

A Review of Road Traffic Accident Prediction Methods

Wang Shunshun, Yan Changshun*, Shao Yong

Faculty of information Technology, Beijing University of Technology, Beijing, China

Email address:

sswang@emails.bjut.edu.cn (Wang Shunshun), yuewuxing@bjut.edu.cn (Yan Changshun), shaoyong@bjut.edu.cn (Shao Yong)

*Corresponding author

To cite this article:

Wang Shunshun, Yan Changshun, Shao Yong. A Review of Road Traffic Accident Prediction Methods. *American Journal of Management Science and Engineering*. Vol. 8, No. 3, 2023, pp. 73-77. doi: 10.11648/j.ajmse.20230803.12

Received: April 20, 2023; **Accepted:** May 23, 2023; **Published:** May 29, 2023

Abstract: With the continuous development of urban traffic and the acceleration of urbanization, traffic accidents have become an important issue for urban safety and social stability. In order to prevent and reduce the occurrence of traffic accidents, traffic accident prediction technology has gradually become a hot spot for research. This paper analyzes road traffic accident prediction techniques from articles included in relevant English journals and provides a detailed introduction to the road traffic accident prediction techniques that are already in existence. This paper introduces the current status of research on traffic accident prediction techniques, including traditional statistical analysis methods, machine learning methods, neural network methods, time series analysis methods and techniques based on spatio-temporal data mining, and analyses the advantages and disadvantages of each road traffic accident prediction method. These methods are able to analyse the influencing factors of traffic accidents, build prediction models, improve prediction accuracy and provide strong support for road traffic accident prevention effects for urban traffic safety. Finally, the main difficulties faced in road traffic accident prediction and the future development trend of road traffic accident prediction is discussed. The work done in this paper can provide necessary theoretical support for relevant researchers and save the time needed for literature review.

Keywords: Road Traffic Accidents, Prediction Models, Time Series Analysis

1. Introduction

According to statistics, millions of people are killed and millions of people are injured in traffic accidents worldwide every year. Especially in developing countries, the incidence of traffic accidents is even higher. Against this background, predicting and preventing traffic accidents has become a hot topic of research. Traffic accidents are a complex problem that is influenced by many factors, such as road conditions, vehicle type, weather conditions, driver behaviour, etc [1]. Traditional traffic accident research methods usually use statistical analysis, but this method has limited prediction accuracy and lacks a comprehensive analysis of multiple factors. With the continuous development of computer technology, traditional statistical analysis methods, machine learning methods, neural network methods, time series analysis methods and techniques based on spatio-temporal data mining are widely used in the field of traffic accident prediction. This paper will introduce the current research

status of traffic accident prediction techniques, including traditional statistical analysis methods, machine learning methods, time series analysis methods, and techniques based on spatio-temporal data mining. By analyzing the advantages and disadvantages of these methods, the application scenarios and development trends of traffic accident prediction techniques can be better understood.

2. Traffic Accident Prediction Methods

Road traffic accident prediction can be divided into macro prediction and micro prediction according to the scope of prediction. Macro prediction of traffic accidents refers to the overall and trend prediction of traffic accidents with a long time (more than one year) or a large spatial area, such as the prediction of regional traffic accident change trends. Microscopic traffic accident prediction refers to the prediction of traffic accidents in a short period of time or at a certain location or road section, such as the prediction of traffic accidents in each month of a year, the prediction of

accidents at intersections, the prediction of accidents on a certain road section, etc. Qualitative forecasting is the use of human experience and judgement to synthesise relevant information and make a qualitative description of the development trend and characteristics of a traffic accident when not much data is available or a short time is needed to make a forecast. Quantitative forecasting is based on historical data and statistics, using mathematical or other analytical techniques to build models that can express quantitative relationships and use them to approximate the likely future performance of the predicted object. Commonly used quantitative forecasting techniques include statistical analysis, machine learning methods, neural network methods, time series analysis and techniques based on spatio-temporal data mining.

