Hindawi
Journal of Advanced Transportation
Volume 2017, Article ID 7170358, 12 pages
https://doi.org/10.1155/2017/7170358

Research Article

Decision Tree-Based Maneuver Prediction for Driver Rear-End Risk-Avoidance Behaviors in Cut-In Scenarios

Manjiang Hu,[1] Yuan Liao,[2] Wenjun Wang,[1] Guofa Li,[3] Bo Cheng,[1] and Fang Chen[2]

[1]*Department of Automotive Engineering, Tsinghua University, Beijing 100084, China*
[2]*Department of Applied IT, Chalmers University of Technology, 412 96 Gothenburg, Sweden*
[3]*Institute of Human Factors and Ergonomics, College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen 518060, China*

Correspondence should be addressed to Wenjun Wang; wangxiaowenjun@tsinghua.edu.cn and Guofa Li; guofali@szu.edu.cn

Received 23 July 2016; Revised 13 November 2016; Accepted 22 December 2016; Published 15 February 2017

Academic Editor: S. C. Wong

Predicting driver rear-end risk-avoidance maneuvers in cut-in scenarios, especially dangerous precrash scenarios, benefits the customization of automatic driving, particularly automatic steering. This paper studies driver rear-end risk-avoidance behaviors in cut-in scenarios on a straight three-lane highway. Data from 24 participants in 1326 valid trials were collected using a motion-based driving simulator. An Eysenck Personality Questionnaire (revised for Chinese participants) was used to obtain the personality traits of the participants. Based on a statistical analysis, the candidate features used in the driver maneuver prediction were determined as a combination of objective risk indicators and driver characteristics. A decision tree-based model was constructed for maneuver prediction in cut-in scenarios. The prediction accuracy of the extracted classification rules was 79.2% for the training data set and 80.3% for the test data set. The most powerful predictive variables were extracted, and their effects on maneuver decisions were analyzed. The results show that driver characteristics strongly influence the prediction of maneuver decisions.

1. Introduction

Autonomous driving systems have been developed from various discrete automated functions (e.g., Adaptive Cruise Control, ACC). More recently, highly automated driving (HAD) systems that automate both longitudinal and lateral motions [1] have been developed. ACC combined with braking and steering interventions has been realized by several automakers, such as Daimler and Volvo [2, 3]. Such systems assist drivers during emergency braking and/or steering maneuvers based on driver input and surrounding sensor information [4]. These combined systems can help avoid collisions in situations where evasive steering is a more appropriate reaction than braking [5–7]. HAD systems provide operational-level assistance in emergency situations. Recent studies of automated steering in the context of HAD have focused on hazard assessment for decision-making [8] and dynamic control for collision avoidance [9–11]. It has been widely understood by researchers and engineers in intelligent driving research field that stages of autonomous driving are divided into 5 levels. They are no automation, function-specific automation, combined function automation, limited self-driving automation, and full self-driving automation. HAD systems refer to the level of limited self-driving automation, where drivers are required to monitor and occasionally to take over.

Issues associated with a user-appropriate design of automatic systems have been considered for decades [12]. With the ever-expanding capabilities of technical systems, such issues are becoming increasingly important [13]. That importance comes from two aspects. One is about capability of single vehicle; before reaching fully autonomous driving, driver needs to monitor and get prepared to take over occasionally. It requires decision-making for HAD systems which is adaptive to drivers' anticipation. The other is about capability of the whole traffic system before all the on-road vehicles become capable of fully autonomous driving. HAD systems are supposed to interact with surrounding vehicles driven by human drivers. In most HAD systems, for example, ACC, some parameters, for example, gap setting, are

user-configurable [14]. According to a recent study, objective metrics for a high-quality automated driving system relative to safety, functionality, and comfort have been identified [15]. Customization of the driving experience is challenging due to the complexity of driving scenarios, the vital balance between efficiency, comfort, and safety, and the prediction of driver intention. To design a better user experience, understanding and predicting driver decision-making in emergency scenarios are important. Such understanding is the key to acceptance by users, therefore leading to high system effectiveness.

Driver risk-avoidance behavior has been extensively studied. Drivers can take two possible corrective actions in response to several emergency situations, that is, accelerating or braking to control speed and steering to control lateral position [16]. The maneuver decision, that is, whether to brake or steer (change lanes), has also been extensively studied. These studies focused on the strategies that drivers use when an unexpected obstacle appears in their path. Last-second braking and steering judgments in a car-following condition have been studied relative to the alert timing for forward collision warnings [17]. Focusing on the car-following scenario, a statistical model was developed to understand the perceptual processes involved in a driver's decision to brake or steer [18]. It was shown that, in imminent rear-end collision situations, most drivers tend only to brake and often do not attempt to avoid obstacles by steering. It has been documented that steering maneuvers are more complex than braking and, therefore, require higher situation awareness and highly developed driving abilities [19]. It has also been found that drivers often unintentionally perform actions that work against automatic steering maneuvers [7].

Along with the development of HAD systems, driving strategy and high-level decision-making processes should be realized for replacing drivers. In this context, understanding driver risk-avoidance strategies is the first step. The next step, which has attracted increasing attention, is to predict driver decisions. Machine-learning is widely applied in this field with the received surrounding environmental sensors. A random-forest machine-learning technique was used to develop red-light-running violation-prediction models using a single objective indicator (time to intersection) [20]. Venkatraman et al. applied cluster analysis and multinomial logistic regression to predict driver decisions in the car-following situation [18]. Generative approaches for maneuver prediction include various filtering methods, such as Kalman and particle filters, hidden Markov models, and dynamic Bayesian networks [21, 22]. Among machine-learning tools, decision rule-based and decision tree-based models are relatively simple and readable with very fast generation [23]. By applying a DT to predict driver maneuver decisions, both prediction results and the reasons behind the results can be obtained.

The cut-in situation is among the most dangerous pre-crash scenarios based on the classification of 37 precrash scenarios proposed by the National Highway Traffic Safety Administration, which is an agency of the United States Department of Transportation [24]. The cut-in scenario is a fairly common event that ACC systems do not handle the way drivers would. Drivers tend to reassume control and may consider the need to intervene as an indication of poor user experience [25–27]. Drivers consider an effective risk-avoidance decision strategy to be important. Thus, given its potential danger, decision alternatives for the cut-in scenario should be given design priority. How does the driver maneuver when another vehicle cuts in front? Do they brake or steer? What are the perceptions and considerations that motivate their decisions? The results of this study provide a reference for decision-making of HAD that will contribute to a better user experience.

