IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received July 20, 2021, accepted August 17, 2021, date of publication September 10, 2021, date of current version September 21, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3111753

Variable Selection and Modeling of Drivers' Decision in Overtaking Behavior Based on Logistic Regression Model With Gazing Information

JUDE C. NWADIUTO [1], SOICHI YOSHINO [2], HIROYUKI OKUDA [1], (Member, IEEE), AND TATSUYA SUZUKI [1], (Member, IEEE)

[1]Department of Mechanical Systems Engineering, Nagoya University, Nagoya 464-8601, Japan
[2]Toyota Research Institute-Advanced Development, Tokyo 103-0022, Japan

Corresponding author: Jude C. Nwadiuto (nwadiuto.jude.chibuike@b.mbox.nagoya-u.ac.jp)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethical Committee of Graduate School of Engineering, Nagoya University under Application No. 18-6, and performed in line with the Declaration of Helsinki.

ABSTRACT This paper investigates the decision-making characteristics of the driver in the overtaking on the highway road. For the research purpose, a novel method was proposed by introducing a logistic regression model accompanied by the statistical test technique, which does not require prior knowledge about the explanatory variables. This study hypothesizes that the driver's gazing behavior is crucial for the decision-making process in driving and hence, the line-of-sight information was introduced to estimate driver's gazing behavior in the model of driver's decision specifically for reproducing the overtaking driving behavior accurately. Consequently, the proposed model realized a high describability on the decision of the driver when performing the overtaking driving task, which is one of the significant advancements of the present study with respect to the past similar studies. This study integrates the perspectives of intelligent vehicle design and cognitive science by revealing which factor the driver pays attention to in a changeable driving environment due to various observable factors. In experiments based on the driving simulator with six human subjects, the overtaking behavior was successfully estimated by specifying a set of variables to reconstruct the driver's behavior and then the proposed model provided a minimum set of necessary variables accompanied with key coefficients. In conclusion, the proposed approach based on a simple logistic regression model demonstrated driving behaviors with an accurate estimation of the driver's intention without the need for prior knowledge, and it may contribute to higher describability for various driving actions in a dynamic environment.

INDEX TERMS Overtaking behavior, decision-making, logistic regression, model selection, statistical test, gazing behavior, line-of-sight information.

I. INTRODUCTION

The automotive industry is currently experiencing significant advances in autonomous driving and advanced driving assistance systems (ADAS) [1]. Autonomous driving have not only advanced from a technological point of view but also from a regulatory point of view [2]. These advances have been spurred on by corresponding advances in sensor technology [3]. The fusion of sensors such as LiDAR (Light Detection and Ranging), radar, camera, ultrasonic, GPS (Global Positioning System), and IMU (Inertial Navigation System) have made automated driving possible. According to the US Society of Automotive Engineers, there are six levels of automation in autonomous driving [4]. To the best of our knowledge, as of 2019, level 3 automation has been the maximum level of automation of commercially available vehicles. However, in places like the US, Netherlands, Germany, and the UK, testing of level 6 (full) automation

The associate editor coordinating the review of this manuscript and approving it for publication was Camelia Delcea.

is allowed under certain conditions. The emergence of open-source driving simulators that support flexible specification of sensor suites and environmental conditions such as the CARLA Simulator [5], and LGSVL Simulator [6] have also contributed to the rapid development of autonomous driving. These simulators give full control of the traffic actors, thus making it possible for researchers to easily simulate different traffic scenarios for testing automated solutions.

Based on these advances, the next step in the autonomous driving technology will be the mixed autonomous driving (i.e., a mixture of human-driven vehicles and automated vehicles), as well as how to reflect personalized driver characteristics and preferences in automated vehicles. Automated vehicles and ADAS need to understand and predict the cognitive and behavioral aspects of the human driver especially for complex driving tasks. This means that driving behavior for the complex driving task must be carefully analyzed and modeled. In other words, developing driver models for the difficult driving tasks, such as overtaking driving maneuver, that are mathematically rigorous enough, and are capable of expressing the dynamical characteristics of the driving behavior will be the key to meeting this advancement.

From the viewpoint of control technology, numerous ideas have been proposed to model the human driving behavior [8]–[16]. In those works, to analyze the driving behavior, the idea of regarding the driver as a sort of controller in combination with linear control theory was adopted. However, a linear controller could be limited in the cases that the driver incorporates not only simple reflexive motion but also decision-making in operating the car. Therefore, it is necessary to introduce a new viewpoint, and a possible and plausible factor to improve is the segmentation and the symbolization of driving behavior by focusing on the concept of 'chunk.' This key idea gives the high-level understanding of the human driving behavior. Examples of chunks considering driving behavior are, 'following the leading car,' 'overtaking,' 'turning right,' 'changing the lane,' 'negotiating with vehicles on other lanes,' and 'stopping at the traffic signal'. Many studies have tried to construct a symbolic model of the driving behavior from this point of view by segmenting the observed driving data [17]–[20], [22]–[27]. In [17] a consolidated fuzzy clustering technique was used to classify real-road car-following driving data into different driving regimes. Reference [18] proposed a unified car-following model to simulate different driving scenarios including traffic at intersections and [19] applied Hidden Markov Model (HMM) to recognise and generate driving patterns based on imitation learning of time series driving data. In [20], driving data was segmented into meaningful chunks of driving scenes using double articulation analyzer and [23] proposed clustering of car-following behavior into segments based on state-action variables. The approach used [23] for mode segmentation is similar to the idea used in this study where the physical meaning of each mode is defined based on some state-action variables.

The typical approach to segment the data is to use the data clustering technique, such as the HMM and its extension methods [18]–[21]. In [21] the expectation-maximization algorithm was used to identify a stochastic switched autoregressive exogenous (SS-ARX) model for the human driving behavior. In the studies introduced so far, methods for automatic segmentation of data were mostly emphasized, while the interpretation of the state (mode) transition (switching mechanism) did not attract much attention as an important point.

Another approach to segmentation is to apply the hybrid dynamical system (HDS) model to driving data [24], [25]. In these works, the mode switching condition could be explicitly extracted from the observed data as the driver's decision making. In addition, the physical meaning of each mode could be revealed as the primitive motion model of that mode. Thus, the HDS model leads to the explicit understanding of the driver's decision making together with the driver's motion control.

Recent works have demonstrated the capabilities of the mode segmentation technique for analyzing and modeling the human driving behavior [24], [28]–[32]. A piecewise autoregressive exogenous (PWARX) model was applied in [24], [28]–[31] to model and understand driving behavior, whereas, Okuda *et al.* [32] demonstrated the efficacy of a probability-weighted ARX (PrARX) model in modeling and analyzing driving behavior.

Such modeling techniques have suggested an importance on paying attention to the information processing mechanism when making a decision by the human driver. In principle, various input variables can be considered to explain the decision making process, and therefore the selection of the proper set of input variables is crucial for designing an appropriate driver model. To achieve high accuracy in model estimation, one option is to use all the explanatory variables from one mode to another. However, this has a risk of increasing the number of parameters extensively. On the other hand, a radical simplification of explanatory variables harms the flexibility of the model, as well as model performance. Since the simplification of the model has a large benefit for the reduction of computational costs and facilitation towards online application, the treatment of explanatory variables is a trade-off issue and requires a careful consideration when used for switching from one mode to another.

Based on this background, the authors have proposed a new approach to model the mode switching condition in the driving behavior without any prior knowledge about the selection of the input variables in [33]. According to the result by Okuda *et al.* [33], a logistic regression model was successfully demonstrated to reproduce the mode switching in driving behaviors. The logistic regression model can represent a relationship between the continuous/discrete input variables and the binary output variable. The mode switching event is expressed as the binary output variable, and the probability of the occurrence is modeled by the logistic function with the input (explanatory) variables. This mathematical feature of the model fits well with the driver's decision making characteristics in a sense of having binary output with

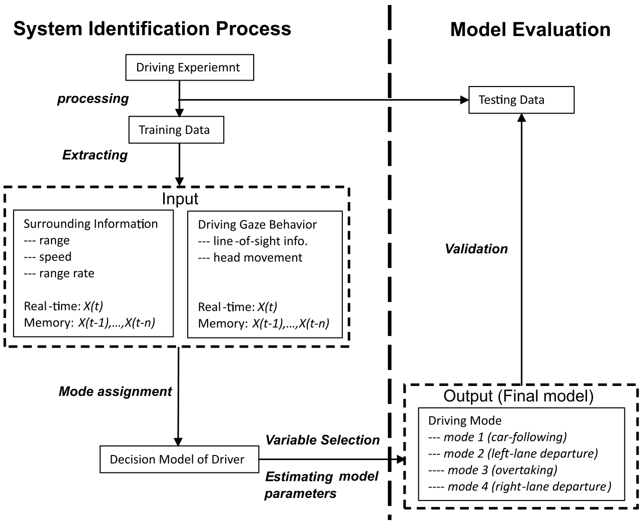

FIGURE 1. Modeling strategy proposed in this paper.

probabilistic uncertainty. In this sense, such a statistical methodology for model selection is valid if the selection of the input variables based on the statistical testing is done properly.

