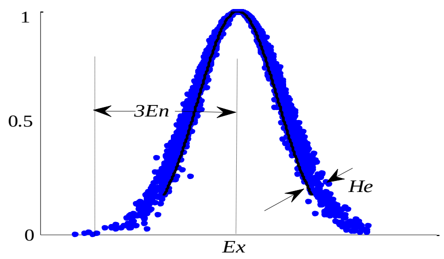

Figure 1. Schematic Diagram of Fitness Cloud Model

Definition 3: Set N as the population size, f_i for the individual fitness, f_{max} is the maximum fitness value; $\overline{f}$ is the average fitness value, and $f_{max} \neq \overline{f}$, so the similarity of population is shown in the Formula (3).

$$S_i = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\left(f_i - \overline{f}\right)/\left(f_{max} - \overline{f}\right)\right)^2} \tag{3}$$

$$\overline{f} = \frac{1}{N} * \sum_{i=1}^{N} f_i \tag{4}$$

The population similarity reflects the diversity of individuals. The higher the similarity, the lower the diversity of individuals, that is, the more similar the individual is, the easier local optimal it is to fall into ; otherwise, the lower the similarity, the better local optimum it is to avoid falling into. Where 0< S_i <1, set a threshold value ε, population similarity is too high for the GEP algorithm to jump out of local convergence when $S_i > \varepsilon$, cloud crossover is used to increase population diversity, so that the GEP algorithm to jump out of local convergence.

3. GEP Algorithm Based on Cloud Model (CM-GEP)

3.1. Cloud Adaptive Strategies

The performance of GEP Algorithm is determined by crossover rate and mutation rate, but the traditional GEP algorithms used fixed rate of mutation and crossover while mining knowledge, ignoring the variation of individual fitness in evolution, hence it is easy to fall into local optimal and premature convergence. GEP Algorithm using cloud adaptive crossover and mutation strategy of cloud model can generate the mutation and crossover rate of cloud generator which is changed with the flexibility of the adaptability, to meet the rapid optimization of ability and avoid the local optimal solution of the problem.

The mechanism of cloud adaptive strategy is: in the early evolution of the population, the rich diversity of individuals, with a fixed rate of mutation rate P_m' and crossover rate P_c'. Later stage of evolution, GEP Algorithm in convergence of the state and for the current fitness is higher than or close to the average fitness of individuals: (I) generating a normal random number En_f' which is denoted by expected value En_f and standard deviation He_f .(II) calculating the crossover rate P_c and mutation rate P_m according to the Formula (5-9) and (10-14) . This approach enables evolution process to continue to generate a meaningful search space without premature convergence at the late stage of genetic manipulation.

Crossover rate (p_c) of the algorithm is shown in the Formula (5-9):

$$E_f = \overline{f} \tag{5}$$

$$En_f = \left(f_{max} - \overline{f} \right) / c_1 \tag{6}$$

$$He_f = En_f / c_2 \tag{7}$$

$$En'_f = RANDN\left(En_f, He_f \right) \tag{8}$$

$$p_c = \begin{cases} k_1 e^{\frac{-\left(f'-E_f\right)^2}{2\left(En'_f\right)^2}}, & f' \geq \overline{f}, \\ P'_c, & f' < \overline{f}, \end{cases} \tag{9}$$

Mutation rate (p_m) of the algorithm is shown in the Formula (10-14):

$$E_f = \overline{f} \tag{10}$$

$$En_f = \left(f_{max} - \overline{f} \right) / c_3 \tag{11}$$

$$He_f = En_f / c_4 \tag{12}$$

$$En'_f = RANDN\left(En_f, He_f \right) \tag{13}$$

$$p_m = \begin{cases} k_2 e^{\frac{-\left(f-E_f\right)^2}{2\left(En'_f\right)^2}}, & f \geq \overline{f}, \\ P'_m, & f < \overline{f}. \end{cases} \tag{14}$$

Among them, f_{max} is the maximum fitness, $\overline{f}$ is the average fitness; f' is the larger fitness value of the two individuals who participated in the crossover, f is the mutation fitness , k_1 and k_2 are constant within [0,1], c_1 □ c_4 are control parameters.

By the Formula (9) and (14), the crossover rate and mutation rate are generated by the normal cloud generator. Normal cloud is a "multi center, two less" distribution, En_f affects its steepness and He_f determines the degree of dispersion of cloud droplets. Therefore, the greater En_f, the greater horizontal width the cloud covered according to the "3 En_f " principle, so more and more excellent individuals can get a smaller crossover and mutation rate. In order to expand the search scope and improve the precision of the algorithm, the algorithm can make the He_f to be large and small. Based on the speed and precision of the algorithm, the experiment is carried out $c_1 = 2.9$, $c_3 = 3.0$, $c_2 = c_4 = 10$ in this paper.