A previous study suggested that 10%–20% of the variance in crashes and more than 35% of the variance in risky driving can be explained by personality variables (e.g., thrill-seeking, impulsiveness, and hostility/anger) [28]. As one important method to quantify the personal traits, the Eysenck Personality Questionnaire (EPQ) conceptualized personality as two biologically based independent dimensions of temperament measured on a continuum, that is, extraversion (E)/introversion and neuroticism (N)/stability [29]. Based on a further study, a third temperament category, psychoticism (P)/socialization, and a fourth scale, the lie, were added to EPQ [30]. EPQ revised for Chinese (EPQ-RSC) is a simplified version [31] to facilitate assessment of a Chinese person's personality traits. EPQ-RSC includes 48 questions related to four dimensions (12 question each, for a maximum of 12 points). Extraversion is characterized by being outgoing, talkative, high on positive affect (feeling good), and in need of external stimulation [32]. Neuroticism is related to experiencing negative affect [33]. Psychoticism refers to the liability to have a psychotic episode (or break with reality) and aggression [34]. It has been shown that accident occurrence is highly affected by personality traits [35, 36]. Dahlen et al. proposed that driver traffic collisions and violations could be predicted using a model involving driver anger, emotional stability, extraversion, openness to experience, agreeableness, and conscientiousness as personality factors [35]. Numerous studies have been conducted to explore the relation between personality traits and risky driving [37, 38]. Results from such studies provide strong support for policy-makers. However, how to apply personality traits to accident prediction and prevention from a technological perspective remains an open question.

As is evident in the preceding literature review, previous studies have had some limitations. User-appropriate design of automatic systems is a key issue for the ever-developing capabilities of autonomous driving systems. However, relative to decision-making and prediction, driver risk-avoidance behavior in cut-in scenarios is not well understood. Currently, maneuver prediction relies on objective indicators, that is, environmental awareness. To better understand and predict driver maneuver decisions, considering the discovered effect of driver personality on risky driving, both objective indicators and driver characteristics should be applied. And driver maneuver prediction in cut-in scenarios is important to the better user experience of HAD.

The remainder of this paper is organized as follows. Data collection and the results of statistical analysis are presented in Section 2. Section 3 describes the construction of the

DT-based model, and Section 4 presents the primary results of the DT model. The results are discussed in Section 5, and conclusions and suggestions for future studies are presented in Section 6.

2. Methodology

This paper presents a decision tree- (DT-) based method for maneuver prediction between braking (BRK) or lane-changing (LC) in cut-in scenarios. The extracted classification rules and their prediction accuracy with the developed model are presented. The data were collected from a motion-based driving simulator with the following scenario. On a three-lane highway, when a low-speed vehicle cuts in from the right lane (RL) to the middle lane (ML), the driver in the ML responds by BRK or moving into the left lane (LL). Usually, time to collision (TTC) is applied to indicate the perceived driving risk. In this simulator-based study, the degree of urgency of the cut-in situation (TTC_c), the adjacent gap time headway (THW) in the LL, the desired speed (v_1), and driver characteristics, including personality traits, age group (A), and gender (G), were analyzed to explain the driver maneuver decision. The DT model was trained and validated using the collected data.

Candidate features were determined based on the analysis. The candidate features are calculated at the start of the cut-in movement and can be classified as objective risk indicators (e.g., TTC_c) and driver characteristics. In addition to A and G, to explore the relation between driver personality and rear-end collision-avoidance maneuver decision, the Eysenck Personality Questionnaire revised for Chinese (EPQ-RSC) was employed to provide three driver personality scales; E, N, and P were used to characterize drivers' personal traits. The results show that the constructed DT-based model performs well for maneuver decision prediction for cut-in scenarios. The most powerful predictive variables were extracted, and their effects on the maneuver decision were analyzed.

2.1. Data Collection

2.1.1. Participants. Twenty-four licensed drivers, in the age bracket of 22–65 years, participated in this experiment. Their driving experience varied from 1 to 28 years. The participants came from a variety of backgrounds, for example, students, factory workers, and engineers. To better explore the age effect, the participants were divided into two groups of 12 by age, that is, younger than 35 (mean = 25.5; SD = 4.5) and older than 45 (mean = 52.2; SD = 6.5). There were six females and six males in each group.

2.1.2. Equipment. The motion-based driving simulator used in this study comprised visual, audio, and motion-simulation units [39]. Vehicle-location and driving-behavior data (speed, acceleration/deceleration, steering angle, etc.) were logged at 60 Hz. There are 5 screens, 3 for front view and 2 for rear view, which can provide 200° of front field view and 55° of back field view while driving for visual simulation. It has six cylinders which can realize 6 degrees of freedom for motion simulation, which can realize angular (roll, yaw, and pitch) and longitudinal movement of ±15° and ±0.4 m, respectively. Sounds of engine, road, and traffic are provided by audio simulation.

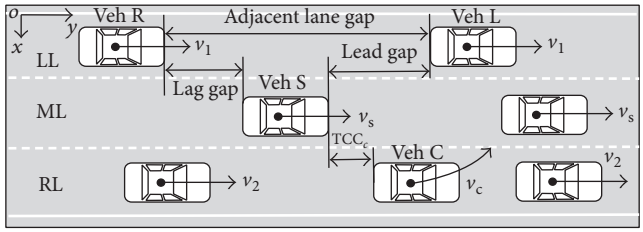

Figure 1: Driving scenario layout.

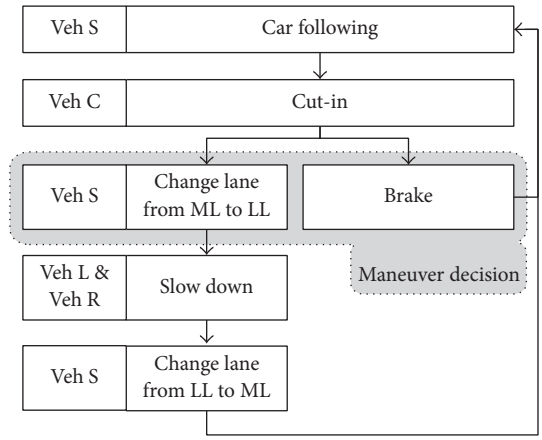

Figure 2: Experiment flow.

2.1.3. Driving Scenario. The driving environment was a straight three-lane urban highway. The general layout of the experimental environment is shown in Figure 1.

Participants were in complete control of the subject vehicle (Veh S). Participants were required to maintain the velocity of Veh S at approximately v_1 in the ML. The cut-in event was realized by a program-controlled vehicle (Veh C) traveling in the RL and triggered by specific cut-in timing, that is, TTC_c, between Veh S and Veh C. The traffic flow in the LL was designed as a series of vehicles with stable velocity v_1 and a specific THW. Each participant underwent 10 minutes of adaptive training prior to beginning the experiment to become familiar with the simulator and the scenario.

The experimental flow is shown in Figure 2. After following a car for approximately 15 s, Veh C cuts in front of Veh S from the RL at a lower speed than Veh S. Vehicles in the RL travel at 20 km/h (v_2), being randomly selected as Veh C to perform cut-in process. Participant was instructed to change lanes or brake to avoid collision with Veh C. Meanwhile, all vehicles in the LL continued driving at speed v_1 and fixed THW.

The time at which Veh C cuts in is triggered by specific TTC_c thresholds. After the cut-in event, the participant could choose to brake or change lanes to avoid an imminent collision as they normally would while actually driving. If

Table 1: Experiment design: impact factor levels.

#	Factor	Level 1	Level 2	Level 3
1	v_1 (km/h)	40	60	80
2	TTC_c (s)	3.5	5.0	6.5
3	Adjacent gap THW (s)	2	4	6

the participant braked, then a period of car-following would repeat the cut-in scenario. If the participant chose to change lanes, the vehicles in the LL slowed down to force the participant to switch back to the ML to allow the experiment to be repeated.

