Trend and Methods of IoT Sequential Data Outlier Detection

Yinuo Wang[1], Tao Shen[1](✉), Siying Qu[2], Youling Wang[2], and Xingsheng Guo[2]

[1] School of EE, University of Jinan, Jinan 250000, China
cse_st@ujn.edu.cn
[2] Jinan Lingsheng Info Tech. Co., Ltd., Huaiyin District, Jinan 250000, China

Abstract. In recent years, the state has made great efforts to develop the transportation industry. With the continuous expansion of the transportation network and the large-scale increase of vehicles, traffic congestion is serious, and traffic accidents occur frequently, which damages the normal traffic order. In order to ensure the overall operation of urban traffic is safer and more coordinated, it is of great practical value to detect abnormal traffic events in urban operation in real-time. Effective traffic incident detection may reduce traffic congestion brought on by traffic incidents, stop the incidence of follow-up accidents, and improve the safety of highway traffic. It has become a general trend to detect and warn about traffic accidents beforehand. This paper aims to build a machine-learning model to study the anomaly detection of traffic accidents. This study detected the number of traffic accidents in different time periods, and the traffic anomalies in 406 days every five minutes were analyzed. The frequent periods of accidents were statistically sorted out, which determined the basic direction for the prevention and detection of traffic accidents, helped to reduce traffic accidents, and improve people's travel experience.

Keywords: Time series · Machine learning · Internet of Things · Outlier detection

1 Introduction

Continuous economic growth in China has resulted in rising living standards for the populace, and the increasing travel demand of residents, the contradiction between urban traffic and urban development needs to be solved urgently. In recent years, China's population and car ownership have increased significantly. Urban traffic problems are becoming more and more prominent, such as road congestion, frequent accidents, etc., which make the problem of low travel efficiency worsen, seriously affecting the quality of residents' travel, making it difficult for residents' needs to be met, and people's happiness has not been greatly improved with the rapid economic development.

This work is supported by Shandong Key Technology R&D Program 2019JZZY021005 and Natural Science Foundation of Shandong ZR2020MF067.

B. Wang et al. (Eds.): ICMTEL 2023, LNICST 535, pp. 386–391, 2024.
https://doi.org/10.1007/978-3-031-50580-5_34

Carrying out scientific research on road traffic management, analyzing and studying road traffic safety and traffic situation, implementing scientific decision-making, using modern management science and modern scientific and technological means, and enhancing the capacity of traffic on roads coordination and control of the important duties of road traffic management and control departments. With the continuous improvement of the current social and economic level, residents' requirements for travel quality are gradually improved. However, urban traffic congestion and traffic safety caused by weather are still the primary causes that restrict the evolution of urban traffic.

Therefore, the abnormal detection of traffic accidents plays an important role in the development of intelligent transportation. Additionally, precise and real-time traffic flow detection can be used to evaluate road planning, lower the number of fatalities from traffic accidents, and offer sensible journey guidance [1]. Traffic accidents at important traffic intersections are detected in a short period to get the most accurate detection results, which can provide a timely and effective reference for traffic control of relevant departments and travel of residents, relieve traffic pressure at different times, make full use of road resources to the maximum extent, and reduce frequent accidents. Therefore, based on the data of traffic accidents per unit of time, this paper processes the data visually arranges and analyzes the abnormal data by using the XGB model [2, 3], obtains the frequency comparison of traffic accidents in different periods, and draws ROC curve to evaluate the prediction results [4].

