

What contributes to driving behavior prediction at unsignalized intersections?



Shun Yang^a, Wenshuo Wang^{b,c}, Yuande Jiang^a, Jian Wu^a, Sumin Zhang^{a,*}, Weiwen Deng^{a,d}

^a Jilin University, Changchun 130022, China

^b Beijing Institute of Technology, Beijing 100081, China

^c University of Michigan, Ann Arbor, MI 48109, USA

^d Beihang University, Beijing 100083, China

ARTICLE INFO

Keywords:

Feature contribution
Unsignalized intersection
Random forest
Driver behavior prediction

ABSTRACT

Safely passing through unsignalized intersections (USI) in urban area is challenging for autonomous vehicles due to high uncertainties of surrounding engaged human-driven vehicles. In order to achieve this, various variables have been selected to estimate and predict the surrounding human driver's behavior. However, it is still not fully clear what variables mostly contribute to driving behavior prediction at USI. This paper investigates the contribution levels of 13 features of human driver's decision-making at USI using a random forest approach. Thirty skilled driver participants are tested in a real-time driving simulator where the traffic scenarios with merging vehicles were designed in different motion styles to mimic real traffic. The experiment results indicate that the relative distance and velocity between merging vehicles have a wider contribution range (i.e., -0.4 – 0.4) than the absolute velocity and distance features (i.e., -0.2 – 0.2) to predict driver behavior. The contribution also varies over the selected feature values and driving conditions. This contribution research gains insight in the influence of different variables on driver behavior prediction at USI, thereby assisting researchers in selecting representative features in self-driving applications.

1. Introduction

One of the biggest challenges for autonomous vehicles driving in urban environment is to pass through unsignalized intersections (USI) safely and efficiently. According to the National Highway Traffic Safety Administration (NHTSA) (Administration et al., 2016), approximately 47% of the total 11,275,000 crashes in the United States in 2015 occurred at intersections and related areas, laying an intractable topic that requires careful solutions (Elmitiny et al., 2010). Moreover, previous research have demonstrated that driver behavior at intersections could significantly affect the traffic intensity and capacity (Ge et al., 2008; Sharma et al., 2011; Xiao et al., 2010; Qi et al., 2017; Shirazi and Morris, 2017). Among all intersection problems, the unsignalized intersection is of great difficulty since few positive indications for drivers are available when entering the intersection (Nagalla et al., 2017; De Beaucorps et al., 2017; Liu et al., 2017; Isele and Cosgun, 2017). In addition, the uncertainty of surrounding drivers' behavior, as shown in Fig. 1, further increases the risk of conflict. In order to safely and efficiently navigate through an unsignalized intersection, it is necessary to understand vehicle dynamics and interpret the intent of other human drivers, and then behave predictably, allowing other drivers to

* Corresponding author.

E-mail address: suminzhang@163.com (S. Zhang).

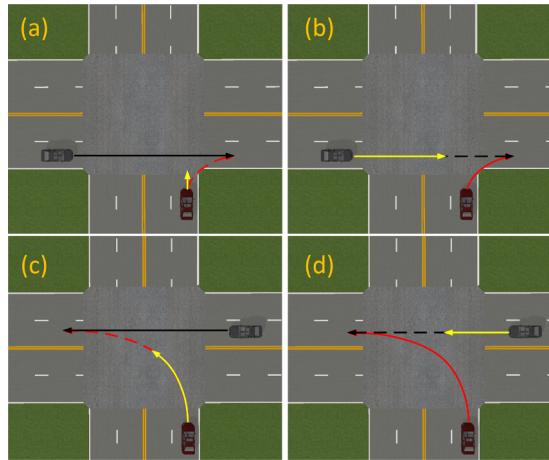


Fig. 1. Human drivers' possible behavior at the USI. The solid line and the dash line are the real and target trajectories of the cars, respectively. (a) The solid black line indicates that the merging vehicle (i.e., gray car) accelerates and passes the intersection while the ego car (i.e., red car) yields. (b) The dash black line shows the merging vehicle decelerates and the ego car exits directly. Two bottom pictures have similar implications, i.e., (c) ego car yielding and (d) ego car exiting directly. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

respond appropriately (Prasetijo et al., 2010; Kye et al., 2015; Hubmann et al., 2017).

Many researchers focused on modeling gap acceptance to describe crossing decisions at the USI (Hamed et al., 1997; Pollatschek et al., 2002; Spek et al., 2006; Guo and Lin, 2011; Devarasetty et al., 2011; Zhou et al., 2017). In order to describe gap acceptance behavior at USI, researchers selected various features, which can be roughly grouped into lane-related features (i.e., lane number, lane size, speed limit), gap-related features (i.e., gap duration, rejected and accepted gap number, mean and total time of rejected/accepted gaps), driver-related features (i.e., driver age, gender). The gap acceptance method has been applied and improved for several years, however it has a couple of drawbacks to explain the importance of every involved feature (Brilon and Wu, 2002).

Towards this end, data-driven approaches with advances of machine learning techniques are utilized to predict the driver behavior using the extracted features from raw data. Tran and Firl (2013, 2014) used Gaussian regression models to predict the maneuver of traffic participants at intersections. They selected the absolute position and velocity of the merging vehicles as the model input, without concerning the intention and interaction between vehicles. Song et al. (2016) develop a continuous hidden Markov model to predict both the high-level motion intention and the low level interaction intentions, they utilized the distance to the intersection, longitudinal velocity/acceleration and yaw rate as the inputs to predict drivers' lateral and longitudinal intentions at USI with a deterministic setting and finally obtained a satisfied prediction performance. However, human drivers usually behave stochastically in real traffic. Kye et al. (2015) defined three categories of variables (i.e., physical variables, behavioral variables, and the intention variable) to represent complicated prediction situations in a simple way, and their intention-aware decision-making algorithm has potential for other shapes of intersections. Liu et al. (2017) proposed a self-developed data analysis method to predict human driver's crossing behavior at USI using velocity, acceleration and distance variables. In addition, for sufficiently describing the dynamic interaction of driver-vehicle-environment systems, an intermediate variable of risk perception was also introduced in Liu et al. (2017). Zyner et al. (2018) treated lateral and longitudinal position with respect to other vehicles, heading angles, speed as input of recurrent neural networks and finally achieved 1.3 s prediction before potential conflict. In Wali et al. (2018) a full Bayesian estimation via Markov Chain Monte Carlo (MCMC) based Gibbs sampling is performed.

The above mentioned research offers certain knowledge for feature selection, thereby benefiting for the driver behavior prediction. However, it has been found that even on the same research topic, researchers use different variables as indicators to describe decision-making processes and predict driver behavior, as shown in Table 1. In order to figure out what are the most contributive variables and how much contribution they make towards driver behavior prediction, in this paper, we applied random forest (RF) approach to analyze driver's decision-making behaviors at the USI. Random forest is widely used in classification and regression tasks. It is faster to train, have fewer parameters to be tuned and can handle a large number of predictors without requiring any variable selection (Yan and Radwan, 2006), which makes it an ideal method for the driving behavior prediction. Further, there are researchers concluding that random forest can outperform SVM and decision trees methods in driving behavior prediction task (Nagalla et al., 2017), which indicates that RF model offers a more precise basis for variable contribution analysis. In this work, instead of using the metrics of feature importance to represent the general influences of features on the prediction result, we developed a method to gauge how much contributions of each feature in every sample offers for decision-making.

The remainders of this paper are organized as follows. Section 2 presents the random forest method for driver behavior prediction. Section 3 introduces the simulation platform, experiment design and the model training procedure. Section 4 discusses and analyzes the experiment results. Conclusions and future work are given in Section 5.

Table 1
Selected Features for Driver Behavior and Intention Prediction at Unsigned Intersections.

