A causal inferencebased speed control framework for discretionary lane-changing processes

Zhen Zhou[1], Yi Zhao[2], Minghao Li[3] and Yuyang Bao[4]

[1]Graduate Student, College of Automotive and Traffic Engineering, Nanjing Forestry University, Nanjing, P.R. China, 210037; Department of Statistics, University of Illinois Urbana-Champaign, Champaign, Illinois, The United States, Email: zhouzhen20210601@163.com

[2]Assistant Professor, College of Automotive and Traffic Engineering, Nanjing Forestry University, Nanjing, P.R. China, 210037 (corresponding author), Email: zhaoyi207@njfu.edu.cn

[3]Graduate Student, College of Automotive and Traffic Engineering, Nanjing Forestry University, Nanjing, P.R. China, 210037, Email: limh@njfu.edu.cn

[4]Graduate Student, College of Science, Wuhan University of Technology, Wuhan, P.R. China, 430070, Email: byy123456@whut.edu.cn

ABSTRACT

The study aims to optimize the speed of discretionary lane-changing behavior. Combining the safety, efficiency and eco-friendly requirements, a causal inference based speed control framework is proposed. First, variables affecting the lane change efficiency are selected and their treatment effects are obtained by conducting causal effect analysis and causal refutation analysis. Second, the conditional average treatment effect is estimated by taking the safety and eco-friendly requirements into consideration and applying the double machine learning version causal forest model. Finally, two controllers are built, and grid search is used to optimize lane change decision making. The results of a numerical experiment show that the framework has a good performance on improving the efficiency and environmental friendliness of a lane change process while ensuring safety. The proposed framework provides a new approach to speed control for discretionary lane-changing behavior. The approach can be well adapted to a vehicle-to-vehicle communication environment and is expected to be applied to vehicle-assisted driving systems in the future.

Keywords: Causal inference, Lane change behavior, Treatment effect, Speed control framework

INTRODUCTION

Vehicle lane changing is the main traffic behavior that triggers traffic flow fluctuations and traffic accidents.Researchers have carried out extensive work on vehicle lane changing behavior, including the study of lane change behavior characteristics in terms of acceptable lane change gap, lane change duration (M. Yang et al., 2019) and lane change probability (Ahmed et al., 2019), the study of lane change behavior modeling based on game theory (Arbis & Dixit, 2019) and deep learning (Zhang et al., 2019), the study of lane change behavior for safety risk assessment and lane change warning (M. Li et al., 2020), and the study of lane change trajectory planning and lane change behavior impact. With the development of cooperative vehicle infrastructure systems (CVISs) and driverless technologies, the characteristics, decisions and impacts of lane-changing behaviors in mixed traffic flow or driverless environments have gradually become a hot research topic.

A connected environment provides information via vehicle-to vehicle and vehicle-to-infrastructure communications. Yasir Ali et al. (Y. Ali et al., 2020) expressed their understanding of discretionary lane-changing behavior in a connected environment, and found that a connected environment has a significant impact on spacing, speed, acceleration noise, and other factors. Lane change models can be classified into three types. The first type is dynamical models (P. Cao et al., 2017; Lin et al., 2019; Luo et al., 2016; Mai et al., 2016). The second type is game theory based lane change models (Y. Ali et al., 2019; Talebpour et al., 2015; Yu et al., 2018). The third type is machine learning based lane change models (Sun et al., 2021; Xie et al., 2019; Zhang et al., 2019). A connected environment has shown promise in addressing the large-scale traffic difficulties related to safety, efficiency, and environmental impact. However, a fully driverless environment or a high degree of CVIS is still a long from before full-scale promotion. Prior to this, assisted driving will be the best way to take full advantage of the traffic information provided by a connected environment. This paper will focus on lane change decision-making for vehicle assisted-driving. The interaction between the lane-changing vehicle and the surrounding vehicle information provides rich information on traffic and vehicle control. In the process of a vehicle lane change, how can the influence relationship between vehicles and accurate surrounding vehicle information provided by a connected environment to be used? How can safety, efficiency, and environmental impact be integrated for the scientific control of lane-changing vehicles? These two problems are the subject of this paper.

There is cause and effect everywhere in transportation science; for example, during the lane change process in the connected environment, the velocity, acceleration, and distance of the surrounding vehicles all interact with each other as cause and effect. Most of the existing models based on deep networks, game theory and accident data do not deeply investigate the intrinsic mechanism of the lane-changing process, nor do they consider the interaction of the variables of lane-changing. Hence, we propose that causal inference can be introduced into lane-changing decision-making.

Causal inference, which is different from correlation analysis, can be used to measure causality, specifically using causal inference o infer whether a change in one event is the cause of a change in other events (Neuberg, 2003). Basic causal inference consists of two steps, identification and estimation; hence, causal inference is often used to discover and quantify the degree of influence of an event. The concept of causal effect was originally introduced by Neyman in 1923 (Ding, 2017), and the causal inference framework was first proposed by Paul W. Holland (Holland, 1986), which is known as the Rubin causal model (RCM). Another popular causal model is the causal diagram model, which was proposed by Pearl (Neuberg, 2003). A causal diagram model can be divided into three steps: causal discovery, causal effect and refuted estimate. It can be proven that a causal diagram model is equivalent to an RCM (Hitchcock, 2001). Due to the ease of interpretation of a causal diagram model, it is more often used in other disciplines, such as social science, computer science, and a range of data-related disciplines.

Causal inference is widely applied in transportation. For example, Li et al. (C. Li et al., 2020) converted risk identification to a causal effect task and “present a novel two-stage risk object identification framework based on causal inference with the proposed object-level manipulable driving model”. You and Han (You Tackgeun and Han, 2020) use causal inference to compute causality in the different frames of traffic accident videos. Cao et al. (D. Cao et al., 2021) proposed a causal-based framework to determine impacted regions and quantify speed reduction. Ali et al. (Q. Ali et al., 2019) use a causal inference model to mine and analyze the causes of traffic safety accidents, including urban road class, urban characteristics, population density and education level. Yang et al. (S. Yang et al., 2021) identified different street impact coefficients on a road system and calculated the causal interaction effects between different streets. Molavipour et al. (Molavipour et al., 2020) explore the interaction of traffic flows at different nodes in a traffic network based entirely on a theoretical proof and aim to construct cause-and-effect diagrams for traffic flow. However, existing traffic causal inference models are only used to quantify, identify and analyze the causal effects of particular road events. Few studies apply causal inference to the decision-making process, and there are also few studies on causal inference in lane change scenarios. In this paper, to study and optimize the mechanism of the lane change process, causal inference is added to the decision-making in a lane change scenario.