Their results also revealed that the model selection with a proper set of input variables might be incomplete without the driver's gazing behavior and its timing when making a decision. Therefore, the new problem must be addressed as how the driver's gazing behavior can be included in the model. In order to consider the driver's gazing behavior, this paper proposes an extended modeling strategy of [33] as shown in Fig.1, which consist of the system identification process and evaluation of the identified model, where the line-of-sight information for estimating driver's gazing behavior is included in the modeling of driver's decision in the overtaking driving behavior. In order to include the gazing behavior information, the variable selection procedure was required to be reformulated formally and verified in experiments with human subjects. To validate the proposed model from a viewpoint of how well it demonstrates higher describability, the overtaking driving behavior was selected for the verification of the model in the sense of how much driver's decision was properly estimated. The combination of methods presented in this study is suitable for multi-disciplinary applications such as intelligent vehicle design, cognitive science, and statistical analysis.

The rest of this paper is summarised as follows. Section II introduces and discusses the proposed method of this study, the experimental setup and the type of driving data collected were discussed in Section III. Application of the proposed model to overtaking driving data is carried out in Section IV, and Section V presents the results. The verification of the proposed model is performed in Section VI, Section VII discussed some useful applications of the method proposed method in this study. Finally, Section VIII concludes this paper in light of the results and also discusses the limitations of the present study as well as highlighting what could be improved in the future.

II. METHODOLOGY

A. LOGISTIC REGRESSION MODEL

Many methods exist to express the human decision, which can be applied to the human driving behavior. These methods include but not limited to, Linear Discriminant Analysis (LDA), Support Vector Machine (SVM) [25], [36], Fuzzy Logic (FL) [22], [34], [35], and Artificial Neural Network (ANN) [37], [38]. In this paper, the Logistic Regression Model (LRM) [39] is used to mathematically express the driver's decision in switching from one mode to another. Logistic regression is the most suitable approach [40] because the identified model coefficients can be used to judge the impact of each explanatory variable which fits the goal of this paper. In addition, using LRM not only provides a measure of how appropriate a predictor (coefficient size) is, but also its direction of association, thereby, making it suitable for analysing the results of statistical variable selection adopted in this study. The LRM is used as a probabilistic statistical classification model to express the relationship between binary dependent variables, also known as the output variables, and the explanatory variables, also known as the input variables. In other words, the output variables are expressed as a function of input variables. This function is referred to as the 'logistic function', and

$$p(y = y(k)|x = x(k)) = \frac{\exp(\beta_0 + \beta x(k))}{1 + \exp(\beta_0 + \beta x(k))}, \tag{1}$$

is used to compute the posterior probability, $p(y = y(k)| x = x(k))$. Here, β_0 and $\beta (= \{\beta_1, \beta_2, \ldots, \beta_n\})$ represents the parameters of the LRM whereas, $x(k)$ and $y(k)$ are the kth sample of the input variable x and binary or output variables y respectively, where the binary variable y can take the value 0 or 1. $k \in 1, 2, \ldots, K$ indexes the measurement.

If the variable "$y(k)$" is set to 0 for the data points observed prior to switching, and to 1 for the data points observed after switching, the condition for mode switching can be expressed as a LRM. The switching can then be expected to happen in the LRM model when the calculated output probability is greater than 0.5. Furthermore, the estimated parameters β specifies the correlation, put in another way, how sensitive the condition is for the corresponding input variables.

Given that the likelihood function for the LRM is

$$L(\beta_0, \beta) = \prod_{i=1}^{k} p(x(k))^{y(k)} (1 - p(x(k)))^{1-y(k)}, \tag{2}$$

for the estimation of the parameters β_0 and β, the log-likelihood $L\_log(\beta_0, \beta)$,

$$\log\{L(\beta_0, \beta)\} = \sum_{i=1}^{k} -\log 1 + e^{\beta_0 + x(k)\cdot\beta} + \sum_{i=1}^{k} y(k)(\beta_0 + x(k)\cdot\beta) \tag{3}$$

is maximized using a Newton-Raphson method.

The advantages of using the Newton-Raphson method are fast convergence to the local minimum [41], and the easiness to obtain the Hessian matrix, which can be calculated from the square of the standard errors of the coefficients. Furthermore, the Wald test, which is used in this paper, requires the Fisher information matrix of the estimated parameters, which is in fact the Hessian matrix around the maximum likelihood estimation(MLE), and thus is readily available for the variable selection without additional computation cost.

B. SELECTION OF INPUT VARIABLES

A critical but necessary problem usually faced when constructing a good regression model is selecting the appropriate input variables x. Various methods exist for variable selection in LRM. These methods are based on information criteria, penalized likelihood, the change-in-estimate criterion, background knowledge, or combinations of the above mentioned methods [43]. However, the combination of a Wald test and bidirectional elimination method is implemented in this work. This method was chosen because computational cost could be reduced since the Fisher information matrix obtained from the LRM maximization using Newton-Raphson is available and need not to be recalculated. By using this method online application could be facilitated. In principle any method for statistical variable selection could be used.

Bidirectional elimination method tests for variables to be selected or removed using iteration by combining forward selection and backward elimination. Given as follows is a short description of the Bidirectional elimination flow used in this paper.

1) FIRST STEP: INITIALIZATION

Initialize i an iteration step as 0. In this case, at first, the LRM $\mathcal{M}^0$ is made only with a constant term (i.e. without any input variables), where only β_0 is the parameter of the model. The remaining set of candidate input variables are denoted by X^i for the ith iteration. $X^1 = \{x_1, x_2, \ldots, x_n\}$ in the first iteration, with n being the number of candidate input variables. Next, proceed to second step.

2) SECOND STEP: VARIABLE ADDITION

In this second step, the iteration step i is incremented. Check if X^i is an empty set, which means there is no more variables to be added, then proceed to third step. Otherwise, select the variable x_s from X^i with the most significance in this fashion. Compute a new model $\mathcal{M}_j^k$ where $x_j \in \{X^i\}$ is added to the current model $\mathcal{M}^{i-1}$, and compare the index

$$\chi^2_{\beta^i_j} = \frac{(\beta^i_j)^2}{\mathrm{Var}(\beta^i_j)} \approx \frac{(\beta^i_j)^2}{\mathcal{H}^{-1}(\beta^i_j)}. \tag{4}$$

In this case, β^i_s is the coefficients for x_s in $\mathcal{M}_j^k$ estimated by MLE. The Wald statistic, $\chi^2_{\beta^i_s}$, is the ratio of the square of the regression coefficient to the square of the standard error of the coefficient. $\mathcal{H}$, which is the computed Fisher information matrix corresponds to the Hessian matrix of the log likelihood around the MLE. The Wald statistic is commonly known to be asymptotically distributed as a χ^2 distribution. Hence, by setting a suitable threshold α, it can be used as an index to test if an input variable is statistically significant or not by using the condition $\chi^2_{\beta^i_s} > \alpha$. It should be noted that in this paper, α was set to a value of 100 based on trial and error. The variable x_s which maximizes $\chi^2_{\beta^i_j}$ in $\mathcal{M}^i_j$, $(j = 1, 2, \ldots, n)$ is accepted as one of the input variables of the model. If the condition, $\chi^2_{\beta^i_s} > \alpha$, is satisfied, $\mathcal{M}^i_s$ is accepted as $\mathcal{M}^i$ and the current model updated. Finally,

$$X^i = X^{i-1} \backslash x_s \tag{5}$$

is used to update the set of candidate input variables. It should be noted that the '\' here represents a set subtraction. Afterwards, proceed to the third step, otherwise, no variable is accepted, hence, terminate the procedure.

3) THIRD STEP: VARIABLE ELIMINATION

Here, the iteration step i is incremented again. If the current model $\mathcal{M}^{i-1}$ does not have any input variable, go back to the second step. Else, test all the input variables in the current model to verify that there is no variable without significance in the model. To do this, for each input variable x_j of the model $\mathcal{M}^{i-1}$, calculate the Wald statistics $\chi^2_{\beta^i_j}$. Find the input variable that has the minimal value of Wald statistics $\chi^2_{\beta^i_s}$ in all $\chi^2_{\beta^i_j}$. If $\chi^2_{\beta^i_s} > \alpha$ is satisfied, proceed to the second step, else, remove x_s from the set of parameters of the model, and add it back to the set of candidate input variables X^i as follows:

$$X^i = X^{i-1} \cup x_s \tag{6}$$

Then repeat the third step.

FIGURE 2. Driving simulator used for experiment.