2 Background

The term "Internet of Things in Automobile" or "Internet of Vehicles" refers to the special use of Internet of Things technology in the automobile and transportation industry [5]. With the advanced RFID sensing technology [6] and mobile communication technology, we can perceive roads and vehicles comprehensively and in real-time [7], and comprehensive intelligent traffic management through the use of big data and artificial intelligence, genuinely tying together cars, road infrastructure, and people to create a cooperative intelligent traffic system. The networking of vehicles can road resource allocation should be optimized, reduce congestion and improve overall traffic efficiency. There are a lot of archival traffic statistics, which are especially suitable for data mining and supervised learning. In [8], a Mahalanobis distance-based method that can handle differential data to predict traffic fluctuation, detect accidents, and utilize it to assist rectification and bolster accident information was enhanced by Sun B et al. The use of fused network conversation data sets as a training and testing strategy for the model is suggested in the literature [9]. The conditional combination of data sets may be used to enhance the performance of algorithms and the accuracy of their detection, according to experiments using real data. The characterization is likely to be applied in various sectors because it was accomplished by broad statistical computation. Regression trees and random forest models are used to forecast traffic flow in urban working areas, according to the literature [10], and when it is compared to the neural network and KNN prediction models [11], the findings suggest that random forest has a better level of prediction accuracy. In literature [12], the deep learning model was initially utilized for predicting traffic flow, and the multi-task learning method was combined to forecasting the volume of traffic on the roads, and the prediction accuracy was better than that of common models. Literature [13] by adopting a node construction scheme under a small-scale traffic network, the road can be effectively transformed into graph network structure data, and two different data sets can be formed by using public large-scale traffic network data. Experiments are carried out on different data sets with effective algorithms, and it is verified that the proposed method of constructing graphs with a graphical neural network improves the dynamic emergency mobility prediction. In the process of data transmission, abnormal data is also valuable [14, 15], because it is necessary to train the model and detect the abnormality when compared with the normal data that has overflowed. A data-driven approach was put out by [16] to address modeling and heteroscedasticity. This method has successfully improved the receiver operating characteristic (AUC) by about 30% [17]. Based on serial data from the Internet of Things, Dynamic Emergency Traffic Prediction; Medical emergency transportation calculates real human life in minutes. Therefore, it is very important to plan the emergency traffic route according to the current state of traffic flow. In the research literature [18], a dynamic graph structure is designed to predict the traffic flow level in conjunction with the graphical neural network (GNN) algorithm [19]. The experimental findings demonstrate that this innovative design offers great prediction accuracy and efficiency.

3 Methods and Materials

3.1 Data Description and Preprocessing

The data source for this study was a dataset of traffic conditions every five minutes from 22:00 March 31, 2013, to 22:00 May 11, 2014, with 116,928 data results read, as shown in Table 1. In this experiment, we counted the number of traffic accidents every five minutes during 406 days, from which 288 feature data points can be generated per day after data processing, and then binary conversion is performed for each feature value. The experiments were set up to judge a day as an abnormal road traffic accident if 10% or more of the data abnormal feature points were generated on that day.

Table 1. Data of partial traffic flow in October 2012

	COUNT_TIME	NUM_TOTAL
0	2013-03-31 22:00:04	10.0
1	2013-03-31 22:05:04	16.0
2	2013-03-31 22:10:04	13.0
3	2013-03-31 22:15:04	2.0
4	2013-03-31 22:20:04	10.0

3.2 Experiments and Results

This experiment uses the XGB model to obtain a feature importance ranking diagram, as shown in Fig. 1. From the analysis in Fig. 1, it can be concluded that traffic accidents occur most often in the early morning hours, which may be caused by driver fatigue from driving for long periods, leading to reduced driver reflexes and poor vision, or low visibility at night. The traffic accident rate is also high around 4:00 or 5:00 a.m. This may be due to drivers getting up early and being unable to concentrate, or foggy weather conditions. During the daytime, traffic accidents are relatively stable and of medium frequency, but the frequency of traffic accidents gradually increases at night. It is inferred that people should be encouraged to drive their vehicles as much as possible during the daytime, avoid driving at night or early morning, and increase road warning lights at night, which can effectively reduce the frequency of traffic accidents.

