Back-calculation of Asphalt Pavement Modulus based on Gene Expression Programming

Yue Hu¹, Ke-zhen Yan², Ying-bin Hu³

¹Master Student, College of Civil Engineering, Hunan University, Changsha, Hunan 410082, China, yuehu@hnu.edu.cn
²Professor, College of Civil Engineering, Hunan University, Changsha, Hunan 410082, China, yankz2004@163.com
³Master Student, College of Civil Engineering, Hunan University, Changsha, Hunan 410082, China, huyingbin@hnu.edu.cn

ABSTRACT: This paper explored the gene expression programming (GEP) to model the back-calculation of asphalt pavement structure for surface layer modulus, base modulus and subgrade modulus, and obtained the function expression correspondingly. Then comparatively analyzed the deviation between theoretical modulus and back-calculation modulus. The motivation theory deflection data used in the analysis was developed from the finite element method under the falling weight deflectometer (FWD) load. Finally, the result showed that it had good precision using GEP algorithm to back-calculate the modulus of asphalt pavement structure layer. The method could not only effectively overcome the shortcomings of traditional method and lowly be affected by initial value, but also could obtain the aim function expression and be convenient for engineering application.

KEY WORDS: Gene expression programming; Modulus back-calculation; Asphalt pavement structure layer

INTRODUCTION

The old pavement layer modulus as one of the considerable design parameter should be determined reasonably for pavement design and rehabilitation. The pavement layer moduli can be obtained by back-calculation method based on pavement deflections tested by the FWD. Hence, it is significant to back-calculate the pavement layer moduli utilizing the deflection values effectively and accurately. Actually back-calculating pavement modulus is a complex nonlinear optimization problem, and can be solved by the mathematical analysis or mechanical analysis [1,2]. The accuracy, uniqueness of the numerical solutions and applicability of the selected mechanics model have become the focus of current research objectives [3]. At present, various methods for back-calculating moduli have been proposed, such as diagram method, regression analysis, iterative method, database retrieval (DBR), genetic algorithms (GAs) and artificial neural networks (ANNs) and so on. However, these methods have their limitations, respectively. The accuracy and versatility of diagram method and regression analysis are dissatisfactory and the database retrieval are highly affected by the initial value, and affection of mesh generation to artificial neural network is not universal [4].

Over the last few decades, machine learning has attracted much attention in both empirical and academic fields for solving civil engineering problems. Genetic programming (GP) is a widely used machine learning method [5]. It has been used to
tackle problems as an extension of GAs. GP solves problems based on the principle of Darwinian nature selection, evolves through the action of genetic operators such as reproduction, crossover and mutation. And GEP is a new descendent of GP. It evolves computer programs of different shapes and sizes encoded in linear strings of fixed size. GEP can be utilized as an effective change to classical GP. There are some applications of GEP in terms of function mining and forecasting time series [6-7].

In this study, the back-calculation models of asphalt pavement layer moduli are established by the function mining in GEP combined with the deflection response of asphalt pavement structure under the FWD. It was indicated that GEP can be utilized for solving civil engineering problems.

GENE EXPRESSION PROGRAMMING

GEP was first invented by Ferreira [8] who utilized the gene expression pattern in biological heredity based on genetic algorithms and genetic programming. There are five main components in GEP: function set like \([+, -, \times, ÷]\), termination set like \([a, b, c, d]\), fitness function, control parameters and termination condition. Unlike the parse-tree representation in GP, GEP uses a fixed length of character strings to represent solutions to the problems. Then the solutions are expressed as computer models in tree-like structures parse trees which are called GEP expression trees (ETs), and the ETs has different sizes and shapes.

\[
01234567\quad \sqrt{(a + b) \times (c - d)}
\]

\[
q^*+abcd
\]

\[
\downarrow
\]

\[
q
\]

\[
+\quad -
\]

\[
a\quad b\quad c\quad d
\]

Figure 1. The expression tree.

\[
q^*+abcd
\]

\[
\downarrow
\]

\[
q
\]

\[
+\quad -
\]

\[
a\quad b\quad c\quad d
\]

\[
01234567\quad q^*+abcd
\]

Figure 2. The \(K\)-expression.

