Research Article

TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Transportation Research Record
1–11

DOI: 10.1177/03611981221128277
journals.sagepub.com/home/trr
SAGE

Predictive Models of Driver Deceleration and Acceleration Responses to Lead Vehicle Cutting In and Out

Jason Hu[1], Brian Tsang-Wei Lin[1], Jim H. Vega[1], and Nathan Ren-Liang Tsiang[1]

Abstract

A common maneuver drivers perform and experience on the road is changing lanes. Autonomous vehicles are required to engage a lane change safely and to react to the other road users' lane changes. To develop autonomous vehicles that change lanes or respond to the lead vehicle's lane changes in a safe and human-like way, one should investigate the factors that affect human driver responses. By reviewing the literature to identify potential factors, this study extracted these factors from a naturalistic driving data set and associated them with driver deceleration and acceleration responses to the lead vehicle's cut-in and cut-out to develop predictive models for the impact of the events on traffic flow. After the events were verified as accurate, the variables associated with the events, including range, range rate, speed, lateral position in the lane, and average acceleration were analyzed using logistic regression, support vector machines (SVM), and two forms of decision trees. In total, 799 cut-in events and 684 cut-out events with the necessary variables were applied for analysis. The significant variables influencing driver behavior were found, and using these variables, the predictive models achieved around 80% accuracy for cut-ins, and 73% accuracy for cut-outs on test data. These results will assist in the future design of autonomous vehicle control to minimize detrimental effects on traffic when changing lanes and safe longitudinal control when responding to a lead vehicle's lane changes, allowing for safe integration with human drivers, and better design of driver assistance systems.

Keywords

pedestrians, bicycles, human factors, driver behavior, human factors in vehicle automation, naturalistic data studies

One of the most common vehicle maneuvers is the action of cutting in or out. This is a frequently performed maneuver at high speeds that demands many considerations and observations by the driver, as well as possibly requiring a response from surrounding vehicles. To assist in the design of systems such as automatic alerts and autonomous vehicles, it is important to understand human behavior patterns in merging, so the automated systems can respond to the maneuvers as human drivers. Autonomous vehicles should also perform in a way that minimizes the impact on following vehicles and traffic to ensure smooth flow of traffic and to minimize the chance of collisions (*1*). To develop such systems more accurately, a comprehensive review on the factors associated with driver responses to the lead vehicle is needed first.

Extensive research has been conducted on mathematical models concerning vehicles cutting in, particularly on the highway with minimal other restrictions. Time headway (THW) and time-to-collision (TTC) were commonly used independent variables as they were quantitative representations of the danger level of a following vehicle (*2–4*). Feng et al. (*2*) modeled a risk perception parameter, defined as THW and TTC, as a function of parameters of the event. Their study focused on the brake initiation of the following vehicle and found that the size of the cutting vehicle and difference in velocity were the significant factors which affected the time of brake initiation. Another important measure of cut-in events was gap acceptance, the range to the lead vehicle (LV) on cutting in. This was critical for ensuring that

[1]Human Factors Group, University of Michigan Transportation Research Institute, Ann Arbor, MI

Corresponding Author:
Brian Tsang-Wei Lin, btwlin@umich.edu

drivers felt comfortable as a response to cut-in events. Das et al. (*5*) showed how lead and lag gaps could be estimated using a multivariate adaptive regressive spline model and obtained a correlation value higher than standard methods with linear regression. Furthermore, Wang et al. (*6*) modeled gap length using a mixed effects model, including in a key variable of the event being either mandatory or discretionary, as well as finding the distribution of the gaps. In a similar fashion, Hou et al. (*7*) determined the safety of gaps using Bayesian classifier and decision trees. The present study intended to use the gap as an independent predictor, which was different from the preceding studies, opting instead to use the following vehicle's acceleration as the main dependent variable.

Other measures of surrogate safety measures (SSM) (*8*) include time integrated TTC (TIT) which better captures vehicle behavior over a period of time. This measure was used by Van Winsum et al. (*9*) in the field of vehicle safety. Another common type of measure is the post encroach time (PET), which measures the difference in time between when two vehicles pass through a common point, used by Allen et al. (*10*). Finally, a third type of measure involves the deceleration rate, including maximum deceleration and deceleration rate to avoid a crash, used by Gettman and Head (*11*). The present study would use the SSM which was closest to this third type but discretized it for classification models.

Another important factor to consider is the effect that a vehicle cutting in will have on preceding vehicles and traffic. To study this factor, a common approach is to build a predictive model for the danger of cut-in events, which is useful for the development of safety and warning systems. Dangerous maneuvers can be directly identified through braking, such as by Bagdadi (*12*), using derivatives of acceleration. However, a more ideal model would be able to expect dangerous maneuvers without requiring them to happen first. To these ends, the parameters of a cut-in event can be used to predict whether the event will have a significant impact on the behavior of the following vehicle, and therefore the danger level of the event. Xie et al. (*4*) modeled the danger level of effects using acceleration, and predicts the artificial label of the event using prior variables and machine learning models. Ma et al. (*13*) took a similar approach, using SVM to predict the danger level based on THW and relative speed. Li et al. (*14*) took a slightly different approach, using support vector regression to predict continuous parameters of events instead of a discrete dependent variable, including integrated TTC. Ma et al. (*15*) developed a simulation for the velocity of the vehicles around a cut-in event that is useful for analyzing the disruption to traffic flow of such an event.

