



Article Traffic Accident Prediction Based on Multivariable Grey Model

Wei Li¹, Xujian Zhao^{1,*} and Shiyu Liu²

- School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang 621010, China; vl729916@aliyun.com
- ² School of Information Engineering, Southwest University of Science and Technology, Mianyang 621010, China; Liusy5256@aliyun.com
- * Correspondence: jasonzhaoxj@Gmail.com

Received: 29 February 2020; Accepted: 29 March 2020; Published: 30 March 2020



Abstract: Owing to frequent traffic accidents and casualties nowadays, the ability to predict the number of traffic accidents in a period is significant for the transportation department to make decisions scientifically. However, owing to many variables affecting traffic accidents in the road traffic system, there are two critical challenges in traffic accident prediction. The first issue is how to evaluate the weight of each variable's impact on the accident. The second issue is how to model the prediction process for multiple interrelated variables. Aiming to solve these two problems, we propose effective solutions to deal with traffic accident prediction. Firstly, for the first issue, we exploit the grey correlation analysis to measure the correlation of factors to accident occurrence. Then, for the second issue, we select the main factors by correlation analysis to establish a multivariable grey model—MGM(1,N) for prediction process modeling. Further, we explore the collinearity between variables and better optimize the predictive model. The experimental results show that our approach achieves best performance than four general-purpose comparative algorithms in traffic accident prediction task.

Keywords: traffic accident prediction; multivariable grey model; grey correlation analysis

1. Introduction

With the rapid development of the automobile industry, the number of cars has increased dramatically. Meanwhile, due to the increase in traffic volume, the ratio of vehicles to traffic roads has been seriously unbalanced, which leads to frequent traffic accidents and casualties [1]. Therefore, it is necessary for traffic management departments to take measures to guarantee people's travel safety and the security of goods transportation [2–4]. However, the road traffic system is normally defined as a complex system with multiple factors, which greatly increases the difficulty for managers to deal with traffic problems. This is especially so, because the occurrence of traffic accidents is characterized by uncertainty [5], which often results in poor management of preventing traffic accidents. Although the occurrence of traffic accidents is very accidental, there is a certain regularity and relative stability for a large number of traffic accidents that occur in a region over a long period of time [6]. Consequently, we can make predictions on the number of possible traffic accidents in the next year through appropriate predictive models in order to make reasonable decision for traffic safety. However, owing to the many variables affecting traffic accidents in the road traffic system, we have to deal with two challenging issues in predicting the number of traffic accidents.

How to evaluate the weight of each factor's impact on the accident? Obviously, various factors have different impacts on the occurrence of traffic accidents, so they should play different roles in accident prediction modeling. Meanwhile, in the traffic system, each factor is normally defined as the variable

sequence which describes the development trend of corresponding factor chronologically. Thus, actually we need to measure the impact of each factor-variable sequence on the accident-variable sequence in order to determine appropriate factors to devise a reasonable model to solve accident prediction.

How do we model the prediction process for multiple interrelated variables? The occurrence of traffic accidents is considered to be correlated with several variables, so a multivariable model is required for accident prediction. Meanwhile, in the traffic system, one variable is not only affected by other variables, but it also affects them. Intuitively, exploring the high correlation between variables in a traffic system is beneficial to reducing the interaction between variables, which is helpful in addressing the issue of collinearity between variables and further improving model prediction accuracy. Consequently, we need to develop an effective approach to solve accident prediction considering the influence of the interaction among variables on the predicting results.

Aiming to solve these problems, in this paper we propose effective solutions to deal with traffic accident prediction. Firstly, we exploit the grey correlation analysis to measure the correlation of factors to accident occurrence. Following this, we select the main factors by correlation analysis to establish a multivariable grey model for prediction theory modeling. Further, we explore the collinearity between variables and better optimize the predictive model.

In summary, we make the following contributions in this paper:

We propose an effective solution to predict the number of traffic accidents by exploring multidimensional time series prediction strategies. Firstly, we use grey correlation analysis to measure the impact of the time series of each factor-variable on the time series of accident-variable. Following this, aiming to solve prediction, we select the main factors by correlation analysis to establish a multivariable grey model—MGM(1,N) which meets the nonlinearity and randomness of practical traffic system.