As an important method for discovering causality between events, a causal inference-based speed control system framework is proposed in this paper to improve the process of discretionary lane-changing. This framework can improve the efficiency of lane change by optimizing the velocity and acceleration of a lane change to ensure the safety and eco-friendliness of the lane change. This framework is expected to be applied to vehicle-assisted driving systems. This paper can provide a causal inference-based model reference for future vehicle-to-vehicle infrastructure communication.

The rest of the article is structured as follows. The METHODS proposes a causal inference based lane change decision framework and introduces the intrinsic mechanism. The DATA DESCRIPTION decribes the basic information of the data, and correlation analysis is conducted. The CASE STUDY shows the results of the causal effect analysis, conditional average treatment effect analysis and lane change decision-making system. Finally, the CONCLUSION summarizes the conclusions and achievements of this study and proposes directions for future research.

METHODS

A lane change decision framework that includes three sections is proposed in this paper. These three sections are summarized as the estimation of the treatment effect, the estimation of the conditional average treatment effect and lane change decision-making. **Figure 1** gives the basic framework of the causal inference-based lane change model. The three sections are summarized below.

(1) The first section investigates the effect of variables during the lane change on the duration events of the lane change. Candidate variables that may affect the duration of the lane change are first selected, and the treatment effects of the candidate variables are estimated. Then by using a placebo refutation test, the final variables are selected and their treatment effects can be obtained.

(2) In the second section, the safety variable time to collision (TTC) and the eco-friendly variable acceleration are set as the condition variables, the velocity is set as the control variable, and the lane change duration is set as the outcome variable. By using a double machine learning version causal forest model, conditional average treatment effect (CATE) can be estimated.

(3) The lane change decision is made in the third section. The CATE of the condition variables that have been estimated in the second section are used as input to build a velocity and acceleration optimizer. The framework uses a grid search on the hyperplane formed by the CATE to obtain the optimized suggested velocity and acceleration.

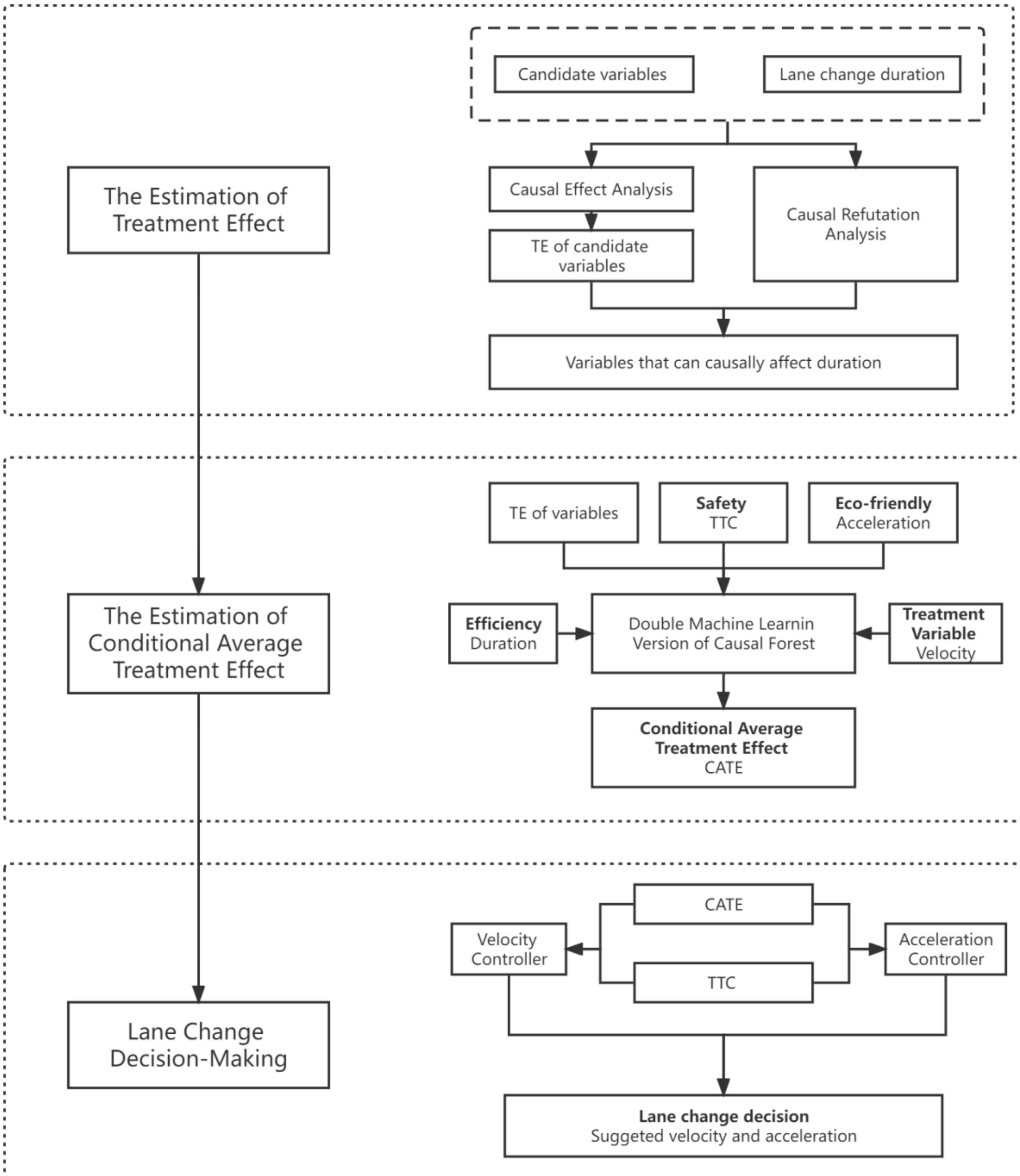

Figure 1 Framework for lane change based causal inference

Estimation of treatment effects and refutation

According to the definition, causal inference can help discover the causality between variables, and it can also enhance the interpretability of a model. The treatment effect is an important tool that can quantify the causal effects between variables, and it can be estimated from treatment variables and outcome variables.

According to causal inference, if a treatment variable is a binary variable, the average treatment effect is

$$E[E(Y|W,T=1)]-E[E(Y|W,T=0)] \tag{1}$$

where T is treatment variable chosen from variable set W and Y is the outcome variable. In this study, the treatment variables are the velocity and acceleration, while the outcome variable is the lane change duration. According to **Equation (1)**, we can estimate $E(Y|W,T)$ using a regression model, which can be denoted as $E(Y|W,T)=\beta_0+\beta_t T+\beta_w W$. The average treatment effect is β_t.

To test the validation of the estimator, a placebo is added to refit the outcome model. We can calculate new effects and compare them with the estimated effects. By definition, "a placebo is any treatment that has no active properties" (Marchant, 2016), if we replce our treatment variables with a placebo, new treatment effects with different distributions can result. By using this property, we can easily determine whether the original variable has a treatment effect.