The discretization of the danger of an event is often done somewhat arbitrarily; this study avoided this problem by splitting the data evenly via custom thresholding of the continuous dependent variable of acceleration. This led to a more even data set and higher accuracy. Furthermore, we considered a large variety of independent variables, and combined methods used by the previous work to form a model based on only the significant variables. Many of the previous studies did not consider multicollinearity and correlation between the independent variables, an important factor that could affect the accuracy and interpretation of the models. Additionally, we applied a greater variety of statistical and machine learning models that were collectively more flexible and that complement each other's strengths and weaknesses. These could also be applied to cut-out events, a subject which was far less frequently studied in the literature. Finally, while many of the existing studies used data of drivers driving on a designated route or analyzed a specific portion of the road, this study would use fully naturalistic driving data collected from the mileage on thousands of different roads to obtain models which were the most broadly applicable.

Studying the distribution of a variety of variables involving such events can be useful in assessing the probability of risk for an event. It can also help in the development of car simulation models and autonomous vehicles. For example, Li et al. (*3*) found that the duration of cut-in follows a lognormal distribution, while Feng et al. (*2*) found that THW during brake initiation also followed a lognormal distribution. Finally, a lognormal distribution was also appropriate for lag gaps, but a gamma distribution was a better fit for lead gaps (*5*). However, the fitting and evaluation models used in many cases were outdated, and we used measures such as goodness of fit that were on average more consistent.

The objectives of this study are as follows. First, study the variables that affect the driver decision-making process for each scenario. Second, develop a model for each scenario that predicts whether the event has a significant impact on the driver's behavior. For cut-in, this corresponds to discretized danger levels posed to the host vehicle by the LV. For cut-out, we evaluate whether the LV was impeding the host vehicle's desired movement. We restricted our attention to cut-in and cut-out events on the highway, to help eliminate the discrete variable of the type of road.

Methods

Data Collection

The study used naturalistic driving data from the Safety Pilot Model Deployment (SPMD) managed by the University of Michigan Transportation Research Institute. The program was held in Ann Arbor, Michigan, starting in August 2012, and included approximately 3,000 pieces of on-board equipment in vehicles driven by participants on thousands of normal commuting trips (*16*). For this research, 130 vehicles instrumented with a MobilEye device and a data acquisition system (DAS) provided the necessary data with the sampling rate of 10 Hz. Video footage captured from several cameras from different angles was available for each timestamp.

Scenario Criteria

When studying the cutting events, a few constraints were included to reduce variation. For cutting in and out, we included the restriction that the host vehicle must not have changed lanes immediately before or after the LV has performed the lane change. This ensured that the host vehicle's lateral position was relatively stable and was mostly reacting in the longitudinal direction. Second, we restricted the curvature of the road through the yaw rate of the host vehicle by an absolute value of 1° per second. Third, we required that the LV performed only one lane change throughout the event. For example, a vehicle that was cutting in might not immediately cut out to an adjacent lane, which filtered double lane changes.

For cut-ins, three key timestamps of the event were identified based on the lateral distance (transversal) to the LV. The start of a cut-in was defined as when the norm of the transversal first crossed over from being greater than 2 m to less than 2 m, which was labeled as the time T_1. This is generally when the wheels of the LV first start to cross over the lane boundary. The end of a cut-in was defined as when the norm of the transversal first crosses over from being greater than 0.5 m to less than 0.5 m (T_3). The time T_2 (between T_1 and T_3) was then defined as when the MobilEye first detected the merging vehicle as being the vehicle fully in the host lane and directly in the front, as opposed to the LV (old) before the cut-in. In other words, after T_2 the cut-in vehicle would become an LV (new).

For cut-outs, three timestamps were defined in a similar way. T_1 was defined as the start of a cut-out, indicating when the norm of the transversal first crossed over from being less than 0.8 m to greater than 0.8 m, which was about the time the LV just crossed the lane boundary. T_3 was defined as the end of a cut-out, indicating when the norm of the transversal first crossed over from being less than 2.5 m to greater than 2.5 m. This allowed T_3 to roughly be the instance when the LV (old) had settled into the adjacent lane. Finally, T_2 for cut-outs was defined analogously as the cut-ins.

Selection of Variables

Table 1 shows the independent variables and their definitions as the predictors of driver responses. The dependent variable was the longitudinal deceleration and acceleration for cut-ins and cut-outs, respectively. For cutting in, the dependent variable Y is defined as Y = 1 if the average deceleration of the host vehicle during the cut-in event is less than $-0.1\,m/s^2$, and Y = 0 otherwise. For cutting out, it is more likely that on seeing the gap to the LV increasing, the host vehicle chose to accelerate. The binary variable Y in this case is, therefore, set to 1 if the average acceleration during the cut-out event is greater than $0.2\,m/s^2$, and Y = 0 otherwise. These values were also set accordingly to partition the data somewhat evenly.

Table 1. Independent Variables

Variable name	Unit	Description
New range	m	Distance to new LV at T_2
New range rate	m/s	Change of range per second at T_2 (negative = approaching)
Old range	m	Distance to LV just before T_2
Old range rate	m/s	Change of range per second just before T_2
Yaw rate	degree/s	Host vehicle angular velocity at T_2
Speed	m/s	Speed at T_2
Lane distance left	m	Distance to the left boundary of the lane at T_2
Duration	s	Value of T_3–T_1
Direction	binary	Direction (left or right) of the cut-in/out
Previous acceleration	m/s^2	Average acceleration of host vehicle between T_1 and T_1-1 s

Note: LV = lead vehicle; T = time.