We present a general method to reduce the influence of the interaction among factor-variables on the predicting results in traffic system. Specifically, we build various MGM(1,N) models according to different factor-variables and further find the optimal model from candidates by measuring the collinearity between variables. The optimal MGM(1,N) model consists of the most influential factor-variables which can best reflect the change of the traffic system.

We conduct experiments on a real traffic statistics dataset for 13 years from China Statistical Yearbook. Compared with four general-purpose algorithms for accident prediction, the experimental results show that our approach achieves best performance than its competitors. In addition, the advantage of our proposal confirms the multivariable grey model has higher prediction accuracy than univariate predictive models in the task of accident prediction.

2. Related Work

Traffic accident prediction is very important for traffic management and decision-making, which has attracted more and more attention in recent years. At present, related work can be roughly divided into the following two aspects: traffic accident prediction based on machine learning and traffic accident prediction based on grey system.

Aimed at the issue of traffic accident prediction, the pioneer work [7] investigated the support vector machine model for predicting the number of road traffic accidents and evaluated the outperformance of their work by comparing with various models in terms of prediction accuracy. In order to achieve an acceptable level of accuracy for traffic accident risk prediction, the FP-tree method was proposed to select most important explanatory variables by Lin's work [8]. In addition, the logistic regression model [9] was also used to improve prediction accuracy while the random forest model was exploited to verify the rationality of the predictive result [10]. Moreover, to analyze the impact of rainfall on urban traffic accidents, Jaroszweski et al. [11] leveraged the weather radar model to rainfall quantification and matched-pairs analysis of traffic accidents and demonstrate divergence by two approaches. Meanwhile, in Theofilatos's work [12], the idea of finite mixture logit models combined with Bayesian model was used to investigate accident likelihood and severity on urban arterials. Additionally, Ren et al. [13]

analyzed the spatial and temporal patterns of traffic accident frequency and proposed a deep learning approach to predict the risk of citywide traffic accident. Considering the spatial heterogeneity challenge in the data, the special deep learning method Hetero-ConvLSTM framework was proposed in work [14], which made reasonably accurate prediction results.

As the extension of basic grey model—GM(1,1), the multivariable MGM(1,N) model was proposed to integrate multi-dimensions by Zhai [15], which can uniformly describe the interactional relationship of multi-variable system [16]. Specifically, it has been successfully applied in many aspects of engineering, such as ship motion [17], social economy [18], and wave prediction [19] and so on. Regarding for the studies on traffic accident prediction based on grey system, the traditional GM(1,1) model was considered to be reflecting the law of dynamic evolution in traffic accident and was applied into predictive analysis for different regions of China [20,21]. In addition, based on the GM(1,1) model, the improved and combined grey models were proposed in references [22] to predict traffic accidents and got more effective results than traditional algorithms. Additionally, in René S's work [23], in order to predict the development of road traffic accidents, the grey system theory was applied to predict traffic accident in Germany.

To our knowledge, so far there are very few works towards traffic accident prediction considering multidimensional time series of influential factors in traffic system. Most previous works only discussed the characteristics of one variable, namely the number of traffic accidents, and did not take the multi-factor effects of the traffic system into consideration. In this paper, we propose a new approach for traffic accident prediction. The unique features of our proposal lie in two designs. First, we exploit the grey correlation analysis to measure the correlation of factors to accident occurrence. Second, we explore the collinearity between variables and establish a multivariable grey model MGM(1,N) for accident prediction.

3. Methodology

In order to predict the number of traffic accidents, we present an effective predictive algorithm based on multivariable grey model MGM(1,N). Figure 1 shows the framework of our predictive system for traffic accidents.



Figure 1. Framework of the predictive system for traffic accidents.

3.1. Correlation Analysis for Accident Impact Factors

Traffic accidents are often affected by multiple factors which have complex relationships among them. Meanwhile, various factors have different influences on the occurrence of accidents. Thus, we aim to measure the impact of each factor variable on the accident variable in order to determine appropriate factors to devise a reasonable model to solve accident prediction. However, in the traffic system, the impact factor for traffic accidents is normally defined as the variable sequence which describes the development trend of corresponding factor chronologically. More specifically, variable sequences represent the time series of various variables we want to simulate and predict in the task of traffic accident prediction. Thus, in the paper, the grey correlation theory is employed to estimate whether the connection among different sequences is compact according to the similarity of sequence curve's geometrical shape and the correlation is measured by calculating the degree of closeness between the accident variable and its' impact factors.

