Multi Layer Hierarchical Fault Prediction Based on Multi Type Data Feature

Li Li, Gao Feng

HE XI UNIVERSITY, Gansu, Zhangye734000, China

Key Words: Traffic accidents, multi-layer hierarchical, Model analysis.

Abstract: For the road traffic accidents happened in a short period of time, the application of multi layer hierarchical prediction method in road traffic accident forcast has higher prediction accuracy compared with other methods. However, we find that the average error in predictions is nearly 5%, which still can't be regarded as the high forecasting accuracy. "Higher" used in the above sentence means relatively high, in this paper, we study on the factors which affect the accuracy of the prediction of multi layer hierarchical method further by analyzing the characteristics of historical data.

Introduction

Based on the information characteristics of a set of data, we put these data into two kinds, namely the smooth type and the wave type. Due to the fact that the multi layer hierarchical prediction method is more suitable for the accident forecast in a short period of time, we use the data of each month of 2012 and 2012 of M province as the historical data, and then intercept different data periods to establish the multi layer hierarchical prediction model respectively. We apply the variance formula $S = \dfrac{1}{n}[(x_1 - x)^2 + (x_2 - x)^2 + \cdots (x_n - x)^2]$ to inspect the data in the table above, and select the data fluctuate smoothly and the data fluctuate fiercely to make a comparison. The two sets of data are displayed in table 4-1 and table 4-2. The variances of the selected two groups of data are 0.00024 and 0.000839 respectively. As far as both two sets of data are concerned, we both build the multi layer hierarchical prediction model with the auto-regression item only to predict. In the forecast of smooth type data, we use the statistical data of 2012 to establish the forecast model of road traffic accidents. We take these data of October and September as the test value of prediction precision. For the forecast of wave type data, we use the statistical data of 2013 to establish the forecast model of road traffic accidents. We take these data of November and December as the test value of the prediction precision. For both of the above two data types, we build the forecast model that only contains the auto-regression item. Based on the characteristics of the above two groups of data, we establish the multi-degree prediction model of the second order, which is shown in the following equation:

$$y(k) = \alpha_1(k)y(k-1) + \alpha_2(k)y(k-2) + e(k) \qquad (1)$$

In the above equation, $\alpha_1(k)$ and $\alpha_2(k)$ are time-varying parameters

$e(k)$ Zero-expect white Noise

Prediction of Smooth Type Data

Table 1 Statistics of Road Traffic Accident Happened in M Provinces in 2012

Month	Accident Number	Death Toll (thousand)	The Number of the Injured
1	1070	0.288	1089
2	1183	0.310	1346
3	1421	0.316	1492
4	1357	0.324	1420
5	1097	0.308	1121
6	1160	0.311	1238
7	1193	0.343	1255
8	1359	0.391	1454
9	1471	0.371	1593

Based on the data given in Table 1, it is assumed that the initial value shown in the following equation of each time-varying parameter is 0.5:

$$\alpha_1(0) = \alpha_2(0) = 0.5;$$

According to the recursive valuation formula of time-varying parameters and the initial value of time-varying parameters, a series of valuations of time-varying parameters can be obtained as shown in Table 2.

Table 2 Sequence Valuation of Time-varying Parameters

k	$\hat{\alpha}_1(k)$	$\hat{\alpha}_2(k)$
3	0.529434	0.527345
4	0.518504	0.516623
5	0.481731	0.480757
6	0.492242	0.491814
7	0.554632	0.553603

To analyze the valuation sequence of the above mentioned $\alpha_1(k)$ and $\alpha_2(k)$, we can find their change law respectively, and based on the characteristics of their valuations sequence we establish forecast formula. As far as both $\alpha_1(k)$ and $\alpha_2(k)$ are concerned, we adopt the parameters of "AR model method by weight, and build their own 2 order AR model". And the prediction results as shown in Table 3.

