The Relationship Between Traffic Flow Forecasting and Traffic Accident Forecasting and the Possible Combination Points

Haoyu Wu

Department of Civil And Environmental Engineering, Hong Kong Polytechnic University, Hong Kong, 999077, China

Abstract. This paper mainly studies the relationship between traffic flow forecasting and traffic accident forecasting and the possible combination points. Firstly, through the introduction of the background, the necessity of combining the two is analyzed. Secondly, the differences between neural networks and traditional machine learning are compared in terms of model structure, feature extraction, applicability, computational resources and time, and model complexity. This paper emphasizes the importance of neural networks for traffic flow forecasting and traffic accident forecasting and introduces the commonly used neural networks. Then, the concept of temporal and spatial characteristics in traffic data is expounded, which opens the relationship between traffic flow forecasting and traffic accident forecasting and the analysis of possible combination points. Through the analysis of the existing research, it is found that the relationship between the two is mainly divided into data complementarity, common goal, and common method, and the possible combination points of the two are analyzed and prospected. Finally, the data and methods of improving traffic flow forecasting are summarized, which is conducive to better traffic flow forecasting and the combination of the two.

1 Introduction

With the development of big data and artificial intelligence technology, the acquisition and analysis of traffic flow and accident data have become more convenient and accurate. Through the analysis of a large number of historical traffic data, it can find the rules and patterns between traffic flow and accidents. The development of real-time data acquisition technology makes it possible to predict accidents with real-time traffic flow. In the case that the traditional traffic management methods are difficult to cope with the increasingly complex traffic conditions, the data-driven approach has become a research hotspot. Traffic flow forecasting and traffic accident forecasting have made remarkable progress in their respective fields, but it is difficult to fully deal with the actual traffic problems in a single research.

Corresponding author: 24101879g@connect.polyu.hk

Traffic flow forecasting and traffic accident forecasting are two key areas in Intelligent Transportation Systems (ITS). The two are closely related, the change in traffic flow directly affects the incidence of traffic accidents, and traffic accidents will in turn have a significant impact on traffic flow. The study of the combination of the two points can understand and predict the dynamic changes in the traffic system more comprehensively, and improve scientific and effective traffic management. Yang et al. used the urban expressway data and traffic flow data of Shanghai and found that the speed and flow had a significant impact on the occurrence of accidents [1]. The low speed and the difference in traffic between lanes will significantly increase the probability of accidents. Liu et al. collected historical data on traffic accidents and traffic flow on the G3001 expressway in Xi 'an to build a traffic accident forecasting model [2]. Adding the improved traffic flow stability coefficient based on the traffic flow parameter setting into the traffic accident forecasting model can not only reduce the calculation time of the model by 15.2% but also improve the accuracy and increase the reliability of the accident forecasting model. Lu et al. found that in the case of a combined analysis of traffic flow and traffic accidents in Singapore, the traffic flow in peak hours increased by about 20%, the average commuting time decreased by 15%, and the accident rate could be reduced by 25% [3].

The combination of traffic flow forecasting and traffic accident forecasting can improve the accuracy of traffic management and help identify potential high-risk areas in advance, so as to optimize traffic flow, reduce accidents, and improve road safety. Therefore, it has become an important research direction to explore the internal relationship between traffic flow forecasting and traffic accident forecasting and the possible combination of the two in the future. Using a literature review and summary, this paper studies the internal relationship between traffic flow forecasting and traffic accident forecasting as well as the possible combination points in the future, aiming to provide direction for the better combination of the two in the future and the optimization of the traffic system.

