 https://doi.org/10.1051/e3sconf/202021303021

Forecast of Road Traffic Accidents grounded on Rolling optimization Grey Markov Model

ChengzhenXu[1], HengZhang[1], YananGong[1], HuiyunSang[1*], GuanglinSun[2] and JingChen[3]

[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

Abstract. This paper recommends the rolling optimization strategy based on the initial data of road traffic accidents, and builds the rolling optimization-grey Markov dynamic prediction model, which can effectively resolve the matter that the precision of accident forecast is influenced by the time benefit of the predicted data. In order to predict the development tendency of road traffic accidents and further improve the prediction precision of random time series, this paper uses Markov chain theory to probe into the transition law between different states. The case study shows that this measure has good forecast precision and practicability in a certain period of time, and can offer reference for road traffic accident forecast and traffic safety warning.

1 Introduction

Currently, the two boundedness of road traffic accident forecast ways are the high requirement of prediction data and the strict requirement of mathematical model precision. It is still difficult to predict road traffic accidents mainly because of the randomness, nonlinearity, volatility and time variation of the initial data. At present, such ways as time series[1], support vector machine (SVM)[2] and Markov method[3] are mostly traffic accident forecast on account of data-driven and statistical analysis. Time series method is hard to establish training and learning model mainly because although it can describe the cyclical rules of data, it is impractical to take notes and predict all data. SVM are hard to be used in practice, mainly because although SVM has high forecast precision, it needs a large of practicing data and uses up a large of energy. Although Markov can reflect the trait of cyclicity of data with strong randomicity, it is hard to identify its state set. Grey GM (1,1) unable reveal the regularity of cyclicity variations of data, mainly because although it only requires a small sample to build the matrix [4], GM (1,1) has a big long-term forecasting residual. On the basis of the above analysis and related research, a grey Markov method based on rolling optimization is put, which unites Markov method with GM (1,1) prediction model. Grounded on rolling data prediction theory[5], this model constantly increases new forecast data and removes old data, and ultimately gains the forecast data in the period of study.

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

2 The construction of the accident prediction model

2.1 Establish GM (1,1) rolling optimization model

t_i is set as a time factor, $x^{(0)} = \left[x^{(0)}(t_1), x^{(0)}(t_2), \cdots, x^{(0)}(t_n)\right]$ as the original time series. Based on the statistics of sequence $\left\{x^{(0)}\right\}$, the traffic accident data of phase m after phase t_n are predicted. Set the predicted time series to $\hat{x}^{(0)} = \left[\hat{x}^{(0)}(t_{n+1}), \hat{x}^{(0)}(t_{n+2}), \cdots, \hat{x}^{(0)}(t_{n+m})\right]$. How to construct GM(1,1) model is that, according to the grey module theory, differential fitting and single-ended function residual identification are carried out. The above model is constructed as follows:
In this model, based on the sequence $\left\{x^{(0)}\right\}$ has been set as a random discrete time series, in order to enhance the regularity of the time series, the first order cumulative sequence $\left\{x^{(1)}\right\}$ is introduced to weaken the randomness of the original sequence. Let the first order cumulative sequence $x^{(1)}$ be:

$$x^{(1)} = \left[x^{(1)}(t_1), x^{(1)}(t_2), \cdots, x^{(1)}(t_n)\right] \quad (1)$$

where $x^{(1)} = \sum_{k=1}^{i} x^{(0)}(t_k), i = 1,2,3, \cdots, n$

Time series $x^{(1)}$ is a monotone increasing function about t. A single-sequence first-order linear dynamic GM(1,1) model is established, and its differential equations are as follows:

$\frac{dx^{(1)}(t)}{dt} + ax^{(1)} = u^{(1)}$ (2)

The approximate function can be obtained by satisfying the initial boundary value with $x^{(0)}(t_1) = x^{(1)}(t_1)$.