Data Reduction

To obtain instances of the relevant scenarios, the general approach was to first query a large quantity of raw data from the database in the Microsoft Structured Query Language (SQL) Server. Next, this raw data was processed to identify likely instances of the relevant scenario. Each candidate instance was then manually reviewed through the video viewing software to verify that the scenario was valid (examples shown in Figure 1). Finally, the verified instances were used to isolate out the timestamps of the raw data that were immediately preceding and following each instance to form a final spreadsheet of useful data.

The primary method of identifying cut-ins and cut-outs was to analyze the distance to the nearest vehicle in the front and observe timestamps where this distance changed dramatically. In particular, timestamps with the distance reducing by over 10 m were considered as candidate cut-in points, whereas those with the distance

Figure 1. Example for a valid lead vehicle cut-in event.

increasing by over 10 m were considered as a candidate cut-out point. Each candidate point was then analyzed in further detail to check that the LV started at the adjacent lane and the host vehicle did not perform a lane change throughout the event.

The initial query of data obtained 1,030 cut-ins and 1,029 cut-outs. On manual verification for the event validity, the numbers of remaining instances were 857 and 866, respectively. However, the research team also found that drivers did not apply the brake if the new LV was far from the host vehicle. Figure 2 shows the average range and range rate for the cut-in cases when drivers did or did not apply the brake between T_1 and T_3. The realistic interpretation of this is that if the new range was greater than 30 to 40 m, the driver might not be responding to the cut-in event at all. Thus, we selected the cut-in instances with the "new range" of less than 30 m from this point onward. To maintain consistency, we also selected cut-out instances with the "old range" of less than 30 m. With these restrictions, the numbers of available events were 799 for cut-ins and 684 for cut-outs.

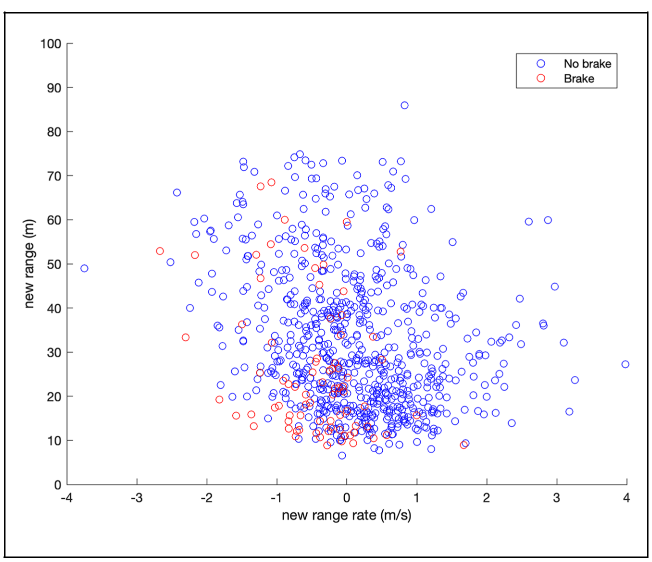

Figure 2. Scatters for range and range rate of braking/no-braking responses.

Modeling Methods

In this study, three classification methods were used to identify driver acceleration/deceleration responses: logistic regression, support vector machine (SVM) with 5-fold cross validation, and decision tree. The use of logistic regression to model the effect of independent variables on a binary dependent variable is common practice for naturalistic vehicle data (e.g., modeling the probability of drivers using one mode of braking over another [*3*]). A variation inflation factor (VIF) was used to evaluate the multicollinearity of a logistic regression. This is a measurement of multicollinearity between independent variables to check for correlation between the independent variables. Generally, the VIF was suggested to be lower than 4 for the variables to be considered independent (*17*).

Support vector machine (SVM) is a useful supervised machine learning model for classification. Xie et al. (*4*) used SVM with iteratively tuned parameters for the cost function to predict the risk of cut-in events, while Ma et al. (*13*) applied SVM with a linear kernel to evaluate the danger of cut-in events. Iranitalab and Khattak (*18*) indicated that in crash prediction, SVM performs better than other machine learning models such as random forests and clustering. To avoid overfitting and to ensure model robustness, 5-fold cross validation was employed.

Decision tree is a popular tool for data mining and supervised learning, such as in predicting driving behavior (*19*). Starting with the most significant variables, a tree is built consisting of a series of variable comparisons that lead to a classification of the data point. The tree is highly flexible and can handle any relationship between the independent variables, without any need for geometric patterns, such as SVM. However, decision trees

Table 2. Logistic Regression Results for Cut-Ins

(a)				(b)
Variable	Coefficient	t	p	Coefficient for the final model
Intercept	0.499	0.33	0.741	na[†]
New range	0.0260	1.40	0.162	na
New range rate	0.898	5.79	0.000*	0.887
Old range	0.0133	3.59	0.000*	0.0153
Old range rate	0.156	2.98	0.003*	0.171
Yaw rate	−0.222	−0.80	0.425	na
Speed	−0.0126	−0.31	0.758	na
Lane distance left	0.594	2.21	0.027*	0.490
Duration	0.102	1.12	0.235	na
Direction	−0.249	−1.08	0.280	na
Previous acceleration	5.83	11.15	0.000*	5.89

Note: *refers to p-value < 0.05; [†]refers to not applicable; na = not applicable.

are known to overfit as a result of the creation of an excessive number of nodes. Therefore, the bagged decision tree (BDT), consisting of the design of many decision trees and a voting system to decide which path to traverse given a data point, is used. Caruana and Niculescu-Mizil (*20*) demonstrated that bagged decision trees generally have a higher accuracy rate than normal decision trees and SVM. In this study, we applied both the normal decision tree and BDT to test their effectiveness and obtain additional measures of accuracy.