We can obtain guidelines for determining whether there is a causal effect between the treatment variable and the outcome variable. First, the estimated effect should not be zero, and the value of estimated effect is proportional to the causal effect of the treatment variable on the outcome variable. Second, the placebo effect should go to 0. For box plots, the average value must not be 0, and the data in the box plot need to be significantly different from the estimated value. If the above two requirements are met, it can be concluded that there is a causal effect between the treatment variable and the outcome variable.

Estimation of the conditional average treatment effect

The conditional average treatment effect (CATE) is a causal tool that can be applied to detect heterogeneity between variables. The input variables of the estimation of conditional average treatment effect are the treatment variable, outcome variable and condition variables, which are the velocity, lane change duration and other variables that can affect the channel change, respectively.

We first assume that the outcome variable and treatment variables are continuous variables, and a tree model is applied to obtain the causal effect of the treatment variables. The conditional regression function can be denoted as $\mu_t(X) = E[Y(t)|X]$, where $\hat{\mu}_t(X) = \mu_t(X, \widehat{W}_t) = E[Y(t)|X, \widehat{W}_t]$ is the estimated outcome regression, X are the condition variables, T is the treatment variable which is chosen from variable set W, and $Y(t)$ is the outcome variable under treatment $T = t$.

To calculate the conditional average treatment effects, in this section, gradient boosting regression is first applied to fit $\mu(X, W) = E[Y|X, W]$ and $f(X, W) = E[T|X, W]$, and local estimates $\widehat{\mu_i}(X, W)$ and $\hat{f}(X, W)$ are obtained. Then, the causal forest proposed by Athey (Athey et al., 2019) is fitted to solve the "local moment equation problem" [29].

$$E[(Y - E[Y|X, W] - \theta(x)(T - E[T|X, W]))(T - E[T|X, W])|X = x] = 0 \tag{2}$$

In this section, we apply the double machine learning version of causal forests to estimate conditional average treatment effects $\hat{\theta}(x)$. According to Athey et al. (Athey et al., 2019) and Keith et al. (Keith Battocchi Eleanor Dillon, 2019), the estimation is

$$\hat{\theta}(x) = argmin_{\theta} \sum_{i=1}^{n} K_x(X_i) \cdot (Y_i - \widehat{\mu_i}(X_i, W_i) - \theta(T_i - \hat{f}(X_i, W_i)))^2 \tag{3}$$

where $K_x(X_i)$ is the similarity metric. We finally obtain the treatment effect estimator by using the double machine learning version of causal forests, which estimator can be denoted as $CATE(X, Y)$.

Decision-making system

Efficiency, safety and eco-friendliness are the three key objectives of traffic management and control and the objectives of system optimization in the context of a CVIS. In the decision system to be built for the lane change scenario, the lane change duration is selected to denote the efficiency, the TTC to denote the safety and the acceleration to denote eco-friendliness. The shorter the lane change duration is, the more efficient the lane change is, and the shorter the impact of the lane change on other vehicles. The TTC needs to be controlled within a safe range during the lane change to reduce the risk of vehicle collisions. From the aspect of energy consumption and emissions, the more stable the velocity is, the less energy consumed and exhaust emissions produced. Therefore, the system should avoid rapid acceleration or deceleration.

In this decision-making system, two controllers are built to determine the suggested velocity and acceleration. According to the definition of the CATE, in this study, the CATE is applied to measure the "elasticity" of the velocity with respect to the duration, and the TTCs are added to the framework to constrain the velocity. Considering traffic safety, the TTCs must be within a reasonable interval *B*, which means $TTCs \in B$. The objective function of the first controller is shown in **Equation (4)**.

$$min_v CATE(TTC) + \lambda_1 F(TTC) \tag{4}$$

where $\lambda \geq 0$, and $F(TTC) = \sum_i TTC_i$. The CATE under different TTCs can be denoted as $CATE(TTC) = CATE(TTC_{ft}, TTC_f,\ TTC_{rt}, TTC_r,\ TTC_{st})$, which can be estimated using the causal forest in **Equation (3).** The second term in **Equation (4)** aims to constrain the velocity so that it does not choose a point that has a high CATE but low velocity.

To avoid high energy consumption and high emissions. The acceleration is added into the algorithm. The second controller aims to find an optimal acceleration. We denote the acceleration as $a \in H$ in this section, and *H* is the covarites set, which is used to prepare for a more general model in the future. From the perspective of efficiency and energy consumption, we use the objective function shown in **Equation (5)** to help find the best suggested covariates set $\widehat{H}$.

$$min_a CATE(a') + \lambda_2 \left| a' - a_j \right| \tag{5}$$

The first term in **Equation (5)** improves the efficiency of *v* toward the duration under a'. The second term aims to help the vehicle transition more smoothly. A decreasing acceleration leads to a decrease in efficiency and an increasing acceleration leads to high energy consumption and high emissions, hence, a stable acceleration can help save energy, and it can also ensure that the lane change process is more stable.

Finally, the complete algorithm for the lane change decision system is shown in **Algorithm 1**.

Algorithm 1 Lane change decision algorithm

Input: Training set $D = (x_{ft}, y_{ft}, x_f, y_f, x_{rt}, y_{rt}, y_r, x_{st}, y_{st})$, $V = (v, v_{ft}, v_f, v_{rt}, v_r)$, *duration*; Test set $D' \in \mathbb{R}_{n\times 6}$, $V' = (v'_{ft}, v'_f, v'_{rt}, v'_r, v'_{st}) \in \mathbb{R}_{n\times 5}$; Control variable set H; Coefficients λ

1: Compute $TTC(D, V) = (TTC_{ft}, TTC_f, TTC_{rt}, TTC_r, TTC_{st})$;
2: Compute $CATE(TTC(D, V), treatment = v, outcome = duration)$;
3: **for** D'_i, V'_i in D', V' **do**
4: Generate evenly spaced samples V' and H' according to V and H;
5: Compute $CATE(TTC(D', V', v'), treatment = v', outcome = duration)$;
6: Search v'_i to find $\hat{v}_i$ to minimize $f(v'_i) = CATE(TTC) + \lambda_1 F(TTC)$, such that $TTC(D', V', \hat{v}_i) \in B$;
7: **for** a', a in H', H **do**
8: Compute $CATE(a', treatment = v', outcome = duration)$;
9: Search a' to find $\hat{a}$ to minimize $g(a') = CATE(a') + \lambda_2 |a' - a|$;
10: **end for**
11: **end for**
12: $\hat{H} = \hat{a}$

Output: $\hat{V}$, $\hat{H}$, $CATE$

In this algorithm, the function to generate an evenly spaced array is in line 4, and the even interval is determined by the range of *T*. By learning the training set, we can obtain the CATE of the velocity for the duration at a specific TTC. To find the optimal solution for the lane change state,two types of optimization problems with regularization terms need to be solved. Finally, the suggested velocity and covariates set can be calculated. With the help of the suggested parameters, the vehicle can obtain an optimal decision to change lanes.

