Modeling Vehicle Merging Behavior in Work Zone Merging Areas During the Merging Implementation Period

Jinxian Weng, Shan Xue, and Xuedong Yan

***Abstract*—Using the classification and regression tree (CART) method, this study aims to model the vehicle merging behavior at work zone merging areas during the merging implementation period. Hereafter, the merging implementation period is defined as the period from the starting time to the completion time of a merging maneuver. From the safety perspective, the times to collision (TTC) between the merging vehicle and its neighboring vehicles are regarded as the factors affecting vehicle merging decisions. The results show that a larger delay, a shorter remaining distance to the work zone, a smaller TTC to the merging lead vehicle, and a higher merging vehicle speed may encourage drivers to complete merging maneuvers early. It is also found that the merging vehicle tends to continue merging when the TTCs between the merging vehicle and its neighboring vehicles in the through lane are too small. In addition, another CART model without the use of TTC is built for comparison. The finding indicates that the use of TTC can not only contribute to a perfect result but also highlight the merging safety situation more clearly. The merging rules of this study are ready to be incorporated into the merging assistance system for guiding a safety merging at work zone merging areas.**

***Index Terms*—Merging, classification and regression tree, time-to-collision, merging implementation period.**

I. Introduction

In recent years, with the development of communication and network technology, internet of things has been applied in various fields, such as medical service, logistics management, agricultural industry, education, intelligent transportation system and etc [1]–[6]. One important application of this advanced technology is Vehicular Ad-hoc Network (VANET) in intelligent transportation system, which is considered as promising solution to many traffic problems. VANETs, as a new communication mode, regards every single vehicle as an information unit, all vehicles equipped with wireless interfaces are able to exchange their information with nearby vehicles as well as fixed roadside infrastructures. So far, many designs based on VANETs have been applied in the traffic engineering field. For instance, an adaptive traffic signal control system, which will alter the timing patterns according to the traffic demand, based on car-to-car communication has been widely used. This system reduces the waiting time of the vehicles at the intersection along with the reduction in queue length [7]. VANETs is also adopted in the vehicle-to-vehicle early-warning system, The principle is that vehicles on a stretch of roads send warning messages to vehicles in the opposite direction, so as to warn them regarding a dangerous situation ahead [8]. Moreover, some intelligent environment-friendly vehicles (i-EFV) ensure their safety on the road by using intelligent information interaction devices based on VANETs [9]. Hence, the use of these systems could provide a more convenient and safer driving.

Note that one of the key components VANETs is the merging decision assistance system [10]. In general, the aim of merging decision assistance system is to advise drivers taking a safer merging behavior for reducing the vehicle crash risk. It depends on judgment rules (i.e., merging rules and non-merging rules) which are preset in the system to tell the driver whether it is a safe merging or not at present time.

The existing merging decision assistance systems can work well in the situation of discretionary lane-changing. In reality, the discretionary merging judgment rules are not applicable for some special road environment, such as work zones on the road. Work zone is defined as a stretch of roadway where road maintenance or construction activities are undertaken. Workers usually close a part of existing lanes in work zone in order to protect their safety. In other words, there is a large number of mandatory merging behaviors occurred in work zone merging area, all vehicles in the merge lane have to merge into the adjacent through lane before passing through the work zone.

The significant increase of the merging maneuvers in work zone merging areas could result in higher traffic crash potential. It was reported that the accident rates can increase by 20%–50% during road construction or maintenance periods [11]–[13]. In addition, the rear-end crash was found to be the major accident type in work zones according to the findings from Garber and Zhao [14]. Srinivasan *et al.* [15] further investigated that the majority of work zone rear-end crashes occur in work zone merging areas, and that the crash accidents in work zone merging areas are likely to be more severe. Therefore, it is important and necessary to explore the specific merging rules in the work zone merging areas for a merging decision assistance

Manuscript received February 16, 2015; revised May 24, 2015; accepted September 4, 2015. Date of publication October 5, 2015; date of current version March 25, 2016. This work was supported by the National Natural Science Foundation of China under Grants 51308037, 51338008, and 71210001. The Associate Editor for this paper was D. Cao. *(Corresponding author: Jinxian Weng.)*

J. Weng is with the College of Transport and Communications, Shanghai Maritime University, Shanghai 201306, China (e-mail: wjx206@gmail.com).

S. Xue and X. Yan are with the MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2015.2477335

system. As depicted in Fig. 1, an information interaction exists between vehicles and roadside infrastructure via VANETs, merging vehicles receive advices according to their real-time traffic environment. If the state of a merging vehicle matches with a merging/non-merging rule, the system will send the corresponding advice to the merging vehicle.

Considering the fact that it usually takes more than one second to complete a merging movement, there is a critical need to model the merging behavior during the entire merging implementation period from the time of starting a merging maneuver to that of completing the maneuver. The rules extracted from the merging behavior during the merging implementation period can be also applied to the merging assistance system for improving the system performance.

The rest of this paper is organized as follows. Section II provides a critical review on the existing relevant literature. Section III describes the data used for building the CART model. Section IV gives the methodology to build the CART model. The CART model results are presented in Section V. Section VI concludes with future directions of work.

II. Literature Review

The merging decision rule is an essential component of a merging decision assistance system. Accurate decision rules contribute to a high-quality assistance system. The decision rules usually are extracted from the merging behavior models. Numerous models related to drivers' merging behavior were developed in previous studies. For example, Kita [16] proposed a game theory based model to calculate the merging probability using the maximum likelihood estimation technique. In addition, Herman [17], Drew [18], and Miller [19] assumed an exponential distribution, a lognormal distribution and a normal distribution for critical gaps, respectively. Moreover, some merging models have been applied in microscopic traffic simulators. Yang and Koutsopoulos [20] established a rule-based lane changing model which has been applied in the microscopic simulator MITSIM. The ARTEMiS traffic simulator adopted intelligent-agent-based techniques to describe drivers' merging behavior [21]. Among these models, most models are the gap acceptance models with an assumption that a driver will take a lane-changing if both the adjacent lag and lead gaps are acceptable [22]–[25]. However, this assumption is inconsistent with the reality. To avoid this inconsistency, Weng and Meng [26] and Jia [27] developed logistic models to estimate the merging probability in work zone merging areas and urban expressway merging sections, respectively, the variables of vehicle speed, gap and remaining distance were considered. Fatema [28] investigated the lane-changing behavior based on acceleration and gap. Apart from parametric statistical approach, many researchers also built non-parametric models which can provide higher prediction accuracy for merging behavior (e.g., [10]).