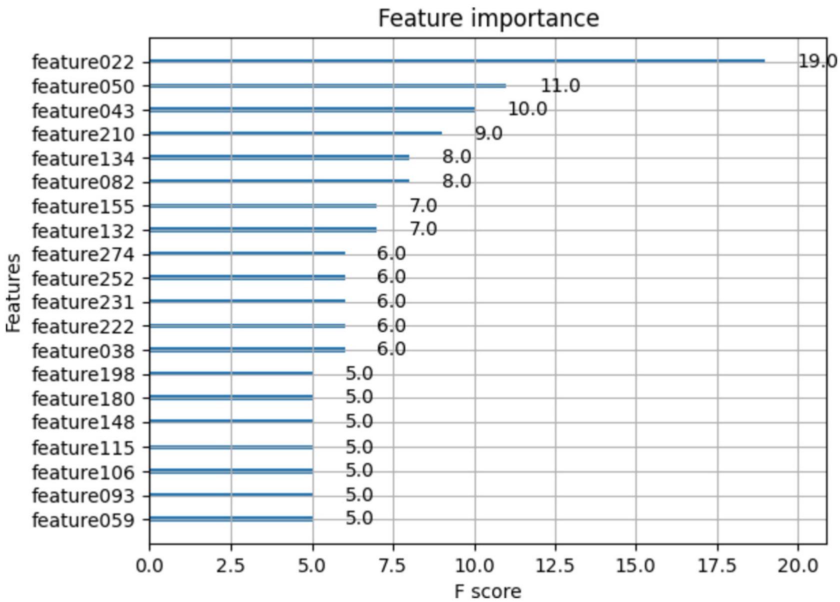

Fig. 1. Feature importance ranking diagram

ROC is a visual tool for evaluating classification models. The AUC, defined as the area under the ROC curve, is generally between 0.5 and 1, and as a value, classifiers corresponding to larger AUCs work better. At the end of this experiment, the AUC-ROC curve was plotted using XGB, as shown in Fig. 2. The AUC value obtained from the experiment is 0.550, which demonstrates the model's ability to anticipate various outcomes, but the prediction is not satisfactory. It is speculated that this may be due to the small number of selected eigenvalues, etc.

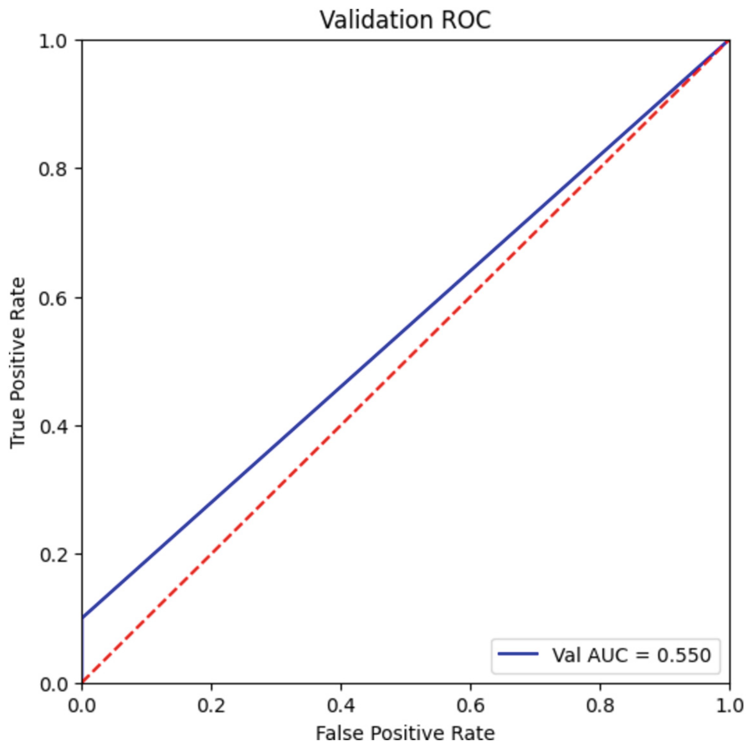

Fig. 2. Roc example XGB

4 Conclusion and Discussion

In this paper, a statistical data set of traffic anomalies over some time is used and the data is processed. Every five minutes is used as a feature point for detection. XGB is used as the experimental model to detect the anomalous values in the dataset, and the ROC curve is finally used to detect the predictive effect of the detection results. However, probably due to the simplicity of the experimental procedure in this experiment, the detection results were not thorough enough and the AUC values were not satisfactory. It is hoped that the experimental process can be improved or more accurate detection methods can be used in future studies to make the prediction more accurate and better reflect the traffic anomalies, which will have more practical guidance for traffic management and people's travel.

