Prediction of Road Traffic Accidents based on Rolling-optimized Grey Markov Mode

Chengzhen Xu¹, Heng Zhang¹, Yanan Gong¹, Huiyun Sang¹*, Guanglin Sun² and Jing Chen³

¹ Department of Logistics Engineering, Shandong Jiaotong University, Jinan, Shandong, 250300, China
² Road Traffic Safety Research Center of the Ministry of Public Security, Beijing, 100000, China
³ School of Architecture and Rail Transit, Xi’an Vocational and Technical College, Xi’an, 710000, China

*Corresponding author’s e-mail: hysang@139.com

Abstract. In order to solve the problem that the accuracy of the accident prediction is affected by the time-effectiveness of the forecast data, the paper introduces the rolling optimization strategy based on the original data of the road traffic accident in the road section, and further establishes the rolling optimization-grey Markov dynamic prediction model. Using the Markov chain theory to explore the transition law between different states, the development trend of road traffic accidents volume is predicted, and the prediction accuracy of the random time series is further improved. Case analysis proves that the method has better prediction accuracy and practicability over a certain period of time, which can provide reference for road traffic accident prediction analysis and traffic safety early warning.

1. Introduction

At present, there are two limitations in road traffic accident prediction methods. The first is that the requirements for prediction data are high, the second is that the accuracy of mathematical models is strict. Due to the randomness, non-linearity, volatility and time-varying of initial data, road traffic accident prediction is still difficult. Currently, traffic accident prediction are mainly based on data-driven and statistical analysis, such as time series method[1], support vector machine method (SVM) [2], Markov method[3], etc. Time series method can describe the periodic law of data, but it is unrealistic to record and predict all the data, so it is difficult to establish training and learning models. SVM has high prediction accuracy, but it needs a lot of training data, which consumes lots of energy and is difficult to be applied in practice. Markov can reflect the periodic variation characteristics of the data with strong randomness, but its determination of the state set is difficult. Grey prediction GM (1, 1) only needs a small sample to establish the model[4], but GM (1, 1) has a large long-term prediction residual, which cannot reveal the regularity of periodic changes of data.

Based on above analysis and relevant research, this paper proposes a method based on Rolling Optimization - Grey Markov which combines Markov method with GM (1, 1) prediction model. Based on rolling data prediction theory[5], the model continuously adds new prediction data and eliminates old data, finally obtains prediction data during the study period.
2. The construction of the accident prediction model

2.1. Establish GM (1, 1) rolling optimization model

Let \( x^{(0)} = [x^{(0)}(t_1), x^{(0)}(t_2), \ldots, x^{(0)}(t_n)] \) be the original time series, where \( t_i \) is the time factor. According to Sequence \( \{x^{(0)}\} \), predicts the traffic accident data of period \( m \) after period \( t_n \). Set the predicted time series to \( \hat{x}^{(0)} = [\hat{x}^{(0)}(t_{n+1}), \ldots, \hat{x}^{(0)}(t_{n+m})] \). The modeling mechanism is as follows:

Model usually sets the sequence \( \{x^{(0)}\} \) as a discrete time series, thus introducing a first order cumulative sequence \( \{x^{(1)}\} \). The randomness of the original sequence is weakened to enhance the regularity of the time series. Let the first order cumulative sequence \( x^{(1)} \) be:

\[
\begin{align*}
x^{(1)} &= [x^{(1)}(t_1), x^{(1)}(t_2), \ldots, x^{(1)}(t_n)] \\
\text{where } x^{(1)} &= \sum_{k=1}^{n} x^{(0)}(t_k), i = 1, 2, 3, \ldots, n
\end{align*}
\]

Time series \( x^{(1)} \) is the monotonically increasing function of \( t \). Then a single-sequence first-orderlinear dynamic GM(1,1) model is established, and the corresponding differential equation is:

\[
\begin{align*}
\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) &= u^{(1)} \\
\end{align*}
\]  

In the GM model, the initial value of the boundary is satisfied with \( x^{(0)}(t_1) = x^{(1)}(t_1) \), from which the approximate function can be obtained:

\[
\hat{x}^{(1)}(t_i) = \left[x^{(0)}(t_i) - \frac{u}{a}\right] \cdot e^{-a(t_i-1)} + \frac{u}{a}, i = 1, 2, \ldots, n
\]