3. Traditional Statistical Analysis Methods

The traditional statistical analysis method is one of the most commonly used methods in traffic accident prediction and focuses on predicting future traffic accidents through the analysis of historical traffic accident data. This approach is based on the assumption that the occurrence of traffic accidents in the past has some similarity to the future situation. Therefore, by analyzing historical data, future traffic accident occurrences can be predicted.

The method based on quantitative theory means that the method uses quantitative theory as its core theoretical basis [2], i.e. traffic accident data is analyzed and processed through statistical methods. In traffic accident prediction, the use of quantitative theory allows factors such as traffic flow, speed and number of lanes to be transformed into specific data indicators for better statistical analysis and predictive modeling. Quantitative theory is widely used in practical applications, not only in the field of traffic safety, but also in many other fields such as finance, environment and healthcare [3].

The main role of grey models is to use limited historical data to make predictions [4] and to improve prediction accuracy. As traffic accidents occur for a variety of reasons, grey models can integrate various factors affecting the occurrence of traffic accidents from multiple dimensions, thus predicting future trends of traffic accidents more accurately. The process of building a grey model can be divided into: data pre-processing, determining the type of model, determining the set of factors, establishing grey differential equations, parameter estimation, etc. Among them, establishing grey differential equations is to use grey differential equations to describe the relationship between the set of factors and determine the structure and form of the model. The grey differential equation is the core of the grey model and is usually in the form of a first or second order differential equation.

The paper [5] proposes a traffic accident prediction method based on a grey prediction model. The method first

analyses historical traffic accident data, and then predicts future traffic accident occurrence trends by building a grey prediction model. At the same time, the researcher also considers factors affecting traffic accidents, such as weather and road conditions, and incorporates them into the model to improve prediction accuracy. The method has high feasibility and accuracy in practical application and can help traffic management departments to take timely measures to reduce the occurrence of traffic accidents.

The application of principal component analysis [6] in traffic accident prediction is mainly through dimensionality reduction processing and analysis of traffic accident data, extracting the main factors affecting the occurrence of traffic accidents and using them to construct a traffic accident incidence prediction model. Specifically, principal component analysis is the conversion of the original data into new combinations of variables that retain as much information as possible about the variance of the original data and satisfy the requirement of linear irrelevance. In traffic accident prediction, researchers can use principal component analysis to process traffic accident data to obtain some important variables (i.e. principal components) that indicate the occurrence of traffic accidents, and use these principal components as input indicators for traffic accident incidence prediction models.

4. Machine Learning Methods

Machine learning is a technique for discovering patterns and regularities between data by learning the data autonomously through computers. Machine learning-based traffic accident prediction techniques are mainly based on analyzing historical traffic data and using machine learning algorithms to build prediction models to predict the probability and impact of future traffic accidents [7]. For example, machine learning algorithms such as support vector machines, decision trees and logistic regression can be used to build traffic accident prediction models so that traffic accidents can be detected and prevented in advance.

Support vector regression (SVR) based road traffic accident prediction is a prediction model based on machine learning methods [8] that can be used to analyse a large amount of historical data to predict the number of road traffic accidents in a certain period of time in the future. Specifically, historical road traffic accident data is used as a training set and trained using the SVR model to obtain a non-linear function on the number of road traffic accidents in relation to other key parameters. This function is then applied to the prediction of future road traffic accident data to obtain more accurate predictions. The model is used to classify, segment and optimize each data point in the sample dataset, resulting in a support vector plane that is used as the basis for prediction of new data points. In predicting road traffic accidents, the SVR model learns and captures the non-linear features in the historical data set to improve prediction accuracy.

The Random Forest Regression (RFR) based road traffic

accident prediction method [9] predicts the number of road traffic accidents at a given time in the future by analyzing and mining historical traffic accident data. Specifically, the model uses a large amount of historical data as the training set, including factors such as traffic flow, vehicle speed, weather and time of day. A random forest model consisting of multiple decision trees is then constructed using the RFR algorithm. In this model, each decision tree captures the complex non-linear relationships in the historical data to better predict future traffic incidents. The RFR algorithm is a decision tree integration algorithm that randomly selects samples and features when generating a random forest to improve the accuracy and generalization of the prediction results.