There are a wide range of tests available for determining the probability distribution that best fits some given data. For example, Wang et al. (*6*) used the Akaike Information Criterion (AIC) to determine the distribution of cut-in lead and lag gaps, and Varotto et al. (*21*) used the Kolmogorov Smirnov test to check that two data sets come from the same distribution. Wax and Kailath (*22*) demonstrated that the Minimum Description Length (MDL) criterion is an asymptotically unbiased measurement, as opposed to the AIC. More precisely, given a set of N observations of data arranged in a vector X and a family of models, the MDL criterion is defined as Equation 1.

$$MDL = -\log f(X|\Theta) + \frac{1}{2} k \log N \quad (1)$$

where
f is the log likelihood of the model,
Θ is the maximum likelihood estimation of the parameters, and
k is the number of free adjustable parameters in Θ.
This criterion penalizes models with many free parameters, as these are biased toward a better fit of the data. For this study, MDL was selected to evaluate the goodness of fit of probability distributions of data and the objective was to select the model that minimized the MDL.

Results

Modeling for Cut-ins

Table 2(a) shows the logistic regression analysis for cut-ins that the variables with significant effect on driver deceleration responses were the old range, new range rate, old range rate, lane distance left, and previous acceleration, by using the significant level of 0.05. Another round of logistic regression was then run only on the significant variables, and found no changes in the significance, with coefficients denoted in Table 2(b). The VIFs for these significant variables were all lower than 1.5 which indicated that no multicollinearity was found. The prediction model for the probability of significant vehicle deceleration for cut-ins is shown in Equation 2.

$$P(Deceleration) = \frac{1}{1 + e^{0.887X_1 + 0.0153X_2 + 0.171X_3 + 0.49X_4 + 5.89X_5}} \quad (2)$$

where X_1 through X_5 respectively represent new range rate, old range, old range rate, lane distance left, and previous acceleration. The overall accuracy of the cut-in model was 81%.

SVM and decision trees (normal and BDT) were then applied, only with the significant variables as independent predictors. A chi-square goodness-of-fit test was conducted to evaluate the consistency of the errors and there was no significant difference between the folds. A linear kernel was found to be optimal for the SVM prediction model. Figure 3 shows the comparison between the original naturalistic data and the prediction with SVM, using the two most significant variables (old range rate and previous acceleration) as an example. Table 3 further shows the errors of the prediction by the three classification methods that SVM and BDT performed better than the decision tree.

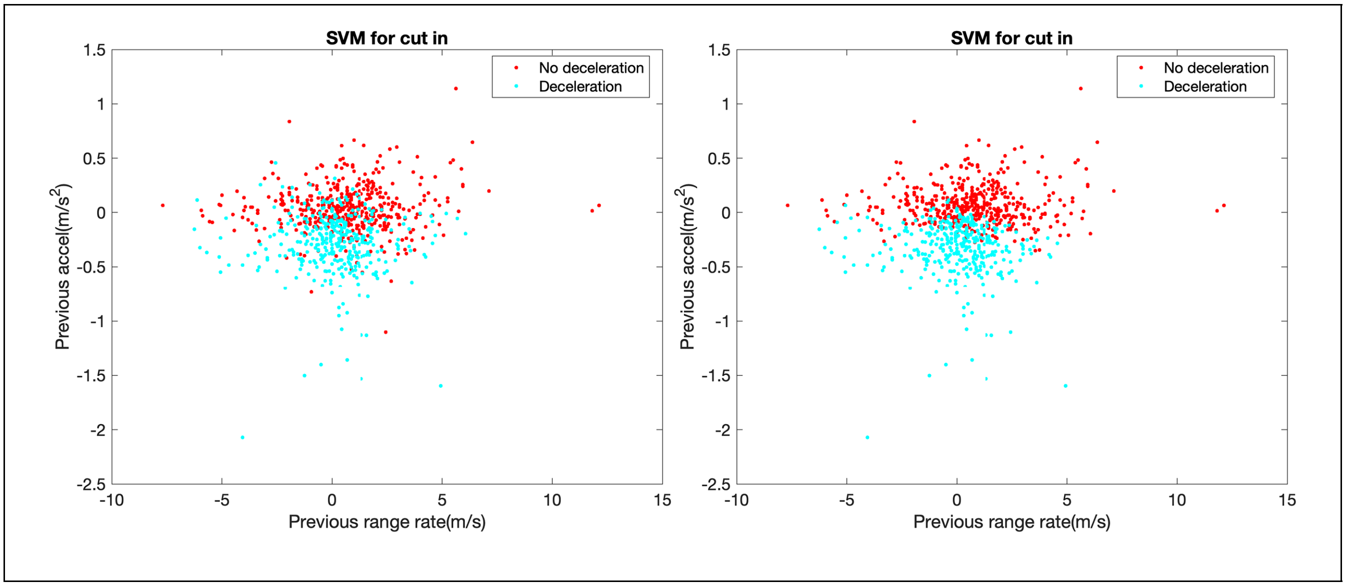

Figure 3. Comparison between the naturalistic data and the prediction with support vector machine (SVM) for cut-ins.