III. EXPERIMENT SETUP

A. TARGET TASK AND DEFINITION OF SURROUNDING CARS

This paper focuses on analysing the overtaking driving behavior on an infinite straight expressway with two lanes. The driving experiment was carried out on a virtual driving environment which was implemented in a driving simulator (DS) as shown in Fig. 2. The angle of the front view of the DS is 180 degrees. The DS has three screens with no gaps in between, together with three mirrors, the rear mirror,

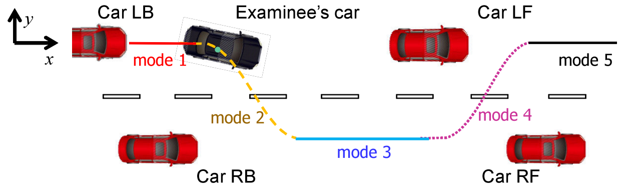

FIGURE 3. Labels of cars and defined tasks considered in this work.

the right door mirror and the left door mirror. Additionally, a real dashboard and HMI, such as the steering, gas/brake pedal are installed.

The speed of the cars on the cruising lane (left-lane) ranges between 70 [km/h] and 90 [km/h] with a distance of 175 [m] between them. Whereas, the running speed of cars on passing lane (right-lane) ranges between 110 [km/hr] and 130 [km/h] at uniform random gap from 1 to 200 [m]. These cars always keep their respective lanes, only the examinee's car changes its lane. During the experiment, the driver is instructed to keep driving on the left-lane unless during the overtaking driving maneuver.

The following describes the label of the surrounding cars according to Fig. 3.

Car LF : Car LF is moving in front of the examinee's car on the left-lane.

Car LB: This is the car behind the examinee's car on the left-lane. The label of Car LF changes to Car LB when the examinee's car overtakes Car LF. This switching of labels occurs in this experiment when the distance between the examinee's car and the overtaken car is 70 [m], thus, enabling the continuity in behavior of the Car LF during the overtake maneuver.

Car RF: Car RF is the car moving in front of the examinee's car on the right-lane.

Car RB: This is the car behind the examinee's car on the right-lane. Unlike, the label switching from Car LF to LB, Car RB switches immediately to Car RF after Car RB is overtaken by the examinee's car.

Shown in Fig. 4 is the sample profiles of the surrounding cars. As can be seen from the Fig. 4, the label of Car RB switches to Car LF when the Car RB passes the examinee's car at times A. Another switching of label takes place at times B, when the examinee's car attains 75 [m] ahead after overtaking Car LF. This type of behavioral data is vital when estimating the condition for mode switching of the primitive driving skills.

It should be noted that the examinee is able to recognise all four cars in the driving simulator, credit to the wide frontal screens and the mirrors.

B. GAZE INFORMATION

In addition to the surrounding car's information described in section III-A, this research utilizes also, the driver's line-of-sight in analysing the overtaking driving behavior. The driver's line-of-sight is expected to estimate the driver's gazing behavior, therefore, the hypothesis follows that,

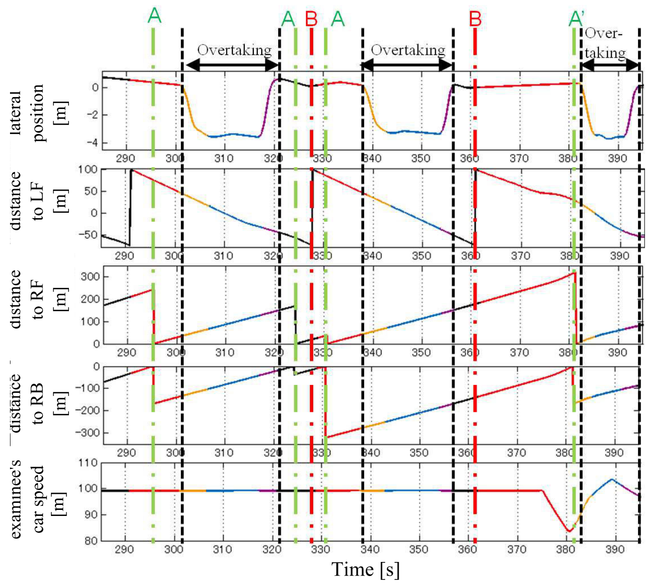

FIGURE 4. Sample profiles of driving data with switching of labels.

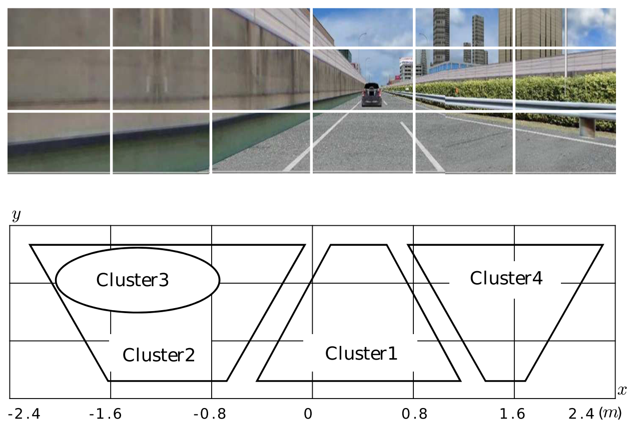

FIGURE 5. The definition of gaze clusters used for the driving experiment.

by including this type of information in the analysis, driver's decision can be more clearly identified. For convenience, the various information obtained as a result of line-of-sight measurement will be referred to as "gaze information". The gaze information data is measured using FaceLab and the feature quantity group. The gaze information here consists of the x and y coordinates of intersection of the line-of-sight and the screen on which the image of the driving simulator is projected, as well as, line-of-sight cluster which is a value derived based on the line-of-sight information by dividing the screen of the driving simulator based on x and y coordinates into four areas as in Fig. 5. Cluster1 shows the front of the driver, Cluster2 shows the left side including the left mirror, Cluster3 corresponds to the area when the driver visually recognizes the rear view mirror, and Cluster4 shows the right side including the right mirror. The coordinates of the screen are (0, 0) at the center of the front screen of the driving

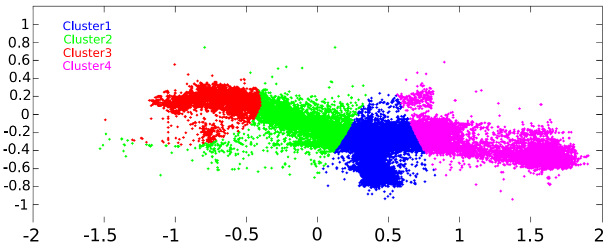

FIGURE 6. The clusters used for the gaze information measurement.

TABLE 1. Profile of experiment participants.

Driver	A	B	C	D	E	F
Age range	20s	20s	20s	40s	40s	40s
Sex	male	male	female	male	male	female
Driving Freq.	rarely	rarely	rarely	daily	daily	daily

simulator, with the x coordinate in the horizontal direction and the y coordinate in the vertical direction.

Fig. 6 shows the color coded distribution of the line-of-sight data obtained from the FaceLab during the experiment. For the moving distance of the line-of-sight, the visual recognition time of each line-of-sight cluster, and the saccade movement time, the cumulative value for the past 3 seconds from that time is used.

C. EXPERIMENTAL PROCEDURE

Total of six drivers with informed consent took part in the driving experiment on the driving simulator. Practice sessions were given to all drivers in advance to make sure that they were skilled enough on virtual driving. The driving experiment for each driver lasted for 2 days with a total of twelve trials, divided into six trials a day, with enough rest between trial intervals. The drivers were asked to drive on the virtual expressway in a safe manner for 10 [min] while keeping a speed of 100 [km/h] for each trial. Additionally, drivers were instructed to keep driving on the left-lane except when overtaking is necessary due to the slow car in front. On average, at least 150 overtaking behaviors in about 120 [min] driving data for each driver was measured which gives on average, 302,000 useful data points (samples) with a time step of 0.02 second for each driver.

The profiles of participants that took part in this driving experiment are given in Table 1, which shows the age range, sex, and the driving frequency of the drivers. Finally, this driving experiment was approved by the Nagoya University's ethical committee and in compliance with the Helsinki declaration.

IV. DEFINITION OF THE VARIABLES

A. DEFINITION OF INPUT VARIABLES

Generally, it is always complicated to decide or determine from candidates of input variables extracted from measured information which input variables are actually meaningful. In the work presented here, first and foremost, several variables are considered as candidates for the input variables in order to model the decision making. Afterwards, the model selection method presented in section II is applied in order to accept some of candidate variables as input variables.

In this paper, two categories of information (the surrounding cars' information and driver's gaze information) are considered to model the overtaking driving behavior of a driver. As part of the surrounding cars' information, the following basic physical information are considered first as the candidates for the input variables:

Distance[m] D_{LF}, D_{RF}, D_{RB}:

The longitudinal distances between examinee's car and surrounding cars,

relative speed[m/s] Vr_{LF}, Vr_{RB}:

The range-rate between examinee's car and surrounding cars,

examinee's car speed[m/s] V:

The driving speed of examinee's car.

The frequency of measure for all the variables is 60 [Hz], and the measured values are denoted as $D_{LF}(k)$ for $(k \in \{1, \ldots, K\})$ where K is the number of measured data or measured samples. Here, for simplicity, the variable $D_{LF}(k)$ is represented by D_{LF}. The variables of interest could be measured easily on our driving simulator.

