Research on Distracted Driving Identification of Truck Drivers Based on Simulated Driving Experiment

Weihao Zhang1*, Huiyan Zhang2

1School of Traffic and Transportation, Beijing Jiaotong University, Beijing, 100044, China
2Department of Sociology, Peking University, Beijing, 100044, China

*Corresponding author’s e-mail: 19120960@bjtu.edu.cn

Abstract. Trucks drivers may be distracted during long-distance transportation, which may threaten the safety of themselves and surrounding vehicles. In this regard, this article first investigates the types of distracted driving that frequently occur in China and select typical distracted driving behaviors to carry out the simulated driving experiment and collect driving performance data. The independent sample T-test was used to select the distracted driving identification indicators. An integrated distracted driving identification model based on Binary Logistic Regression and Fisher discriminant analysis was established. The accuracy of the integrated model in identifying the driver's distracted driving state is 94.79%. This research can provide a basis for the driver assistance systems that reflect the degree of driver distraction.

1. Introduction

Road freight transportation has made a huge contribution to China's economic development. But at the same time, large trucks on the road are also one of the dangerous sources of traffic accidents. Relevant studies have shown that, taking dangerous goods tank vehicles as an example, the driver is the primary factor leading to accidents, and the accidents caused by them account for 67% of the total number of accidents[1]. In the process of driving a vehicle, in addition to main driving tasks such as vehicle control and lane-keeping, the driver also performs distracting driving behaviors, such as drinking water, eating, calling, and sending text messages. Distracted driving is defined as[2]: The driver turns his attention to tasks or activities that are not related to the main driving task, which leads to the decline of the driver's vision, consciousness, decision-making, and operating ability.

Many scholars have conducted a lot of research on the impact of distracted driving behavior on vehicle performance. Tarabay Rana et al.[3] designed several driving simulation experiments with different levels of difficulty. The study found that the driver will adopt adjustment behaviors such as reducing the speed of the vehicle, to be able to drive the vehicle and perform additional tasks at the same time. Kountouriotis et al.[4] compared and analyzed the changes in the operating behavior of drivers in cognitive distraction and normal driving and found that compared with normal driving, the steering acceleration, steering wheel slew rate and steering entropy of the driver in the distracted state are all obvious increase.

In the identification of distracted driving behavior, the commonly used indicators mainly include physiological information[5], eye movement[6], driving performance indicators[7] and mixed indicators[8] (a combination of the above multiple indicators). Although there have been a large number of cases of distracted driving at home and abroad, the research objects are mostly focused on small
vehicles, and there are few studies on trucks, especially dangerous goods vehicles.

2. Materials and Methods

This paper takes the dangerous goods vehicle as the research object. Firstly, the typical distracted driving behavior of truck drivers is selected by Naturalistic Driving Study (NDS), and then the traffic accidents related to distracted driving are summarized through General Estimates System (GES) and Crash Report Sampling System (CRSS). Combined with the frequent and dangerous of distracted driving, the distracted driving behavior of watching a mobile phone is selected as the research object, and the truck driver simulated driving experiment is designed and carried out. The independent sample T-test is used to compare the difference of driving performance between looking at the mobile phone while driving and normal driving, so as to provide a basis for the selection of distracted driving identification model indicators. Then, this paper establishes an integrated distracted driving identification model based on Binary logistic Regression and Fisher discriminant analysis, and verifies the effectiveness of the model based on the correct rate.

2.1. Selection of typical distracted driving behavior

First of all, the distracted behavior was investigated by using Naturalistic Driving Study (NDS), and 15 dangerous goods tank vehicles were selected for 5 consecutive days by taking the GPS surveillance video of a transportation enterprise in Ningbo, Zhejiang Province, and the statistical time was up to 360 hours. During this period, the cumulative time of distracted driving is 54439s, in which the number of distractions and their proportion are shown in figure 1. As can be seen from the picture, mobile phones account for the largest proportion of all statistical distracted behaviors.