Table 3 Prediction Results of Time-varying Parameter

k	$\hat{\alpha}_1(k)$	$\hat{\alpha}_2(k)$
8	0.583308	0.562028
9	0.530308	0.523028

Based on the above data, we can get the forecast values and errors of road traffic accidents shown in Table 4

Table 4 Forecast Values and Errors of Road Traffic Accidents

Month（Year 2013）	Actual value/one thousand people	Accident forecast/one thousand people	Relative error / %/%
8	0.391	0.375	4.13
9	0.371	0.387	4.25
Average relative error		4.19	

Prediction of Wave Type Data

Table 5 Road Traffic Accident Statistics of M Provinces in 2013

Month	Accident Number	Death Toll	The Number of Injured
4	1787	0.389	1749
5	1665	0.367	1710
6	1733	0.341	1816
7	1403	0.314	1484
8	1399	0.374	1670
9	1617	0.391	1812
10	1412	0.403	1487
11	1697	0.598	1646
12	2130	0.718	2019

Based on the data given in Table 5, the initial value of time-varying parameter in each period is taken as 0.5, and it is shown in the following equation:

$$\alpha_1(0) = \alpha_2(0) = 0.5;$$

According to the recursive valuation formula of the time-varying parameters and the initial values of time-varying parameters, a series of valuations of time-varying parameters can be obtained as in Table 6.

Table 6 Value Sequence of Time-varying Parameters

k	$\hat{\alpha}_1(k)$	$\hat{\alpha}_2(k)$
3	0.452523	0.449677
4	0.445264	0.441865
5	0.567299	0.574393
6	0.564901	0.57238
7	0.522235	0.531568

By analyzing the valuation sequences of $\alpha_1(k)$ and $\alpha_2(k)$, we can find their change law respectively,

and based on their own law we can establish the forecast formula. As far as both $\alpha_1(k)$ and $\alpha_2(k)$ are

concerned, we adopt the parameters of "AR model method by weight, and build their own 2 order AR model".

Based on the above data, we can get the forecast value and the error of road traffic accidents as shown in Table 7

Table 7 Forecast Value and Error of Road Traffic Accident

Month （Year 2013）	Actual value/one thousand people	Accident forecast/one thousand people	Relative error / %/%
11	0.598	0.562	6.04
12	0.718	0.68	5.4
Average relative error		5.72	

Comparative Analysis of the Prediction Precision of the Two Data

We find that in the application of using multi layer hierarchical forecast model to forecast the future traffic accidents, the different characteristics of the historical data directly influence the changes of the prediction accuracy, and with the historical data becoming more and more smooth, the prediction accuracy becomes higher and higher, and at the same time it points out the direction for us while we use the multi-degree prediction method to predict. Different prediction methods are chosen based on the different characteristics of the data to improve the accuracy of prediction.

The Change of Predictors

In the process of forecasting, the change of the predictors includes two types. One refers to we conduct the prediction analysis by choosing different predictors to reestablish model, while the other refers to we make each factor on the same magnitude level by proper handling of predictors. Through the study, we find that even though there is an obvious correlation between certain forecast factor and forecast quantity, but because the quantity is very small compared with other factors, as far as this forecast factor is concerned, the contribution rate for the equation will still be very small, even insignificant, so this phenomenon is clearly very unreasonable[05]. Therefore, in order to improve the prediction accuracy, we must process the forecast factors in a mathematical way, thus we should make them be on the same magnitude level.

In the above mentioned forecast, we selected the population density, per capita GDP, the highway passenger traffic, the highway freight volume, the road network density and per capita car ownership as the factors and carried out the relevant inspections. At the same time, the historical data also prove that the six factors are closely related to the road traffic accidents. Therefore, the change of predictors referred by us doesn't mean that we should choose different predictors, but it means that we should conduct the mathematical process on the existing factors. Commonly used processing methods mainly include the flattening, the geometric error and the standardization. Through the comparative analysis, we find that the effect of standardization toward the factors is good, so the standardization of each factor is used to compare the changes of prediction accuracy while we use the application multi-layer hierarchical prediction model to predict the road traffic accidents.

The Influence of Factor Changes on Accuracy

In the same way, we conduct the comparative analysis of data types from M province between2002 to 2013, and by the factors change conditions before and after, we compare the change conditions of the prediction accuracy. Here, we use the standardized variable formula:

$$u_i^{'}(k) = \frac{u_i(k) - \overline{u}_i}{S_i} \quad (i = (1,2,\cdots,m; k = 1,2,\cdots,n)) \tag{2}$$

In the above equation：
$$\overline{u}_i = \frac{1}{n}\sum_{k=1}^{n} u_i(k)$$

$$S_i = \sqrt{\frac{1}{n}\sum_{k=1}^{n}[u_i(k) - \overline{u}_i]^2}$$

M is the number of factors, n is the number of data sample, u_i is the mean value of the samples,

S_i is the mean square error.