Considering the fact that it takes more than one second to complete a lane change maneuver, many researchers presented various dynamic models to describe the dynamic lane-changing behavior. For example, Ahmed [29] proposed the dynamic gap acceptance model to replicate the real-time lane changing behavior. In [30], Oliver and Pentland used hidden Markov models (HMM) to recognize the lane change state. Smith *et al.* (2003) [31] used four states to reflect the severity level of lane-change behavior according to relative longitudinal distance and speed. McCall *et al.* [32] used the latitudinal position and head motion to infer lane-change intent. Meng and Weng [33] developed Cellular Automata models incorporating the dynamic lane-changing behavior. These models are able to determine when and where lane-change maneuver starts and ends.

Nevertheless, there still exist two limitations in previous models mentioned above. First, the majority of models ignored the effect of merging time elapsed after a merging action being triggered. Second, previous models only considered vehicle speeds and gaps as major contributory factors affecting drivers' merging behavior. This may lead to the difficulty in explaining drivers' merging behavior under some special circumstances. For instance, these models may ask the merging vehicle to make a lane change when the relative speed between the merging and through lead vehicles is large but their gap distance is very small. However, from the safety perspective, the merging vehicle driver may not make a lane change under this situation because of the high accident crash risk in reality. To overcome this shortcoming, a surrogate safety measure (SSM) combining vehicle speed and gap is selected as the explanatory variable in our model. The SSM can reflect the crash risk level between two adjacent vehicles based on their real-time trajectory data [34], [35]. For example, Gettman and Head [36] considered the deceleration rate, maximum speed and maximum speed standard deviation as three surrogate safety measures to investigate the traffic accident risk. Cunto and Saccomanno [37] employed the deceleration rate to avoid the crash (DRAC) to evaluate the individual vehicle risk. Weng and Meng [38] selected the time-to-collision (TTC) to measure the rear-end crash risk. Gao *et al.* [39] analyzed the freeway work zone safety using two safety surrogate measures including the TTC and DRAC.

In summary, the above discussion clearly indicates that the literature regarding the vehicle merging behavior in work zone merging areas during the entire merging implementation period is rather limited. Therefore, the objective of this study is to model the vehicle merging behavior in work zone merging areas during the merging implementation period. From the safety perspective, the use of relative vehicle speeds and vehicle gaps may produce unacceptable safety problems. Theoretically, the much higher relative vehicle speed but a shorter vehicle gap may still produce a bigger lane changing/merging probability according to previous models. However, this may result in an impending collision. Hence, it is more realistic to use the SSMs between the merging vehicle and neighboring vehicles as influencing factors to model the vehicle merging behavior. If the merging assistance system incorporates the merging decision rules during the merging implementation period, the vehicle crash risk at work zone merging areas can be greatly reduced.

The contributions of this study are three-fold. First, this study makes an initial attempt to model the vehicle merging behavior in work zone merging areas. Our model focuses on the entire merging implementation period and the time elapsed since a merging action being triggered is also considered as an influencing factor. Second, instead of the direct use of vehicle

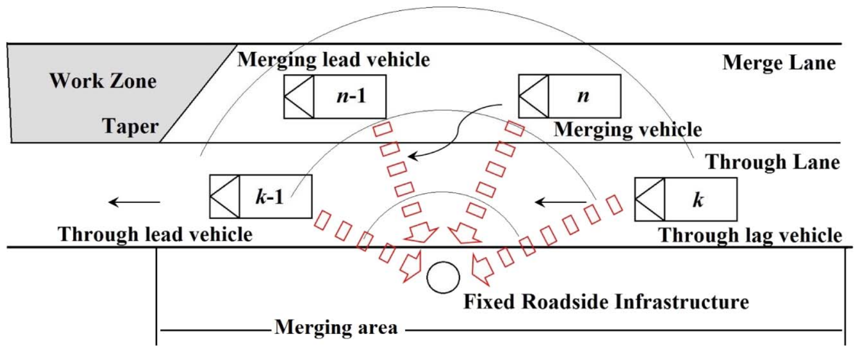

Fig. 1. Principle of a work zone merging decision assistance system.

Fig. 2. The layout of Ang Mo Kio work zone site.

speeds and space gaps, the SSMs (e.g., TTC) are regarded as influencing factors affecting the merging behavior, which were rarely taken into account in previous studies related to the analysis of lane changing/merging behavior. Third, the merging decision rules adapting to work zone traffic environment can be extracted from our merging behavior model, which can provide technical supports for the work zone merging assistance system.

III. Data Description

To model the vehicle merging behavior at work zone merging areas during the merging implementation period, we conducted a field survey on a work zone site located on the Ang Mo Kio Avenue 3 in central Singapore (Fig. 2). The work crew also closed the fast lanes for removing trees and there were only two lanes available for the traffic. Vehicles in the fast lane had to merge into the adjacent through lane to continue their trips. Therefore, the blocked fast lane can be considered as the merge lane, and vehicles in the fast lane are termed merging vehicles. Traffic engineers posted a speed limit of 70 km/h for this work zone site. We assume that a merging vehicle desires to change lane at time t and completes its merging maneuver (define that a merging vehicle's centerline enter into the through lane) at time $t + T$. Hereafter, T is the vehicle merging duration. The entire merging duration can be divided into several time intervals in this study, as shown in Fig. 3. There are two choices for a merging vehicle in the whole merging maneuver: (i) "continue merging" if a merging vehicle driver thinks that it is unsuitable to enter the adjacent through lane under the present condition or (ii) "complete merging" if he/she completes his/her merging maneuver safely within the next time interval. The merging vehicle's state will be changed into "complete merging" from "continue merging" at time $t + T$.

The vehicle merging behavior data at work zone merging areas were collected by video cameras in the good weather condition. We extracted vehicle trajectory data from the videotapes every second using Premiere Pro 3.0 software which can display 30 frames per second, with an error of 0.03 second. Due to the analyst's visual judgment error for vehicle positions, the total possible error could be up to 0.1 second [40]. The speed of a vehicle was determined by calculating the time taken by the vehicle to cover two lane-markers' distance in the video. As deemed by Strong *et al.* [41], this measurement method could yield data of comparable quality to the radar-speed measurement method, which produces the potential error of ±0.8 km/h. According to the videotapes, traffic volume approaching the site ranged from 815 vph to 1644 vph and the average traffic speed ranges from 18 km/h to 66 km/h. Finally, we extracted all the records containing the information regarding vehicle speeds, positions and types of the merging vehicle and its neighboring vehicles and time elapsed after a merging action being triggered.