2.1.4. Experimental Design. A driver's maneuver decision is highly influenced by perceived risk, which can be described using the objective indicators v_1, THW, and TTC_c. The speed of vehicles in the LL (v_1) influences LC difficulty. Vehicle speed in the LL (v_1) and THW indicate traffic flow density, which also influences LC difficulty. TTC_c indicates the emergency level of the cut-in process, which determines the amount of time drivers have to respond to the cut-in event. Based on our previous lane-changing behavior study of Chinese drivers, the TTC value used in this study was intended to realize 90% coverage of normal lane-changing behavior (i.e., 3.5–6.5 s).

To explore the maneuver decision on different levels of perceived driving risk, an orthogonal experiment design was adopted. As shown in Table 1, three within-subject factors, that is, v_1, THW, and TTC_c, were applied, resulting in a seven-trial repeated-measures design for each combination of the orthogonal experiment L (3^{3-1}). Note that nine emergency situations occurred and were counterbalanced across participants to eliminate the learning effect.

2.2. Data Description. In this section, collected data is illustrated through descriptive statistical analysis. ANOVA was applied to show the relationship between maneuver decision and impact factors including experiment parameters and driver characteristics (see Sections 2.2.1 and 2.2.2). Then, in Section 2.2.3, extraction of cut-in event and original signals were presented. And lastly, those features, fed to the construction of DT model, were summarized.

2.2.1. Maneuver Decisions and Independent Variables. From 24 participants, 1326 valid instances were collected. There were nine emergency situations; for each participant, each emergency situation had a single LC rate (frequency of LC divided by 7 repetitions), which occurred in 216 instances (208 valid instances) in total. Driver LC rates in the cut-in scenarios are summarized in Table 2. Younger participants tended to change lanes more frequently than older participants (25.81% versus 15.29%; $F(1, 206) = 6.79$; $p = 0.01$). Male participants had slightly higher tendency to change lanes than female participants (23.82% versus 17.16%; $F(1, 206) = 2.552$; $p = 0.112$). The combination of impact factor levels influenced participant maneuver decisions significantly. The LC rate was significantly influenced by v_1 ($F(2, 205) = 6.254$; $p = 0.002$) and THW ($F(2, 205) = 6.604$; $p = 0.002$), and TTC_c had no significant effect on LC maneuver decisions ($F(2, 205) = 0.341$; $p = 0.711$). Note that the three factors were applied to create different levels of urgency in the cut-in scenarios. For each single instance, the event's details changed within a particular scenario. To predict driver maneuver decisions, more detailed indicators should be involved.

Table 2: Maneuver decision results (mean value).

Independent variables		Levels	LC rate (%)
Age group		Young	25.81
		Older	15.29
Gender		Male	23.82
		Female	17.16
Emergency situation	v_1 (km/h)	1	11.89
		2	20.24
		3	29.17
	TTC_c (s)	1	21.56
		2	21.84
		3	18.12
	Adjacent gap THW (s)	1	10.89
		2	23.01
		3	28.03
Average		—	20.51

The BRK decision includes a braking maneuver followed by an LC, which accounted for only 3.5%; the LC decision includes the LC maneuver followed by a BRK, which accounted for 7.9%. In this study, maneuver prediction is simplified into a binary classification for that only data collected before the timing that maneuver decision is made is available. As can be seen from the overall decision results, compared to the BRK decision, the LC decision rate was only 20.1%, which implies that the LC maneuver in cut-in situations was adopted by a minority of the participants. The collected data (i.e., instances) are extremely imbalanced, and the LC decision instances should be considered in rare cases; however, such cases are relevant to risk-avoidance intervention.

2.2.2. Personality Traits. The participants' personality traits measured by EPQ-RSC are shown in Figure 3. The average N score was quite low and did not significantly vary among the 24 participants. P and E scores covered a wider range than N score. The personality traits also varied by A and G. N, P, and E independent variables were affected by A and G; however, only P was influenced significantly by G ($F(1, 23) = 4.343$; $p = 0.05$).

The LC rate was significantly affected by E ($F(7, 200) = 3.38$; $p = 0.002$), N ($F(4, 203) = 3.384$; $p = 0.01$), and P ($F(7, 200) = 2.197$; $p = 0.036$). Thus, all personality traits, A, and G should be considered when selecting candidate features for maneuver prediction.

2.2.3. Original Signals and Instance Extraction. The original signals including steering angle (°), speed (km/h), throttle and brake pedal positions (%), acceleration (m/s^2), and

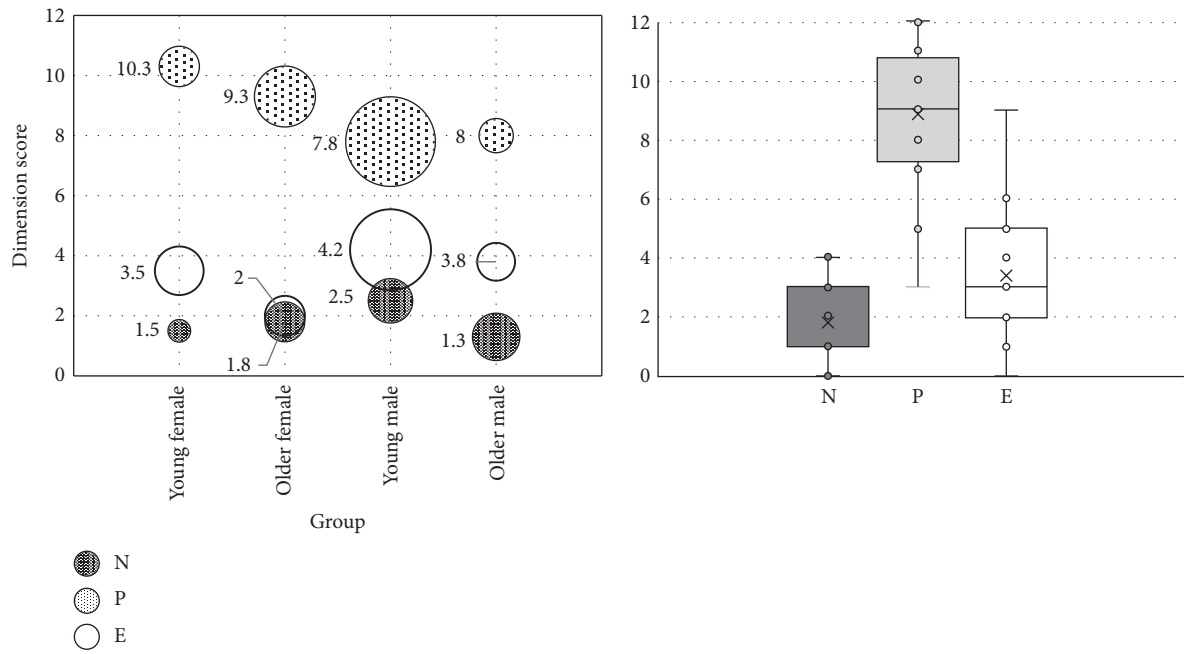

Figure 3: EPQ-RSC measures (N: neuroticism; P: psychoticism; E: extraversion; bubble width indicates one unit of standard deviation).

location (m, x; m, y) of the host vehicle and surrounding vehicles were logged by the driving simulator.

