Accident Analysis and Prevention 42 (2010) 101–111

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Classification analysis of driver's stop/go decision and red-light running violation

Noor Elmitiny[a,*], Xuedong Yan[b], Essam Radwan[c], Chris Russo[a], Dina Nashar[a]

[a] Department of Civil & Environmental Engineering, University of Central Florida, 4000 Central Florida Blvd., Orlando, FL 32816, United States
[b] Center for Transportation Research, University of Tennessee, 309 Conference Center Bldg, Knoxville, TN 37996-4133, United States
[c] CATSS, University of Central Florida, United States

ARTICLE INFO

Article history:
Received 23 May 2008
Received in revised form 12 July 2009
Accepted 14 July 2009

Keywords:
Signalized intersections
Yellow signal change
Stop/go decisions
Red-light running
Classification tree
Video-based system

ABSTRACT

When the driver encounters a signal change from green to yellow, he is required to make a stop or go decision based on his speed and the distance to the stop bar making the wrong decision will lead to a red-light running violation or an abrupt stop at the intersection. In this study, a field data collection was conducted at a high-speed signalized intersection, where a video-based system with three cameras was used to record the drivers' behavior related to the onset of yellow. Observed data include drivers' stop/go decisions, red-light running violation, lane position in the highway, positions (leading/following) in the traffic flow, vehicle type, and vehicles' yellow-onset speeds and distances from the intersection. Further, classification tree models were applied to analyze how the probabilities of a stop or go decision and of red-light running are associated with the traffic parameters. The data analysis indicated that vehicle's distance from the intersection at the onset of yellow, operating speed, and position in the traffic flow are the most important predictors for both the stop/go decision and red-light running violation. This study illustrates that the tree models are helpful to recognize and predict how drivers make stop/go decisions and partake in red-light running violations corresponding to the traffic parameters.
© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

When the driver encounters a signal change from green to yellow he is required to make a stop or go decision based on his speed and the distance to the stop bar. A driver who is too far away from the intersection and decides not to stop may run the red light. On the other hand an over-defensive driver may decide to stop even though they are close enough to pass through safely. If a following driver makes a conflicting decision, this may lead to rear end crash. This dilemma increases the probability of rear end or angle crashes at intersections (red-light running violation is entering the intersection while the signal head is showing red phase). On a national basis red-light running and rear-end crashes result in substantial numbers of severe injuries and significant property damage. It has been estimated that red-light running results in 260 000 crashes each year, of which approximately 750 were fatal (Retting et al., 2002). Porter and England observed 5112 drivers entering six traffic-controlled intersections in three cities and found that 35.2% of observed light cycles had at least one red-light runner prior to the onset of opposing traffic (Porter and England, 2000). Rear-end crashes account for 29.6% of all reported crashes in the United States (NHTSA "Traffic Safety Facts", 2005), but constitute more than 40.2% percent of all reported crashes in Florida according to an analysis of 1531 signalized intersections history in the state of Florida (Abdel-Aty et al., 2005). A prior study indicated that most rear-end crashes at a signalized intersection occur when two successive drivers approaching the intersection make conflicting decisions when the yellow signal appears (Mahalel and Prashker, 1987). Therefore, properly recognizing driver behavior associated with the signal change at signalized intersections is crucial for improving traffic operation and safety.

The yellow signal change is used to warn approaching drivers of an imminent change in right-of-way at intersections. At the onset of the yellow signal indication, drivers who are close to the intersection may clear the intersection before the signal indication changes to red, while drivers who are far enough from the intersection should stop at the intersection. Drivers' incorrect decisions to cross the intersection at the onset of the yellow change may lead to RLR violations or traffic conflicts with the vehicles in front of them, whose drivers decide to stop at the intersections. A previous study (Baugley, 1988) categorized three groups of red-light violators: the first type is drivers who could have cleared the intersection before the red, but were delayed either by their own indecision or by slower traffic in front of them (Retting et al., 2002); the second

* Corresponding author.
E-mail addresses: nelmitiny@yahoo.com (N. Elmitiny), yxd22222@yahoo.com (X. Yan), aeradwan@mail.ucf.edu (E. Radwan), chrisrusso89@yahoo.com (C. Russo), nashardina@hotmail.com (D. Nashar).

0001-4575/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2009.07.007

type is drivers in the dilemma zone (Porter and England, 2000); and the third type is aggressive drivers who could have stopped comfortably, but chose to run the red light deliberately (NHTSA, "Traffic Safety Facts" 2005). In this study we attempt to find out the factors that affect the driver decision and what can possibly cause stop/go decision hesitation.

Many studies reported that dilemma and option zones existing upstream of intersections at the onset of the yellow signal are associated with larger variability in the drivers' stop/go decisions (Mahalel and Prashker, 1987; Zimmerman and Bonneson, 2004; Srinivasa et al., 2005). When the driver is going at a speed lower than the speed limit an option zone is created, i.e. an area where the driver can stop or cross successfully. When a driver is traveling higher than the speed limit a dilemma zone is created, i.e. the driver can neither stop without slamming on the brakes or cross safely without running the red light (Papaioannou, 2007). When vehicles are located in an option zone, drivers can either easily stop before the stop line or successfully clear the intersection before the onset of the red signal. The option zones existence may contribute to rear-end conflicts due to the variability of the drivers' stop/go decisions. On the other hand, drivers who are in a dilemma zone can neither stop, nor cross the stop line before the signal turns red. Therefore, the dilemma zone's existence may result in both rear-end conflicts and RLR violations.

In numerous studies that focused upon driver behavior associated with the signal change, the probability for drivers' stop/go decisions was modeled as a function of the space or potential time (gap) from the stop line using the logistic regression (logit) technique (Newton et al., 1997; Köll et al., 2003; Yan et al., 2007; Papaioannou, 2007; Chang et al., 1985). Based on the function, most drivers will either stop when they have a shorter distance from the intersection or cross the intersection when they have a longer distance from the intersection. The time or space point where 50% of the drivers choose to stop or cross may result in situations where the stop/go decision is least predictable. Furthermore, the area between the 10th and 90th percentiles of the stopping probability function was described as an indecision zone. A longer indecision zone indicates a larger variability in drivers' stop/go deci-