Secondly, variables relating to the 'Risk feeling index' of the driver are considered as candidates for the input variables. Numerous works have tried to quantify and propose the risk feeling of drivers, however, in this work the following variables were chosen as the typical cognitive information relating to the driver's risk feeling index:

Time to collision[s] TTC_*:

This risk evaluation index is calculated by $TTC_{LF} = D_{LF}/Vr_{LF}$,

time headway[s] THW_*:

$THW_{LF} = D_{LF}/V$ computes this risk evaluation index,

KdB[-] KdB_{LF}:

KdB as a risk feeling index is the projection of the back of the frontal car unto the retina of the driver. Generally speaking, large values of KdB signifies high level of dangerous situation. KdB is computed by $KdB_{LF} = 10 \log\{4.0 \times 10^7 \times Vr_{LF}/D_{LF}\}$.

Finally, the following variables $(x_{21}, \ldots, x_{28})$ relating to the driver's gaze information are considered as candidates for the input variables:

eye movement distance[m] D_{gaze}:

This is the distance measured from the driver's eye movement by observing the gaze movement on the screen,

x coordinate on the screen [m] X_{gaze}:

The x coordinate on the screen of driver's gaze,

y coordinate on the screen [m] Y_{gaze}:

The y coordinate on the screen of driver's gaze,

TABLE 2. Candidates of input variables.

x_1	Examinee's car speed (V)
x_2	Square of V (V^2)
x_3	Distance to car LF (D_{LF})
x_4	Relative speed to car LF (Vr_{LF})
x_5	Inverse of Vr_{LF} ($1/Vr_{LF}$)
x_6	Square of Vr_{LF} (${Vr_{LF}}^2$)
x_7	Time to collision to car LF (TTC_{LF})
x_8	Inverse of TTC_{LF} ($1/TTC_{LF}$)
x_9	Time head way to car LF (THW_{LF})
x_{10}	Inverse of THW_{LF} ($1/THW_{LF}$)
x_{11}	KdB to car LF (KdB_{LF})
x_{12}	Distance to car RF (D_{RF})
x_{13}	Time head way to car RF (THW_{RF})
x_{14}	Inverse of THW_{RF} ($1/THW_{RF}$)
x_{15}	Distance to car RB (D_{RB})
x_{16}	Relative speed to car RB (Vr_{RB})
x_{17}	Inverse of Vr_{RB} ($1/Vr_{RB}$)
x_{18}	Square of Vr_{RB} (${Vr_{RB}}^2$)
x_{19}	Time to collision to car RB (TTC_{RB})
x_{20}	Inverse of TTC_{RB} ($1/TTC_{RB}$)
x_{21}	Eye movement distance (D_{gaze})
x_{22}	x coordinate on the screen (X_{gaze})
x_{23}	y coordinate on the screen (Y_{gaze})
x_{24}	Front cluster viewing time (T_{FC})
x_{25}	Left cluster viewing time (T_{LC})
x_{26}	Rear mirror cluster viewing time (T_{rearC})
x_{27}	Right cluster viewing time (T_{RC})
x_{28}	Saccade movement time (T_S)

cluster viewing time [s] $T_{FC}, T_{LC}, T_{RMC}, T_{RC}$:
The time taken by the driver to view front cluster, left cluster, rear mirror cluster, and right cluster respectively,

saccade movement time [m] T_S:
Time taken by the driver for saccade movements.

These glance metrics were selected based on the proposals of [44]–[46] for measuring driver's gaze behavior. Reference [44] proposed gaze-point-coordinates related metrics such as x and y coordinates on the screen, while [45] and [46] proposed time related metrics (such as cluster viewing time) and directional related metrics (such as front cluster) respectively as relevant measures for eye glance behavior. In addition, the selected surrounding cars' information was also based on [45] which gives a summary of the necessary information surrounding the ego vehicle for analysing driving behavior. All the candidates for input variables, $x(k) = \{x_1(k), x_2(k), \ldots, x_{28}(k)\}$, where $k \in \{1, 2, \ldots, K\}$, used in this analysis are listed in Table 2. Note that the square and the inverse of some of the above mentioned variables are included as part of candidates for the input variables with the intention of capturing the behavior of these variables over time.

B. MODES DEFINITION IN THE OVERTAKING BEHAVIOR

In this analysis, the overtaking behavior is segmented into five modes. The transformation or segmentation of the experimental driving data is a critical task, thus, this process should be considered deeply. There has been several techniques as pointed out in Section I for solving this problem. Since the main purpose of this paper is to identify the mode switching condition after segmenting the data, manual segmentation of the data to specify the number of modes is assumed. As such, the modes in overtaking behavior as in Fig. 3 are defined as follows:

mode 1: following the leading car
Approaching and following Car LF without lane change,

mode 2: lane change
Changing driving lane to the right-lane,

mode 3: passing car
Passing car LF without lane change,

mode 4: return
Returning to the left-lane,

mode 5: free driving
Driving after returning to the left-lane until the next leading Car LF is seen.

In this paper, mode 5 is excluded from the analysis because it not a part of the overtaking behavior. The modes specified above are obtained manually by considering some specific dynamics of the of the examinee's car as follows:

Mode 1

$$y > -2.0[\mathrm{m}] \tag{7}$$

Mode 2

$$\theta_{yaw} < -0.017[\mathrm{rad}] \quad \text{and} \quad \theta_{steer} > 1[\mathrm{deg}] \tag{8}$$

Mode 3

$$\theta_{yaw} > -0.005[\mathrm{rad}] \tag{9}$$

Mode 4

$$\theta_{yaw} > 0.017[\mathrm{rad}] \quad \text{and } \theta_{steer} < -1[\mathrm{deg}] \tag{10}$$

Here, y, θ_{yaw}, θ_{steer} are the lateral position, yaw angle, and steering angle of the examinee's car respectively. These conditions were set by taking into account the physical significance of each mode. In addition, it should be noted that these dynamics were not included as candidates for the input variables to model the driver's decision making, but used only to define the modes. These values of y, θ_{yaw}, θ_{steer} used in defining the modes were chosen after analysing the driving experiment data, specifically, the values just before the ego vehicle departs the left lane, the values after the vehicle arrives at the destination (right) lane and the values just before the car departs the right lane.

Using the predefined conditions, all data samples obtained from the driving experiment are assigned to one of the modes. Finally, all measurement data set $\mathcal{D}$ is divided into five data sets, $\mathcal{D}_i$ $(i = 1, 2, \ldots, 5)$.

TABLE 3. LRM variable selection result for mode 1 → mode 2 switch.

Subject	A	B	C	D	E	F
x_1 (V)	0	15.81	0	0	0	20.10
x_2 (V^2)	0	-11.02	0	0	0	-14.17
x_3 (D_{LF})	0	-0.26	0	-0.36	0	8.06
x_4 (Vr_{LF})	0	0	0	0	-1.79	0
x_5 $(1/Vr_{LF})$	0	0	0	0	0	0
x_6 (Vr^2_{LF})	0	0	0	0	0	0
x_7 (TTC_{LF})	0	0	0	0	0	0
x_8 $(1/TTC_{LF})$	1.74	0	0	0	11.16	1.33
x_9 (THW_{LF})	-6.99	-2.40	-3.46	-4.68	0	-8.99
x_{10} $(1/THW_{LF})$	0	0	0	-2.32	6.92	0
x_{11} (KdB_{LF})	0	3.60	0	2.70	0	0
x_{12} (D_{RF})	0	0	0	0	0	-7.07
x_{13} (THW_{RF})	-0.73	0	0	-0.85	0	5.63
x_{14} $(1/THW_{RF})$	-1.19	-0.66	-0.48	-0.94	-0.63	-3.17
x_{15} (D_{RB})	-0.55	0	0	0	0	-0.48
x_{16} (Vr_{RB})	0	0	0	0	0	0
x_{17} $(1/Vr_{RB})$	0	0	0	0	0	0
x_{18} (Vr^2_{RB})	0	0	0	0	0	0
x_{19} (TTC_{RB})	-0.70	0	0	-0.23	0	0
x_{20} $(1/TTC_{RB})$	-2.73	-2.85	0	-4.09	-4.97	-3.45
x_{21} (D_{gaze})	0	0.50	0	0.15	0	0
x_{22} (X_{gaze})	0	0	0	0	0.35	0
x_{23} (Y_{gaze})	0	0	0	0	0.23	0
x_{24} (T_{FC})	0	0	-0.55	0.18	0	-0.48
x_{25} (T_{LC})	0.46	0	0	0	0	0
x_{26} (T_{rearC})	0	0	0	-0.39	0	0
x_{27} (T_{RC})	-0.12	0	0	0.39	0.43	0
x_{28} (T_S)	0	0	0	0	0	0

C. DEFINITION OF OUTPUT VARIABLES

A binary variable specifying the switching of the mode is used as the output variable of the LRM. To express the mode switching from mode i to mode $i+1$, four different models $\mathcal{M}_i$ $(i = 1,2,3,4)$ are identified using the data sets $\mathcal{D}_i$ and $\mathcal{D}_{i+1}$. $y_i(k) \in \mathcal{Y}_i$ is specified as follows:

$$y_i(k) = \begin{cases} 0 & \cdots \quad \text{if } x(k) \in \mathcal{D}_i \\ 1 & \cdots \quad \text{if } x(k) \in \mathcal{D}_{i+1} \end{cases} \tag{11}$$

where $y_i(k) = 0$ corresponds to the data $x(k)$ measured before mode switching from i to $i+1$, whereas, $y_i(k) = 1$ corresponds to the data $x(k)$ measured after the mode is switched from i to $i+1$. The model $\mathcal{M}_i$ is identified by the method described in section II using $\mathcal{D}_i$ and $\mathcal{D}_{i+1}$ as the input variables and $\mathcal{Y}_i$ as output variables. It should be noted that the data used was normalized prior to model identification and as such the magnitudes of the selected variables can be used directly to interpret the importance of that variable.