We collected 1562 records from 283 merging vehicles in the work zone site. In other words, the data sample was comprised of 283 records with the choice of "complete merging" (i.e., $Y(n) = 1$) and 1279 records with the choice of "continue merging" (i.e., $Y(n) = 0$).

However, as mentioned above, the separate use of the variables including speed and space gap may not fully reflect the risks between the merging vehicle and its neighboring vehicles during the merging implementation period. Hence, the variable of time-to-collision (TTC), which is a surrogate safety measure combining vehicle speeds and space gap, is used. Hereafter, the TTC is defined as the time that remains until a collision between two vehicles would have occurred if the collision course and speed difference remain unchanged. It can be regarded as a measure to explain the risk level between two vehicles. The TTC between the merging vehicle n and merging lead vehicle $n-1$, denoted by $TTC(n-1)$, can be computed by

$$TTC(n-1) = \frac{X(n) - X(n-1) - L(n-1)}{V(n) - V(n-1)}, \ \forall V(n) > V(n-1) \tag{1}$$

where $X(n-1)$ is the longitudinal position of the lead vehicle, $V(n-1)$ is the lead vehicle speed and $L(n-1)$ is the length of lead vehicle.

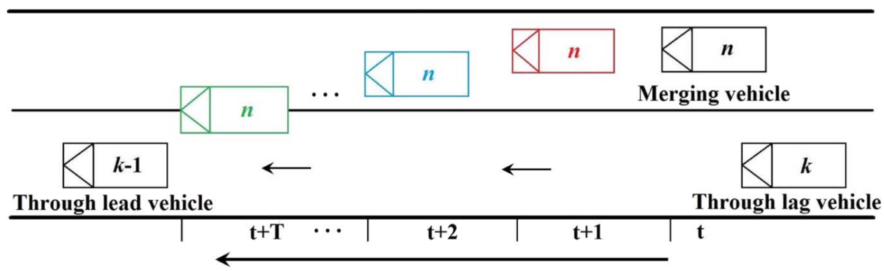

Fig. 3. Merging maneuver of a merging vehicle.

TABLE I
DESCRIPTIONS OF EXPLANATORY VARIABLES

Variable	Descriptions
$V(n)$	The speed of merging vehicle *n*
$X(n)$	The remaining distance of the merging vehicle *n* to the work zone
$C(n)$	The type of merging vehicle *n* (*1* for car; *2* for heavy vehicle)
$TTC(n-1)$	The *TTC* between merging vehicle *n* and merging lead vehicle *n*-1
$C(n-1)$	The type of merging lead vehicle *n*-1 (*1* for car; *2* for heavy vehicle; *3* for no lead vehicle)
$TTC(k)$	The *TTC* between merging vehicle *n* and through lag vehicle *k*
$C(k)$	The type of through lag vehicle *k* (*1* for car; *2* for heavy vehicle)
$TTC(k-1)$	The *TTC* between merging vehicle *n* and through lead vehicle *k*-1
$C(k-1)$	The type of through lead vehicle *k*-1 (*1* for car; *2* for heavy vehicle)
$D(n)$	The merging time already elapsed (i.e., the time elapsed of the merging vehicle *n* since a merging action starts)
$Y(n)$	The decision of the merging vehicle *n*

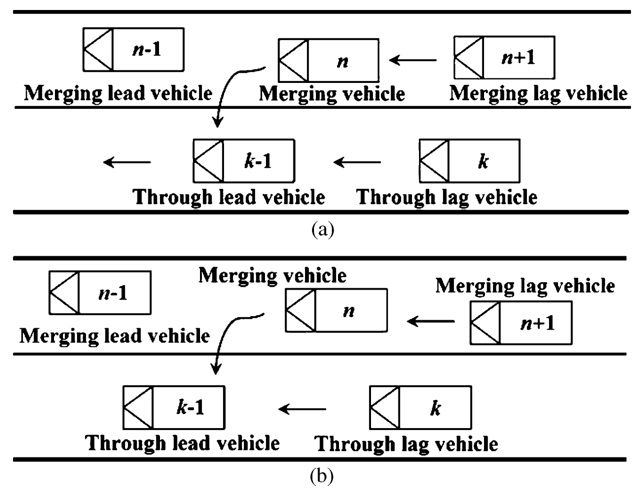

Fig. 4. Two cases that the merging vehicle cannot complete merging within next time interval: (a) $X(n) - X(k-1) - L(k-1) < 0$; (b) $X(k) - X(n) - L(n) < 0$.

Similarly, we can calculate the $TTCs$ between the merging vehicle n and its neighbor vehicles in the through lane (i.e., $TTC(k)$ and $TTC(k-1)$). The explanatory variables used for modeling the merging behavior are shown in Table I. Note that the TTC can be calculated only for the condition that the lag vehicle is faster than the lead one, hence an infinite large value (=99 in this study) will be assigned to the TTC when the speed of the lag vehicle is less than or equal to the corresponding lead vehicle speed. Fig. 4(a) shows a case of $X(n) - X(k-1) - L(k-1) < 0$ while Fig. 4(b) shows another case where $X(k) - X(n) - L(n) < 0$. It is impossible for the merging vehicle n to complete the merging maneuver within next time interval under both situations because an accident must occur if the merging vehicle persists in entering the through lane. Hence, the corresponding TTCs are equal to zero for the two cases (i.e., $TTC(k) = 0$, $TTC(k-1) = 0$), suggesting that the merging vehicle has a probability of 100% being collided with the through lead and/or lag vehicles if it takes the choice of "complete merging."

IV. METHODOLOGY

The majority of driving behavior models have been developed using parametric statistical approaches. However, some researchers also built non-parametric models which can provide higher prediction accuracy for merging behavior (e.g., [42] and [43]). More importantly, these non-parametric models are more suitable to extract merging decision rules. Therefore, as a typical non-parametric method, the CART method [44] is applied in this study. The CART identifies mutually exclusive and exhaustive subgroups of a population whose members share common characteristics. Because of low requirement for underlying data and high accuracy for prediction, the CART approach has been widely discussed and applied. In general, the CART can be classified into two categories depending on the target variable. One is called the regression tree when the target variable is continuous and the other is a classification tree for a categorical target variable. In this study, the CART is actually a classification tree because of the dichotomous target variable (i.e., "complete merging" or "continue merging").