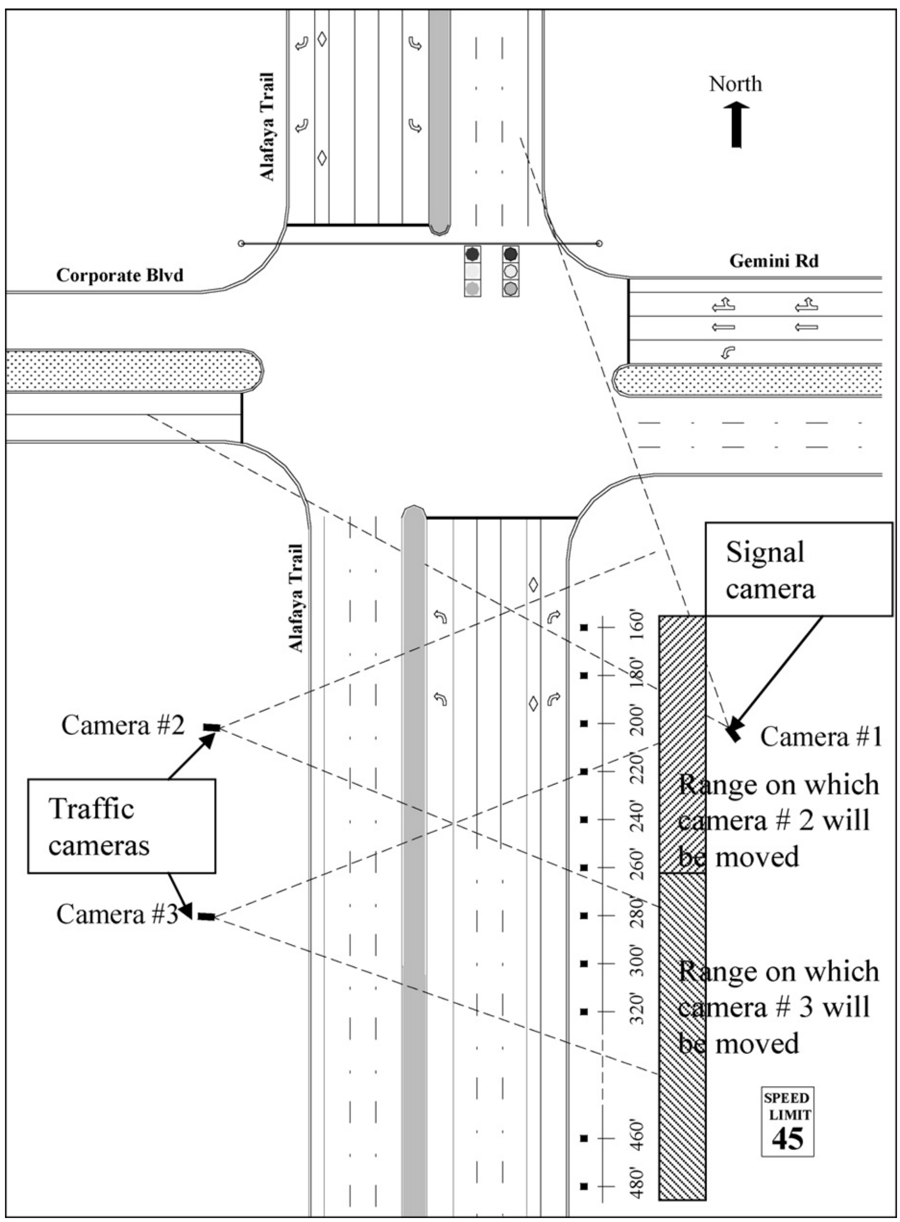

Fig. 1. Video-based system for data collection.

(a) Snapshots from camera #1

(b) Snapshots from cameras #2 and #3

Fig. 2. Data extraction from videos using the Adobe Premiere Pro software.

sions and higher rear-end crash risk (Parsonson, 1978; Sheffi and Mahmassani, 1981).

In this study, a field data collection was conducted at a high-speed signalized intersection, where a three-camera video-based system was used to record the drivers' stop/go decisions related to the yellow signal change, RLR violations, vehicles' yellow-onset speeds and distances from the intersection, lane positions in the highway, positions in the traffic flow, and vehicle types. The objective of this study is to investigate how drivers' stop/go decisions and RLR violations are associated with the traffic parameters like: distance, speed, yellow-entry time, lane position, leading/following and vehicle type. Both probabilities of stop/go decision and RLR are modeled based on the classification tree technique.

2. Research method

2.1. Observation site description

A high-speed intersection (see Fig. 1) named Alafaya Trail & Corporate Blvd., located in a typical Central Florida (Orlando) suburban area, was selected for this study. The northbound approach of the intersection was chosen to observe drivers' behavior after the green phase terminated. The approach has three through lanes, one left-turn lane, and one right-turn lane. The length of the yellow phase is 4.3 s and the length of the all-red phase is 1 s. Yellow and all-red intervals were compliant with the ITE values for the speed limit and the intersection width (FDOT Traffic Engineering Manual, 2005). The speed limit posted upstream of the northbound approach is 72.4 km/h (45 mph); however, another speed limit at only a 304.8 m (1000 ft) distance downstream from the intersection is 50 mph. This design results in a high percent of drivers traveling around 80.5 km/h (50 mph) at the intersection. Generally, the speeding drivers facing the yellow signal are more likely to fall into the dilemma zone so as to run red lights (Papaioannou, 2007). Therefore, a high RLR violation rate at the approach was expected.

2.2. Data collection using video-based system

A three-camera video-based system was used to record the driving behavior associated with the signal change (three cameras were

Table 1
Descriptive statistics for operating speed (SPEED) at the onset of the yellow.

Factor	Sub-level	*N*	Mean (mph)	Std. deviation (mph)	Minimum (mph)	Maximum (mph)
ST_GO[a]	Stop	692	76.9 (47.8)	7.7 (4.8)	40.9 (25.4)	101.0 (62.8)
	Go	600	80.1 (49.8)	8.0 (5.0)	48.6 (30.2)	105.1 (65.3)
RLR[b]	No	1065	78.1 (48.5)	7.9 (4.9)	40.9 (25.4)	101.1 (62.8)
	Yes	227	79.7 (49.5)	8.7 (5.4)	48.6 (30.2)	105.1 (65.3)
LD_FL[c]	Leading	579	79.5 (49.4)	8.0 (5.0)	52.6 (32.7)	105.1 (65.3)
	Following	713	77.4 (48.1)	8.0 (5.0)	40.9 (25.4)	101.1 (62.8)
L_POSITION[c]	Right	470	76.4 (47.5)	8.5 (5.3)	40.9 (25.4)	105.1 (65.3)
	Middle	458	80.0 (49.7)	7.7 (4.8)	52.6 (32.7)	101.1 (62.8)
	Left	364	78.9 (49.0)	7.6 (4.7)	48.6 (30.2)	101.1 (62.8)
V_TYPE[d]	PC	726	78.9 (49.0)	8.0 (5.0)	50.5 (31.4)	101.8 (62.8)
	LTV	538	77.9 (48.4)	8.0 (5.0)	40.9 (25.4)	105.1 (65.3)
	LSV	28	73.2 (45.5)	7.4 (4.6)	57.9 (36.0)	84.8 (52.7)
Total		1292	78.4 (48.7)	8.0 (5.0)	40.9 (25.4)	105.1 (65.3)

[a] Driver's stop/go decision.
[b] Whether the going vehicle ran a red light or not.
[c] Whether the vehicle was in a leading position or a following position in the traffic flow.
[d] Vehicle type [passenger car (PC), light truck vehicle (LTV), larger size vehicle (LSV)].

set on the side of the road, the cameras and research team were hidden in the bushy area along the road side to avoid affecting the driver behavior at the intersection). As shown in Fig. 1, camera #1 was pointed toward the traffic signal heads to film signal phasing status and vehicles around the stop line; cameras #2 and #3 were positioned in the left side of the approach and pointed perpendicular to the highway to film the vehicles approaching toward the intersection during the signal change. Moreover, a marker system composed of a series of yellow flags at 20-ft intervals along the approach was utilized as relative coordinates as shown in Fig. 2. This technique was used to help researchers clearly identify the vehicles' distances from the intersection and accurately extract vehicles' speeds at the onset of the yellow in the video analysis later.