Table 3. Number of Errors for Each Fold: Cut-Ins

	Number of prediction errors			
Fold	SVM	Decision tree	BDT	Total cases
1	34	45	36	159
2	30	41	26	160
3	33	34	25	160
4	25	30	30	160
5	31	32	36	160
Total	153	182	153	799

Note: SVM = support vector machine; BDT = bagged decision tree.

Modeling for Cut-outs

Table 4(a) shows the logistic regression analysis for cut-ins that the variables with significant effect on driver deceleration responses were the new range, old range, speed, lane distance left, and previous acceleration. Another round of logistic regression was then run only on the significant variables, from which no changes were found in the significance and VIF were lower than 1.5 that was the same as the results for cut-ins. The final coefficients for the classification model are shown in Table 4(b) and the prediction model for the probability of significant vehicle acceleration for cut-outs is shown in Equation 3.

$$P(Acceleration) = \frac{1}{1 + e^{-6.98-0.0091X_1 + 0.076X_2 + 0.205X_3-4.45X_4}} \quad (3)$$

where X_1 through X_4 respectively represent new range, old range, speed, and previous acceleration. The overall accuracy of the cut-out model was 73%.

In the same manner as for cut-ins, SVM and decision tree models were run on the cut-outs. With the optimal linear kernel for SVM, Figure 4 shows the comparison between the original naturalistic data and the prediction with the two most significant variables (speed and previous acceleration) as an example. Among the three classification methods, the results shown in Table 5 indicate that the SVM performed the best, followed by BDT and the decision tree.

Classification Model Comparisons

Confusion matrices were created to compare the predictive results using the four classification methods, as shown in Table 6. In general, the predictions for cut-ins were more accurate than for cut-outs. In addition, the miss rates were almost 20% higher than false positive rates for cut-out prediction, but the difference decreased to less than 10% for cut-in. The accuracies of the classification for driver deceleration responses to LV cut-ins by logistic regression, SVM, decision tree, and BDT were 81.0%, 80.9%, 77.2%, and 80.9%, respectively. For LV cut-outs, the accuracy for identifying driver acceleration responses by the four models decreased to 73.2%, 72.4%, 65.9%, and 70.8%. In both cases, the models by logistic regression, SVM, and BDT had similar performance on accuracy, false negative rate (miss rate), and false positive rate (fall-out) that was better than the decision tree model.

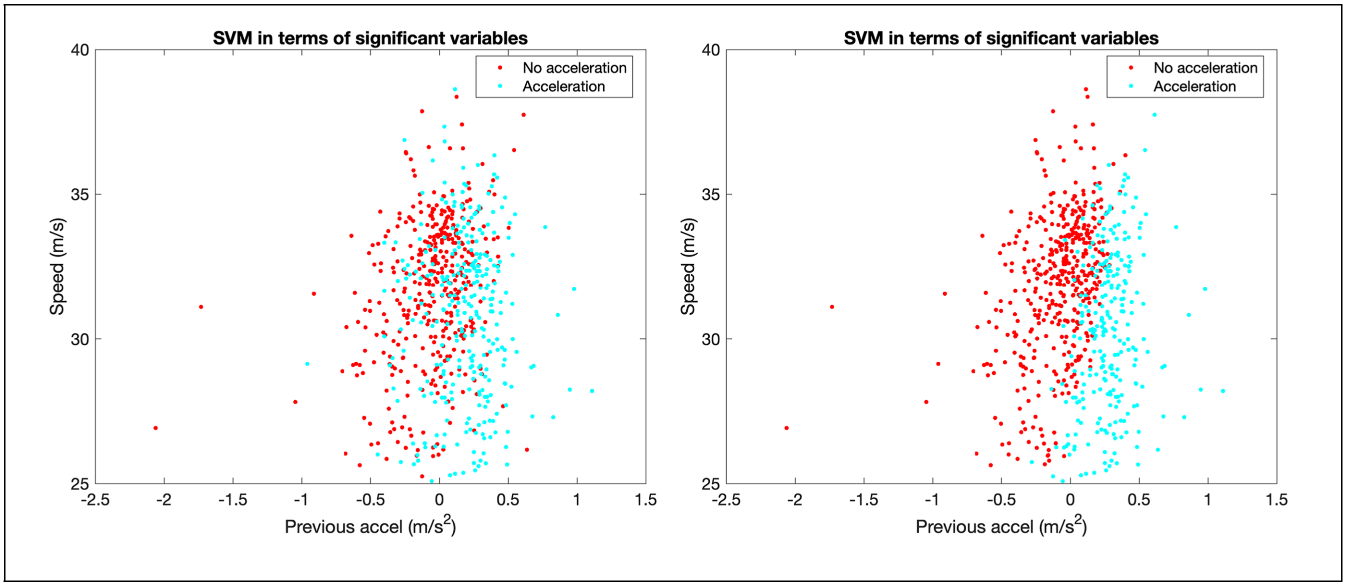

Figure 4. Comparison between the naturalistic data and the prediction with support vector machine (SVM) for cut-outs.