A single cut-in event was defined as starting 3 s prior to the movement of the cut-in vehicle and ending 3 s after the head direction of the cut-in vehicle returned to its initial state. An example of Veh S signals recorded in a BRK situation is shown in Figure 4. The movement start of the cut-in vehicle is defined as the timing where the change of Veh C lateral position exceeds 0.1 m.

2.2.4. Candidate Features. The cut-in trajectory is illustrated in Figure 5. The surrounding vehicles, including Veh R, Veh L, and Veh C, are significantly affected by the maneuver decisions of participants in the cut-in scenarios.

To predict the maneuver decision, we focused on the start time of the cut-in vehicle movement. To better understand the maneuver decision, the temporal and spatial relations between the host vehicle and surrounding vehicles should be comprehensively described. Based on that consideration and the results of our statistical analysis, 10 candidate features are proposed (Table 3). As previously mentioned, the experimental scenarios were designed with three independent variables, that is, v_1, TTC_c, and THW, and the range of the objective risk indicators was limited by these three factors.

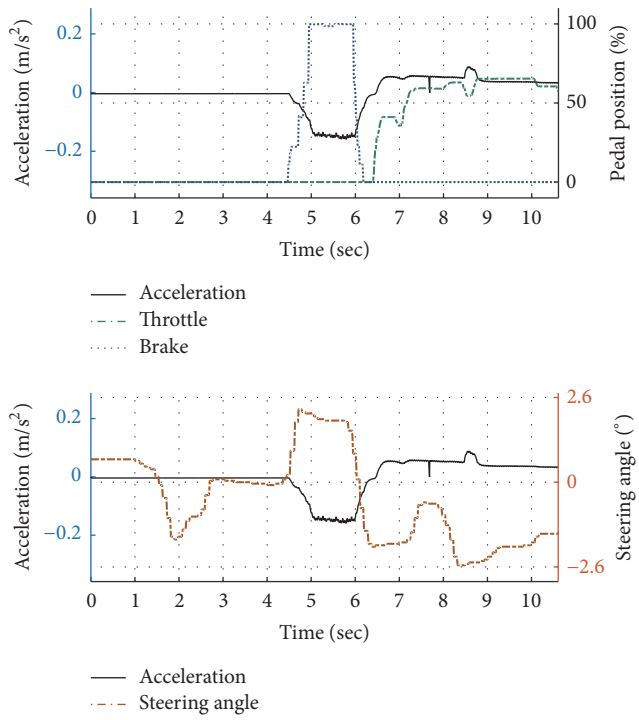

Figure 4: Veh S signals recorded in a BRK situation.

3. Classifier Construction

3.1. DT Structure. In a DT, the decision result can be traced from the top root node to bottom leaf nodes. Initially, all data are stored in the root node. The data are partitioned based on a series of binary splits. There are two types of DTs, that is, a classification tree—where the target decision is categorical—and a regression tree—where the predicted decision is numerical. In this study, we employ a classification DT.

The DT classifies instances by recursively partitioning the predictor space. Classification and regression trees (CART) [40] are commonly used in data mining. Typically, tree methods are nonlinear and nonparametric and are easy

Table 3: Candidate features (calculated based on the start time of the cut-in vehicle movement).

#	Feature	Description	Type
1	D_F	Distance between Veh L and Veh S	Objective risk indicators
2	D_L	Distance between Veh R and Veh S	
3	TTC_F	TTC between Veh L and Veh S	
4	TTC_F	TTC between Veh R and Veh S	
5	V_s	Velocity of Veh S	
6	N	Neuroticism	Driver characteristics
7	P	Psychoticism	
8	E	Extraversion	
9	A	Age group	
10	G	Gender group	

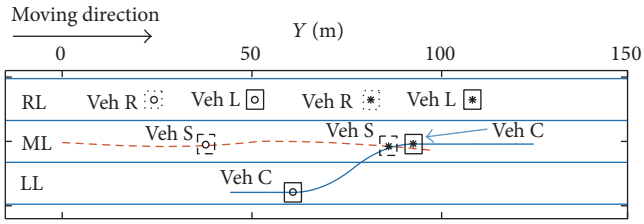

Figure 5: Cut-in trajectory in a BRK situation.

to interpret. In DT-based classification, a hierarchical tree structure is established by a certain split criterion to represent the data set with smaller and more homogeneous groups. To investigate driver decision strategies, DT-based models can predict maneuvers for new cases. Moreover, such models can reveal patterns behind driver rear-end risk-avoidance behaviors in cut-in scenarios.

The CART algorithm has two steps, that is, tree growing and pruning. In this study, the classification trees were produced based on the CART algorithm using the Gini diversity index as the split criterion. This index measures the impurity of a data partition (D) as follows:

$$\text{Gini}(D) = 1 - \sum_{i=1}^{m} p_i^2, \tag{1}$$

where p_i is the probability that a partition in D belongs to class C_i estimated by $|C_{i,D}|/|D|$ and the sum is computed over m classes [41]. The reduction in impurity that would be incurred by a binary split on a possible split-point of A is expressed as follows.

$$\Delta\text{Gini}(A) = \text{Gini}(D) - \text{Gini}_A(D). \tag{2}$$

The Gini diversity index minimizes the impurity of a node group by finding the best split-variable value. In our study, to set the prune criterion, one standard error was adopted as the specified maximum difference in risk. The minimum number of cases for the parent nodes was 10, and the minimum number of cases for the child nodes was 1. To avoid overfitting, the DT structure was automatically pruned to the smallest subtree based on two standard errors as the specified maximum difference in risk. A series of IF–THEN classification rules can be extracted from the CART structure. Such IF–THEN rules reveal the driver maneuver decision pattern and can be easily understood and applied. In addition to extracted rules, variables can be ranked in order of importance and applied to identify the key predictive features. The variable importance VI of X (i.e., VI(X)) is expressed in terms of a normalized quantity relative to the variable having the highest measure of importance [42]. The importance value ranges from 0 to 100. The variable with the greatest importance is scored as 100 as follows:

$$\text{VI}(X) = \frac{\sum_{t\in T} \Delta\text{Gini}(A_t, X)}{\max_X \sum_{t\in T} \Delta\text{Gini}(A_t, X)} \times 100, \tag{3}$$

where T denotes the number of DT nodes and t denotes the specific node upon which the variable X works as the split factor.

Using the candidate features shown in Table 3, a single binary classification tree model was developed to predict the driver maneuver decisions in the cut-in scenarios. In the DT maneuver decision model, the classification result gives the predicted maneuver decision, denoted as MA_DEC (BRK = 0; LC = 1). In this study, IBM SPSS Statistics 19 was used for data training and testing.