The three cameras were synchronized by pointing the three cameras to a single point and recording traffic as it passed through. The cameras were then left running and each camera was positioned correctly at its location. Later this common event was used to determine the fractional difference in time between the three videos on the computer. Since the software we used can display 30 frames per second, the difference between the videos can be determined to an accuracy of 1/30 s.

Cameras #2 and #3 could record 80-ft distance range each. To cover the required distance to obtain stop/go probability as a function of approaching distance from 160 ft to 480 ft upstream of the intersection, camera #2 was moved to cover the range of 160–320 ft and camera #3 was moved to cover the range from 320 ft to 480 ft span as shown in Fig. 1. Each camera covered an 80 ft span for 1 h of data collection and then was moved to cover another 80 ft span within its range until the required sample size was reached.

2.3. *Extracting data from videos*

A total of 36 1-h videos including 28 off-peak hours (1:30 pm to 4:30 pm) and 8 peak hours (4:30 pm to 6:00 pm) were filmed during weekdays. Adobe Premiere Pro software was used to upload and compress the videos for computer storage in Window's WMV format. The video data collection methodology was selected because it may produce higher quality traffic data than manual methods. With a video rate of 30 frames per second, the error caused by the video program is 0.03 s for the event-times data. Due to the analyst's visual judgment error for vehicle positions, the total possible error for the event-times data could be up to 0.1 s (Bonneson and Fitts, 1995).

Two computers were used simultaneously for video analysis in the Adobe Premiere Pro software. One computer was used to analyze the videos from camera #1. As shown in Fig. 2a, the snapshots indicate the moment of signal change status and the moment at which a vehicle is entering the intersection during the yellow or red. From the time-elapse photography, researchers can accurately record yellow or red entry time for the vehicle entering the intersection after green, vehicle size, which lane the vehicle was traveling in, and whether the vehicle was the leading or following vehicle in the traffic flow. The other computer was used to analyze the videos from cameras #2 and #3. As shown in Fig. 2b, the snapshots indicate the vehicles' positions at the moment that the signal displays changes from green to yellow. Based on the marker system (with predefined distances from the intersection) and the corresponding reference lines (with a perspective effect of the two-dimensional graph), researchers can identify a vehicle's distance from the intersection. The vehicle's speed was determined during the video analysis by calculating the time taken by the car to cover 80 ft distance in the video measured to the accuracy of one frame in the video. A previous study indicated that the video-based speed measurements yield data of comparable quality to the radar speed measurements (Strong et al., 2003).

Using the method of extracting data from videos, the behavior of a total of 1292 vehicles' behavior was analyzed and recorded. The size of sample does not include vehicles forced to stop by the vehicle in front. For each vehicle whose driver had a chance to make a stop/go decision at the onset of the yellow, the data extracted from the videos were organized into the following variables:

- DISTANCE (in ft): vehicle's distance from the intersection at the onset of the yellow indication;
- SPEED (in mph): vehicle's operating speed at the onset of the yellow;
- ST_GO: driver's stop/go decision (stop = 0; go = 1);
- Y_TIME (in seconds): time elapsed from the onset of the yellow until the vehicle entered the intersection, if the vehicle crossed the intersection;
- RLR: whether the going vehicle ran a red light or not (no = 0; yes = 1);
- LD_FL: whether the vehicle was in a leading position or a following position in the traffic flow (leading = 0; follow = 1); if headway was shorter than 1 s the car was considered following in the platoon;

Table 2
Analysis of variance table for SPEED.

Source	Type III sum of squares	df	Mean square	*F*	Sig.
Corrected model[*]	3962.368(a)	7	566.053	25.315	.000
Intercept	467346.922	1	467346.922	20900.949	.000
ST_GO[a]	1729.169	1	1729.169	77.333	.000
RLR[b]	21.836	1	21.836	.977	.323
LD_FL[c]	1017.471	1	1017.471	45.504	.000
L_POSITION[d]	890.048	2	445.024	19.903	.000
V_TYPE[e]	485.301	2	242.650	10.852	.000
Error	28710.345	1284	22.360		
Total	3098143.040	1292			
Corrected total	32672.713	1291			

[*] R-squared = .121 (adjusted R-squared = .116).
[a] Driver's stop/go decision.
[b] Whether the going vehicle ran a red light or not.
[c] Whether the vehicle was in a leading position or a following position in the traffic flow.
[d] Vehicle's lane position (left lane, middle lane, right lane).
[e] Vehicle type [passenger car (PC), light truck vehicle (LTV), larger size vehicle (LSV)].

- L_POSITION: the vehicle's lane position (left lane = 0; middle lane = 1; right lane = 2);
- V_TYPE: vehicle type [passenger car (PC) = 0; light truck vehicle (LTV) = 1; larger size vehicle (LSV) = 2];

3. Observation results and data analyses

3.1. Operating speed

Table 1 lists the statistical summary of the operating speeds of vehicles at the onset of the yellow. The mean operating speed at the approach is 78.4 km/h (48.7 mph), slightly higher than the 72.4 km/h (45 mph) speed limit. In the subsequent statistical analysis, an ANOVA was used to investigate differences between factors (see Table 2). The ANOVA results show that traffic factors including ST_GO ($p < 0.001$), LD_FL ($p < 0.001$), L_POSITION ($p < 0.001$), V_TYPE ($p < 0.001$) have significant effects on the operating speed, but the mean speed of red-light runners is statistically similar to that of non-red-light runners ($p = 0.323$).

Comparisons among sub-levels in the traffic factors indicate that the mean speed for vehicles with go decisions is higher those that with stop decisions (M = 49.8 mph, SD = 5.0 mph) and (M = 47.8 mph, SD = 7.7 km/h 4.8 mph) respectively. The mean speed for the leading vehicles is higher than that for the following vehicles (M = 49.4 mph, SD = 5.0 mph) and (M = 48.1 mph, SD = 5.0 mph) respectively. The mean speed of vehicles traveling in the left lane and the middle lane is higher than that in the right lane (M = 49.0 mph, SD = 4.7 mph), (M = 49.0 mph), SD = 74.7 mph) and (M = 47.5 mph, SD = 5.3 mph) respectively. The mean speeds of light truck vehicles and passenger cars are very close to each other, but apparently higher than that of large-size vehicles (M = 48.4 mph, SD = 5.0 mph), (M = 49.0 mph, SD = 5.0 mph) and (M = 45.5 mph, SD = 4.6 mph) respectively. Furthermore, although the operating speed is independent of RLR, the mean speed of red-light runners (M = 49.5 mph, SD = 5.4 mph) is statistically higher ($p < 0.001$) than that of vehicles with stop decisions (M = 47.8 mph, SD = 4.8 mph).