References

1. Wang, F.Y.: Parallel control and management for intelligent transportation systems: concepts, architectures, and applications. IEEE Trans. Intell. Transp. Syst. **11**(3), 630–638 (2010)
2. Lee, E.H.: Exploring transit use during COVID-19 based on XGB and SHAP using smart card data. J. Adv. Transp. **2022** (2022)

3. Zhang, L.: Retraction note: management of offshore oil pollution and logistics transportation based on decision tree. Arab. J. Geosci. **14**(24) (2021)
4. Gomez, C., Guardia, A., Mantari, J.L., Coronado, A.M., Reddy, J.N.: A contemporary approach to the MSE paradigm powered by Artificial Intelligence from a review focused on Polymer Matrix Composites. Mech. Adv. Mater. Struct. **29**(21) (2022)
5. Tripathi, B., Sharma, R.K.: Modeling bitcoin prices using signal processing methods, Bayesian optimization, and deep neural networks. Comput. Econ. (2022)
6. Lv, L.: RFID data analysis and evaluation based on big data and data clustering. Comput. Intell. Neurosci. **2022** (2022)
7. Shimi, A., Ebrahimi Dishabi, M.R., Abdollahi, A.M.: An intelligent parking management system using RFID technology based on user preferences. Soft. Comput. **26**(24), 13869–13884 (2022)
8. Sun, B., Cheng, W., Bai, G., et al.: Correcting and complementing freeway traffic accident data using Mahalanobis distance-based outlier detection. Techn. Gaz. **24**(5), 1597–1607 (2017)
9. Sun, B., Geng, R., Zhang, L., et al.: Securing 6G-enabled IoT/IoV networks by machine learning and data fusion. EURASIP J. Wirel. Commun. Netw. **2022**(1), 1–17 (2022)
10. Hou, Y., Edara, P., Sun, C.: Traffic flow forecasting for urban work zones. IEEE Trans. Intell. Transp. Syst. **16**(4), 1761–1770 (2015)
11. Pandey, P., Bandhu, K.C.: A credit risk assessment on borrowers classification using optimized decision tree and KNN with Bayesian optimization. Int. J. Inf. Technol. **14**(7) (2022)
12. Huang, W., Song, G., Hong, H., et al.: Deep architecture for traffic flow prediction: deep belief networks with multitask learning. IEEE Trans. Intell. Transp. Syst. **15**(5), 2191–2201 (2014)
13. Sun, B., Geng, R., Xu, Y., et al.: Prediction of emergency mobility under diverse IoT availability. EAI Endorsed Trans. Pervasive Health Technol. **8**(4), e2–e2 (2022)
14. Chen, M., Shao, H., Dou, H., Li, W., Liu, B.: Data augmentation and intelligent fault diagnosis of planetary gearbox using ILoFGAN under extremely limited samples. IEEE Trans. Reliab. (2022). https://doi.org/10.1109/TR.2022.3215243
15. Yan, S., Shao, H., Xiao, Y., Liu, B., Wan, J.: Hybrid robust convolutional autoencoder for unsupervised anomaly detection of machine tools under noises. Robot. Comput. Integr. Manuf. **79**, 102441 (2023)
16. Sun, B., Ma, L., Shen, T., et al.: A robust data-driven method for multiseasonality and heteroscedasticity in time series preprocessing. Wirel. Commun. Mob. Comput. **2021** (2021)
17. Kaminka, G.A., et al.: AUC maximization in Bayesian hierarchical models. Front. Artif. Intell. Appl. **285** (2016)
18. Sun, B., Geng, R., Shen, T., et al.: Dynamic emergency transit forecasting with IoT sequential data. Mob. Netw. Appl., 1–15 (2022)
19. McDonnell, K., Abram, F., Howley, E.: Application of a novel hybrid CNN-GNN for peptide ion encoding. J. Proteome Res. (2022)