Traffic accident prediction based on Markov chain cloud model

Xiaobin Ma¹, Jinyang Zhang¹, Peijie Huang¹, Huiyun Sang⁴, Guanglin Sun² and Jing Chen³

¹ Logistics Engineering Department, Shandong Jiaotong University, Jinan, Shandong Province, 250000, 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. A dynamic prediction model based on Markov chain and cloud model is established to predict the volume of road traffic accidents which is under the guidance of stochastic process and cloud theory referring to the characters of the road traffic accidents and time series data. First of all, we establish the evaluation model based on cloud model, and get the relative error range of the observed value and the predicted value. Then we can use Markov chain to correct the relative error. The result shows that this method balances both randomicity and fuzziness and has higher accuracy. Therefore, it can be used for analysing the trend of road traffic accidents in different traffic conditions and provide evidence for safety early warning and corresponding accident prevention countermeasures formulation in relative road segments.

1. Introduction
Currently, there are lots of prediction models such as Time Series Model[1,2], Gray Model[3,4], Neural Network Model[5,6], which have been widely used to predict road traffic accidents and achieved good prediction results. But all are imperfect. Time Series Model can describe the changing periodic law of data, however, it is unrealistic to record and predict all the data about relative road segments thus it is difficult to establish training and leaning model. The neural network method needs a large number of training data, and takes a lot of energy, therefore it is difficult to apply in practice, and its prediction accuracy is difficult to be guaranteed. Further, it is easy to have over fitting phenomenon when there are many learning samples. GM (1,1) needs a small number of samples to establish the model[7,8], but it requires the smoothness of original data, and the prediction period is short, so it can not reveal the law of periodic change of data. Whereas, the qualitative concept of cloud model can grasp the overall law, and transform the quantitative data into qualitative knowledge that divided as per historical periods using specific algorithms. Then, it activates the prediction rules of the corresponding qualitative knowledge to realize the uncertainty prediction of the observation value, according to the prediction time’s membership degree of different periods, so it has a better prediction effect. Markov chain can predict the development trend of the error according to the known error state, so as to realize the correction of the error. Therefore, it has certain advantages for the prediction of road traffic accident data series affected by various factors. Based on above analysis, this paper proposes a road traffic accident...
prediction method based on the cloud model and Markov chain [9,10]. Prediction analysis based on cloud model

1.1. Basic concepts of cloud model
Definition 1: Cloud is an uncertain transformation model between a qualitative concept and a numerical representation of the concept. It is assumed that X is a quantitative domain expressed by a precise numerical value and T is a qualitative concept related to this domain. The membership degree \( \mu(x) \) of element x in X to the qualitative concept expressed by T is not an accurate number, but a random number with stable tendency. \( \mu(x) \in [0,1] \), the distribution of this random membership degree in the domain is called membership cloud, which is called cloud for short. Amongst, each x, is called a cloud droplet[3].

Definition 2: Let U be a quantitative domain expressed by precise numerical value, and C is a qualitative concept on U. It assumes that quantitative value \( x \in U \), and \( x \) is a random realization of qualitative concept C, if x satisfies that \( x \sim N(Ex, En'^2) \) amongst \( En'^2 \sim N(En, He^2) \), and the certainty of x to C satisfies \( \mu(x) = \exp \left( -\frac{(x-Ex)^2}{2y^2} \right) \), then the distribution of x on the domain U is called normal cloud.

1.2. Predication method based on cloud model
In order to present the distributional law and uncertainty of the current sequence, proposes that to excavate the trend of data from existing data. Based on this, the observed value of the time series can be divided into two parts: historical data set H and current data set C.

Assumed that in the time series data set H, ai is the value with time property A (such as the ai-th study period ), bi is the value of numerical attribute B that at a certain time point of ai, the study purpose is to predict the value bt of the numerical attribute B at a certain point at in the future.

It is therefore necessary to excavate prediction knowledge from the database D firstly.

The implementation mechanism of predicted rule is as shown in figure 1, the algorithmic flow is as follows:

1) According to the established cloud model parameters, constructs the x-conditional cloud generator of A and y-conditional cloud generator of B respectively (i=1,2, ..., n).
2) Input the prediction time-antecedent value ai into x-conditional cloud generator of A and produced cloud droplet (ai, ui) (i=1,2,..., n).
3) Input the membership degree ui of cloud droplet (ai, ui) into y-conditional cloud generator of B and produced cloud droplet (bi, ui).
4) Output prediction value bi.

What needs illustration is that the cloud droplet and output value in the algorithm are not unique or definite, thus this rule based on cloud model can realize uncertain reasoning. Taking cloud B with the maximum membership \( u_i \) as the corresponding prediction knowledge, which is called historical cloud.

Historical cloud \( Bt1 \) (Ex1, En1, He1) can reveal the quasi-periodic law. Introduce current cloud \( It2 \) (Ex2, En2, He2) to represent the current trend, of which \( It2 \) (Ex2, En2, He2) can be obtained from current dataset C.