DATA DESCRIPTION

The data we use to conduct the case study are from previous work proposed by Dong et al. (Dong et al., 2017), and we also use NGSIM (Alexiadis et al., 2004) to extract the lane change behavior.

In this section, we add different types of extra variables that are generated by the lane change behavior to the data set. The description of these variables is shown in **Table 1.**

TABLE 1 Variable Descriptions

Variables	Description	Variables	Description
len	Vehicle length	f_len	Front vehicle’s length in current lane
v	Vehicle velocity	r_len	Rear vehicle’s length in current lane
a	Vehicle acceleration	ft_len	Front vehicle’s length in target lane
v_after	Vehicle velocity after changing lane	rt_len	Rear vehicle’s length in target lane
a_after	Vehicle acceleration after changing lane	st_len	Overlap vehicle’s length in target lane
f_v	Front vehicle’s velocity in current lane	x_ft	The x distance with front vehicle in target lane
r_v	Rear vehicle’s velocity in current lane	y_ft	The y distance with front vehicle in target lane
ft_v	Front vehicle’s velocity in target lane	x_rt	The x distance with rear vehicle in target lane
rt_v	Rear vehicle’s velocity in target lane	y_rt	The y distance with rear vehicle in target lane
st_v	Overlap vehicle’s velocity in target lane	x_f	The x distance with front vehicle in current lane
f_a	Front vehicle’s acceleration in current lane	y_f	The y distance with front vehicle in current lane
r_a	Rear vehicle’s acceleration in current lane	x_r	The x distance with rear vehicle in current lane
ft_a	Front vehicle’s acceleration in target lane	y_r	The y distance with rear vehicle in current lane
rt_a	Rear vehicle’s acceleration in target lane	x_st	The x distance with overlap vehicle in target lane

st_a	Overlap vehicle's acceleration in target lane	y_st	The y distance with overlap vehicle in target lane
duration	The duration of lane changing		

There are 544 observed lane change vehicles and 31 variables for each lane change, and each observation has a specific lane change trajectory. **Table 2** shows the linear correlation of the duration with the other variables. The largest absolute linear correlation value is -0.183 obtained between the acceleration and duration. All variables have little linear correlation. In the real world, when a lane change occurs, most variables must have a strong association with the duration. For example, there is definitely a causal relationship between the velocity and duration. Therefore, in the next section, we will use causal effects to quantify the relationship between the variables and duration.

TABLE 2 Correlation Between Each Variable and Duration

Variable	Value	Variable	Value	Variable	Value	Variable	Value
len	-0.021	ft_v	-0.131	st_a	-0.088	y_ft	0.068
v	-0.152	rt_v	-0.127	f_len	0.0008	x_rt	0.015
a	-0.089	st_v	-0.099	r_len	-0.015	y_rt	-0.065
v_after	-0.183	f_a	0.014	ft_len	0.074	x_f	0.093
a_after	0.006	r_a	0.028	rt_len	-0.055	y_f	-0.100
f_v	-0.098	ft_a	0.046	st_len	-0.070	x_r	0.049
r_v	-0.114	rt_a	-0.008	x_ft	-0.025	y_r	-0.072

CASE STUDY

Causal effect analysis

A causal diagram is a visualization tool that can visualize the causal relationship among variables, and it is a directed acyclic graph (DAG) composed of nodes and arrows. Nodes are composed of the treatment variable *T*, the outcome variable *Y* and other variables *W*. The direction of the arrow is the direction of our hypothetical causal relationship. **Figure 2** is the causal diagram that shows the causal relationships of the variables in **Table 1**. Since the causal relationships between these variables are a priori knowledge, the construction of the causal graphs is completed through manual selection.

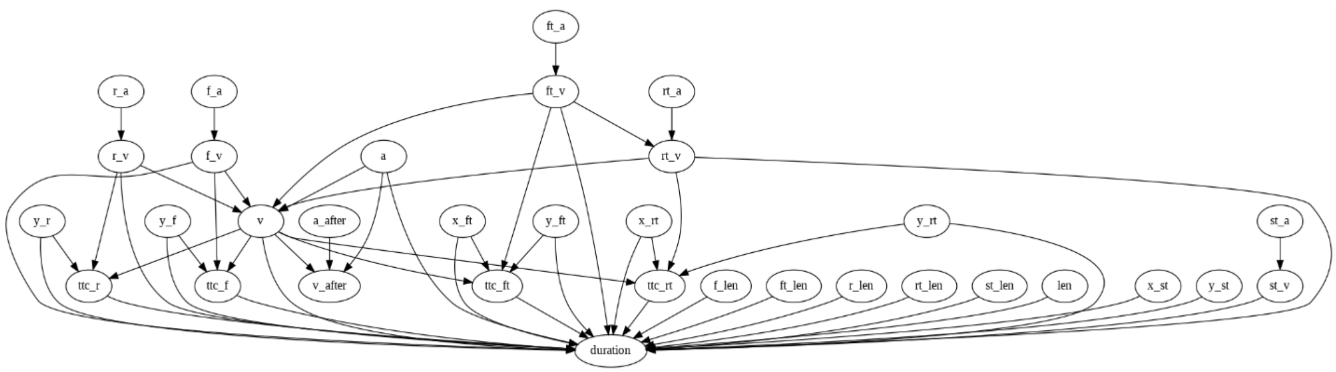

Figure 2 Hypothetical causal relationships between variables

In this section, the average treatment effect of the potential treatment variables is calculated. The Python package "Dowhy" (Sharma & Kıcıman, 2020) is used to construct a causal diagram and calculate the average treatment effect. According to the diagram, the outcome variable is the duration, and the treatment variables are shown in **Figure 4**.

