International Journal of Multimedia and Ubiquitous Engineering
Vol.11, No.9 (2016), pp.355-362
http://dx.doi.org/10.14257/ijmue.2016.11.9.35

Application of GEP Algorithm in Prediction of Subgrade Settlement

Gao Hui[1,3] and Huang Jun[2,3]

[1]School of Electrical and Information Engineering, Heilongjiang Institute of Technology, Harbin, Heilongjiang, China 150050
[2]Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu, Sichuan, China 610031
[3]Chengdu Donglu Traffic Science and Technology Co., Ltd. Chengdu, Sichuan, China 610031
Gaohui_dl@sina.com

Abstract

The development law of subgrade settlement is highly nonlinear, unstable and uncertain, so the prediction of subgrade settlement is a difficult problem. In this paper, a prediction model of subgrade settlement based on gene expression programming algorithm (GEP) is proposed, and the algorithm compared with the gray level GM (1,1) model and BP neural network model. The results show that the prediction accuracy of subgrade settlement prediction model that is put forward is better than that of GM (1,1) model and BP neural network model, which provides a new method for subgrade settlement prediction and is of practical significance.

***Keywords**: subgrade settlement; prediction model; Gene expression programming; GM model; BP neural network model*

1. Introduction

Subgrade is the foundation of highway and railway, which directly affects the quality of construction, and relates to the safe and efficient operation of highway and railway. Subgrade settlement is an important factor which influences the quality of highway and railway construction [1]. In the construction process to ensure construction quality we must promptly grasp the changes of the subgrade settlement and reasonable prediction of subgrade settlement, building roadbed settlement prediction model according to the observation data has important meaning.

Based on observation data, the settlement model of subgrade [2] is mainly included: the empirical formula method, gray system method and neural network method. Above methods exist a variety of problems in applications, it results in each method has its special applicable scope. In order to overcome the defects of the existing forecasting methods in common use, people do a lot of research on the subgrade settlement prediction model: Such as in literature [3] using genetic algorithm and artificial neural network which is constructed by the distance of genetic neural network model for the prediction of subgrade settlement. Genetic algorithm is easy to appear local optima problem because of the diversity of the population problem; Literature [4] proposed GM gray algorithm for highway subgrade settlement deformation prediction. The algorithm can better reflect the trend of deformation for short-term deformation monitoring, and convergence is poor for the long-term deformation data; By analyzing the expression of Weibull model, this paper [5] analyzes the sensitivity of the subgrade settlement, and puts forward a good forecasting model, but the model has too strong dependence for data.

In this paper, it’s based on the expression into the GEP algorithm which has powerful

ISSN: 1975-0080 IJMUE

function of mining capacity and nonlinear modeling capabilities [6-8] to propose a gene expression programming algorithm (GEP) forecast of subgrade settlement prediction model. Through the original sample data, the model is to find the optimal nonlinear function, and the prediction model based on the observed data is obtained. The validity of the prediction model is verified by the engineering example, and the correlation prediction model is compared with it, which shows that the proposed method has higher prediction accuracy and better practical value.

2. GEP Principle

Gene expression programming (called GEP)which is an evolutionary algorithm that can be used for reference to the law of gene expression in biological genetics by Portuguese scientists Candida Ferreira in 2001. GEP algorithm is developed on the basis of genetic algorithm (GA) and genetic program (GP), which is a new type of genetic algorithm for the separation of genotype and phenotype. GEP individuals is formed by multiple fixed length gene of linear strings, and then these individual represented as expression trees (Expressing trees, ETS), which overcome the lack of functional complexity of GA system and the genetic manipulation is difficult to carry out in GP system, so as to obtain the better function discovery and prediction ability, and greatly enhance the ability to solve subgrade settlement prediction of high-speed railway. The relationships between GEP, GA and GP are shown in Figure 1.