3.2. Cloud Crossover Strategy

In GEP algorithm, with the increase of the evolution generation, the diversity of the population decreases and the similarity increases gradually, this single gene model will not only slow down the evolution ,but also lead to the stagnation of evolution, the most important situation is that single gene model weaken local optimum prematurely ,then the search performance of the algorithm will not high . Therefore, GEP algorithm uses cloud crossover strategy of cloud model when the threshold is exceeded by population similarity and selects excellent individuals by evaluating the fitness of stepfather and their offspring fitness after crossover. If the fitness of the two generation individuals is different from that of the two parents, the offspring of the individual is a valid individual, otherwise it is invalid, this cross is called cloud cross. GEP algorithm increases population diversity by the cloud crossing strategy, and it can accelerate the evolution process and avoid the local optimization problem.

The basic idea of CMGEP algorithm using the cloud crossover strategy is that: in the early stage of evolution, the population diversity is rich, and the algorithm adopts normal crossover mode .When there is no evolution in the population, the individual is too similar to the new individual, it shows that there is no effective individual and need to introduce cloud crossover strategy to operate: (1)when the highest fitness of successive T is kept constant and the population similarity is $S_i > \varepsilon$, the cloud crossover is used by the individuals of the population;(2) the offspring of the cross generation and the non-substituted parent are all involved in the evaluation and selection. According to the sort of fitness, the greatest fitness of individuals are selected to form a new group.

4. Prediction Model of CMGEP Algorithm

In order to verify the effectiveness and efficiency of CMGEP algorithm, this paper selects 21 groups of subgrade settlement data which were from a certain section of Subgrade the second high iron project of China's Lanxin line, from to the DK649+500D K650+000 pier as the experimental data sets. In the experiment, 15 groups of railway subgrade settlement data were selected as training data, and the 6 groups were predicted as the test data. The parameter settings of GEP algorithm are shown in Table 1.

Table 1. Parameter Settings

Parameter name	Parameter value	Parameter name	Parameter value
Evolutionary generation	1000	IS transposition rate	0.1
Population size	40	RIS transposition rate	0.1
Gene number	5	Gene transposition rate	0.1
Head length	6	Inversion rate	0.01
Linking function	+	Mutation rate	0.04 or Dynamic mutation rate
Fitness function	See Formula (1)	Single-Point recombination Rate	0.3 or Dynamic crossover rate
Selection operator	Roulette wheel selection	Two-Point recombination Rate	0.3 or Dynamic crossover rate

4.1. Prediction Model

According to the above parameters, CMGEP algorithm can get a better prediction model after several runs, and select the largest fitness of chromosome as the optimal solution. In this experiment, the fitness of the optimal solution is 0.99103. and the chromosome is：QS-+*a?a?aa??aa?a??SCE-E*?a?aaa?????a?SSS?a???aaa???a???a ST^LS-aa??aaa??a?aa?Q-L*C*?aa?aa?a???a.

The chromosome is transformed into an expression and simplified. Then the expression is used the "+" linking function to connect each other, and the subgrade settlement prediction formula as shown in Formula (15). t represents of time in the formula.

$$Y=-0.491+\sqrt{\left|sin(\frac{t}{-0.662})+0.205*t\right|}+sin\left(cos\left(2.905+0.205*t\right)\right) \quad (15)$$

4.2. Comparative Analysis of Prediction Results

A series of experiments were carried out to verify the feasibility of GEP algorithm and the superiority of CMGEP algorithm. The comparison of evolution generation and average fitness of GEP and CMGEP algorithms is shown in Figure 2.

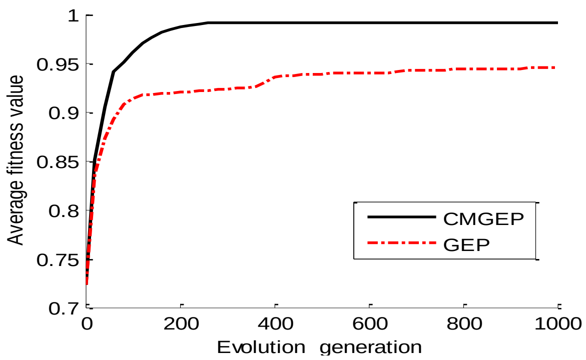

Figure 2. Comparison of Average Evolutionary Fitness of the Two Algorithms

Figure 2, shows that in the early stages of population evolution, the evolutionary curves of GEP and CMGEP algorithms are basically similar. GEP algorithm began to converge when it reaches a certain value, and cannot jump out of the local area. However CMGEP algorithm can promptly adjust the rate of mutation and crossover according to the changes of individual fitness, and the cloud cross is used to increase the diversity of the population, so that the individual fitness can be further improved. Thus, CMGEP algorithm is superior to GEP Algorithm, it was about 4.8% better than average fitness of GEP Algorithm and has strong global optimization ability. The feasibility and superiority of the CMGEP algorithm used in subgrade settlement prediction are proved.