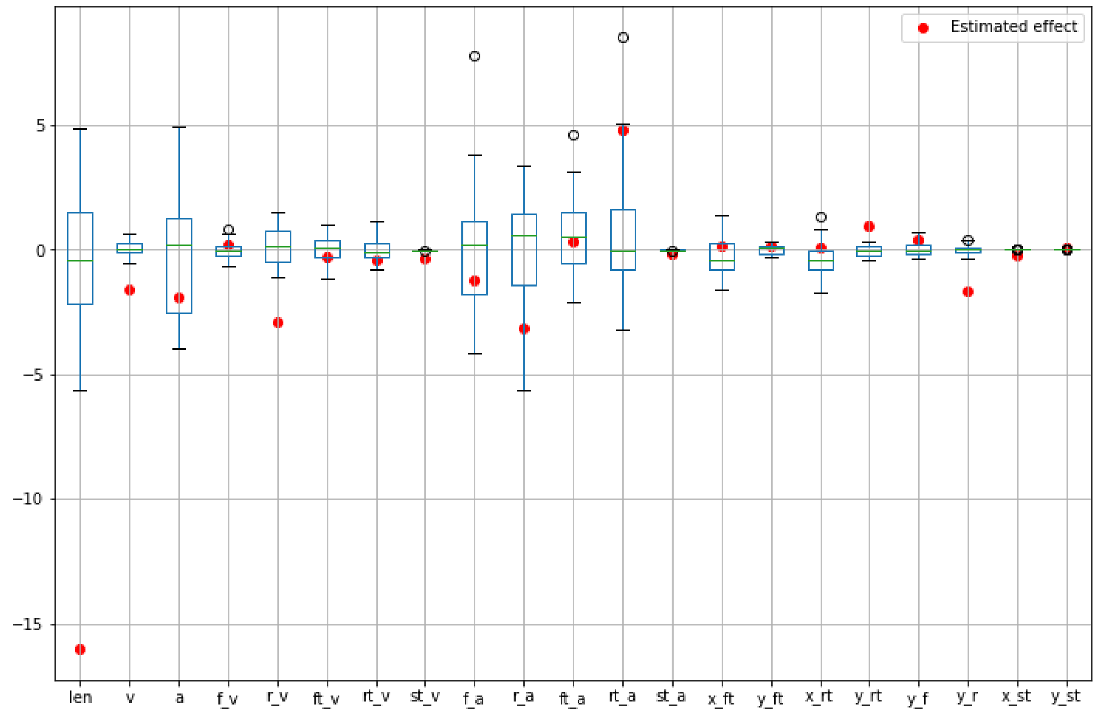

Figure 3 Boxplot of causal effects using placebo

The back-door criterion is adopted as the identification criterion, and regression is applied to estimate the causal effect. In the last step, a placebo treatment as a refutation method is applied to obtain a new treatment effect. The results of the estimated treatment effects are shown in **Figure 3**. The red dots are the estimated effects of different variables. The boxplots are the treatment effects that are generated from 30 randomized placebo tests.

According to the rules of the placebo refutation test in causal inference, the causal effect between each variable and the outcome variable can be estimated. From **Figure 3**, the estimated effect of “len”, “v”, “r_v”, “st_v”, “st_a”, “y_rt”, “x_st” and “y_r” have a significant difference from the effect using placebo treatment, and the effect using placebo treatment tends to be zero. It can be concluded that these variables have a causal effect on the outcome variable, which is the duration of the lane change process.

Classifying the variables according to the different level reveals that “len”, “v”, “r_v”, “y_rt” and “y_r” have a higher causal effect on the outcome variable, and “st_v”, “x_st” and “st_a” have a lower causal effect on the outcome variable.

Using the positive and negative causal effects for identification, we find that “len”, “v”, “r_v”, “st_v”, “st_a”, “x_st” and “y_r” have a negative causal effect on the outcome variable, while “y_rt” has a positive causal effect on the outcome variable.

Conditional average treatment effect under the TTCs and acceleration

In this section, the double machine learning version of causal forests is used to estimate conditional average treatment effects under the TTCs and acceleration. Api from “EconML” (Keith et al., 2019) is used to calculate the CATEs. There are five parameters in this step: “TTC_ft”, “TTC_f”, “TTC_rt”, “TTC_r” and “TTC_st”. These five variables are condition variables X, which have been mentioned in **Equation (2)**. The velocity is the treatment variable while duration is the outcome variable.

The calculation of the different condition variables is divided into four parts: front vehicles, rear vehicles, overlap vehicles in target lane and acceleration. Two 3D surface plots and two 2D plots are constructed to show the conditional average treatment effects under different TTCs and accelerations. The results are shown in **Figure** 4.

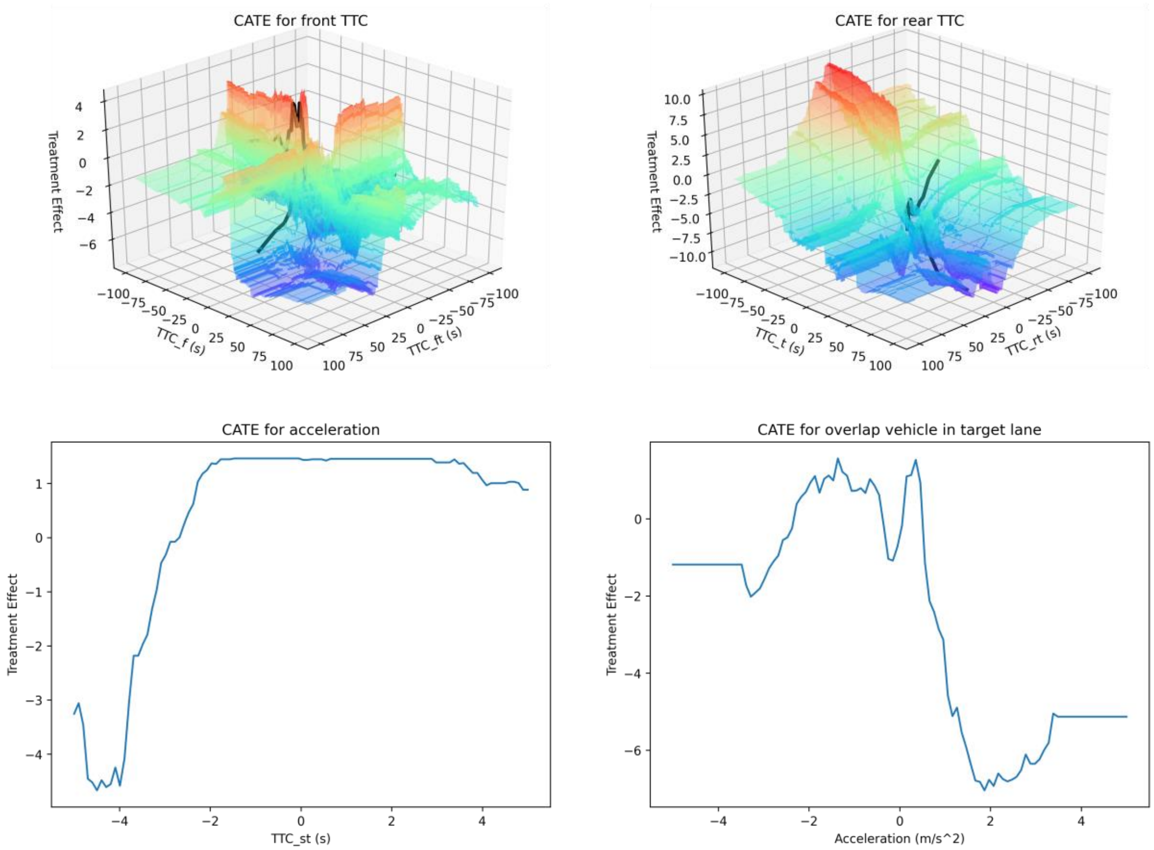

Figure 4 Treatment effects for front, and rear vehicles, overlap vehicles and acceleration

In the lane change scenario, different TTCs or accelerations can affect the CATE differently. In **Figure 4**, a positive value in the 3D and 2D plot represents the positive "elasticity" of the velocity with respect to the duration at a specific TTC or acceleration, implying that an increase in the velocity causes an increase in the duration, which means the velocity has a positive causal effect on the duration. Negative values represent a negative causal effect of the velocity on the duration at a specific TTC or acceleration.