2 Spatiotemporal modeling and predictive techniques

2.1 Introduction of neural network and its role in traffic flow and traffic accident forecasting

2.1.1 Introduction to mainstream neural networks

Mainstream neural networks mainly include Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Transformer, Long Short-term Memory Networks (LSTM), and so on. CNN is primarily used to process data with a grid structure, such as images. The local features are extracted through convolution and pooling operations, and it has the characteristics of parameter sharing and migration invariance. Great success has been achieved in the tasks of image classification, target detection, and image segmentation. RNN is suitable for processing sequential data, such as text, time series, etc. It has loop joins to capture time dependencies in sequence data. The transformer is built on a self-attention mechanism and does not rely on circular joins or convolution operations. It can be calculated in parallel to speed up training. Excellent at handling long-distance dependencies, it is often used for tasks such as machine translation, language modeling, and text generation. LSTM is also a kind of neural network. Through its unique input gate, forget gate, and output gate mechanisms, it can effectively control the flow of information and capture important features in long time series, thereby avoiding gradient disappearance and explosion problems. LSTM is widely used for tasks such as machine translation, sentiment analysis, stock forecasting, weather forecasting, and traffic flow forecasting.

2.1.2 Differences between neural network and machine learning

The neural network plays an important role in traffic flow and traffic accident forecasting. Through multi-layer perceptron, CNN, RNN, and other models, it can process complex non-linear traffic data, realize short and long-term traffic flow forecasting, and evaluate traffic accident risk. They have the advantages of high nonlinear mapping capability and automatic feature extraction, which can adapt to different types and scales of data, significantly improve the accuracy and reliability of forecasting, and provide strong technical support for the development of intelligent transportation systems.

There are some differences between neural networks and traditional machine learning methods, and they each have different advantages. Mainly from the following five aspects to discuss the differences between the two. First, the difference in model structure. Traditional machine learning method usually has a clear formula or structure and are relatively simple. The neural network is composed of multiple nodes to form a multi-layer network. The structure is more complex and features can be learned automatically. Second, is the difference in feature extraction. Traditional machine learning relies on domain knowledge and experience. A neural network can automatically learn features from data, especially when processing complex data such as images and audio, deep neural network can automatically extract high-level features through multi-level structures, reducing the dependence on manual feature engineering. Third, differences in applicability and performance. Traditional machine learning methods are suitable for problems with relatively small amounts of data and low feature dimensions, with fast training speed and easy interpretation and understanding. A neural network is particularly suited to a mass of data, performing well on these tasks. Fourth, the difference in computing resources and time. Traditional machine learning method typically has lower computational resource requirements and shorter training times. Training neural networks, especially deep learning models, usually requires a lot of computational resources and time. Fifth, difference in model complexity. The traditional machine learning method model is relatively simple and easy to explain. For example, regression coefficient of linear regression, branch rules of decision tree, etc. However, the neural network model is highly complex, and the relationship between weights and nodes is difficult to explain intuitively.

Because of the differences between them, each has its own advantages. The advantage of neural networks is that they can automatically extract complex features from data and have a strong ability to deal with unstructured data. The modeling ability of complex nonlinear relationships is strong, and the forecasting accuracy is high. The advantage of traditional machine learning is that it is simple and efficient, suitable for small data sets and rapid development. The model is interpretable and easy to understand and trust. Therefore, neural networks are often used to predict traffic flow and traffic accidents.

2.2 Relationship and possible combination points of traffic flow forecasting and traffic accident forecasting

2.2.1 Temporal and spatial characteristics of traffic data

Spatiotemporal characteristics in traffic data refer to attributes or properties about time and space. These characteristics include changes in traffic flow over time, such as daily, weekly, and seasonal variations, as well as traffic density in different locations. By analyzing these temporal and spatial characteristics, it can better understand the operation of the traffic system, predict the occurrence and duration of traffic congestion, and then carry out more reasonable traffic planning.

