Hindawi
Journal of Control Science and Engineering
Volume 2020, Article ID 4206919, 10 pages
https://doi.org/10.1155/2020/4206919

Research Article

Traffic Accident Prediction Based on LSTM-GBRT Model

Zhihao Zhang, Wenzhong Yang, and Silamu Wushour

College of Information Science and Engineering, Xinjiang University, Urumqi 830046, China

Correspondence should be addressed to Wenzhong Yang; ywz_xy@163.com

Received 9 September 2019; Accepted 15 November 2019; Published 5 March 2020

Academic Editor: Juan-Albino Méndez-Pérez

Road traffic accidents are a concrete manifestation of road traffic safety levels. The current traffic accident prediction has a problem of low accuracy. In order to provide traffic management departments with more accurate forecast data, it can be applied in the traffic management system to help make scientific decisions. This paper establishes a traffic accident prediction model based on LSTM-GBRT (long short-term memory, gradient boosted regression trees) and predicts traffic accident safety level indicators by training traffic accident-related data. Compared with various regression models and neural network models, the experimental results show that the LSTM-GBRT model has a good fitting effect and robustness. The LSTM-GBRT model can accurately predict the safety level of traffic accidents, so that the traffic management department can better grasp the situation of traffic safety levels.

1. Introduction

"By 2020, half the number of global deaths and injuries from road traffic accidents" is one target of the Sustainable Development Goals (SDGs) published by the United Nations (UN) in 2015 [1]. The country's attention to traffic safety continues to increase. Applying traffic accident situation prediction results to traffic planning can improve traffic safety. Many experts and scholars have predicted some indicators of traffic accidents [2, 3]. The research methods are mainly divided into three categories, statistical regression method [4], grey prediction [5], and neural network model method.

Statistical regression methods include time series prediction and many classic traffic accident experience models (Smid model, I. Agalal model, Japanese model, and Beijing model). Yannis et al. [6] proposed an autoregressive nonlinear time-series modelling of traffic fatalities in Europe. Kumar and Toshniwal [7] proposed a novel framework for time series data of road traffic accidents, which segments the time series data into different clusters for trend analysis. Ihueze and Onwurah [8] analyzed road traffic crashes in Anambra State, Nigeria, with the intention of developing accurate predictive models for forecasting crash frequency in the state using autoregressive integrated moving average (ARIMA) and autoregressive integrated moving average with explanatory variables (ARIMAX) modelling techniques. The regression model is simple and convenient to calculate, and it can predict short-term data changes. The essence of the regression model is the linear fit to the data. However, the results predicted by the model are one sided and weak in anti-interference ability. Due to the randomness of traffic accidents themselves, there are many influencing factors. Therefore, the reliability of its prediction results is not guaranteed.

The grey prediction model can predict a small number of samples, and the principle is simple, the operation speed is fast, and the testability is strong. The grey prediction model can make short-term and medium-term macropredictions for data with little fluctuation. The essence of the model is to find the dynamic relationship between the road traffic accident sequence data. However, the grey theory is modeled for a class of series that conforms to the condition of a smooth discrete function, and the grey system model describes only a process that monotonically increases or decays exponentially over time. Shi et al. [9] proposed a sequence GM (1, 1) model with strong exponential law to predict traffic accidents, but the model can only describe the monotonous change process. Hosse et al. [10] applied a Grey Systems Theory MGM (1, 4) in order to predict the development of road traffic accidents in Germany until 2025 based on the market diffusion of electronic stability program

(ESP). Liu and Wu [11] proposed a grey Verhulst prediction model for road traffic accidents, which is suitable for nonmonotonic wobble development sequences or S-shaped sequences with saturation. Zhao et al. [12] proposed a model that weighted and combined a variety of grey prediction methods. Although the prediction accuracy has been improved, its essence is a linear combination of the original data and there are still shortcomings in the medium- and long-term prediction.

The neural network prediction method has strong nonlinear mapping ability, high robustness, and powerful self-learning ability and has been widely used in many fields. He and Guo [13] proposed a traffic accident prediction model based on the BP neural network. The model can implement any nonlinear mapping, especially suitable for complex internal mechanisms. The shortcomings of the BP neural network model include slow training convergence, long training time, and easy to fall into the saddle point. Liwei et al. [14] proposed a grey neural network model. The grey theory compensated for the shortcomings of data mining for small sample data distortion, while the neural network compensated for the shortcomings of grey theory that can only be used for short-term prediction. Although the model improves the training speed, the accuracy of the model prediction results is low and the deviation is too large.

This paper proposed an LSTM-GBRT model for traffic accident prediction. The LSTM layer captures time-dependent information in the data; the GBRT model has the advantage of high robustness of ensemble learning for model training.

2. Related Theory

2.1. Long Short-Term Memory. The LSTM [15] model proposed by Hochreiter et al. is a variant of the recurrent neural network (RNN). It builds a specialized memory storage unit that trains the data through a time backpropagation algorithm. It can solve the problem that the RNN has no long-term dependence. The schematic diagram of the LSTM structure is shown in Figure 1.

