

Research Article

How Do Vehicles Make Decisions during Implementation Period of Discretionary Lane Change? A Data-Driven Research

Qiangru Shen ¹, Yujie Ni,¹ Hui Cao ¹, Wangping Qian ¹ and Gen Li ²

¹School of Transportation and Civil Engineering, Nantong University, Nantong 226019, China

²College of Automobile and Traffic Engineering, Nanjing Forestry University, Nanjing 210037, China

Correspondence should be addressed to Wangping Qian; qianwangping@my.swjtu.edu.cn and Gen Li; ligen@njfu.edu.cn

Received 5 July 2022; Revised 19 October 2022; Accepted 24 November 2022; Published 14 February 2023

Academic Editor: Zhenzhou Yuan

Copyright © 2023 Qiangru Shen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To investigate and compare the lane changing behavior of passenger cars and heavy vehicles during the implementation period (defined as the interval from the start time to the end time of a lane change maneuver), this study applies the gradient boosting decision tree (GBDT) method to model the lane changing behavior of heavy vehicles and passenger cars, respectively. Results show that the lane change models vary with the vehicle types and lane change directions. Different factors are considered by the drivers of passenger cars and heavy vehicles when implementing lane changes to different directions. Partial dependence plots of GBDT models reveal that the influence of independent variables on lane changing behavior is nonlinear and complicated, which means that the same variable leads to various effects on the lane change decision across different vehicle types and lane change directions. In contrast with other state-of-the-art methods, the proposed method can obtain more accurate results. The findings indicate that it is necessary to build specific lane change models based on vehicle types and lane change directions for microscopic traffic simulators and autonomous vehicles.

1. Introduction

Recently, several projects of cooperative vehicle infrastructure systems have been undertaken all over the world. Among the numerous projects, the vehicular ad hoc network (VANET) has an important role in the upcoming “super smart highway.” In a VANET, all vehicles and roadside infrastructures are equipped with wireless interfaces and sensors. Therefore, vehicles are able to perceive the surrounding traffic conditions and communicate with roadside units and other vehicles, which can help vehicles realize autonomous driving. However, it will take a long time to realize full autonomous driving [1–3]. Advanced driver assistance systems (ADASs) and connected and autonomous vehicles (CAVs) will play important roles in the near future. Accurate and robust lane change decision-making, planning, and controlling systems are extremely important in developing ADAS and CAVs to improve traffic safety, promote fuel economy, ease urban traffic congestion, and optimize the utilization of roadways [4, 5].

It is stated in several studies that lane change behavior has negative impacts on traffic safety operations and is responsible for traffic breakdown in certain situations [6–9]. Thus, one of the most important tasks of driving assistance system is to help drivers take a safe lane change behavior and reduce the crash risk. Such task is realized by a subsystem of the driving assistance system, named as lane change assistance system or merging assistance system, which can tell the driver whether it is safe or not to change lanes at the present time based on certain decision rules [10]. Figure 1 shows a schematic of a lane change assistance system via VANETs.

However, most of the existing lane change assistance systems are designed for passenger cars, which are very different from heavy vehicles (e.g., trucks) in size, performance, and maneuverability. According to a report of the Federal Motor Carrier Safety Administration [11], large trucks accounted for 3429, 3622, and 3864 fatal crashes in 2014, 2015, and 2016, respectively. The percentage of truck-involved fatal crashes in all vehicle fatal

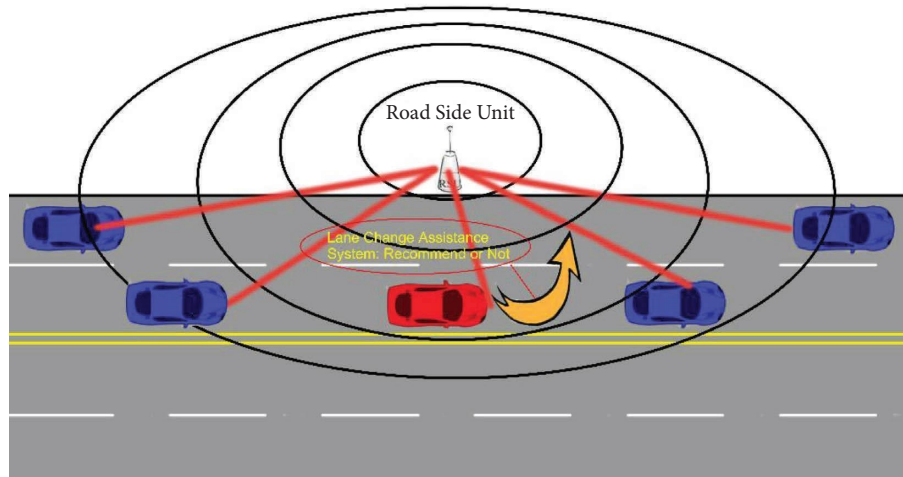


FIGURE 1: Lane change assistance system via VANETs.

crashes is much higher than the percentage of trucks in vehicle ownership. Moreover, about half of the fatal crashes were two-vehicle crashes that involved one large truck and one nontruck vehicle type. As an active vehicle interaction behavior, lane changes are one of the main sources of two-vehicle crashes of one large truck and one nontruck vehicle. Thus, it is rather urgent to build lane change models for heavy vehicles, which can be applied to lane change assistance systems to help drivers make safer lane change decisions.

Most lane change models are designed for making decisions. Weng et al. [10] stated that it is necessary to model the merging behavior during the entire merging implementation period because it usually takes several seconds to execute the lane change maneuver. As depicted in Figure 2, the blind areas of heavy vehicles are much larger than passenger cars during the lane change implementation period, making heavy vehicles face more risks. Hence, it is necessary to build a lane change model for heavy vehicles during implementation. In this study, GBDT is proposed to analyze the lane changing behavior. As a nonparametric method, the GBDT has several advantages over the parametric models and has been successfully applied in many fields [12–15]. Firstly, GBDT can provide higher modelling accuracy than parametric models. Secondly, no predetermined assumptions related to data distributions are needed for GBDT. Thirdly, GBDT can handle a large number of explanatory variables at the same time. GBDT is an ensemble learning technique that combines the classification and regression tree (CART) and the boosting technique. The basic scheme of GBDT is to use a series of weak CART to achieve better results than a single strong learner [16, 17]. Therefore, GBDT retains the excellent performance and good interpretability of CART and, at the same time, overcomes the deficiency of CART that it is easily disturbed by perturbations in the training data [18]. The partial effect plots provided by GBDT can be directly used to understand the nonlinear effects of influencing variables, which is also one of the most important reasons why we choose GBDT.

Thus, we can use GBDT to deeply analyze lane changing decisions.

The remaining of this paper is organized as follows. Section 2 provides a state-of-the-art review of the existing studies. Section 3 gives the methodology for how to build a GBDT model. Section 4 describes the data used in this paper. Results and discussion are presented in Section 5. Finally, the concluding remarks are presented in Section 6.

2. Literature Review

Gipps [19] is believed to put forward the earliest comprehensive framework of lane change behavior. In Gipps' model, the lane change decision was determined by some fixed rules, such as safety, route, the location of permanent obstructions, the presence of heavy vehicles, and speed advantage. Then, several studies soon widely used similar frameworks [20–23]. This kind of lane change model is called the rule-based model, which has the advantage of simplicity. The rule-based models are also widely applied to microscopic traffic simulators, such as VISSIM, Paramics, and ARTEMIS. However, these models are not easy to calibrate. Game theory was used by Kita [24] to describe the give-way behavior during merging process. In recent years, more complicated rules were developed to better model the lane change behavior. For example, free, forced, and cooperated lane changes were proposed and modelled separately [22]. Furthermore, considering the unobserved plans of drivers, a framework of latent plans was proposed by Choudhury et al. [25]. In all the above models, gap acceptance is considered the most important part. However, it has been criticized by several studies for the inconsistency in reality [9, 26–28]. Thus, discrete choice models were proposed to overcome the deficiency [29–31]. The output of discrete choice models is the probability of a lane change maneuver. Recently, driver heterogeneity has drawn much attention and been incorporated in discrete choice models such as mixed logit, finite mixture of logistic regression, and mixed probit models [6, 32]. Nevertheless, the methods used

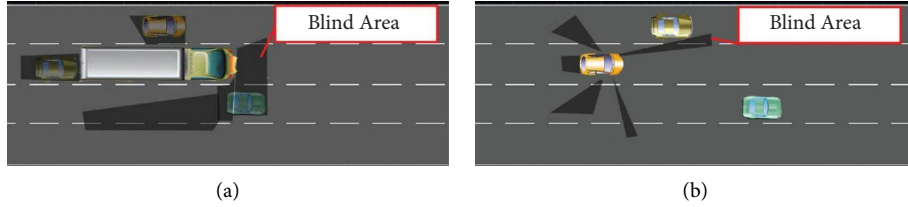


FIGURE 2: Blind area of (a) heavy vehicles and (b) passenger cars.

in the abovementioned studies are all parametric statistical approaches, having some limitations in dealing with the complex nonlinearity in human behaviors [3], which can be better addressed by artificial intelligence or data-driven methods. Balal et al. [34] proposed a fuzzy logic model and achieved promising results. Combining the Bayesian network and decision tree, Hou et al. [34] built a mandatory lane change model for autonomous driving. Classification and regression tree (CART) was also applied to lane change models [10, 35]. Xie et al. [36] proposed a lane change model based on deep learning. Moridpour et al. [37] modelled the lane change decision using a fuzzy logic method for heavy vehicles.