Suppose that the i-th variable sequence with n time points in the traffic system is denoted by Formula (1):

$$x_i = \{x_i(k) | k = 1, 2, \cdots, n\} = (x_i(1), x_i(2), \cdots, x_i(n))$$
(1)

Further, there are m variable sequences in the traffic system, which includes the factor-variable sequences which describe the development trend of corresponding factors chronologically and the accident-variable sequence which represents the time series of the accident variable. Here the sequence in the traffic system is defined by Formula (2).

$$x_i = \{x_i(k) | k = 1, 2, \cdots, n\} = (x_i(1), x_i(2), \cdots, x_i(n)), \ i = 1, 2, \cdots, m$$
(2)

Then the grey correlation degree between two sequences is represented by Formula (3):

$$\xi_{i}(k) = \frac{\min_{s} \min_{t} \left| x_{0}(t) - x_{s}(t) \right| + \rho \max_{s} \max_{t} \left| x_{0}(t) - x_{s}(t) \right|}{\left| x_{0}(k) - x_{i}(k) \right| + \rho \max_{s} \max_{t} \left| x_{0}(t) - x_{s}(t) \right|}$$
(3)

where $\xi_i(k)$ is used to calculate the grey correlation coefficient at point k, which is similar to other correlation coefficient, such as the Pearson correlation coefficient. Furthermore, $\rho \in [0, 1]$ refers to the resolution coefficient. Additionally, in this formula, $\min_s \min_t |x_0(t) - x_s(t)|$ and $\max_s \max_t |x_0(t) - x_s(t)|$ respectively represent the two-stage minimum difference and the two-stage maximum difference. Based on Formula (3), the relevance of the sequence x_0 to the reference sequence x_i is represent by Formula (4):

$$r_i = \frac{1}{n} \sum_{k=1}^n \xi_i(k) \tag{4}$$

where r_i is the correlation degree between the factor-variable sequence and the accident-variable sequence.

3.2. Multivariable Grey Model MGM(1,N) for Accident Prediction

In the paper, we first consider the issue of accident prediction as a multiple time series prediction problem in order to describe the multivariable characteristics of the traffic systems. Secondly, although the occurrence of traffic accidents has certain contingency, the number of accidents is closely related to its past and present situation because there is a certain regularity and relative stability for a large number of traffic accidents that occur in a region over a long period of time. Namely, the number of traffic accidents can be predicted through the comprehensive analysis of its present situation and history. Based on these two hypothesis, we utilize the multivariable grey model MGM(1,N), greatly satisfying the multivariable stochastic process for the traffic system, to deal with accident prediction.

In the transportation system, one variable is not only affected by other variables, but also affecting other variables. It objectively reflects the laws that systems are interrelated and mutually influential. The MGM(1,N) model aims to uniformly describe the variables from the perspective of the system and to construct a traffic accident impact system, which can reflect the regular pattern. The specific modeling process is as follows:

Specifically, if there are *n* time series in a traffic system defined as $x_i^{(0)}(k)_i$, $(i = 1, 2, \dots, n)$, the accumulation sequence is represented by Formula (5).

$$x_i^{(1)}(k)_i = \sum_{j=1}^k x_i^{(0)}(j) \ (i = 1, 2, \cdots, n)$$
(5)

The multivariable grey model MGM(1,N) simulates prediction by establishing n first-order differential equations as shown in Formula (6).

$$\frac{dx_1^{(1)}}{dt} = a_{11}x_1^{(1)} + a_{12}x_2^{(1)} + \dots + a_{1n}x_n^{(1)} + b_1
\frac{dx_2^{(1)}}{dt} = a_{21}x_1^{(1)} + a_{22}x_2^{(1)} + \dots + a_{2n}x_n^{(1)} + b_2
\vdots
\frac{dx_n^{(1)}}{dt} = a_{n1}x_1^{(1)} + a_{n2}x_2^{(1)} + \dots + a_{nn}x_n^{(1)} + b_n$$
(6)