Year	Reference	Research topic	Selected features (variables)	Method
1997 2007	(Hamed et al., 1997) (Kaysi and Abbany, 2007)	Gap acceptance behavior Aggressive merging behavior	Velocity of mainstream, lane number, lane size, conflicting flow volume Age, gender, car performance, car model, number of rejected gaps, total waiting time at head of queue, major-traffic speed	Regression model Binary probit model
2011	(Devarasey et al., 2011)	Gap acceptance for left turning	Gap duration, total waiting time, time to turn, distance to next signal downstream	Binary probit model
2013 2013 2014	(Tran and Firj, 2013) (Rossi et al., 2013) (Liu et al., 2014)	Intention prediction Gap acceptance behavior Crossing behavior	Position and velocity of merging vehicles Type of time interval (lag or gap), interval size, vehicle type, vehicle waiting time Velocity of merging vehicles, distances between vehicles and merging points, velocity and merging distance difference, merging time of vehicles	Two dimensional gaussian process regression Transfer Scaling, Bayesian Updating and Combined Transfer Estimation Classification tree
2016 2017 2017 2018	(Song et al., 2016) (Liu et al., 2017) (Zhou et al., 2017) (Zyner et al., 2018)	Crossing behavior Crossing behavior Gap acceptance for left turning Intention prediction	Distance to intersection, lateral and longitudinal velocity, yaw rate Velocity, acceleration and distance of merging vehicles, risk perception factor Driver gender and age, lane number, speed limit, gap duration, rejected gap number, mean and total time interval of rejected gaps Relative position to merging vehicles, heading angle, velocity	Hidden Markov model A self-developed data analysis method Binary decision and correlated logit model Recurrent neural networks

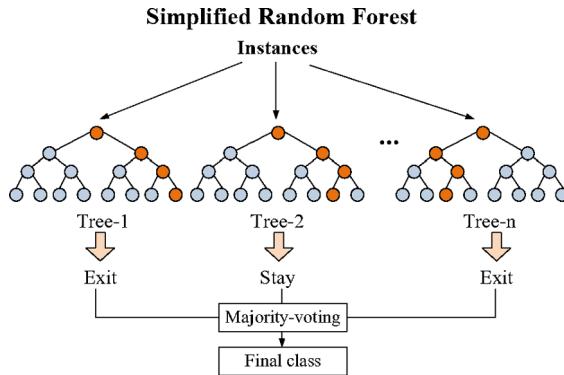


Fig. 2. Voting processes of the random forest.

2. Methodology

In this paper, we selected the RF method to model and predict drivers' decision-making behavior at USI under different merging scenarios. We first introduce the basic theory of RF and then illustrate a method to calculate the feature contribution based on the RF algorithm.

2.1. Random forest

RF is a decision tree-based approach that is popular in diversified research fields, it combines Breiman's bagging idea (Breiman, 1996) and Ho's random subspace methods (Ho, 1995) to construct a collection of decision trees with controlled variations (Washington et al., 2010). In the research field of traffic safety, the decision tree has been widely used, such as classifying driver's stop/go decision and red-light running violation (Elmitiny et al., 2010), analyzing driver injury severity in truck-involved accidents (Chang and Chien, 2013), identifying key factors of transit service quality (de Oña et al., 2012). On the other hand, RF is also being widely used to analyze the gap acceptance behavior (Nagalla et al., 2017), imitate driving skills of human being at free roads (Cichosz and Pawełczak, 2014), and recognize driving events (Cao et al., 2017).

The basic idea of RF is to build a multitude of weak decision tree classifiers in parallel and then combine them to form a single, strong learner by averaging/voting their individual predictions. Fig. 2 shows the voting process of RF. In application, a training set for growing trees is randomly selected from a sample set, the remaining samples, termed an out-of-bag (OOB) set, are used to estimate random forest's goodness-of-fit. Then, trees are grown to the maximum extent possible without pruning. The procedure of tree-learning algorithms is as follows:

- First, at each node, the best split is selected by examining a random subset of features rather than the complete feature set.
- Then, split each node and go down to extend its children nodes till terminal nodes reached.
- Once all the trees are constructed through the above steps, final predictions are weighted and summed up from individual predictions of the trees.

For a classification task, the classified result from each tree will vote for the final result. The random forest usually outperforms the decision tree since the over-fitting case can be effectively avoided by tuning its hyper-parameters such as tree depth, number of trees, number of features, etc.

2.2. Random forest interpretation

In this paper, the relationship between decision-making prediction model and selected features is explained using the feature contribution concept (Kuz'min et al., 2011; Palczewska et al., 2013). The feature contribution is computed based on two aspects:

1. Only one feature is used for splitting on each node level of a decision tree;
2. The difference between the classification/regression values in the parent node and the child node can be simply calculated and considered as change caused by this feature.

For each decision tree in the random forest, it has one root node (i.e., Node 0 in Fig. 3). At the root node of one tree, the mean probability value of each class represents the base level of this tree. These values are generally calculated from the training set obtained by bootstrapping. Each node has its children nodes and parent node, for example, Node 1 has children nodes of Node 3 and 4, its parent node is Node 0. The child node level is always reached by meeting one splitting rule, the rule is only relevant to one feature in data samples. If we summarize the probability increments caused by the feature, we can obtain the contribution of it to the investigated classification value. In another work, the feature contribution in our work is calculated from the proportion of historical

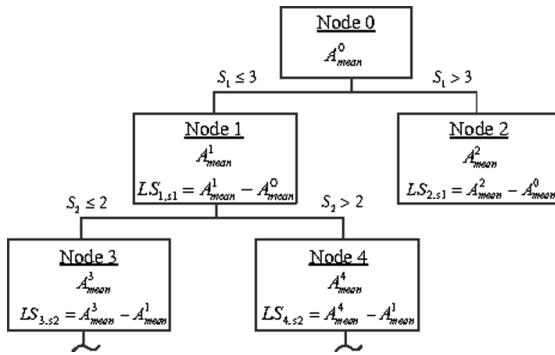


Fig. 3. Local increment calculation of feature contributions in one decision tree. (A_{mean} : mean value of training set samples in the node, S_i : features, LS_{Si} : local increment of classification value caused by S_i feature).

training samples that meet different splitting rule, which represents a priori probability. The judgement from one feature can lead one sample to fall into some children node and change its priori probability belonging to possible categories, we take this priori probability change as contribution, which means it can contribute to our judgement/prediction of which category this sample should be. As to the contribution value itself, it is the increment or decline of the probability that we predict some sample to in a specific category. In the random forest model, the probability increments caused by one feature over all trees are calculated and averaged to get its contribution to prediction.

2.2.1. All training sets run down each tree

In each node of the tree, the mean value (A_{mean}) is calculated for all samples in training sets by $A_{mean} = \frac{1}{m} \sum_{k=1}^m A_k$ (1) where A_{mean} is the mean classification values for each class in the tree node (as A_{mean}^0 for the root node in Fig. 3), m is the number of samples in this tree node, A_k is the classification value of the k -th sample for each class in this tree node.

2.2.2. Samples fall into each tree node using its splitting rule

The difference between averaged classification values in the parent and child nodes represents a local increment of contribution of the corresponding feature, which is included in the splitting rule of this node.

$LS_{i,j} = A_{mean}^C - A_{mean}^P$ (2) where $LS_{i,j}$ represents the contribution of the j -th feature at i -th node, as shown in Fig. 3, the contribution of feature S_1 calculated from Node 1 is $LS_{1,S_1} A_{mean}^C$ is the mean classification value of local training set samples included in current child node (i.e., Node 1); A_{mean}^P is the mean classification value of local training set samples included in the parent node (i.e., Node 0). Then the contribution of features can be calculated at each node of all trees in random forest model. Fig. 3 shows step 1 and step 2.

2.2.3. Local increments of contribution included in the rules of nodes

To determine the overall contribution of one feature for a certain sample, one should summarize local increments of contribution of this feature included in the rules of nodes, which contain the given samples. The final summation of local increments of contribution made by each feature is divided using the overall number of trees in the forest, i.e.,

$S_{k,j} = \frac{1}{T} \sum_{i=1}^n LS_{i,j}$ (3) where $S_{k,j}$ is the overall contribution of the j -th feature for the k -th sample, T is the overall number of trees in the RF model, n is the overall number of tree nodes in all trees of RF, which contain the k -th sample and have the j -th feature in their splitting rule, $LS_{i,j}$ is the local increment of contribution of the j -th feature of the i -th node.