In terms of how existing models model and capture spatiotemporal characteristics in traffic data, existing models usually use and combine a variety of techniques to model and capture spatiotemporal features in traffic data. These technologies mainly include (1) Time series analysis and forecasting: using time series models such as Autoregressive Integrated Moving Average (ARIMA), Prophet, etc., to analyze the characteristics of time changes in traffic data. These models are able to predict future traffic flows and congestion. He used the combined model of ARIMA and LSTM to predict short-term traffic flow at intersections, and the results showed that the ARIMA model could meet the forecasting of the linear part of traffic flow data [4]. (2) Spatial data analysis: spatial statistical methods and geographic information system technology are used to analyze the spatial characteristics of traffic density and flow distribution in different locations. (3) Deep learning: Zhu uses a combination model of residual network and graph attention network to extract spatial features in traffic flow in-depth, and uses bidirectional simple cycle units to extract temporal features [5]. (4) Spatiotemporal data mining: Such as trajectory analysis, spatiotemporal clustering, and association rule mining, are used to identify and understand spatiotemporal patterns and associations in traffic data

2.2.2 Relationship between traffic flow forecasting and traffic accident forecasting

Traffic flow forecasting is the forecasting of traffic conditions of a certain section of road or a region, usually based on historical traffic data, real-time monitoring data, weather information, and so on. The common methods of traffic flow forecasting include time series analysis, machine learning models, simulation models, and so on. Traffic accident forecasting is to predict the probability and location of traffic accidents that may occur in the future. Common methods include statistical analysis, machine learning, spatial analysis, etc. [6]. Traffic flow forecasting involves analyzing and predicting the volume of vehicle traffic on the road, helping to optimize traffic mobility and reduce congestion. Traffic accident forecasting, on the other hand, focuses on identifying risk factors and areas that may lead to accidents, as well as taking steps to reduce the incidence of accidents in advance.

Existing studies have extensively discussed the relationship between traffic flow characteristics and traffic safety. Zhong et al. explored the correlation between traffic flow characteristic parameters and traffic safety from multiple aspects through statistical analysis methods such as multivariate polynomial ratios [7]. The results show that there is a statistical relationship between traffic flow, saturation, speed difference of large vehicles the proportion of large vehicles, and the probability and severity of traffic accidents. Besides, the characteristic data of traffic flow and traffic accidents are collected and analyzed, and the generalized linear regression method is adopted to establish a traffic accident forecasting model for a four-lane expressway in China. The model quantitatively describes the impact of traffic characteristic parameters on highway safety from a microscopic point of view and reveals that the distribution of highway traffic accidents is mainly affected by dynamic traffic flow parameters at the mathematical level of Objective facts and their characteristic laws. Aiming at the problem of expressway traffic flow under traffic accidents, Le et al proposed a long short-term memory network model (Multi-Scale Dynamic Convolution Long Short-Term Memory Network) based on the maximum deviation similarity criterion to predict expressway traffic flow under traffic accidents [8]. Through simulation experiment analysis, the MDSC-LSTM neural network model is effective in predicting traffic flow under traffic accidents, and it is concluded that large traffic flow may cause congestion or even accidents, and when an accident occurs, it may have a certain impact on the traffic flow around the accident point.

The relationship between traffic flow forecasting and traffic accident forecasting mainly includes the following three aspects. First, is data complementarity. The concept of data complementarity refers to the fact that traffic flow data, such as vehicle flow, speed, density, and traffic accident data, such as the time, place, and severity of the accident, can complement and influence each other. The change in traffic flow may not only directly affect the probability of accidents and the severity of accidents, but also indirectly affect the accident risk by affecting the efficiency of road use. At the same time, the occurrence of accidents may, in turn, affect the dynamic change of traffic flow through traffic jams or road closures, and then produce a chain effect, affecting the traffic status of the surrounding roads. Through in-depth analysis of the complex interaction between these data, it can reveal the deeper causal relationship and influence mechanism in the traffic system, and provide theoretical support and data basis for formulating more effective traffic management strategies. Second, common goals. Both aim to improve road safety and efficiency. Traffic flow forecasting focuses on analyzing and predicting traffic volume and mobility, with the main goal of reducing traffic congestion and improving overall traffic management efficiency. By understanding future traffic flow trends, traffic management departments can take measures in advance to optimize traffic signals and allocate road resources, thereby improving road capacity and driving efficiency. Traffic accident forecasting, on the other hand, focuses on predicting the occurrence and severity of accidents, with the goal of reducing the incidence and impact of accidents. Through the analysis of historical accident data and environmental factors, high-risk areas and time periods can be identified, and preventive measures can be taken, such as improving road design, strengthening law enforcement, and enhancing driver safety awareness, in order to reduce the occurrence of traffic accidents and the resulting losses. Although the specific priorities of the two are different, they are all working in the same direction to build a safer and more efficient transportation system. Third, a common approach. CNN and bidirectional Long short-term memory networks can be used effectively in traffic flow forecasting and traffic accident forecasting. For example, in traffic flow forecasting, CNN can process image data captured by traffic cameras; In accident forecasting, CNN can process high-dimensional features such as road conditions and traffic signals. Bidirectional Long Short-Term Memory Networks are used to capture time dependencies in time series data. In traffic flow forecasting, BiLSTM can analyze the change in traffic flow with time. In accident forecasting, BiLSTM can deal with the time pattern of accident occurrence and other relevant time characteristics.