To understand the CATE of a specific vehicle at different velocities with respect to the duration, we add solid black lines that denote one single car under different velocities in the graphs. We take a single vehicle as an example and use the rear and front vehicle TTCs as covariates to simulate changes in the

CATE. The results are shown in the first two graphs of **Figure 4**. The solid black lines show the variation in the vehicle velocity CATE under different TTCs.

For different cars, the paths of the TTCs in the images are different. Not all vehicles can reach the lowest CATE in the surface plot, and the value depends on the location and velocity of the vehicle. Considering the safe range of the TTCs and the causal effect of the velocity on the duration, a lower CATE does not mean a better lane change decision. When the velocity is low, even if the CATE is very low, the efficiency of the lane change can be very poor, which will cost the vehicle more fuel to reduce the duration of the lane change. Hence, the value of the velocity needs to be taken into consideration, and this is the purpose of the second term in **Equation (4)**. The change in acceleration should also be taken into consideration, because of the efficiency and energy consumption, a lower CATE also does not mean a better lane change decision.

From the experiments, the CATE is not linearly related to the duration of lane change, and the CATE cannot be used as the only condition to evaluate a good or bad lane change decision. The process of finding the optimal lane change decision is to find a suitable point in a six-dimensional hyperplane. To obtain the optimal lane change decision, we apply **Algorithm 1** to the test data set and conduct further experiments.

Lane change decision-making

According to **Algorithm 1**, after obtaining the TTCs and the treatment effect, the next step is to find the velocity that minimizes **Equation (4)** and find the condition variables *T* that minimize **Equation (5)** using a grid search. In this system, the suggested velocity and acceleration are calculated to balance the safety and efficiency according to the hyperparameters. Hence, according to the TTC distribution (Murata et al., 2016), the TTC threshold is set to be 2.5s in this paper.

To obtain the optimal lane change duration, the estimated lane change duration under different λ_1 and λ_2 is first calculated, and the results are shown in **Figure 5**. A surface plot is applied to show the different durations. It is clear that the suggested duration decreases as λ_1 increases, and the trajectory of λ_2 is more complex; it reaches its minimum value at $\lambda_2 = 2$ or 3.

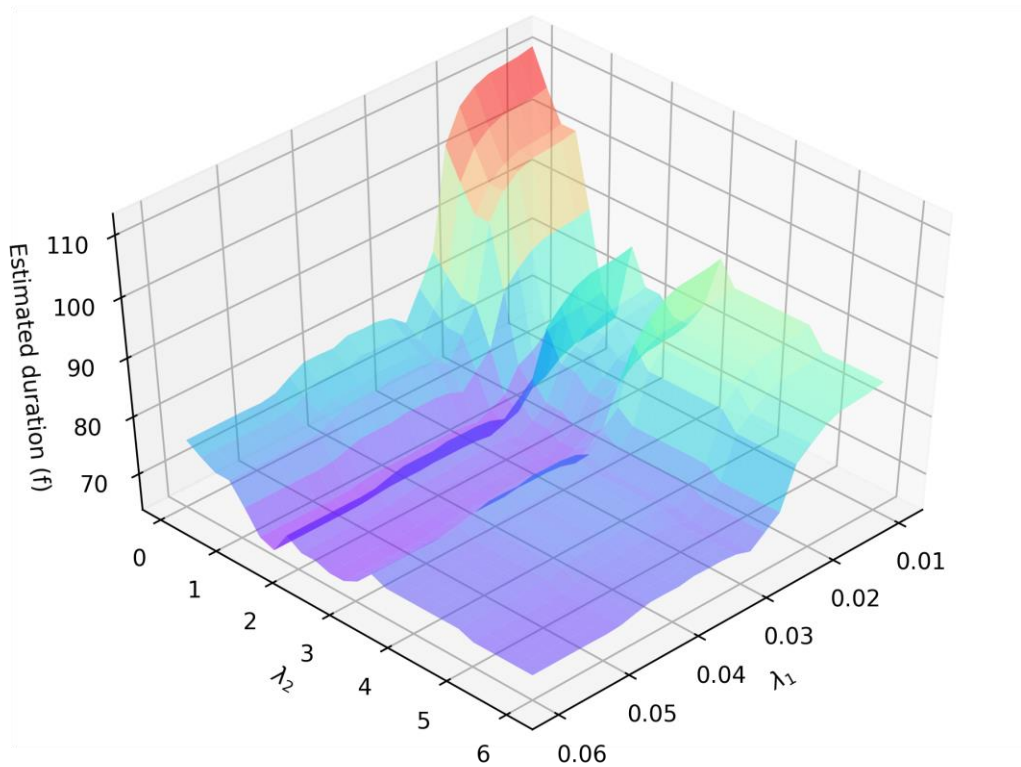

Figure 5 Estimated duration at different λ_1 and λ_2

Several experiments were implemented to calculate the average velocity, average acceleration, CATE and average duration before and after using the framework. The candidate values of the parameters are $\lambda_1 = 0.1, 0.2, 0.3, 0.4, 0.5$, and $\lambda_2 = 2, 3$. According to Figure 7, the lowest estimated duration can be obtained when λ_2 is approximately 2 or 3. Furthermore, the specific values at different λ_1 and λ_2 values are summarized in **Table 3**.