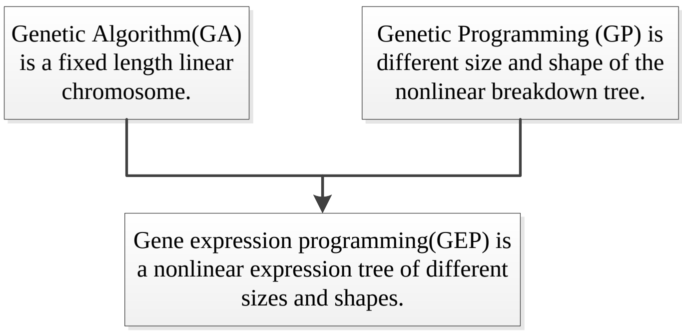

Figure 1. Relationship between GEP, GA and GP

2.1. GEP Code

GEP coding of each gene is a fixed length of the linear symbol string, known as GEP chromosome. A chromosome can be composed of a number of genes, the gene is composed of the effective part of K expression and the filling part. In order to ensure the effectiveness of the gene encoding, encoding is divided into three parts of the head , the tail and constant domains e. The head is formed by function symbols (FS) and end symbol (TS) ,and the tail can only contain terminal symbols. Function symbols including +, -, *, /, sin, cos, tan, log, exp, *etc.* And terminators include program variables, constants, and functions with no parameters .The length of the head and tail of the gene are expressed by symbol of h and t, in which the head length h is generally set by the user, and the tail length t can only be obtained by the formula, as shown in the formula (1).

$$t = h*(n-1)+1, \tag{1}$$

Where h represents the length of the head of the gene, T represents the tail length, n represents the number of parameters that require the most parameters of the function symbols.

GEP encoding rules: according to the semantic the expression is converted to an expression tree (ET), in accordance with the order from left to right one by one to read the gene character, then according to the rules of grammar the expression trees are traversed layer by layer from top to bottom, and left to right .Then this sequence of symbols is k-expression. For example: function is $f=\sqrt{(a+b)*(c-d)}$, function set F={Q, +, -, *}, end point set T={a, b, c, d}, Expression tree ET as shown in Figure 1.

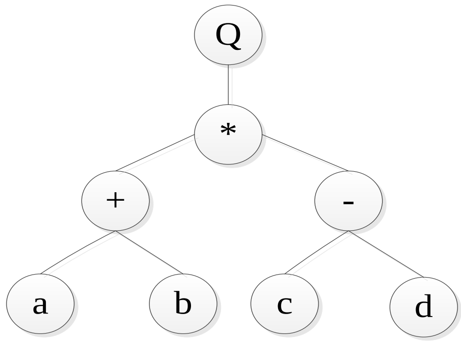

Figure 1. Expression Tree (ET)

The K- expression obtained after layer by layer traversal is: "Q*+-abcd". Obviously n=2, assuming h=4, then the tail length t=5, then gene length is 9.Because k-expression only has 8-bit coding region, the remaining portion of the tail are filled by random elements of the terminator said non coding regions, so the gene coding is" Q*+ abcd**a**", Where the tail elements are represented in black.

2.2. Fitness Evaluation

The fitness function is an assessment of the environmental adaptability of individuals in the population, which controls the direction of the evolutionary algorithm. Two error based evaluation models of traditional GEP algorithms have their own inherent shortcomings. In statistics, it is more to use the coefficient of determination to evaluate the degree of agreement of the settlement data of the railway subgrade shape with time. The coefficient is recorded as $R^2(R-Squared)$, and the calculation formula of which is shown in formula(2).

$$R^2=1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\overline{y})^2} \tag{2}$$

Where y_i is the actual observed value, $\overline{y}$ is the average value of the observed value, and $\hat{y}_i$ is the predicted value. R^2 indicates the fitting degree of the predicted value to the actual data, the value range of which is [0,1]. It is closer to the observed value and the predication performance is better when R^2 is closer to 1.

2.3. The Flow of GEP Algorithm

GEP algorithm is similar to the genetic algorithm, and its concrete steps are as follows:

Step one, the initial population is randomly generated. and the individual in the population is a linear string with fixed length, and the symbol of the string is composed of the function and the ending point of the problem.

Step second, chromosome decoding. The process of the chromosome decoding which uses an expression tree to represent each individual, and then to execute each program, and to evaluate their fitness.

Step third, retaining the current best individual. The roulette selection method is used to select individuals, and then to retain the best individual.

Step fourth, performing genetic operations .It includes replication, variation, interpolation, recombination, and so on.