Among them, \( a \) and \( u \) are the undetermined coefficients of the initial statistical data of the accident, which represent the development coefficient and the grey control quantity in the modeling process. They can be calculated by the least square method:

\[
\hat{a} = [x(A)^T x(A)]^{-1} \cdot x(A)^T \cdot Y_N
\]

where \( \hat{a} = [a, u]^T \)

\[
x(A) = \begin{bmatrix}
-\frac{1}{2} [x^{(1)}(2) + x^{(1)}(1)] & 1 \\
-\frac{1}{2} [x^{(1)}(3) + x^{(1)}(2)] & 1 \\
\vdots & \vdots \\
-\frac{1}{2} [x^{(1)}(N) + x^{(1)}(N-1)] & 1
\end{bmatrix}
\]

\[
Y_N = [x^{(0)}(2), x^{(0)}(3), \ldots, x^{(0)}(N)]^T
\]

Therefore, the formula (7) is obtained as:

\[
\hat{x}^{(1)}(t_i) = \left(x^{(0)}(1) - \frac{u}{a}\right) e^{-a(t_i-1)} + \frac{u}{a}
\]

Based on the above steps, the time series function of predicted value of the traffic accident amount can be obtained by the regressive inverse operation:

\[
\begin{align*}
\hat{x}^{(0)}(t_i) &= \hat{x}^{(1)}(t_i) - \tilde{x}^{(1)}(t_{i-1}) \\
\tilde{x}^{(0)}(t_i) &= (1 - e^a) \cdot \left(x^{(0)}(1) - \frac{u}{a}\right) e^{-a(t_i-1)}
\end{align*}
\]

(3) Residual prediction GM(1,1) model

\( \hat{x}^{(0)}(t_i) \) is the factor prediction value of \( t_i \), and then the residual sequence is constructed as:

\[
e^{(0)}(t_i) = \{x^{(0)}(t_i) - \hat{x}^{(0)}(t_i)\}
\]
The sequence \( \{ \hat{\epsilon}^{(0)}(t_i) \} \) can be found by the GM(1,1) model, which is similar to the formulas (2)-(10). Add \( \hat{\epsilon}^{(0)}(t_i) \) of relevant sample to the original prediction sample \( \hat{x}^{(0)}(t_i) \) and achieve the revised prediction data \( \hat{x}^{(0)}(t_i) \):

\[
\hat{x}^{(0)}(t_i) = \hat{x}^{(0)}(t_i) + \hat{\epsilon}^{(0)}(t_i)
\]  

(11)

2.2. Markov Residual Modified GM (1,1) Model

Markov residual correction

The Markov method is based on state transition prediction, and the residual sequence is divided into \( h \) states. The residual value can only be expressed as positive and negative: \( E1 \) and \( E2 \). Set the state transition probability of state \( E_1 \) at time \( t_i \) to state \( E_1 \) at time \( t_j \) after \( k \) steps of transition is \( p_{ij}(k) \). Markov’s state transition matrix \( p_{ij}(k) \) can be calculated and obtained as:

\[
P(k) = [p_{ij}(k)]_{2\times 2} \quad i, j = 1,2
\]  

(12)

Defines the first step state transition probability vector \( p(1) = [p_1 \quad p_2] \)  

(13)

where, \( p_1 \) is the probability of state \( E1 \), and \( p_2 \) is the probability of state \( E2 \). After the state is transferred by \( k \) steps, we can get:

\[
p^{(k)}(1) = [p_1^{(k)} \quad p_2^{(k)}] = p(1) \cdot P(k) \]  

(14)

The Markov residual correction result is shown in equation (15):

\[
\hat{x}_1^{(0)}(t_i) = \hat{x}_1^{(0)}(t_i) + s(t_i) \cdot \hat{\epsilon}^{(0)}(t_i)
\]  

(15)

In the formula, \( \hat{x}_1^{(0)}(1) = \hat{x}_1^{(0)}(1) \), and \( s(t_i) \) is the residual symbol.

In step \( k \), when \( p_1 > p_2 \), let the residual symbol \( s(t_i) = 1 \); and when \( p_1 < p_2 \), let the residual symbol \( s(t_i) = -1 \).

2.3. Rolling Optimization Strategy

Considering that the prediction accuracy is greatly affected by the timeliness of the data, this paper uses "rolling data prediction principle" to remove the first period of data \( \hat{x}^{(0)}(t_{n+1}) \) of the original sequence and add the calculated new predicted value \( \hat{x}^{(0)}(t_{n+1}) \), re-construct the time series in the subsequent prediction process and then to predict \( \hat{x}^{(0)}(t_{n+m+1}) \) of the next period. The process circulates and can calculate a series of new forecast data, which is drawn as Fig.1.

