Prediction of city traffic accidents based on grey Markov chain model

Shengneng Hu

School of Civil Engineering & Communication, North China University of Water Resources and Electric Power, Zhengzhou 450011, China

Abstract

According to the characteristics of nonlinear and stochastic volatility in urban traffic accident prediction system, Combining Grey model with Markoff chain, established according to grey Markov prediction model for city traffic accident. It can not only give full play to the characteristics of gray system prediction, but also use Markoff chain to predict the volatility data. In the algorithm based on Grey Markov chain model, to correct the error of the prediction results, based on the historical data of traffic accidents in Chinese Xinyang city, the number of traffic accidents from 2016 year to 2018 year was predicted, the results show that when using Grey Markov chain model, the average relative error from the Grey model 5.14% down to 1.34%, the error reduction of 2.12%. The model not only has high prediction accuracy, but also can reflect the general trend of the development and change of the data sequence and the inherent law of each state of the system, it can be used to describe the stochastic volatility.

Keywords: markoff chain, prediction model, traffic accident, grey model.

1. INTRODUCTION

Road traffic system is a dynamic system based on human, vehicle and road, there are many factors that affect traffic safety, and the mechanism is complex, so the occurrence of road traffic accidents is very random and random. When the traditional regression analysis, time series, BP neural network and other methods for traffic accident prediction, because of the limitations of each model, the prediction results are often biased. Regression coefficient is difficult to determine, and the sample size is larger; The time series method cannot reflect the influence of the random external environment due to the few factors considered, Therefore the accuracy and practicability of the forecasting results are poor; BP neural network prediction rules need to input more comprehensive data, the method converges slowly, and the program is complicated.

When using the regression analysis method, need to use historical data to find the key factors that lead to traffic accidents, and quantify the impact of various key factors. In the prediction of the future situation, according to these rules have been found, in the case of known or expected factors, to determine the occurrence of traffic accidents. Sluipas introduces a variety of road factors to analyze the influence factors of highway traffic accidents, and analyzes the data of Lithuania high grade highway (SLIUPAS, 2015). Cheng Wei et al have expanded, an adaptive neural network fuzzy system is introduced for the prediction of traffic accident loss, and the number of traffic accidents, deaths and injuries are also introduced as input variables (Cheng et al., 2014). Liu Xiuqing et al. Introduced radial basis function neural network (ANN) into traffic accident prediction, which has a strong advantage over BP network in approximation ability, classification ability and learning speed, with the total population, the number of motor vehicles, the number of roads, GDP and other macroeconomic data as input variables, the number of deaths and economic losses as the output, the data fitting effect is better(Liu et al., 2009). Qin Liyan and others established a neural network model for road traffic accident prediction based on genetic algorithm and BP neural network prediction model, analysis of factor evaluation index China road safety status and the main influence to traffic accident deaths as the evaluation index (output variables), the amount of motor vehicle road mileage and the per capita GDP as input variables (factors), were used to test 2009 ~ 2014 the road traffic accident data (Qin et al., 2016).

The traffic accident is random, and the occurrence of the traffic accident is very random and fuzzy. The causes of traffic accidents including many uncertain factors, such as traffic flow, vehicle speed, personnel status, collection and identification of these factors have great difficulty, if using regression analysis techniques, often requires a lot of factors to construct and large sample data set, so the prediction of many cases analysis is not
feasible. Therefore, some scholars from the development of the accident itself to explore the law, the use of more mature time series analysis method to predict traffic accidents. Zhang Jie et al. use the ARIMA model to predict the number of deaths per year (Zhang et al., 2015). Wang Zhen and Zhang Xingqiang put together ARIMA and fuzzy neural network, the idea is simple, there are two methods to fit the data, and then assign the respective weights to determine the weights of the method is to minimize the sum of squared error (Wang et al., 2010).

The traditional time series model is relatively simple and cannot reflect the inherent complexity of road traffic accidents, some scholars use the theory of grey system. The grey system theory regards the stochastic process as a time dependent gray process, the prediction of road traffic accident can be regarded as a grey system, which can be predicted by the theory of grey system (Zhao et al., 2014; Zhang et al. 2015; Li, 2011). The grey model GM(1,1) takes the change of traffic accident index in some areas as the grey quantity changing with time, The exponential function is used to fit the traffic accident index value of each period, and the law of its development with time is found, and predict the future target value by this rule. This method is suitable for small samples and poor data. Because the road traffic system is a dynamic time-varying system, road traffic accident is a kind of grey system, which is a characteristic of the grey system, and it is a non-stationary random process (DANG, 2009; Zhao, 2013). So the grey Markov chain prediction model can be established by using the characteristics of grey prediction and Markov prediction, grey prediction is used to reveal the general trend of the road traffic accident time series change, Markov prediction is used to determine the transfer law of residual state. The grey Markov chain prediction model of road traffic accidents can effectively utilize the information given by the historical data of road traffic accidents, which can greatly improve the prediction accuracy of the large data column with random fluctuations (Stephen, 2014; Hu, 2013;).