Most of the above studies focused on lane change decision and treated the lane change behavior as an instant event. Considering that it takes several seconds for a vehicle to complete a lane change behavior, more and more researchers realized that it is necessary to build models that can describe the whole lane change implementation period [9, 38]. Some of them focused on the longitudinal acceleration and deceleration behavior during the implementation period based on modified car-following models [39–43] or data-driven models [7]. The lane change trajectory was also investigated in some studies [44]. Besides, several studies tried to model the decisions during lane change implementation period based on the dynamic gap acceptance model [45], hidden Markov models [46], and cellular automata models [47]. CART was recently applied by Weng et al. [10] to model the merging behavior in the work zone during implementation.

Nevertheless, some limitations still exist in the literature. Firstly, most of the previous studies focused on the lane change behavior of passenger cars but neglected heavy vehicles. Moreover, the limited studies about heavy vehicles were mostly based on the NGSIM dataset, which contains less than 50 lane change maneuvers of heavy vehicles [8, 37, 39, 48]. Secondly, drivers may either change to the left or right lane, resulting in different behaviors [37, 49]. However, the decisions behind different behaviors were ignored in previous studies. Thirdly, compared with parametric methods, data-driven techniques can improve prediction accuracy, while most are black-box methods and cannot be used to understand the internal mechanism of lane change behaviors.

To address the above shortcomings, GBDT is used to model and analyze the lane change decision during implementation. Different from other black-box methods, GBDT can not only achieve satisfying results but also

provide ways to explore the internal mechanism of the trained model. GBDT has been successfully used in fields of transportation and produced promising results [3, 18, 50–52]. The main contributions of this study contain three aspects. Firstly, a data-driven method is applied to model the lane change decision during the implementation period and understand the influence of different variables on the lane change decision behaviors. Secondly, the lane changing behavior of heavy vehicles is explored and compared to passenger cars based on a large-scale dataset. Thirdly, the lane change direction is considered, and lane changes to different directions are modelled separately to investigate the different influencing factors.

3. Methodology

3.1. Gradient Boosting Decision Trees. Boosting technique is the key to GBDT, and it generates a series of weak learners sequentially and iteratively. In each iteration, the weak learner is trained based on the residual of the previous one. Let X and y denote the vector of input variables and the response variable, respectively, and the trained learner $F(X)$ is the weighted summation of all weak learners as follows:

$$F(X) = \sum_{m=1}^M \rho_m \cdot h_m(X, \alpha_m), \quad (1)$$

where $h_m(X, \alpha_m)$ is the base function, α_m is the corresponding parameter, ρ_m is the step-size parameter, m is the number of the current iteration, M is the number of total iterations, and X is the vector of input variables.

In each iteration, the boosting technique will estimate ρ_m and α_m by minimizing the loss function

$$(\rho_m, \alpha_m) = \arg \min_{\beta, \alpha} \sum_{i=1}^N L(y_i, F_{m-1}(x_i) + \rho h(X_i, \alpha)), \quad (2)$$

where N is the number of training samples. $L(y, F(X))$ is the loss function which reflects the accuracy of the training model. Different forms of loss functions can be used in GBDT when addressing different problems. For example, the squared loss function is usually used in regression problems. The lane change decision problems can be regarded as typical classification problems. Similar to logistic regression model, the log-likelihood loss function is used in this paper:

$$L(y, F(X)) = \ln(1 + e^{-yF(X)}) \quad y \in \{-1, 1\}. \quad (3)$$

No matter which loss function is used, it is not easy to accurately estimate ρ_m and α_m according to (2). To address this problem, Friedman [53] proposed an approximation algorithm based on the assumption that the loss function always declines fastest in the direction of negative gradient:

$$g_m = - \left[\frac{\partial L(y, F(X))}{\partial F(X)} \right]_{F(X)=F_{m-1}(X)}. \quad (4)$$

Two ways were adopted by Friedman [53] to improve the generalization of GBDT. Firstly, a shrinkage parameter (or learning rate) η was used to scale the contribution of each tree:

$$F_m(X) = F_{m-1}(X) + \eta \cdot \rho_m \cdot h_m(X, \alpha_m). \quad (5)$$

A lower learning rate can achieve a better result but requires more trees to converge. Thus, it is necessary to choose proper η to seek the balance between the precision and computing burden. The second way was the adoption of random sampling of training data. In each iteration, a subsample fraction $S_{f_r d}$ is used to draw from the training sample without replacement to speed up the modelling time and guard against over-fitting.

A simple or weak CART is trained in each iteration to improve the model. The trained trees are only allowed to grow to a small size J , called tree complexity, referring to the number of splits. A larger J can help GBDT capture more complex interactions among variables. However, it will also degrade the generalization ability.

3.2. Relative Importance and Partial Plots of Variables in GBDT. The interpretability is an important advantage of GBDT over other nonparametric methods, such as NN and SVM. GBDT can reveal the internal mechanism of the training model in two ways: rank the influences of independent variables on response predictions and draw the partial dependence functions of the independent variables on response predictions.

For a single CART, one can get the relative importance of the variable X^k by [53]

$$I_k^2(T_m) = \sum_{j=1}^J E_j^2 1_j(X^k), \quad (6)$$

where T_m is the decision tree with J leaf nodes in the m^{th} round iteration. $1_j(X^k)$ is the indicator function indicating whether the specific feature variable X^k is chosen as split variable at node j in the decision tree T_m . E_j^2 is the performance improvement if X^k is selected as the splitting variable at node j . $I_k^2(T_m)$ is the relative importance of the variable X^k in the decision tree T_m . Then, the relative importance of the variable X^k in GBDT can be obtained by

$$I_k^2(F) = \frac{1}{M} \sum_{m=1}^M I_k^2(T_m), \quad (7)$$

where $I_k^2(F)$ is the relative importance of the variable X^k in the final GBDT F .

The partial dependence plot of a subset of variables X_S on the target variable can be obtained by

$$F(X_S) = E_{X_C} [F(X_S)] = \frac{1}{N} \sum_{i=1}^N F(X_S, X_{iC}), \quad (8)$$

where X_C is the complementary set of X_S , $X_S \cup X_C = X_K$ is the set of input variables in the model, and X_{iC} is the value of X_C in training sample ($i = 1, \dots, N$).

4. Data Preparation

4.1. The HighD Dataset. To model and compare the lane change behavior of heavy vehicles and passenger cars, an enriched vehicle trajectory dataset called highD dataset is used in this study. This dataset was provided by Krajewski et al. [54] and was initially used to study autonomous driving. A drone with a 4K-resolution camera was used to shoot the traffic flow at a great height within a freeway section of about 420 meters (shown in Figure 3(a)). Traffic flows in both directions at 6 locations around Cologne in Germany were recorded during 2017 and 2018. One of the most important highlights of this dataset is the amount of data. The highD dataset contains 60 videos in total, and each video lasts 17 minutes. About 90000 passenger cars and 20000 heavy trucks are tracked [54], while the commonly used NGSIM dataset only has 8860 passenger cars and 278 trucks. The highD dataset has been used for car-following analysis by Kurtc [55] and achieved promising results. The speed analysis by Kurtc [55] showed that there was mostly free traffic in highD dataset, but it still contained a considerable dataset showing impeded traffic or even jams with stop-and-go waves because of its large sample size. For detailed information about the highD dataset, one can refer to Krajewski et al. [54]. The highD dataset can be downloaded from <https://www.highd-dataset.com/>.