Table 4. Logistic Regression Results for Cut-Outs

(a)				(b)
Variable	Coefficient	t	p	Coefficient for the final model
Intercept	−7.25	−4.95	0.000*	−6.98
New range	−0.0102	−2.79	0.005*	−0.0091
New range rate	−0.0245	−0.45	0.655	na†
Old range	0.0756	4.02	0.000*	0.076
Old range rate	0.0023	0.07	0.940	na
Yaw rate	0.0670	0.27	0.791	na
Speed	0.207	5.41	0.000*	0.205
Lane distance left	−0.0315	−0.13	0.896	na
Duration	0.130	1.18	0.240	na
Direction	−0.0885	−0.48	0.631	na
Previous acceleration	−4.44	−10.23	0.000*	−4.45

Note: *refers to p-value $>$ 0.05; †refers to not applicable; na = not applicable.

Discussion

This study has provided a naturalistic basis for determining the impact of an LV cut-in or cut-out on the following vehicle. In the logistic regression model, positive coefficients of a variable indicate that an increase of the variable was correlated with a lower probability of the response of deceleration. For cut-ins, a higher range rate of the new LV, range and range rate of the original LV, distance to the left boundary, and the previous acceleration were all correlated with a lower probability of the driver choosing to decelerate. One explanation for the prevalence of the previous vehicle dynamics is that higher range and range rate to the original LV allow the cutting vehicle to rapidly accelerate on cutting in. A common occurrence is for an LV to be accelerating on an adjacent lane to overtake the host vehicle, which is only possible if the vehicle in front is either a significant distance away, or currently accelerating.

Table 5. Number of Errors for Each Fold: Cut-Outs

	Number of prediction error			
Fold	SVM	Decision tree	BDT	Total cases
1	47	44	32	136
2	34	37	40	137
3	33	46	46	137
4	40	51	45	137
5	35	56	40	137
Total	189	233	203	684

Note: SVM = support vector machine; BDT = bagged decision tree.

Table 6. Confusion Matrices for the Four Classification Methods

Cut-in					
Logistic regression	Predicted deceleration	Predicted no-deceleration	SVM	Predicted deceleration	Predicted no-deceleration
Deceleration	273 (34.2%)	82 (10.3%)	Deceleration	276 (34.5%)	79 (9.9%)
No-deceleration	70 (8.8%)	374 (46.8%)	No-deceleration	74 (9.3%)	370 (46.3%)
Decision tree	Predicted deceleration	Predicted no-deceleration	BDT	Predicted deceleration	Predicted no-deceleration
Deceleration	259 (32.4%)	96 (12.0%)	Deceleration	282 (35.3%)	78 (9.8%)
No-deceleration	95 (11.9%)	349 (43.7%)	No-deceleration	76 (9.5%)	363 (45.4%)
Cut-out					
Logistic regression	Predicted acceleration	Predicted no-acceleration	SVM	Predicted acceleration	Predicted no-acceleration
Acceleration	191 (27.9%)	110 (16.1%)	Acceleration	181 (26.4%)	120 (17.5%)
No-acceleration	73 (10.7%)	310 (45.3%)	No-deceleration	69 (10.1%)	314 (45.9%)
Decision tree	Predicted acceleration	Predicted no-acceleration	BDT	Predicted acceleration	Predicted no-acceleration
Acceleration	188 (27.5%)	125 (18.3%)	Acceleration	178 (26.0%)	125 (18.3%)
No-acceleration	113 (16.5%)	258 (37.7%)	No-acceleration	72 (10.5%)	309 (45.2%)

Note: SVM = support vector machine; BDT = bagged decision tree.

For cutting out, higher speed and old range were correlated with lower probability of acceleration, whereas higher new range and previous acceleration were correlated with higher probability of acceleration. The former of these correlations makes sense, as an LV traveling at higher speed is less likely to impede the host vehicle's desired travel, and drivers are more likely to follow with a shorter distance. A greater distance to the new LV following the cut-out event will give more freedom to the host vehicle to freely accelerate. Notably, the speed of the host vehicle was only significant for cut-out, but not for cut-in. A possible explanation for that is the restriction that the vehicle be traveling above 25 m/s on the highway: for cut-in, the driver has a higher priority to react to the event and avoid collision, emphasizing relative speed, whereas for cut-out, the driver will be more likely to consider the speed limit when accelerating, as a driver traveling near the speed limit is unlikely to accelerate.

The accuracy for the cut-out acceleration prediction was significantly lower than that of the cut-in deceleration prediction. One likely explanation for this is that following an LV cut-out, the host vehicle has more freedom to maneuver, and may choose to accelerate or maintain current speed with human randomness. In contrast, on having an LV cut-in with range under 30 m, a previously accelerating host vehicle has much less freedom, especially when eliminating possibilities of lane change. Furthermore, it may be useful for future studies to analyze the acceleration of the host vehicle for a period of time after the event has occurred to determine if some drivers have a delayed reaction to the event.

In all cases, the accuracy results of the logistic regression, SVM, and BDT were consistent, while the decision tree performs notably worse. The latter result was to be expected, as decision trees commonly experience overfitting (*23*). The consistency of the methods demonstrates the robustness of the methods and shows that inherent noise and randomness of the data is largely responsible for the inaccuracies. SVM in this case performed well, as the data was well separated by a linear kernel. Further improvements would likely require consideration of new variables and querying of new data.

The present study builds on and improves the results from the existing literature on responses to lane changes. Feng et al. (*2*) showed that drivers tend to brake earlier with increasing relative velocity, consistent with our findings that higher range rate was correlated with higher probability of significant deceleration by the driver. Das et al. (*5*) and Hou et al. (*7*) found that the most significant predictors of driver response were relative range and range, and Wang et al. (*6*) also found range rate to be a significant variable. This was consistent in our findings, as it was necessary to restrict the range to the newly cut-in vehicle to under 30 m for any variables to be significant, indicating general driver unresponsiveness to a more distant cut-in event. Additionally, the p value for the new range rate was one of the lowest out of the studied independent variables.