2. CITY TRAFFIC ACCIDENT PREDICTION ALGORITHM BASED ON GREY MARKOV MODEL

2.1 Grey GM (1,1) model

The grey model is used to accumulate the original data, and the mean generation sequence and the matrix B and Y are established, then the mathematical model is established by the least square regression and differential method. Finally, the results obtained by the model are obtained by the reduction data (Shen, 2013).

According to the value of the model at each moment, the original data sequence is shown in the formula (1)

\[ x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)\} \tag{1} \]

Original data sequence is accumulated to get the formula (2)

\[ x^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(n)\} \tag{2} \]

Make a sequence of sequences, and generate a sequence

\[ z^{(1)}(k) = \frac{1}{2}\left[x^{(1)}(k) + x^{(1)}(k-1)\right] \tag{3} \]

Using type (1) and formula (3), establish matrix Y and B

\[ Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(k-1) \\ x^{(0)}(k) \end{bmatrix}, \quad B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(k-1) & 1 \\ -z^{(1)}(k) & 1 \end{bmatrix} \tag{4} \]
Parameters are estimated by least squares, and the value of a and B is obtained

\[
\hat{a} = (B^T B)^{-1} B^T Y = \begin{bmatrix} a \\ b \end{bmatrix}
\] (5)

Determine the form of the model, and restore the gray prediction value, as shown in formula (6), formula (7)

\[
\hat{x}^{(1)}(k) = x^{(0)}(1) - \frac{b}{a} e^{-a(k-1)} + \frac{b}{a}
\] (6)

\[
\hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1)
\] (7)

2.2 Markov chain model

The Markoff chain is based on the observation of the discrete state, an empirical dominated estimate of the transition probability of a parameter stochastic process. Firstly, the state of the original data is divided, then calculate the transition-probability matrix, and finally get the future predictive value (Pei, 2013).

2.2.1 State partition

According to the relative error between the predicted value and the actual value of the grey model, the relative error is divided into r state. Set residual sequence as \( e=(\varepsilon(1), \varepsilon(2), \ldots, \varepsilon(n)) \). In the type: \( \varepsilon(i)=x^{(0)}(i)-\hat{x}^{(0)}(i) \), \( i=(1, 2, ..., n) \). First, the maximum absolute value of residuals is calculated. by using the averaging method, let \( g=\max/3, -3g, -g, g \) and \( 3g \) are used to divide the residuals into 3 states. \( E_1:(-3g, g); E_2:(-g, g); E_3:(g, 3g) \). Then a sequence of states can be obtained from the range of each value of the residual sequence, \( \varphi=(\varphi(1), \varphi(2), ..., \varphi(n)).\) The algorithm flow chart is shown in Figure 1.

![Figure 1. Determining residual state algorithm flow](image)

2.2.2 State transition probability matrix

Assuming that \( P \) is the state I to j the M step transfer probability, \( M \) is the state I to the state of the j m transfer times, \( M \) is the number of I states, the state transition probability matrix, such as formula (8)

\[
R^{(m)} = \begin{bmatrix}
P_{11}^{(m)} & P_{12}^{(m)} & \cdots & P_{1r}^{(m)} \\
P_{21}^{(m)} & P_{22}^{(m)} & \cdots & P_{2r}^{(m)} \\
\vdots & \vdots & \ddots & \vdots \\
P_{r1}^{(m)} & P_{r2}^{(m)} & \cdots & P_{rr}^{(m)}
\end{bmatrix}
\] (8)
Formula: $p_{ij}^{(m)} = \frac{M_{ij}^{(m)}}{M_I}$.

2.2.3 Calculated predicted value

Assume that the time series is in state $j$ at time $k$, according to the state $j$ residual interval median $[w^j, w^{j*}]$, and grey prediction value $\hat{x}^{(0)}(k)$, the grey Markov model prediction value $\hat{y}(k+1)$ can be obtained, such as formula (9)

$$\hat{y}(k+1) = \left[1 + \frac{(\omega_{j-} + \omega_{j+})}{2}\right]x^{(0)}(k)$$

(9)

2.3 Test of model accuracy

After the establishment of the grey prediction model, the practicability of the model and the accuracy of the model are verified. After calculating the residuals, the average relative error, the mean square error ratio, the small error probability and other indicators, can find the scale of the gray prediction model accuracy test (see Table 1), the accuracy of the model GM(1,1) could be judged.