Among the three locations of highD dataset, the data collected at location 1 are used in this study because of three reasons:

- (1) Simplicity of the section: This section is a basic segment of freeway with three lanes per direction (shown in Figure 3(b)). It is about 2 kilometers from an upstream on-ramp and 2 kilometers from a downstream off-ramp. Thus, more LCs of heavy vehicles exist at location 1, and they can be generally regarded as discretionary lane changes (DLCs) because of the location.
- (2) Diverse traffic conditions: 37 of the 60 recordings were collected at location 1, and they covered both free flow and congestion, which could provide enough data under different traffic conditions for analysis.
- (3) Popular speed limitations: the speed limit at location 1 is 120 km/h, which is very common in other countries and can provide more insightful observations.

It should be pointed out that it is possible that some drivers may preallocate long before the off-ramp. However, according to van Beinum et al. [56], most vehicles preallocate by changing lanes after the exit sign on the side of



FIGURE 3: Schematic diagrams of (a) the covered section and (b) the location of the section.

the motorway. For example, in the Netherlands, the first and second exit signs are normally positioned at about 1200 m and 600 m upstream of an off-ramp, respectively. Thus, vehicles start to preallocate at about 1000 m upstream of an off-ramp, and 600 m is the location where the change in lane flow distribution is almost at its peak [56, 57]. In Germany, the first exiting sign is positioned at about 2000 m upstream of an off-ramp, which is similar to China. According to Zhang et al. [58], 85% of the vehicles preallocate after 1200 m upstream of an off-ramp. Actually, only one vehicle preallocates before 2000 m upstream of an off-ramp. Thus, the preallocations are neglected in this study.

The trajectories of vehicles that only made lane change one time are then extracted from the dataset. However, some vehicles may initiate lane changing when they first enter the segment, and some did not finish lane changing when they left the segment. Such trajectories of vehicles were filtered

out. At last, trajectories of 2905 passenger cars and 433 heavy vehicles were extracted from the dataset.

Previous studies showed that vehicles had different behaviors when they changed to the different lanes (faster lane or slower lane) [37, 39, 49]. Thus, the lane change direction is distinguished in this study. For the rest of this paper, we use LCLL and LCRL to refer to DLC to the left lane (faster lane) and right lane (slower lane), respectively. The distribution of different DLCs for passenger cars and heavy vehicles is shown in Table 1.

In Table 1, lanes 1, 2, and 3 denote the rightmost, middle, and leftmost lanes, respectively. One can find that 1355 passenger cars and 213 heavy vehicles change lanes to the left, while 1550 passenger cars and 220 heavy vehicles change lanes to the right. It can also be found that most heavy vehicles made lane change in the middle and rightmost lanes, while most passenger cars made lane change in the middle and leftmost lanes.

TABLE 1: The numbers of different DLC samples.

| Vehicle type | LC direction | Original lane # | Target lane # | Numbers | Total |
|---------------|--------------|-----------------|---------------|---------|-------|
| Passenger car | LCLL | 1 | 2 | 312 | 1355 |
| | | 2 | 3 | 1043 | |
| | LCRL | 2 | 1 | 452 | 1550 |
| | | 3 | 2 | 1098 | |
| Heavy vehicle | LCLL | 1 | 2 | 198 | 213 |
| | | 2 | 3 | 15 | |
| | LCRL | 2 | 1 | 205 | 220 |
| | | 3 | 2 | 15 | |

4.2. Data Extraction. The lane change implementation period can be divided into several time intervals. At each time interval t , the driver will either choose to continue lane change ($y^t = 0$) or complete it ($y^t = 1$ defined as time interval that the front center of the vehicle crosses the lane line). Figure 4 shows the decision-making process during lane change implementation period. Figure 4 shows the number of time intervals elapsed when the front center of the vehicle crosses the lane line. PL and PF denote the putative leading and following vehicles in the target lane. L and F denote the leading and following vehicles in the original lane. In this study, the time interval is determined as one second, which has also been used in previous studies [10, 34]. The number of observations collected for passenger and heavy vehicles is presented in Table 2.

During the lane change process, the lane change vehicles are influenced by the traffic flows in both the original lane and the target lane, as shown in Figure 5. The main factors affecting LC vehicles' decision making are the speeds, relative speeds, and gaps in both original and target lanes. Weng et al. [10] stated that previous studies considered the above variables separately but ignored the interaction between them. Thus, a surrogate safety measure combining vehicle speeds and space gap, called time-to-collision (TTC), was used by Weng et al. [10]. In this study, the TTC is also considered as a candidate variable, which is defined as

$$\text{TTC} = \frac{x_L - x_F - L}{V_F - V_L}, \quad (9)$$

where x_L and x_F are the longitudinal position coordinates of leading and following vehicles, respectively, V_L and V_F are the speeds of leading and following vehicles, respectively, and L is the length of the leading vehicle.

It should be pointed out that TTC is negative when the following vehicle moves slower than the leading one, which means that the collision will never occur. In addition, when the speed of the following vehicle is equal to or slightly higher than the leading vehicle, TTC will be infinite or too large. In order to restrict these situations, we will set the TTC range to $(0, 100\text{s}]$, that is, when TTC is negative or greater than 100 s, it is configured as 100 s. Another situation is that $x_L - x_F - L$ might be negative before the lane change vehicle encroaches the target lane. In this situation, the subject vehicle cannot complete the lane change maneuver. Thus, the TTC is set to 0.

Previous studies [7, 10, 26, 34] showed that the main factors affecting the decision making of lane change are the

speeds, gaps, relative speeds, TTCs, and vehicle types of the lane changing vehicle and surrounding vehicles. The candidate variables and definitions of influencing factors are shown in Table 3. Previous studies showed that some of these variables share high collinearity [59]. One will make great efforts to deal with the collinearity, which, however, is not the problem in boosting methods because one will try to avoid refocusing on a feature when a specific link between this feature and the outcome has been learned [60].

5. Application and Results

5.1. Parameter Determination. It can be found in Table 2 that the data for the four models are all imbalanced. The observations $y^t = 0$ are about 3 to 4 times the observations of $y^t = 1$. Previous machine learning techniques, such as CART or SVM, are quite sensitive to the balance of data. Thus, under-sampling or over-sampling techniques are needed before modelling. However, over-sampling may lead to over-fitting and under-sampling can certainly cause information loss. Fortunately, the GBDT method can naturally deal with this problem. To reduce the information loss and avoid over-fitting simultaneously, ensemble learning techniques are recommended by He and Garcia [61] because of the random sampling for the training data. The subsample fraction S_{frd} for observations of $y^t = 0$ and $y^t = 1$ will be set to different values to ensure the balance of the training data in each iteration. Specifically, S_{frd} for observations of $y^t = 1$ will be set to 1, and S_{frd} for observations of $y^t = 0$ will be set according to the sample proportion.

Besides S_{frd} , several parameters need to be optimized before developing the final models. A large learning rate will result in over-fitting problem, and 0.001 was recommended in many previous studies as it can generate the final model with lower predictive deviance and a reasonable tree size [62, 63]. Thus, for all four models, the learning rate is fixed at 0.001.

The tree complexity J is another important parameter of GBDT and should be carefully selected because a large J can reflect the unknown interaction information among the independent variables but may lead to over-fitting at the same time [53]. To select the optimal tree complexity J , a series of experiments are conducted by increasing the value of J from 2 to 10. We conduct a fivefold cross-validation procedure, which randomly splits the training data into five equal subsets and use each subset as the test data while the remaining subsets are used to train the model. The negative

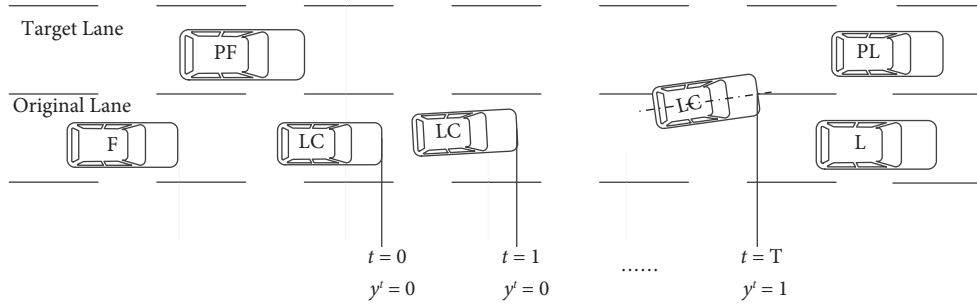


FIGURE 4: Decision making during lane change execution process.

