Enhancing Traffic Safety: A Deep Learning Approach for Predicting and Understanding Traffic Violations

Hao Yang, Angela Yao

ABSTRACT: Traffic violations pose a significant challenge to our society as they may lead to traffic congestion, property damage, environmental pollution, personal injuries, and even death. Therefore, understanding the factors contributing to the occurrence of traffic violations, comprehending the spatiotemporal patterns of these violations, and predicting future violations play critical roles in effective traffic management and accident reduction. This paper thoroughly examines the factors contributing to traffic violations. Based on this examination, we propose a deep neural network to predict violation types when given information about location, time, and contextual factors. We conduct a case study in Montgomery County, Maryland, with this model, shedding light on its potential application in other regions.

KEYWORDS: *traffic violations, deep learning, violation prediction, spatiotemporal analysis, topic modelling*

Introduction

Traffic violations encompass a wide range of behaviors that violate traffic laws and regulations. These behaviors can lead to various problems, including traffic congestion, significant property damage, loss of life, environmental pollution, and more. Therefore, it is of great importance to study traffic violations. However, there is a lack of comprehensive studies on this subject. Two critical demands necessitate the study of traffic violations. First, there is a pressing need to thoroughly examine the factors contributing to traffic violations. Drivers can be encouraged to drive carefully and safely when they fully understand the factors that influence traffic violations. Second, it is essential to predict the types of traffic violations in specific times, locations, and environmental contexts. Accurate predictions will assist traffic management departments in making swift responses to reduce property damage and save lives.

Many studies have investigated the contributing factors of traffic violations. However, most of them focus primarily on demographic characteristics of drivers, such as gender, race, and age. They rarely take into account contextual factors, such as weather conditions,

road types, whether it is a road junction, traffic lights, school zones, or proximity to shopping malls and gas stations.

Numerous scholars have attempted to predict traffic accidents, utilizing classical machine learning methods such as K-Nearest Neighbor (KNN), Bayesian networks, and decision trees. Most of these studies treat prediction as a binary classification task, determining whether a traffic accident will occur in the future. In recent years, a few deep learning methods have been developed that aim to predict the frequency of traffic accidents and associated risks. However, there is a notable scarcity of studies examining traffic violations, which significantly contribute to accidents.

To address these limitations, this paper conducts a thorough examination of the factors influencing traffic violations. Based on these factors, we construct a deep neural network to predict the type of violation given the time, location, and contextual factors. We then apply this model in a case study conducted in Montgomery County, Maryland. Such a model can be adapted for use in other areas, offering substantial potential for enhancing public safety and reducing economic losses.

Method

Contributing Factors

The foremost thing is to investigate and understand what the factors are affecting the traffic violations. Haddon (1968) introduced the "Haddon matrix model", in which he considered three sets of contributing factors, namely human, vehicle and environment factors. Many studies (Miaou and Lum 1993, Shinar et al., 2001, Fosgerau 2005, Keyes et al., 2012) have demonstrated that personal characteristics such as the driver's race, gender, age, education level and income are related to traffic violation behaviours, such as seat belt usage and alcohol driving, and furtherly related to the traffic accidents caused.

Previous studies (Al-Ghamdi 2002) have also suggested that the characteristics of the vehicle, such as the age of the car, the use of the car, and the vehicle type, are important factors that are significantly associated with accident severity. In addition, multiple environmental factors are related to traffic accidents (Feng et al., 2020), including the type of the roads, the time of the day, the weather, the road conditions, season and year, street lighting and many others.

Prediction of traffic violation types

One of the critical goals of this study is to predict the type of traffic violation when given the location, time and some context information. Here we used three machine learning models as the base line modes, decision tree, neural network, and ridge regression. In addition, we are constructing a deep learning model which takes temporal and spatial dimensions into consideration. Here we use a simple MLP to encode the time information and the characteristics. Then for the location embedding, we used the space2vec (Mai et

al., 2020), then we used three layers of MLP for the final prediction of traffic violation topic. The structure of the neural network is shown as Figure 1.