In order to solve MGM(1,N) model, the corresponding variables and parameters are calculated by Formulas (7) and (8), respectively.

$$\begin{aligned} x^{(0)} &= \left[x_1^{(0)}(k), x_2^{(0)}(k), \cdots, x_n^{(0)}(k) \right]^T \\ x^{(1)} &= \left[x_1^{(1)}(k), x_2^{(1)}(k), \cdots, x_n^{(1)}(k) \right]^T \end{aligned}$$
(7)

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}, \quad B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$
(8)

Based on the initializations of model variables and parameters represented by Formula (7) and (8), the n first-order differential equations, as shown in Formula (6), can be described by Formula (9).

$$\frac{dX^{(1)}}{dt} = AX^{(1)} + B \tag{9}$$

Further then, we revise Formula (9) into a continuous time response function defined by Formula (10) and further discretized by Formula (11).

$$X^{(1)}(t) = e^{At} X^{(1)}(0) + A^{-1}(e^{At} - 1) \cdot B$$
(10)

$$x_i^{(0)}(k) = \sum_{j=1}^n \frac{a_{ij}}{2} (x_j^{(1)}(k) + x_j^{(1)}(k-1)) + b_i \quad (i = 1, 2, \cdots, n; \ k = 2, 3, \cdots, m)$$
(11)

Additionally, in the paper, the least square method, defined by Formula (12), is utilized to solve the n first-order differential equations represented by Formula (9).

$$\hat{a}_{i} = \begin{pmatrix} \hat{a}_{i1} \\ \hat{a}_{i2} \\ \vdots \\ \hat{a}_{in} \\ b_{i} \end{pmatrix} = (L^{T}L)^{-1}L^{T}Y_{i} \quad i = 1, 2, \cdots, n$$
(12)

Finally, based on the differential equations results represented by Formula (13) and the recognized parameter matrices \hat{A} and \hat{B} defined by Formula (14), we solve the MGM(1,N) model for accident prediction, as shown in Formula (15).

$$L = \begin{pmatrix} \frac{1}{2}(x_{1}^{(1)}(2) + x_{1}^{(1)}(1)), & \frac{1}{2}(x_{2}^{(1)}(2) + x_{2}^{(1)}(1)), & \cdots, & \frac{1}{2}(x_{n}^{(1)}(2) + x_{n}^{(1)}(1)), & 1\\ \frac{1}{2}(x_{1}^{(1)}(3) + x_{1}^{(1)}(2)), & \frac{1}{2}(x_{2}^{(1)}(3) + x_{2}^{(1)}(2)), & \cdots, & \frac{1}{2}(x_{n}^{(1)}(3) + x_{n}^{(1)}(2)), & 1\\ & & \vdots\\ \frac{1}{2}(x_{1}^{(1)}(m) + x_{1}^{(1)}(m-1)), & \frac{1}{2}(x_{2}^{(1)}(m) + x_{2}^{(1)}(m-1)), & \cdots, & \frac{1}{2}(x_{n}^{(1)}(m) + x_{n}^{(1)}(m-1)), & 1 \end{pmatrix}$$

$$Y_{i} = (x_{i}^{(0)}(2), x_{i}^{(0)}(3), \cdots, x_{i}^{(0)}(m))$$
(13)

$$\hat{A} = \begin{pmatrix} \hat{a}_{11} & \hat{a}_{12} & \dots & \hat{a}_{1n} \\ \hat{a}_{21} & \hat{a}_{22} & \dots & \hat{a}_{2n} \\ & & \vdots \\ \hat{a}_{n1} & \hat{a}_{n2} & \dots & \hat{a}_{nn} \end{pmatrix}, \quad \hat{B} = \begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \vdots \\ \hat{b}_n \end{pmatrix}$$
(14)

$$\hat{X}^{(1)}(k) = e^{\hat{A}(k-1)}X(1) + \hat{A}^{-1}(e^{\hat{A}(k-1)} - I)\cdot\hat{B},
\hat{X}^{(0)}(1) = X^{(0)}(1)
\hat{X}^{(0)}(k) = \hat{X}^{(1)}(k) - \hat{X}^{(1)}(k-1) \qquad k = 2, 3, \cdots$$
(15)