Based on the U.S. GES and CRSS database, the number of traffic accidents related to distracted driving recorded in the database from 2016 to 2018 is counted. As shown in Figure 2, although "Distracted by things outside the car, Distracted by things inside the car, and Careless" accounts for a large proportion, it is difficult to design corresponding simulated driving experiments. Combined with the results of the Naturalistic Driving Study, this research finally chose “Watching mobile phone” as a distracted driving behavior to design a simulated driving experiment. For the convenience of presentation, this article will use distracted driving to refer to the driving behavior of watching a mobile phone while driving.
Figure 2. Number of accidents caused by different types of distracted driving

2.2. Truck driver simulated driving experiment

2.2.1. Experimental scheme
A total of 30 drivers of dangerous goods transport vehicles with rich driving experience were recruited to carry out the experiment. Before carrying out the simulated driving experiment, all participants are required to fill out a questionnaire, including personal driving habits, driving experience, traffic accident experience, physical and psychological status, and so on. The average age of the participants was 37.7 years (standard deviation is 3.91), and the age range was 28-53 years; all participants had qualified driving licenses and more than five years of driving experience, and all participants did not have any visual and psychological problems. The driver first filled out the questionnaire, and in order to make them familiar with the operation of the driving simulator, the driver will be arranged to carry out driving adaptability training before the experiment. After the formal test begins, the driver is required to drive the dangerous goods tank vehicle normally. When the message appears on the screen, the driver takes out the mobile phone and browses any web page for about 10 seconds. Figures 3 and 4 show the experimental scene diagram and the distracted driving state diagram of the driver, respectively.

2.2.2. Selection of distracted driving identification indicators
In this experiment, the driving simulator produced by FORUM 8 is used to carry out the simulated driving experiment, which can output the performance data of the driving vehicle and the vehicles within a certain radius. This paper selects the average vehicle speed, vehicle speed standard deviation (SD), lateral acceleration mean, lateral acceleration SD, longitudinal acceleration mean, longitudinal acceleration SD, steering wheel input value mean, steering wheel input value SD, steering wheel rotation rate mean, steering wheel rotation rate SD, brake pedal input value mean, brake pedal input value SD, lane deviation mean, lane deviation SD as an optional set of distracted driving identification indicators. This paper further compares the difference of driving performance between distracted driving and normal driving through independent sample T-test, and selects the indicators with significant differences.
as the input parameters of the distracted driving identification model.

2.3. Truck driver distracted driving Identification Model based on the simulated driving experiment

2.3.1. Distracted driving Identification Model based on Binary Logistic Regression

The model constructed in this paper takes the selected distracted driving identification indicators as input and outputs two states of normal driving or distracted driving (represented by 0 and 1 respectively). The mathematical model of Binary Logistic Regression is:

\[
\text{Logit} \, P = \ln \frac{P}{1-P} = \beta_0 + \sum \beta_i x_i
\]

(1)

\[
\frac{P}{1-P} = \exp \left( \beta_0 + \sum_{i=1}^{k} \beta_i x_i \right)
\]

(2)

\[
P = \frac{\exp \left( \beta_0 + \sum_{i=1}^{k} \beta_i x_i \right)}{1 + \exp \left( \beta_0 + \sum_{i=1}^{k} \beta_i x_i \right)}
\]

(3)

Where P indicates the probability of identifying distracted driving behavior. \( x_i \) represents distracted driving identification indicators. The Regression coefficient \( \beta_i \) represents the logarithmic change of the ratio of the occurrence of events to the probability of non-occurrence when the independent variable \( x_i \) changes one unit.

2.3.2. Fisher discriminant model

LDA (Linear Discriminant Analysis), also known as Fisher discriminant analysis, is a classical linear discriminant method. Its principle is shown in figure 5. For the two classification problem, it tries to find a straight line L to project the training set samples onto the line, so that the same kind of samples are as close as possible, while the different kinds of samples are as far away as possible.

![Figures 5. Fisher discriminant Analysis principle diagram](image)

3. Results

3.1. Indicators extraction of distracted driving identification model

Since the candidate indicators set for distracted driving identification established in 2.2.2 has many indicators. Taking all of them as input variables may cause slow or overfitting problems. In this regard,
this paper compares the difference between the driver's driving performance under distracted driving and normal driving by independent sample T-test and selects significant indicators for model training. Table 1 shows that compared with normal driving, distracted driving has significant differences in average vehicle speed, lateral acceleration SD, steering wheel input value SD, steering wheel rotation rate SD, lane deviation mean, and lane deviation SD. Specifically, compared with normal driving, the driver will reduce the vehicle speed as a compensation for distracted driving; but the SD of steering wheel input value and steering wheel rotation rate when distracted are significantly higher. Reflected in vehicle trajectory, lateral acceleration SD and lane deviation SD during distracted driving are significantly higher than those of normal driving; this indicates that distraction affects driver lateral control ability and results in large fluctuation of vehicle lateral motion; Moreover, lane deviation mean during distracted driving was significantly different from that of normal driving.