3.2. Undersampling and Model Validation. As previously mentioned, the collected data set was imbalanced (Section 2.2.1). However, rare cases are often of great interest and great value [43], which is true in this study. In particular, the LC decision deserves more attention due to its greater complexity. To avoid reducing classifier performance due to such imbalance, an undersampling method was adopted for the model training.

Among 1326 valid trials, there were 267 instances of LC decisions. The undersampling process randomly extracted 267 additional cases from 1059 BRK decision cases. Thus, the undersampled data set had 534 instances (50% LC, 50% BRK). Two-thirds of the data set were used to train the model, and one-third was used to validate the model via randomized extraction. To better evaluate the constructed model, the remaining 792 BRK decision cases were also tested.

Table 4: DT classification rules.

Node	Rules: IF...	THEN	Probability
3	IF (D_F <= 17.00) AND (TTC_F <= 5.83)	BRK	89.2%
8	IF (D_F <= 17.00) AND (TTC_F > 5.83) AND (A == 1)	BRK	79.2%
13	IF (D_F <= 17.00) AND (TTC_F > 5.83) AND (A == 0) AND (N <= 2.5)	BRK	100%
14	IF (D_F <= 17.00) AND (TTC_F > 5.83) AND (A == 0) AND (N > 2.5)	LC	72.2%
11	IF (D_F > 17.00) AND (E > 5.5) AND (G == 1)	LC	91.3%
12	IF (D_F > 17.00) AND (E > 5.5) AND (G == 0)	BRK	87.5%
18	IF (D_F > 17.00) AND (E <= 5.5) AND (A == 1) AND (N > 1.5)	BRK	83.7%
16	IF (D_F > 17.00) AND (E <= 5.5) AND (A == 0) AND (N > 2.5)	LC	85.7%
19	IF (D_F > 17.00) AND (E <= 5.5) AND (A == 0) AND (N <= 2.5) AND (D_F <= 25.94)	BRK	92.3%
20	IF (D_F > 17.00) AND (E <= 5.5) AND (A == 0) AND (N <= 2.5) AND (D_F > 25.94)	LC	57.5%
22	IF (D_F > 17.00) AND (E <= 5.5) AND (A == 1) AND (N <= 1.5) AND (E > 3.5)	BRK	100%
21	IF (D_F > 17.00) AND (A == 1) AND (N <= 1.5) AND (E <= 3.5)	LC	65.9%

Table 5: Classification results.

Maneuver decision	Train data set		Test data set	
	Number	Correct classified	Number	Correct classified
BRK	178	122 (68.5%)	89	67 (75.3%)
LC	178	160 (89.9%)	89	76 (85.4%)
In total	356	282 (79.2%)	178	143 (80.3%)

4. Results

4.1. DT-Based Classification Rules and Prediction Performance. The DT diagram (23 terminal nodes; 21 nodes) for the maneuver prediction in a cut-in situation is shown in Figure 6. The classification rules generated by the DT-based model for the maneuver decision are shown in Table 4. Note that the probabilities of the extracted rules range from 57.5% to 100%. In Table 4, the root variable, that is, the objective risk indicator, is the distance between the host vehicle and lead vehicle in the LL (D_F). If D_F ≤ 17 m, the time to collision between the host vehicle and the lead vehicle in the LL (TTC_F), A, and N score are involved in the driver maneuver prediction. If D_F > 17 m, the maneuver decision depends on E score, A, and N score. As D_F increases, LC becomes increasingly difficult. This root variable indicates that driver maneuver preference only matters when the perceived risk is not too high.

To predict the maneuver decision when D_F ≤ 17 m, more objective indicators must be involved. For example, in Node 3, if D_F ≤ 17 m and TTC_F is less than 5.83 s, the probability of BRK reaches 89.2% for all participants. Note that the classification rule for Node 3 is the only rule that does not include any measures of driver characteristics. If TTC_F > 5.83 s, the maneuver decision is affected by A and N. In Nodes 8 and 13, where the LC risk is not as high as in Node 3, both older and younger drivers with relatively lower N scores are predicted to make a BRK decision. Only young drivers with an N score > 2.5 will choose LC (Node 14).

When D_F > 17 m, more nodes are required in prediction, which indicates that the rules determining driver maneuver preference are complicated and are dominated by driver characteristics. In this situation, if drivers have an E score > 5.5, males will choose LC and females will choose BRK (Nodes 11 and 12). For drivers with an E score < 5.5, older drivers with an N score > 1.5 tend to brake (Node 18). However, the decisions made by older drivers with an N score < 1.5 are affected by their E scores (Nodes 21 and 22). Drivers with an E score > 3.5 tend to choose BRK, while those with an E score < 3.5 tend to choose LC. Predicting young drivers' maneuver decisions is complex (Nodes 16, 19, and 20). Node 16 implies that a high N score indicates a greater probability of LC. For young drivers with a low N score, if the perceived LC risk is acceptable, then they tend to choose LC (Node 20, D_F > 25.94).

To summarize the extracted rules, the LC decision correlates to low risk, male drivers, young drivers, high N score, and high E score.

The classification results for the constructed DT model are shown in Table 5. The overall correct rate of the established DT model reached 79.2% for the training data set and 80.3% for the test data set. As for the remaining 792 BRK decision cases, the number of correct cases was 627 (64.6%). It indicates that the extracted classification rules are reasonable and applicable.

4.2. Variable Importance for Maneuver Prediction. Variable importance produced by CART can be applied to identify the key features that can be used to understand and predict driver maneuver decisions in cut-in scenarios. The normalized importance of candidate features is shown in Figure 7. According to the extracted rules for maneuver prediction, E score, D_F, N score, G, TTC_F, and A are the features of the constructed model. It is evident that driver characteristics, that is, E and N, strongly influence the maneuver decision.