V. MODEL IDENTIFICATION RESULTS AND ANALYSIS OF OBTAINED MODELS

This section presents the analysis and discussion of the results of the identified drivers' mode switching condition.

A. DECISION MAKING FOR SWITCHING FROM MODE 1 TO 2 (LEFT-LANE DEPARTURE)

The decision for the driver departing from the left-lane to the right-lane (i.e., switching from mode 1 to 2) is investigated for drivers A to F. Table 3 shows the selected variables and corresponding estimated parameters of the identified logistic regression model. All the variables were normalized prior to the model identification, hence, the magnitude of a parameter can be taken to be a reflection of the significance of its corresponding variable.

The risk feeling indexes relating to time headway $(THW : x_9, x_{14})$ and time-to-collision $(TTC : x_{20})$ were a common selection among drivers. Time headway (THW) can be seen as the spacing a driver tries to keep depending on the current speed, whereas, the inverse of TTC expresses, and is proportional to the driver's collision risk to a surrounding vehicle of interest. Therefore, it is natural that these risk feeling indexes were selected commonly as variables a driver consider for lane departure. Variables x_{20} $(1/TTC_{RB})$ and $x_9(THW_{LF})$ were selected among five drivers when switching from following the leading vehicle to left-lane departure. The inverse of TTC and inverse of THW are risk feeling indexes that increase as the driver's vehicle approaches the surrounding vehicles. The negative values of the parameters of x_{20} $(1/TTC_{RB})$ and x_{14} $(1/THW_{RF})$ indicates that the probability of mode transition decreases as the right front vehicle and the right rear vehicle becomes closer. The parameter value of variable x_9 (THW_{LF}) outweighs that of variable x_3 (D_{LF}), this implies that a driver does not simply change from the right-lane based on the distance to the leading vehicle, but rather puts more consideration on the amount of THW change. For drivers B and E, the variables x_1 and x_2 related to the vehicle speed are very large. Considering the positive value of x_1 parameter, it can be said that these drivers accelerate significantly at a high frequency when changing lanes.

For Driver A, most of the risk feeling index variables were selected as important decision variables. In addition, distance to the right-lane back car x_{15}, right cluster (which includes the right mirror) viewing time x_{27} and left cluster viewing time x_{25} were also selected. These added selections implies that this driver's consideration is deeply affected by the vehicles behind. The selection of both x_{27} and x_{25} at first glance might seem contradictory, however, careful consideration shows that this is a perfect demonstration of a driver (A) making decision based on memory (in this case, the choice of x_{25}).

Driver B from the choice of D_{LF}, KdB_{LF} and THW_{LF} seems to focus on keeping a safety distance while following. This explains the rapid acceleration looking at coefficients of x_1 and x_2 when changing lanes.

For Driver C, the variables x_9, x_{14} and x_{24} were selected as important decision variables. The selection here suggests that this driver focuses more on what is happening in front. More so, the driver makes lane change decision based on spacing-distance versus speed relationship. Intuitively, this type of driver will tend to negotiate more for lane change using the turn signal rather than planning ahead of time.

Numerous variables relating to driving risk feeling index and driver gaze information were chosen for driver D. From the selection, it can be said that this driver considers carefully all the surrounding cars during the lane change decision. Driver D can be considered a conservative driver.

TABLE 4. LRM variable selection result for mode 2 → mode 3 switch.

Driver	A	B	C	D	E	F
x_1 (V)	1.40	0	0	0	0	0
x_2 (V^2)	0	0	0	0	0	0
x_3 (D_{LF})	0	0	0	0	-6.78	-4.92
x_4 (Vr_{LF})	-2.11	0	0	0	0	0
x_5 ($1/Vr_{LF}$)	0	0	0	0	0	0
x_6 (Vr_{LF}^2)	0	0	-0.58	0	0	0
x_7 (TTC_{LF})	0	1.97	0	0	0	0
x_8 ($1/TTC_{LF}$)	-0.40	0.24	0	-0.40	0	0
x_9 (THW_{LF})	-12.25	-3.34	-1.99	-4.80	0	0
x_{10} ($1/THW_{LF}$)	-3.41	0	0.59	0	0	0
x_{11} (KdB_{LF})	0.53	-0.48	-0.59	0	0	0
x_{12} (D_{RF})	0	-8.22	-0.95	0	-4.43	0.86
x_{13} (THW_{RF})	0	7.67	0	0	4.73	0
x_{14} ($1/THW_{RF}$)	-0.37	-1.92	-2.50	-1.03	-1.17	-0.98
x_{15} (D_{RB})	0	0	1.17	0	0	0
x_{16} (Vr_{RB})	0	1.12	0	0	0	0
x_{17} ($1/Vr_{RB}$)	0	0	0	0	0	0
x_{18} (Vr_{RB}^2)	0	0	0	0	0	0
x_{19} (TTC_{RB})	0	0	0.16	0	0	0
x_{20} ($1/TTC_{RB}$)	0	0	-2.02	0	0	0
x_{21} (D_{gaze})	0	0	0	0	0	0
x_{22} (X_{gaze})	0	0	0	0	0	0
x_{23} (Y_{gaze})	0	0.70	0	0	0	0.31
x_{24} (T_{FC})	0	0.55	0.36	0.41	0	0
x_{25} (T_{LC})	0	-0.31	0.60	0	0	0
x_{26} (T_{rearC})	0	0	0	0	-0.48	0
x_{27} (T_{RC})	0	0	0.35	0	0	0
x_{28} (T_S)	0	0	0	0	0	0

TABLE 5. LRM variable selection result for mode 3 → mode 4 switch.

Driver	A	B	C	D	E	F
x_1 (V)	0	0	0	0	0	0
x_2 (V^2)	0	0	0	0	0	0
x_3 (D_{LF})	0	0	-19.28	0	-21.02	-22.27
x_4 (Vr_{LF})	0	0	0	0	0	3.07
x_5 ($1/Vr_{LF}$)	0	0	0	0	0	2.59
x_6 (Vr_{LF}^2)	0	0	-1.19	0	0	-1.82
x_7 (TTC_{LF})	0	0	0	0	0	0
x_8 ($1/TTC_{LF}$)	0	0	0	0	0	0
x_9 (THW_{LF})	0	0	0	0	0	0
x_{10} ($1/THW_{LF}$)	0	0	0	0	0	0
x_{11} (KdB_{LF})	0	0	0	0	0	0
x_{12} (D_{RF})	-0.19	0	0	0	0	0
x_{13} (THW_{RF})	0.40	0	0	0	0	0
x_{14} ($1/THW_{RF}$)	0	0	0	0	0	0
x_{15} (D_{RB})	0	0	0	0	0	0
x_{16} (Vr_{RB})	0	0	0	0	0	0
x_{17} ($1/Vr_{RB}$)	0	0	0	0	0	0
x_{18} (Vr_{RB}^2)	0	0	0	0	0	0
x_{19} (TTC_{RB})	0	0	0	0	0	0
x_{20} ($1/TTC_{RB}$)	0	0	0	2.44	0	0
x_{21} (D_{gaze})	0	1.02	0	0	0.72	0
x_{22} (X_{gaze})	0	0	0	0	0	0
x_{23} (Y_{gaze})	0	0	0	0	0	0
x_{24} (T_{FC})	0	0	0	-0.28	-0.58	-0.38
x_{25} (T_{LC})	-0.30	-0.40	0	-0.45	0	0
x_{26} (T_{RMC})	0	0	0	0.99	0	0
x_{27} (T_{RC})	-0.36	-0.71	0	0	0	0
x_{28} (T_S)	0	0	0	0	0	0

Driver E is the only driver where the relative speed to the left-lane frontal vehicle x_4 was selected. This insinuates that the speed of the frontal vehicle is important for this driver when deciding to change lanes. Additionally, the chosen risk feeling index variables together with the selected gaze information variables suggest that this driver appropriately checks the surrounding vehicles for lane change decision.