3.2. Stop or go decision and RLR

A driver's stop/go decision at the onset of the yellow change is one of the essential behaviors at signalized intersections. Incorrect go decisions or decision hesitation may result in RLR violations, which may increase the exposure to angle collisions. During the data-collection period the following was observed: among the 1292 drivers who encountered the yellow change 601 drivers made stop decisions, 691 drivers made go decisions, and 227 RLR violations were observed. The driver was identified to have encountered yellow change if the vehicle was located between the range of 160–480 ft upstream of the intersection at the onset of yellow. A pilot study found that closer than 160 ft are not likely to stop at the intersection and drivers further away than 480 ft will make the stop, the field range shown in Fig. 1. Fig. 3 illustrates distributions for the stop/go decision and RLR violations by yellow-onset distance interval from the intersection. Fig. 3a shows that as the distance from the intersection increases, the probability of a stop decision increases and the probability of a go decision decreases. When drivers are located in the distance interval of 280–320 ft, the probabilities of both stop and go decisions are close to 0.5. This means that the 280 ft, 320 ft area is considered the zone where the driver shows the largest variability in their stop/go decision. Fig. 3b shows that most red-light runners were concentrated in the distance interval of 85.3 m (280 ft) to 115.8 m (380 ft). When the green signal terminates, drivers close to the intersection are less likely to run red lights unless their vehicle's speed is very slow; also, drivers far away from the intersection are also less likely to run red lights.

3.3. Tree-based classification analysis of stop/go decision and RLR

Classification trees, also called decision trees, are among popular statistical tools that emerged from the field of machine learning and data mining. A classification tree classifies observations by recursively partitioning the predictor space. The resultant model can be expressed as a hierarchical tree structure. Due to its nonparametric nature and easy interpretation, decision trees have received wide popularity in a variety of fields, especially since the introduction of the Classification and Regression Trees (CART; Breiman et al., 1984). The advantage of using classification trees over many of the other methods is the effectiveness to construct classifications of driver behavior through segmenting the data set into smaller and more homogeneous groups. For a classification tree, the target variable takes its values from a discrete domain, and for each leaf node the classification tree associates a probability for each class (i.e. value of the target variable). In tree-structured representations, a set of data is represented by a node, and the entire dataset is represented as a root node. When a split is made, two child nodes are formed, which correspond to partitioned data subsets. If a node is not to be split any further, it is called a terminal node that is associated with a group membership; otherwise, it is an internal node. The tree structure is constructed following a set of decision rules applied sequentially. Each decision rule is used to form branches (i.e. split-

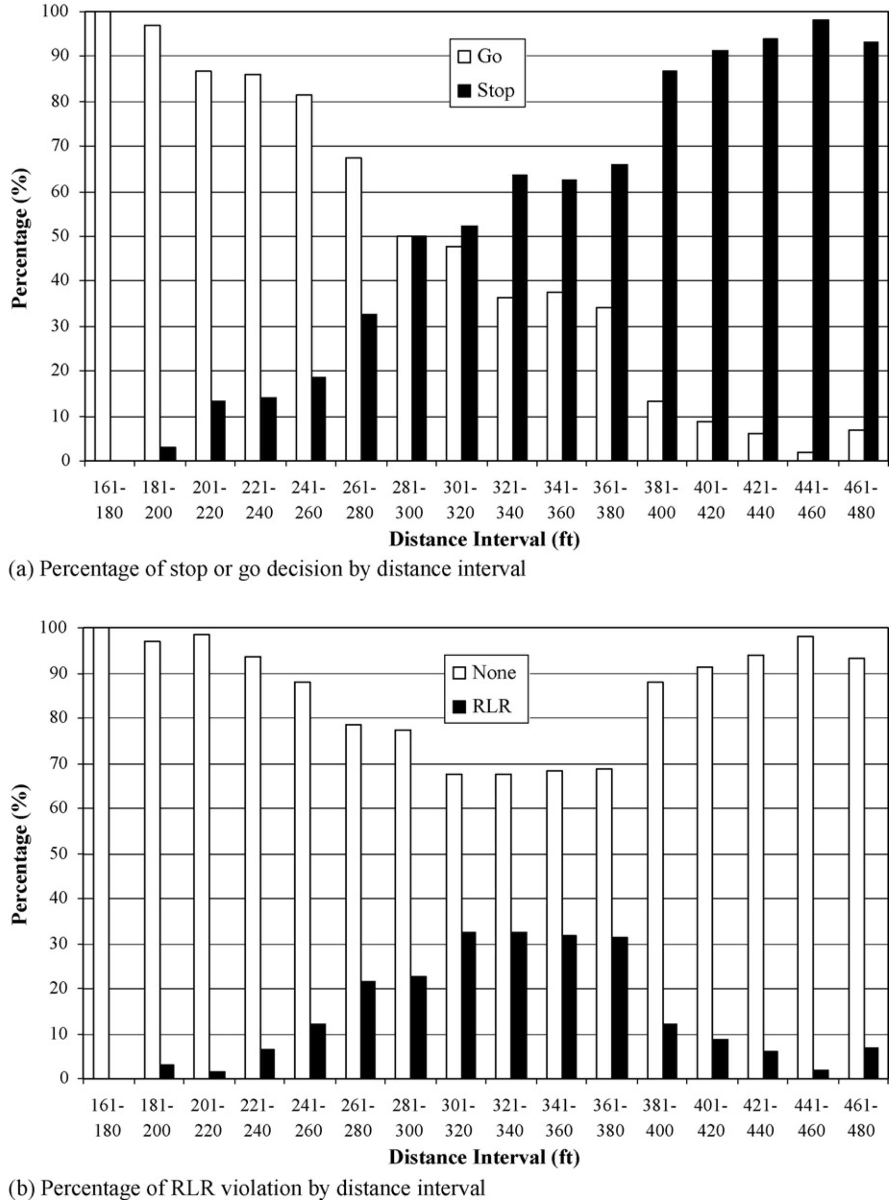

Fig. 3. Distributions of stop/go decision and RLR violation by distance interval from the intersection.

ting) connecting the root node to the terminal node at a certain level of the tree. More detailed descriptions of these decision tree algorithms are beyond the scope of the present study. For further discussions of tree methodology, the reader is referred to Breiman et al. (1984).