Carrying out inertia weighting of the historical cloud based on current cloud, and obtains the integrated cloud \( S(Et, En, He) \). The definition of the integrated cloud is as follows:
The integrated cloud $S$ ($E_t$, $E_n$, $H_e$) can integrate two different prediction knowledge. By replacing historical Cloud B with the integrated cloud $S$, traffic accident prediction rules is therefore generated.

2. Markov chain modified model

Markov chain prediction method is based on the prediction of state transition, dividing the random sequence into several states to predict the future states. Therefore, it is assumed that the relative error between the measured value and the predicted value is divided into $n$ state intervals, denoted as $S = \{S_1, S_2, \cdots, S_n\}$, the relative error is transferred to the state $S_j$ by $k$-step transition from the state $S_i$, and the state transition probability $p_{ij}(k)$ is:

$$p_{ij}(k) = \frac{f_{ij}(k)}{f_i} \quad (i = 1, 2, \cdots, n)$$

The error interval which is most likely to contain the predicted value calculated by cloud model can be judged and modified according to formula (3). Firstly, selected the states of $n$ known relative errors which are nearest to the prediction time, and then obtained the number of transition steps of the predicted value, and then extracted the row vectors corresponding to the initial state from the state transition matrix corresponding to the transition step number to form a new probability matrix. The sum of the $i$-th column vector of the new probability matrix represents the probability of the predicted value in the state $S_i$. The midpoint of the state $S_i$ with the maximum probability is taken as the most probable relative error, that is:

(1) If $p_{ij} = \max\{p_{ij} | p_{ij} \in P(k)\}, j = 1, 2, \cdots, n$, then the state change of accident volume is most likely to transit from state $E_k$ to state $E_j$.

(2) If two or more probabilities are the same in the row vector $p_{ij}, j = 1, 2, 3, \cdots, n$ and both are the maximum values, then the transition direction of the next step cannot be determined. In this case, it is necessary to calculate the second or above step state transition probability matrix.

Generally, the mid-value can be used as the prediction value of the future, then the single value of the prediction of the accident volume of the future time corresponding to the time interval can be obtained according to this state. The predicted values of the combined model are as follows:

$$F(x) = \frac{f(x)}{1 + \Delta}$$

$$\Delta = \frac{\Delta_u + \Delta_d}{2}$$

In the above formula, $F(x)$ is the revised predicted value by Markov model; $\Delta$ is the average relative error; $f(x)$ is the predicted value of the cloud model; $\Delta_u, \Delta_d$ are the upper and lower bounds of the state interval where $f(x)$ locates respectively.
3. Accuracy test of combined forecasting model

(1) average relative error $\lambda$

$$\lambda = \left(1 - \frac{1}{n} \sum_{t=1}^{n} \left[ \frac{|\delta(t)|}{x(t)} \right] \right) \times 100\%$$  (7)

(2) Mean square error ratio $C$

$$C = \frac{S_2}{S_1}$$  (8)

(3) Probability of minor error $Q$

$$Q = \left\{ \left| \frac{\delta(t) - \bar{\delta}(t)}{\delta(t)} \right| \leq 0.6745 \cdot S_1 \right\}$$  (9)

In the above formula: $\delta(t)$ and $\bar{\delta}(t)$ separately represent the residual sequence and average values of residual sequence, $S_1$ and $S_2$ separately represent the standard deviation of original sequence and residual sequence.

The precision of the grey Markov combined prediction model is generally divided into 4 grades.

| Model precision grade | Average relative error $\lambda$ | Mean square error ratio $C$ | Probability of minor error $Q$ |
|-----------------------|---------------------------------|----------------------------|-------------------------------|
| Level 1 (excellent)   | $0.01$                          | $\leq 0.35$                 | $0.95$                        |
| level 2 (good)        | $0.05$                          | $>0.35$–$0.50$              | $0.80$                        |
| level 3 (qualified)   | $0.10$                          | $>0.50$–$0.65$              | $0.70$                        |
| level 4 (unqualified) | $0.20$                          | $>0.65$–$0.80$              | $0.60$                        |

4. Case analysis

In order to research the trend of road traffic accidents, chooses the observed values form January 1, 2018 to January 1, 2019 in a subordinate districts and country under Jinan City and set them as the original sample, to as show in Figure 2 (each week serves as an observation period). Using the data from January 2, 2019 to September 1, 2019 (each week serves as an observation period) to predict based on the cloud model. Amongst all data, the first 31 periods are used as fitting samples to obtain relative error data, and the last 4 periods are used as prediction samples of the combined model.