The standard LSTM can be expressed as follows. Each step t and its corresponding input sequence are $X = \{x_1, x_2, \ldots, x_t\}$, the input gate is t, the forget gate is i_t, and the output gate is f_t. Memory cell state c_t controls data memory and oblivion through different gates. The formula is as follows:

$$\begin{aligned} i_t &= \sigma\left(W_i x_t + U_i h_t\right), \\ f_t &= \sigma\left(W_f x_t + U_f h_t\right), \\ o_t &= \sigma\left(W_o x_t + U_o h_t\right), \\ \widetilde{c}_t &= \tanh\left(W_c x_t + U_c h_t\right). \end{aligned} \tag{1}$$

The memory cell state c_t^j of the unit time t of the jth LSTM is as follows:

$$c_t^j = i_t^j \circ \widetilde{c}_t + f_t^j \circ c_{t-1}^j. \tag{2}$$

After the memory cell state is updated, calculate the current hidden layer h_t^j:

$$h_t^j = o_t^j \circ \tanh\left(c_t^j\right), \tag{3}$$

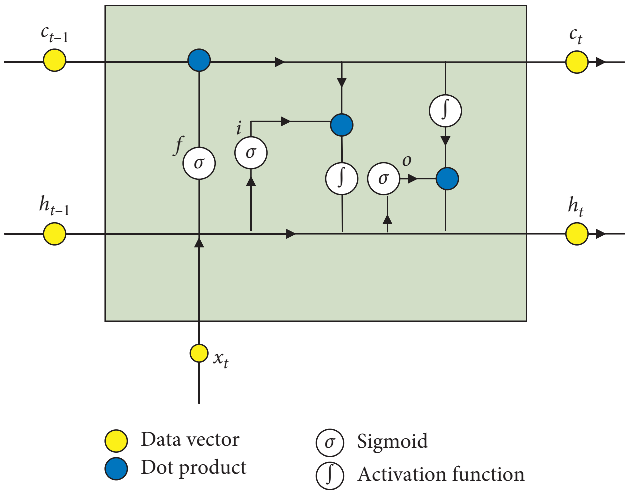

Figure 1: LSTM structure diagram.

where W is the weight matrix of the input, U is the state transition weight matrix, σ is the sigmoid function, tanh is the hyperbolic tangent function, h_t is the hidden state vector of the output, $\widetilde{c}_t$ is the new cell state after the adjustment and update, and "$\circ$" indicates point multiplication. The three types of gates jointly control the information entering and leaving the memory cell state, and the input gates adjust new information into the memory cells; the forgetting gate controls how much information is stored in the memory cells and how much information can be output by the output gate definition. The gate structure of the LSTM allows the information in the time series to form a balanced long short-term dependency.

2.2. Boosting Ensemble Learning Framework. GBRT model is a boosting [16] type ensemble learning algorithm. Ensemble learning is a technical framework that combines multiple different models to perform the corresponding tasks in order to achieve more efficient and accurate arrival. Currently used ensemble learning frameworks include bagging, boosting, and stacking. The training process of the boosting framework is stepped, the base model is trained in order, and the training set of the base model is transformed according to a certain summary strategy. Then, the prediction results of all the base models are linearly integrated to produce the final prediction result. Figure 2 is a schematic diagram of the boosting ensemble learning framework.

The overall model based on the boosting framework can be described by a linear combination:

$$F(x) = \sum_{i}^{m} h_i(x), \tag{4}$$

where $h_i(x)$ is the product of the base model and its weight. The training goal of the overall model is to approximate the predicted value $F(x)$ to the true value y, that is, to make the predicted value of each base model approximate the partial true value to be predicted. h_t is tested by using training examples, and the weight of misclassified instances is increased. The researchers came

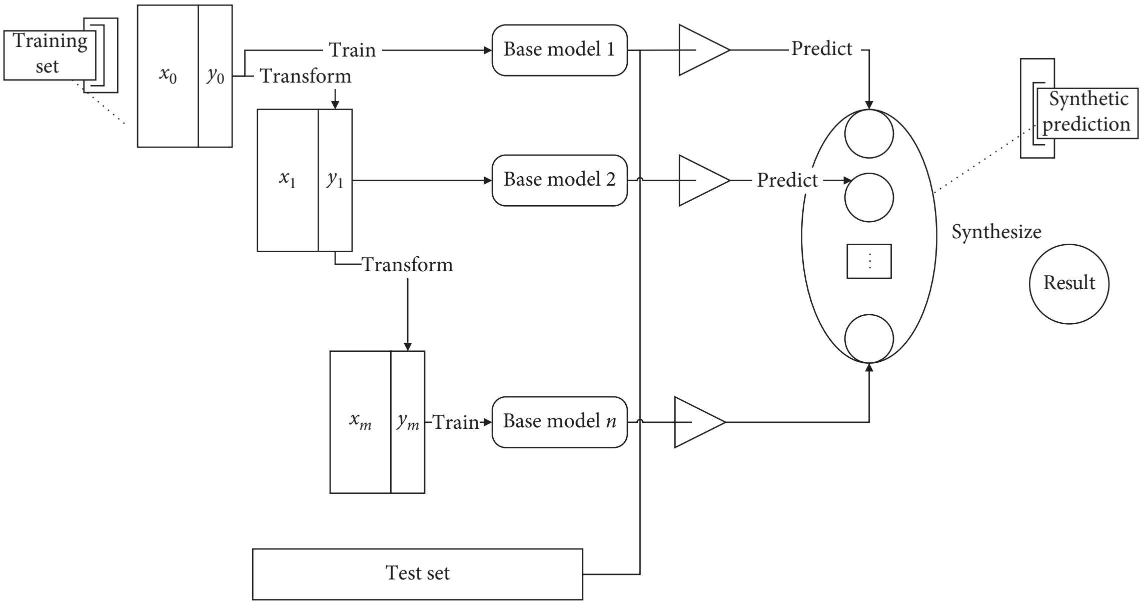

Figure 2: Boosting ensemble learning framework.