In this study, the classification tree analyses were carried out using the SPSS software package (version 13.0; SPSS Inc., Chicago, IL, USA). The classification trees were developed based on the CART approach, and the Gini criterion (or index) was used as a measure of split criteria. With the CART method, one can avoid over fitting the model by “pruning the tree.” In this study, the tree is trimmed automatically to the smallest sub-tree based on one standard error as the specified maximum difference in risk. Since the data size is not very large, the minimum number of cases for parent nodes was set as 30 and the minimum number of cases for child nodes was set as 10. Additionally, the cross-validation method (10-folds) was used to assess how well the tree structure generalizes to a larger population. The Gini index equation for a discrete probability function $f(y)$, where y_i $i = 1$ to n, are the points with nonzero probabilities and which are indexed in increasing order ($y_i < y_{i+1}$):

$$G = 1 - \frac{\sum_{i=1}^{n} f(y_i)(S_{i-1} + S_i)}{S_n} \tag{1}$$

where

$$S_i = \sum_{j=1}^{i} f(y_i) y_i \tag{2}$$

$S_0 = 0$ (initial value of S_{i-1}) and y_i = the points with nonzero probabilities indexed in increasing order.

Based on these inputs, two binary classification tree models were developed for the driver’s stop/go decision and RLR, respectively. For the stop/go decision model, the target variable is ST_GO (stop = 0; go = 1); for the RLR model, the target variable is RLR (no = 0; yes = 1). The classification results are depicted and discussed in the following.

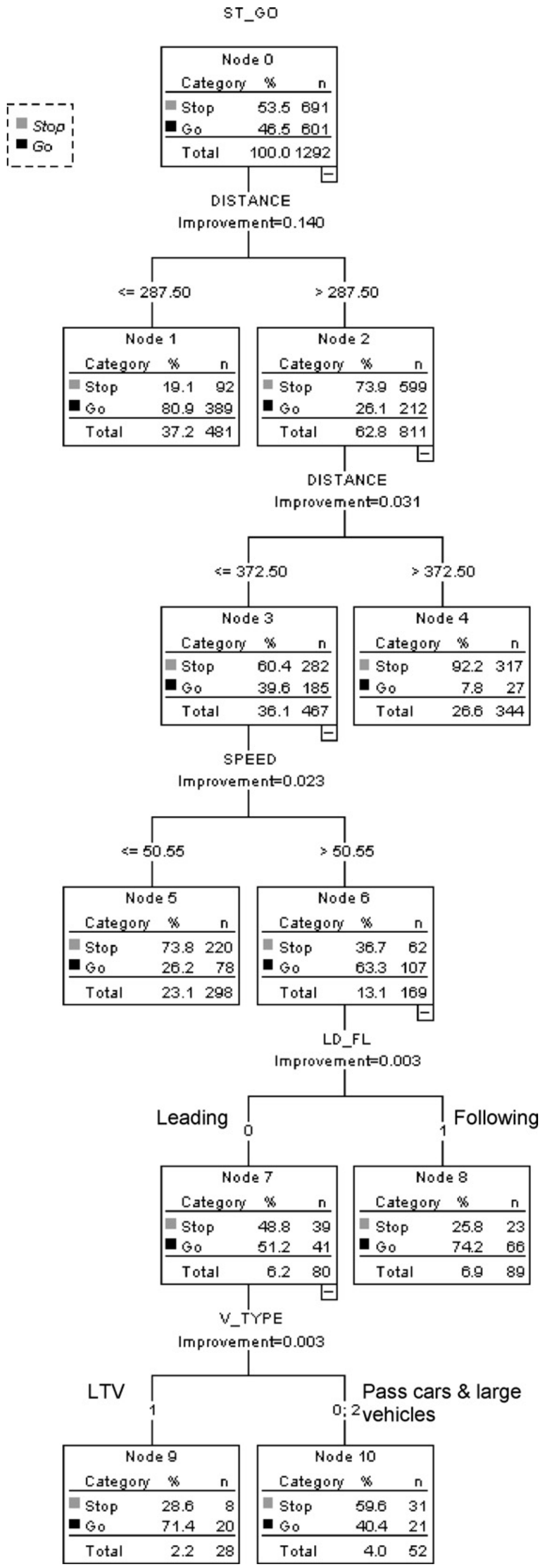

Fig. 4. Classification tree diagram for the stop/go decision model.

3.3.1. *Stop/go decision model*

Fig. 4 illustrates the classification tree diagram for the stop/go decision model, which includes six terminal nodes. The model generates the following classification rules for the stop/go decision:

- If the vehicles' yellow-onset distances are smaller than 287.5 ft, most of drivers (80.9%, 389 veh) would cross the intersection.
- If the vehicles' yellow-onset distances are larger than 372.5 ft, most of drivers (92.2%, 317 veh) would stop at the intersection.
- If the vehicles' yellow-onset distances are between 287.5 ft and 372.5 ft, the operating speed plays an important role in the stop/go decision. If operating speeds are lower than 50.55 mph, drivers are more likely to stop (73.8%, 220 veh); on the contrary, if the drivers are speeding (higher than 50.55 mph) at the onset of the yellow, they are more likely to cross the intersection (63.3%, 107 veh).
- An interesting finding shows that speeding drivers are more likely to cross the intersection when they are following drivers in the traffic flow than when they are the leading driver (74.2%, 66 veh vs. 51.2%, 41 veh).
- Furthermore, for speeding drivers who are the leading driver in the traffic flow, vehicle types have a significant effect on their stop/go decisions: light truck vehicles are more likely to cross the intersection than passenger cars and large-size vehicles (71.4%, 20 veh vs. 40.4%, 21 veh).

Additionally, Fig. 5 shows the independent variable importance to the stop/go decision model. According to the importance order in the figure, DISTANCE, SPEED, and LD_FL are the most important variables to predict drivers' stop/go decisions; however, V_TYPE and L_POSITION are less significant.

3.3.2. *RLR model*

Fig. 6 illustrates the classification tree diagram for the RLR model, which includes seven terminal nodes. To describe the trees in English roles for example node 0 has 53.5% stop and 46.5% go and best split is using the variable "Distance". If vehicle "Distance" is smaller than or equal to 287.5 ft (87.6 m) the tree categorize it as node 1 meaning that this vehicle have 80.9% chance to go and 19.1% to stop. If the vehicle "Distance" is higher than 287.5 ft (87.6 m) the tree categorize it as node 2 73.9% stop and 26.1% go this node is considered not pure enough and requires additional splitting. The splitting goes on accordingly till all the terminal nodes reach desired