5. Neural Network Methods

Neural networks are a technique for constructing complex non-linear mapping relationships by learning and training on data [10]. Neural network-based traffic accident prediction techniques are mainly based on analyzing historical traffic data [11] and using neural network algorithms to build prediction models to predict the probability and impact of future traffic accidents [12]. For example, neural network algorithms such as multilayer perceptron and recurrent neural networks can be used to build traffic accident prediction models so that traffic accidents can be detected and prevented in advance.

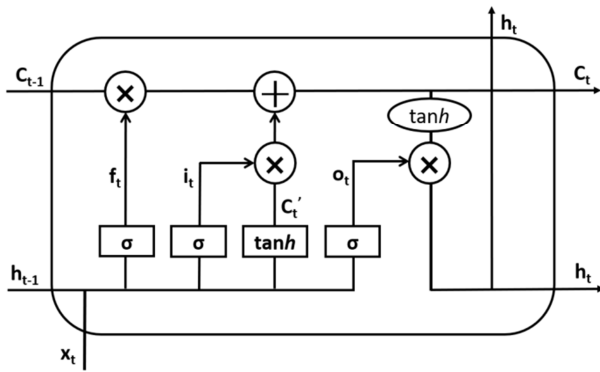


Figure 1. LSTM structure.

The LSTM (Long Short Term Memory) based road traffic accident prediction method [13] constructs an LSTM neural network model by learning and training historical traffic accident data, and uses the model to predict the occurrence of future traffic accidents [14]. The internal structure of the LSTM model is shown in Figure 1. The LSTM model can better handle sequential data, avoiding the gradient disappearance or explosion problems that occur in traditional neural networks when dealing with long sequential data. The LSTM model can better handle sequential data, avoiding problems such as gradient disappearance or explosion that occur in traditional neural networks when handling long sequential data [15]. In traffic accident prediction, the time series of historical traffic accident data can be used as input

data and the number of future traffic accidents as output data, and the LSTM neural network model can be used to learn and train in order to obtain a neural network model suitable for traffic accident prediction [16]. During the training process, the LSTM neural network automatically captures the trends and periodic changes in the number of traffic accidents and makes predictions based on them.

5.1. Forget Gate

The Forget Gate, as the name suggests, determines how much of the previous state I should forget based on the previous time step. For C_{t-1} , it first looks at the output h_{t-1} from the previous stage and the input x_t from this stage, and then uses a sigmoid function to determine how much of C_{t-1} to forget. A sigmoid value of 1 means more importance is given to retaining C_{t-1} , while a value of 0 means completely forgetting C_{t-1} . The formula is as follow:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

5.2. Input Gate

It first takes the output h_{t-1} from the previous stage and the input x_t from this stage, and then uses a sigmoid function to control how much to incorporate into the main storyline of C_t , which is the meaning of the first formula. Then, an alternative \tilde{C}_t is created, and tanh is used to control how much of \tilde{C}_t to include in C_t . Then, by multiplying the two parts together, the overall influence on C_t is determined, along with the impact of the previous Forget Gate, which can be written as follows:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (4)$$

5.3. Output Gate

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t * \tanh(C_t) \quad (6)$$

First, a sigmoid function is used to determine which part of C_t needs to be output, which is represented by the first formula o_t . Next, we put C_t into tanh to determine the part that will be output, and then multiply it by o_t to obtain the final output.

6. Time Series Approach

Time series analysis is a technique for predicting future trends in data change by analyzing and modeling time series data. Traffic accident prediction techniques based on time series analysis are mainly based on analyzing historical traffic data and using time series analysis algorithms to build predictive models to predict the probability and impact of future traffic accidents. For example, time series analysis algorithms such as ARIMA, ARCH and GARCH can be used

to build traffic accident prediction models so that traffic accidents can be detected and prevented in advance.