Step fifth, the algorithm to terminate the judgment. When the loop condition of the algorithm satisfies the stopping condition, the program outputs the optimal solution and exits the algorithm; otherwise, it returns to the second step.

The flow chart of the algorithm is shown in Figure 2.

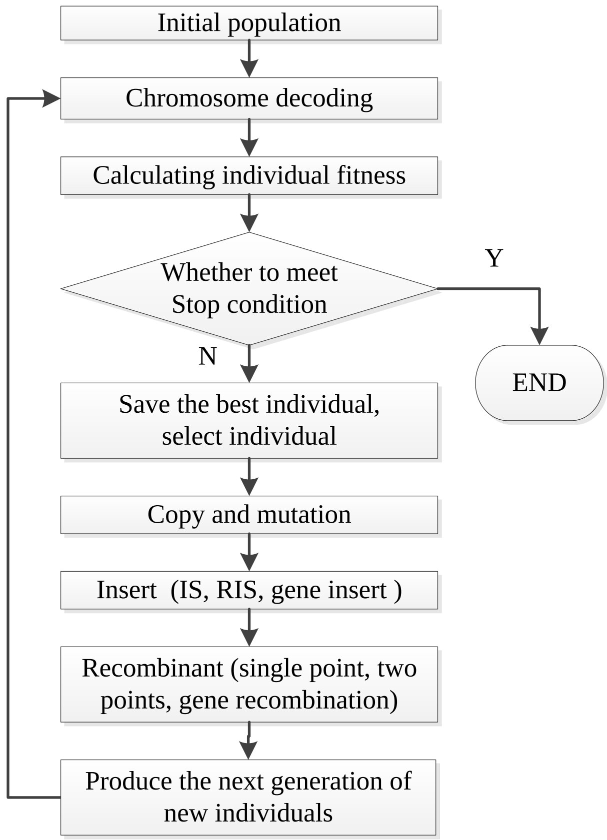

Figure 2. Flow Chart of GEP Algorithm

3. Prediction Model of GEP Algorithm

To verify the GEP algorithm effectiveness compared to traditional prediction algorithms, this paper selected 21 groups of settlement data from LanXin Dual high-speed rail project as the experimental data set. At the same time select the traditional GM (1,1) algorithm and BP neural network algorithm as a compared predictive model experiment, mean absolute error and average relative error were calculated to evaluate the effect of error prediction.

3.1. Parameter Settings

With VC++ and Matlab as the experimental platform. Based on GEP algorithm of subsidence data modeling, mainly has the following five steps:

First, select the fitness function, select determining factor R^2 (See equation 2)

The second step, select the set of functions F and terminators set T to produce chromosome. F={+, -, *, /, ^, L, E, Q, S, C, T} , Wherein ^ represents exponent, L represents the natural logarithm, E indicates an exponent, Q represents prescribing, S represents the sine function, C represents a cosine function, T represents the tangent function. Terminator set T={t}, where t is the variable that represents the number of days of observation.

The third step is to determine the structural gene, the length of the gene's head is 6, and the length of the tail is 7. *i.e.* the length of each gene is 13.

The fourth step is to select the connection function, the sub expression tree for each genetic makeup is connected with "+" to form a chromosome.

The fifth step to determine the genetic operators parameters. GEP algorithm parameters shown in Table 1.

Table 1. GEP Algorithm Parameters

parameter name	Parameter Value
The maximum evolution generation	1000
Population size	40
String interpolation rate (IS)	0.1
Root string interpolation rate (RIS)	0.1
Gene interspersed rate	0.1
Inversion probability	0.01
Mutation rate	0.044
Single-point recombination rate	0.3
Double-point recombination rate	0.3
Select operator	Roulette wheel selection

3.2. Subgrade Settlement Formula

It is found that the maximum fitness for the 1000 generation is 0.991230, and the optimal prediction function expression of the output is shown in the formula (3). Where t represents the number of observations.

$$\begin{aligned} Y = 0.064180 &+ sqrt(abs((\sin((t./(-0.661794)))-(t.*(-0.205176))))) \\ &+ \sin(\cos((\exp(\exp(0.64180))-(t.*(-0.205176))))) \\ &+ \sin(\tan(10.\wedge(\log(abs(\sin(t-t))))))+\sin(\sin(\sin((-0.68047)))) \end{aligned} \quad (3)$$

3.3. Model Evaluation

The prediction results of GEP model are compared with the gray model GM(1,1) and BP neural network model (due to the prediction model requires a lot of computing data, not listed herein), and the model prediction results are shown in Table 2.