Figure1. The layout of rolling optimization strategy

3. Accuracy test of combined forecasting model

Let the residuals of the original time series and the predicted series be \( \delta(t_i) \), then the mean value \( \overline{\delta}(t_i) \) and variance \( S_\delta \) are as follows.

\[
\overline{\delta}(t_i) = \frac{1}{n} \sum_{i=1}^{n} \delta(t_i), \quad S_\delta^2 = \frac{1}{n} \sum_{i=1}^{n} [\delta(t_i) - \overline{\delta}(t_i)]^2
\]  

(16)
Mean value $\overline{x}(t_i)$ and variance $S_o$ of the original time series are:

$$\overline{x}(t_i) = \frac{1}{n} \sum_{t=1}^{n} x^{(0)}(t_i), \quad S_o^2 = \frac{1}{n} \sum_{t=1}^{n} [x^{(0)}(t_i) - \overline{x}(t_i)]^2$$  \quad (17)

Let the mean square error ratio $C$ be:

$$C = \frac{S_\delta}{S_o}$$  \quad (18)

| Accuracy grade | Grade 1 (excellent) | Grade 2 (good) | Grade 3 (qualified) | Grade 4 (unqualified) |
|----------------|---------------------|----------------|---------------------|-----------------------|
| Mean square error ratio $C$ | $\leq 0.35$ | >0.35−0.50 | >0.50−0.65 | >0.65−0.80 |

### 4. Case study

#### (1) Predictive value calculation

Based on the statistical data of traffic accidents in 9 statistical periods of a subordinate administrative district of Jinan, the rolling optimization-greys Markov method is used to predict the traffic accidents in this district in the next few years. primary number of traffic accidents is shown in Tab 2.

| Statistical period | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|--------------------|---|---|---|---|---|---|---|---|---|
| Accidents No.(hundred) | 0.62 | 0.87 | 1.02 | 0.8 | 1.8 | 1.6 | 1.66 | 1.35 | 1.76 |

Based on the rolling optimization strategy, the first original data is sequentially removed, and the new calculated prediction value is added. According to Grey Markov theory, using Matlab to realize the algorithm. The results are achieved as in Tab. 3. And the value of $C$ is 0.1675, i.e. $C<0.35$ showing the excellent prediction accuracy. It is proved that road traffic accident prediction model is effective and can be used for subsequent prediction to reflect the development trend of road traffic accidents.

| Statistical period | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|--------------------|---|---|---|---|---|---|---|---|---|
| Results(hundreds) | 0.62 | 0.79 | 0.99 | 0.77 | 1.79 | 1.49 | 1.67 | 1.49 | 1.84 |

### 5. Conclusion

According to the characteristics of road traffic accidents and time series data, it is guided by the theory of gray system and Markov chain theory, a rolling optimization-grey Markov dynamic prediction model is established for road traffic accident volume prediction. The main contributions of this paper are as follows:

1. The rolling optimization strategy has better improved the problem that the prediction accuracy is greatly affected by the timeliness of the data. Each prediction replaces the first-period data in the time series. Moreover, the method is in line with the actual situation, it is relatively simple and easy to cooperate.

2. The method provided in this paper is to predict the number of traffic accidents. The accuracy of the prediction results depends to a large extent on the quality and quantity of the observed samples. However, for the case of limited data, based on the rolling optimization strategy, a small amount of data can be used to complete the forecast, and case studies show that higher prediction accuracy can be achieved.
Acknowledgments
Authors wishing to acknowledge the anonymous reviewers for their careful, insightful, and constructive comments that led to an improved version of this paper. This work is partially supported by Technical Research Plan of the Ministry of Public Security (2018JSYJB05).

References
[1] Li S, Cui Z, Zhao J. Research and Application of Dynamic Triple Exponential Smoothing Method Based on Matlab Adaptation—Taking the Prediction of National Road Traffic Accidents as an Example [J]. Mathematics in Practice and Theory, 2018, 48(12): 169-177.
[2] Tang Z, Dong W, Wang Z. Research on SVR-based urban road scraping accident assessment model [J]. Journal of Transportation Engineering and Information, 2019, 17(01): 1-4.
[3] Cao J, Shi S, Cao H, et al. Study on the road accident and traffic accident prediction and relationship of hazardous chemicals based on GM(1,1)-Markov [J]. China Safety Production Science and Technology, 2019, 15(01): 26-31.
[4] Jiang S, Song K, Xie W, et al. Data Analysis of Road Traffic Accidents Based on Grey Correlation and Apriori Algorithm [J]. Highway Engineering, 2019, 44(04): 67-73.
[5] Xu G. Statistical Forecasting and Decision Making (Fifth Edition) [m]. Shanghai University of Finance and Economics Press. 2016.