There are two main parameters in GEP: the chromosomes and the ETs. We call the procedure of information decoding from the chromosomes to the ETs as translation. And the translation is based on a set of rules [9]. The genetic code operator is very simple: there is a one-to-one relationship between the symbols of the chromosome and the functions or terminals they represent. Both language of the genes and the language of ETs are used in GEP for the rules simply to determine the spatial organization of the functions and terminals in the ETs and the type of interaction between sub-ETs. It is possible to infer the phenotype given the sequence of a gene immediately thanks to the simple rules that determine the structure of ETs and their interactions, and vice versa. This is called Karva notation or a \(K\)-expression. The Fig.1 to 2 show examples of ETs and \(K\)-expression, respectively.
The genetic operators in GEP consists of nine main components: replication, mutation, pour string, insertion sequence, root insertion sequence, gene transposition, one-point recombination, two-point recombination, gene recombination [8]. They iterate on a particular method utilize the character strings of genes. Such as mutation occur in the chromosome anywhere. It can change any symbol to another in the heads, but in the tails only terminals can change into terminals. And an entire gene is exchanged during crossover in gene recombination. The exchanged genes occupy the same position in the parent chromosomes and are randomly chosen. One important application of GEP is symbolic regression, where the goal is to find an expression that performs well for all fitness cases within a certain error of the correct value. The basic evaluation criteria of GEP relies on the fitness function. The larger the fitness value, the higher the back-calculation accuracy. There are some fitness functions: absolute error, relative error, etc. Eqs. 1-5 are their expressions:

\[ f_k = \sum_{k=1}^{m} \left( M - |y_k - \bar{y}_k| \right) \]. \hspace{1cm} (1)

\[ f_k = \sum_{k=1}^{m} \left( M - \left| \frac{y_k - \bar{y}_k}{\bar{y}_k} \right| \times 100 \right) \]. \hspace{1cm} (2)

fitness \( = R^2 = 1 - \frac{SSE}{SST} \). \hspace{1cm} (3)

\[ SSE = \sum_{k=1}^{m} (y_k - \bar{y}_k)^2 \]. \hspace{1cm} (4)

\[ SST = \sum_{k=1}^{m} (y_k - \bar{y})^2 \]. \hspace{1cm} (5)

where M is the range of selected, \( y_k \) is the k set of measured value and \( \bar{y}_k \) is the back-calculation value, R is correlation coefficient, SSE is the sum-of-squared residuals and SST is the total sum-of-squared deviations, \( \bar{y} \) is average of all data.

**CALCULATION AND DISCUSSION**

**Experimental database**

A total 100 sets of experimental data used in this study were from a previous research [10]. In order to ensure the accuracy of the back-calculation model, a database of road surface deflection basin and pavement layer parameters was firstly created by the three-dimensional dynamic finite element method. The pavement layer parameters were listed in Table 1. Different kinds of asphalt pavement layer parameters were assembled by changing surface layer thickness, base thickness, surface layer modulus, base modulus and subgrade modulus respectively.
Table 1. Parameters of pavement layer.

<table>
<thead>
<tr>
<th>Pavement layer</th>
<th>Variable parameter value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface layer thickness[m]</td>
<td>0.12, 0.14, 0.16, 0.18, 0.20, 0.22, 0.24</td>
</tr>
<tr>
<td>Surface layer modulus[MPa]</td>
<td>1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, 2800, 3000, 3200, 3400, 3600, 3800, 4000</td>
</tr>
<tr>
<td>Base thickness[m]</td>
<td>0.20, 0.24, 0.28, 0.32, 0.36, 0.40, 0.44</td>
</tr>
<tr>
<td>Base modulus[MPa]</td>
<td>400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, 2800, 3000, 3200, 3400</td>
</tr>
<tr>
<td>Subgrade modulus[MPa]</td>
<td>30, 45, 60, 75, 90, 105, 120, 135, 150</td>
</tr>
</tbody>
</table>

Modeling

The data of GEP was divided randomly into two different subsets: the training data subset containing 80 data points (80%) and the testing data subset consisting of 20 data points (20%). The back-calculation models of the surface layer modulus, the base modulus and the subgrade modulus were established respectively because deflection in each location had affected by the pavement layer parameters differently. Nine input parameters, $C_1$-$C_7$, $H_1$ and $H_2$, were used to create the GEP model. They were listed in Table 2.