The Table 2 shows that the MAP of GEP prediction model is 0.187, MAPE is 4.44% , is better than the MAP and MAPE of GM (1,1) prediction model and BP neural network prediction model. It shows that the predicted value of GEP algorithm prediction model is close to the actual value, and the prediction accuracy is the highest. In order to better compare the fitting results of the three models, We draw the comparison chart according to the actual data for the predicted results of the three models, the comparison of three predicted model and actual value shown in Figure 3.

From Figure 2, compared with GM (1,1) prediction model and BP neural network model, the predicted value of GEP model is more close to the measured value, and the fitting degree is high, and the prediction accuracy of BP neural network is equal to that of the prediction model. Figure 3 shows that GEP prediction model of the relative error is smaller, the curve of model in near zero value fluctuates and amplitude is smaller, it is more accurate than the other two models for forecasting precision.

Table 2. Prediction Results of Three Models

Sample	Measured value (mm)	GEP model (mm)	GM model (mm)	BP model (mm)
1	0.57	0. 891	1.152	0.767
2	2.45	2.379	2.604	2.32
3	4.30	4.258	3.973	4.35
4	4.62	4.987	5.259	4.65
5	6.36	6.012	6.462	6.042
6	7.32	7.109	7.581	7.017
7	8.66	8.334	8.617	8.239
8	8.85	9.246	9.570	9.436
9	10.00	10.179	10.440	9.783
10	11.55	11.337	11.226	11.247
11	11.49	11.523	11.929	11.432
12	11.94	12.379	12.549	12.434
13	13.22	13.183	13.086	13.181
14	13.20	13.318	13.539	13.00
15	13.46	13.643	13.909	13.542
16	14.00	14.201	14.196	13.959
17	14.25	14.316	14.399	14.155
18	14.40	14.695	14.519	14.249
19	14.25	14.479	14.556	14.337
20	14.53	14.674	14.510	14.336
21	14.71	14.861	14.38	14.275
	MAE	0.187	0.397	0. 244
	MAPE	4.4%	9.51%	4.74%