up with a greedy solution: train only one base model at a time, and in each iteration, focus on one base model training problem:

$$F^i(x) = F^{i-1}(x) + h_i(x). \tag{5}$$

Fit the residual. Introducing an arbitrary loss function and fitting the inverse gradient

$$F^i(x) = F^{i-1}(x) + \arg\min \sum_j^n L\left(y_j, F^{i-1}\left(x_j\right)\right) + h_i\left(x_j\right). \tag{6}$$

2.3. Gradient Boosted Regression Trees Model. For a given data set with n examples and m features $D = \{(x_i, y_i)\}\,(|D| = n, x_i \in \mathfrak{R}^m, y_i \in \mathfrak{R})$, a tree ensemble model uses K additive functions to predict the output.

$$\hat{y}_i = \phi(x_i) = \sum_{K=1}^{K} f_k(x_i), \quad f_k \in \Gamma, \tag{7}$$

where $\Gamma = \left\{f(x) = w_{q(x)}\right\}(q\colon \mathfrak{R}^m \longrightarrow \mathrm{T}, \omega \in \mathfrak{R}^{\mathrm{T}})$ is the space of regression trees. Here, q represents the structure of each tree that maps an example to the corresponding leaf index, T is the number of leaves in the tree, and each f_k corresponds to an independent tree structure q and leaf weights w. Unlike decision trees, each regression tree contains a continuous score on each of the leaf, and we use w_i to represent score on the ith leaf.

3. Data Source

3.1. Road Safety Impact Factor Data. As we all know, the occurrence of traffic accidents is caused by the combination of factors such as people, vehicles, roads, and the environment. People include pedestrians and drivers; vehicles include motor vehicles and nonmotor vehicles on the road; road conditions are the condition of the road; environment refers to the natural environment and social environment, and the social environment includes political, economic, cultural, and other factors. On the premise of collecting data, we should consider as much as possible the relevant data of the accident. The data used in this paper include gross national product (GDP) (100 million yuan), per capita GDP (yuan), gross national income (RMB 100 million), road mileage (10,000 kilometers), highway mileage (10,000 kilometers), number of civilian vehicles (10,000 vehicles), number of drivers (10,000 people), passenger traffic (10,000 people), road passenger traffic (10,000 people), total population at the end of the year (10,000 people), male population (10,000 people), female population (10,000 people), urban population (10,000 people), rural population (10,000 people), and the total number of deaths from traffic accidents per year (person). The data used are from the 1997–2016 data of the National Bureau of Statistics of China. The data are shown in Table 1.

3.2. Prediction Index for Road Safety Level. The measures of traffic safety level generally include the number of accidents, deaths, injuries, and property losses. To ensure the accuracy of the data, indicators such as the number of accidents, the number of injured people, and economic losses are subject to subjective influence and the accuracy is difficult to judge. The statistics on the number of deaths are true and reliable, difficult to falsify, and comparable. Therefore, this article uses the number of deaths as a predictor of traffic safety levels to predict the number of deaths.

3.3. Variable Correlation Analysis. If the information in the data is uncorrelated or noisy, the quality of the predictions may be affected [17]. In this paper, by comparing the chi-square value and the Pearson correlation coefficient to

Table 1: Sample of road traffic safety raw data for 1997–2016.

Year	GDP	Gross national income	Per capita GDP	Road mileage	Highway mileage	Number of civilian vehicles	Number of drivers	Passenger traffic	Highway passenger traffic	Total population	Male population	Female population	Urban population	Rural population	Traffic accident death toll
1997	79715	78802.9	6481	122.64	0.48	1219.09	2619.25	1326094	1204583	123626	63131	60495	39449	84177	73861
1998	85195	83817.6	6860	127.85	0.87	1319.30	2974.06	1378717	1257332	124761	63940	60821	41608	83153	78067
1999	90564	89366.5	7229	135.17	1.16	1452.94	3361.12	1394413	1269004	125786	64692	61094	43748	82038	83529
2000	100280.1	99066.1	7942	167.98	1.63	1608.91	3746.51	1478573	1347392	126743	65437	61306	45906	80837	93853
2001	110863.1	109276.2	8717	169.8	1.94	1802.04	4462.68	1534122	1402798	127627	65672	61955	48064	79563	105930
2002	121717.4	120480.4	9506	176.52	2.51	2053.17	4827.08	1608150	1475257	128453	66115	62338	50212	78241	109381
2003	137422	136576.3	10666	180.98	2.97	2382.93	5368.07	1587497	1464335	129227	66556	62671	52376	76851	104372
2004	161840.2	161415.4	12487	187.07	3.43	2693.71	7101.64	1767453	1624526	129988	66976	63012	54283	75705	107077
2005	187318.9	185998.9	14368	334.52	4.1	3159.66	8017.76	1847018	1697381	130756	67375	63381	56212	74544	98738
2006	219438.5	219028.5	16738	345.7	4.53	3697.35	9317.24	2024157.64	1860487	131448	67728	63720	58288	73160	89455
2007	270232.3	270844	20505	358.37	5.39	4358.36	10567.15	2227761.21	2050680	132129	68048	64081	60633	71496	81649
2008	319515.5	321500.5	24121	373.02	6.03	5099.61	12276.8	2867891.96	2682114	132802	68357	64445	62403	70399	73484
2009	349081.4	348498.5	26222	386.08	6.51	6280.61	13740.73	2976897.83	2779081	133450	68647	64803	64512	68938	67759
2010	413030.3	411265.2	30876	400.82	7.41	7801.83	15129.89	3269508.17	3052738	134091	68748	65343	66978	67113	65225
2011	489300.6	484753.2	36403	410.64	8.49	9356.32	17416.76	3526318.73	3286220	134735	69068	65667	69079	65656	62387
2012	540367.4	539116.5	40007	423.75	9.62	10933.09	20028.52	3804034.9	3557010	135404	69395	66009	71182	64222	59997
2013	595244.4	590422.4	43852	435.62	10.44	12670.14	21742.7	2122991.55	1853463	136072	69728	66344	73111	62961	58539
2014	643974	644791.1	47203	446.39	11.19	14598.11	24812.07	2032217	1736270	136782	70079	66703	74916	61866	58523
2015	689052	686449.6	50251	457.73	12.35	16284.45	28012.99	1943271	1619097	137462	70414	67048	77116	60346	58022
2016	743585	740598	53935	469.63	13.1	18574.54	30328.77	1900194	1542758	138271	10825	67456	79298	58973	63093