Table 2. The input and output variables of GEP model.

<table>
<thead>
<tr>
<th>Model name</th>
<th>Output</th>
<th>Input</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEP-1</td>
<td>$E_l$</td>
<td>$H_1$ $H_2$ $C_1$ $C_2$ $C_3$ $C_4$ $C_5$ $C_6$ $C_7$</td>
</tr>
<tr>
<td>GEP-2</td>
<td>$E_b$</td>
<td>$H_1$ $H_2$ $C_1$ $C_2$ $C_3$ $C_4$ $C_5$ $C_6$ $C_7$</td>
</tr>
<tr>
<td>GEP-3</td>
<td>$E_s$</td>
<td>$H_1$ $H_2$ $C_1$ $C_2$ $C_3$ $C_4$ $C_5$ $C_6$ $C_7$</td>
</tr>
</tbody>
</table>

Note: $E_l$ is the layer modulus, $E_b$ is the base modulus and $E_s$ is the subgrade modulus. $H_1$ and $H_2$ for the surface and base layer thickness; $C_1$-$C_7$ for deflection values of different distance from the loading point.

Results and analysis

The back-calculation models were obtained by matlab programming according to the principle of the GEP algorithm. The GEP-based formulations of the pavement layer moduli were as given below. And the parameters of the models were listed in Table 3.

$$E_s = \left[ \frac{(C_6 - C_2) \times (H_1 + H_2)}{C_1 - C_4} \right]^2 - C_4 + 852.68877 + \frac{C_6 \times (H_2 - C_6)}{C_5 - C_1} + 67.54 - 7.4492515 \times C_6$$ (6)
\[ E_0 = (-0.0158339)\times C_4 \times C_2 - C_5 + \frac{(C_6 - C_4)\times C_4}{C_4 - C_3 + 0.640906} + 1353.6745 + C_7 - H_1 \times H_2 + \frac{(C_6 - C_4)\times C_1}{C_4 - C_3 + 9.9179136} \]  

\[ E_s = \frac{8.8652014\times C_1 - H_1}{C_7} + \frac{H_2}{C_4 - C_3 + H_1 - 7.9684747} - H_2 - \frac{64.80242\times C_1}{C_1 - 8.1003025} + \frac{0.1084811\times C_1 + 14441.541}{C_6} \]  

Table 3. Parameters of GEP approach model.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>GEP-1</th>
<th>GEP-2</th>
<th>GEP-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population size</td>
<td>50</td>
<td>65</td>
<td>55</td>
</tr>
<tr>
<td>Number of generation</td>
<td>60000</td>
<td>72000</td>
<td>80000</td>
</tr>
<tr>
<td>Genes</td>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Head size</td>
<td>8</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>Fitness</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Function set</td>
<td>+x^{\frac{2}{3}}</td>
<td>+x^{\frac{1}{a}}</td>
<td>+x^{3}</td>
</tr>
<tr>
<td>Mutation rate</td>
<td>0.044</td>
<td>0.044</td>
<td>0.044</td>
</tr>
<tr>
<td>Pouring string rate</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>IS rate</td>
<td>0.05</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>RIS rate</td>
<td>0.05</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Gene transposition rate</td>
<td>0.2</td>
<td>0.3</td>
<td>0.35</td>
</tr>
<tr>
<td>One-point recombination rate</td>
<td>0.3</td>
<td>0.2</td>
<td>0.125</td>
</tr>
<tr>
<td>Two-point recombination rate</td>
<td>0.3</td>
<td>0.2</td>
<td>0.125</td>
</tr>
<tr>
<td>Gene recombination rate</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Linking function</td>
<td>addition</td>
<td>addition</td>
<td>addition</td>
</tr>
</tbody>
</table>

Note: In the function symbol set, \(^2\) delegates squared, \(^3\) represents to take cube, \(1 / a\) represents to take reciprocal.