$\tilde{x}^{(1)}(t_i) = \left[x^{(0)}(t_i) - \frac{u}{a}\right] \cdot e^{-a(t-1)} + \frac{u}{a}, i = 1,2,\cdots,n$ (3)

Among them, a and u are included in the undetermined coefficients of the initial accident statistics, which represent the development coefficient and the grey control quantity in the modeling process. They can be figured out by the least square way:

$\hat{a} = [x(A)^T x(A)]^{-1} \cdot x(A)^T \cdot Y_N$ (4)

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

$$x(A) = \begin{bmatrix} -\frac{1}{2}[x^{(1)}(2) + x^{(1)}(1)] & 1 \\ -\frac{1}{2}[x^{(1)}(3) + x^{(1)}(2)] & 1 \\ \vdots & \vdots \\ -\frac{1}{2}[x^{(1)}(N) + x^{(1)}(N-1)] & 1 \end{bmatrix} \quad (5)$$

$Y_N = [x^{(0)}(2), x^{(0)}(3), \cdots, x^{(0)}(N)]^T$ (6)

Therefore, the formula (7) is achieved as:

$\tilde{x}^{(1)}(t_i) = \left(x^{(0)}(1) - \frac{u}{a}\right) e^{-a(t_i-1)} + \frac{u}{a}$ (7)

According to the above, the time series function of predicted value of the traffic accident amount can be achieved by the regressive inverse operation:

$\tilde{x}^{(0)}(t_i) = \tilde{x}^{(1)}(t_i) - \tilde{x}^{(1)}(t_{i-1})$ (8)

$\tilde{x}^{(0)}(t_i) = (1 - e^a) \cdot \left(x^{(0)}(1) - \frac{u}{a}\right) e^{-a(t_i-1)}$ (9)

Based on the basis that $\tilde{x}^{(0)}(t_i)$ is the factor prediction value of t_i, the residual sequence is constructed as:

$\varepsilon^{(0)}(t_i) = \{x^{(0)}(t_i) - \tilde{x}^{(0)}(t_i)\}$ (10)

The sequence $\{\hat{\varepsilon}^{(0)}(t_i)\}$can be found by the GM(1,1) model, which is similar to the formulas (2)-(10). Adds $\hat{\varepsilon}^{(0)}(t_i)$ of relevant sample to the original prediction sample $\tilde{x}^{(0)}(t_i)$ to get the revised prediction data $\hat{x}^{(0)}(t_i)$:

$\hat{x}^{(0)}(t_i) = \tilde{x}^{(0)}(t_i) + \hat{\varepsilon}^{(0)}(t_i)$ (11)

2.2 Markov Residual Modified GM (1,1) Model

(1) Markov residual correction

Markov prediction is a state transition prediction method. So, it is presumed that the residual sequence is divided into h states. Since the residual value can only be demonstrated as affirmative and depressed states: *E1* and *E2*, thereby *h=2* can be confirmed. Set the state transition probability of state E_i at time t_i to state E_j at time t_j after k steps of transition is $p_{ij}(k)$, Markov's state transition matrix $p_{ij}(k)$ can be figured out and achieved 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)

In which, 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 displayed in equation (15):

$\hat{x}_1^{(0)}(t_i) = \tilde{x}_1^{(0)}(t_i) + s(t_i) \cdot \hat{\varepsilon}^{(0)}(t_i)$ (15)

In the formula, $\hat{x}_1^{(0)}(1) = \tilde{x}_1^{(0)}(1)$, 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

Given time sequence $\{\hat{x}^{(0)}(t_i)\}$, predicts the quantity of road traffic accidents in period m after period t_n. The data results will be distorted due to the timeliness of the original data. Therefore, in the next data processing, through the principle of mobile data prediction, the original first-period sequence value $\tilde{x}^{(0)}(t_{n+1})$ will be replaced by the newly calculated predicted value $\hat{x}^{(0)}(t_{n+1})$ by which the next data $\hat{x}^{(0)}(t_{n+m+1})$ [11,12] for the first period will be predicted. Thereby, according to the application of mobile optimization strategies, the model can be established from the latest predicted values, while keeping the amount of data obtained from the sequence constant, which is drawn as Fig.1.