1) Calculating the residual error $\epsilon(k)$, relative error $\Delta(k)$ and average relative error $\overline{\Delta}$ of the original data sequence

$$\epsilon(k) = x^{(0)}(k) - \hat{x}^{(0)}(k)$$

(10)

$$\Delta(k) = \left|\frac{\epsilon(k)}{x^{(0)}(k)}\right|$$

(11)

$$\overline{\Delta} = \frac{1}{n} \sum_{k=1}^{n} \Delta(k)$$

(12)

2) The standard deviation $S_1$, $S_2$ of the residuals and original data are calculated respectively. According to $S_1$, $S_2$, the mean square error ratio $C$ and the small error probability $P$ are calculated respectively.

Table 1 Accuracy test of grey model

<table>
<thead>
<tr>
<th>Accuracy class</th>
<th>Relative error($\Delta$)</th>
<th>Mean variance ratio($C$)</th>
<th>Small error probability($P$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>$\Delta \leq 0.01$</td>
<td>$C \leq 0.35$</td>
<td>$P \geq 0.95$</td>
</tr>
<tr>
<td>Medium</td>
<td>$0.01 &lt; \Delta \leq 0.05$</td>
<td>$0.35 &lt; C \leq 0.50$</td>
<td>$0.80 &lt; P \leq 0.95$</td>
</tr>
<tr>
<td>Qualified</td>
<td>$0.05 &lt; \Delta \leq 0.1$</td>
<td>$0.5 &lt; C \leq 0.65$</td>
<td>$0.7 &lt; P \leq 0.8$</td>
</tr>
<tr>
<td>Poor</td>
<td>$0.1 &lt; \Delta \leq 0.2$</td>
<td>$C &gt; 0.65$</td>
<td>$P &lt; 0.7$</td>
</tr>
</tbody>
</table>

$$S_1 = \sqrt{\frac{1}{n} \sum_{k=1}^{n} \left[x^{(0)}(k) - \bar{x}\right]^2}$$

(13)

$$S_2 = \sqrt{\frac{1}{n} \sum_{k=1}^{n} \left[\epsilon(k) - \bar{\epsilon}\right]^2}$$

(14)

147
3. EXAMPLE APPLICATION

Markoff chain is used to correct the prediction error of grey model GM(1,1), based on the number of traffic accidents in Xinyang city, China from 2009 year to 2015 year, the number of traffic accidents in Xinyang city from 2016 to 2018 was predicted.

3.1 Establishment of GM (1, 1) model

The grey GM (1, 1) model is established as follows:

(1) The original data sequence is $x^0 = \{1047, 1068, 872, 902, 876, 846, 895\}$;

(2) The data were accumulated sum: $x^{(0)} = \{1047, 2115, 2987, 3889, 4765, 5611, 6506\}$;

(3) Establish mean and generate sequence $z^{(1)}(k)$, $z^{(1)}(k) = \{1581, 2551, 3438, 4327, 5188, 6058.5\}$, matrix $Y$ and $B$ are respectively:

$$
Y = \begin{bmatrix}
1068 \\
872 \\
\vdots \\
895
\end{bmatrix},
B = \begin{bmatrix}
-1581 & 1 \\
-2551 & 1 \\
\vdots & 1 \\
-6058 & 1
\end{bmatrix}
$$

(4) The parameters are estimated by least squares, and the value of $a$ and $b$ is obtained

$$
\hat{a} = (B^T B)^{-1} B^T Y = \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0.031 \\ 1032.155 \\ 452 \end{bmatrix}
$$

(5) The values of the parameters $a$ and $b$ are introduced into the formula (6), and the model is shown as the formula (17).

$$
\hat{x}^{(1)}(k) = 32547.78797 - 31500.78797 e^{-0.0317124 - 1}
$$

According to formula (17), and formula (6) calculated data, the grey prediction value of the number of traffic accidents in Xinyang city from 2009 year to 2015 year is obtained, the results are shown in table 2. Prediction results show, the predicted value of the model is monotonically decreasing, and the relative error of the model is larger in 2010 year and in the 2011 year, respectively 7.96% and 9.29%. Finally, the grey prediction value of the number of casualties from 2016 to 2017 year was found to be 813, 788, and 763 respectively.