Table 1. Independent sample T-test results of distracted driving and normal driving

| Indicators                  | distracted driving | normal driving | P     |
|-----------------------------|--------------------|----------------|-------|
| the average vehicle speed   | 63.7305            | 78.0994        | 0.001 |
| vehicle speed SD            | 2.5321             | 1.5026         | 0.370 |
| lateral acceleration mean   | -0.0172            | -0.0124        | 0.902 |
| lateral acceleration SD     | 0.1005             | 0.0432         | 0.024 |
| longitudinal acceleration mean | -0.6189          | -0.2345        | 0.180 |
| longitudinal acceleration SD | 0.4647            | 0.3096         | 0.652 |
| steering wheel input value mean | 0.0003           | 0.0000         | 0.555 |
| steering wheel input value SD | 0.0017           | 0.0005         | 0.0006|
| steering wheel rotation rate mean | 0.0000         | -0.0001        | 0.756 |
| steering wheel rotation rate SD | 0.0057           | 0.0018         | 0.0017|
| brake pedal input value mean | 0.0370            | 0.0376         | 0.938 |
| brake pedal input value SD  | 0.0102             | 0.0110         | 0.956 |
| lane deviation mean         | -0.1124            | 0.3078         | 0.030 |
| lane deviation SD           | 0.2925             | 0.0647         | 0.025 |

By comparing the difference between the driving performance of distracted driving and normal driving vehicles, this paper uses the average vehicle speed, lateral acceleration SD, steering wheel input value SD, steering wheel rotation rate SD, lane deviation mean, and lane deviation SD as the identification indicators of distracted driving behavior.

3.2. Distracted driving Identification Model based on Binary Logistic Regression.

This paper selects more than 300 samples of distracted driving and normal driving. 70% of the samples are selected as the training set and the remaining 30% as the test set. The Hosmer test shows that the model fits well with the observations (P=0.598>0.05). Table 2 shows the classification of the model in the test set. After calculation, the accuracy of the model is 90.6%.

Table 2. Classification results of the Binary Logistic Regression model

| Reality               | normal driving | distracted driving |
|-----------------------|----------------|-------------------|
| normal driving        | 39             | 4                 |
| distracted driving    | 5              | 48                |

The constructed truck driver's distracted driving identification model based on Binary Logistic Regression is

\[
P = \frac{e^{0.647-0.05X_1-109.35X_2+1712.04X_3+2875.96X_4-0.15X_5-0.21X_6}}{1+e^{0.647-0.05X_1-109.35X_2+1712.04X_3+2875.96X_4-0.15X_5-0.21X_6}}
\]

In the formula, \(X_1 \sim X_6\) respectively represent the average vehicle speed, lateral acceleration SD,
steering wheel input value SD, steering wheel rotation rate SD, lane deviation mean, and lane deviation SD. For a sample, bring the driving performance data \( x_1 \sim x_6 \) into the model, if \( P > 0.5 \), it is determined that the driver is distracted. By counting most of the samples with incorrect predictions, its P-value is in the range of 0.4 to 0.6. It can be seen that when P is close to 0.5, the probability of judging distracted and undistracted driving is similar, and wrong predictions may be made.

3.3. Fisher discriminant model

This paper selects more than 300 samples of drivers with distracted driving and normal driving. 70% of the samples are selected as the training set and the remaining 30% as the test set. The Fisher discriminant model is as follows, where the normal driving label is A and the distracted driving label is B.

\[
A: Z_A = -62.62 + 1.57x_1 - 398.29x_2 + 33331.37x_3 - 1441.86x_4 + 4.45x_5 + 13.28x_6
\]

\[
B: Z_B = -52.53 + 1.40x_1 - 414.88x_2 + 31510.84x_3 + 130.47x_4 + 3.84x_5 + 8.83x_6
\]

Put the test sample into equations (1) and (2) to get \( A: Z_A \) and \( B: Z_B \). If \( Z_A < Z_B \), it is determined that the driver is distracted. Table 3 shows the classification of the model in the test set. After calculation, the accuracy of the model is 86.45%. In the same way, by counting most of the samples with incorrect predictions, it is found that \( A: Z_A \) and \( B: Z_B \) are similar, which makes it easier to make wrong predictions.