Fig. 5 shows an example of CART. The Node 1, which is called the root node, contains all records. The CART approach employs a recursive partitioning technique to select variables and best splits by means of a splitting criterion, which measures the quality of each possible split. According to the selected Independent Variable 1, the root node branches into two descendent (child) nodes (i.e., Nodes 2 and 3).

It should be pointed out that several splitting criteria are available for determining the splits, including Gini-index, Chi-square test, Entropy reduction and so on. Comparing with

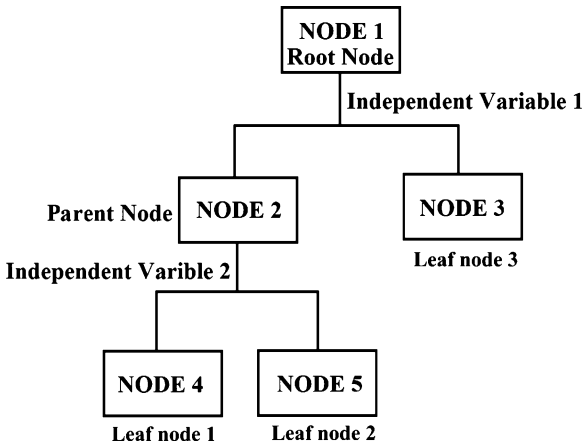

Fig. 5. The example of a CART.

Gini-index and Chi-square test splitting criteria, it will take a relatively longer time for the entropy to reaches its peak values in the Entropy reduction criterion. In other words, the Entropy reduction would be a stricter splitting criterion for a complex and chaotic set. Considering the unbalanced data structure as well as the combination of binary and continuous variables, the Entropy reduction is selected as the splitting criterion to ensure a more accurate split in this study. Each child node continues to be split if there is still a variable that can split the node best according to the chosen splitting criterion. Node 2, as a parent node, can be split into another two child nodes (Nodes 4 and 5) based on the Independent Variable 2. The splitting procedure will stop when one of the following rules is satisfied: (1) the number of records in the node is less than the preset minimum number; (2) the node impurity is less than the preset minimum value of the splitting criterion; (3) the current tree depth has reached the preset maximum tree depth. If the node cannot be split, it is called a leaf node. In Fig. 5, Nodes 3, 4, and 5 are considered as leaf nodes.

Note that an over-fitted tree which could result in a biased or inaccurate prediction for new datasets may be produced if the CART contains a large number of leaf nodes. Therefore, we have to remove some redundant branches to ensure the accuracy of the model. There are two different tree pruning techniques including the pre-pruning and post-pruning approaches. The cost complexity post-pruning approach is adopted in this study. The general principle of this approach can be briefly described as follows. Denote the error rate of tree B for the sample S as $err(B, S)$. Here, the error rate is defined as the percentage of vehicle maneuvers misclassified by the tree. Let the function $prune(B, b)$ represents the pruned tree after removing the subtree b from the tree B. It should be pointed out that we only select the subtree b which can minimize $(err(prune(B, b), S) - err(B, S))/(|leaves(B)| - |leaves \times (prune(B, b))|)$ for new datasets (i.e., validation data).

Although the tree can be easily understood by humans, it may not be easily recognized by computers. Hence, the rules need to be extracted from the pruned CART so that they can be incorporated into the merging assistance system. One rule can be created for each path from the root node to a leaf node in the form of an "if—then" rule. All variable values along a path form a conjunction in the rule antecedent (the "if" part), whereas the leaf node determines the predicted class, forming the rule consequence (the "then" part).

V. Results and Discussions

Considering that the records that a merging vehicle enter into the through lane ($Y(n) = 1$) account for a small proportion (17.89%). In order to eliminate biased results caused by unbalanced data, a data pre-processing should be performed before building a CART. In general, the under-sampling for reducing majority data [45], [46] and over-sampling for expanding minority data [47], [48] are two classic methods to deal with the issue of unbalanced data. Moreover, Gustavo [49] proposed 10 improved methods based on these two basic balance methods. Compared with the under-sampling method, the over-sampling method may increase the probability of over-fitting problem. In this study, we applied the random under-sampling method to reduce non-merging records by one-quarter (i.e., 320 non-merging records). Additionally, the parameters including the minimum number of observations in a node, the maximum tree depth are also preset. In order to ensure an accurate classification, the minimum number of records in a leaf node is set to be five and the number of records required for a split is taken to be 10. The maximum tree depth is assigned by 10. In addition, the total records are divided into two parts randomly, including 80% of the collected records as training data to produce a CART and the remaining 20% of the collected records as validation data for pruning the produced CART. That is, 483 records are used to build the CART and another 121 records are used for validating the built CART.

The misclassification rate for the training data is 0.1139 comparing to 0.1074 for the validation data. The final CART containing 17 leaf nodes is depicted in Fig. 6. From the figure, we can find that the time elapsed since a merging action being triggered ($D(n)$) is the most significant factor influencing the vehicle merging behavior during the merging implementation period. More specifically, $D(n)$ is the first variable selected to split the root node into two descendent nodes. The records with a value of $D(n)$ lower than 2.5 seconds are gathered in the left child node, forming Node 2 and the right child node (Node 5) contains the remaining records. According to the splitting criterion, the variable $TTC(n-1)$ is further selected to split Node 2. The records with a $TTC(n-1)$ smaller than 5.51 seconds are directed to the left, while the other records with a $TTC(n-1)$ larger than 5.51 seconds are directed to the right. It should be pointed out that the right node is a leaf node named Leaf Node 1. In Leaf Node 1, we can easily observe that there are 123 records in total, including 122 "zero" and 1 "one" ("complete merging"). This indicates that 99.2% (=122/123) of the merging vehicles prefer to continue their merging maneuvers, when the merging time elapsed are smaller than 2.5 seconds and $TTCs$ associated with the merging lead vehicle are larger than 5.51 seconds. This suggests that the merging vehicles have enough patience when the vehicle crash risk between the merging vehicle and lead vehicle is small at the initial stage of merging. Hence, they are not anxious to enter the through lane as quickly as possible under this situation.