filter the features, the prediction results can be optimized. The formula for calculating the chi-square value is as follows:

$$\chi_d^2(V_k) = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(f_{ij} - F_{ij}\right)^2}{F_{ij}}, \tag{8}$$

where k is the variable of the kth group, r is the variable number, c is the target variable number, d is the degree of freedom = $(r - 1) * (c - 1)$, f_{ij} is the observation frequency of the variable V_k, and F_{ij} is the expected frequency of the variable V_k.

The Pearson correlation coefficient is calculated as follows:

$$R = \frac{\sum(X - \overline{X})(Y - \overline{Y})}{\sqrt{\sum(X - \overline{X})^2(Y - \overline{Y})^2},} \tag{9}$$

where R is the correlation coefficient, X is the independent variable, Y is the dependent variable, $\overline{X}$ is the mean of the independent variable, and $\overline{Y}$ is the mean of the dependent variable.

The chi-square test can calculate the degree of deviation between samples, and the greater the score of the chi-square test, the more obvious the association exists. The Pearson correlation coefficient can roughly give the degree of correlation between variables, and the absolute value of the Pearson coefficient indicates the degree of correlation. According to the chi-square score and the Pearson coefficient in Table 2, we removed the variable with the smallest chi-square score (highway mileage) and removed the variable with the smallest Pearson coefficient (road passenger traffic). Finally, 12 related independent variables and death toll were used as input variables, a total of 13.

3.4. Model Performance Evaluation Index. In this paper, error rate (E) and root mean square error (RMSE) were used to compare the predicted deviation degree, and root mean square logarithmic error (RMSLE) and decision coefficient (R-square) were used to measure the fitting capacity of the model.

The error rate and root mean square error formula are as follows:

$$\begin{aligned} E &= \frac{Y_0 - Y_p}{Y_0}, \\ \text{RMSE} &= \sqrt{\frac{\sum_{i=1}^{n}\left(Y_0 - Y_p\right)^2}{n}}. \end{aligned} \tag{10}$$

The formula for the logarithmic error and the coefficient of determination of the root mean square is as follows:

$$\begin{aligned} \text{RMSLE} &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\log\left(Y_0 + 1\right) - \log\left(Y_p + 1\right)\right)^2}, \\ R^2 &= 1 - \frac{\sum_{i=1}^{n}\left(Y_0 - Y_p\right)^2}{\sum_{i=1}^{n}\left(Y_0 - Y_{\text{mean}}\right)^2}, \end{aligned} \tag{11}$$

where n is the number of samples, Y_0 is the original value, Y_p is the predicted value, and Y_{mean} is the sample mean.

4. LSTM-GBRT Modelling Methodology

The LSTM neural network is capable of capturing time-dependent information and has an excellent effect on time series prediction, but it is insufficient in predicting inflection point data. The GBRT model is a typical representative of the ensemble learning algorithm, and the model is robust. In this paper, the LSTM-GBRT model is proposed by combining the two methods. The LSTM neural network is used to extract the features with time-dependent information. The features are trained by the GBRT model to predict traffic accidents. The structure of the LSTM-GBRT model is shown in Figure 3.

4.1. Normalization. The raw data are processed using min-max normalization to eliminate dimensional differences. A linear transformation of the original data causes the result to fall into the [0, 1] interval, and the conversion formula is as follows:

$$X = \frac{x - \min}{\max - \min}, \tag{12}$$

where max represents the maximum value of the feature in the sample data and min represents the minimum value of the feature in the sample data; x represents raw data, and X represents normalized data.

4.2. LSTM Layer Hidden Unit Number. There is no clear theoretical guidance for determining the number of nodes in the hidden layer. In general, use the following formula to select the number of nodes:

$$N = \sqrt{n + m} + a, \tag{13}$$

where N is the number of hidden nodes; n is the number of input nodes; m is the number of output nodes; and a can take a constant of 1 to 10.

In this paper, there are 13 input nodes and 1 output node. According to formula (7), the number of hidden nodes is 5~13. Try a different number of hidden layer nodes using 1 layer of LSTM and judge the degree of deviation according to the error rate and root mean square error, so as to select the number of hidden layer nodes.

The experimental results of the test set show that the LSTM model using 11 hidden nodes has the smallest RMSE value and the best prediction effect. The detailed error rate and root mean square error results of the test set are shown in Table 3.