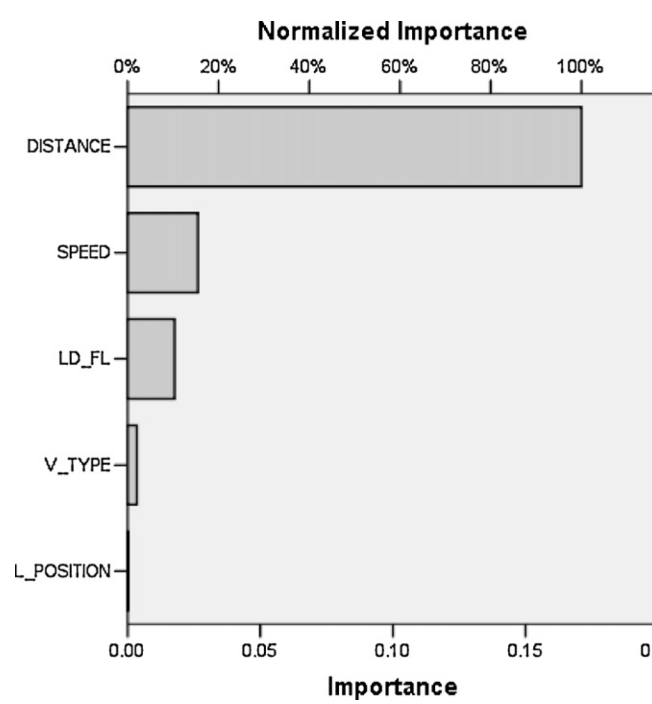

Fig. 5. Independent variable importance for the stop/go decision model.

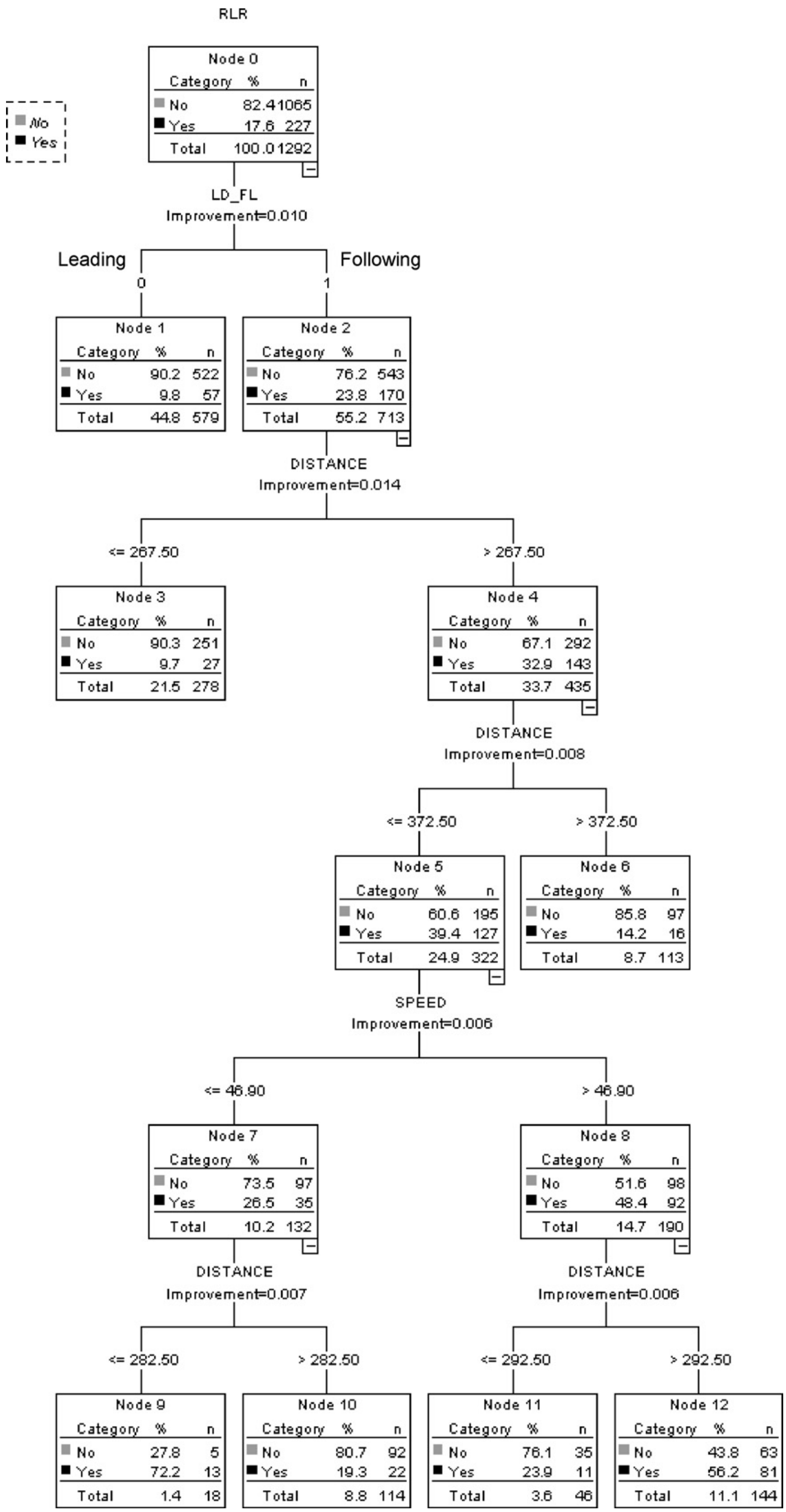

Fig. 6. Classification tree diagram for the RLR model.

purity. The model generates the following classification rules to predict the RLR probability:

- If the vehicles are in the leading position in the traffic flows, the drivers are less likely to run the red light (9.8%, 57 veh).
- For the following vehicles, if the yellow-onset distances are smaller than 267.5 ft, few drivers (9.7%, 27 veh) run red lights; if the yellow-onset distances are larger than 372.5 ft, 8.7% drivers would run the red light, a more severe RLR violation (8.7%, 113 veh).

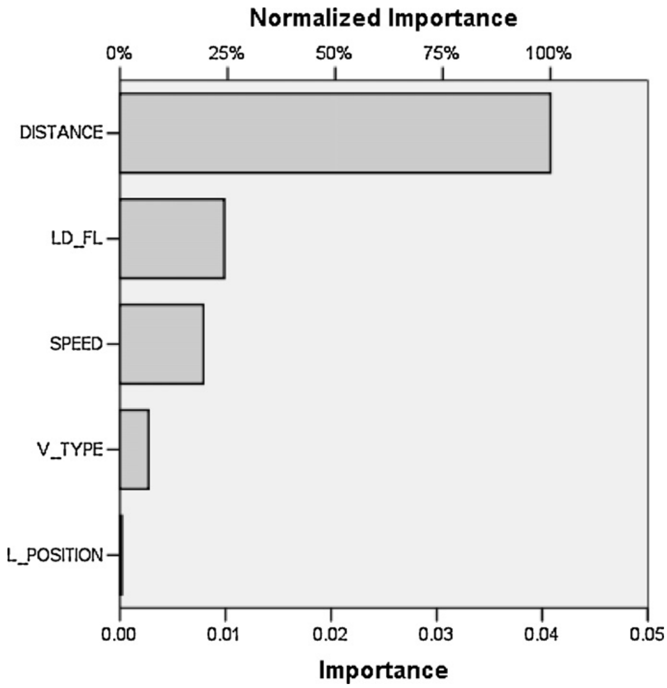

Fig. 7. Independent variable importance for the RLR model.