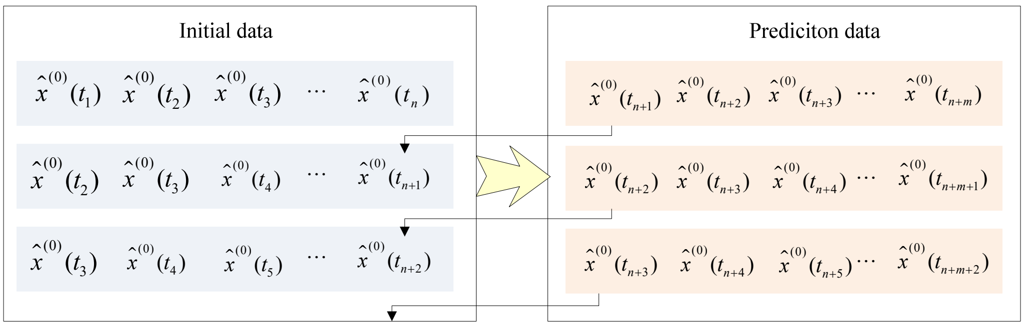

Figure1. The layout of rolling optimization strategy

3 Accuracy test of combined forecasting model

Assuming that the residual of the original time series and the predicted series is $\delta(t_i)$, then the average value is $\overline{\delta}(t_i)$and variance $S_\delta$are as follows.

$$\overline{\delta}(t_i) = \frac{1}{n}\sum_{i=1}^{n}\delta(t_i),\ S_\delta^2 = \frac{1}{n}\sum_{i=1}^{n}\left[\delta(t_i) - \overline{\delta}(t_i)\right]^2 \quad (16)$$

The mean value $\overline{x}(t_i)$and variance S_o of the temporal series temporal series are:

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

Let the error of mean square ratio *C* be:

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

Table 1. grade of accuracy inspection

Accuracy grade	Grade 1 (excellent)	Grade 2 (good)	Grade 3(qualified)	Grade 4(unqualified)
Mean square error ratio *C*	≤0.35	>0.35~0.50	>0.50~0.65	>0.65~0.80

4 Case study

(1) Predictive value calculation

According to the statistical data of traffic accidents in 9 statistical periods of subordinate administrative organs in Jinan area, the rolling optimization-grey Markov method is used to forecast the traffic accidents in the next few years. Table 2 shows the main number of traffic accidents.

Table 2. The primary number of traffic accidents

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 trundling optimization strategy, the first original data is deleted in sequence, and new calculated predicted values are added. According to grey Markov theory, the algorithm is realized by MATLAB. The consequence is achieved as in Tab. 3. And the value of *C* is 0.1675, i.e. C<0.35 showing the excellent precision of prediction.

Table3. The predicted value based on rolling optimization - Grey Markov

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

In the light of traits of road traffic accidents and time series data, under the guidance of gray system theory and Markov chain theory, a rolling optimization-gray Markov dynamic prediction model of road traffic acci dent volume prediction is built. The main contribution s of this paper are as follows:

(1) The rolling optimization strategy resolves the matter that the forecast precision is quite influenced by the softening data of time line. Each forecast displaces the data for the first period in the time series. furthermore, the method accords with the reality circumstance and is comparatively briefness and easy accessibility.

(2) The way offered in this paper is to forecast the number of traffic accidents. The quality and quantity of the observed samples must determine the accuracy of the predicted results to some extent. Nevertheless, for limited data, grounded on the rolling optimization strategy, a small number of data can be used to complete the prediction. The example study shows that high forecast precision can be gained.

Acknowledgments

The authors wish to thank the anonymous reviewers for their serious, insightful and constructive opinions, which contributed to the improved version of this paper. This work has been partially supported by the 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.