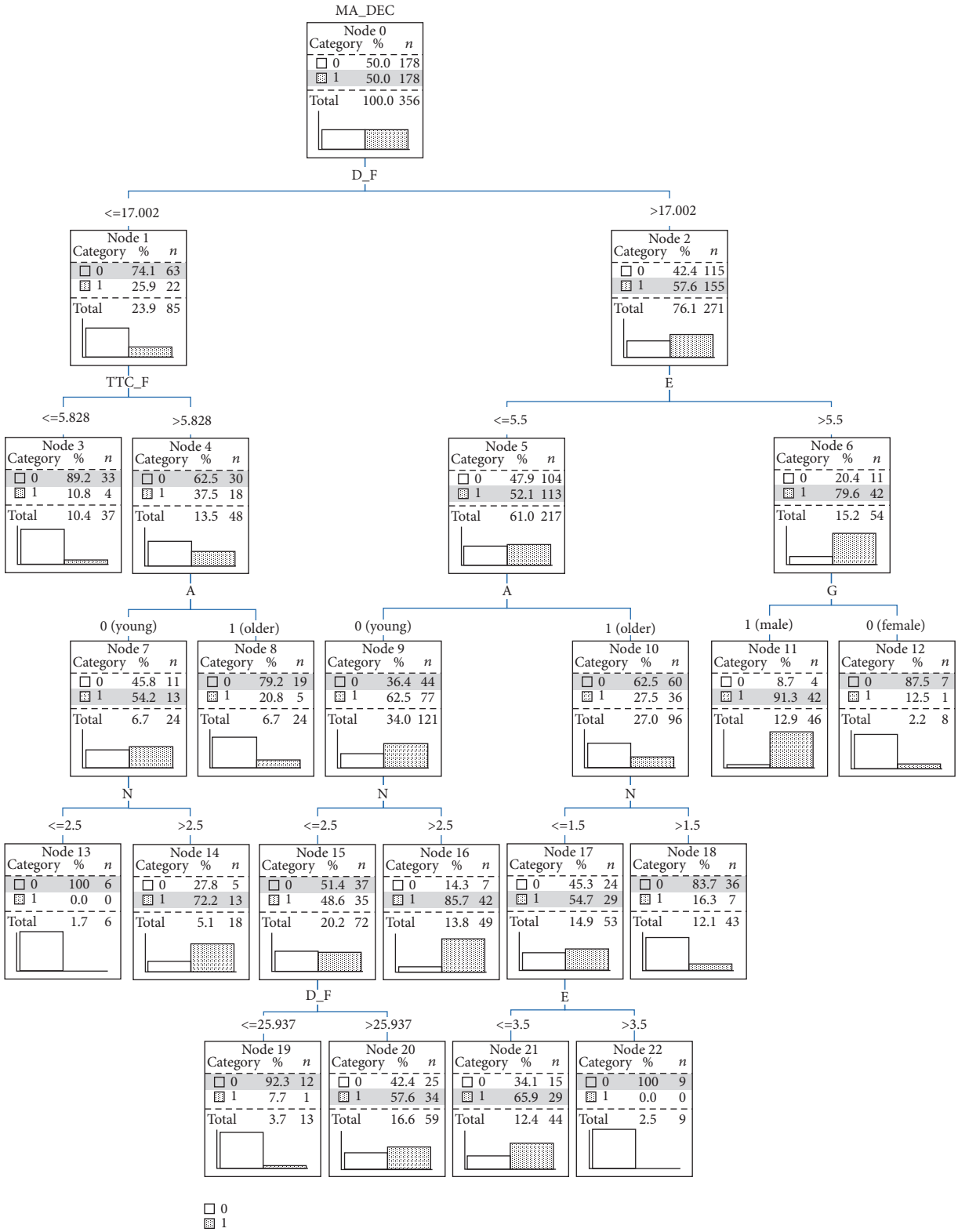

Figure 6: DT diagram for the maneuver decision model for cut-in scenarios.

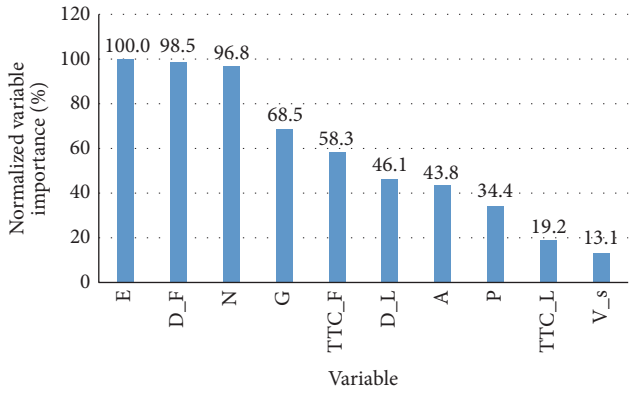

Figure 7: Ranked importance of the independent variables in the maneuver decision model.

The triggering factor is D_F, that is, the perceived main indicator of the LC risk. TTC_F is an indicator of imminent collision risk between the host vehicle and lead vehicle in the LL. Both TTC_F and D_F profile the risk and difficulty behind the LC maneuver. As mentioned above, E and N scores are important predictive variables, particularly when the perceived LC risk is widely recognized as low. G and A also affect the maneuver decision significantly. With the exception of those with a relatively high E score, female drivers tend to avoid the LC maneuver. Note that the effects of A are more complicated than those of G. Generally, older drivers are more likely to brake than younger drivers.

5. Discussion

In this simulator-based study, driver rear-end risk-avoidance behaviors in cut-in scenarios on a straight three-lane highway have been explored. Based on a statistical analysis, candidate features were identified in combination with objective risk indicators and driver characteristics. A DT-based model was constructed to predict driver maneuvers in cut-in scenarios. The prediction accuracy of the extracted classification rules was 79.2% for the training data set and 80.3% for the test data set. The most powerful predictive variables were extracted, and their effects on maneuver decisions were analyzed.

The proposed DT-based model performs well for both BRK and LC decisions with reasonable candidate features. The DT diagram for cut-in scenario maneuver prediction has 12 terminal nodes/extracted rules and 21 nodes. The root variable is the D_F, which is an objective risk indicator. Driver maneuver preference matters only when the perceived risk is not too high. In this study, D_F is considered an important risk indicator for driver maneuver decision-making. In addition to D_F and E and N scores, G, TTC_F, and A are included in the final classification tree. Based on the extracted rules, the LC decision correlates to low driving risk, male drivers, young drivers, high N, and high E. The results of the statistical analysis based on the orthogonal experiment data show that the LC rate was significantly affected by A ($F(1, 206) = 6.79$; $p = 0.01$), G ($F(1, 206) = 2.552$; $p = 0.112$), E ($F(7, 200) = 3.38$; $p = 0.002$), N ($F(4, 203) = 3.384$; $p = 0.01$), and P ($F(7, 200) = 2.197$; $p = 0.036$). The VI results produced by the CART algorithm are consistent with the results obtained by the statistical analysis.

Seen from accident data, steering is associated with higher crash severity. As indicated by a previous study based on NASS GES database for the years 2005–2009 in United States, compared to nonavoidance maneuver, steering increases crash severity from 30% in damage only accidents to more than 60% in fatal accidents [44]. Many factors, including road, driver, reaction time, TTC, and hazard perception, contribute to the selection of steering maneuver. Hence, the prediction of driver's maneuver decision is challenging and it requires an integrated angle.

Drivers' maneuver decisions in cut-in scenarios are strongly affected by both driving environment and driver characteristics. The decision preference is dominated by driver characteristics within a perceived safe zone. Once the perceived driving risk is above a certain level, most drivers will not choose to change lanes. The driver decision preference only matters in the safe zone, where decisions are strongly affected by the driver's individual characteristics, including personality traits. Consideration of driver preference can contribute to good user experience in comfortable automated driving. It was also summarized that results based on manual data are a good indicator of the automated driving style perceived to be comfortable [15]. In addition, this study implies that the boundary of the perceived safe zone is also part of driver preference.

As indicated by the extracted important variables, driver characteristics strongly influence the prediction of driver maneuver decisions in cut-in scenarios. LC behavior in cut-in scenarios is widely recognized as an indication of aggressive driving as it involves greater risk [19]. The correlation between extraversion and openness toward new experience, as well as the aggressive behavior of drivers, has been previously reported [35, 45, 46]. Drivers feel anger and express aggression while driving, while N is found to indirectly correlate with aggressive driving through anger [47]. Drivers with high N scores and high E scores may be easy to provoke and may choose to change lanes when stimulated by the cut-in vehicle, which has been confirmed by the rules extracted by the proposed DT-based model.