(6) Test the accuracy of the model GM (1, 1). Using the formula (10) ~ (16) can be calculated, the average relative error is 4.32%, the posterior error ratio is about 61.34%, the small error probability is about 0.7143. According to the standard of table 1, it can be seen that the accuracy of the model is 3, which can be used for traffic accident prediction, but the accuracy is low, further optimization is needed to improve the accuracy of the model.
Table 2 Actual traffic accident casualties and grey model prediction value

<table>
<thead>
<tr>
<th>Year</th>
<th>Actual casualty/people</th>
<th>Grey model predictive value</th>
<th>Residual</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>1047</td>
<td>1047</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2010</td>
<td>1068</td>
<td>983</td>
<td>85</td>
<td>7.96</td>
</tr>
<tr>
<td>2011</td>
<td>872</td>
<td>953</td>
<td>-81</td>
<td>-9.29</td>
</tr>
<tr>
<td>2012</td>
<td>902</td>
<td>923</td>
<td>-21</td>
<td>-2.33</td>
</tr>
<tr>
<td>2013</td>
<td>876</td>
<td>894</td>
<td>-18</td>
<td>-2.05</td>
</tr>
<tr>
<td>2014</td>
<td>846</td>
<td>866</td>
<td>-20</td>
<td>-2.36</td>
</tr>
<tr>
<td>2015</td>
<td>895</td>
<td>839</td>
<td>56</td>
<td>6.26</td>
</tr>
</tbody>
</table>

3.2 Establishment of Markov chain model

3.2.1 State partition

Because of the small number of samples in this study, according to the mean value, the error can be divided into three states, respectively, using $E_1$, $E_2$, $E_3$, as shown in table 3.

According to the state division in Table 3, the number of casualties in the traffic accident was divided from 2009 to 2015 year. The results are shown in table 4.

Table 3 State division of traffic accident casualties

<table>
<thead>
<tr>
<th>State</th>
<th>$E_1$ (%)</th>
<th>$E_2$ (%)</th>
<th>$E_3$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error range</td>
<td>-9.29~3.54</td>
<td>-3.54~2.21</td>
<td>2.21~7.96</td>
</tr>
</tbody>
</table>

Table 4 State division of actual casualties in traffic accidents

<table>
<thead>
<tr>
<th>Year</th>
<th>Actual casualty/people</th>
<th>Grey model predictive value</th>
<th>Error</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>1047</td>
<td>1047</td>
<td>0</td>
<td>$E_2$</td>
</tr>
<tr>
<td>2010</td>
<td>1068</td>
<td>983</td>
<td>7.96</td>
<td>$E_3$</td>
</tr>
<tr>
<td>2011</td>
<td>872</td>
<td>953</td>
<td>-9.29</td>
<td>$E_1$</td>
</tr>
<tr>
<td>2012</td>
<td>902</td>
<td>923</td>
<td>-2.33</td>
<td>$E_2$</td>
</tr>
<tr>
<td>2013</td>
<td>876</td>
<td>894</td>
<td>-2.05</td>
<td>$E_2$</td>
</tr>
<tr>
<td>2014</td>
<td>846</td>
<td>866</td>
<td>-2.36</td>
<td>$E_2$</td>
</tr>
<tr>
<td>2015</td>
<td>895</td>
<td>839</td>
<td>6.26</td>
<td>$E_3$</td>
</tr>
</tbody>
</table>

3.2.2 Constructing transition probability matrix

By the formula (18) to calculate one step, two step, three step transition probability matrix:

\[
R^{(1)} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1/2 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} 
R^{(2)} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1/2 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} 
R^{(3)} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1 & 0 \end{bmatrix} \]  

(18)

3.2.3 Predictive value calculation

Using the formula (9) to fit the number of traffic accidents in Xinyang city from 2007 to 2013 year. For example, in 2010 year, the gray prediction value was 983, in the state of $E_3$, by calculating formula $\hat{y}=983*[1+0.5*(2.21%+7.96%)])$, grey Markov chain forecast in 2010-year value of 1033 people were obtained. Similarly, the predicted values of the remaining years are obtained, and the residuals and errors of the two models are shown in table 5. From table 5, the relative error of the predictive value of the grey GM (1, 1) in 2010 year and 2011year were 7.96% and 9.29% respectively, and the grey Markov chain GM (1, 1) predicted values
and the relative error were reduced to 3.28% and 2.29%. From table 2 and 4 it can be seen that the grey GM (1, 1) model predictive value showed a smooth decline trend, but the grey Markov chain GM (1, 1) model value fluctuated, the number of casualties is more close to the actual value, the prediction results are more reliable.