Research Implications

This study has several implications on the existing literature and future studies. Ma et al. (*13*) demonstrated that time headway (THW, calculated by distance divided by host vehicle velocity) and relative velocity to the LV are accurate predictors for the danger level of a cut-in event, where the danger level was defined as driver reaction time and longitudinal acceleration. However, the threshold for danger levels beyond the base level was very high, so the vast majority of data fell under the lowest level. The present study extends this work by examining many more possibly significant independent variables, as well as setting a threshold of disturbance met by far more cut-in events. The considered variables included enough information to derive the THW, and under the highway restriction found differing results. Similarly, Xie et al. (*4*) use clustering on a graph of average and maximum deceleration to assign danger levels. However, the unsupervised clustering method appears to be somewhat arbitrary, as the data points do not naturally fall under evident clusters, and the strong linear relationship between average and maximum deceleration lead to the risk levels being nearly discretized by random thresholds of average deceleration. The present study consistently classifies events as affecting the host vehicle above or below the average, with a consideration of all variables shown to be significant.

In applying the models to the design of autonomous vehicles, two perspectives may be taken. First, from the perspective of a vehicle looking to perform a discretionary lane change, an autonomous vehicle should strive to do so in a manner that avoids a significant acceleration response from the vehicle behind, possibly through a speed or range change, as this will help in maintaining steady traffic flow, as well as minimize collision risk to unalert drivers. On the other hand, from the perspective of a vehicle behind another vehicle performing the lane change, the models serve as a guide for an appropriate response for the autonomous vehicle. These are most effective in situations where the desired response, either accelerating or braking sharply, is not immediately obvious as a result of an imminent threat of collision, as the classification is based on a very slight value of acceleration and is thus most effective when the appropriate response consists of a low acceleration. For events with a much greater threat of collision, other models, such as those concerning the avoidance of an immediate collision with a low TTC, would be more applicable to determine the absolute value of the necessary acceleration of the autonomous vehicle.

Limitations

Given the nature of the data acquisition systems and system of experimentation, a few limitations should be noted. The first involves the choices of independent variables that were made. Notable variables that were not included in the analysis included the weather (may increase the acceptable gap [*6*]), road conditions, influence of vehicles other than the LV for cutting in and out, and speed limit. Furthermore, though the presence of nearby vehicles other than the LV in a cut-in or out maneuver may affect driver decision making, several different variables would need to be considered, possibly one for each nearby vehicle. This would cause thinning of data as well as possible overfitting. Lastly, speed limit is possibly an important factor in influencing whether a driver decides to include deceleration or acceleration as part of their collision avoidance maneuver.

Conclusions and Future Work

The use of the SPMD naturalistic database allows for the investigation of the relevant events in natural settings. Combinations of variables set thresholds for parameters of cut-in and cut-out events that determine whether or not the event has a significant impact on driver behavior. The study quantifies the intuitive decision-making process drivers naturally make that has been developed throughout the years of driving experience.

Knowledge of the conditions under which a vehicle cutting in has a significant behavioral impact on the preceding driver can help in the development of autonomous vehicles. The developed predictive models allow for an evaluation by an autonomous vehicle planning a lane change of the impact on the following vehicles in the designated lane. Unless the cut-in is mandatory, autonomous vehicles should avoid making a lane change that prompts the preceding driver to perform a significant deceleration, as it has a detrimental effect on the natural flow of traffic. Alert systems on vehicles can also be developed that indicate possibly dangerous cut-in events.

For the development of autonomous vehicles, this research provided adjustable classification models by selecting different thresholds. The confusion matrices were based on the thresholds that maximized the summation of true positive and true negative rates. The developer should weight the trade-offs between misses (lower thresholds) and false alarms (higher thresholds) by considering cost, safety, policy, and end users. Having human drivers in autonomous vehicles to subjectively evaluate the thresholds can help select more acceptable thresholds.

Future work could involve modeling the dynamics of the vehicle performing the lane change in a way that incorporates these minimal traffic disruption models, as well as modeling the lateral movement behavior of the driver in addition to the longitudinal movement as in our work. This could also include further time-series data analysis. Additionally, external environment factors such as road conditions, lane count, and weather can be considered in the models, which could improve the prediction accuracies. Finally, the research team plans to expand this study to experiments with human subjects in an autonomous vehicle to identify the optimal thresholds.

Acknowledgments

The authors would like to thank the University of Michigan Transportation Research Institute for supporting the naturalistic driving data. Jim Vega was not available to confirm coauthorship, but the corresponding author Brian T. W. Lin affirms that Jim Vega contributed to the paper and vouches for Jim Vega's coauthorship status.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: Jason Hu, Brian Lin, Jim Vega, Nathan Tsiang; data collection: Jason Hu, Jim Vega, Nathan Tsiang; analysis and interpretation of results: Jason Hu, Brian Lin, Jim Vega, Nathan Tsiang; draft manuscript preparation: Jason Hu, Brian Lin. All authors reviewed the results and approved the final version of the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Brian Tsang-Wei Lin https://orcid.org/0000-0003-0425-7586
Jim H. Vega https://orcid.org/0000-0001-7721-2480