Generally, prediction function is evaluated from two aspects, One is the fitting degree of the training set prediction, the two is the error of the test set prediction. Figure 3, shows the prediction of the two algorithms training data, Figure 4 shows the relative error of the predicted values of the two algorithms training data. Combined with the two figure, we can see that the prediction of CMGEP algorithm training data is more close to the true value, and can evolve closer to the perfect solution.

Figure 5, shows GEP algorithm and CMGEP algorithm for the prediction of the test set. From the figure can be seen that the predicted value of CMGEP algorithm is more close to the true value. That is to say, CMGEP algorithm can overcome the local optimization problem, and its prediction accuracy is more accurate than GEP algorithm, which shows the excellent performance of the algorithm.

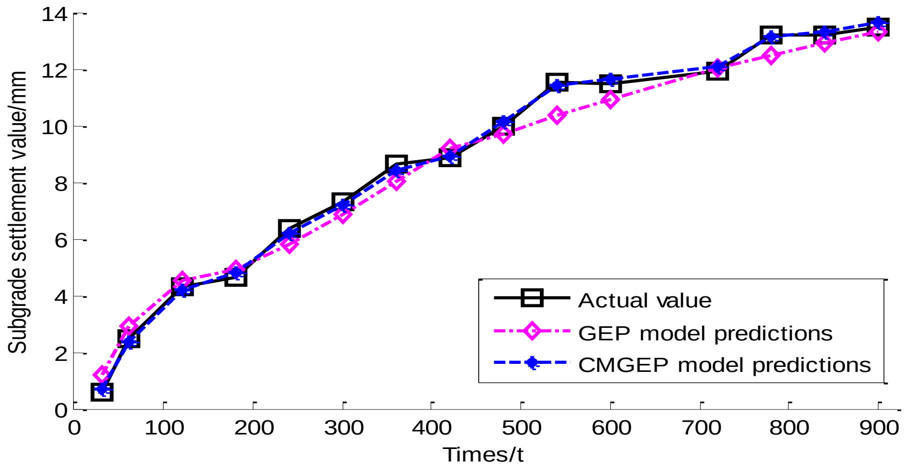

Figure 3. Comparison of Prediction Value of GEP and CMGEP Algorithm Training Data

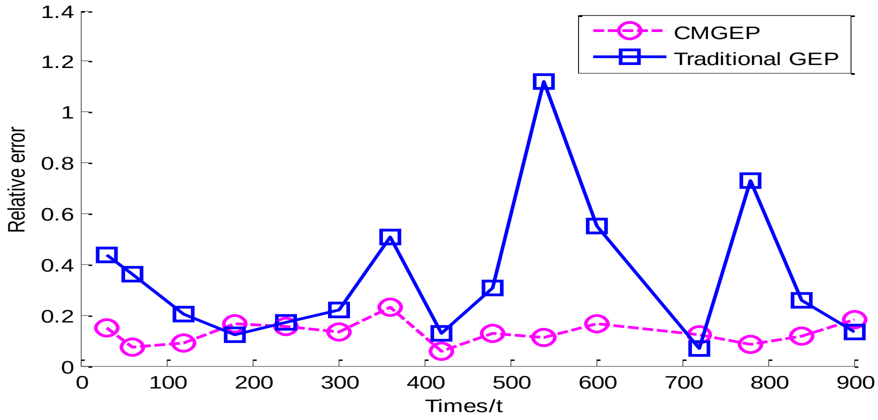

Figure 4. The Relative Error of the Predictive Value of GEP and CMGEP Training Data