3. Experiment and data collection

In this section, we collected the driving data from a high-fidelity and real-time simulation platform and analyzed the decision-making process without putting drivers in jeopardy.

3.1. Simulation platform

A high-fidelity and real-time driving simulation platform was used in this research for data collection. The platform consists of six main parts (Fig. 4): input devices, dSPACE Simulator 1006, PanoSim-RT software, Industrial Personal Computer (IPC), 6 degree-of-freedom motion platform with a 120 curvy screen, and human drivers. The input devices were used to collect operation signals of drivers such as the steering wheel angle, brake pedal displacement, and throttle opening. To improve the driver's sense of immersion, we chose Senso-Wheel to emulate the feedback steering torque where the torque value was calculated from models of vehicle dynamics and steering systems, and generated by a torque motor. The dSPACE Simulator receives the data from the input devices and solves the vehicle dynamic equations transferred from the IPC. The PanoSim-RT is an integrated driving simulation software that provides virtual driving scene, high-fidelity vehicle dynamics model, and other models like traffic, testing field weather/light conditions as well as various types of environmental sensors (e.g., radar, camera, antenna for wireless communication, GPS). The IPC

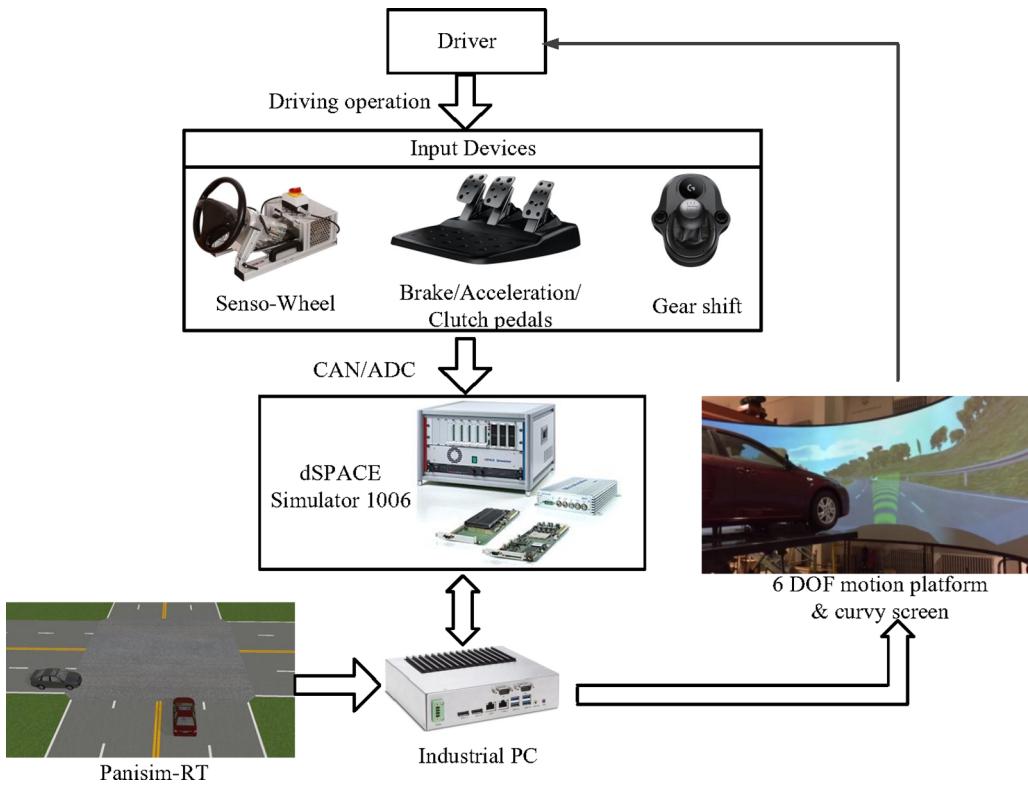


Fig. 4. High-fidelity driving simulation system.

receives scene and vehicle dynamic data from Panosim-RT and sends all these data to dSPACE simulator for calculating. All the result data from dSPACE simulator is then sent to 6 degree-of-freedom motion platform and curvy display screen to generate the realistic driving motions and scenarios. Human drivers then choose how they should react according to the motion they feel and the scenario they see.

During the testing process, parameters including the throttle opening, braking pressure, steering wheel angle, vehicle position, longitudinal and lateral speed, yaw rate, roll angle, and acceleration were calculated and recorded.

3.2. Unsignalized intersection scenario

To make the driving scenarios similar to what the driver participants would encounter in the real traffic, the layout of USI was designed in the Panosim-RT simulation platform and two representative scenarios were selected. A typical intersection was designed, where each leg of the intersection was set to be a two-lane two-way road, with each lane width of 3.7 m, the central line of the road was set to be solid double yellow lines, the central line of each lane was dashed white line and the edge lines were solid white lines. There no traffic light indicates right-of-way so that drivers need to choose action at the intersection according to their driving experience and driving style.

The intersection created in Panosim-RT is illustrated in Fig. 5. We set the right direction as the positive direction of x axis and the up direction as the positive direction of y axis. The red car represents the ego vehicle and the gray car represents the merging vehicle. The red marks on the two sub-figures are the merging points of two merging vehicles, representing the possible collision sites of the ego and merging vehicles if both keep going. In the left figure of Fig. 5, the ego vehicle is going to turn right and the merging vehicle is going straight. A fluctuation was also added to the longitudinal and lateral motion of merging vehicle, speed V_{y2} for the straight-going vehicle. In addition, since the direction definition is important for the variable statement in the following part and for the classification, we define the directions as follows. For two vehicles in the left sub-figure, both of their velocities along the x axis are positive, therefore, the ego vehicle has the positive velocity along the y axis and the merging vehicle has a negative V_{y2} . The definition for velocities in the right sub-figure is in the same way.

3.3. Data collection procedure

In order to obtain a representative data set, thirty drivers (15 males and 15 females), with more than 5 years of driving experience and ages of between 25 and 45, are chosen. The data samples are collected while they drive through the USI. Before the trials, each driver participant is asked to drive in the simulator for more than 30 min to get used to the devices and operation skills. During the

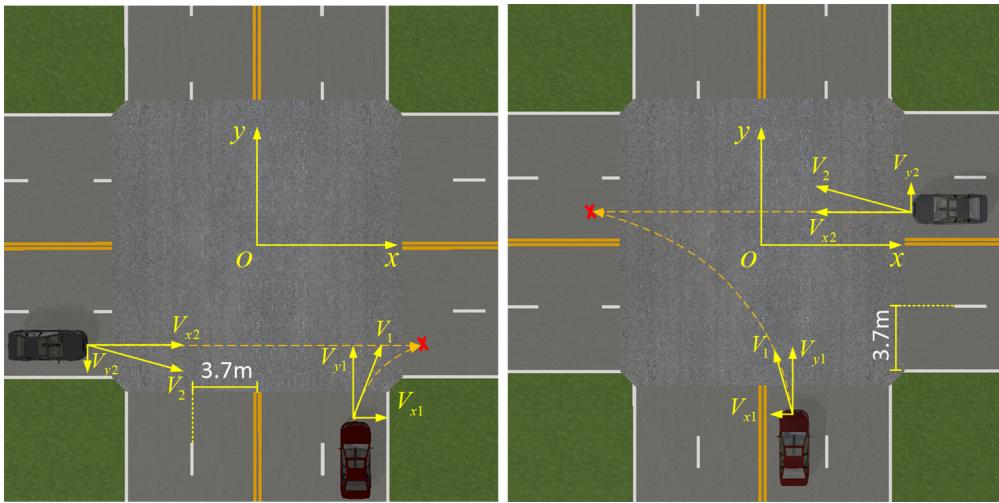


Fig. 5. Illustration of the designed USI, defaulting axis direction statement and two testing conditions. In the left figure, the ego vehicle (i.e., red car) is about to turn right and merging vehicle (gray car) going straight; in the right figure, the ego vehicle is about to turn left and merging vehicle going straight. The red marks in the figure are merging points. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

experiment, driver participants are told to turn left or right at the next coming unsignalized intersection, but they still need to choose the timing for “exit” or “yield” according to their driving experience and the dynamic driving environment since there is no guidance from traffic infrastructure. The participants are told to pay attention to the merging vehicle to avoid collision, and they are also told that the vehicle might not appear in the view if they are too far away so that participants don’t need to wait for them. If there is collision between the ego vehicle and merging vehicle, the data collection procedure is stopped and the data is discarded. For minimizing possible fatigue or habituation related constraints, driver participants are asked to take a 1-min break for every 5-min trial.