References

1. Friedrich, B.The Effect of Autonomous Vehicles on Traffic. In *Autonomous Driving* (J. Gerdes, M. Maurer, B. Lenz, and H. Winner, eds.), Springer, Berlin, Heidelberg, 2016.
2. Feng, Z., X. Ma, X. Zhu, and Z. Ma. Analysis of Driver Brake Behavior Under Critical Cut-In Scenarios. *Proc., 2018 IEEE Intelligent Vehicles Symposium (IV)*, Changshu, China, 2018.
3. Li, S., P. Li, Y. Yao, X. Han, Y. Xu, and L. Chen. Analysis of Drivers' Deceleration Behavior Based on Naturalistic Driving Data. *Traffic Injury Prevention*, Vol. 21, No. 1, 2020, pp. 42–47.
4. Xie, G., H. Qin, M. Hu, D. Ni, and J. Wang. Modeling Discretionary Cut-In Risks Using Naturalistic Driving Data. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 65, 2019, pp. 685–698.
5. Das, A., M. N. Khan, and M. M. Ahmed. Nonparametric Multivariate Adaptive Regression Splines Models for Investigating Lane-Changing Gap Acceptance Behavior Utilizing Strategic Highway Research Program 2 Naturalistic Driving Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2020. 2674: 223–238.
6. Wang, X., M. Yang, and D. Hurwitz. Analysis of Cut-In Behavior Based on Naturalistic Driving Data. *Accident Analysis & Prevention*, Vol. 124, 2019, pp. 127–137.
7. Hou, Y., P. Edara, and C. Sun. Modeling Mandatory Lane Changing Using Bayes Classifier and Decision Trees. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 15, No. 2, 2014, pp. 647–655.
8. Hydén, C., and L. Linderholm. The Swedish Traffic-Conflicts Technique. In *International Calibration Study of Traffic Conflict Techniques* (E. Asmussen, ed.), Springer, Berlin, Heidelberg, 1984. https://doi.org/10.1007/978-3-642-82109-7_12.
9. Van Winsum, W., K. A. Brookhius, and D. Waard. A Comparison of Different Ways to Approximate Time-to-Line Crossing (TLC) During Car Driving. *Accident Analysis & Prevention*, Vol. 32, 2000, pp. 47–56.
10. Allen, B. L., B. T. Shin, and D. J. Cooper. Analysis of Traffic Conflicts and Collision. *Transportation Research Record: Journal of the Transportation Research Board*, 1978. 667: 67–74.
11. Gettman, D., and L. Head. Surrogate Safety Measures From Traffic Simulation Models. *Transportation Research Record: Journal of the Transportation Research Board*, 2003. 1840: 104–115.
12. Bagdadi, O. Assessing Safety Critical Braking Events in Naturalistic Driving Studies. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 16, 2013, pp. 117–126.
13. Ma, X., Z. Ma, X. Zhu, J. Cao, and F. Yu. *Driver Behavior Classification Under Cut-In Scenarios Using Support Vector Machine Based on Naturalistic Driving Data*. SAE Technical Paper 2019-01-0136. Society of Automotive Engineers, Detroit, 2019.
14. Li, M., Z. Li, C. Xu, and T. Liu. Short-Term Prediction of Safety and Operation Impacts of Lane Changes in Oscillations With Empirical Vehicle Trajectories. *Accident Analysis & Prevention*, Vol. 135, 2020, p. 105345.
15. Ma, Y., Z. Lv, P. Zhang, and C.-Y. Chan. Impact of Lane Changing on Adjacent Vehicles Considering Multi-Vehicle Interaction in Mixed Traffic Flow: A Velocity Estimating Model. *Physica A: Statistical Mechanics and its Applications*, Vol. 566, 2021, p. 125577.

16. Bezzina, D., and J. Sayer. *Safety Pilot Model Deployment: Test Conductor Team Report*. DOT HS 812 171. National Highway Traffic Safety Administration, Washington, D.C., 2015.
17. Hair, J. F., W. C. Black, B. J. Babin, and R. E. Anderson. *Multivariate Data Analysis*, 7th ed. Prentice Hall, Upper Saddle River, NJ, 2010.
18. Iranitalab, A., and A. Khattak. Comparison of Four Statistical and Machine Learning Methods for Crash Severity Prediction. *Accident Analysis & Prevention*, Vol. 108, 2017, pp. 27–36.
19. Wang, X. Y., and X. Y. Yang. Study on Decision Mechanism of Driving Behavior Based on Decision Tree. *Journal of System Simulation*, Vol. 20, No. 2, 2008, pp. 415–420.
20. Caruana, R., and A. Niculescu-Mizil. An Empirical Comparison of Supervised Learning Algorithms. *Proc., 23rd International Conference on Machine Learning - ICML '06*, Pittsburgh, PA, 2006.
21. Varotto, S. F., R. Jansen, F. Bijleveld, and N. van Nes. Driver Speed Compliance Following Automatic Incident Detection: Insights From a Naturalistic Driving Study. *Accident Analysis & Prevention*, Vol. 150, 2021, p. 105939.
22. Wax, M., and T. Kailath. Detection of Signals by Information Theoretic Criteria. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 33, No. 2, 1985, pp. 387–392.
23. Leiva, R. G., A. F. Anta, V. Mancuso, and P. Casari. A Novel Hyperparameter-Free Approach to Decision Tree Construction That Avoids Overfitting by Design. *IEEE Access*, Vol. 7, 2019, pp. 99978–99987.