The extracted classification rules indicate that the effect of personality measures on the driver maneuver decision varies according to different combinations of A and G. Female drivers, even with high E score, tend to brake. Older drivers, even with relatively higher N scores and average E scores, also tend to brake. These results indicate that female and older drivers share the same pattern relative to the maneuver decision in cut-in scenarios. A previous study found that the youngest drivers had the highest rate of involvement in all police-reported crashes, and men had a higher risk than women of experiencing a fatal crash [36, 48, 49]. Those results indicate that young male drivers tend to aggressively drive, which has also been confirmed by this study.

The results of this study contribute to user-appropriate design of HAD systems that are developing from operational intervention into earlier stage. Hence, in this study, driver decision strategy of risk-avoidance in cut-in scenario was explored in the aspects of understanding and prediction. The

results of this study can be applied to create user-adaptive automated driving program for semiautonomous driving. For full autonomous driving, since that one autonomous vehicle would never be isolated from the whole traffic system, automatic systems, which enable autonomous driving, need to consider surrounding vehicles of lower capability than them. Therefore, in that context, how human would avoid the situation in this study should be considered in the decision-making of such systems.

One limitation of this study was the small sample size (24 participants). The risk of overfitting existed; however, some countermeasures were applied, for example, downsampling, tree pruning, counterbalanced experiment design, and completely balanced A and G factors. Due to the limited sample size, the personality traits measured by EPQ-RSC were not well balanced between the various participants. The other limitation of this study is using driving simulator to imitate real traffic situations. Driver behaviors obtained from driving simulator and real traffic have differences; however, the impact tendency of specific factors should be the same. Applying the model produced in this study, driver behaviors in both driving simulator and real traffic can be described, although with different parameters. And that is why we should adjust the model parameters based on real traffic condition.

6. Conclusions

We have proposed and tested a DT-based driver maneuver decision prediction method. The independent predictive features were a combination of objective risk indicators and driver characteristic features. The following conclusions were obtained:

(1) The proposed DT-based model achieves the correct rate of 79.2% for the training data set and 80.3% for test data set, indicating that the extracted classification rules are reasonable and applicable.

(2) Driver decision preference matters only in a perceived driving safe zone, where decisions are strongly affected by driver characteristics, particularly personality traits.

(3) Driver characteristics are a strong influence in the prediction of maneuver decisions. The lane-change decision is frequently observed in low-risk driving conditions, male drivers, young drivers, drivers with high N scores, and drivers with high E scores.

The results of this study contribute to user-appropriate design of HAD systems that are developing from operational intervention into early stage. In the future, we will focus on improving method in a real driving environment. To further assess the generality and functionalities required for the application of this improved method, actual driving data will be collected. In addition, studies having more participants should be conducted to obtain a deeper understanding of the effects of personality traits on risk-avoidance decision-making.

Abbreviations

NHTSA:	National Highway Traffic Safety Administration
EPQ-RSC:	Eysenck Personality Questionnaire revised for Chinese
DT:	Decision tree
TTC:	Time to collision
THW:	Time headway
LL:	Left lane
ML:	Middle lane
RL:	Right lane
LC:	Lane-changing
BRK:	Braking
CART:	Classification and regression trees
M:	Mean value
SD:	Standard deviation
MA_DEC:	DT model classification result (BRK = 0; LC = 1).

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

The first two authors, Manjiang Hu and Yuan Liao, have equally contributed to this research work.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant no. 51505247) and International Science & Technology Cooperation Program of China (Contract no. 2016YFE0102200).

References

[1] J. C. F. De Winter, R. Happee, M. H. Martens, and N. A. Stanton, "Effects of adaptive cruise control and highly automated driving on workload and situation awareness: a review of the empirical evidence," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 27, pp. 196–217, 2014.

[2] Daimler, 2013, https://www.daimler.com/innovation/autonomous-driving/special/technology-cars.html.

[3] Volvo, 2013, https://www.media.volvocars.com/global/en-gb/media/pressreleases/49875/volvo-cars-reveals-world-class-safety-and-support-features-to-be-introduced-in-the-all-new-xc90-in-2.

[4] A. Eckert, B. Hartmann, M. Sevenich, and P. Rieth, "Emergency steer & brake assist: a systematic approach for system integration of two complementary driver assistance systems," in *Proceedings of the 22nd International Technical Conference on the Enhanced Safety of Vehicles*, Washington, DC, USA, June 2011.

[5] T. Dang, J. Desens, U. Franke, D. Gavrila, L. Schäfers, and W. Ziegler, "Steering and evasion assist," in *Handbook of Intelligent Vehicles*, A. Eskandarian, Ed., pp. 759–782, Springer, New York, NY, USA, 2012.

[6] M. Itoh, T. Horikome, and T. Inagaki, "Effectiveness and driver acceptance of a semi-autonomous forward obstacle collision avoidance system," *Applied Ergonomics*, vol. 44, no. 5, pp. 756–763, 2013.

[7] M. Heesen, M. Dziennus, T. Hesse et al., "Interaction design of automatic steering for collision avoidance: challenges and potentials of driver decoupling," *IET Intelligent Transport Systems*, vol. 9, no. 1, pp. 95–104, 2015.

[8] M. Brännström, E. Coelingh, and J. Sjöberg, "Decision-making on when to brake and when to steer to avoid a collision," *International Journal of Vehicle Safety*, vol. 7, no. 1, pp. 87–106, 2014.

[9] A. Eidehall, J. Pohl, F. Gustafsson, and J. Ekmark, "Toward autonomous collision avoidance by steering," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 1, pp. 84–94, 2007.

[10] D. Fernandez Llorca, V. Milanes, I. Parra Alonso et al., "Autonomous pedestrian collision avoidance using a fuzzy steering controller," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 390–401, 2011.

[11] J. Shah, R. Schaefer, and M. C. Best, "Integrated collision avoidance by active intervention of brake and steer for intelligent vehicles," in *Proceedings of the FISITA 2014 World Automotive Congress*, Maastricht, The Netherlands, June 2014.

[12] D. A. Norman, "The 'problem' with automation: inappropriate feedback and interaction, not 'over-automation'," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 327, no. 1241, pp. 585–593, 1990.

[13] M. Maurer, J. C. Gerdes, B. Lenz, and H. Winner, *Autonomous Driving*, Springer Berlin Heidelberg, Berlin, Germany, 2016.

[14] A. Rosenfeld, Z. Bareket, C. V. Goldman, D. J. LeBlanc, and O. Tsimhoni, "Learning drivers' behavior to improve adaptive cruise control," *Journal of Intelligent Transportation Systems*, vol. 19, no. 1, pp. 18–31, 2015.

[15] H. Bellem, T. Schönenberg, J. F. Krems, and M. Schrauf, "Objective metrics of comfort: developing a driving style for highly automated vehicles," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 41, pp. 45–54, 2016.

[16] L. D. Adarns and P. Place, "Review of the literature on obstacle avoidance maneuvers: braking versus steering," Report UMTRI-94-19, The University of Michigan Transportation Research Institute, Ann Arbor, Mich, USA, 1994.