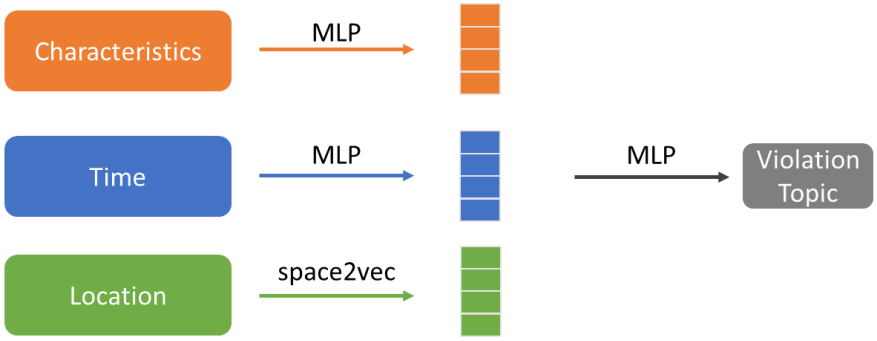

Figure 1: Deep neural network architecture

Case Study

Data

The dataset is obtained from the Montgomery Couty, Maryland. It contains traffic violation information from all electronic traffic violations issued in the County. As shown in Figure 2, we plot the traffic violations on the map.

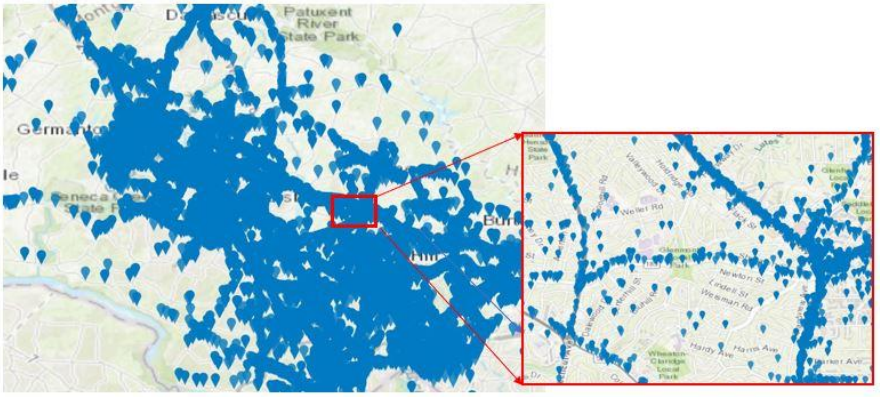

Figure 2: Spatial distribution of traffic violations

Here, we include multiple contributing factors, such as the gender, race, and driver's license state of the driver, the year, mode, and color of the vehicle, the time of day, season, and year, as well as the locations and road type. However, due to time constraints, in this case, we have not included the weather conditions or whether it is a work zone or school zone. More information will be included in future work.

Semantic analysis of violation description

One column in this dataset contains descriptions of traffic violations. It describes the traffic violation in a short sentence, resulting in approximately 2000 unique descriptions. We employ topic modeling, LDA (Blei et al., 2003), to assign each description to a small number of topics. As shown in Figure 3, the word cloud provides insights into the semantic explanations of each topic. In conclusion, there are seven topics: Failure to Yield Right of Way, Parking Violation, Traffic Control Device Violations, License & Plate Violation,

Reckless, Impaired, and Distracted Driving, Seatbelt and Child Restraint Violations, and Speeding.

Figure 3: Word cloud for the seven topics

As shown in Figure 4, we also plot the spatial distribution of each violation topic. As the maps demonstrate, there is no significant variance between the spatial patterns of various topics.