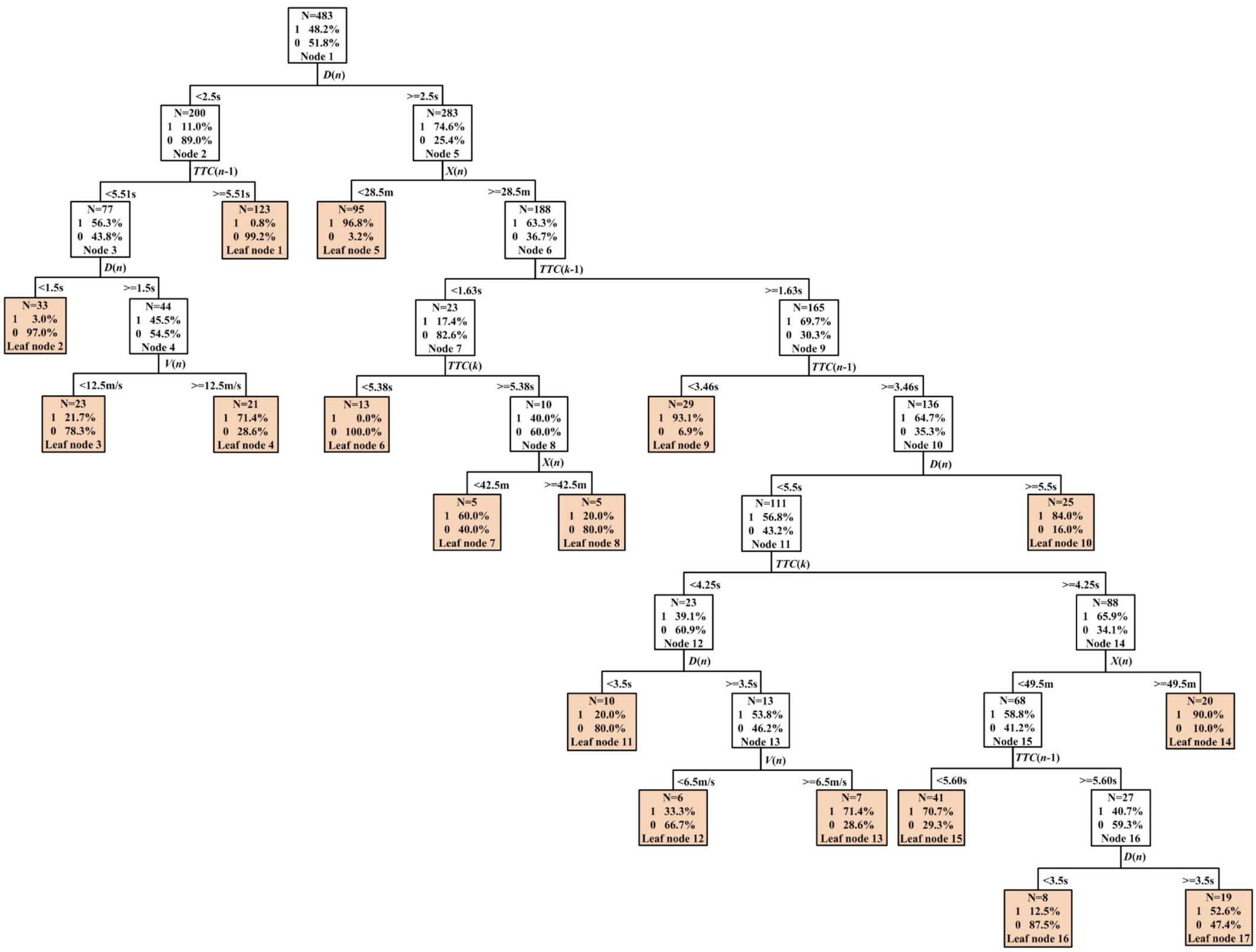

Fig. 6. The structure of the CART model.

The CART further splits Node 3 into Leaf Node 2 and Node 4 based on the variable of $D(n)$ again. The rule extracted from Leaf Node 2 is that the majority of merging vehicles choose to "continue merging" if the values of $D(n)$ are less than 1.5 seconds, even $TTC(n-1)$ are less than 5.51 seconds. The tree keeps splitting in the left until Leaf node 3 and 4 appear. The records with $V(n) < 12.5$ m/s are gathered in Leaf node 3 and the left are classified in Leaf Node 4. The rules reflected in these two leaf nodes means that most drivers never spend much time before complete their merging in a higher speed.

The root node completes its split in the left and produces four leaf nodes (i.e., Leaf Nodes 1–4). Likewise, the root node can be further split in the right. As shown on the right-hand side of the tree in the Fig. 6. The structure of right subtrees is much more complex and there are 13 leaf nodes in the right. The variable $X(n)$ plays an important role in splitting Node 5, the records in this node have two paths to split. In the left is Leaf Node 5, whose $X(n)$ is shorter than 28.5 meters. The remaining records are classified into Node 6, which is further split based on the variables $TTC(k-1)$, $TTC(k)$, $TTC(n-1)$, $X(n)$, $V(n)$ and $D(n)$ into Leaf Nodes 6–17. For example, three variables $TTC(k-1)$, $TTC(k)$ and $X(n)$ accomplish the left split of Node 6. More specifically, Leaf Node 6 is split by the variable $TTC(k)$ and Leaf Node 7 and 8 are determined by $X(n)$. The merging vehicles in Leaf Node 6 choose to continue their merging actions within next time interval when the $TTCs$ associated with the through lead vehicle and through lag vehicle are less than 1.63 seconds and 5.38 seconds, respectively. It shows that smaller $TTCs$ between the merging vehicle and its neighboring vehicles in the through lane may decrease the likelihood of entering the through lane earlier. In addition, the remaining distance of the merging vehicle to the work zone can also influence the vehicle merging behavior. From Leaf Nodes 7 and 8, the merging vehicles perform a stronger willingness to complete merging maneuver when they are closer to the work zone. No variables associated with the vehicle type are selected for the CART in this study. One possible reason for this may be that all vehicles in the work zone merging area will slow down and the drivers may obey traffic rules more cautiously. In this situation, the differences between cars and heavy vehicles almost disappear.