- If the following vehicles are located in the distance interval between 267.5 ft to 372.5 ft, the operating speed is significantly associated with the RLR probability, and drivers are categorized into four speed-distance subgroups.
  - If the vehicle speeds are lower than 46.9 mph and located in the distance interval between 267.5 ft to 282.5 ft, 72.2% (13 veh) of drivers run the red light, presumably due to the of dilemma zone effect;
  - If the vehicle speeds are lower than 46.9 mph and located in the distance interval between 282.5 ft to 372.5 ft, only 19.3% (22 veh) of drivers would run the red light;
  - If the vehicle speeds are higher than 46.9 mph and located in the distance interval between 267.5 ft to 292.5 ft, 23.9% (11 veh) of drivers would run the red light; and
  - If the vehicle speeds are higher than 75.5 km/h (46.9 mph) and located in the distance interval between 89.2 m (292.5 ft) to 113.5 m (372.5 ft), 56.2% (81 veh) drivers would run the red light. This subgroup includes 81 RLR observations, which account for 35.5% (81/228) of RLR violations.

Table 4
Analysis of variance table for Y_TIME.

Source	Type III sum of squares	df	Mean square	F	Sig.
Corrected model[*]	299.091(a)	7	42.727	297.828	.000
Intercept	54.347	1	54.347	378.821	.000
LD_FL[a]	2.493	1	2.493	17.376	.000
L_POSITION[b]	.457	2	.228	1.592	.204
V_TYPE[c]	.954	2	.477	3.323	.037
SPEED[d]	26.100	1	26.100	181.931	.000
DISTANCE[e]	290.862	1	290.862	2027.444	.000
Error	85.073	593	.143		
Total	9787.750	601			
Corrected total	384.164	600			

[*] R-squared = .779 (adjusted R-squared = .776).
[a] Whether the vehicle was in a leading position or a following position in the traffic flow.
[b] Vehicle's lane position (left lane, middle lane, and right lane).
[c] Vehicle type [passenger car (PC), light truck vehicle (LTV), larger size vehicle (LSV)].
[d] Vehicle's operating speed at the onset of the yellow.
[e] The distance separating the vehicle from the stop bar at the onset of the yellow.

Fig. 7 shows the independent variable importance to the RLR model. According to the importance order, DISTANCE, LD_FL, and SPEED are the most important variables to predict the likelihood for RLR. The difference from the stop/go decision model is that LD_FL plays a more important role than SPEED in the RLR model.

3.4. Yellow-entry time

Yellow-entry time is an important measure to analyze the RLR tendency at an intersection. At this intersection, the yellow phase length is 4.3 s, and therefore, any entry times larger than this value indicate a RLR violation. Furthermore, the later a vehicle enters the intersection after the red, the more likely the conflicting traffic will be moving through the intersection, and the more likely a RLR crash will occur (Washburn and Courage, 2004). The descriptive statistics of yellow-entry time are presented in Table 3. The ANOVA result (see Table 4) shows that factors including LD_FL ($p < 0.001$) and V_TYPE ($p = 0.037$) are significantly associated with the yellow-entry time, but L_POSITION ($p = 0.215$) is not ($p = 0.204$). It was found that the average entry time for the leading vehicles in traffic flows ($M = 3.8$ s, SD = 0.8 s) is shorter than that for the following vehicles ($M = 4.0$ s, SD = 0.8 s); and the average entry times of light truck vehicles ($M = 4.0$ s, SD = 0.8 s) and passenger cars ($M = 3.9$ s, SD = 0.8 s) are similar, but slightly shorter than that of large-size vehicles ($M = 4.2$ s, SD = 0.7 s). Additionally, the ANOVA result shows that DISTANCE ($p < 0.001$) and SPEED ($p < 0.001$) as two continuous variables are significant covariates associated with the yellow-entry time. Intuitively, yellow-entry

Table 3
Descriptive statistics for yellow-entry time (Y_TIME).

Factor	Sub-level	N	Mean km/h (mph)	Std. deviation km/h (mph)	Minimum km/h (mph)	Maximum km/h (mph)
LD_FL[a]	0	182	6.1 (3.8)	1.3 (0.8)	3.4 (2.1)	9.3 (5.8)
	1	419	6.4 (4.0)	1.3 (0.8)	3.9 (2.4)	11.6 (7.2)
L_POSITION[b]	0	209	6.4 (4.0)	1.3 (0.8)	3.4 (2.1)	11.3 (7)
	1	219	6.4 (4.0)	1.3 (0.8)	3.9 (2.4)	11.6 (7.2)
	2	173	6.3 (3.9)	1.1 (0.7)	4.0 (2.5)	9.7 (6)
V_TYPE[c]	0	325	6.3 (3.9)	1.3 (0.8)	3.4 (2.1)	11.6 (7.2)
	1	260	6.4 (4.0)	1.3 (0.8)	3.9 (2.4)	9.7 (6)
	2	16	6.8 (4.2)	1.1 (0.7)	4.0 (2.5)	8.9 (5.5)
Total		601	6.4 (4.0)	1.3 (0.8)	3.4 (2.1)	11.6 (7.2)

[a] Whether the vehicle was in a leading position or a following position in the traffic flow.
[b] Vehicle's lane position (left lane, middle lane, right lane).
[c] Vehicle type [passenger car (PC), light truck vehicle (LTV), larger size vehicle (LSV)].

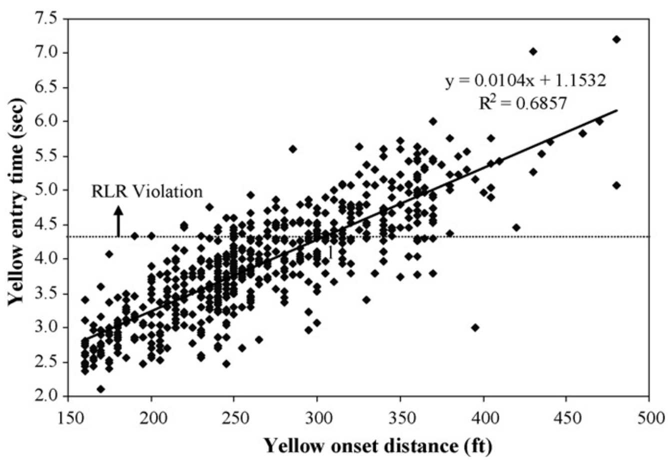

Fig. 8. Relationship between yellow-entry time and vehicle's yellow-onset distance.

time is positively related to the yellow-onset distance. Fig. 8 illustrates a relationship between yellow-entry time and, vehicle's yellow-onset distance. However, yellow-entry time is negatively related to the operating speed. Given a same yellow-onset distance from the intersection, the vehicles traveling at higher speeds would result in shorter yellow-entry times than those traveling at lower speeds.