For driver F, the distances, safety distance versus current speed relationship, and collision to the surrounding cars are important factors for mode 1 to 2 switching. This driver tends to accelerate rapidly when changing lanes, this might explain why only one gaze information variable was selected.

Generally speaking, from these results, the THW and TTC risk feeling indexes have strong influence on the lane departure decision in the overtaking driving behavior. Furthermore, depending on the driver, different additional variables are considered when making lane departure decision. The variable selection results here insinuates that, though there are common characteristics between drivers, personalized preferences are largely expressed in the lane departure decision of the overtaking driving behavior.

B. DECISION MAKING FOR SWITCHING FROM MODE 2 TO 3 (OVERTAKING)

Table 4 shows the selected variables for the LRM to express the mode switching from mode 2 to 3.

When the driving task was switched from right-lane change to overtaking, the variables x_9 and x_{14} with large parameter values were commonly selected among most of the drivers. This result is similar to the case of switching from the left-lane to right-lane (i.e., mode 1 to mode 2). From this, it can be said that the environmental information recognized by the driver is similar between the task of switching to overtaking and the task of switching to lane departure. The difference is that the variable related to the vehicle speed is no longer selected, but instead the variable x_{12} that represents the range to the frontal vehicle is selected for drivers B, C, E, and F. The negative values of the parameters of drivers B, C, and E implies that for these drivers, the probability of changing lanes increases as the distance from the vehicle ahead increases, and vice-versa for driver F.

C. DECISION MAKING FOR SWITCHING FROM MODE 3 TO 4 (LANE RETURN)

The selected variables and estimated parameters of the model for mode switching from 3 to 4 (returning to the left-lane) are shown in Table 5.

The first thing to recognize is that in this result, fewer number of variables were selected. This indicates that left-lane return maneuver requires less information compared to the left-lane departure driving task. For drivers C, E, and F, the variable x_3 has a large negative parameter value, meaning that they are more likely to switch to the left-lane as the range to the left-lane frontal vehicle increases. For drivers A, B, and D, the variable x_{25}, which is the left-cluster viewing time is a common selection. This selection implies that these drivers are constantly checking visually, the situation of the vehicle on the destination lane during the lane-return task. The selected variables vary greatly among drivers, therefore, it can be concluded that the lane return task is more of a personalized preference driving maneuver.

TABLE 6. Mode switching estimation result (mode 1 to mode 2).

Driver	A	B	C	D	E	F
Matching rate (Mode 1)	0.97	0.94	0.86	0.94	0.98	0.93
Matching rate (Mode 2)	0.90	0.89	0.87	0.97	0.94	0.87
Correct switching rate [%]	87.88	86.21	64.00	93.10	100.00	85.19
Switching delay time [s]	0.91	1.19	1.65	0.83	0.58	1.74

TABLE 7. Mode switching estimation result (mode 2 to mode 3).

Driver	A	B	C	D	E	F
Matching rate (Mode 2)	0.98	0.89	0.74	0.88	0.84	0.86
Matching rate (Mode 3)	0.86	0.74	0.86	0.87	0.92	0.93
Correct switching rate [%]	100.00	87.50	50.00	92.80	61.53	81.81
Switching delay time [s]	0.55	1.25	1.69	1.38	1.50	1.33

VI. MODEL VERIFICATION BASED ON ESTIMATION PERFORMANCE

This section verifies the validity of the identified model by mode switching estimation performance testing. To evaluate the identified model, switching estimation using the model is performed on a testing data set as shown in Fig.1.

Table 6, Table 7, and Table 8 show the result of the estimated mode switching from mode 1 to 2, mode 2 to 3, and mode 3 to 4 respectively. The result of each table shows the matching rate of each mode in respective mode switching estimation. In addition, the correct switching rate and the switching delay time are shown. The correct switching rate is the percentage the time instant of the calculated mode switching belongs to the time interval of 3 seconds plus or minus of the actual mode switching time. This means that a sample is part of the correct sample when it is on the same side of the algorithm fitting segmentation point and the ground truth segmentation point. The 3 seconds time interval here was chosen by trial and error by observing the measured driving data and also by careful consideration of the overtaking driving behavior. From the three tables, it could be seen that good identification for each mode in each switching task was achieved, furthermore, high correct switching rate was achieved relatively for all the drivers.

In addition, the result of the correct switching rate in percentage where only the surrounding environment was considered in the model without including the driving behavior information in form of gaze information is shown in Table 9 for all the drivers. Comparing this to the results of Table 6 - 8, it can be seen that the correct switching rate improved for all the drivers in all the modes with the inclusion of the gaze information.

The sample profiles for three overtaking trials of the four modes defined to represent the overtaking driving behavior is shown in Fig.7. The black solid line represents the true mode transition from the data, whereas the dashed line represents the mode switching estimated using the identified model. It can be seen that the estimated mode switching agrees relatively well with the true mode transition.

Finally, Table 10 shows the estimation result for the overtaking trials for all the drivers. The result shows that for drivers A to F there is an 80% chance or more of properly

TABLE 8. Mode switching estimation result (mode 3 to mode 4).

Driver	A	B	C	D	E	F
Matching rate (Mode 3)	0.94	0.96	0.88	0.76	0.98	0.99
Matching rate (Mode 4)	0.97	0.52	0.95	0.94	0.97	0.92
Correct switching rate [%]	100.00	70.00	100.00	91.00	100.00	100.00
Switching delay time [s]	0.25	1.00	1.60	0.35	0.13	0.33

TABLE 9. Mode switching estimation result (correct switching rate [%]) without including driving gazing behavior information.

Driver	A	B	C	D	E	F
Mode 1 to 2	81.80	75.90	60.00	79.30	96.90	70.40
Mode 2 to 3	92.00	84.20	8.30	85.00	53.80	63.60
Mode 3 to 4	100.00	50.00	81.80	25.00	100.00	100.00

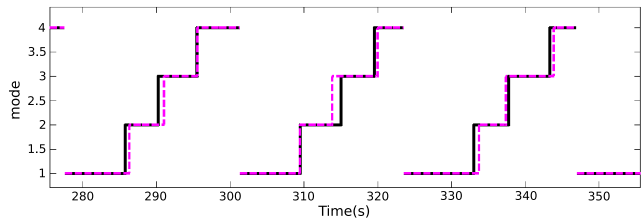

FIGURE 7. Sample profile of model estimating the overtaking decision.

estimating their overtaking driving behavior. For drivers D and F, the estimation rate of mode 3 was lower than that of the other modes. Since the logistic regression models are connected linearly, there was a concern that if the estimation rate of one mode is relatively bad, the estimation rate of the next mode would also be affected. However, judging from Driver D and Driver F where consecutive estimation rates are 0.60 to 0.88 and 0.74 to 0.96 respectively, which are relatively quite low and high. This indicates that the proposed model does not depend only on the accuracy of individual logistic regression models.

Comparing the results (Table 6 - 8) of the work presented here to the results (Table 9) of the previous work [33], where the experiment setting and drivers were the same and only the surrounding vehicles' information was collected and used to model the driver's overtaking decision making. In this study, by including the gazing behavior information, there was improvement in the correct switching rate for all the drivers in all the modes defined for the overtaking driving task. Compared to the previous work, it can be said that this study estimates the driver's decision for the overtaking driving task better. Thus highlighting one of the contributions/improvements of this study. In [47], HMM was proposed for detecting risky lane changes using integrated modeling of driver gaze and vehicle operation behavior. Although the modeling result was good, however, better result that is explainable could be achieved with less computational requirement if the idea of variable selection introduced in this paper was used. For example, using variable selection during the lane changing to reduce to the most influential variables as input for the Hidden Markov Model. In addition, most of the studies that have used gazing behavior

TABLE 10. Estimation of the overtaking decision by the model.

Driver	A	B	C	D	E	F
Mode 1	0.93	0.94	0.51	0.97	0.98	0.81
Mode 2	0.80	0.82	0.61	0.79	0.86	0.90
Mode 3	0.88	0.93	0.72	0.60	0.88	0.74
Mode 4	0.93	0.56	0.97	0.88	0.96	0.96
Total	0.89	0.85	0.64	0.81	0.93	0.84

to analyze driving behavior have not explored the use variable selection to add additional meaning to modeling results, therefore, the study presented here advances the knowledge of using gazing information to model the human driving behavior.

VII. DISCUSSION

The useful applications of the proposed model are summarized as follows:

- In understanding the overtaking behavior especially expressing individual driver characteristics, the method presented here which is a combination of the logistic regression model and statistical model selection, is well suited. The methodology presented here is able to identify and differentiate the common characteristics and personalities respectively among drivers in the decision making for overtaking driving. Therefore, the idea in this study can be directly leveraged by controller designs for implementing personalized automated driving. Additionally, by segmenting the overtaking driving behavior into modes and using only the most influential explanatory variables in each mode, computational cost is reduced, facilitating online applications.
- The work presented here is suitable for application in mixed traffic that comprises of human driven vehicles and automated vehicles which is seen as the realistic evolution of automated driving. By applying the scheme developed here the overtaking intention of a human driver can be predicted and communicated to the surrounding automated vehicles, ensuring a safer traffic.
- In design of advanced driving assistance systems (ADAS), by using the combination of the driver's gaze information and surrounding vehicles' information, a better and more reliable system could be developed. For example, in the overtaking driving maneuver, by using the gaze information of the driver to identify the driver's intention to overtake or change lanes, and using the information of the surrounding vehicles' to judge if it is safe or not to overtake, a warning or assistance system could be developed.