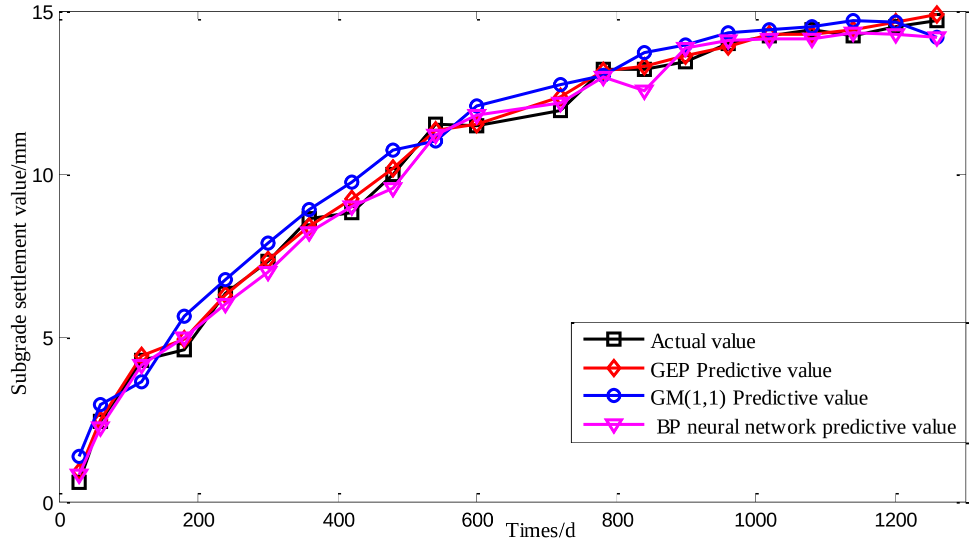

Figure 3. Comparison of the Three Predicted Model and the Actual Value

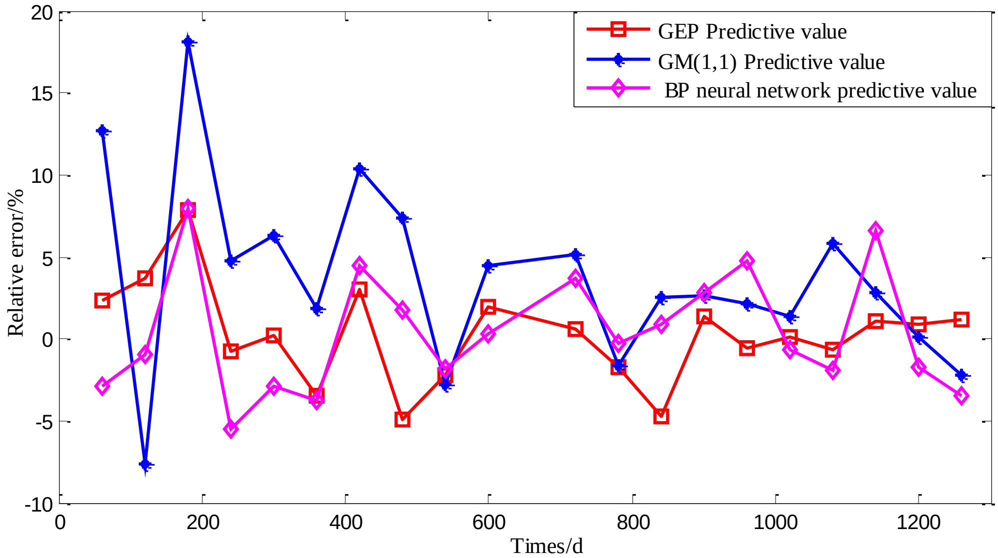

Figure 4. Relative Error of Three Prediction Model

4. Conclusions

In this paper, based on the existing problems of commonly used subgrade settlement prediction model, using the GEP algorithm to establish the prediction model of subgrade settlement. It is characterized by the full use of data mining capacity and nonlinear modeling ability of GEP algorithm, and which uses the observation data of subgrade settlement to reveal the evolution law of subgrade settlement, and then realize the prediction of subgrade settlement. Example analysis shows that, compared with conventional prediction model of subgrade settlement, the proposed method has better predictive power, and its predictive accuracy is also higher than other forecasting models.

Acknowledgements

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

References

[1] S. X Chen, X. G. Wang and L. F. Jiang, "Settlement characteristics and engineering significance of subgrade surface for railway passenger dedicated line", Rock and Soil Mechanics ,vol. 31, no. 3, **(2010)**, pp. 702-706.
[2] W. Z Liu, H. Y. Dai and Q. Wang, "The Programming of Prediction for Subgrade Settlement of High-speed Railway and Analysis of Instance", Bulletin of Surveying and Mapping, vol. 3, no. 4, **(2013)**, pp. 62-64.
[3] Y. Y. Guo, L. G. Sun and Z. R. Su, "Application of Improved BP Neural Network in Prediction of Settlement of Roadbed", Investigation of science and technology, vol. 12, no. 5, **(2010)**, pp. 28-31.
[4] S. Hu, "Subgrade Settlement Prediction of Highway based on Grey Model", Highways & Automotive Applications, vol. 5, no. 1, **(2013)**, pp. 115-117.
[5] T. Zhang, G. J. Cai and Y. S. Liu, "Settlement prediction of soft ground based on Weibull model", Chinese Journal of Geotechnical Engineering, vol. 35, no. 2, **(2013)**, pp. 212-216.
[6] C. Ferreira, "Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence (2nd Edition) [M]", Springer, **(2006)**.
[7] J. Zuo and J. C. Tang, "Time series prediction based on gene expression programming", International Conference for Web Information Age, **(2004)**, pp. 1125-1132.
[8] D. B. Zhang, J. G. Zhang and S. Peng, "Application of gene expression programming on combination forecasting modeling", Systems Engineering-Theory & Practice, vol. 32, no. 3, **(2012)**, pp. 568-573.