2.2.3 Possible combination points

The possible combination points of traffic flow forecasting and traffic accident forecasting can be considered from the following aspects. First, data. Combining traffic flow data, real-time traffic monitoring data, sensor data, and accident data, such as historical accident records and emergency incident reports, can provide a more comprehensive basis for analysis. Specifically, traffic flow characteristics such as peak hour flow, road congestion index, etc., can reflect road usage and congestion degree in different time periods. The accident characteristics, such as weather conditions and road conditions, can reveal the external environmental factors of the accident. For example, real-time traffic monitoring data and sensor data can provide information on vehicle flow, speed, and density to help us identify areas of traffic flow that are likely to be congested and at high risk. Historical accident records and emergency reports can provide information on the location and time period of frequent accidents, revealing which road sections are more likely to have accidents under certain conditions [9]. By combining these data, it can find the correlation between traffic flow and accident occurrence. Further, it can construct a traffic accident forecasting model based on multi-source data. Combining traffic flow characteristics with accident characteristics can improve the accuracy of the forecasting model. This not only helps to identify potential accidence-prone areas in advance and formulate more effective traffic management and control strategies, but also provides decision support for traffic management departments, optimizes resource allocation, improves road safety and traffic efficiency, and promotes the development of intelligent transportation systems. Second, method. CNNS and BiLSTM show great power when processing traffic data because they are able to handle both spatial and temporal features. Specifically, CNN is good at extracting spatial features from data and is able to identify patterns and structures in traffic flow. For example, CNN can extract spatial information such as traffic flow density and speed distribution from traffic surveillance videos or high-resolution traffic flow images. BiLSTM is good at handling time series data and is able to capture time dependencies and dynamic changes in the data. For example, BiLSTM can analyze the time series patterns in traffic flow data and identify the traffic variation rules during peak, off-peak and off-peak hours. The combination of CNN and BiLSTM is applied to the merged traffic data set, which can realize the simultaneous forecasting of traffic flow and accident probability. In terms of specific operation, the spatial features, such as traffic density and congestion of different sections, are extracted by CNN first. Then, BiLSTM is used to process the time series information of these spatial features to capture the changing trend and pattern of traffic flow over time. This combination of spatial and temporal features allows the model to more accurately predict changes in traffic flow and potential accident probabilities.

3 Data types and technologies to improve the accuracy of traffic flow forecasting

Researchers use a variety of data types and techniques to improve the accuracy of traffic flow forecasting. In terms of data, it mainly includes the following types. In terms of data, it is mainly divided into the following seven types of data (1) Historical traffic flow data. (2) Real-time traffic data: such as Global Positioning System (GPS) data. Remote sensing data, etc. (3) Raster data (4) External data: such as weather data, socio-economic data, road and infrastructure data, environmental data, etc. (5) Figure structure data. (6) Data generated by vehicles or sensors: such as taxi track data, and mobile phone signaling data. (7) Point of Information Distribution data: location information of restaurants, shops, entertainment venues, tourist attractions, etc. In terms of technology, it is mainly divided into the following seven technologies (1) Convolutional neural network technology. (2) Residual technique. The above two technologies can be used to construct a spatio-temporal information fusion module, extract spatio-temporal characteristics, and use it for the forecasting of subsequent traffic. (3) Fully connected networks and convolution: more efficient collection of meteorological data. (4) Convolutional Long Short-Term Memory (ConvLSTM): it is a model that combines CNN and LSTM, especially used to process spatiotemporal data [10]. (5) Visual converters: capturing potential long-distance spatial dependencies between urban areas. (6) Geographic Information System (GIS) technology, time series analysis. (7) Big data analysis technology: such as data mining technology.