VIII. CONCLUSION

This paper presented a new methodology for modeling the driver's decision making in the overtaking driving behavior without prior knowledge of the explanatory variables. The overtaking decision which includes switching of multiple driving modes was expressed mathematically by the combination of a logistic regression model and a statistical model selection method based on Wald statistics. By using the line-of-sight information in particular, which can be regarded as the driving gazing behavior, in combination with the surrounding environment information, it was possible to construct the overtaking decision making by estimating switching of modes and its timing in an accurate manner based on the proposed model. In addition, investigation was carried out on the explanatory variables extracted by the variable selection method to clarify the relationship between the driver's selected variables and mode switching, and also, the individual differences and similarities of the models were discussed.

For the overtaking driving behavior in the problem setting as shown in Fig. 3, in the first step which is left-lane departure, the risk feeling indexes relating to time headway ($THW : x_9,\ x_{14}$) and time-to-collision ($TTC : x_{20}$) were a common selection among drivers, on the other hand, the selected gaze information for each driver was different highlighting personalized preferences in the lane-departure task. In the decision making for overtaking (mode 2 to 3), variables ($THW : x_9,\ x_{14}$) relating to the time headway were a common selection among drivers. This is similar to the behavior in the left-lane departure mode, however, the distance to the frontal vehicle plays an important role in this mode switching. Finally, for the decision to return to the left lane, higher percentage of gaze features was selected (different for each driver) when compared to the surrounding environment features. This suggests that the drivers make decision based on visual observations of the destination lane during the lane return task. In summary, the variable selection results insinuate that, though there are common characteristics between drivers, personalized preferences are mainly expressed in the decision for the overtaking driving task.

The usefulness of the proposed modeling framework for enabling autonomous driving and design of ADAS was presented in VII, and can be summarized in the following way. The result can be directly used to design controllers for autonomous lane-change/overtaking where reflecting driver's preferences is required. In addition, this work can enable mixed (i.e., a mixture of human-driven vehicles and automated vehicles) autonomous driving by predicting the overtaking intention of the human driven vehicle and communicating this prediction where necessary. Finally, by using the gazing behavior, particularly the variables that a driver uses for decision making, the driver's overtaking intention can be estimated and then the surrounding environment can be used to judge if it is safe or not to overtake, and thus a personalized warning system can be developed.

In this study, individual differences and similarities by comparing models of multiple drivers was analyzed. However, it is considered that the driving behavior model also differs depending on the driving skill of the driver. In the future, it will be necessary to evaluate the differences in obtained models due to driving skills by analyzing the data of drivers with various driving skill tests. In addition, even though the focus of this study was not in generalising or grouping drivers into classes, but rather identifying the differences, thus explaining the decision to use only 6 (six) subjects, further experiment should be conducted in the future to increase the number of subjects. Furthermore, in the environment assumed here, the driving behavior of other vehicles was limited to straight forward driving, and the driving behavior of the driver was limited to overtaking. A natural extension of this work is to consider in the future, more complicated driving situations for both the driving and the surrounding vehicles, for example, a situation where another vehicle merges or changes lanes. In addition, driving simulator data was used in this study, and such data may lack the complexity of real-road driving, therefore, in the future, a real-road overtaking driving experiment could be conducted to further verify the real-world applicability of the proposed model.

REFERENCES

[1] N. Sousa, A. Almeida, J. Coutinho-Rodrigues, and E. Natividade-Jesus, "Dawn of autonomous vehicles: Review and challenges ahead," *Proc. Inst. Civil Eng. Municipal Eng.*, vol. 171, no. 1, pp. 3–14, Mar. 2018, doi: 10.1680/jmuen.16.00063.

[2] C. Bartneck, C. Lütge, and A. Wagner, "Autonomous vehicles," in *An Introduction to Ethics in Robotics and AI* (Springer Briefs in Ethics). Cham, Switzerland: Springer, 2020, Accessed: May 18, 2021, doi: 10.1007/978-3-030-51110-4_10.

[3] S. Campbell, N. O'Mahony, L. Krpalcova, D. Riordan, J. Walsh, A. Murphy, and C. Ryan, "Sensor technology in autonomous vehicles: A review," in *Proc. 29th Irish Signals Syst. Conf. (ISSC)*, Jun. 2018, pp. 1–4, doi: 10.1109/ISSC.2018.8585340.

[4] *Taxonomy and Definitions for Terms Related to Driving Automation Systems for on-Road Motor Vehicles*, document J3016-202104, SAE International, SAE, Warrendale, PA, USA, 2021. Accessed: May 18, 2021. [Online]. Available: https://www.sae.org/standards/content/j3016_202104/

[5] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proc. 1st Annu. Conf. Robot Learn.*, Mountain View, CA, USA, Nov. 2017, pp. 1–16.

[6] G. Rong, B. H. Shin, H. Tabatabaee, Q. Lu, S. Lemke, M. Mozeiko, E. Boise, G. Uhm, M. Gerow, S. Mehta, E. Agafonov, T. H. Kim, E. Sterner, K. Ushiroda, M. Reyes, D. Zelenkovsky, and S. Kim, "LGSVL simulator: A high fidelity simulator for autonomous driving," in *Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2020, pp. 1–6, doi: 10.1109/ITSC45102.2020.9294422.

[7] P. Angkititrakul, C. Miyajima, and K. Takeda, "Modeling and adaptation of stochastic driver-behavior model with application to car following," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2011, pp. 814–819.

[8] S. D. Keen and D. J. Cole, "Bias-free identification of a linear model-predictive steering controller from measured driver steering behavior," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 434–443, Apr. 2012.

[9] L. Pariota, G. N. Bifulco, and M. Brackstone, "Linear dynamic model for driving behavior in car following," *Transp. Sci.*, vol. 50, pp. 1138–1763, Aug. 2016.

[10] Y. Jiao, R. Cheng, and H. Ge, "A new continuum model considering driving behaviors and electronic throttle effect on a gradient highway," *Math. Problems Eng.*, vol. 2020, pp. 1–22, May 2020.

[11] J. Hu, Y. Zhang, and R. Zhao, "Considering prevision driving behavior in car-following model," in *Proc. 4th Int. Conf. Sensors, Mechatronics Autom. (ICSMA)*, Nov. 2016, pp. 409–412.

[12] J. Wang, F. Sun, R. Cheng, and H. Ge, "An extended car-following model considering the self-stabilizing driving behavior of headway," *Phys. A, Stat. Mech. Appl.*, vol. 507, pp. 347–357, Oct. 2018.

[13] Z. Wen-Xing and Z. Li-Dong, "A new car-following model for autonomous vehicles flow with mean expected velocity field," *Phys. A, Stat. Mech. Appl.*, vol. 492, pp. 2154–2165, Feb. 2018.

[14] W.-X. Zhu and H. M. Zhang, "Analysis of mixed traffic flow with human-driving and autonomous cars based on car-following model," *Phys. A, Stat. Mech. Appl.*, vol. 496, pp. 274–285, Apr. 2018.

[15] H. Ou and T.-Q. Tang, "An extended two-lane car-following model accounting for inter-vehicle communication," *Phys. A, Stat. Mech. Appl.*, vol. 495, no. 1, pp. 260–268, Apr. 2018.

[16] M. Treiber and A. Kesting, "The intelligent driver model with stochasticity—New insights into traffic flow oscillations," *Transp. Res. B, Methodol.*, vol. 117, pp. 613–623, Nov. 2018, doi: 10.1016/j.trb.2017.08.012.

[17] X. Ma and I. Andreasson, "Behavior measurement, analysis, and regime classification in car following," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 1, pp. 144–156, Mar. 2007.

[18] X. Chen, L. Li, and Y. Zhang, "A Markov model for headway/spacing distribution of road traffic," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 4, pp. 773–785, Dec. 2010.

[19] W. Takano, A. Matsushita, K. Iwao, and Y. Nakamura, "Recognition of human driving behaviors based on stochastic symbolization of time series signal," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2008, pp. 167–172.

[20] K. Takenaka, T. Bando, S. Nagasaka, T. Taniguchi, and K. Hitomi, "Contextual scene segmentation of driving behavior based on double articulation analyzer," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2012, pp. 4847–4852.

[21] S. Sekizawa, S. Inagaki, T. Suzuki, S. Hayakawa, N. Tsuchida, T. Tsuda, and H. Fujinami, "Modeling and recognition of driving behavior based on stochastic switched ARX model," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 4, pp. 593–606, Dec. 2007.