To avoid the interference of other vehicles during the decision-making process of the ego vehicle, we simplified the merging case where the ego vehicle encountered just one straight-going vehicle without other objects. Two test conditions were designed (Fig. 5):

- the ego vehicle turns right and merging vehicle is also going straight to right;
- the ego vehicle turns left and merging vehicle also going straight to left.

In order to introduce the behavior uncertainty to the merging vehicle, its longitudinal speed was set time-varying by adding random noise of $\pm 5\%$ of the setting speed. For the lateral movement, sinusoid with peak value 0.2 m and frequency 0.1 Hz was also added. The initial position of the ego vehicle was set to be 20 m away from the intersection. The initial positions of the merging vehicle were set to 50 m, 40 m, 30 m, 20 m, 10 m far away from the intersection. For each position, the initial longitudinal velocities of merging vehicle were set to 2 m/s, 4 m/s, 6 m/s and 8 m/s. The velocity greater than 8 m/s was beyond reasonable for a normal intersection, so we do not discuss these conditions in this work. The road in the experiments was set to be flat, the tire-road adhesion was in good condition (i.e., adhesion coefficient 0.8), and the weather was set to be sunny so guaranteeing sufficient illumination for driving.

In each testing condition, driver participants were asked to drive through the intersection for 20 times. In total, 24,000 trajectories from 30 drivers were collected (20 times \times 5 positions \times 4 velocities \times 30 drivers \times 2 test conditions). The whole data collection procedure took more than two weeks. Data set that consists of 21,709,520 data points was collected in more than 120 h with sampling frequency of 50 Hz.

3.4. Training and test procedure

3.4.1. Data preprocessing and labeling

We define the merging point for extracting useful data points. For each merging event, if either of two vehicles in the experiment reaches to the merging point, it means the merging behavior ends and no more decisions are needed, thus the rest of data points are discarded. Here, the merging point is a virtual point and cannot always be obtained exactly if the trajectories of two vehicles’ centroid do not intersect (as shown in Fig. 6), we define the two points (i.e., M_1 and M_2) in the vehicle trajectories that lead to the minimum distance as the merging points.

After the valid data points were extracted, we obtained 7,685,272 data points and labeled them. Since drivers were not asked for labeling the merging decision when they were under trials, we manually label the data off-line. In our work, for every merging process, we took out the last x coordinate of the ego and merging vehicles from the valid data, denoted as X_e and X_m . According to the

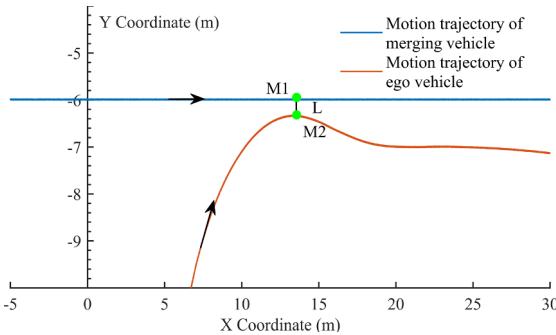


Fig. 6. Illustration of the merging point. The red and blue lines represent the trajectory of merging vehicle and ego vehicle, respectively. Two black arrows denote the motion direction. L is the shortest distance between two vehicle trajectories. M_1 and M_2 are chosen to be merging points for two vehicles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

absolute value of the coordinates and the lateral velocity direction of merging vehicle (V_{mx}), the data points of every merging sample was labeled by -1 for left turn, 0 for yield decision, and 1 for right turn, as shown in [Table 2](#).

3.4.2. Feature selection

These selected features were calculated by using the variables from ego and merging vehicles and then divided into four main categories including: (1) velocity and relative velocity between two vehicles, (2) relative distance between two vehicles, (3) relative distance between merging point and vehicles, (4) relative acceleration. [Table 3](#) lists 13 features and their definitions.

3.4.3. Model training

After the 13 features were figured out, training samples were generated by combining these features with their labels. Then all training samples were randomly divided into 2 parts, 70% of datasets were used for training and 30% of dataset were used for test. An ensemble of 50 decision trees were grown to form a RF classifier to predict the driving decision. For each decision tree, 5 features were randomly selected from each sample. The training procedure was performed on a laptop with 3.6 GHz Intel(R) Core(TM) i7 processor and 4 GB installed memory. The algorithm was conducted in Matlab R2017a. It takes 1416.17 s to train models on average.

3.5. Random forest applied on driving behavior prediction

In the driving behavior prediction application, the original data set includes all the collected samples during the driving process, where each sample is the combination of features, such as vehicle speed, relative distance between ego vehicle (the vehicle we mainly focus on) and nearest traffic participant, etc. selected for predicting the driving behavior. Since the random forest is the combination of many decision trees, for growing every one of them, subsets of the original data set (i.e., 5000 samples for the example in [Fig. 7](#)) are randomly selected with replacement. Since all the samples are well labelled in training data set, we can get exact proportion of three prediction categories, i.e., left turn, yield and right turn. Taking one sample out of the training data set, it has an initial probability to belong each of them (i.e., 32%, 37% or 31%). For the root node, a certain number of features are randomly selected to classify the 5000 samples, and the feature that can best split all the samples (i.e., the ego vehicle speed in this example) is chosen to establish the splitting rule for the root node. After splitting, the parent nodes will generate two children nodes, so all the nodes except the root node have their parent node and sibling nodes. These nodes may inherit samples that result in the positive judgement (i.e., 2586 samples as example) of the splitting rule from their parent node while their sibling nodes get all other samples (i.e., 2414 samples) that cannot meet the rule. In the children node, the proportion of three categories changes, which means if one sample meets the judgement condition of the splitting rule and falls into one children node, its probability belonging to one prediction category also changes. The nodes keep splitting in this way until the terminating depth is reached or all the samples in the current node belong to only one class.

Taking [Fig. 7](#) for example, after the training process, we can get the random forest with a decision tree like above. If we take one sample with ego vehicle speed 10 and relative distance to the left vehicle 20 from the real driving condition, it would end up in the left-most leaf with probabilities of 0.08, 0.17, and 0.75 for left turn, yield and right turn. All these probability values are the

Table 2
Conditions For Labeling.

Last x coordinates	V_{mx}	Label
$ X_e > X_m $	+	Right turn (1)
$ X_e < X_m $	+	Yield (0)
$ X_e > X_m $	-	Left turn (-1)
$ X_e < X_m $	-	Yield (0)

Table 3
Features for Training.