ARIMA models (differential autoregressive moving average models) can remove random fluctuations in the data while analyzing it, thus predicting future trends more accurately. In traffic accident forecasting, ARIMA models are widely used to predict future traffic accident rates [17]. Specifically, ARIMA models are applicable to single-variable time series data and can be used to predict the number and timing of future traffic accidents through the analysis and modeling of historical traffic accident data. In traffic accident forecasting, ARIMA models typically consist of three components [18]:

Differencing: In an ARIMA model, differencing refers to the first-order or multi-order differencing of time series data to remove trends and seasonality from the data and to make the data more consistent with smoothness requirements.

Autoregression (AR): Autoregression refers to the use of correlations in historical data to predict future values. In an ARIMA model, the autoregressive term represents the relationship between the current value and the value at the previous moment.

Moving average (MA): Moving average is the use of the mean value in the historical data to predict future values. In an ARIMA model, the moving average term represents the relationship between the current value and the noise error at the previous moment.

It should be noted that when using the ARIMA model for forecasting, constant model validation and optimisation is required to ensure the accuracy and reliability of the forecasting results.

7. Methods Based on Spatio-Temporal Data Mining

Spatio-temporal data mining is a technique for discovering relationships and patterns between data through analysis and mining of geospatial data. Traffic accident prediction technology based on spatial data mining is mainly through the analysis of historical traffic data and geographic information data, and the use of spatial data mining algorithms to establish prediction models to predict the probability of occurrence and impact of future traffic accidents.

Spatio-temporal graph convolutional neural network (STGCN) is a deep learning model based on graph convolutional operations, suitable for modeling and prediction against spatio-temporal data. STGCN currently has some applications in traffic flow prediction, and scholars are actively exploring more applications of STGCN in the field of transportation. In the field of road traffic accident prediction, STGCN can use the temporal and spatial information in historical traffic data to capture the spatio-temporal dependencies between traffic data through graph convolution operations and map them into fixed-length feature vectors. These feature vectors can be used to train

machine learning models, such as support vector machines, random forests, neural networks, etc., for road traffic accident prediction. Architecture of graph spatio-temporal network is shown in Figure 2. Compared to traditional models, STGCN has the following advantages:

Ability to accurately capture spatio-temporal dependencies: Using graph convolution operations, STGCN is able to accurately capture the spatio-temporal dependencies between traffic data, avoiding the problem of inability to handle spatio-temporal correlations in traditional models.

Better prediction accuracy: STGCN can extract rich feature information, effectively eliminating noise and redundant information, which can improve prediction accuracy.

Better model interpretability: STGCN can explain the causes and patterns of traffic accidents based on the spatial structure and time series in traffic data, improving the interpretability of prediction results.

Good adaptability: STGCN is able to handle spatial data at different scales and is applicable to different road networks and traffic flows, and has good adaptability and generalization ability.

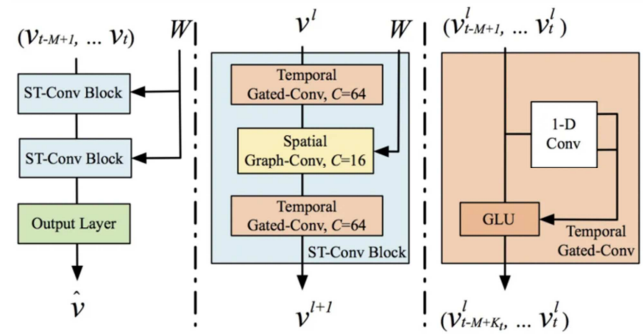


Figure 2. Architecture of graph spatio-temporal network.