[17] R. J. Kiefer, D. J. Leblanc, and C. A. Flannagan, "Developing an inverse time-to-collision crash alert timing approach based on drivers' last-second braking and steering judgments," *Accident Analysis and Prevention*, vol. 37, no. 2, pp. 295–303, 2005.

[18] V. Venkatraman, J. D. Lee, and C. W. Schwarz, "Steer or brake? Modeling drivers' collision avoidance behavior using perceptual cues," in *Proceedings of the Transportation Research Board 95th Annual Meeting*, No. 16-6657, 2016.

[19] A. Schieben, S. Griesche, T. Hesse, N. Fricke, and M. Baumann, "Evaluation of three different interaction designs for an automatic steering intervention," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 27, pp. 238–251, 2014.

[20] A. Jahangiri, H. Rakha, and T. A. Dingus, "Red-light running violation prediction using observational and simulator data," *Accident Analysis & Prevention*, vol. 96, pp. 316–328, 2016.

[21] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*, MIT Press, Cambridge, Mass, USA, 2005.

[22] A. Jain, H. S. Koppula, B. Raghavan, and A. Saxena, *Know Before You Do: Anticipating Maneuvers via Learning Temporal Driving Models*, Department of Computer Science, Cornell University, Ithaca, NY, USA, 2015.

[23] M. Kantardzic, *Data Mining: Concepts, Models, Methods, and Algorithms*, John Wiley & Sons, 2nd edition, 2011.

[24] W. G. Najm, J. D. Smith, and M. Yanagisawa, *Pre-Crash Scenario Typology for Crash Avoidance Research*, DOT HS, 2007.

[25] N. Strand, J. Nilsson, I. C. M. Karlsson, and L. Nilsson, "Exploring end-user experiences: self-perceived notions on use of adaptive cruise control systems," *IET Intelligent Transport Systems*, vol. 5, no. 2, pp. 134–140, 2011.

[26] A. F. L. Larsson, "Driver usage and understanding of adaptive cruise control," *Applied Ergonomics*, vol. 43, no. 3, pp. 501–506, 2012.

[27] A. F. L. Larsson, K. Kircher, and J. A. Hultgren, "Learning from experience: familiarity with ACC and responding to a cut-in situation in automated driving," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 27, pp. 229–237, 2014.

[28] D. J. Beirness, "Do we really drive as we live? The role of personality factors in road crashes," *Alcohol Drugs Driving*, vol. 9, no. 3, pp. 129–143, 1993.

[29] H. J. Eysenck and S. B. G. Eysenck, *Manual of the Eysenck Personality Questionnaire (Junior and Adult)*, Hodder and Stoughton, 1975.

[30] S. B. G. Eysenck, H. J. Eysenck, and P. Barrett, "A revised version of the psychoticism scale," *Personality and Individual Differences*, vol. 6, no. 1, pp. 21–29, 1985.

[31] Q. Mingyi, W. Guocheng, Z. Rongchun, and Z. Shen, "Development of the revised Eysenck Personality Questionnaire short scale for Chinese (EPQ-RSC)," *Journal of Chinese Psychology Acta Psychological Sinica*, vol. 32, no. 3, pp. 317–323, 2000.

[32] C. R. Bartol and A. M. Bartol, *Criminal Behavior: A Psychosocial Approach*, Prentice Hall, Upper Saddle River, NJ, USA, 1999.

[33] S. B. Moeller, P. Bech, L. Kessing, E. L. Mortensen, S. F. Austin, and J. D. Bukh, "A psychometric validation analysis of Eysenck's neuroticism and extraversion scales in a sample of first time depressed patients," *Journal of Depression & Anxiety*, vol. 4, no. 4, Article ID 1000202, 2015.

[34] H. J. Eysenck, *Neuroticism and Psychoticism. Individual Differences and Psychopathology: Physiological Correlates of Human Behaviour*, 2016.

[35] E. R. Dahlen, B. D. Edwards, T. Tubré, M. J. Zyphur, and C. R. Warren, "Taking a look behind the wheel: an investigation into the personality predictors of aggressive driving," *Accident Analysis and Prevention*, vol. 45, pp. 1–9, 2012.

[36] J. Yang, F. Du, W. Qu, Z. Gong, and X. Sun, "Effects of personality on risky driving behavior and accident involvement for chinese drivers," *Traffic Injury Prevention*, vol. 14, no. 6, pp. 565–571, 2013.

[37] R. Elvik, "Speed limits, enforcement, and health consequences," *Annual Review of Public Health*, vol. 33, pp. 225–238, 2012.

[38] J. I. Castillo-Manzano and M. Castro-Nuño, "Driving licenses based on points systems: efficient road safety strategy or latest fashion in global transport policy? A worldwide meta-analysis," *Transport Policy*, vol. 21, pp. 191–201, 2012.

[39] Y. Liao, S. E. Li, W. Wang, Y. Wang, G. Li, and B. Cheng, "Detection of driver cognitive distraction: a comparison study of stop-controlled intersection and speed-limited highway," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 6, pp. 1628–1637, 2016.

[40] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olsen, *Classification and Regression Trees*, CRC Press, 1984.

[41] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*, Elsevier, Amsterdam, The Netherlands, 2011.

[42] IBM Knowledge Center, 2016, http://www.ibm.com/support/knowledgecenter/SS3RA7_15.0.0/kc_gen/com.ibm.spss.modeler.help_com.ibm.spss.model-er.help_toc-gen28.html.

[43] G. M. Weiss, "Mining with rarity: a unifying framework," *ACM SIGKDD Explorations Newsletter*, vol. 6, no. 1, pp. 7–19, 2004.

[44] S. Kaplan and C. G. Prato, "Observed and unobserved correlation between crash avoidance manoeuvers and crash severity," *International Journal of Injury Control and Safety Promotion*, vol. 23, no. 4, pp. 413–426, 2016.

[45] J. A. Benfield, W. J. Szlemko, and P. A. Bell, "Driver personality and anthropomorphic attributions of vehicle personality relate to reported aggressive driving tendencies," *Personality and Individual Differences*, vol. 42, no. 2, pp. 247–258, 2007.

[46] R. Fernandes, R. F. S. Job, and J. Hatfield, "A challenge to the assumed generalizability of prediction and countermeasure for risky driving: different factors predict different risky driving behaviors," *Journal of Safety Research*, vol. 38, no. 1, pp. 59–70, 2007.

[47] D. Jovanović, K. Lipovac, P. Stanojević, and D. Stanojević, "The effects of personality traits on driving-related anger and aggressive behaviour in traffic among Serbian drivers," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 14, no. 1, pp. 43–53, 2011.

[48] D. L. Massie, K. L. Campbell, and A. F. Williams, "Traffic accident involvement rates by driver age and gender," *Accident Analysis & Prevention*, vol. 27, no. 1, pp. 73–87, 1995.

[49] W. Qu, Y. Ge, C. Jiang, F. Du, and K. Zhang, "The dula dangerous driving index in China: an investigation of reliability and validity," *Accident Analysis & Prevention*, vol. 64, pp. 62–68, 2014.