3.3. Variables Selection for Model Optimization

Although correlation analysis for accident impact factors is able to be used for measuring the impact of each factor variable on the accident variable, the correlation between factor variables in traffic system is not considered in the task. However, the model prediction accuracy is greatly affected by the correlation between factor variables in traditional regression prediction task. Thus, in the paper we consider the issue of variables selection for model optimization inspired by the regression task. Algorithm 1 shows the details of the variables selection algorithm. Practically, there are three steps to fulfill variables selection in model optimization task. Firstly, we rank the grey correlation of each impact factor as discussed in 3.1 (line 2); secondly, we select the factors whose grey correlation coefficient is larger than the threshold from all the impact factors and establish different MGM(1,N) models with these factors respectively (lines 3 to 7); and then, we use the metric MAPE to measure the effectiveness of each model on the validation set in order to find the optimal solution quickly (lines 8 to 13).

Algorithm 1. Variables Selection for Model Optimization.

Input: W: The sequence of grey correlation degree of impact factors.

Output: *S*: The set of selected impact factors which optimize the performance of the prediction model.

Preliminary: *W* is represented by $[w_1, ..., w_n] \in (0, 1)$, which w_i is the grey correlation degree of the *i*th impact factor. 1. **begin**

- 5. Find the max number *k*, which meets $w_k > \theta_0$;
- 6. end for
- 7. Construct the sequence $W' = [w_1, \dots, w_k] \in (0, 1);$
- 8. for each w_i in W' do
- 9. Measure *MGM*(1, *j*) model by metric *MAPE*;
- 10. Update the *MAPE* minimum of *MGM*(1,*N*) model;
- 11. end for
- 12. **return** $s = \operatorname{argmin}MAPE(MGM(1, N));$

j∈N

13. Retrieve the corresponding impact factor from the sequence $[w_1, \ldots, w_s]$ and **return** *S*.

14. end

^{2.} Rank the sequence *W* in descending order;

^{3.} Set a threshold θ_0 ;

^{4.} for each w_i in W do

3.4. Model Evaluation

Aiming to compare proposed method with comparative methods, we perform model evaluations for prediction task. Specifically, we measure the accuracy of our model in terms of different metrics, namely MAPE, MAE and RMSE. The specific evaluation indexes are defined as follows:

$$MAPE = \frac{1}{n} \sum_{k=1}^{n} \left| \frac{x_i^{(0)}(k) - \hat{x}_i^{(0)}(k)}{x_i^{(0)}(k)} \right| \times 100\%$$
(16)

$$MAE = \frac{1}{n} \sum_{k=1}^{n} \left| x_i^{(0)}(k) - \hat{x}_i^{(0)}(k) \right|$$
(17)

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^{n} \left(x_i^{(0)}(k) - \hat{x}_i^{(0)}(k) \right)^2}$$
(18)

where $x_i^{(0)}(k)$ and $\hat{x}_i^{(0)}(k)$ respectively represent the ground truth of the variable sequence and the simulation result by the predictive model.

4. Experiment

In this section, we discuss the performance evaluation of our algorithms. We aim to measure the effectiveness of our proposal for traffic accident prediction.

4.1. Dataset

In the paper, the dataset comes from China Statistical Yearbook, which consists of the traffic-related statistical data of China in total and all provinces or cities individually from 2004 to 2016. Here four impact factors, namely the number of private cars, the number of taxis, the number of road operating cars and the urban population, affecting the number of traffic accidents are considered in our work. The details of the dataset are as shown in Table 1.