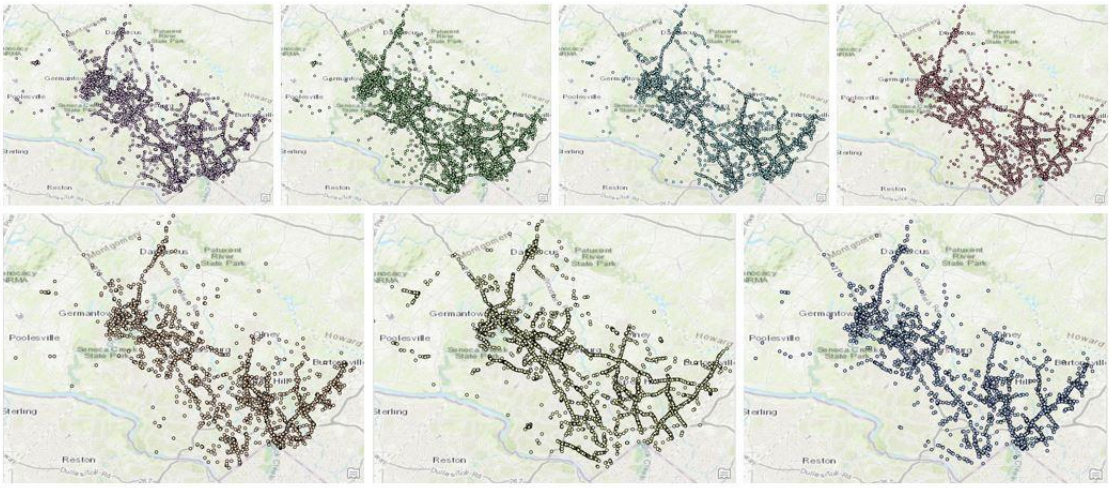

Figure 4: Spatial distributions of violations in the 7 topics

Violation Type Prediction

In this study, we use three machine learning models to predict the topic of violation. As shown in Table 1, here we did the ablation analysis, training with different input features and compare the performances among these models. We can find the Ridge regression model performs best among the three models.

Table 1: Ablation analysis of the three base models.

	Accuracy	Precision	Recall	F1-score
Extra Tree: Location + Time	0.3166	0.3755	0.3166	0.3104
Extra Tree: Location + Time	0.3109	0.2933	0.3109	0.2999

+Context Factors				
Random Forest: Location + Time	0.3218	0.3145	0.3218	0.3175
Random Forest: Location + Time +Context Factors	0.3361	0.3181	0.3361	0.3249
Ridge Regression: Location + Time	0.3248	0.2283	0.3248	0.2312
Ridge Regression: Location + Time +Context Factors	0.3283	0.2649	0.3284	0.2359

Conclusions

From the results, we can observe that, the increase of context factors will incur a slightly better performance of the models. The possible reason of the poor performance of the baseline models and the unsignificant improvement after adding more factors is that, we simply used the dummy variables to represent time and the projected location values (longitude and latitude) as input features, without considering any other temporal and spatial relationships. It demonstrates the need that we should add more contributing factors and construct a new deep neural network to incorporate the spatial and temporal relationships.

References

Miaou, S. P., & Lum, H. (1993). Modeling vehicle accidents and highway geometric design relationships. Accident Analysis & Prevention, 25(6), 689-709.

Keyes, K. M., Liu, X. C., & Cerda, M. (2012). The role of race/ethnicity in alcohol-attributable injury in the United States. Epidemiologic reviews, 34(1), 89-102.

Shinar, D., Schechtman, E., & Compton, R. (2001). Self-reports of safe driving behaviors in relationship to sex, age, education and income in the US adult driving population. Accident Analysis & Prevention, 33(1), 111-116.

Fosgerau, M. (2005). Speed and income. Journal of Transport Economics and Policy (JTEP), 39(2), 225-240.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), 993-1022.

Mai, G., Janowicz, K., Yan, B., Zhu, R., Cai, L., & Lao, N. (2020). Multi-scale representation learning for spatial feature distributions using grid cells. arXiv preprint arXiv:2003.00824.

Feng, M., Zheng, J., Ren, J., & Liu, Y. (2020, February). Towards big data analytics and mining for UK traffic accident analysis, visualization & prediction. In Proceedings of the 2020 12th International Conference on Machine Learning and Computing (pp. 225-229).

Hao Yang, PhD Candidate, Department of Geography, University of Georgia, Athens, GA 30605

Xiaobai A. Yao, Professor, Department of Geography, University of Georgia, Athens, GA 30605