Name	Definition
V1	Speed of the ego vehicle
V2	Speed of the merging vehicle
R_Vx	Relative absolute x velocity of two vehicles
R_Vy	Relative absolute y velocity of two vehicles
R_V	Relative absolute velocity of two vehicles
R_x	x coordinate difference between two vehicles
R_y	y coordinate difference between two vehicles
R_D	Relative distance of two vehicles
Merge_D1	Distance between ego vehicle and merging point
Merge_D2	Distance between merging vehicle and merging point
R_M_D	Difference between Merge_D1 and Merge_D2
R_Ax	Relative x acceleration of two vehicles
R_Ay	Relative y acceleration of two vehicles

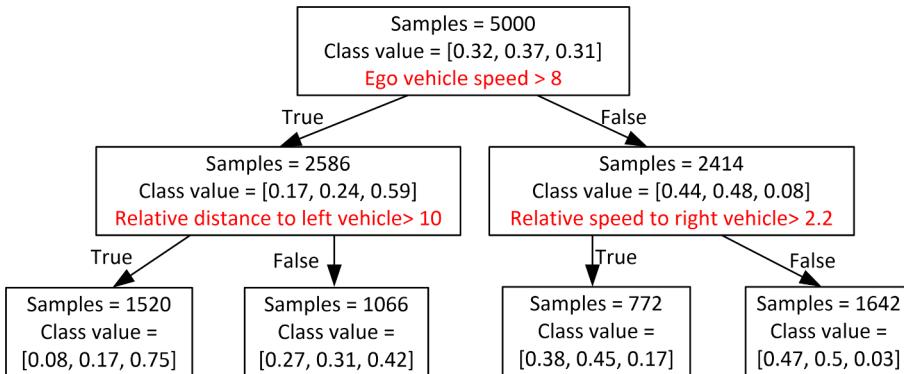


Fig. 7. Application example of single decision tree on driving behavior prediction task. The training set of this tree is randomly chosen from original training data set. The class values are the proportion of left turn samples, yield and right turn samples in the corresponding nodes.

proportion of historical training samples that meet the splitting rule, which represents a priori probability. We can see that this sample have the probability of 31% for right turn in the root node, but after falling in the final children node, the probability for right turn increase to 75%. It is the “ego vehicle speed” and “relative distance to left vehicle” features that contribute to the prediction probability increment. The contribution of “ego vehicle speed” to the right turn decision can be calculated as: $0.59 - 0.31 = 0.28$ and the contribution of “relative distance to left vehicle” is: $0.75 - 0.59 = 0.16$, the contribution values represent the increment of the probability that we predict this example sample to belong Right Turn category. We only show a single path of the tree, for the whole RF, all the contribution relevant to one certain feature should be added and averaged to obtain the final contribution value.

4. Results and analysis

In this section, we will analyze the distributions of the collected data to describe the drivers’ decision-making at USI, and then evaluate the trained RF models. Finally, we obtained the contribution of each feature for each sample by applying the interpretation method to the trained random forest model and test datasets.

4.1. Analysis of collected data

The collected data mainly recorded the decision-making process of driver participants at USI. Therefore, the distribution of the collected raw data can reflect the drivers’ driving characteristics (i.e., speed and acceleration preferences) during the decision-making process. Fig. 8 shows the density distribution of longitudinal and lateral velocity/acceleration of the ego vehicle.

The top plots in Fig. 8 demonstrate that Vx distributes from -10 m/s to 5 m/s and Vy from 0 m/s to 10 m/s . The distribution range for the negative Vx is wider than its positive range, which implies that for turning left, the drivers tend to drive at higher velocity to leave the intersections rapidly than turning right. Most of the Vx data point concentrates around zero because the yield decision leads to a longer data recording time, which in turn contributes to the increasing number of data point with value 0. For Vy, most of the data are distributed in the range of 2 m/s and 8 m/s , which implies that the driver participants prefer to drive at a relatively low speed in near-intersection area. There are also some data around 0 for Vy, it is also because the yield operation of drivers to avoid merging vehicle.

The bottom plots in Fig. 8 demonstrate that the Ax is distributed from -8 m/s^2 to 5 m/s^2 and Ay from -10 m/s^2 to 5 m/s^2 . For

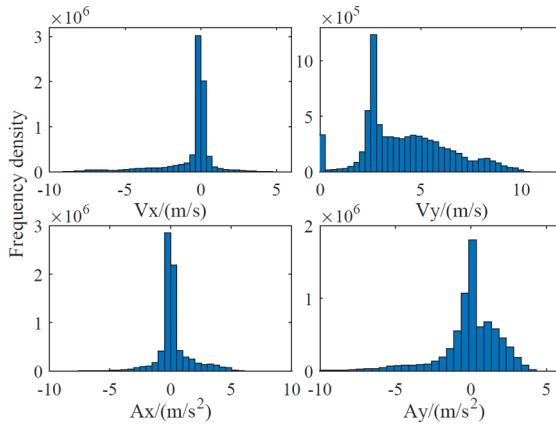


Fig. 8. Density distribution of longitudinal and lateral velocity (top), and acceleration (bottom) of the ego vehicle.

A_x , the negative proportion is quite large, this is because the distance between ego vehicle and merging point during left turn is greater than right turn, leading to the drivers' eagerness to accelerate and leave the intersections quickly. The A_y distributes in a wider range compared with A_x since the driver participants sometimes take strong braking for yielding.

4.2. Evaluation of the trained RF model

An interesting characteristic of ensemble methods is the built-in possibility of using the left-out samples to estimate model accuracy. Instead of using general performance index, the out-of-bag (OOB) error is introduced to represent the random forest error. During the training process, about one-third of the data were left out of the bootstrap samples and not used to reconstruct the decision tree. These data, termed OOB samples, were used for model test. Error estimated on these OOB samples is the OOB error. Fig. 9 shows the OOB error under the increasing number of grown trees.

From Fig. 9, it can be seen that the OOB error decreases as the number of grown trees increases since the increasing number of grown trees always performs a better feature splitting task. When the number of grown decision tree reaches to 9, the OOB error dropped to 0.01, which indicates 99% of prediction accuracy. When the tree number is 20, the error is about 0.002; finally, we choose the tree number to be 50 so that the error can be well converged to the tree number.

4.3. Feature contribution analysis

After finishing the training process, various samples in the test datasets were chosen for test and the feature contributions were then calculated using our proposed interpretation approach as shown in Section 2.2.

4.3.1. Analysis of example samples

We chose 3 example samples to illustrate how every feature contributes to the prediction of merging decisions. Fig. 10 shows the feature contributions and corresponding predicted decision for Left Turn (LT), Fig. 11 is for Yield condition and Fig. 12 for Right Turn (RT) conditions. The true label and the prediction score for each possibility are placed top of the figure, contribution of 13 features for each example condition is listed below. The value of every feature is shown in the left rectangles and the contribution of features is represented by the length of blue bar. Analysis for the three samples are shown as below:

- **Left Turn:** From the length of contribution bar, the relative distance along x direction between ego and merging vehicle (i.e., R_x) is

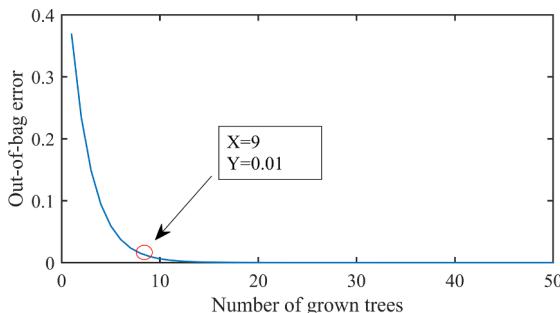


Fig. 9. The OOB classification error of the random forest model.

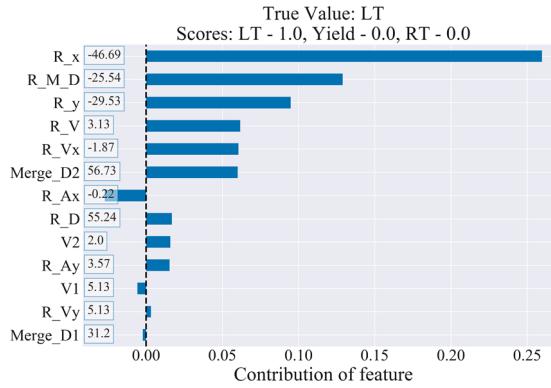


Fig. 10. Feature contributions of Left turn (LT) example sample.