4.3. LSTM Layer Depth. Since there are only 19 records in this example, the model depth is too high, which will cause the data to be overfitting. The experiment uses

Table 2: Chi-square value and Pearson coefficient value.

Variable	Chi-square value	Pearson coefficient value
GDP	3050901	−0.799
Gross national income	3047539	−0.799
Per capita GDP	212068.8	−0.801
Highway mileage	1030.5	−0.728
Highway mileage of high-speed roads	53.8	−0.749
Number of civilian vehicles	90673.9	−0.764
Number of drivers	119582.7	−0.766
Passenger traffic	5040066	−0.555
Road passenger traffic	5131653	−0.498
Total population at the end of the year	2764.561	−0.649
Male population	1353.25	−0.598
Female population	1433.637	−0.696
Urban population	48875.14	−0.693
Rural population	16934.48	+0.716

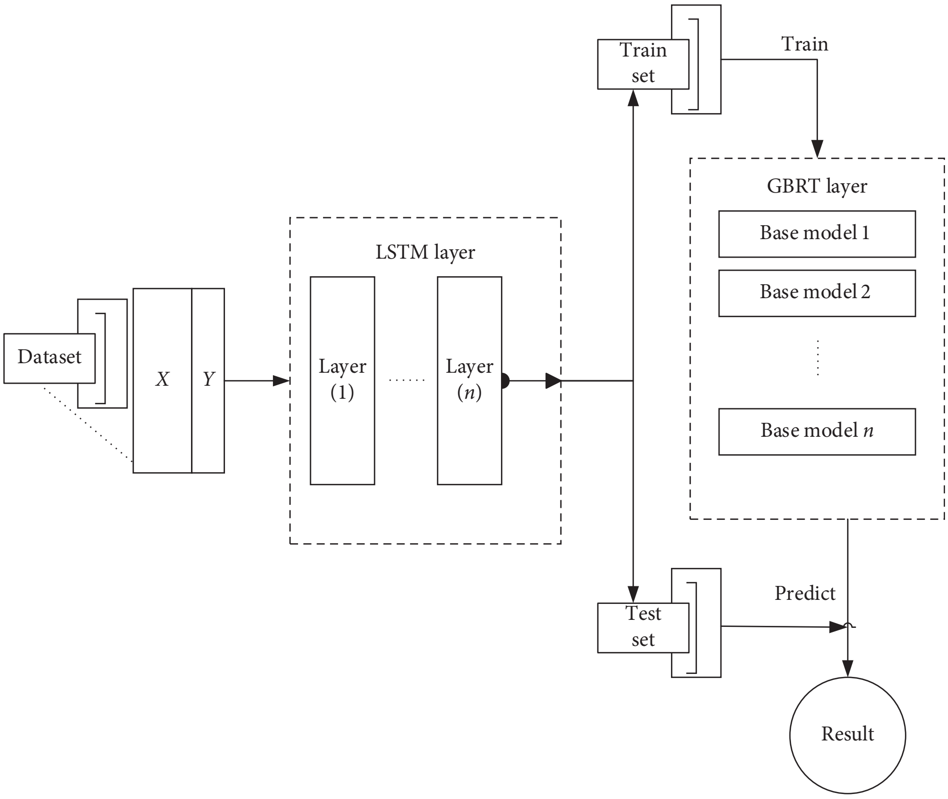

Figure 3: The structure of the LSTM-GBRT model.

1~5 layer models for comparison, and 11 hidden nodes are used for each layer. After training, the model performance is judged by calculating the root mean square logarithmic error and the decision coefficient of all records. The fitting results are shown in Table 4.

The smaller the RMSLE model, the better the fitting effect. The closer the R-square is to 1, the stronger the ability of the variable to interpret y and the model fits the data better. According to the results in Table 4, the 2-layer LSTM model has the best fitting ability.

4.4. GBRT Layer Regularization. The regularization formula is as follows:

$$L(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k), \tag{14}$$

where $\Omega(f) = \Upsilon T + (1/2)\lambda\|\omega\|^2$. Here, l is a differentiable convex loss function that measures the difference between the prediction $\hat{y}_i$ and the target y_i; the second term Ω penalizes the complexity of the model (i.e., the regression tree functions). The additional regularization term helps to smooth the final learnt weights to avoid overfitting.

Table 3: Death prediction for different hidden nodes.

Hide_size	E (2013)	E (2014)	E (2015)	E (2015)	RMSE
5	−0.0053	0.0085	−0.0001	−0.0963	3052.086
6	−0.0120	0.0111	−0.0077	−0.1138	3629.307
7	−0.0004	0.0059	−0.0054	−0.1021	3228.962
8	−0.0035	0.0141	0.0041	−0.0941	3002.066
9	0.0067	0.0189	0.0187	−0.0743	2476.743
10	0.0084	−0.0028	−0.0100	−0.1025	3256.069
11	0.0122	0.0031	0.0073	−0.0742	2378.019
12	0.0094	−0.0092	−0.0169	−0.1093	3505.351
13	0.0055	−0.0016	−0.0040	−0.1008	3186.757

Table 4: The RMSLE and R-square of different layer models.

Layer number	1	2	3	4	5
RMSLE	0.0333	0.0267	0.0301	0.0314	0.0356
R-square	0.9747	0.9843	0.9804	0.9822	0.9810

4.5. Hyperparameters of LSTM-GBRT Model. The hyperparameters of the LSTM layer include the number of network layers, the number of hidden cells in the layer, the learning rate, and the optimizer type, and the parameter settings are shown in Table 5.