Built time cloud with data according to the calculating procedure in chapter 2.2. Road traffic volume shows a certain trend of rising volatility as per Figure 2, which can be divided into four prediction intervals [10-30), [30-50), [50-70), [70-80), recorded as section 1, 2, 3, 4. Set the prediction value of data

![Figure 2. The traffic accidents volume](image-url)
in each period as historical data set H and constructed historical cloud B (Ex, En, He). Then distinguished the road traffic accident's membership interval in every historical moment and calculate its membership degree. Take the prediction of road traffic accidents volume in phase 32, 2019 as an example, based on current cloud, the historical cloud is processed by inertia weighting, to obtain the characteristic parameters therefore generated the integrated cloud S (Et, En, He), ss shown in Table 2.

| characteristic value | Interval 1 | Interval 2 | Interval 3 | Interval 4 |
|----------------------|-----------|-----------|-----------|-----------|
| Et                   | 24        | 42        | 59        | 72        |
| En                   | 22.1      | 19.6      | 26.4      | 18        |
| He                   | 3.5       | 4.2       | 4         | 2.9       |

Then carrying out prediction based on the integrated cloud and repeat for several times to get enough cloud droplets. Determine the prediction set of accidents volume at the prediction time according to the region where the cloud droplets are located. Obtained all prediction values in the prediction period and furthermore, obtained the predicted value based on average value of cloud droplets. Afterwards, introduced prediction value into current cloud, and the next prediction is completed by continuously updating current cloud model.

Use Markov chain to correct the error of prediction data and divide the state of relative error into four state intervals, as shown in Table 3.

| Status No. | Relative error interval % |
|------------|---------------------------|
| S1         | [-2,-1)                   |
| S2         | [-1,0)                    |
| S3         | [0,1)                     |
| S4         | [1,2)                     |

Input the relative error of cloud model prediction value into Markov chain. Take the relative error states of four research periods closest to the prediction time and calculate the steps required for the four known states transferring to the prediction research period. Four state transition matrices are calculated as follows.

\[
P(1) = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{6} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \quad P(2) = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \\
\begin{pmatrix} 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 \end{pmatrix} \quad P(3) = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix} \\
\begin{pmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \end{pmatrix} \quad P(4) = \begin{pmatrix} \frac{1}{5} & \frac{2}{5} & 0 & 0 \\ \frac{1}{5} & \frac{2}{5} & 0 & 0 \\ \frac{3}{5} & \frac{2}{5} & \frac{2}{5} & \frac{2}{5} \\ 0 & \frac{3}{4} & \frac{1}{4} & 0 \end{pmatrix} \]

Figure 3. Four state transition matrices

Then we take the corresponding row vectors in each state transition matrix, and finally sum them by columns to get the most likely state of relative error. The specific results are shown in Table 4.
Table 4. State of relative error

| Prediction period | Transfer steps | state  | S1    | S2    | S3    | S4    |
|-------------------|----------------|--------|-------|-------|-------|-------|
| 32 period         | 4              | S2     | 0.31  | 0.69  | 0     | 0     |
| 33 period         | 3              | S1     | 0.25  | 0.5   | 0.25  | 0     |
| 34 period         | 2              | S2     | 0.2   | 0.2   | 0.2   | 0.4   |
| 35 period         | 1              | S3     | 0     | 0.6   | 0.4   | 0     |
| Add up to         |                |        | 0.76  | 1.99  | 0.85  | 0.4   |

It can be seen from table 4 that the relative error of the first prediction period in 2019 is most likely in S2 state. From this we can calculate that the corrected value is 31, which belongs to interval 2. In this way, we can get the prediction values of other prediction periods, and then get the final prediction results of road traffic accidents based on the cloud model modified by Markov chain. As shown in Table 5.

Table 5. Accuracy test

| Prediction period | Measured value | Evaluation value |
|-------------------|----------------|------------------|
| The 32ed          | 31             | 28               |
| The 33d           | 34             | 35               |
| The 34th          | 28             | 28               |
| The 35th          | 24             | 21               |

According to the prediction results, the predicted C value is less than 0.5, and the prediction fitting accuracy is good. It proves that the cloud model based on Markov chain is effective and can be used for subsequent prediction to reflect the development trend of road traffic accidents.

5. Conclusion

Based on the existing literature, through in-depth analysis of their research ideas, and using results of relevant research results in econometrics field for reference, according to the characteristics of road traffic accidents and time series data, under the guidance of cloud model and Markov chain theory, a dynamic prediction model based on cloud model and Markov chain is established to predict road traffic accidents volume. The prediction accuracy is improved after the correction by Markov chain. The method provides a new idea for road traffic accident prediction.

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] Z Tang, W Dong, Z Wang. (2019) Research on the assessment model of urban road friction accident based on SVR. Journal of transportation engineering and information, 17:1-4.
[2] B Zhu, X Fu, S Yang, Z Zhu. (2018) Prediction model of traffic accident time and space influence based on nonlinear regression and BP neural network.
[3] S Jiang, K Song, W Xie, C Pan. (2019) Data analysis of road traffic accidents based on grey correlation and Apriori algorithm. Highwei engering, 44:67-73.
[4] Wang Dangli, Lv Xue, Wang Xuejia, Gao Ruijiang. (2018) Combined prediction model of water traffic accidents based on Iowa operator. Journal of ShangHai Maritime University, 39:39-40.
[5] J Yang, G Wang, Q Liu, Y Guo, Y Liu, W Gan, Y Liu. (2018) Process and conjecture of normal cloud model [J]. Journal of Computer Science, 41:724-744.
[6] B Fu, J Wang, N Zhang, W Xv. (2018) Short term passenger flow prediction of Urban Rail Transit Based on cloud model. Urban rail transit research, 21:61-65.