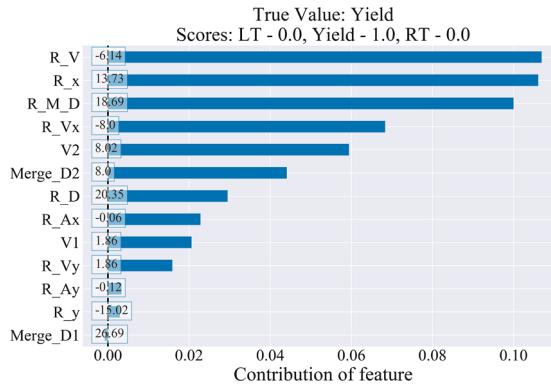


Fig. 11. Feature contributions of Yield example sample.

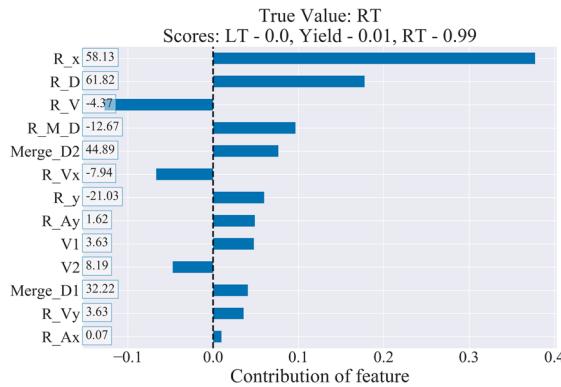


Fig. 12. Feature contributions of Right turn (RT) example sample.

the most contributive feature for predicting LT decision, followed by R_{M_D} . The distances between ego vehicles and merging point (i.e., $Merge_D1$) have little contributions to the prediction of final decision in this sample. From value of all 13 features, the R_x value (i.e., -46.69) implies that ego vehicle is quite far away from the merging vehicle; from the value of R_{M_D} , ego vehicle is 25.54 m nearer than merging vehicle to the merging point. These two features contribute much to the prediction of driver making LT decision. The contribution of R_{Ax} is negative, that is because the acceleration of ego vehicle is only slightly larger than merging vehicle (i.e., 0.22 m/s^2), which is not profitable for making LT decision.

- **Yield:** For the yield sample, the relative velocity between ego and merging vehicles (i.e., R_V) contributes most to the decision prediction, followed by R_x and R_{M_D} . Features such as accelerations and distance between vehicles and merging point contribute less. The prediction rarely depends on the y direction relevant features (i.e., R_r, R_{Ay}). However, since there is no feature with dominant value (i.e., big distance value between vehicles or very big ego vehicle velocity value, etc.), the contribution of all the features in this sample are within 0.15.
- **Right Turn:** For right turn sample, R_x is also the most contributive feature, the value of R_x (i.e., 58.13) makes it a good indicator

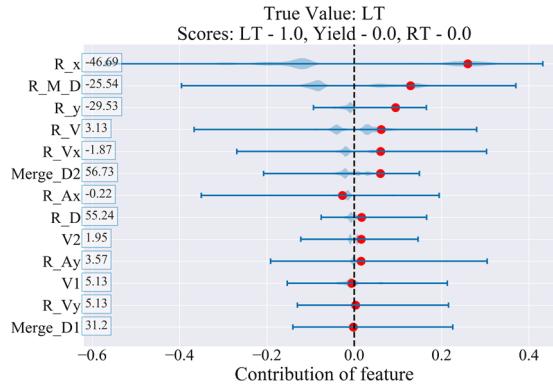


Fig. 13. Contribution distributions of 13 variables for Left Turn condition among all the samples.

for predicting that the driver will take the turning decision. From R_D and R_{M_D} value, it can be seen that the ego vehicle is far away from the merging vehicle and closer (12.67 m closer) to the merging point, so these two feature have 0.18 and 0.1 contribution to the decision prediction. The related velocity between two vehicles (i.e., R_V) is -4.37 m/s , which implies that the merging vehicle is much faster than ego vehicle, so the contribution of it to predict turning behavior of ego vehicle is -0.12 . This feature value also results in a dropping prediction confidence of RT decision. The value of R_{Vx} and $V2$ also indicate that a fast approaching vehicle will lead to a negative contribution to predict the turning decision. There is 0.01 probability that the driver will choose to yield in this condition.

4.3.2. Overall contribution distribution

Above analysis method help us gain insights into how much each feature contributes to the example merging decision. Since contribution for all samples has been obtained, in order to get comprehensive understanding of above samples among all samples, we illustrate the contribution distribution of features in Fig. 13, Fig. 14, Fig. 15. The feature values of example samples are shown in the left rectangles, the orange dots in the figure represent the feature contribution of sample and the blue area illustrates the contribution distribution of each features. The height of blue area represents the density corresponding to contribution values. From the three figures, it can be seen that the contribution distribution varies over different merging condition, which indicates that different feature choices should be made under various scenarios. Details for the three sample are as follows:

- **Left Turn:** From Fig. 13, we can infer that the chosen LT sample is representative because two of the most contributive features in this sample fall in the relatively high density area. There are also some high density areas with negative contributions, which contributes to the prediction case of not turning left. The contribution of last three features, i.e., $Merge_D1$, R_{Vy} and $V1$, mostly distribute near zero, which indicates that they contribute little to the decision prediction. Therefore we could exclude these three features when predicting traffic participants' LT decision in practical application. R_{Ay} and R_{Ax} is of little contribution to the decision in this sample; however, the distribution ranges of them are wide which implies the driving decision can be inferred depending on them sometimes in other cases.
- **Yield:** The example sample of Yield condition is also representative as most of the features are in the high contribution density area. The top four contributive features in this sample also have the widest contribution range. The maximum of the contribution of the top four features are all over 0.3. The final five features have little contributions to the prediction in this sample and the distribution range is also narrow, which indicates that they have limited contribution to Yield decisions. Therefore, these features

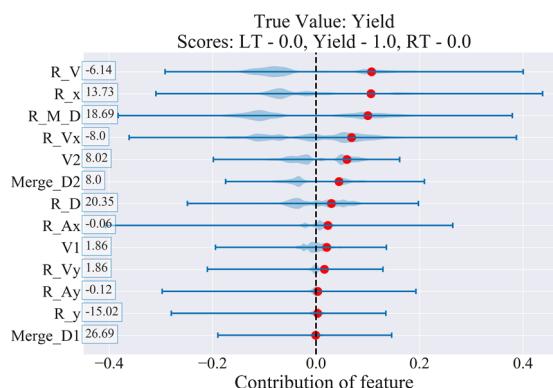


Fig. 14. Contribution distributions of 13 variables for Yield condition among all the samples.

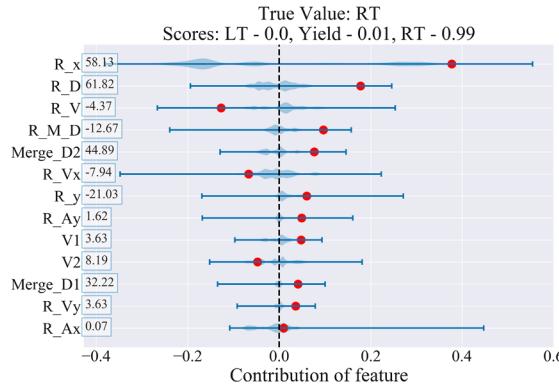


Fig. 15. Contribution distributions of 13 variables for Right Turn condition among all the samples.

can also be excluded in practice.

- **Right Turn:** For the right turn condition, the R_x has dominant influence on the decision prediction. Most of the data points of R_x have contribution of 0.3 or -0.2 . Besides the distance-related features, the relative speed and acceleration also contribute much to the right turn decision prediction. The contribution bars of $V1$ and $V2$ are shorter than R_V and R_{Vx} , indicating that for right turn decision prediction, the relative velocity features are more important than the velocity features. R_{Vy} and $Merge_D1$ contribute little in this sample, the contribution bars of them are also short, which means these two features are of little use in predicting the RT decision.