Table 5 Prediction results of Grey Markov chain model

<table>
<thead>
<tr>
<th>Year</th>
<th>Actual casualty /people</th>
<th>Grey Markov model prediction value</th>
<th>Residual</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>1047</td>
<td>1047</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2010</td>
<td>1068</td>
<td>1033</td>
<td>35</td>
<td>3.28</td>
</tr>
<tr>
<td>2011</td>
<td>872</td>
<td>892</td>
<td>-20</td>
<td>-2.29</td>
</tr>
<tr>
<td>2012</td>
<td>902</td>
<td>917</td>
<td>-15</td>
<td>-1.66</td>
</tr>
<tr>
<td>2013</td>
<td>876</td>
<td>888</td>
<td>-12</td>
<td>-1.37</td>
</tr>
<tr>
<td>2014</td>
<td>846</td>
<td>860</td>
<td>-14</td>
<td>-1.65</td>
</tr>
<tr>
<td>2015</td>
<td>895</td>
<td>882</td>
<td>13</td>
<td>1.45</td>
</tr>
</tbody>
</table>

The difference between the two models is tested and compared with the posterior difference, \( C(GM(1,1) \text{model}) = \frac{S_1}{S_0} = 0.227 \), \( C(\text{Grey Markov model}) = \frac{S_1}{S_0} = 0.094 \). Scientific sense, from the data fitting precision: the fitting precision of 2 prediction models are more suitable, but in comparison, the grey Markov model is the highest precision. The simulated data and actual data line chart in Figure 2, which can be seen from the line graph and the actual value is most consistent with the modified Markov chain.

Figure 2. Actual value and broken-line chart of three kinds of mode

As can be seen from table 4, the predicted value of casualties in 2015 year is in the state of \( E_3 \), and the initial row vector is \( V_0 = (0, 0, 1) \). Therefore, \( R(1)V_0 = (1, 0, 0) \). It shows that the number of casualties caused by traffic accidents in 2016 year is \( E_1 \), and the number of casualties is 761 by the formula (9), which is predicted to be about 2015 year. Similarly, it can be obtained the state and the predicted value of the number of traffic accident casualties in 2017, 2018 year, as shown in table 6.

Table 6 Prediction value of traffic accident casualties from 2016 to 2018 year

<table>
<thead>
<tr>
<th>Year</th>
<th>Grey model predictive value</th>
<th>State</th>
<th>State interval median</th>
<th>Grey Markov model prediction value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>813</td>
<td>( E_1 )</td>
<td>-6.37%</td>
<td>761</td>
</tr>
<tr>
<td>2017</td>
<td>788</td>
<td>( E_2 )</td>
<td>-0.665%</td>
<td>783</td>
</tr>
<tr>
<td>2018</td>
<td>763</td>
<td>( E_2 )</td>
<td>-0.665%</td>
<td>758</td>
</tr>
</tbody>
</table>

3. CONCLUSION

(1) Firstly, the grey GM (1, 1) model is established, on the basis of the number of traffic accidents from 2009-2015 year in Chinese Xinyang city, the number of traffic accidents in the city from 2016 to 2018 year was predicted. But the prediction accuracy of the model is low, and the average relative error is higher, the accuracy of the model is only three level, and the accuracy of the model can be improved.

(2) The Markoff chain model is established by using the Markoff chain, though mean state division, establishing the state transition-probability matrix, the grey GM(1, 1) model prediction error was corrected, the prediction accuracy of the model has been improved obviously, the average relative error is significantly reduced, and the
accuracy of the model is improved to one level. It is found that the grey Markov chain model is more reliable than single gray model, the prediction results are more close to the actual.

(3) When the original data fluctuation is large, the simple use of the grey GM(1, 1) model may be random will erase the original data, the grey Markoff model gives full play to the advantages of the grey prediction model and the prediction model of the Markoff, and takes into account the influence of the two factors of the change trend and the relative fluctuation on the prediction results, It is a more accurate and practical prediction method for the original series with strong fluctuation and strong randomness. At the same time, the grey Markov model also considers all kinds of random factors on the system state transition rule, the information provided by historical data is fully exploited, the state division and the transfer probability matrix are determined in the prediction, which improves the accuracy and reliability of the prediction results.

(4) At the same time, there are still some imperfections in the model. The results of the model prediction will be affected by the division of the state interval and the number of prediction steps, and sometimes the results will be different because of the different state. In theory, the state division is fine, the prediction accuracy is high, but it will reduce the number of samples in each state. Therefore, the prediction results of the grey Markov model accuracy is affected by the state division to a great extent. Therefore, it is necessary to determine the relevant parameters for specific problems, so that the model can achieve the best results.

ACKNOWLEDGMENTS

Thanks to Xinyang city road management department to provide the relevant data, thanks to the colleagues in the project team to deal with a large number of data.

REFERENCES