TABLE 2: Sample size for passenger cars and heavy vehicles.

| Vehicle type | LC direction | $y^t = 0$ | $y^t = 1$ | Total |
|---------------|--------------|-----------|-----------|-------|
| Passenger car | LCLL | 735 | 213 | 948 |
| | LCRL | 783 | 220 | 1003 |
| Heavy vehicle | LCLL | 5315 | 1355 | 3960 |
| | LCRL | 5186 | 1550 | 6736 |

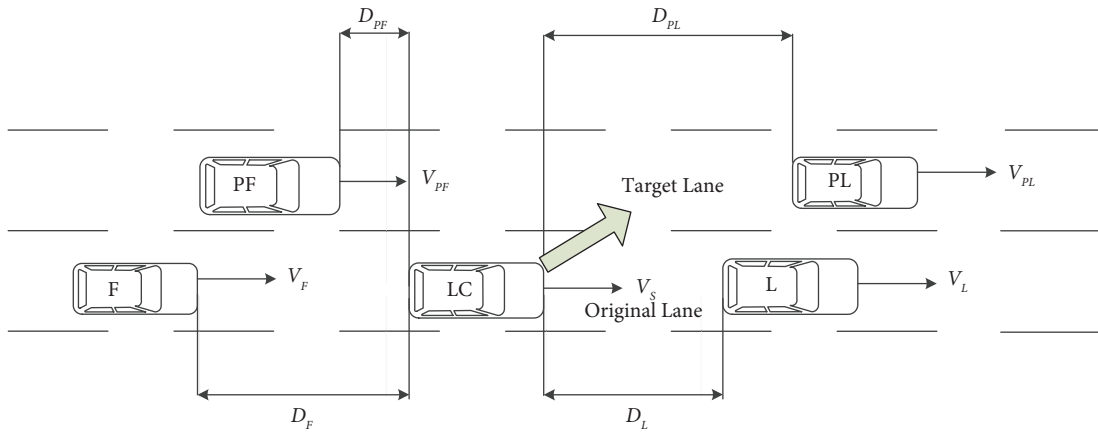


FIGURE 5: Schematic diagram of data extraction.

TABLE 3: Candidate variables and definitions.

| Candidate variables | Descriptions |
|-----------------------|--|
| τ | LC time that has already elapsed |
| V_{LC} (m/s) | Speed of LC vehicle |
| ΔV_{PL} (m/s) | Relative speed between merging vehicle and PL vehicle |
| ΔV_{PF} (m/s) | Relative speed between merging vehicle and PF vehicle |
| ΔV_L (m/s) | Relative speed between LC vehicle and leading vehicle in the original lane |
| ΔV_F (m/s) | Relative speed between LC vehicle and following vehicle in the original lane |
| D (m) | Gap size between PL and PF vehicles |
| D_{PL} (m) | Gap size between merging vehicle and PL vehicle |
| D_{PF} (m) | Gap size between merging vehicle and PF vehicle |
| D_{PL} (m) | Gap size between merging vehicle and leading vehicle in the original lane |
| D_{PF} (m) | Gap size between merging vehicle and following vehicle in the original lane |
| TTC_{PL} | TTC between merging vehicle and PL vehicle |
| TTC_{PF} | TTC between merging vehicle and PF vehicle |
| TTC_L | TTC between merging vehicle and leading vehicle in the original lane |
| TTC_F | TTC between merging vehicle and following vehicle in the original lane |
| δ_{PL} | PL vehicle type (0 for passenger car and 1 for heavy vehicle) |
| δ_{PF} | PF vehicle type (0 for passenger car and 1 for heavy vehicle) |
| δ_L | Leading vehicle type (0 for passenger car and 1 for heavy vehicle) |
| δ_F | Following vehicle type (0 for passenger car and 1 for heavy vehicle) |

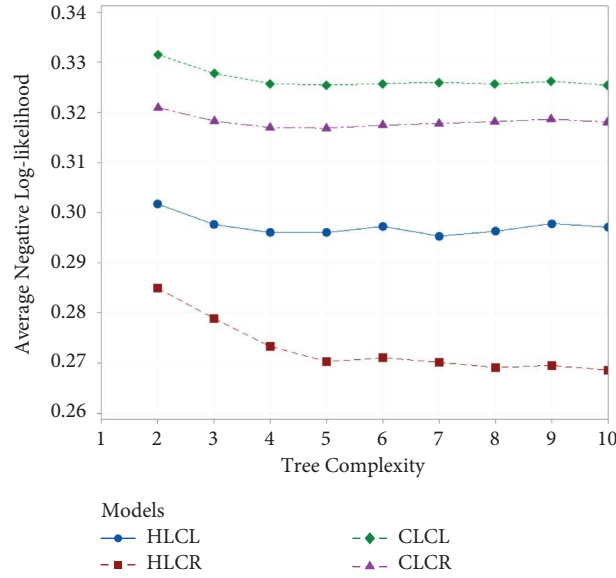


FIGURE 6: Average negative log-likelihood of the four datasets with different tree complexities.

TABLE 4: Optimal parameters for models of passenger cars and heavy vehicles.

| Vehicle type | LC direction | η | S_{frd} | | J | M |
|---------------|--------------|--------|------------------|-----------|-----|------|
| | | | $y^t = 1$ | $y^t = 0$ | | |
| Passenger car | CLCL | 0.001 | 1 | 0.255 | 5 | 3774 |
| | CLCR | 0.001 | 1 | 0.299 | 5 | 4997 |
| Heavy vehicle | HLCL | 0.001 | 1 | 0.290 | 7 | 3093 |
| | HLCR | 0.001 | 1 | 0.281 | 10 | 1814 |

TABLE 5: Prediction accuracy of GBDT and CART models.

| Vehicle type | Direction | y^t | GBDT | | CART | |
|---------------|-----------|------------------------|--------|--------|--------|--------|
| | | | 0 | 1 | 0 | 1 |
| Passenger car | LCLL | Number of observations | 3960 | 1355 | 3960 | 1355 |
| | | Accuracy | 83.21% | 83.32% | 82.95% | 79.93% |
| | | Overall accuracy | 83.24% | | 82.18% | |
| | LCRL | Number of observations | 5186 | 1550 | 5186 | 1550 |
| | | Accuracy | 83.69% | 83.61% | 82.86% | 82.65% |
| | | Overall accuracy | 83.69% | | 82.81% | |
| Heavy vehicle | LCLL | Number of observations | 735 | 213 | 735 | 213 |
| | | Accuracy | 84.90% | 84.04% | 79.73% | 81.22% |
| | | Overall accuracy | 84.70% | | 80.06% | |
| | LCRL | Number of observations | 783 | 220 | 783 | 220 |
| | | Accuracy | 84.29% | 84.09% | 82.25% | 82.73% |
| | | Overall accuracy | 84.25% | | 82.35% | |

average log-likelihood (or cross-entropy) is used to determine the best J :

$$-\text{ave.}LL = \frac{1}{N} \sum_{i=1}^N \ln(1 + e^{-y_i F(X_i)}). \quad (10)$$

Figure 6 describes the negative average log-likelihood of the four models with different tree complexities. It is interesting to find that the best numbers of J for the four models are different. In summary, Table 4 illustrates the optimal combinations of parameters for the lane change

models of passenger cars and heavy vehicles. The numbers of trees M in the final trained models are also presented in Table 4.

5.2. Accuracy of the Model. The GBDT model is compared with CART, which was proposed by Weng et al. [10]. The detailed process can be found in Weng et al. [10]. Table 5 shows the prediction accuracy of GBDT and CART models. One can find that the GBDT method outperforms CART for all the types of lane change, indicating that it has important prospects for GBDT in building lane change models.

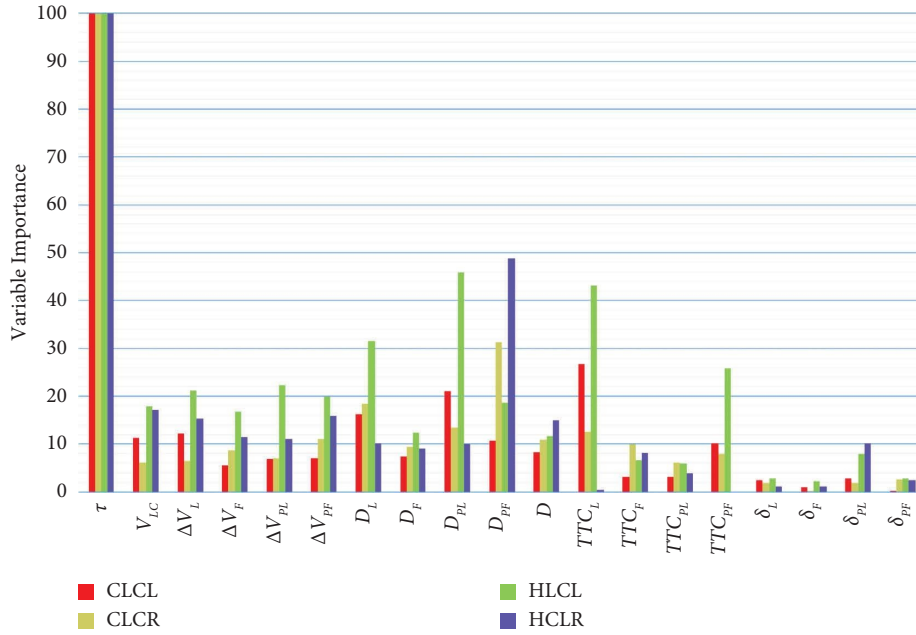


FIGURE 7: Relative importance of candidate variables in GBDT models.