8. Conclusion

This paper has systematically reviewed and summarized the current state of research in road traffic accident prediction, introduced the development process of road traffic accident prediction methods, as well as analyzed the advantages and disadvantages of various methods and the differences between them, illustrating what specific problems these methods have solved and in which aspects they have brought improvements. The article systematically introduces the methods and theories related to road traffic accident prediction from the perspectives of time series analysis, machine learning and neural networks, while also covering the details of practical operations such as data pre-processing and feature extraction. Currently, road traffic accident data prediction is more inclined to use techniques based on spatio-temporal graph networks. In the future, spatio-temporal graph neural networks can be used for more extensive applications in road traffic accidents, exploring more spatial information as well as more accurate attributes to predict road traffic accident indicators more accurately,

which has certain guiding significance for enhancing road safety.

References

- [1] Yannis, G., et al., Use of accident prediction models in road safety management--an international inquiry. *Transportation research procedia*, 2016. 14: p. 4257--4266.
- [2] Kelly, F., *The mathematics of traffic in networks. The Princeton companion to mathematics*, 2008. 1 (1): p. 862--870.
- [3] Keqiang, H., Y. Jibao and W. Sijing, Analysis of dynamic factors of debris landslide by means of the model of quantitative theory—using the Xintan landslide, China, as an example. *Environmental geology*, 2005. 48: p. 676--681.
- [4] Li, W., X. Zhao and S. Liu, Traffic accident prediction based on multivariable grey model. *Information*, 2020. 11 (4): p. 184.
- [5] Tongyuan, H. and W. Yue, Forecasting Model of Urban Traffic Accidents Based on Gray Model GM (1, 1). 2007, IEEE. p. 438--441.
- [6] Molla, M. M., M. L. Stone and E. Lee, Identification of road traffic fatal crashes leading factors using principal components analysis. *Journal of URISA, North Dakota State University, Fargo, ND (Under Review)*, 2015.
- [7] Pourroostaei Ardakani, S., et al., Road Car Accident Prediction Using a Machine-Learning-Enabled Data Analysis. *Sustainability*, 2023. 15 (7): p. 5939.
- [8] Yixuan, S., et al., Urban traffic accident time series prediction model based on combination of ARIMA and information granulation SVR. *Journal of Tsinghua University (science and technology)*, 2014. 54 (3): p. 348--353.
- [9] Yan, M. and Y. Shen, Traffic accident severity prediction based on random forest. *Sustainability*, 2022. 14 (3): p. 1729.
- [10] Wenqi, L., L. Dongyu and Y. Menghua, A model of traffic accident prediction based on convolutional neural network. 2017, IEEE. p. 198--202.
- [11] Shaik, M. E., M. M. Islam and Q. S. Hossain, A review on neural network techniques for the prediction of road traffic accident severity. *Asian Transport Studies*, 2021. 7: p. 100040.
- [12] Ogwueleka, F. N., et al., An artificial neural network model for road accident prediction: a case study of a developing country. *Acta Polytechnica Hungarica*, 2014. 11 (5): p. 177--197.
- [13] Zhang, Z., W. Yang and S. Wushour, Traffic accident prediction based on LSTM-GBRT model. *Journal of Control Science and Engineering*, 2020. 2020: p. 1--10.
- [14] Roh, Y. J. and S. H. Bae, Forecasting of traffic accident occurrence pattern using LSTM. *Journal of the Korea Institute of Intelligent Transportation Systems*, 2021. 20 (3): p. 59--73.
- [15] Staudemeyer, R. C. and E. R. Morris, Understanding LSTM--a tutorial into long short-term memory recurrent neural networks. *arXiv preprint arXiv: 1909.09586*, 2019.
- [16] Li, P., M. Abdel-Aty and J. Yuan, Real-time crash risk prediction on arterials based on LSTM-CNN. *Accident Analysis & Prevention*, 2020. 135: p. 105371.
- [17] Avuglah, R. K., K. A. Adu-Poku and E. Harris, Application of ARIMA models to road traffic accident cases in Ghana. *International journal of statistics and applications*, 2014. 4 (5): p. 233--239.
- [18] Gavirangaswamy, V. B., et al., Assessment of ARIMA-based prediction techniques for road-traffic volume. 2013. p. 246--251.