4 Conclusion

This paper introduces some mainstream neural networks and their functions in traffic flow and traffic accident forecasting, then analyzes the relationship between these two types of forecasting, and prospects the possible combination points of them. The conclusion of this paper is that the relationship between the two types of forecasting mainly includes complementary data, common goals, and common methods. In terms of existing and possible future combinations, they can be combined in terms of data and methods. There are also some limitations in this study. For example, it is not considered that there is a certain difficulty in data combination, traffic flow data, and accident data usually come from different sources, data format and quality may be inconsistent, and the inconsistency of data may lead to the limited effect of the forecasting model. In addition, traffic accidents have great randomness and uncertainty. Although some patterns can be identified by statistical methods, many accidents are accidental and difficult to predict. Based on the findings of this paper, future research could consider developing efficient data processing and analysis techniques to acquire and process large amounts of data in real-time to support real-time forecasting. And strengthen the interdisciplinary cooperation in the fields of traffic engineering, computer science, data science, psychology, etc., comprehensively apply the knowledge and methods of various disciplines, and enhance the comprehensiveness and innovation of research. In general, this paper provides an indirect view of the possible intersection of traffic flow forecasting and traffic accident forecasting and provides a basis for future research and practice. Hopefully, this paper will be useful for the development of related fields and trigger more meaningful discussions.

References

1. K. Yang, R. Yu, X. Wang, Accident risk assessment based on traffic flow data from lane aggregation, J. Tongji Univ. (Nat. Sci. Ed.) **44**, 1567-1572 (2016).
2. X.-L. Liu, J. Shan, T.-Z. Liu, C. Rao, T. Liu, Real-time risk prediction of expressway traffic accident based on traffic flow stability coefficient, Traffic Inf. Saf. **4**, 71-81 (2022).
3. P. Lu, W. Bi, A comparative study on security governing of smart city: A perspective of taking Singapore and Shanghai for example, J. Sinol. **17**, 128-144 (2023).
4. K. He, Research on short-term intersections traffic flow prediction and application based on ARMI-LSTM combined model, Master's thesis, Xihua Univ. (2023).
5. Y. Zhu, Spatiotemporal modeling and prediction of traffic flow based on deep learning, Master's thesis, Nanjing Univ. of Posts and Telecommunications (2023).
6. F. Alhaek, W. Liang, T. M. Rajeh, M. H. Javed, T. Li, Learning spatial patterns and temporal dependencies for traffic accident severity prediction: A deep learning approach, Knowl.-Based Syst. **286**, 111406 (2024).
7. L. Zhong, Y. Chen, X. Zhang, R. Zhou, K. Wu, N. Zhao, B. Ren, Research on highway running situation prediction method based on multi-parameter regression of short-term traffic flow, Highway Eng. (2016).
8. B. Le, Y. G. Cai, H. Cai, J. G. Wang, Short-term traffic flow prediction under traffic accidents, J. Dongguan Univ. Technol. **5**, 44-49+54 (2020).
9. Z. Yang, W. Zhang, J. Feng, Predicting multiple types of traffic accident severity with explanations: A multi-task deep learning framework, Saf. Sci. **146**, 105522 (2022).
10. P. Cao, F. Dai, G. Liu, J. Yang, B. Huang, A survey of traffic prediction based on deep neural network: Data, methods and challenges, in M. R. Khosravi, Q. He, H. Dai (Eds.), Cloud Computing, Springer Int. Publ., 17-29 (2022).