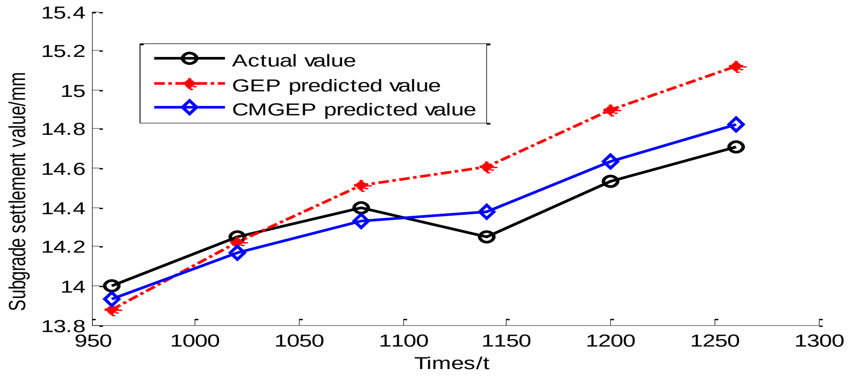

Figure 5. Comparison of Predictive Value of GEP and CMGEP Test Set

The above results are compared with the single experiment. In order to obtain more accurate prediction results, the method which is a number of times to do the experiment to get the average value is applied to the experiment. Under the premise of the experimental conditions, the 10 prediction experiments were carried out, and then the average results of the predicted values were taken, and GEP algorithm and CMGEP algorithm were evaluated by the error of the predictions. After experiment, the average value of the predictions of the two algorithms are obtained. The results are shown in Table 2.

Table 2. Prediction Results of GEP and CMGEP Subgrade Settlement

Serial number	Real value	CMGEP	GEP
16	14.00	13.934	13.797
17	14.25	14.069	14.230
18	14.40	14.300	14.630
19	14.25	14.409	14.887
20	14.53	14.645	15.003
21	14.71	14.880	15.391
Absolute error		0.13	0.37
Relative error		0.94%	2.6%

From the Table 2, the average absolute error of CMGEP prediction was 0.13 (<0.2),the mean relative error is 0.94% (<1%), comparing to the absolute error and relative error of GEP algorithm which improves the 0.24 and 1.66%.Experiments show that the algorithm of CMGEP is better than traditional algorithm of GEP. The algorithm of CMGEP can jump premature convergence to obtain more accurate predictions in the superiority of the settlement prediction of roadbed.

5. Conclusion

Traditional GEP algorithm uses a fixed rate of mutation and crossover rates to affect the population evolution rate of convergence and accuracy of forecasting. In this paper, via the cloud model is introduced in GEP algorithm, putting forward to a cloud model GEP algorithm (CMGEP).In the evolutionary stage, according the crossover rate and mutation rate to adjust the fitness of dynamic changes, a cloud adaptation strategy is proposed to overcome the local optimum, though using the population similarity to judge diversity, and putting forward the strategy which can increase the effective individual. Compared with the traditional GEP algorithm, the CMGEP algorithm improves the convergence rate and precision, and it has better prediction ability. In the future work, we will study the application of CMGEP algorithm besides function prediction in other fields.

Acknowledgments

This work was supported by Science and Technology Supporting Plan Project of Sichuan Province of China (No. 2015FZ0056).

References

[1] Q. Tan, L. Li and S. Wang, “Measurement and Analysis of High-speed Railway Subgrade Settlement in China: A Case Study”, Sensors & Transducers, vol. 170, no. 5, **(2014)**, pp. 184.
[2] F. C. Gene, “Expression Programming: Mathematical Modeling by an Artificial Intelligence”, (2nd Edition), Springer, **(2006)**.
[3] J. Zhong, L. Luo and W. Cai, “Automatic rule identification for agent-based crowd models through gene expression programming”, Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems. International Foundation for Autonomous Agents and Multiagent Systems, **(2014)**, pp. 1125-1132.

[4] T. Y. Li, C. J. Tang and J. Wu, "Adaptive Population Diversity Tuning Algorithm for Gene Expression Programming", Journal of University of Electronic Science and Technology of China, vol. 39, no. 2, **(2010)**, pp. 279-283.
[5] Q. Li, C. Chen and W. Wang, "An adaptive multi-phenotype GEP algorithm", Natural Computation (ICNC), 2012 Eighth International Conference on. IEEE, **(2012)**, pp. 631-634.
[6] K. S. Li, H. D. Cheng and W. S. Zhang, "NCRS - GEP algorithm and its application in statistical problems", Journal of Wuhan University of Technology, vol. 36, no. 12, **(2014)**, pp. 122-130.
[7] K. Zhang and H. G. Hu, "GEP based on frog leaping and it's application in the function founding", Industrial Control Computer, vol. 10, **(2015)**, pp. 83-85.
[8] Z. Xu, P. Zhu and J. Mei, "Research on backward cloud model with uncertainty", ISSN 1796-2056 vol. 8, no. 11, **(2013)** November, pp. 2556.