Finally, a set of merging decision rules which can be identified by computers are extracted from 17 leaf nodes of the CART. These rules can be incorporated into the work zone vehicle merging assistance system to ensure safer merges. The detailed rules are shown in Table II. It can be seen that nine

TABLE II
MERGING DECISION RULES

No.	Antecedent	Consequent	Accuracy
1	$D(n) < 2.5s$; $TTC(n-1) \geq 5.51s$	Continue merging	99.2%
2	$D(n) < 1.5s$; $TTC(n-1) < 5.51s$	Continue merging	97.0%
3	$1.5s \leq D(n) < 2.5s$; $TTC(n-1) < 5.51s$; $V(n) < 12.5m/s$	Continue merging	78.3%
4	$1.5s \leq D(n) < 2.5s$; $TTC(n-1) < 5.51s$; $V(n) \geq 12.5m/s$	Complete merging	71.4%
5	$D(n) \geq 2.5s$; $X(n) < 28.5m$	Complete merging	96.8%
6	$D(n) \geq 2.5s$; $X(n) \geq 28.5m$; $TTC(k-1) < 1.63s$; $TTC(k) < 5.38s$	Continue merging	100.0%
7	$D(n) \geq 2.5s$; $28.5m \leq X(n) < 42.5m$; $TTC(k-1) < 1.63s$; $TTC(k) \geq 5.38s$	Complete merging	60.0%
8	$D(n) \geq 2.5s$; $X(n) \geq 42.5m$; $TTC(k-1) < 1.63s$; $TTC(k) \geq 5.38s$	Continue merging	80.0%
9	$D(n) \geq 2.5s$; $X(n) \geq 28.5m$; $TTC(n-1) < 3.46s$; $TTC(k-1) \geq 1.63s$	Complete merging	93.1%
10	$D(n) \geq 5.5s$; $X(n) \geq 28.5m$; $TTC(n-1) \geq 3.46s$; $TTC(k-1) \geq 1.63s$	Complete merging	84.0%
11	$2.5s \leq D(n) < 3.5s$; $X(n) \geq 28.5m$; $TTC(n-1) \geq 3.46s$; $TTC(k-1) \geq 1.63s$; $TTC(k) < 4.25s$	Continue merging	80.0%
12	$3.5s \leq D(n) < 5.5s$; $X(n) \geq 28.5m$; $TTC(n-1) \geq 3.46s$; $TTC(k-1) \geq 1.63s$; $TTC(k) < 4.25s$; $V(n) < 6.5m/s$	Continue merging	66.7%
13	$3.5s \leq D(n) < 5.5s$; $X(n) \geq 28.5m$; $TTC(n-1) \geq 3.46s$; $TTC(k-1) \geq 1.63s$; $TTC(k) < 4.25s$; $V(n) \geq 6.5m/s$	Complete merging	71.4%
14	$2.5s \leq D(n) < 5.5s$; $X(n) \geq 49.5m$; $TTC(n-1) \geq 3.46s$; $TTC(k-1) \geq 1.63s$; $TTC(k) \geq 4.25s$	Complete merging	90.0%
15	$2.5s \leq D(n) < 5.5s$; $28.5m \leq X(n) < 49.5m$; $3.46s \leq TTC(n-1) < 5.6s$; $TTC(k-1) \geq 1.63s$; $TTC(k) \geq 4.25s$	Complete merging	70.7%
16	$2.5s \leq D(n) < 3.5s$; $28.5m \leq X(n) < 49.5m$; $TTC(n-1) \geq 5.6s$; $TTC(k-1) \geq 1.63s$; $TTC(k) \geq 4.25s$	Continue merging	87.5%
17	$3.5s \leq D(n) < 5.5s$; $28.5m \leq X(n) < 49.5m$; $TTC(n-1) \geq 5.6s$; $TTC(k-1) \geq 1.63s$; $TTC(k) \geq 4.25s$	Complete merging	52.6%

rules refer to "complete merging" and eight rules refer to "continue merging." The CART performs better in predicting "continue merging" decisions than in predicting "complete merging" decisions (94.1% versus 84.0%). For a vehicle merging decision assistance system, the consequence of misleading a "continue merging" advice as a "complete merging" advice is much severer than misleading a "complete merging" as a "continue merging." The former may lead to a disastrous accident, whereas the latter only results in a longer time delay. Note that the developed CART is based on time scan because it can predict merging behaviors for all merging vehicles every second.

For the purpose of model comparison, we develop another CART model with no use of $TTCs$ in this study. In order to distinguish the two CART models, the first model with the use of $TTCs$ is named CART_I and the second one is called CART_II. The parameters and data for the two CARTs are the same except for the use of variables. The variables used for building CART_II are slightly different from those used to build CART_I. More specifically, the TTC related variables are removed while six new variables are added for the CART_II, including $X(n-1)$, $V(n-1)$, $X(k)$, $V(k)$, $X(k-1)$ and $V(k-1)$. These newly added variables represent the relative distances and speeds between the merging vehicle and merging lead vehicle, through lag vehicle and through lead vehicle, respectively.

Results show that the structure of CART_II is significantly different from CART_I. Comparing with 17 leaf nodes in the CART_I, there are only four leaf nodes in the CART_II. Table III shows the selected variables and their importance in the two CART models. In statistical analysis, variable importance is a relative measure of how much each of variables used in the model contributes to reducing the prediction error of the model. In general, the high importance of a variable indicates that this variable could make a big contribution in reducing the prediction error. Note that the variable importance depends on the priority of variables, which can further influence model results. For example, it can be seen from Table III that the importance of the variable $D(n)$ is the highest (1.000) in both CART models. This suggests that the time elapsed since a merging action being triggered is the most important factor in improving the prediction accuracy of merging behaviors. If the variable $D(n)$ is not selected first to split the root node, it will cause different CART model results and the importance of this variable will be lower than 1.0.

TABLE III
IMPORTANCE OF INPUT VARIABLES

CART_I		CART_II	
Variable	Importance	Variable	Importance
$D(n)$	1.000	$D(n)$	1.000
$X(n)$	0.410	$X(k-1)$	0.486
$TTC(n-1)$	0.354	$X(n)$	0.272
$TTC(k-1)$	0.322		
$V(n)$	0.244		
$TTC(k)$	0.204		

Table III shows that there are only three variables (i.e., $D(n)$, $X(k-1)$ and $X(n)$) selected for the CART_II. It should be pointed out that the CART_II could not reflect the merging safety absolutely. For instance, the CART_II contains a rule that drivers should complete merging as $D(n) < 2.5$ s and $X(k-1) \geq 6.5$ m. If the merging vehicle adopts this rule but travels at a speed higher than the through lead vehicle, there will be a high probability of rear-end crash between the merging and through lead vehicles. In addition, the CART_II has lower overall prediction accuracy than the CART_I (85.7% versus 88.6%).

Nevertheless, it should be pointed out that there still exist several limitations in this study. First, the developed CART does not take into account the effects of driver characteristics (e.g., age, gender and driving experience) on the merging behavior. Second, the merging behavior may be also affected by different lane closure location while only the fast lane closure is considered in this study.