After standardized treatment, each factor becomes a quantity having no unit, so we can conduct comparison among factors easily. Therefore, through the standardized treatment of factors, we can establish multi layer hierarchical prediction model, and this model could effectively eliminate the contribution difference caused by relative magnitude difference among factors; besides, the value change of each sample of the same variable is quite smooth, and the corresponding sequence change of time-varying parameter is quite smooth. Each factor after processing is shown in Table 8.

Table 8 Each Related Factors after Standardized Processing

Population density	density of road network	Per capita GDP	Per car quantity	Highway passenger transport volume	highway freight volume
-1.6489	-0.8311	-1.0862	-1.0929	-0.9628	-1.0104
-1.4114	-0.7770	-1.0280	-1.0443	-0.8134	-0.9582
-0.8574	-0.7342	-0.9339	-0.9914	-0.6927	-0.7998
-0.6200	-0.7274	-0.8368	-0.8899	-0.6153	-0.7162
-0.3825	-0.6846	-0.7109	-0.8105	-0.5415	-0.5752
-0.2242	-0.6531	-0.4957	-0.6913	-0.5199	-0.4672
0.0923	-0.6305	-0.1528	0.1118	-0.2822	-0.2757
0.4089	-0.5945	0.1910	0.4648	-0.1220	-0.0389
0.7255	1.2425	0.5856	0.9017	0.0599	0.2449
1.0421	1.3507	1.0139	1.0032	0.3119	0.7185
1.2795	1.4746	1.5769	1.2547	1.9072	1.6343
1.5961	1.5648	1.8769	1.7842	2.2709	2.2437

In the same way, we build multi layer hierarchical prediction model of road traffic accidents as the following equation

$$y(k) = \alpha_1(k)y(k-1) + \alpha_2 y(k-2) + \sum_{j=1}^{6} \beta_j(k)u_j(k) \tag{3}$$

In the above equation: the road traffic accident value of moment y (k)-k ,

$\alpha_1(k)$, $\alpha_2(k)$, $\beta_1(k)$, $\beta_2(k)$, $\beta_3(k)$, $\beta_4(k)$, $\beta_5(k)$, $\beta_6(k)$ – varying parameters while forecasting

$u_1(k), u_2(k), u_3(k), u_4(k), u_5(k), u_6(k)$ − certain influence index of road traffic safety impact. Selection of initial value is accomplished through the establishment of multiple linear regression model. Under the principle of minimum square method, we use the Matlab tool to conducted the fitting analysis on the model parameters to get the regression coefficient values, and then we take those obtained regression coefficient values as the initial values of corresponding parameters. After calculating, the initial values of the time-varying parameters are shown respectively:

$$\alpha_1(0) = -0.8954; \quad \alpha_2(0) = 2.0592; \quad \beta_1(0) = -2.2940;$$

$$\beta_2(0) = 0.0474; \quad \beta_3(0) = 1.0178; \quad \beta_4(0) = 0.04727;$$

$$\beta_5(0) = -1.4114; \quad \beta_6(0) = 1.3220;$$

The Estimated Values of Time-varying Parameters of the Model

Table 9 Estimated Value Sequences of Time-varying Parameters

k	$\hat{\alpha}_1(k)$	$\hat{\alpha}_2(k)$	$\hat{\beta}_1(k)$	$\hat{\beta}_2(k)$	$\hat{\beta}_3(k)$	$\hat{\beta}_4(k)$	$\hat{\beta}_5(k)$	$\hat{\beta}_6(k)$
3	-0.8857	2.0664	-2.3035	0.0392	1.0074	0.4616	-1.4191	1.3131
4	-0.8970	2.0565	-2.2965	0.0475	1.0169	0.4718	-1.4121	1.3213
5	-0.8967	2.0567	-2.2966	0.0474	1.0168	0.4716	-1.4122	1.3211
6	-0.8965	2.0569	-2.2966	0.0472	1.0166	0.4714	-1.4123	1.3210
7	-0.8954	2.0580	-2.2965	0.0465	1.0165	0.4716	-1.4126	1.3207
8	-0.8947	2.0589	-2.2961	0.0459	1.0166	0.4720	-1.4128	1.3207
9	-0.8939	2.0598	-2.2953	0.0472	1.0173	0.4730	-1.4127	1.3209
10	-0.8937	2.0599	-2.2950	0.0476	1.0175	0.4733	-1.4126	1.3211