5.3. Results and Discussion. The variable importance can be easily obtained in GBDT. Figure 7 describes the relative importance of the four trained models. Similar to previous studies on merging behavior [10], we can find that the time elapsed τ is the most important variable in all four models. However, the ranks of the relative importance of other variables are significantly different among the four models.

From Figure 7, one can find that the relative importance values of several variables are rather low, such as the vehicle type, indicating some redundant or irrelevant variables in the GBDT models. Therefore, a forward-step-wise feature selection process is applied in this study. Detailed steps can be found in Genuer et al. [64].

Table 6 shows the selected variables for the four models, and some conclusions can be drawn from it:

- (1) It is not surprising to find that τ is still the most important variable in all of these four models. Moreover, 5 to 8 variables remained in the four models.
- (2) All the vehicle type-related variables are dropped in the models of passenger car, which means the vehicle type does not influence the decision of drivers of passenger cars after initiating lane change.
- (3) The ranks of the first to fourth variables do not change after variable selection, and the ranks of other variables also do not change much, indicating the GBDT method's stability.

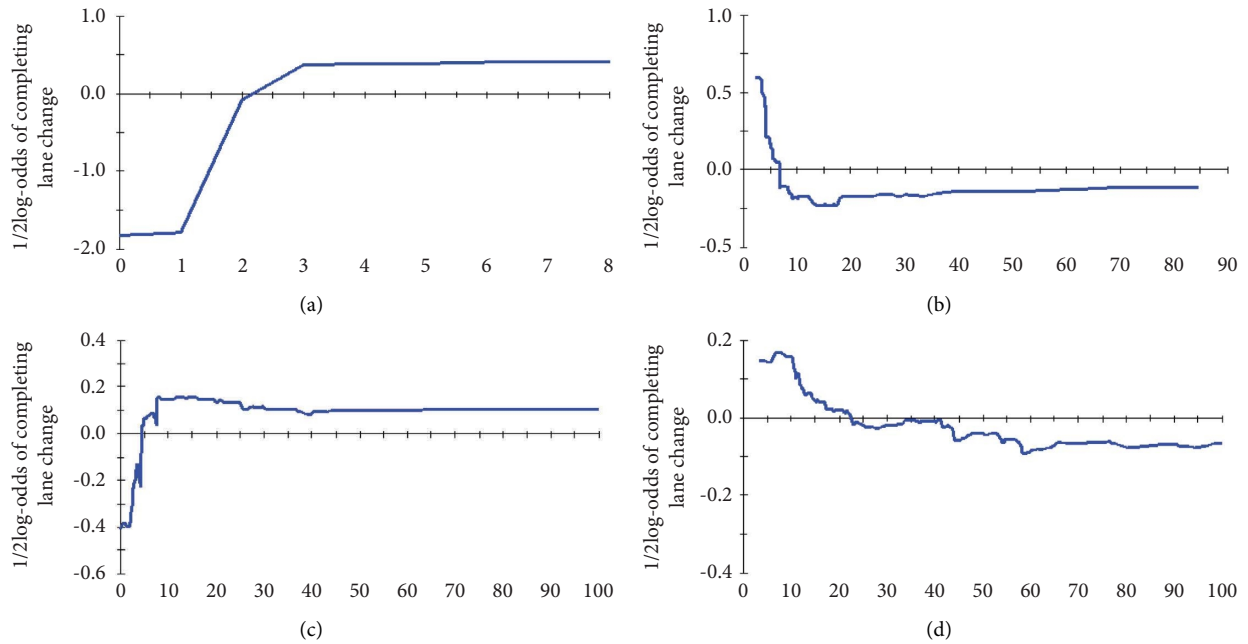
It can be found in Table 5 that TTC_L , D_{PL} , and D_L rank second to fourth for LCLL of passenger cars with the importance value of 29.37, 26.06, and 20.84, while D_{PL} , TTC_L , and D_L rank second to fourth for LCLL of heavy vehicles with the importance value of 55.86, 55.43, and 42.62. It means that when performing lane changes to the left lane,

the considered variables of passenger cars are very similar to those of heavy vehicles. Most variables are related to vehicles in front of the lane change vehicle, which means both passenger cars and heavy vehicles pay more attention to the vehicles in front of them in both the original and the target lane when vehicles change to the left lane. This is probably because when changing to the left lane, the lane change vehicles pursue a speed advantage and are always moving faster than their leading vehicles. Thus, they should pay more attention to vehicles in front of them to avoid crash.

Another interesting finding is that D_{PF} ranks second for LCRL of both passenger and heavy vehicles with importance values of 34.21 and 64.34. It can be observed that the importance values of the variable ranked third for LCRL of both passenger cars and heavy vehicles are much smaller than those of D_{PF} , which means both passenger cars and heavy vehicles pay more attention to the putative following vehicles when they change to the right lane. This is because the traffic flow in the right lane is generally slower than the original lane, and drivers will leave more front space when initiating lane change, which will lead to a small putative following space. The third and fourth important variables for LCRL of passenger cars are D_L and D_{PL} , which means the space gap-related variables are most important for passenger car drivers to perform CLCR. However, this is not the case for heavy vehicles, whose ΔV_F ranks third and fourth. It means that when heavy vehicle drivers execute lane changes to the right lane, they will keep adjusting the speeds as well as the space gaps. Two TTC variables also remained after variable selection for both LCLL and LCRL of passenger cars, which means drivers of passenger cars will face situations that are more complicated when changing to the right lane, and safety is more important than other types of lane changes.

TABLE 6: Selected variables of four models.

| Variables | Passenger cars | | | | Heavy vehicles | | | |
|-----------------------|----------------|------------------|------|------------------|----------------|------------------|------|------------------|
| | LCLL | | LCRL | | LCLL | | LCRL | |
| | Rank | Importance value | Rank | Importance value | Rank | Importance value | Rank | Importance value |
| τ | 1 | 100 | 1 | 100 | 1 | 100 | 1 | 100 |
| V_{LC} (m/s) | — | — | — | — | 5 | 38.75 | 3 | 43.39 |
| ΔV_{PL} (m/s) | 6 | 13.61 | — | — | — | — | 6 | 37.31 |
| ΔV_{PF} (m/s) | 7 | 11.98 | — | — | — | — | 5 | 40.57 |
| ΔV_L (m/s) | 5 | 17.47 | — | — | — | — | — | — |
| ΔV_F (m/s) | — | — | 5 | 14.41 | 6 | 36.20 | 4 | 41.53 |
| D (m) | — | — | — | — | — | — | — | — |
| D_L (m) | 4 | 20.84 | 3 | 21.12 | 4 | 42.62 | — | — |
| D_F (m) | — | — | — | — | — | — | — | — |
| D_{PL} (m) | 3 | 26.06 | 4 | 16.81 | 2 | 55.86 | — | — |
| D_{PF} (m) | 5 | 22.99 | 2 | 34.21 | — | — | 2 | 64.34 |
| TTC_{PL} | 8 | 5.26 | — | — | — | — | — | — |
| TTC_{PF} | — | — | 7 | 10.69 | — | — | — | — |
| TTC_L | 2 | 29.37 | 6 | 14.36 | 3 | 55.43 | — | — |
| TTC_F | — | — | — | — | — | — | — | — |
| δ_{PL} | 9 | 4.27 | — | — | — | — | 7 | 16.53 |
| δ_{PF} | — | — | — | — | — | — | 8 | 4.35 |
| δ_L | — | — | — | — | — | — | — | — |
| δ_F | — | — | — | — | — | — | — | — |

FIGURE 8: Partial plots of (a) τ , (b) TTC_L , (c) D_{PL} , and (d) D_L for LCLL of passenger car.

Comparing the lane changes of passenger cars and heavy vehicles, one can find that the importance values of variables except τ for heavy vehicles are much larger than those for passenger cars. For example, the importance values of the second important variables for LCLL and LCRL of passenger cars are 29.37 and 34.21, respectively, while they are 55.86 and 64.34 for heavy vehicles, respectively. It means the drivers of heavy vehicles pay more attention to the variables related to surrounding traffic. This is probably because of heavy trucks' large size and poor maneuverability. As we all

know, heavy vehicles' acceleration and deceleration performance is much worse than passenger cars. The size of the heavy vehicle is much larger, which makes the drivers more cautious about performing lane changes and naturally results in more attention to their surrounding traffic.