VI. Conclusion

This study modeled the vehicle merging behavior from the time of starting a merging maneuver to that of completing this maneuver using the CART approach. The vehicle trajectory data used to train and validate the CART were collected from a work zone site located in Singapore. As the target variable, the vehicle merging decision during the merging implementation period in the CART can be explained by influencing factors including the merging vehicle speed, vehicle types and TTCs between the merging vehicle and its neighboring vehicles, the remaining distance to the work zone and merging time elapsed. A CART with 17 leaf nodes was finally produced to describe the relationship between the merging decision and influencing factors.

The produced CART is able to provide acceptable prediction accuracy on both training and validation data, especially in predicting the "continue merging" decisions. Six factors including the merging time elapsed, the remaining distance to the work zone, merging vehicle speed and the TTCs between the merging vehicle and neighboring vehicles were the important factors influencing vehicle merging behavior during the merging implementation period. More specifically, a larger merging time elapsed, a shorter remaining distance to work zone, a smaller TTC to the merging lead vehicle and a higher speed of merging vehicle may encourage vehicle to make "complete merging" decision earlier. On the other hand, if TTCs between the merging vehicle and neighboring vehicles in the through lane are too small, the merging vehicle would like to continue its merging maneuver. For the comparison purpose, we developed another CART model using space gaps and neighboring vehicle speeds instead of $TTCs$. The comparison results show that our CART model with the use of TTCs can provide higher prediction accuracy.

In order to make full use of our results, we extract the information of the CART into 17 merging decision rules, which can be further incorporated into the work zone merging assistance system to guide a safety merging. Note that the results cannot transfer to car following rules, because all data in this study come from vehicles which are staying merging state. These model results reflect the lateral movement behavior rather than the longitudinal car-following behavior. In addition, it should be pointed out that our model results are only applicable for the work zone site where the fast lane is closed because the merging behavior may vary with different lane closure configurations. Due to data limitations, we did not consider other influencing factors (e.g., driver characteristics, traffic density and cooperative behavior of lag drivers) that might also affect vehicle merging decisions in this study. Therefore, future studies will take into account these factors in order to develop a more perfect model to explain the vehicle merging behavior during the merging implementation period.

References

[1] F. Hu, X. Dan, and S. Shaowu, "On the application of the Internet of things in the field of medical and health care," in *Proc. IEEE GreenCom iThings/CPSCom*, 2013, pp. 2053–2058.
[2] C. Louis and J. Eksteen, "The internet of things—Promise for the future? An introduction," in *Proc. IEEE Conf. IST-Africa*, 2011, pp. 1–9.
[3] P. Fan and Z. Guang-zhao, "Analysis of the business model innovation of the technology of internet of things in postal logistics," in *Proc. IEEE 18th Int. Conf. IE EM*, 2011, pp. 532–536.
[4] L. Li, "Application of the internet of thing in green agricultural products supply chain management," in *Proc. IEEE ICICTA*, 2011, vol. 1, pp. 1022–1025.
[5] Y. Yu, J. Wang, and G. Zhou, "The exploration in the education of professionals in applied internet of things engineering," in *Proc. IEEE 4th ICDLE*, 2010, pp. 74–77.
[6] T. Zhang, "The internet of things promoting higher education revolution," in *Proc. IEEE 4th MINES* , 2012, pp. 790–793.
[7] N. Maslekar, M. Boussedjra, J. Mouzna, and H. Labiod, "VANET based adaptive traffic signal control," in *Proc. 73rd VTC—Spring*, 2011, pp. 1–5.
[8] Q. Yu and D. Liu, "Disseminate warning message in VANETs based on predicting the interval of vehicles," in *Proc. IEEE 5th FCST*, 2010, pp.559–564.
[9] K. Li, T. Chen, Y. Luo, and J. Wang, "Intelligent environment-friendly vehicles: Concept and case studies. *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 1, pp. 318–328, Mar. 2012.
[10] M. Killat, G. Tristan, and H. Hannes, "Enabling traffic safety assessment of VANETs by means of accident simulations," in *Proc. IEEE 19th Int. Symp. PIMRC*, 2008, pp. 1–6.
[11] J. W. Hall and V. M. Lorenz, "Characteristics of construction zone accidents," *Transp. Res. Rec.*, vol. 1230, pp. 20–27, 1989.
[12] N. J. Garber and T. H. Woo, "Accident characteristics at construction and maintenance zones in urban areas," Virginia Transp. Council, Charlottesville, VA, USA, Rep. VTRC 90-R12, 1990.
[13] M. Zhao, "Crash characteristics at work zones," Dept. Transp., Washington, DC, USA, Rep. UVACTS-15-0-48, 2001.
[14] N. J. Garber and M. Zhao, "Crash characteristics at work zones," Virginia Transp. Council, Charlottesville, VA, USA, Rep. VTRC 02-R12, 2002.
[15] S. Srinivasan, G. Carrick, K. Heaslip, and S. Washburn, "Analysis of crashes in freeway work-zone queues: A case study," Southeastern Transp. Center, Knoxville, TN, USA, Rep. 07-UF-R-S3, 2007.
[16] H. Kita, "A merging-giveway interaction model of cars in a merging section: A game theoretic analysis," *Transp. Res. A, Policy Pract.*, vol. 33, no. 3/4, pp. 305–312, Apr. 1999.
[17] R. Herman and G. H. Weiss, "Comments on the highway crossing problem," *Oper. Res.*, vol. 9, no. 6, pp. 828–840, Nov./Dec. 1961.
[18] D. R. Drew, L. R. LaMotte, J. H. Buhr, and J. A. Wattleworth, "Gap acceptance in the freeway merging process," Texas Transp. Inst., College Station, TX, USA, 1967.
[19] A. J. Miller, "Nine estimators of gap acceptance parameters," in *Proc. 5th Int. Symp. Theory Traffic Flow Transp.*, 1972, pp. 215–235.
[20] Q. Yang and H. N. Koutsopoulos, "A microscopic traffic simulator for evaluation of dynamic traffic management systems," *Transp. Res. C, Emerging Technol.*, vol. 4, no. 3, pp. 113–129, Jun. 1996.
[21] P. Hidas, "Modeling vehicle interactions in microscopic simulation of merging and weaving," *Transp. Res. C, Emerging Technol.*, vol. 13, no. 1, pp. 37–62, Feb. 2005.
[22] K. I. Ahmed, M. Ben-Akiva, H. N. Koutsopoulos, and R. G. Mishalani, "Models of freeway lane changing and gap acceptance behavior," in *Proc. 13th Int. Symp. Theory Traffic Flow Transp.*, 1996, pp. 501–515.
[23] G. Lee, "Modeling gap acceptance at freeway merges," M.S. thesis, Dept. Civil Environ. Eng., Cambridge, MA, USA, 2006
[24] T. Toledo, "Driving behaviors: Models and research directions," *Transp. Rev.*, vol. 27, pp. 65–84, 2007.
[25] T. Toledo, H. N. Koutsopoulos, and M. Ben-Akiva, "Estimation of an integrated driving behavior model," *Transp. Res. C, Emerging Technol.*, vol. 17, no. 4, pp. 365–380, Aug. 2009.