Analyzing the estimated value sequences of time-varying parameters $\hat{\alpha}_1(k), \hat{\alpha}_2(k), \hat{\beta}_1(k), \hat{\beta}_2(k), \hat{\beta}_3(k), \hat{\beta}_4(k), \hat{\beta}_5(k)$ and $\hat{\beta}_6(k)$ respectively, based on their own laws, prediction formula can be established. Here we use the "averaging method[32]" to forecast the time-varying parameters. And the prediction results are shown in Table 10.

Table 10 Prediction Results of Time-varying Parameters

k	$\hat{\alpha}_1(k)$	$\hat{\alpha}_2(k)$	$\hat{\beta}_1(k)$	$\hat{\beta}_2(k)$	$\hat{\beta}_3(k)$	$\hat{\beta}_4(k)$	$\hat{\beta}_5(k)$	$\hat{\beta}_6(k)$
11	-0.8510	2.0849	-2.2020	0.1467	1.1319	0.5278	-1.3028	1.4299
12	-0.8592	2.0844	-2.2434	0.1229	1.0800	0.5132	-1.3316	1.3965

Based on the prediction results of time-varying parameters in Table 13, the forecast formula- $y(k)$ of forwards h steps is used to forecast the traffic accidents of the corresponding year.

Forecast value and error of road traffic accidents are shown in Table 11.

Table 11 Forecast Value and Error of Road Traffic Accidents

Year	Actual accident value / ten thousand	Accident forecast/ten thousand	Relative error/%
2012	0.503	0.524	4.07
2013	0.452	0.432	4.38
Average relative error		4.23	

Table 12 The Fourth Comparison Table of Road Traffic Accident Forecast Value

Factors before processing			Factors after processing		
Actual value	Predicted value	Relative error	Actual value	Predicted value	Relative error
0.503	0.473	-6.03%	0.503	0.524	4.07
0.452	0.417	-7.79%	0.452	0.432	4.38
Average error	6.91%		Average error	4.23%	

From the comparison tables above, it can be found that based on the same historical data of the same year we can conduct the prediction job, and the prediction accuracy results are very different. It suggests that the change of the predicting factors influence profoundly on the prediction accuracy results. When there is great difference among the quantity values of each predictor, in order to improve the accuracy of prediction we must properly process the factors.

Conclusion

The rapid increase of the cars puts more serious test to the traffic safety. With the continuous traffic accidents in our country, the traffic casualties and losses do serious harm to the economic development of our country, and the traffic safety issues have been paid much attention by the relevant departments. How to reduce the harm of traffic accidents and to minimize the damage of traffic accidents become the main topics of some scholars' researches. The best way to deal with the traffic problem is to avoid the accidents happen but take measures to remediate after the accidents. Thus making the accurate prediction of traffic accidents and effectively preventing them become particularly important.

Based on the analysis of the various accidents forecast methods of road traffic accidents at home and abroad, in this paper, we conduct the comparative analysis, and put forward the multi layer prediction method of road traffic accidents, and adopt this method to forecast the accidents happened in M province. In the process of prediction, by the comparison and analysis, we have carried on the thorough discussion on the influential factors of the prediction accuracy of multi layer hierarchical prediction model.

References

[1] Jia Xinzhang, Li Jingyuan. Electronic Industry, Vol. 6 (2004) No 53, p.25-26
[2] Peng Sue, Wang Yunhui, Wang Qunyong. Electronic Industry, Vol. 12 (2005) No 27, p.74-76

[3] Qian Xiyuan, Jing Jianfen, Hou XuSiem. Computer Engineering, Vol. 30 (2004) No 19, p.144-145

[4] Wang Kuailiang. Mining machinery, Vol. 29 (2008) No 27, p.21-23

[5] Zhang Gongxu, Sun Jing. New Quality Management, Vol. 8 (2003) No 27, p.57-60

[6] M. Colledani, T. Tolio. CIRP Journal of Manufacturing Science and Technology, Vol. 4 (2011) No 27, p. 281—289.