Year	Private Car (Million)	Taxi	Road Operating Car (Million)	Population (Million)	Traffic Accidents
2004	14.8166	903,734	10.6718	1299.88	517,889
2005	18.4807	936,973	7.3322	1307.56	450,254
2006	23.3332	928,647	8.0258	1314.48	378,781
2007	28.7622	959,668	8.4922	1321.29	327,209
2008	35.0139	968,811	9.3061	1328.02	265,204
2009	45.7491	971,579	10.8735	1334.50	238,351
2010	59.3871	986,000	11.3332	1340.91	219,521
2011	73.2679	1,002,306	12.6375	1347.35	210,812
2012	88.3860	1,026,678	13.3989	1354.04	204,196
2013	105.0168	1,053,580	15.0473	1360.72	198,394
2014	123.3936	1,074,386	15.3793	1367.82	196,812
2015	140.9910	1,092,083	14.7312	1374.62	187,781
2016	163.3020	1,102,563	14.3577	1382.71	212,846

Table 1. The traffic-related statistical data of China from 2004 to 2016.

4.2. Results of Grey Correlation Analysis

Firstly, we take equation $\frac{x_i - \bar{x}}{\sigma}$ to eliminate the scale difference between the dimension and the unit of each factor, where \bar{x} and σ are the mean value and the standard deviation of the variable sequence x_i respectively. Following this, the Formulas (3) and (4) defined in Section 3.1 are used to calculate the correlation coefficient of each impact factor and find the main factor variable that has a great influence on reference series.

The results of Table 2 show that the number of private cars as the factor has a smaller impact on the number of traffic accidents. Thus, we do not consider this factor in model selection through setting a correlation threshold while the other three factors, namely taxi, road operating car, and population are taken as independent influencing variables of traffic accident. Consequently, we build different MGM(1,N) models respectively in terms of the results of correlation analysis. Here we use the metric MAPE to measure the effectiveness of each model. The evaluation results, as shown in Table 3, indicate that considering the factors having the most influence on the number of accidents will optimize the model performance. The reason is that the collinearity between different variables will greatly affect the prediction ability of the MGM(1,N) model. For example, the increase of population will affect the rise of the number of taxis, so the redundancy among variables will make it difficult for predictive model to work well. Additionally, the most influential factor-variable is considered to greatly reflect the change of the traffic system. Therefore, the MGM(1,2) model with the most influential factor-variable, namely the number of road operating cars, is considered in the paper.

Impact Factor	Degree of Correlation		
Private car	0.6216		
Taxi	0.8623		
Road operating car	0.8837		
Population	0.8739		

Table 2. Influence degree of each impact factor on traffic accident.

Model MGM(1,2) Independent Variables Road Operating Car		MGM(1,3) Road Operating Car and Population	MGM(1,4) Road Operating Car, Population and Taxi	
MAPE	1.48%	1.52%	3.79%	

4.4. Results of Traffic Accident Prediction

For the evaluation of overall performance, we compare our algorithm with four existing methods, including GM(1,1) model, MGM(1,5) model, linear regression model and BP neural networks. Actually, the statistics of China from 2004 to 2012 are used as training data, while data from 2013 to 2016 are used for model evaluation. Meanwhile we measure the model effectiveness in terms of the metric MAE, MAPE and RMSE through calculating the difference between the fitted value and ground truth.

The result as shown in Table 4 indicates that our model achieves better results compared with other algorithms in the task of China traffic accident prediction. Specifically, our method gets the minimal deviation measured by the evaluation metrics both in Table 4. Meanwhile the evaluation results show that tradition grey model GM(1,1) and MGM(1,5) are not applicable for our task because of the complexity and nonlinearity in transportation system. In addition, our approach performs better than linear regression model and BP neural networks with respect to the metrics MAE and MAPE in prediction task. The reason is that our model is more suitable for small sample data while the machine learning approaches cannot work well on it.

Additionally, in order to evaluate the performance of our proposal and comparative methods in terms of generalization, we conduct comparative experiments on two different corpus selected from China Statistical Yearbook as mentioned in Section 4.1, namely the traffic-related statistical data of Zhe Jiang province and Chong Qing city. The results are shown in Table 5, where we use the same metrics as in the national data measurement in Table 4, including MAE, MAPE, and RMSE, to evaluate the performance. As Table 5 shows, our model is superior to the other four approaches in terms of all the metrics.