[22] K. Younsi, P. Loslever, J.-C. Popieul, and P. Simon, "Fuzzy segmentation for the exploratory analysis of multidimensional signals: Example from a study on driver overtaking behavior," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 41, no. 5, pp. 892–904, Sep. 2011.

[23] B. Higgs and M. Abbas, "Segmentation and clustering of car-following behavior: Recognition of driving patterns," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 1, pp. 81–90, Feb. 2015.

[24] R. Terada, H. Okuda, T. Suzuki, K. Isaji, and N. Tsuru, "Multi-scale driving behavior modeling using hierarchical PWARX model," in *Proc. 13th Int. IEEE Conf. Intell. Transp. Syst.*, Sep. 2010, pp. 1638–1644.

[25] T. Akita, S. Inagaki, T. Suzuki, S. Hayakawa, and N. Tsuchida, "Analysis of vehicle following behavior of human driver based on hybrid dynamical system model," in *Proc. IEEE Int. Conf. Control Appl.*, Oct. 2007, pp. 1233–1238.

[26] A. Bender, G. Agamennoni, J. R. Ward, S. Worrall, and E. M. Nebot, "An unsupervised approach for inferring driver behavior from naturalistic driving data," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3325–3336, Dec. 2015, doi: 10.1109/TITS.2015.2449837.

[27] G. Li, Y. Chen, D. Cao, X. Qu, B. Cheng, and K. Li, "Extraction of descriptive driving patterns from driving data using unsupervised algorithms," *Mech. Syst. Signal Process.*, vol. 156, Jul. 2021, Art. no. 107589, doi: 10.1016/j.ymssp.2020.107589.

[28] T. Akita, T. Suzuki, S. Hayakawa, and S. Inagaki, "Analysis and synthesis of driving behavior based on mode segmentation," in *Proc. Int. Conf. Control, Autom. Syst.*, Oct. 2008, pp. 2884–2889.

[29] J. Nwadiuto, H. Chin, H. Okuda, and T. Suzuki, "Analysis and modeling of real world car-following driving in downtown area based on variable structured piecewise linear model," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 1211–1216, doi: 10.1109/ITSC.2019.8916962.

[30] D.-A. Nguyen, J. Nwadiuto, H. Okuda, and T. Suzuki, "Modeling car-following behavior in downtown area based on unsupervised clustering and variable selection method," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2020, pp. 3714–3720, doi: 10.1109/SMC42975.2020.9282910.

[31] J. C. Nwadiuto, H. Okuda, and T. Suzuki, "Driving behavior modeling based on consistent variable selection in a PWARX model," *Appl. Sci.*, vol. 11, no. 11, p. 4938, May 2021, doi: 10.3390/app11114938.

[32] H. Okuda, N. Ikami, T. Suzuki, Y. Tazaki, and K. Takeda, "Modeling and analysis of driving behavior based on a probability-weighted ARX model," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 1, pp. 98–112, Mar. 2013, doi: 10.1109/TITS.2012.2207893

[33] H. Okuda, S. Yoshino, T. Suzuki, and T. Kawai, "Identification of mode switching condition in overtaking behavior using variable-free logistic regression model," in *Proc. IEEE 18th Int. Conf. Intell. Transp. Syst.*, Sep. 2015, pp. 1539–1544.

[34] H. Takahashi, "A study on designing driver's operation support system using driving environment-driver-vehicle interaction model," in *Proc. Int. Conf. IEEE World Congr. Comput. Intell. Fuzzy Syst.*, vol. 1, May 1998, pp. 177–182.
[35] S. Li, "Simulation of car-following decision using fuzzy neural networks system," in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, Oct. 2003, pp. 140–145.
[36] G. Xu, L. Liu, and Z. Song, "Driver behavior analysis based on Bayesian network and multiple classifiers," in *Proc. IEEE Int. Conf. Intell. Comput. Intell. Syst.*, Oct. 2010, pp. 663–668.
[37] H. Wei, W. Ross, S. Varisco, P. Krief, and S. Ferrari, "Modeling of human driver behavior via receding horizon and artificial neural network controllers," in *Proc. 52nd IEEE Conf. Decis. Control*, Dec. 2013, pp. 6778–6785.
[38] A. Wahab, C. Quek, C. K. Tan, and K. Takeda, "Driving profile modeling and recognition based on soft computing approach," *IEEE Trans. Neural Netw.*, vol. 20, no. 4, pp. 563–582, Apr. 2009.
[39] D. W. Hosmer and S. Lemeshow, "Introduction to the logistic regression model," in *Applied Logistic Regression*, 2nd ed. Hoboken, NJ, USA: Wiley, 2000, pp. 1–30, doi: 10.1002/0471722146.ch1.
[40] S. Sperandei, "Understanding logistic regression analysis," *Biochemia Medica*, vol. 24, no. 1, pp. 12–18, 2014, doi: 10.11613/bm.2014.003.
[41] C. Herzet, X. Wautelet, V. Ramon, and L. Vandendorpe, "Iterative synchronization: EM algorithm versus Newton-raphson method," in *Proc. IEEE Int. Conf. Acoust. Speed Signal Process.*, May 2006, p. 4, doi: 10.1109/icassp.2006.1660988.
[42] T. Wada, S. Doi, K. Imai, N. Tsuru, K. Isaji, and H. Kaneko, "Analysis of drivers' behaviors in car following based on performance index for approach and alienation," in *Proc. SAE Tech. Paper Ser.*, Apr. 2007, pp. 547–552.
[43] G. Heinze, C. Wallisch, and D. Dunkler, "Variable selection—A review and recommendations for the practicing statistician," *Biometrical J.*, vol. 60, no. 3, pp. 431–449, May 2018.
[44] A. Kar and P. Corcoran, "Performance evaluation strategies for eye gaze estimation systems with quantitative metrics and visualizations," *Sensors*, vol. 18, no. 9, p. 3151, Sep. 2018, doi: 10.3390/s18093151.
[45] K. M. Bach, M. G. Jæger, M. B. Skov, and N. G. Thomassen, "Interacting with in-vehicle systems: Understanding, measuring, and evaluating attention," in *Proc. Brit. Comput. Soc. Conf. Hum. Comput. Interact.*, Sep. 2009, pp. 1–5.
[46] C. Miyajima, S. Yamazaki, T. Bando, K. Hitomi, H. Terai, H. Okuda, T. Hirayama, M. Egawa, T. Suzuki, and K. Takeda, "Analyzing driver gaze behavior and consistency of decision making during automated driving," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2015, pp. 1293–1298, doi: 10.1109/IVS.2015.7225894.
[47] M. Mori, C. Miyajima, T. Hirayama, N. Kitaoka, and K. Takeda, "Integrated modeling of driver gaze and vehicle operation behavior to estimate risk level during lane changes," in *Proc. 16th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2013, pp. 2020–2025, doi: 10.1109/ITSC.2013.6728526.

JUDE C. NWADIUTO was born in Imo, Nigeria, in 1992. He received the B.E. and M.E. degrees in automotive engineering from Nagoya University, in 2016 and 2018, respectively, where he is currently pursuing the Ph.D. degree with the Department of Mechanical Systems Engineering.

His research interests include modeling and analysis of human driving behavior, and the design of personalized assistance systems for automated driving.

SOICHI YOSHINO received the B.E. degree from Nagoya University, Nagoya, Japan, in 2015, and the M.E. degree from The University of Tokyo, Japan, in 2018. His research interest includes modeling and analysis of human driving behavior.

HIROYUKI OKUDA (Member, IEEE) was born in Gifu, Japan, in 1982. He received the B.E. and M.E. degrees in advanced science and technology from Toyota Technological Institute, Japan, in 2005 and 2007, respectively, and the Ph.D. degree in mechanical science and engineering from Nagoya University, Japan, in 2010.

He is currently an Assistant Professor with the Department of Mechanical Science and Engineering, Nagoya University. His research interests include the areas of system identification of hybrid dynamical system and its application to the modeling and analysis of human behavior and human-centered system design of autonomous/human-machine cooperative systems.

Dr. Okuda is a member of the IEEJ, SICE, and JSME.

TATSUYA SUZUKI (Member, IEEE) was born in Aichi, Japan, in 1964. He received the B.S., M.S., and Ph.D. degrees in electronic mechanical engineering from Nagoya University, Japan, in 1986, 1988, and 1991, respectively.

He is currently a Professor with the Department of Mechanical Systems Engineering, and an Executive Director of the Global Research Institute for Mobility in Society (GREMO), Nagoya University. His current research interests include the areas of analysis and design of human-centric intelligent mobility systems, and integrated design of transportation and smart grid systems.

Dr. Suzuki is a member of the SICE, ISCIE, IEICE, JSAE, RSJ, JSME, and IEEJ. He won the Best Paper Award in International Conference on Autonomic and Autonomous Systems 2017 and the Outstanding Paper Award in International Conference on Control Automation and Systems 2008. He also won the Journal Paper Award from IEEJ, SICE, and JSAE, in 1995, 2009, and 2010, respectively.