[26] J. Weng and Q. Meng, "Modeling speed-flow relationship and merging behavior at work zone merging areas," *Transp. Res. C, Emerging Technol.*, vol. 19, no. 6, pp. 985–996, Dec. 2011.
[27] H. Jia, Y. Tan, and L. Yang, "Modeling vehicle merging behavior in urban expressway merging sections based on logistic model," in *Proc. IEEE Int. Conf. TMEE*, 2011, pp. 656–659.
[28] T. Fatema and Y. Hassan, "Probabilistic design of freeway entrance speed change lanes considering acceleration and gap acceptance behavior," in *Proc. Transp. Res. Board Annu. Meet.*, Washington, DC, USA, 2012, pp. 1–16.
[29] K. I. Ahmed, "Modeling drivers' acceleration and lane changing behavior," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, 1999.
[30] N. Oliver and A. P. Pentland, "Graphical models for driver behavior recognition in a SmartCar," in *Proc. IEEE Intell. Veh. Symp.*, 2000, pp. 7–12.
[31] D. L. Smith, R. Glassco, J. Chang, and D. Cohen, "Feasibility of modeling lane-change performance," Nat. Highway Traffic Safety Admin., Washington, DC, USA, Tech. Rep. 2003-01-0280, 2003.
[32] J. C. McCall, D. P. Wipf, M. M. Trivedi, and B. D. Rao, "Lane change intent analysis using robust operators and sparse Bayesian learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 3, pp. 431–440, Sep. 2007.
[33] Q. Meng and J. Weng, "An improved cellular automata model for heterogeneous work zone traffic," *Transp. Res. C, Emerging Technol.*, vol. 19, no. 6, pp. 1263–1275, Oct. 2011.
[34] C. Oh, S. Park, and S. G., Ritchie, "A method for identifying rear-end collision risks using inductive loop detectors," *Accident Anal. Prevention*, vol. 38, no. 2, pp. 295–301, Mar. 2006.
[35] C. Oh and T. Kim, "Estimation of rear-end crash potential using vehicle trajectory data," *Accident Anal. Prevention*, vol. 42, no. 6, pp. 1888–1893, Nov. 2010.
[36] D. Gettman and L. Head, "Surrogate safety measures from traffic simulation models," US Dept. Transp., Fed. Highway Admin., Washington, DC, USA, Rep. FHWA-RD-03-050, 2003.
[37] F. Cunto and F. Saccomanno, "Calibration and validation of simulated vehicle safety performance at signalized intersections," *Accident Anal. Prevention*, vol. 40, no. 3, pp. 1171–1179, May. 2008.
[38] J. Weng and Q. Meng, "Rear-end crash potential estimation in the work zone merging area," *J. Transp. Eng.*, vol. 48, no. 3, pp. 238–249, Apr. 2014.
[39] L. Gao, X. Meng, and H. Xu, "Freeway work zone rear-end conflict study based on two traffic conflict indicators," in *Proc. ICTE*, 2013, pp. 558–563.
[40] J. A. Bonneson and J. W. Fitts, "Traffic data collection using video-based systems," *Transp. Res. Rec.*, vol. 1477, pp. 31–40, 1995.
[41] C. Strong, S. Lowry, and P. McCarthy, "Collecting vehicle-speed data by using time-elapse video recording equipment," *Transp. Res. Rec.*, vol. 1855, pp. 97–104, 2004.
[42] Y. Hou, P. Edara, and C. Sun, "Modeling mandatory lane changing using Bayes classifier and decision trees," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 2, pp. 647–655, Apr. 2014.
[43] Q. Meng and J. Weng, "Classification and regression tree approach for predicting drivers' merging behaviour in short-term work zone merging areas," *J. Transp. Eng.*, vol. 138, no. 8, pp. 1062–1070, Aug. 2012.
[44] L. Breiman, R. Friedman, and A. Olshe, "Classification and regression trees," Wadsworth Brooks/Cole, Pacific Grove, CA, USA, 1984.
[45] P. E. Hart, "The condensed nearest neighbor rule," *IEEE Trans. Inform. Theory*, vol. 14, no. 3, pp. 515–516, May 1968.
[46] D. Skalak, "Prototype and feature selection by sampling and random mutation hill climbing algorithms," in *Proc. 11th Mach. Learn. Conf.*, 1994, pp. 293–301.
[47] E. DeRouin, J. Brown, H. Beck, L. Fausett, and M. Schneider, "Neural network training on unequally represented classes," in *Intelligent Engineering Systems Through Artificial Neural Networks*. New York, NY, USA: ASME, 1991, pp.135–145.
[48] D. Lewis and J. Catlett, "Heterogeneous uncertainty sampling for supervised learning," in *Proc. 11th Int. Conf. Mach. Learn.*, 1994, pp. 144–156.
[49] B. Gustavo EAPA, P. Ronaldo C, and M. M. Carolina, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD Explorations Newslett.*, vol. 6, no. 1, pp. 20–29, 2004.

Jinxian Weng received the Ph.D. degree in transportation engineering from National University of Singapore. He is currently a Professor with the College of Transport and Communications, Shanghai Maritime University, Shanghai, China. His research interests include traffic simulation, work zone operations, traffic safety modeling, and driver behavioral analysis. Dr. Weng is an Associate Editor of *Journal of Transportation Safety and Security* and an Advisory Board Member of *International Journal of Transportation*.

Shan Xue received the B.Eng. degree in transportation engineering from Beijing Jiaotong University, Beijing, China, where he is currently working toward the M.Eng. degree in traffic safety engineering. His research interests include driver behavior modeling and traffic safety analysis.

Xuedong Yan received the Ph.D degree in civil engineering from the University of Cental Florida, Orlando, FL, USA. He is currently a Professor with the School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China, where he is also the Executive Director of the MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology. His research interests include traffic safety, driving simulation, urban planning, and management.