Table 5 shows that different variables are considered in four models, indicating that the decision behavior during the lane change implementation period varies between vehicle types and lane change directions. Compared to passenger

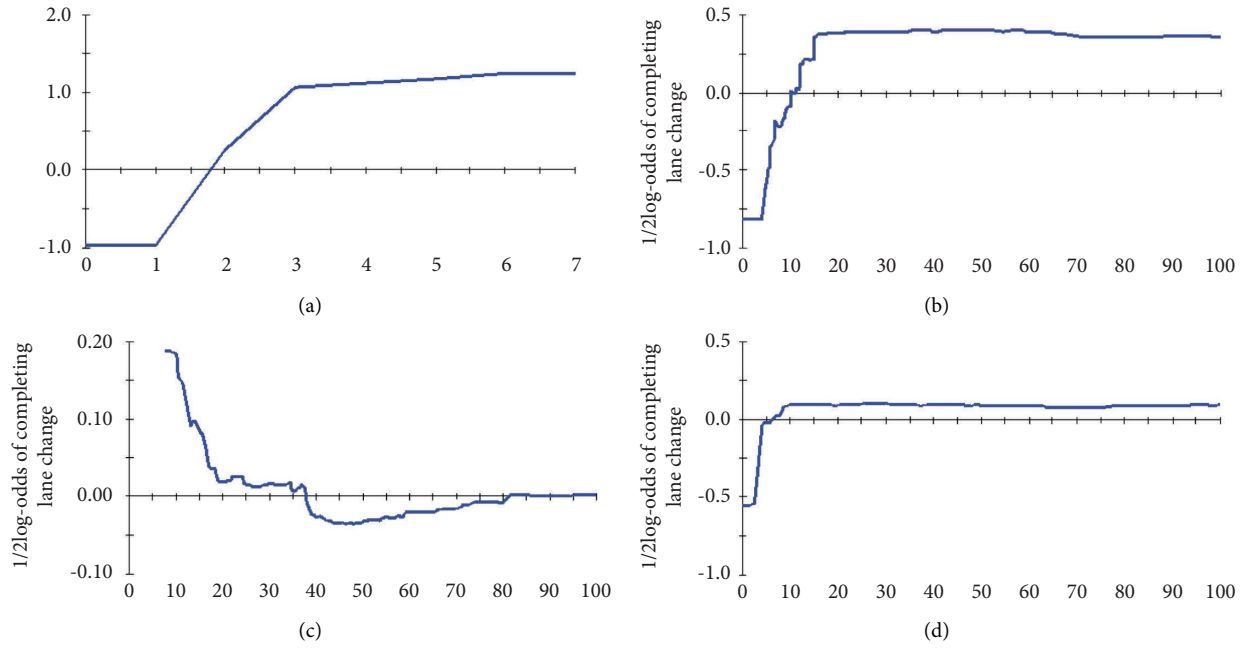


FIGURE 9: Partial plots of (a) τ , (b) D_{PF} , (c) D_L , and (d) D_{PL} for LCRL of passenger car.

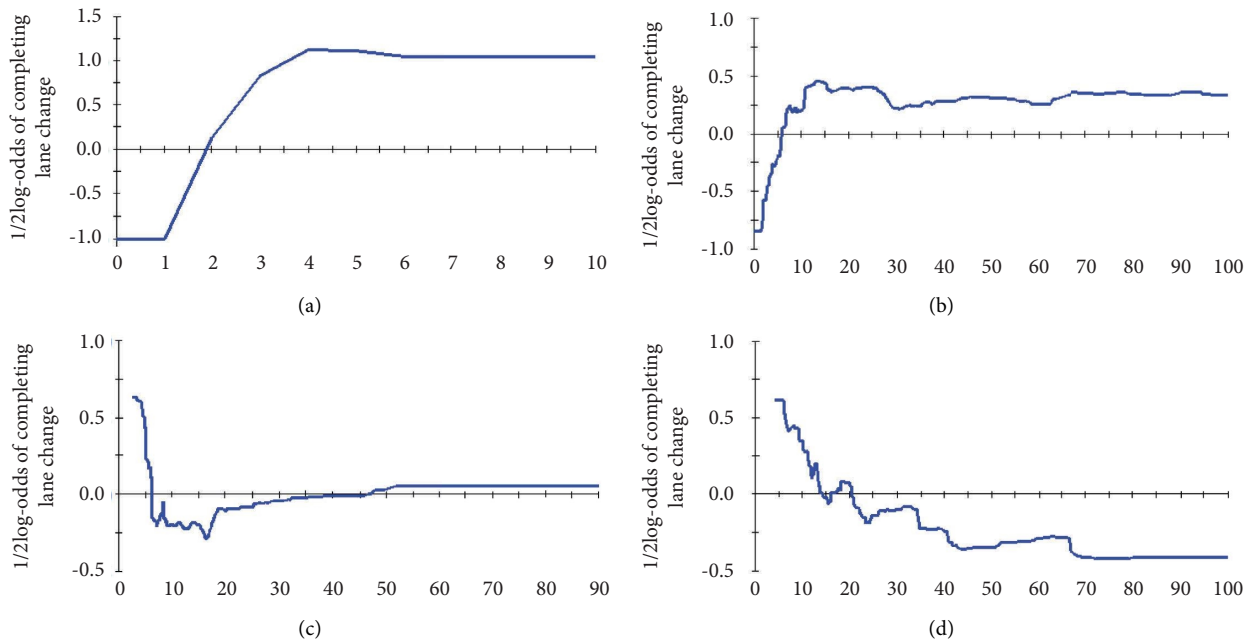


FIGURE 10: Partial plots of (a) τ , (b) D_{PL} , (c) TTC_L , and (d) D_L for LCLL of heavy vehicle.

cars, heavy vehicle drivers pay more attention to their surrounding traffic characteristics.

GBDT can easily explain the complicated relationships between the independent and dependent variables by drawing the partial dependence plots, which can be used for sensitivity analysis. The partial dependence plots of the first four variables in four models are provided in Figures 8–11. It should be pointed out that the Y-axis in the partial dependence plots in this study refers to the average 1/2 log odds of completing lane change.

As shown in Figures 8–11, it can be found that the relationship between independent variables and the output of GBDT models is highly nonlinear. Considering that τ has the most significant influence on decision behavior, we will discuss the effects of τ in this section for demonstration. The effects of other variables can be easily observed from Figures 8–11. For passenger cars, the probability of successful lane change significantly increases when τ increases from 1 second to 3 seconds, which means most passenger cars complete lane change

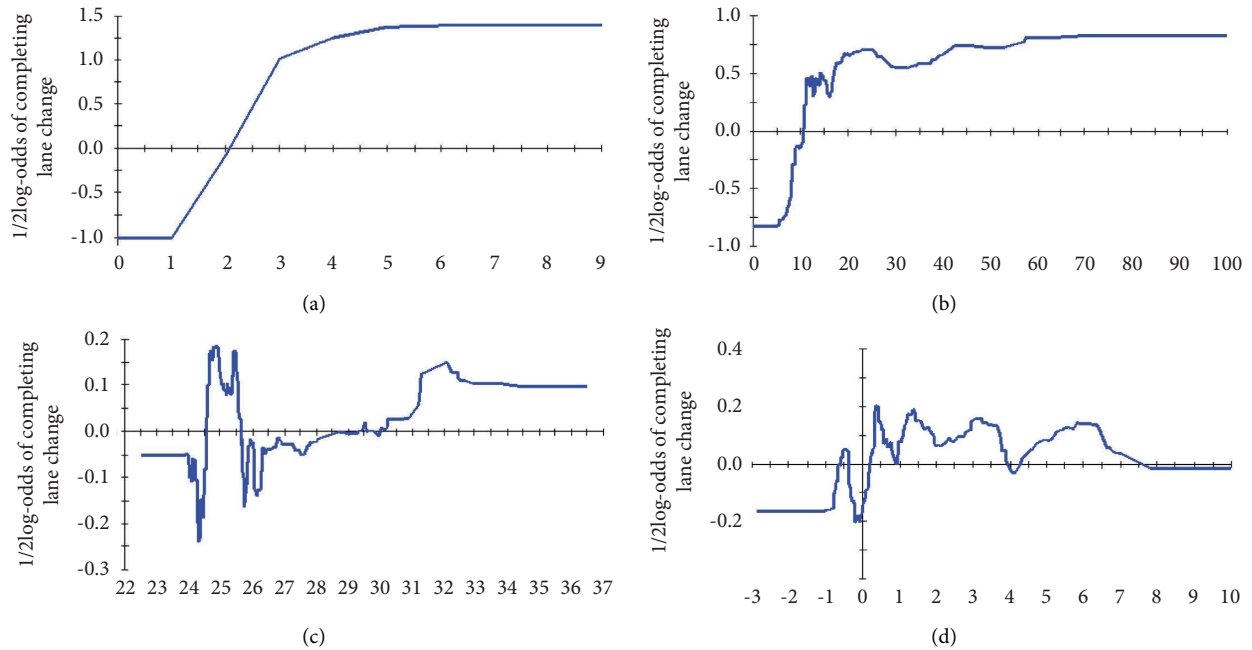


FIGURE 11: Partial plots of (a) τ , (b) D_{PF} , (c) V_{LC} , and (d) ΔV_{PF} for LCRL of heavy vehicle.

within 3 seconds after initiating lane change. However, the effective interval of τ is larger for heavy vehicles, indicating that heavy vehicles will take a longer time to complete lane change.