The ‘goodness-of-fit’ statistics for the GEP model predictions were performed using statistical parameters which include the correlation coefficient (R), relative mean absolute error (RMAE) and mean absolute percentage error (MAPE) [11]. There was a high linear correlation when the absolute value of R was greater than 0.9, but it was known that the R value alone was not a well indicator of prediction accuracy of a model. Definitions of these evaluation criteria are provided as follows, respectively:

\[ R = \frac{\sum_{i=1}^{n} (y_i^f - \bar{y}^f)(y_i^b - \bar{y}^b)}{\sqrt{\sum_{i=1}^{n} (y_i^f - \bar{y}^f)^2 \sum_{i=1}^{n} (y_i^b - \bar{y}^b)^2}} \]
\[ RMAE = \sqrt{\sum_{i=1}^{n} \left| y_i' - y_i^b \right|} / n \]  

(10)

\[ MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i' - y_i^b}{y_i'} \right| \times 100 \]  

(11)

Where \( y_i' \) is the theoretical value, \( y_i^b \) is the predicted value, \( \overline{y_i'} \) is average of the theoretical values, \( \overline{y_i^b} \) is average of the predicted values and \( n \) is total number of sample. The statistical parameters of the sets of the GEP models are listed in Table 4.

| Table 4. Correlation coefficient of GEP back-calculation model. |
|-------------------|-------------------|-------------------|-------------------|
|                   | Total set         | Training set       | Testing set       |
|                   | R       | RMAE   | MAPE   | R       | RMAE   | MAPE   | R       | RMAE   | MAPE   |
| GEP-1             | 0.997   | 7.112  | 3.549  | 0.991   | 7.417  | 4.207  | 0.997   | 5.754  | 0.949  |
| GEP-2             | 0.983   | 11.652 | 5.922  | 0.972   | 8.781  | 6.067  | 0.910   | 13.937 | 5.340  |
| GEP-3             | 0.999   | 1.133  | 1.499  | 0.997   | 0.843  | 1.207  | 0.999   | 1.891  | 2.667  |

Figs. 2-4 were the distribution diagrams for the back-calculation modulus and the numerical results of the pavement layer moduli, which defined as the theoretical modulus. As we could see that the values of R of these models for all the sets were higher than 0.9, and the best value of R was 0.999. It meant that the theoretical values and the predicted values showed a high linear correlation. As the statistical values of RMAE and MAPE from the training set in GEP-1 were found as 7.417 and 4.207%, respectively. The values in the testing set were found as 5.754 and 0.949%, respectively. The values of RMAE and MAPE were low, so the distinction between the theoretical modulus and the predictive modulus was small, it meant that the predict effect of the GEP-1 model was very well. Similarly, the values of RMAE and MAPE in GEP-2 and GEP-3 were not high. The best values of RMAE and MAPE were 0.843 and 1.207% in the training set of GEP-3 which was the back-calculation model of the subgrade modulus. All of the statistical values in Table 4 showed that the back-calculation models had high precision and were suitable and accurate. The GEP models were easy to get and calculated quickly, and it could effectively prevent and overcome the shortcomings of existing methods.

![Figure 2. The back-calculation result of GEP-1 model.](image)
CONCLUSIONS

This paper studies an efficient approach for back-calculating the pavement layer moduli. The proposed models are based on the tested results collected from the literatures. The results obtained from the model show a high linear correlation between the numerical results and the predicted modulus according to the statistical values of R, RMAE and MAPE. GEP is simple, reliable and can effectively avoid the impact of the initial value, and is a new modular back-calculation method should be widely applied. GEP is a good technique for used in the prediction of pavement modulus. The GEP-based models for asphalt pavement back-calculation can obtain the function expression of each pavement layer moduli quickly. It can met engineering requirements. And it may open a new area for simple and accurate method to many other civil engineering problems.

REFERENCES