TABLE 3 Change in velocity, CATE and average duration before and after using the framework

		Without using framework				Using framework			
λ_1	λ_2	$\bar{v}$	$\bar{a}$	CATE	Average Duration	$\bar{v}$	$\bar{a}$	CATE	Average Duration
0.01	2	7.165	0.196	−2.029	72.043	7.819	0.664	−5.926	82.941

0.01	3	7.165	0.196	−2.029	72.043	7.819	0.545	−5.633	89.189
0.02	2	7.165	0.196	−2.029	72.043	7.797	0.664	−5.924	79.091
0.02	3	7.165	0.196	−2.029	72.043	7.797	0.545	−5.638	77.991
0.03	2	7.165	0.196	−2.029	72.043	7.791	0.664	−5.926	68.973
0.03	3	7.165	0.196	−2.029	72.043	7.791	0.545	−5.633	67.792
0.04	2	7.165	0.196	−2.029	72.043	7.777	0.664	−5.931	68.981
0.04	3	7.165	0.196	−2.029	72.043	7.777	0.545	−5.631	67.809
0.05	2	7.165	0.196	−2.029	72.043	7.817	0.664	−3.341	68.979
0.05	3	7.165	0.196	−2.029	72.043	7.817	0.545	−3.048	67.823

The proposed framework is evaluated in three aspects under different hyperparameters λ_1 and λ_2. Five candidate values of λ_1 and two candidate values of λ_2 are considered. The lowest average duration after using the framework of 67.792 is obtained when $\lambda_1 = 0.03$ and $\lambda_2 = 3$, and the average duration decreases from 72.043 to 67.792 when using the framework. The average velocity increases from 7.165 to 7.791, an increase of 8.74%, and the CATE decreases from -2.029 to -5.633.

However, the average acceleration increases from 0.196 to 0.545. The increase is unexpected and an in-depth discussion is conducted. A test data set that contains 50 vehicles is used to test the suggested acceleration; λ_1 is set to 0.03, and λ_2 is set to 3. **Figure 6** shows the optimization results of the 50 vehicle lane changes. The red dots are the suggested acceleration, while the blue dots are the real acceleration. From **Figure 6**, it is clear that the real acceleration of most vehicles is consistent with the suggested acceleration, and only 8% of the vehicles have a significant increase in acceleration compared to the real acceleration.

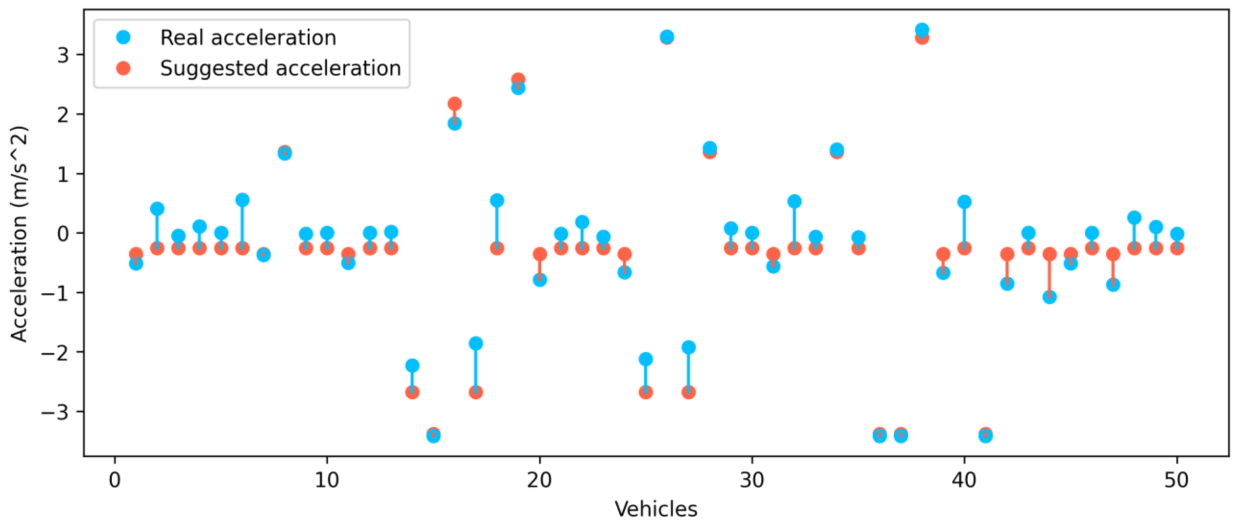

Figure 6 Change in the velocity and acceleration before and after using the framework

Overall, the decision system can optimize the lane change process by adjusting the velocity and acceleration to achieve the optimal lane change duration, and all decisions are made under safety (TTC) conditions. The proposed model can well balance the efficiency (duration) and environmental protection (acceleration) during a lane change. Moreover, the decision system also has the ability to be generalized in the future by considering more parameters.

CONCLUSIONS

This study focuses on the discretionary lane-changing process. A causal inference-based speed control framework was proposed to improve lane change behavior from the aspects of efficiency, safety and eco-friendliness, which are denoted by the velocity, TTC and acceleration, respectively. The following contributions are made from this study:

(1) By estimating the treatment effect of candidate variables and performing a placebo refutation test, the effect between variables during a lane change is quantified. Then, the final variables can be selected and their treatment effects can be obtained.

(2) By applying a double machine learning version of the causal forest model, the CATE is estimated. Based on the CATE, the lane change control pronlem is transferred into an optimization problem of the TTC safety variable, the eco-friendly acceleration variable and the control velocity variable.

(3) By applying the proposed lane change framework, the final validation results show that the framework can improve the efficiency and environmental friendliness of lane change while ensuring safety.

Along the stream of this study, several elements of future research can be identified. (1) To obtain a more general and accurate framework, extra lane change parameters such as lane change angle or interaction variables of the target vehicle with the surrounding vehicles can also be added to the algorithm. (2) The framework proposed in this paper can be improved according to specific CVIS scenarios, and it can be applied to the lane change system optimization module in collective intelligent decision making and cooperative control of traffic groups to improve the performance of lane change decision-making.

DATA AVAILABILITY STATEMENT

Some or all data, models, or codes that support the findings of thisstudy are available from the corresponding author upon reasonablerequest.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: Yi Zhao, Zhen Zhou; data collection: Zhen Zhou, Minghao Li, Yuyang Bao; analysis and interpretation of results: Yi Zhao, Zhen Zhou; draft manuscript preparation: Zhen Zhou, Minghao Li, Yuyang Bao. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

Ahmed, I., Xu, D., Rouphail, N., & Karr, A. (2019). Lane Change Rates at Freeway Weaving Sites: Trends in HCM6 and from NGSIM Trajectories. *Transportation Research Record: Journal of the Transportation Research Board*, *2673*(5), 627–636. https://doi.org/10.1177/0361198119841281

Alexiadis, V., Colyar, J., Halkias, J., Hranac, R., & McHale, G. (2004). The next generation simulation program. *ITE Journal (Institute of Transportation Engineers)*, *74*, 22–26.

