Exploration of Contributing Factors to Traffic Accident Severity Based on Random Forest Model

㈱高速道路総合技術研究所　正会員　○劉　星委
㈱高速道路総合技術研究所　正会員　邢　健
早稲田大学　正会員　佐々木　邦明

1. Abstract and introduction

Traffic accidents cause a lot of loss of life and property. Understanding the causes of traffic accidents and predicting the level of traffic accident severity can assist relevant agencies to do efficient traffic management. This study uses the accident severity as an indicator and applied the accident data from Gotenba to Tokyo on the Tomei Expressway in 2019. We analyze the potential causes of each influencing factor and predicted the severity of traffic accidents.

Firstly, the accident data cleansed with information on drivers, cars, as well as the surrounding environment where the accidents occurred. Afterward, a random forest model was used for the cleansed data, applying grid search to optimize the performance of the model. Eventually, SHAP (**SH**apley **A**dditive ex**P**lanations) visualized the independence performance on each feature and their contributions to traffic accident severity.

2. Random forest and application

Tomei expressway is an important expressway connecting the chubu-region and Tokyo area. We selected data of the section from Gotenba to Tokyo (inbound), which contains 12 interchanges in total. After data cleansing and filtering, 701 accident data have been obtained. The features we have chosen in this study are: hour of day, day of week, driving error, speed, accident location, age of driver, gender of driver, vehicle type, weather, lane, congestion situation, and traffic flow rate.

We employ an ensemble machine learning approach – random forest – to predict the severity of traffic accidents, and to reveal the relationships between the different types of traffic features and traffic accidents on expressways.

We constructed the random forest methods in the Python environment with the Scikit-learn package. The datasets for accident cases were randomly split into a training subset and a testing subset, accounting for 70% and 30% of the entire data, respectively. In addition, we used the grid search and best prediction results model was obtained with n=150, d=4, m=6.

3. Results of random forest with the explainable tool

In what follows, the result calculated by random forest will be interpreted by the global and local relative importance of independent variables by SHAP, an approach that is applied to interpret the prediction results. Due to space limitations, the mathematical explanation of SHAP is not shown here. SHAP shows convincing results in both explanation and visualization in many fields.

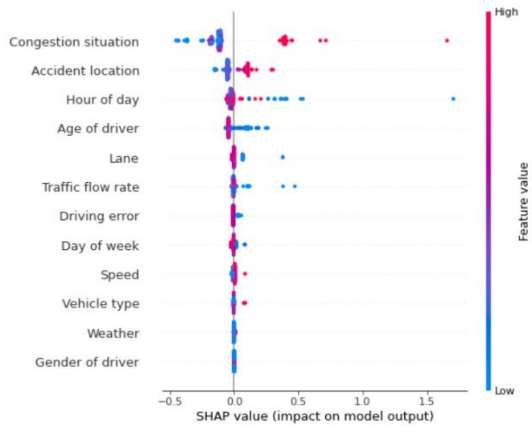

Fig.1 Relative importance of independent variables

The global importance calculated by the average absolute SHAP values per variable reflects the contribution of each feature to the prediction result (Fig.1). The horizontal location shows the effect of the feature associated with a higher or lower prediction. The color shows whether that variable is high or low for this feature.

It can be seen from Fig.1 that congestion situation and accident location have a higher effect on the prediction, however, the gender of drivers has hardly any influence. As

Keywords: Random forest, traffic accident analysis, severity

連絡先　〒194-8508　東京都町田市忠生 1 丁目 4-1（株）高速道路総合技術研究所（NEXCO 総研）交通研究室
TEL：042-791-1679，FAX：042-791-1099，　E-mail：liu@ri-nexco.co.jp

for the drivers' individual information, the age of driver lists the fourth and the distribution of its color shows that drivers with younger ages have a positive contribution to cause severity.

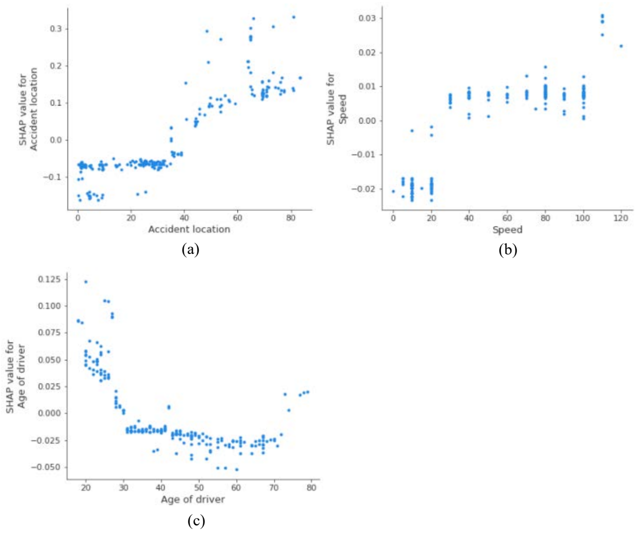

Fig. 2 The SHAP value plot on accident severity

Afterwards, we will analyze the relationship between accident severity and each feature in depth. As for the accident location (Fig.2(a)), it represents the distance coordinates where the accident occurred. When location is less than 40KP, the local effects are at a lower level and this effect appears obvious from 40KP and the local peak arrived when accident location is around 45KP and 65KP. It reminds us to be particularly aware of the travel conditions around segments 40~50KP and 60~70KP. Fig.2 (b) demonstrates the effects of speed on accident severity. We notice that when speed exceeds 100km/h, the local effect on severity increases a lot which reminds us to keep the driving speed in a reasonable range. There has the same tendency as the effects of age of driver shown in Fig.2 (c). Drivers who are younger than 30 or older than 70 are more likely to be affected than other ages, especially novice drivers.

Furthermore, we perform a microscopic analysis to shift our focus on each accident case shown in Fig.3. In Fig.3, f(x) is the predicted accident severity for the current accident. The red labels represent features that make severity increase, and the blue labels are on the contrary.

Fig.3 (a) and Fig.3 (b) provide the result with high accident severity, 1.46 and 1.63. They both have features of congestion situation=2, contributing a lot to the accident severity. Congestion situation=2 represents this accident vehicle is located at the end of queue.

Fig. 3 The force plot for representing accidents ((a), (b), (c), (d) from upside)

In low predicted severity cases Fig.3 (c) and Fig.3 (d), they have more common features which have negative effects on severity, including making the same driving error and no congestion situation. One of the obvious differences is the age of driver, the driver in case Fig.3 (d) is 23 years old which has a higher possibility to cause the accident.

4. Conclusion

In sum, this study primarily measures the relationships between accident features and accident severity on expressway, showing some insights from the non-linearity varieties shown in SHAP. We suggest that future management strategies can pay attention to the specific segments in order to reduce the severity. The one area is located at the end of congested areas and the other area is the roadway between 40~50KP and 60~70KP. These two segments have a higher possibility to occur the accident.

Another is about the drivers, in addition to the elderly, it is also necessary to strengthen safety driving education for younger people under 30 years old. Compared with the elderly, young people are more likely to be injured, probably due to the high speed. The results also emphasized the need to speed limitation, especially for cars with speeds higher than 100km/h.

Reference

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324