The hyperparameters of the GBRT layer include the learning rate, the number of estimators, the maximum depth of the tree, the number of split nodes in the sample, the minimum sample required for the leaf nodes, and the loss function. This paper uses GridResearchCV to automatically find the optimal superparameters. The final parameter settings are shown in Table 6.

5. Comparative Analysis of Experiments

5.1. Experimental Environment. The experimental environment of this example is TOSHIBA satellite S40-A laptop, CPU: Intel(R) Core(TM) i3-3217U CPU at 1.80 GHz, running memory is 10 G, operating system is Windows 10 Enterprise Edition 2016 long-term service version, development environment. To use the PyCharm integrated development tool of Python 3.5 language, use the neural network model such as LSTM provided by Keras and use the GBRT model provided by skit-learn.

5.2. Experimental Design and Analysis of Results. Experiments include traditional regression models, neural network models, and integration model types of experiments. The experimental items are multivariate nonlinear regression (MUL), BP neural network model (BP), long- and short-term memory neural network model (LSTM), gradient boosted regression trees model (GBRT), and LSTM-GBRT model. The 15 data from 1998 to 2012 were used as training sets, and the four data from 2013 to 2016 were used as test sets. Use the data from the previous year as an input sample to predict the number of traffic accident deaths in the coming year. Figure 4 is a trend chart of actual traffic accident deaths from 1998 to 2016.

Table 5: Hyperparameters value of the LSTM layer.

Hyperparameters	Values
Learning rates	0.02
LSTM layer depth	2
LSTM layer hidden unit number	11
Optimizer	Adam

Table 6: Hyperparameters value of the GBRT layer.

Hyperparameters	Values
Learning rates	0.2
The depth of the tree	4
min_samples_split	3
min_samples_leaf	1
Number of estimators	140

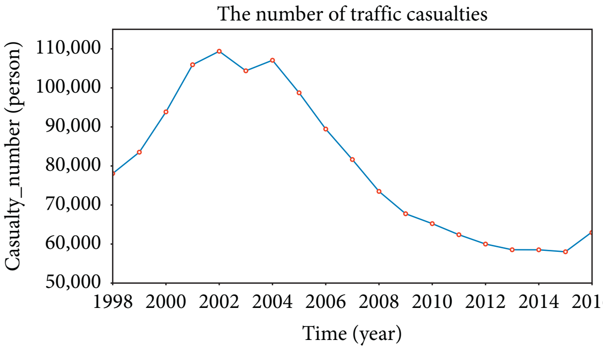

Figure 4: Trend chart of actual traffic accident deaths in China from 1998 to 2016.

After experimental training, the prediction results of each model in the test set are shown in Table 7.

The prediction results in the test set show that the BP neural and MUL regression models have no obvious regularity, and the prediction accuracy is not high.

The accuracy of LSTM in 2013, 2014, and 2015 was extremely high, and the deviation in 2016 suddenly increased. The trend of the actual number of deaths in Figure 4 is analyzed. 2016 is the year of the inflection point in the time period, and the trend of the first three years is consistent with the trend of the training data set, indicating that LSTM has an excellent prediction effect on the same trend. Conversely, when the forecast is the inflection point of the trend, the performance will suddenly drop. It also proves that the LSTM model can indeed learn the time-dependent information in the data.

The prediction results of the GBRT model and the LSTM-GBRT model did not fluctuate particularly among large samples, and the overall prediction effect remained stable. Many of the predicted values of the GBRT model and the LSTM-GBRT model are the same. We analyzed that the base model of the GBRT model is a regression tree, and the data fluctuations in the test set in 2013–2016 are small. In addition, the result for 2015–2016 moves away from the real data because the LSTM layer included in the LSTM-GBRT model learned time-dependent information, resulting in poor prediction of the trend inflection point.

Table 7: Test set prediction results.

Year	True value	P (LSTM)	P (MUL)	P (BP)	P (GBRT)	P (LSTM-GBRT)
2013	58539	59000.99	56518.18	60185.61	59998.38	58553.24
2014	58523	58738.28	53302.67	60485.92	61364.23	58553.24
2015	58022	58209.44	52670.03	60221.46	61364.23	58553.24
2016	63093	57777.69	51228.89	59990.11	61364.23	58553.24