It can also be found from the partial dependence plots that the drivers of heavy vehicles are easier to be affected by surrounding traffic characteristics. The partial dependence plots show that the influence of independent variables on lane change decisions is nonlinear and complicated. For example, the partial plots of V_{LC} for both LCLL and LCRL show high volatility. Different speeds indicate different traffic conditions, which result in different lane change behavior. One can also find that the effect of V_{LC} on lane change decision is quite different between LCLL and LCRL in Figures 10 and 11. It means different rules are used by drivers when changing to different lanes.

The above results show that the lane change behavior significantly varies with the vehicle types and lane change directions. Specific models for different vehicle types and lane change directions are needed for microscopic traffic simulators. Besides, the results also indicate that lane change rules should be carefully designed for ADAS and CAV of different vehicle types.

6. Conclusions

This study employs GBDT to build the decision model during the lane change implementation. This paper uses vehicle trajectory data collected on German Highways to validate the GBDT method. Unlike previous studies, this study modelled lane changes of different vehicle types and directions separately.

The results showed that the elapsed time τ is the most important variable in all models. Other variables are

found to play different roles in different models. Partial dependence plots of GBDT are drawn to reflect the effects of variables on lane change decisions. It is shown that the influence of independent variables on decision behaviors is nonlinear and complicated, which means that the same variable has different effects on the lane change decision for different vehicle types and lane change directions. Compared with other state-of-the-art methods, GBDT can produce more accurate prediction results, making it a promising tool for autonomous driving.

The results of this study indicate that it is necessary for microscopic traffic simulators, ADASs, and CAVs to build specific lane change models for different vehicle types and lane change directions. Nonetheless, there are also some limitations in this study. Firstly, the driving environment type is very limited since all the data are collected on highways. The applicability of the proposed method of lane change behaviors on urban roads needs to be tested. Secondly, driver characteristics are not considered in this study. Furthermore, more data will be collected to test the generalization of the model built in this study.

Data Availability

Some or all data, models, and codes that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study was supported by the Natural Science Fund for Colleges and Universities in Jiangsu Province under grant no. 21KJB580014, Science and Technology Innovation Fund for Youth Scientists of Nanjing Forestry University under grant no. CX2019021, Nantong Science and Technology Plan Project under grant no. JC2020142, and Large Instruments Open Foundation of Nantong University under grant no. KFJN2260. The authors would like to thank Robert Krajewski and Julian Bock (RWTH Aachen University, Aachen, Germany) for providing the highD trajectory data for this study and experimenter Siyuan Liu in the Analysis and Test Center of Nantong University.

References

- [1] Q. Chen, H. Huang, Y. Li et al., "Modeling accident risks in different lane-changing behavioral patterns," *Analytic methods in accident research*, vol. 30, Article ID 100159, 2021.
- [2] T. Chen, X. Shi, and Y. D. Wong, "A lane-changing risk profile analysis method based on time-series clustering," *Physica A: Statistical Mechanics and Its Applications*, vol. 565, Article ID 125567, 2021.
- [3] G. Li, S. Fang, J. Ma, and J. Cheng, "Modeling merging acceleration and deceleration behavior based on gradient-boosting decision tree," *Journal of Transportation Engineering, Part A: Systems*, vol. 146, no. 7, Article ID 5020005, 2020.
- [4] M. Yang, X. Wang, and M. J. T. Quddus, "Examining lane change gap acceptance, duration and impact using naturalistic driving data," *Transportation Research Part C: Emerging Technologies*, vol. 104, pp. 317–331, 2019.
- [5] Y. Yang, Z. Yuan, and R. Meng, "Exploring traffic crash occurrence mechanism toward cross-area freeways via an improved data mining approach," *Journal of Transportation Engineering, Part A: Systems*, vol. 148, no. 9, Article ID 4022052, 2022.
- [6] G. Li, "Application of finite mixture of logistic regression for heterogeneous merging behavior analysis," *Journal of Advanced Transportation*, vol. 2018, Article ID 1436521, 9 pages, 2018.
- [7] M. Li, Z. Li, C. Xu, and T. Liu, "Short-term prediction of safety and operation impacts of lane changes in oscillations with empirical vehicle trajectories," *Accident Analysis and Prevention*, vol. 135, Article ID 105345, 2020.
- [8] S. Moridpour, G. Rose, and M. Sarvi, "Effect of surrounding traffic characteristics on lane changing behavior," *Journal of Transportation Engineering*, vol. 136, no. 11, pp. 973–985, 2010.
- [9] Z. Zheng, "Recent developments and research needs in modeling lane changing," *Transportation Research Part B: Methodological*, vol. 60, pp. 16–32, 2014.
- [10] J. Weng, S. Xue, and X. Yan, "Modeling vehicle merging behavior in work zone merging areas during the merging implementation period," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 4, pp. 917–925, 2016.
- [11] Fmcsa, *Large Truck and Bus Crash Facts 2016*, US Department of Transportation, Washington, DC, USA, 2016.
- [12] C. Qi, A. Fourie, and X. Zhao, "Back-analysis method for slope displacements using gradient-boosted regression tree and firefly algorithm," *Journal of Computing in Civil Engineering*, vol. 32, no. 5, Article ID 4018031, 2018.
- [13] X. Wang, C. Shao, C. Yin, and L. Guan, "Disentangling the comparative roles of multilevel built environment on body mass index: evidence from China," *Cities*, vol. 110, Article ID 103048, 2021.
- [14] C. Yin and C. Shao, "Revisiting commuting, built environment and happiness: new evidence on a nonlinear relationship," *Transportation Research Part D: Transport and Environment*, vol. 100, Article ID 103043, 2021.
- [15] J. Zhou, X. Li, and H. S. Mitri, "Classification of rockburst in underground projects: comparison of ten supervised learning methods," *Journal of Computing in Civil Engineering*, vol. 30, no. 5, Article ID 4016003, 2016.
- [16] M. Kearns, "Thoughts on hypothesis boosting," *Machine Learning Class Project*, vol. 45, p. 105, 1988.
- [17] R. E. Schapire, "The strength of weak learnability," *Machine Learning*, vol. 5, no. 2, pp. 197–227, 1990.
- [18] Y. Zhang and A. Haghani, "A gradient boosting method to improve travel time prediction," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 308–324, 2015.
- [19] P. G. Gipps, "A model for the structure of lane-changing decisions," *Transportation Research Part B: Methodological*, vol. 20, no. 5, pp. 403–414, 1986.
- [20] L. Bloomberg and J. Dale, "A comparison of the VISSIM and CORSIM traffic simulation models," in *Proceedings of the Institute of Transportation Engineers Annual Meeting*, pp. 52–60, St. Louis, MO, USA, August 2011.
- [21] P. Hidas, "Modelling lane changing and merging in microscopic traffic simulation," *Transportation Research Part C: Emerging Technologies*, vol. 10, no. 5–6, pp. 351–371, 2002.
- [22] P. Hidas, "Modelling vehicle interactions in microscopic simulation of merging and weaving," *Transportation Research Part C: Emerging Technologies*, vol. 13, no. 1, pp. 37–62, 2005.
- [23] Q. Yang and H. N. Koutsopoulos, "A microscopic traffic simulator for evaluation of dynamic traffic management systems," *Transportation Research Part C: Emerging Technologies*, vol. 4, no. 3, pp. 113–129, 1996.
- [24] H. Kita, "A merging-giveway interaction model of cars in a merging section: a game theoretic analysis," *Transportation Research Part A: Policy and Practice*, vol. 33, no. 3–4, pp. 305–312, 1999.
- [25] C. F. Choudhury, M. E. Ben-Akiva, T. Toledo, G. Lee, and A. Rao, "Modeling cooperative lane changing and forced merging behavior," in *Proceedings of the 86th Annual Meeting of the Transportation Research Board*, Washington, DC, USA, January 2007.
- [26] G. Li and J. Cheng, "Exploring the effects of traffic density on merging behavior," *IEEE Access*, vol. 7, pp. 51608–51619, 2019.
- [27] D. J. Sun and L. Elefteriadou, "A driver behavior-based lane-changing model for urban arterial streets," *Transportation Science*, vol. 48, no. 2, pp. 184–205, 2014.
- [28] D. J. Sun and L. Elefteriadou, "Lane-changing behavior on urban streets: a focus group-based study," *Applied Ergonomics*, vol. 42, no. 5, pp. 682–691, 2011.
- [29] H. Kita, "Effects of merging lane length on the merging behavior at expressway on-ramps," in *Proceedings of the 12th International Symposium on the Theory of Traffic Flow and Transportation*, pp. 37–51, Berkeley, CA, USA, July 1993.
- [30] F. Marczak, W. Daamen, and C. Buisson, "Merging behaviour: empirical comparison between two sites and new theory development," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 530–546, 2013.
- [31] J. Weng and Q. Meng, "Modeling speed-flow relationship and merging behavior in work zone merging areas,"

- Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 985–996, 2011.
- [32] H. Zhou, Y. Sun, X. Qin, X. Xu, and R. Yao, “Modeling discretionary lane-changing behavior on urban streets considering drivers’ heterogeneity,” *Transportation letters*, vol. 12, no. 3, pp. 213–222, 2020.
- [33] E. Balal, R. L. Cheu, and T. Sarkodie-Gyan, “A binary decision model for discretionary lane changing move based on fuzzy inference system,” *Transportation Research Part C: Emerging Technologies*, vol. 67, no. 2, pp. 47–61, 2016.
- [34] Y. Hou, P. Edara, and C. Sun, “Modeling mandatory lane changing using Bayes classifier and decision trees,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 2, pp. 647–655, 2014.
- [35] Q. Meng and J. Weng, “Classification and regression tree approach for predicting drivers’ merging behavior in short-term work zone merging areas,” *Journal of Transportation Engineering*, vol. 138, no. 8, pp. 1062–1070, 2012.
- [36] D.-F. Xie, Z.-Z. Fang, B. Jia, and Z. J. T. He, “A data-driven lane-changing model based on deep learning,” *Transportation Research Part C: Emerging Technologies*, vol. 106, pp. 41–60, 2019.
- [37] S. Moridpour, M. Sarvi, G. Rose, and E. Mazloumi, “Lane-changing decision model for heavy vehicle drivers,” *Journal of Intelligent Transportation Systems*, vol. 16, no. 1, pp. 24–35, 2012.
- [38] M. Rahman, M. Chowdhury, Y. Xie, and Y. He, “Review of microscopic lane-changing models and future research opportunities,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 4, pp. 1942–1956, 2013.
- [39] S. Moridpour, M. Sarvi, and G. Rose, “Modeling the lane-changing execution of multiclass vehicles under heavy traffic conditions,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2161, no. 1, pp. 11–19, 2010.
- [40] M. Sarvi and M. Kuwahara, “Microsimulation of freeway ramp merging processes under congested traffic conditions,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 3, pp. 470–479, 2007.
- [41] T. Toledo, H. N. Koutsopoulos, and M. Ben-Akiva, “Integrated driving behavior modeling,” *Transportation Research Part C: Emerging Technologies*, vol. 15, no. 2, pp. 96–112, 2007.
- [42] X. Wan, P. J. Jin, F. Yang, J. Zhang, and B. Ran, “Modeling vehicle interactions during merge in congested weaving section of freeway ramp,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2421, no. 1, pp. 82–92, 2014.
- [43] X. Wan, P. J. Jin, H. Gu, X. Chen, and B. Ran, “Modeling freeway merging in a weaving section as a sequential decision-making process,” *Journal of Transportation Engineering, Part A: Systems*, vol. 143, no. 5, Article ID 5017002, 2017.
- [44] Q. Wang, Z. Li, and L. Li, “Investigation of discretionary lane-change characteristics using next-generation simulation data sets,” *Journal of Intelligent Transportation Systems*, vol. 18, no. 3, pp. 246–253, 2014.
- [45] K. I. Ahmed, *Modeling Drivers’ Acceleration and Lane Changing Behavior*, Massachusetts Institute of Technology, Massachusetts, MA, USA, 1999.
- [46] N. Oliver and A. P. Pentland, “Graphical models for driver behavior recognition in a smartcar,” in *Proceedings of the IEEE Intelligent Vehicles Symposium*, pp. 7–12, IEEE, Dearborn, MI, USA, October 2000.
- [47] Q. Meng and J. J. T. Weng, “An improved cellular automata model for heterogeneous work zone traffic,” *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 1263–1275, 2011.
- [48] S. Moridpour, E. Mazloumi, and M. Mesbah, “Impact of heavy vehicles on surrounding traffic characteristics,” *Journal of Advanced Transportation*, vol. 49, no. 4, pp. 535–552, 2015.
- [49] M. Park, K. Jang, J. Lee, and H. Yeo, “Logistic regression model for discretionary lane changing under congested traffic,” *Transportmetrica: Transport Science*, vol. 11, no. 4, pp. 333–344, 2015.
- [50] X. Ma, C. Ding, S. Luan, Y. Wang, and Y. Wang, “Prioritizing influential factors for freeway incident clearance time prediction using the gradient boosting decision trees method,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 9, pp. 2303–2310, 2017.
- [51] X. Wang, C. Yin, J. Zhang, C. Shao, and S. Wang, “Nonlinear effects of residential and workplace built environment on car dependence,” *Journal of Transport Geography*, vol. 96, Article ID 103207, 2021.
- [52] F. Zhang, X. Zhu, T. Hu, W. Guo, C. Chen, and L. Liu, “Urban link travel time prediction based on a gradient boosting method considering spatiotemporal correlations,” *ISPRS International Journal of Geo-Information*, vol. 5, no. 11, p. 201, 2016.
- [53] J. H. Friedman, “Greedy function approximation: a gradient boosting machine,” *Annals of Statistics*, pp. 1189–1232, 2001.
- [54] R. Krajewski, J. Bock, L. Kloeker, and L. Eckstein, “The highd dataset: a drone dataset of naturalistic vehicle trajectories on German highways for validation of highly automated driving systems,” in *Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2118–2125, IEEE, Maui, HI, USA, November 2018.
- [55] V. Kurtc, “Studying car-following dynamics on the basis of the highd dataset,” *Transportation Research Record*, vol. 2674, 2020.
- [56] A. van Beinum, H. Farah, F. Wegman, and S. Hoogendoorn, “Driving behaviour at motorway ramps and weaving segments based on empirical trajectory data,” *Transportation Research Part C: Emerging Technologies*, vol. 92, pp. 426–441, 2018.
- [57] A. van Beinum, M. Hovenga, V. Knoop, H. Farah, F. Wegman, and S. Hoogendoorn, “Macroscopic traffic flow changes around ramps,” *Transportmetrica: Transport Science*, vol. 14, no. 7, pp. 598–614, 2018b.
- [58] L. Zhang, C. Chen, J. Zhang, S. Fang, J. You, and J. Guo, “Modeling lane-changing behavior in freeway off-ramp areas from the shanghai naturalistic driving study,” *Journal of Advanced Transportation*, vol. 2018, Article ID 8645709, 10 pages, 2018.
- [59] E. Balal, R. L. Cheu, T. Gyan-Sarkodie, and J. Miramontes, “Analysis of discretionary lane changing parameters on freeways,” *International Journal of Transportation Science and Technology*, vol. 3, no. 3, pp. 277–296, 2014.
- [60] T. Chen and C. Guestrin, “Xgboost: a scalable tree boosting system,” in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, San Francisco, CA, USA, August 2016.

- [61] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.
- [62] C. Ding, X. J. Cao, and P. Næss, "Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo," *Transportation Research Part A: Policy and Practice*, vol. 110, pp. 107–117, 2018.
- [63] S. Touzani, J. Granderson, and S. Fernandes, "Gradient boosting machine for modeling the energy consumption of commercial buildings," *Energy and Buildings*, vol. 158, pp. 1533–1543, 2018.
- [64] R. Genuer, J.-M. Poggi, and C. J. P. Tuleau-Malot, "Variable selection using random forests," *Pattern Recognition Letters*, vol. 31, no. 14, pp. 2225–2236, 2010.