4.3.3. Trend of feature contribution

From the analysis above, we obtain the intuitive knowledge of the contribution distribution of all 13 features. Here, by extracting all the contribution of each single feature, we can get insight into how the corresponding contribution changes with varying the feature values. Fig. 16 shows the contribution changes of 4 highly contributive features.

- **Contribution Ranges:** As illustrated in Fig. 16, each individual feature has various contributions for different driving condition. Considering all merging scenarios as a whole, the contribution range of R_x is -0.5 to 0.5 , R_{M_D} is -0.4 to 0.4 , while R_{Vx} and R_V gain contribution value ranging from -0.3 to 0.3 . For a single feature R_x , when in the left turn condition, R_x makes contribution of -0.5 to 0.3 and in right turn condition, it has R_x totally opposite range of -0.3 to 0.5 ; however, when in yield condition, the contribution range of R_x shrinks to -0.2 to 0.3 . These results imply that this distance-related feature contributes more to the prediction of turning decision, but not for the decision of Yield.
- **Contribution Trend between Different Features:** From Fig. 16, it can be seen that R_{Vx} and R_V obtain similar changing trends of contribution. For the LT condition, the changing trend of their contribution is monotonously increasing and while decreasing for the Yield condition. For RT condition, the changing trend is not pronounced. However, its changing trend is totally different from R_x and R_{M_D} .
- **Contribution Trend for Single Features:** For a single feature, the changing trend varies between different driving conditions. Taking R_x for example, its trend of contribution for LT and RT are obviously opposite. For the yield condition, the contribution increases first and then decrease as R_x increases. The various changing trends in different driving condition can be explained as follows:
 - (a) For LT condition, the negative R_x indicates the merging vehicle is at the right side of the ego vehicle, which benefits the ego vehicle to make left turn. As R_x increases to near zero, which means two vehicles are very close in x direction, most of the drivers would choose to yield, so the contribution of the feature to LT decreases close to zero. When R_x is positive which implies that the merging vehicle at the left side is running to the right, it would threaten the ego vehicle to turn left. Therefore the contribution for predicting a left turning decision drops to a negative value.
 - (b) For the Yield condition, if R_x is in the range of -6 m to 12 m, the drivers tend to yield for the safety concerns, leading to a large contribution value. The positive and negative thresholds are different because the ego vehicle has asymmetric distance to the merging points when taking left and right turn. When R_x is out of the range, the high value of R_x will contribute to the turning decision prediction, so the contribution drops to below zero.
 - (c) For RT condition, the changing trend can be explained with the same idea of LT condition.

5. Conclusion

This paper investigated the contribution of 13 different features to driver behavior prediction at unsignalized intersections. We modeled and analyzed driver merging behavior using the features of the ego and merging vehicles based on the random forest approach. The driving data of 30 skilled drivers were collected and analyzed. Experimental results demonstrated that when the number of grown decision trees increases to 9, the out-of-bag error of the random forest will drop to 0.01. The performance of driving decision prediction rarely depends on $V1$, $Merge_D1$, and R_{Vy} and the top 4 contributive features are R_x , R_{Vx} , R_V and R_{M_D} . In

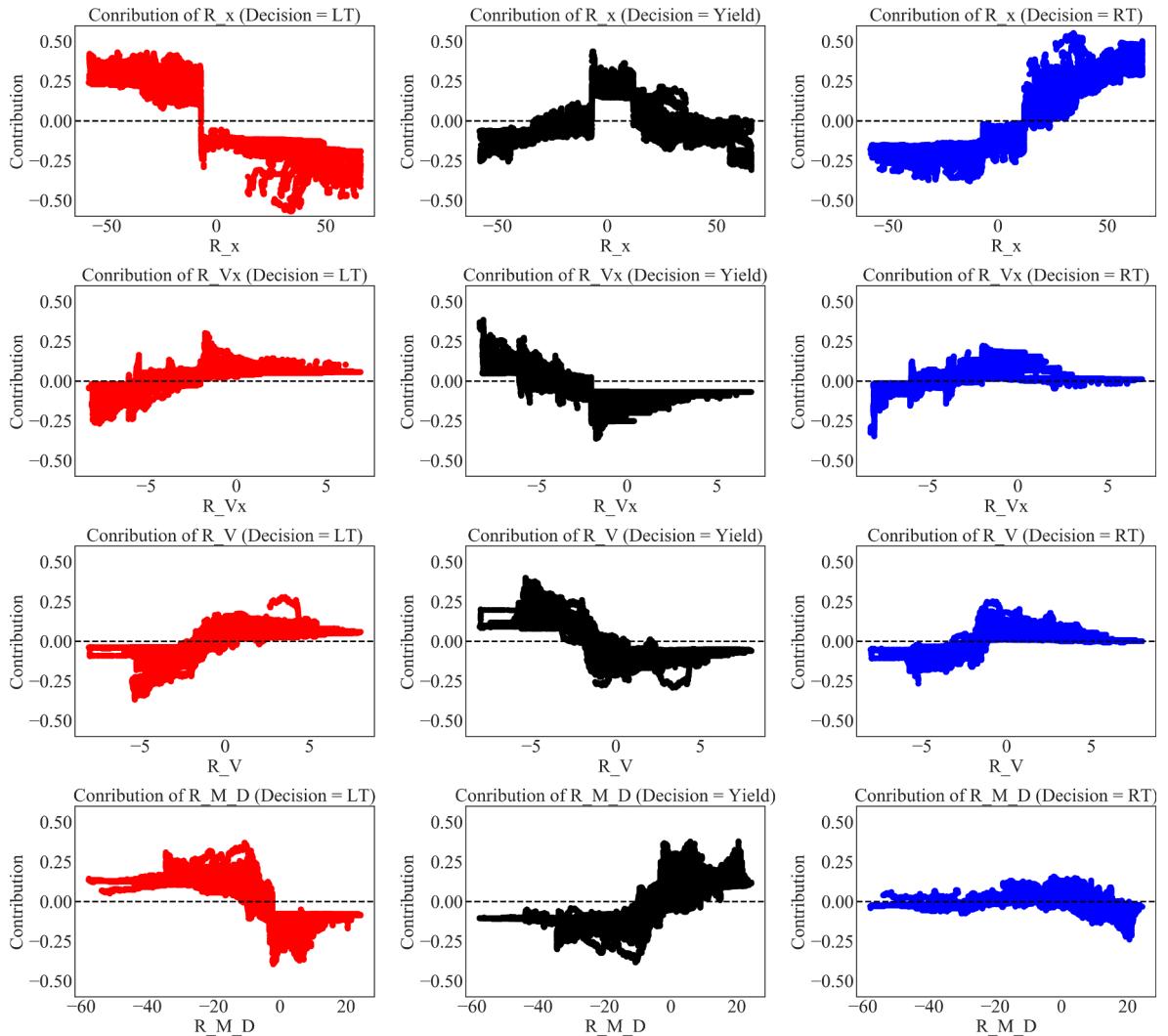


Fig. 16. Changing trends of 4 highly contributive features. From top to bottom are the contribution trends of R_x , R_{Vx} , R_V , R_{M_D} , respectively. The red scatter subplots are the changing trend for left turn conditions, the black scatter subplots are yield conditions and the blue ones are right turn conditions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

addition, we also found that (1) the contribution ranking of features for driver's decision-making prediction varies over the driving conditions and (2) the features that could make more contributions to prediction would present clearer changes of contribution trends. This work could finally provide researchers and engineers a guideline to choose the right features for specific research topics in driver behavior prediction at unsignalized intersections. Researchers should first collect test data, then select relevant features, and use the method introduced in this paper to analyze the contribution of each feature, finally select the best features for security control.

Acknowledgment

The authors would like to thank Guojun Wang from College of Automotive Engineering, Jilin University, Changchun, China. He provided us much help in the experiments. The research is partially supported by National Key R&D Program of China (2016YFB0100904) and National Natural Science Foundation of China under grants of U1564211, 51475206 and 51605185.