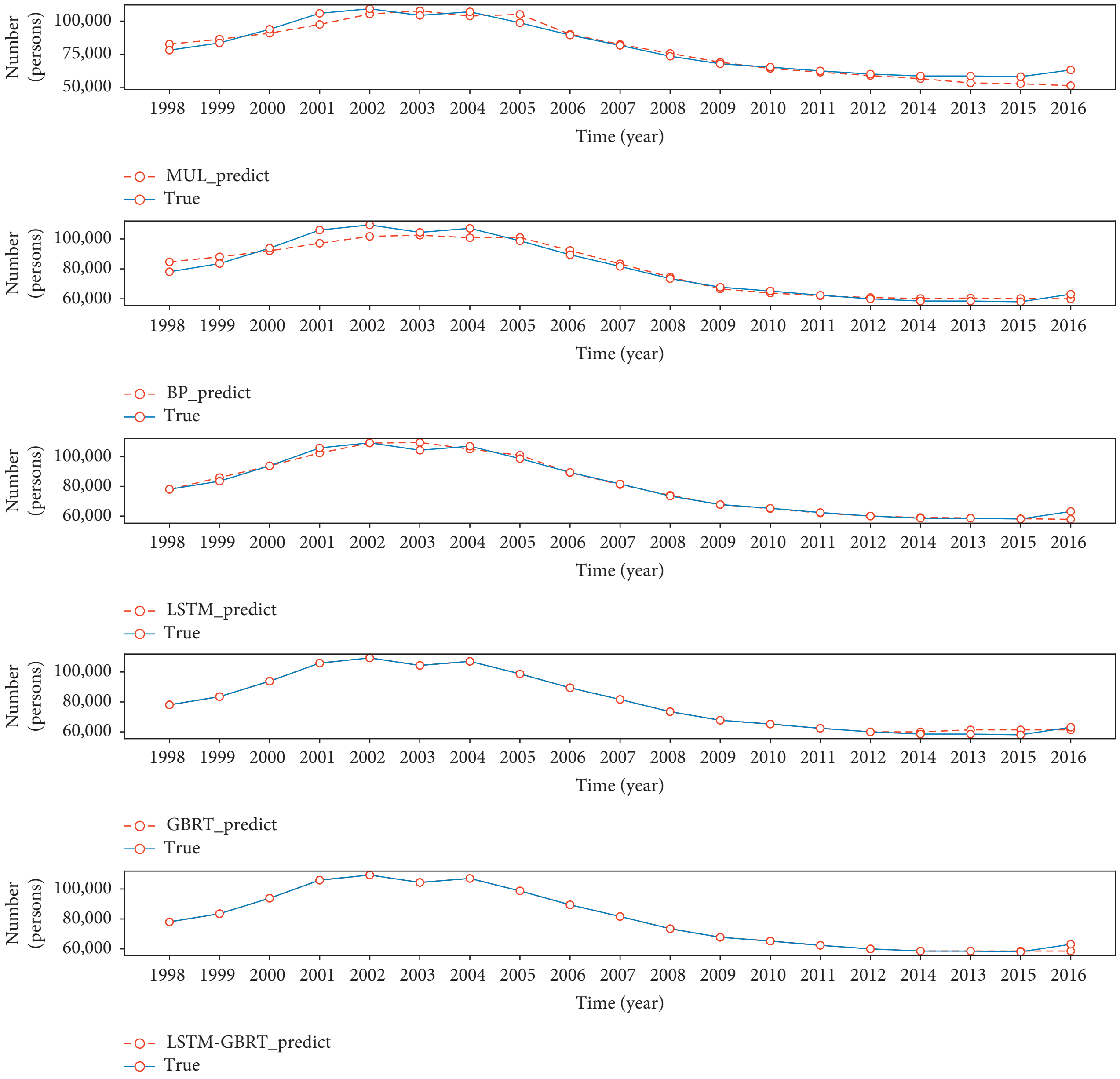

Figure 5: Fitting prediction results for each model.

Figure 5 shows the actual death toll in 1998–2016 and the fitted prediction results for each model.

After observing the prediction results of each model, we evaluate the effect of fitting all the data of each model. In addition to the performance indicators mentioned in Section 3.4, we add the training time-consuming indicators of the model for comparison. The performance indicators are shown in Table 8.

The performance indicators of different models in Table 8 show that the LSTM-GBRT model has the smallest RMSLE value, the best model fitting effect, the R-square value is closest to 1, and the variable has the strongest interpretation ability for the predicted value, but the training time is the longest. GBRT model training time is the shortest. The prediction performance of the GBRT model is not bad but slightly lower than the LSTM-GBRT

Table 8: Performance indicators of each model.

Indicators	MUL	BP	LSTM	GBRT	LSTM-GBRT
RMSLE	0.0665	0.0431	0.0267	0.0189	0.0172
R-square	0.9372	0. 9543	0.9843	0.9961	0.9967
Train time (s)	0.1851	5.4276	6.6909	0.1296	7.4275

model. The performance of the LSTM neural nework is lower than the GBRT model, and the performance of the MUL regression model and the BP neural network model is poor.

In terms of training time, the MUL regression model and the GBRT model have the shortest training time because their essence is a linear combination of mathematical data; the LSTM-GBRT model has the slowest training time, and LSTM model training time is very close to BP neural network training time. The training time of the neural network is obviously higher than that of the former because the training of the neural network model needs to construct a complex network structure.

5.3. Robustness Analysis. The occurrence of a traffic accident is influenced by many factors. The predictive model can predict the complex and variable conditions more stably, which indicates that the model has better robustness.

When analyzing the robustness of the model in this experiment, two aspects should be considered: first, internal factors, whether there are abnormal fluctuations in the model training data; second, external factors and policies proposed at the social level have promoted or inhibited the predicted data. Regardless of internal factors and external factors, the core is in the data. For model training data, the role of external factors is still indirectly affecting the data required for training, and then the effect of prediction is reflected. The model which is difficult to control is the external factor.

In this case, the model uses annual periodic data and policy factors have a short period of action and can cause less data fluctuation, so the robustness is better. When the model uses more sophisticated data, the influence of data fluctuation will increase. Firstly, anomaly data should be analyzed visually, the uniformity of each variable should be observed, and the uneven data should be processed, such as log function. Secondly, the abnormal variables are divided into two or more types of processing strategies. After the correlation analysis is completed, two or more models are established to train and predict. Finally, the prediction results of the multiple models are accumulated. Model training for specific data classification can also improve the accuracy of prediction, thus enhancing the robustness of the model.