4. Conclusions and discussions

Using a video-based system, drivers' behavior associated with the signal change was observed at a high-speed signalized intersection. The data collection included drivers' stop/go decisions, vehicles' yellow-onset speeds and distances from the intersection, lane position in the highway, position in the traffic flow, and vehicle type in relation to the yellow signal change. The data analysis indicated that a vehicle's yellow-onset distance, operating speed, and position in the traffic flow are the most important predictors for both the stop/go decision and RLR violations.

Drivers tend to cross the intersection as the yellow-onset distances decrease. However, the yellow-onset distance is not linearly correlated with the probability of RLR. The reason is that only a few aggressive drivers intentionally run red lights at yellow-onset distances too far away from the intersection. Table 1 shows the speeds and standard deviations for the different variables. Drivers are more likely to run red lights at distances where it seems possible to beat the signal change. If the vehicles' yellow-onset distances are between 267.5 ft and 372.5 ft, the operating speed plays an important role in the stop/go decision and RLR. More than 50% of speeding drivers would make go decisions at these yellow-onset distances. Further, more than 50% of speeding drivers would run red lights at the yellow-onset distance interval between 292.5 ft and 372.5 ft, and these speeding red-light runners constitute a substantial portion of the overall RLR violations. Therefore, effectively lowering drivers' speeds at the approach to the intersection may lead to significantly less RLR violations. Moreover, installing green-extension loop detectors within the distance interval may contribute to reducing the chance that drivers encounter the yellow onset in this region, thus resulting in a lower RLR rate.

Drivers that ran the red light were mostly located between the distances of 61.3 m (201 ft) to 128.0 m (420 ft) (90% of the red-light runners) as shown in Fig. 3. This indicates that an effective counter measure located in this area would have a higher probability of reducing red-light running. Also from Fig. 3 we can observe that all the drivers beyond about 360 ft who did not stop ran the red light at the intersection.

An interesting finding of this study is that under the same conditions, the drivers who are in the following positions in the traffic flow are more likely to make go decisions and run the red light than those who are in the leading positions. In this case, if the leading driver in the traffic flow makes a conservative stop decision at the onset of yellow, the following driver's rear-ending crash risk may greatly increase. Furthermore, the analysis indicated that the following drivers with go decisions tend to have longer yellow-entry time than the leading drivers with go decisions. A intersections with higher traffic volumes, drivers would have more chances to be in a following position in the traffic flow at the onset of yellow, thus the RLR rate and crash risk would increase.

The study shows a different way to model red-light running behavior and the stop/go decision Although the logit model is acceptable to use when the dependent variable (stop/go) is binary. However, it was found from the analysis that drivers' stop/go decisions may be dependent of vehicle's yellow-onset distance, speed, and other traffic parameters and logistic regression assumes that a link function can be used to relate the probabilities of group membership to a linear function of the predictor variables. Classification trees offer a simple way to model driver behavior without making any assumptions of normality.

Finally, this study offers a new approach to rank and test the importance of different variables and how are they associated with the light running phenomena through using decision classification tree models to analyze the data. The study generated two models: the first one illustrated the conditions that were associated with the stop/go decision and the other explained the conditions associated with e red-light running violations.

The results are based on data collected at a four legged intersection with speed limit of 45 mph and thus the results should not be generalized over a wider range of intersection types or speed limits without considering the effect of different conditions on the result. However, the tree-based model offers good verbal explanations which would make examining other conditions easier.

Acknowledgments

The authors would like to acknowledge the Florida Department of Transportation (FDOT) for its sponsorship of this research project. The recommendations of this study are those of the authors and they do not represent views of FDOT.

References

Abdel-Aty, M., Lee, C., Wang, X., Nawathe, P., Keller, J., Kowdla, S., Prasad, H., 2005. Identification of intersections' crash profiles/patterns. FDOT Final Report.

Baugley, C.J., 1988. Running the red at signals on high-speed roads. Traffic Engineering Control 29, 415–420.

Bonneson, J.A., Fitts, J.W., 1995. Traffic data collection using video-based systems. Transportation Research Record 1477, 31–40.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Chapman and Hall.

Chang, M.-S., Messer, C.J., Santiago, A.J., 1985. Timing traffic signal change intervals based on driver behavior. Transportation Research Record: Journal of the Transportation Research Board 1027, 20–30.

Köll, H., Bader, M., Axhausen, K.W., 2003. Driver behavior during flashing green before amber: a comparative study. Accident Analysis & Prevention 36, 273–280.

Mahalel, D., Prashker, J.N., 1987. A behavioral approach to risk estimation of rear-end collisions at signalized intersections. Transportation Research Record 1114, 96–102.

National Highway Traffic Safety Administration (NHTSA), 2006. Traffic Safety Facts 2005. <http://www-nrd.nhtsa.dot.gov/pdf/nrd-30/NCSA/TSFAnn/TSF2005EE.pdf> (accessed February 2007).

Newton, C., Mussa, N.R., Sadalla, K.E., Burns, K.E., Matthias, J., 1997. Evaluation of an alternative traffic light change anticipation system. Accident Analysis & Prevention 29, 201–209.

Papaioannou, P., 2007. Driver behavior. Dilemma zone and safety effects at urban signalized intersections in Greece. Accident Analysis & Prevention 39, 147–158.

Parsonson, P.S., 1978. Signalization of high speed. Isolated intersection. Transportation Research Record 681, 34–42.

Porter, B.E., England, K.J., 2000. Predicting red-light running behavior: a traffic safety study in three urban settings. Journal of Safety Research 31 (Spring 1), 1–8.
Retting, R.A., Chapline, J.F., Williams, A.F., 2002. Changes in crash risk following re-timing of traffic signal change intervals. Accident Analysis & Prevention 34, 215–220.
Sheffi, Y., Mahmassani, H., 1981. A model of driver behavior at high speed signalized intersections. Transportation Science 15, 50–61.
Srinivasa, S.R., Carroll, M.J., Hassan, C., 2005. Performance of advance warning for end of green system for high-speed signalized intersections. Transportation Research Record 1925, 176–184.
Strong, C., Lowry, S., McCarthy, P., 2003. Collecting vehicle-speed data by using time-lapse video recording equipment. Transportation Research Record 1855, 97–104.
Traffic Engineering Manual, Chapter 3 Section 3.6, 2005. Florida Department of Transportation. <http://www.dot.state.fl.us/TrafficOperations/Operations/Studies/TEM/TEM.shtm>.
Washburn, S.S., Courage, K.G., 2004. Investigation of Red Light Running Factors. Southeast Transportation Center, the University of Tennessee, Knoxville, Tennessee.
Yan, X., Radwan, E., Guo, D., 2007. Effect of the pavement-marking countermeasure to improve signalized-intersection safety. ITE Journal., in press.
Zimmerman, K., Bonneson, J.A., 2004. Intersection safety at high-speed signalized intersections: number of vehicles in the dilemma zone as a potential measure. Transportation Research Record 1897, 126–133.