Year	Ground Truth –	GM(1,1) Model	MGM(1,5) Model	Linear Regression	BP Neural Network	Our Model
		Fitted Value	Fitted Value	Fitted Value	Fitted Value	Fitted Value
2013	198,394	152,680	192,351	204,321	203,372	199,773
2014	196,812	134,123	186,158	204,317	203,188	202,247
2015	187,781	117,822	180,781	206,867	202,976	208,406
2016	212,846	103,502	175,094	208,365	202,952	217,798
М	IAPE	35.88	7.48	4.77	4.62	4.19
N	ЛАЕ	71,925.50	15,362.25	9294.67	9110.77	8097.75
RMSE		75,614.24	20,150.92	10,906.59	9927.52	10,969.89

Table 4. Evaluation results for different predictive models on the statistical data of China.

Table 5. Evaluation results for different predictive models on the statistical data of Zhe Jiang province and Chong Qing city in China.

	Chong Qing City			Zhe Jiang Province		
Region	MAPE (%)	MAE	RMSE	MAPE (%)	MAE	RMSE
GM (1,1)	23.55	1191.58	1196.59	26.30	4310.30	4343.10
MGM (1,5)	87.00	4223.76	5166.16	37.06	5867.46	6739.37
LR	11.62	566.93	659.01	8.79	1383.98	1634.24
BPNN	16.26	798.28	894.36	20.08	3139.53	3882.37
Our Model	5.76	286.19	322.09	8.20	1279.53	1595.61

4.5. Results of Trend Prediction

In order to evaluate the overall performance of our proposal and comparative methods in time series prediction, we demonstrate the trend prediction results for country and regional data sets respectively. As shown in Figures 2–4, the trend curve by our model best fits the ground truth curve in the whole time period on different corpus, which suggests our proposal can greatly indicate the dynamic variability of the traffic system. Note that although the regression model and neural network-based approach also perform well in the task of time series prediction, the results of machine learning method are sensitive to the size of training sets.



Figure 2. Result of trend prediction for Chong Qing city.



Figure 3. Result of trend prediction for Zhe Jiang province.



Figure 4. Result of trend prediction for the country of China.

5. Conclusions

In this paper, we propose an effective approach to deal with traffic accident prediction. In particular, we exploit the grey correlation analysis to measure the correlation of factors to accident occurrence. Following this, we explore the collinearity between variables and establish a multivariable grey model MGM(1,N) for accident prediction. We conduct comparative experiments with different existing methods on a real data set containing the Chinese traffic-related statistics from 2004 to 2016 to measure the performance of our proposal, and the results demonstrate the superiority of our proposal with respect to various metrics.