Ali, Q., Yaseen, M. R., & Khan, M. T. I. (2019). The causality of road traffic fatalities with its determinants in upper middle income countries: A continent-wide comparison. *Transportation Research Part A: Policy and Practice*, *119*, 301–312. https://doi.org/10.1016/j.tra.2018.12.002

Ali, Y., Zheng, Z., Haque, Md. M., & Wang, M. (2019). A game theory-based approach for modelling mandatory lane-changing behaviour in a connected environment. *Transportation Research Part C: Emerging Technologies*, *106*, 220–242. https://doi.org/10.1016/j.trc.2019.07.011

Ali, Y., Zheng, Z., Mazharul Haque, Md., Yildirimoglu, M., & Washington, S. (2020). Understanding the discretionary lane-changing behaviour in the connected environment. *Accident Analysis & Prevention*, *137*, 105463. https://doi.org/10.1016/j.aap.2020.105463

Arbis, D., & Dixit, V. v. (2019). Game theoretic model for lane changing: Incorporating conflict risks. *Accident Analysis & Prevention*, *125*, 158–164. https://doi.org/10.1016/j.aap.2019.02.007

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, *47*(2). https://doi.org/10.1214/18-AOS1709

Cao, D., Wu, J., Dong, X., Sun, H., Qu, X., & Yang, Z. (2021). Quantification of the impact of traffic incidents on speed reduction: A causal inference based approach. *Accident Analysis & Prevention*, *157*, 106163. https://doi.org/10.1016/j.aap.2021.106163

Cao, P., Hu, Y., Miwa, T., Wakita, Y., Morikawa, T., & Liu, X. (2017). An optimal mandatory lane change decision model for autonomous vehicles in urban arterials. *Journal of Intelligent Transportation Systems*, *21*(4), 271–284. https://doi.org/10.1080/15472450.2017.1315805

Ding, P. (2017). A Paradox from Randomization-Based Causal Inference. *Statistical Science*, *32*(3). https://doi.org/10.1214/16-STS571

Dong, C., Zhang, Y., & Dolan, J. M. (2017). Lane-change social behavior generator for autonomous driving car by non-parametric regression in Reproducing Kernel Hilbert Space. *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 4489–4494. https://doi.org/10.1109/IROS.2017.8206316

Hitchcock, C. (2001). CAUSALITY: MODELS, REASONING AND INFERENCE. *The Philosophical Review*, *110*(4), 639–641. https://doi.org/10.1215/00318108-110-4-639

Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, *81*(396), 945–960. https://doi.org/10.1080/01621459.1986.10478354

Keith Battocchi Eleanor Dillon, M. H. G. L. P. O. M. O. V. S. (2019). *EconML: A Python Package for ML-Based Heterogeneous Treatment Effects Estimation*.

Li, C., Chan, S. H., & Chen, Y.-T. (2020). Who Make Drivers Stop? Towards Driver-centric Risk Assessment: Risk Object Identification via Causal Inference. *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 10711–10718. https://doi.org/10.1109/IROS45743.2020.9341072

Li, M., Li, Z., Xu, C., & Liu, T. (2020). Short-term prediction of safety and operation impacts of lane changes in oscillations with empirical vehicle trajectories. *Accident Analysis & Prevention*, *135*, 105345. https://doi.org/10.1016/j.aap.2019.105345

Lin, D., Li, L., & Jabari, S. E. (2019). Pay to change lanes: A cooperative lane-changing strategy for connected/automated driving. *Transportation Research Part C: Emerging Technologies*, *105*, 550–564. https://doi.org/10.1016/j.trc.2019.06.006

Luo, Y., Xiang, Y., Cao, K., & Li, K. (2016). A dynamic automated lane change maneuver based on vehicle-to-vehicle communication. *Transportation Research Part C: Emerging Technologies*, *62*, 87–102. https://doi.org/10.1016/j.trc.2015.11.011

Mai, T., Jiang, R., & Chung, E. (2016). A Cooperative Intelligent Transport Systems (C-ITS)-based lane-changing advisory for weaving sections. *Journal of Advanced Transportation*, *50*(5), 752–768. https://doi.org/10.1002/atr.1373

Marchant, J. (2016). Placebos: Honest fakery. *Nature*, *535*(7611), S14–S15. https://doi.org/10.1038/535S14a

Molavipour, S., Bassi, G., iić, M., Skoglund, M., & Johansson, K. H. (2020). Causality Graph of Vehicular Traffic Flow. *ArXiv*, *abs/2011.11323*.

Murata, E., Usui, T., Nogi, K., & Takahashi, H. (2016, April 5). *Study on TTC Distribution when Approaching a Lead Vehicle*. https://doi.org/10.4271/2016-01-1452

Neuberg, L. G. (2003). CAUSALITY: MODELS, REASONING, AND INFERENCE, by Judea Pearl, Cambridge University Press, 2000. *Econometric Theory*, *19*(04). https://doi.org/10.1017/S0266466603004109

Sharma, A., & Kıcıman, E. (2020). DoWhy: An End-to-End Library for Causal Inference. *ArXiv*, *abs/2011.04216*.

Sun, Q., Wang, C., Fu, R., Guo, Y., Yuan, W., & Li, Z. (2021). Lane change strategy analysis and recognition for intelligent driving systems based on random forest. *Expert Systems with Applications*, *186*, 115781. https://doi.org/10.1016/j.eswa.2021.115781

Talebpour, A., Mahmassani, H. S., & Hamdar, S. H. (2015). Modeling Lane-Changing Behavior in a Connected Environment: A Game Theory Approach. *Transportation Research Procedia*, *7*, 420–440. https://doi.org/10.1016/j.trpro.2015.06.022

Xie, D.-F., Fang, Z.-Z., Jia, B., & He, Z. (2019). A data-driven lane-changing model based on deep learning. *Transportation Research Part C: Emerging Technologies*, *106*, 41–60. https://doi.org/10.1016/j.trc.2019.07.002

Yang, M., Wang, X., & Quddus, M. (2019). Examining lane change gap acceptance, duration and impact using naturalistic driving data. *Transportation Research Part C: Emerging Technologies*, *104*, 317–331. https://doi.org/10.1016/j.trc.2019.05.024

Yang, S., Ning, L., Cai, X., & Liu, M. (2021). Dynamic Spatiotemporal Causality Analysis for Network Traffic Flow Based on Transfer Entropy and Sliding Window Approach. *Journal of Advanced Transportation*, *2021*, 1–17. https://doi.org/10.1155/2021/6616800

You Tackgeun and Han, B. (2020). Traffic Accident Benchmark for Causality Recognition. In H. and B. T. and F. J.-M. Vedaldi Andrea and Bischof (Ed.), *Computer Vision – ECCV 2020* (pp. 540–556). Springer International Publishing.

Yu, H., Tseng, H. E., & Langari, R. (2018). A human-like game theory-based controller for automatic lane changing. *Transportation Research Part C: Emerging Technologies*, *88*, 140–158. https://doi.org/10.1016/j.trc.2018.01.016

Zhang, X., Sun, J., Qi, X., & Sun, J. (2019). Simultaneous modeling of car-following and lane-changing behaviors using deep learning. *Transportation Research Part C: Emerging Technologies*, *104*, 287–304. https://doi.org/10.1016/j.trc.2019.05.021