6. Conclusion

The prediction of traffic accidents is of great significance. The future traffic accident trend forecasting work can help the traffic management department to grasp the trend dynamics in time, discover the rules of traffic accidents, formulate laws and regulations according to the rules, make scientific decisions, and construct the traffic system reasonably.

This paper proposes a road traffic accident prediction model based on the LSTM-GBRT model. Compared with the traditional regression model, the traditional BP neural network model, the LSTM neural network model, and the GBRT model, the experimental results show that the LSTM-GBRT model has the strongest ability to fit the data and the variable has the best interpretability for the predicted value. The model has a good predictive ability for the trend of road traffic safety level and can provide more accurate forecast data for the traffic management department, so that the traffic management department can better grasp the situation of traffic safety levels.

The model proposed in this paper also has defects. (1) Data collection, model training lacks relevant data on environmental factors. Due to the large randomness of road traffic accidents, its occurrence is affected by many factors. The environmental data belong to spatiotemporal data, which are difficult to collect, and the data of annual accident traffic accidents are not easy to quantify, so the weather environment factors are lacking in the model training data. (2) In terms of inflection point prediction, since the inflection point of the trend is unlikely to be discovered by the model in advance, the forecasting ability of the inflection point of the possible future trend is poor.

This paper takes China's annual traffic accident data as the research object, the proposed prediction task of the model is relatively macroscopic, and the predictability of microdata needs further experiment. This paper considers adding more relevant features, but in the macrodata forecasting work, related features are difficult to obtain or difficult to quantify. In the future, when taking microtraffic accident data as the research object, consider adding more features.

Data Availability

The raw data we used were official open data published by the UK Department of Transportation, and our experimental data were filtered from raw data online available at https://data.gov.uk/dataset/road-accidents-safety-data. The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was funded by the Xinjiang Uygur Autonomous Region Natural Science Fund Project "Research on Highway

VANET Early Warning Information Broadcast Transmission Mechanism” (2017D01C042).

References

[1] United Nation, *Transforming Our World: The 2030 Agenda for Sustainable Development*, https://sustainabledevelopment.un.org/post2015/transformingourworld, 2015.

[2] F. La Torre, M. Meocci, L. Domenichini, V. Branzi, N. Tanzi, and A. Paliotto, “Development of an accident prediction model for Italian freeways,” *Accident Analysis & Prevention*, vol. 124, pp. 1–11, 2019.

[3] M. Taamneh, S. Alkheder, and S. Taamneh, “Data-mining techniques for traffic accident modeling and prediction in the United Arab Emirates,” *Journal of Transportation Safety & Security*, vol. 9, no. 2, pp. 146–166, 2017.

[4] F. Zong, H. G. Xu, and H. Y. Zhang, “Prediction for traffic accident severity: comparing the Bayesian network and regression models,” *Mathematical Problems in Engineering*, vol. 2013, Article ID 475194, 9 pages, 2013.

[5] J. L. Deng, “Control problems of grey systems,” *Systems & Control Letters*, vol. 1, no. 5, pp. 288–294, 1982.

[6] G. Yannis, C. Antoniou, and E. Papadimitriou, “Autoregressive nonlinear time-series modeling of traffic fatalities in Europe,” *European Transport Research Review*, vol. 3, no. 3, pp. 113–127, 2011.

[7] S. Kumar and D. Toshniwal, “A novel framework to analyze road accident time series data,” *Journal of Big Data*, vol. 3, no. 1, p. 8, 2016.

[8] C. C. Ihueze and U. O. Onwurah, “Road traffic accidents prediction modelling: an analysis of Anambra State, Nigeria,” *Accident Analysis & Prevention*, vol. 112, pp. 21–29, 2018.

[9] Y. Shi, Y. Lin, Y. Zou, L. Jing, and L. Wu, “The prediction model on Chinese traffic deaths based on the grey topology,” *Mathematics in Practice and Theory*, vol. 43, no. 20, pp. 110–116, 2013.

[10] R. S. Hosse, U. Becker, and H. Manz, “Grey systems theory time series prediction applied to road traffic safety in Germany,” *IFAC-PapersOnLine*, vol. 49, no. 3, pp. 231–236, 2016.

[11] S. B. Liu and C. W. Wu, “Road traffic accident forecast based on optimized grey Verhulst model,” in *Proceedings of the 2016 Joint International Information Technology, Mechanical and Electronic Engineering*, Xi’an, China, October 2016.

[12] L. Zhao, X. U. Hongke, and H. Cheng, “Road traffic accidents prediction based on optimal weighted combined model,” *Computer Engineering & Applications*, 2013.

[13] M. He and X. C. Guo, “The application of BP neural network principal component analysis in the forecasting the road traffic accident,” in *Proceedings of the Second International Conference on Intelligent Computation Technology and Automation*, Zhangjiajie, China, October 2009.

[14] H. U. Liwei, T. Zhang, F. Guo, and Z. Chen, “Traffic accident split rate of vehicle types prediction and prevention strategies study based on gray BP neural network,” *Journal of Wuhan University of Technology*, 2018.

[15] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[16] J. H. Friedman, “Stochastic gradient boosting,” *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367–378, 2002.

[17] H. Wang, A. Parrish, R. K. Smith, and S. Vrbsky, “Variable selection and ranking for analyzing automobile traffic accident data,” in *Proceedings of the 2005 ACM Symposium on Applied Computing*, Santa Fe, NM, USA, December 2005.