Author Contributions: Conceptualization, W.L., X.Z. and S.L.; Data curation, W.L.; Methodology, W.L. and X.Z.; Resources, W.L.; Supervision, X.Z.; Validation, W.L., X.Z. and S.L.; Writing—original draft, W.L. and S.L.; Writing—review & editing, W.L. and X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Humanities and Social Sciences Foundation of the Ministry of Education, grant number 17YJCZH260, and the CERNET Innovation Project, grant number NGII20180403.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Ki, Y.; Lee, D. A traffic accident recording and reporting model at intersections. *IEEE Trans. Intell. Transp. Syst.* 2007, *8*, 188–194. [CrossRef]
- 2. Zhidan, L.; Zhenjiang, L.; Kaishun, W.; Mo, L. Urban traffic prediction from mobility data using deep learning. *IEEE Netw.* **2018**, *32*, 40–46.
- 3. Pecherkova, P.; Nagy, I. Analysis of Discrete Data from Traffic Accidents. In Proceedings of the 2017 IEEE International Conference on Smart City Symposium Prague(SCSP), Prague, Czech Republic, 25–26 May 2017.
- 4. Addi, A.M.; Tarik, A.; Fatima, G. An approach Based on Association Rules Mining to Improve Road Safety in Morocco. In Proceedings of the 2016 IEEE International Conference on Information Technology for Organizations Development (IT4OD), Fez, Morocco, 30 March–1 April 2016.
- 5. Zou, T.; Li, H.; Li, Y.; Cai, M.; Xiao, J. Methods for calculating intervals of reconstructed results in traffic accidents with different types of interval traces. *J. Shanghai Jiaotong Univ.* **2017**, *22*, 555–561. [CrossRef]
- 6. Xu, Q.; Zheng, J.; Sun, C. Digital Modeling Analysis of Urban Road Traffic Capacity under the Condition of Traffic Accidents. In Proceedings of the 2017 29th Chinese Control and Decision of the Conference (CCDC), Chongqing, China, 28–30 May 2017.
- Cheng, R.; Tian, X.; Zhang, M. Prediction Model for Road Traffic Accident Based on Support Vector. In Proceedings of the Education Science and Development of the International Conference (ICITIA), Guangzhou, China, 21–22 December 2018.
- 8. Lin, L.; Wang, Q.; Adel, W. A novel variable selection method based on frequent pattern tree for real-time traffic accident risk prediction. *Trans. Res. Part C Emerg. Technol.* **2015**, *55*, 444–459. [CrossRef]
- Tao, L.; Zhu, D.; Yan, L. The Traffic Accident Hotspot Prediction Based on the Logistic Regression Method. In Proceedings of the 2015 IEEE International Conference on Transportation Information and Safety (ICTIS), Wuhan, China, 25–28 June 2015.
- Cheng, R.; Zhang, M.; Yu, X. Prediction Model for Road Traffic Accident Based on Random Forest. In Proceedings of the Education Science and Development of the International Conference (ICESD), Jakarta, Indonesia, 12–13 January 2019.
- 11. Jaroszweski, D.; McNamara, T. The influence of rainfall on road accidents in urban areas: A weather radar approach. *Travel Behav. Soc.* **2014**, *1*, 15–21. [CrossRef]
- 12. Theofilatos, A. Incorporating real-time traffic and weather data to explore road accident likelihood and severity in urban arterials. *J. Safety Res.* **2017**, *61*, 9–21. [CrossRef] [PubMed]
- 13. Ren, H.; Song, Y.; Wang, J.; Hu, Y.; Lei, J. A deep learning approach to the citywide traffic accident risk prediction. *arXiv* **2017**, arXiv:1710.09543.
- 14. Yuan, Z.; Zhou, X.; Yang, T. Hetero-ConvLSTM: A Deep Learning Approach to Traffic Accident Prediction on Heterogeneous Spatio-Temporal Data. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 984–992. [CrossRef]
- 15. Zhai, J.; Sheng, J.M. The grey model MGM(1,n) and its application. *Syst. Eng. Theory Pract.* **1997**, *17*, 109–113.
- 16. Guo, X.J.; Liu, S.F. A multi-variable grey model with a self-memory component and its application on engineering prediction. *Eng. Appl. Artif. Intell.* **2015**, *42*, 82–93. [CrossRef]
- Shen, J.H.; Wang, S.J.; Hu, M.M. The Prediction of Nonlinear Ship Motion Based on Multi-Varible Grey Model MGM(1,n). In Proceedings of the 2007 IEEE International Conference on Mechatronics and Automation, Harbin, China, 5–8 August 2007.
- 18. Xiong, P.P.; Dang, Y.G.; Wu, X.H.; Li, X.M. Combined model based on optimized multi-variable grey model and multiple linear regression. *Syst. Eng. Electron.* **2011**, *22*, 615–620. [CrossRef]
- 19. Truong, D.Q.; Ahn, K.K. Wave prediction based on a modified grey model MGM(1,1) for real-time control of wave energy converters in irregular waves. *Renew. Energy* **2012**, *43*, 242–255. [CrossRef]
- 20. Wang, L.; Zhu, J.; Lu, H.; Zheng, Y. Forecasting of Traffic Accident in Shanxi Province Based on Grey System Theory. In Proceedings of the 2012 IEEE International Conference on Remote Sensing, Nanjing, China, 1–3 June 2012.

- 21. Wang, L.; Zheng, Y. Comparison of Macro Prediction for Traffic Accident between Beijing and Tianjin. In Proceedings of the 2012 IEEE International Conference on Computational and Information Sciences, Chongqing, China, 17–19 August 2012.
- 22. Zhao, L. Traffic accident prediction based on equal dimension and new information unbiased grey markov model. *Computer Eng. Appl.* **2013**, *49*, 35–383.
- 23. René S, H.; Becker, U.; Manz, H. Grey Systems Theory Time Series Prediction applied to Road Traffic Safety in Germany. *IFAC-Papers OnLine* **2016**, *49*, 231–236. [CrossRef]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).