References

- Administration, N.H.T.S., et al., 2016. 2015 motor vehicle crashes: overview. *Traffic Saf. Facts Res. Note* 2016, 1–9.
 Breiman, L., 1996. Bagging predictors. *Mach. Learn.* 24, 123–140.
 Brilon, W., Wu, N., 2002. Unsignalized intersections-a third method for analysis. In: *Transportation and Traffic Theory in the 21st Century: Proceedings of the 15th International Symposium on Transportation and Traffic Theory*, Adelaide, Australia, 16–18 July 2002. Emerald Group Publishing Limited, pp. 157–178.

- Cao, W., Lin, X., Zhang, K., Dong, Y., Huang, S.L., Zhang, L., 2017. Analysis and evaluation of driving behavior recognition based on a 3-axis accelerometer using a random forest approach. IPSN 303–304.
- Chang, L.Y., Chien, J.T., 2013. Analysis of driver injury severity in truck-involved accidents using a non-parametric classification tree model. Saf. Sci. 51, 17–22.
- Cichosz, P., Pawełczak, L., 2014. Imitation learning of car driving skills with decision trees and random forests. Int. J. Appl. Math. Comput. Sci. 24, 579–597.
- De Beaucorps, P., Streubel, T., Verroust-Blondet, A., Nashashibi, F., Bradai, B., Resende, P., 2017. Decision-making for automated vehicles at intersections adapting human-like behaviour. In: IV'17-IEEE Intelligent Vehicles Symposium.
- de Oña, J., de Oña, R., Calvo, F.J., 2012. A classification tree approach to identify key factors of transit service quality. Expert Syst. Appl. 39, 11164–11171.
- Devarasetty, P.C., Zhang, Y., Fitzpatrick, K., 2011. Differentiating between left-turn gap and lag acceptance at unsignalized intersections as a function of the site characteristics. J. Transp. Eng. 138, 580–588.
- Elmitiny, N., Yan, X., Radwan, E., Russo, C., Nashar, D., 2010. Classification analysis of driver's stop/go decision and red-light running violation. Accid. Anal. Prev. 42, 101–111.
- Ge, X., Lu, J., Xiang, Q.J., Wang, P.Y., 2008. Study on driving behavior and traffic conflict at highway intersection. In: Power Electronics and Intelligent Transportation System, 2008. PEITS'08. Workshop on. IEEE, pp. 492–495.
- Guo, R.J., Lin, B.L., 2011. Gap acceptance at priority-controlled intersections. J. Transp. Eng. 137, 269–276.
- Hamed, M., Easa, S., Batayneh, R., 1997. Disaggregate gap-acceptance model for unsignalized t-intersections. J. Transp. Eng. 123, 36–42.
- Ho, T.K., 1995. Random decision forests. In: Proceedings of 3rd International Conference on Document Analysis and Recognition. IEEE, pp. 278–282.
- Hubmann, C., Becker, M., Althoff, D., Lenz, D., Stiller, C., 2017. Decision making for autonomous driving considering interaction and uncertain prediction of surrounding vehicles. In: Intelligent Vehicles Symposium (IV), 2017 IEEE. IEEE, pp. 1671–1678.
- Isele, D., Cosgun, A., 2017. To go or not to go: A case for q-learning at unsignalized intersections.
- Kaysi, I.A., Abbany, A.S., 2007. Modeling aggressive driver behavior at unsignalized intersections. Accid. Anal. Prev. 39, 671–678.
- Kuz'min, V.E., Polishchuk, P.G., Artemenko, A.G., Andronati, S.A., 2011. Interpretation of qsar models based on random forest methods. Mol. Inf. 30, 593–603.
- Kye, D.K., Kim, S.W., Seo, S.W., 2015. Decision making for automated driving at unsignalized intersection. In: Control, Automation and Systems (ICCAS), 2015 15th International Conference on. IEEE, pp. 522–525.
- Liu, M., Lu, G., Wang, Y., Wang, Y., Zhang, Z., 2014. Preempt or yield? an analysis of driver's dynamic decision making at unsignalized intersections by classification tree. Saf. Sci. 65, 36–44.
- Liu, M., Chen, Y., Lu, G., Wang, Y., 2017. Modeling crossing behavior of drivers at unsignalized intersections with consideration of risk perception. Transp. Res. Part F: Traffic Psychol. Behav. 45, 14–26.
- Nagalla, R., Pothuganti, P., Pawar, D.S., 2017. Analyzing gap acceptance behavior at unsignalized intersections using support vector machines, decision tree and random forests. Procedia Comput. Sci. 109, 474–481.
- Palczewska, A., Palczewski, J., Robinson, R.M., Neagu, D., 2013. Interpreting random forest models using a feature contribution method. In: Information Reuse and Integration (IRI), 2013 IEEE 14th International Conference on. IEEE, pp. 112–119.
- Pollatschek, M.A., Polus, A., Livneh, M., 2002. A decision model for gap acceptance and capacity at intersections. Transp. Res. Part B: Methodol. 36, 649–663.
- Prasettijo, J., Ali, M.M., Wu, N., 2010. Critical speed prediction at unsignalized intersection under mixed traffic condition. Traffic Transp. Stud. 2010, 823–837.
- Qi, L., Zhou, M., Luan, W., 2017. Impact of driving behavior on traffic delay at a congested signalized intersection. IEEE Trans. Intell. Transp. Syst. 18, 1882–1893.
- Rossi, R., Meneguzzi, C., Gastaldi, M., 2013. Transfer and updating of logit models of gap-acceptance and their operational implications. Transp. Res. Part C: Emerg. Technol. 28, 142–154.
- Sharma, A., Bullock, D., Peeta, S., 2011. Estimating dilemma zone hazard function at high speed isolated intersection. Transp. Res. Part C: Emerg. Technol. 19, 400–412.
- Shirazi, M.S., Morris, B.T., 2017. Looking at intersections: a survey of intersection monitoring, behavior and safety analysis of recent studies. IEEE Trans. Intell. Transp. Syst. 18, 4–24.
- Song, Weilong, Xiong, Guangming, Chen, Huiyan, 2016. Intention-aware autonomous driving decision-making in an uncontrolled intersection. Math. Prob. Eng. 2016, 1–15. <https://doi.org/10.1155/2016/1025349>.
- Spek, A., Wieringa, P.A., Janssen, W., 2006. Intersection approach speed and accident probability. Transp. Res. Part F: Traffic Psychol. Behav. 9, 155–171.
- Tran, Q., Firl, J., 2013. Modelling of traffic situations at urban intersections with probabilistic non-parametric regression. In: Intelligent Vehicles Symposium (IV), 2013 IEEE. IEEE, pp. 334–339.
- Tran, Q., Firl, 2014. Online maneuver recognition and multimodal trajectory prediction for intersection assistance using non-parametric regression. In: Intelligent Vehicles Symposium Proceedings, 2014 IEEE. IEEE, pp. 918–923.
- Wali, Behram, Khattak, Asad J., Bozdogan, Hamparsum, Kamrani, Mohsen, 2018. How is driving volatility related to intersection safety? A Bayesian heterogeneity-based analysis of instrumented vehicles data. Transp. Res. Part C: Emerg. Technol. 92, 504–524.
- Washington, S.P., Karlaftis, M.G., Mannerling, F., 2010. Statistical and Econometric Methods for Transportation Data Analysis. CRC Press.
- Xiao, Y., Ran, Q., Yang, J., Wang, Z., 2010. Analysis and modeling of crossing behavior at urban intersections in china. J. Transp. Eng. 137, 121–127.
- Yan, X., Radwan, E., 2006. Analyses of rear-end crashes based on classification tree models. Traffic Inj. Prev. 7, 276–282.
- Zhou, H., Ivan, J.N., Gårder, P.E., Ravishanker, N., 2017. Gap acceptance for left turns from the major road at unsignalized intersections. Transport 32, 252–261.
- Zyner, A., Worrall, S., Nebot, E., 2018. A recurrent neural network solution for predicting driver intention at unsignalized intersections. 1 1. IEEE Rob. Autom. Lett. PP. <https://doi.org/10.1109/LRA.2018.2805314>.