Table 3. Classification results of Fisher discriminant model

| Reality       | Forecast Result |
|---------------|-----------------|
| normal driving| 38              |
| distracted driving | 5          |
| distracted driving | 8          |
| normal driving | 45              |

3.4. Integrated identification model of driver distracted driving based on Binary Logistic Regression model and Fisher discriminant model

Based on the research in 3.2 and 3.3 of this article, when P is close to 0.5, the Binary Logistic Regression model is prone to make wrong predictions. When \( A: Z_A \) and \( B: Z_B \) are similar, the Fisher discriminant model is easy to make wrong predictions. The reason is that, in reality, when the driver is distracted, his attention for controlling the vehicle will be distracted. The degree of distracted driving is different, so the distracted attention will be more or less. Therefore, when the driver is less distracted, in order to improve the accuracy of the distracted driving identification model, this paper establishes an integrated identification model of driver distracted driving based on Binary Logistic Regression and Fisher discriminant analysis. For a sample, the integrated identification model uses \( x_1 \sim x_6 \) as the input of the two models in 3.1 and 3.2 respectively. When both models predict that the driver is distracted, the output is "obvious distraction"; when only one of the models predicts that the driver is distracted, the output is "slight distraction"; when both models predicts that the driver is driving normally, it outputs "normal driving". In order to verify the accuracy of the integrated model, the results "obvious distraction" and "slight distraction" are counted as distracted driving, and the test set data is brought into the integrated model. The accuracy of the model is 94.79%. Table 4 shows the classification of the integrated model in the test set. This result shows that the integrated identification model can improve the accuracy of distracted driving identification.

Table 4. Classification results of the integrated identification model

| Reality       | Forecast Result |
|---------------|-----------------|
| normal driving| 41              |
| distracted driving | 2          |
| distracted driving | 3          |
| normal driving | 50              |

According to the output result, the warning information output by the driving assistance system can
be designed. When the integrated identification model outputs "slight distraction", the system sends out a reminder message, such as: "Please concentrate on driving"; and when the model outputs "obvious distraction", the system sends out a warning message, such as: "You have been distracted, please pay attention"; the driving assistance system that reflects the driver’s distraction level can effectively reduce the adverse effects of distracted driving on the driving safety.

4. Conclusions
Taking distracted driving behavior of truck drivers as subjects, we designed the simulated driving experiment to collect drivers' performance data during normal driving and distracted driving, respectively. Then Binary logistic Regression model and Fisher discriminant model were established respectively for distracted driving behavior identification. Aiming at the problem that a single model may easily make the wrong prediction when identifying slight distraction behavior. This paper integrates two models and designs a driving assistant system which reflects the degree of driving distraction according to the output result of the integrated model.

Considering that most truck drivers have rich driving experience, this group of people will often form their unique driving style or driving habits during long-distance transportation. The formation of this style and habits may be related to the driver’s gender, age, driving experience, and personality. Subsequent research in this paper will extract indicators that characterize the driver’s driving style or driving habits, and strive to dig out the relationship between driver’s characteristics and driving behavior, hoping to further improve the accuracy of the distracted driving identification model.

References
[1] Shen X Y, Li X N, Xie P, Xie C. (2012) Statistical analysis on 886 road HAZMAT transportation accidents by the tank truck. J. Journal of Safety Science and Technology, 8(11): 43–48.
[2] Flaherty M R, Kim A M, Salt M D, et al. (2020) Distracted Driving Laws and Motor Vehicle Crash Fatalities. 145(6).
[3] Rana T, Maya A Z. (2018) Assessing the effects of auditory-vocal distraction on driving performance and physiological measures using a driving simulator. J. Transportation Research Part F: Traffic Psychology and Behaviour, 58: 351-364.
[4] Kountouriotis G K, Spyridakos P, Carsten O M J, et al. (2016) Identifying cognitive distraction using steering wheel reversal rates. J. Accident Analysis & Prevention, 96: 39-45.
[5] Mousa K W, Murugappan M, Badlishah A. (2013) Subtractive Fuzzy Classifier Based Driver Distraction Levels Classification Using EEG. J. Journal of Physical Therapy Science, 25(9):1055-1058.
[6] Zhou Y, Fu R, Yuan W, et al. (2018) Research on drivers’ cognitive distracted recognition model based on random forest. China Safety Science Journal, 28(01): 20–25.
[7] Zhang H, Qian D L, Shao C F , Chen Q M, Shan Q C. (2018) Drivers Distraction States Identification in Simulating Driving Environment. China Journal of Highway and Transport, 31(04):43-51.
[8] Liao Y. (2015) Driver Distraction Detection through Driving Performance and Eye Movement: from Feature Extraction to Classifier Design [D]. Tsinghua University.