Research Article

Research on Driver Behavior in Yellow Interval at Signalized Intersections

Zhaosheng Yang,1,2,3 Xiujuan Tian,2 Wei Wang,1,2,3 Xiyang Zhou,2 and Hongmei Liang2

1 State Key Laboratory of Automobile Dynamic Simulation, Jilin University, Changchun 130022, China
2 College of Transportation, Jilin University, Changchun 130022, China
3 Jilin Province Key Laboratory of Road Traffic, College of Transportation, Jilin University, Changchun 130022, China

Correspondence should be addressed to Wei Wang; wwei@jlu.edu.cn

Received 11 June 2014; Accepted 5 August 2014; Published 26 August 2014

Academic Editor: Huimin Niu

Copyright © 2014 Zhaosheng Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Vehicles are often caught in dilemma zone when they approach signalized intersections in yellow interval. The existence of dilemma zone which is significantly influenced by driver behavior seriously affects the efficiency and safety of intersections. This paper proposes the driver behavior models in yellow interval by logistic regression and fuzzy decision tree modeling, respectively, based on camera image data. Vehicle’s speed and distance to stop line are considered in logistic regression model, which also brings in a dummy variable to describe installation of countdown timer display. Fuzzy decision tree model is generated by FID3 algorithm whose heuristic information is fuzzy information entropy based on membership functions. This paper concludes that fuzzy decision tree is more accurate to describe driver behavior at signalized intersection than logistic regression model.

1. Introduction

Yellow interval plays a significant role in the operation and security of signalized intersections. Traffic accidents in yellow interval account for more than half of the whole traffic accidents at signalized intersections according to statistics. When vehicles approach intersection in the initiation of yellow, drivers need to make a decision to stop or cross through according to the state of vehicle, distance to stop line, vehicle speed, road condition, and other information. The process is a complexly fuzzy and uncertain decision-making process belonging to uncertain decision problems.

In recent years, increasing attention has been given to driver behaviors and decision-making process research in yellow interval at signalized intersections. Elmitiny et al. [1] used classification tree models to analyze the probabilities of stopping or crossing and red-light running associated with the observed traffic data and concluded that the distance from the intersection at the onset of yellow, operating speed, and position in the traffic flow were the most important predictors for both the stopping/crossing decision and red-light running violation. Koll et al. [2] proposed a parking discrete model, emphasizing the impact of speed, distance, and time on decision. Through researching, they found stopping probability was a function of the time a driver expected he would take to stop line. Papaioannou [3] divided drivers into conservative, ordinary, and adventurous three categories and built a binary choice model which calculated the probability of stopping and crossing considering vehicle speed, drivers’ age, and gender as well as dilemma zone. Elmitiny et al. [4] studied driver behavior by decision-making tree modeling and made correlation analysis with traffic flow parameters based on camera data including drivers’ decision, lane position, vehicle type, speed, and weather through a red light. Moore and Hurwitz [5] proposed fuzzy logic as a tool for modeling driver behavior in the dilemma zone and three models related to the speed and position of the vehicles were developed which could predict driver behavior with a very high accuracy. Chiou and Chang [6] researched the effects of green and red vehicle signal countdown display on drivers’ responses and on safety and efficiency aspects which included late-stopping ratio, start-up delay, and discharge headway.
Long et al. [7] investigated the impact of countdown timer on driver behavior after the onset of yellow at signalized intersections in China. Binary logistical regression analysis was introduced to compare the difference of driver's stopping or crossing decisions at intersections with and without countdown timer display. Gates and Noyce [8] researched the influences of vehicle type on various aspects of dilemma zone driver behavior, including brake response time, deceleration rate, and red light running occurrence. Hurwitz et al. [9] characterized driver behavior to understand how and where drivers made their decision to stop or proceed approaching a signal and comprehension related to type 2 dilemma zones for the purpose of defining these boundary conditions.

This paper analyzes the problem of dilemma zone and researches driver behavior in yellow interval at signalized intersections based on camera image data. Logistic regression and fuzzy decision tree are applied to describe driver behavior in yellow interval, respectively, based on the distance to stop line, vehicle speed, and installation of countdown timer. The results show that fuzzy decision tree model is more accurate to predict drivers’ decision in yellow interval at signalized intersections. With the knowledge of driver behavior in yellow interval, traffic engineers can optimize the signal timing and decrease the influence of dilemma zone to improve the efficiency and operations of intersections.

2. Driver Behavior in Yellow Interval Modeling

The research of driver behavior in yellow interval has significant effect on preventing dilemma zone, increasing the safety and optimizing the signal timing of intersections. This paper utilizes logistic regression and fuzzy decision tree methods considering the characteristic of driver behaviors and compares the prediction precision of the two methods.

2.1. Problem Description. When vehicles approach the signalized intersections in yellow interval, they are often easily caught in an area where vehicles can neither stop safely nor cross through smoothly which is defined as “dilemma zone.” The notion of dilemma zone was first referenced by Gazis et al. [10] in their paper in 1960. The zone is often referred to as type 1 dilemma zone. The reason of this scenario is poor design of intersections mainly associated with inappropriate signal timing and unreasonable detector placement.

The second category of dilemma zone was formally proposed by the Southern Section of ITE [11] in the technical committee report in 1974. Type 2 dilemma zone is often referred to as “indecision zone.” Two types of dilemma zone can be described in Figure 1.

The research of dilemma zone is so important because the existence of dilemma zone negatively influences the safety of the intersection. The size of dilemma zone has a close relationship with vehicle speed, vehicle position, acceleration and deceleration of vehicles, road conditions, drivers' own decisions, and other factors. The choice of drivers has a greater influence and also has certain relevance with other factors. Therefore, researching driver behavior models in yellow interval has great significance for avoiding dilemma zone and safety of the whole signalized intersections.

2.2. Driver Behavior Models Formulation

2.2.1. Data Collection. In good weather conditions, the data about related influencing factors of drivers’ behaviors at the signalized intersection during workday rush hours was collected with video camera. Firstly, two intersections that were similar in traffic flow, signal control program, and geometric condition of roads were chosen in Changchun which locates in northeast China, but one of them was installed with a countdown timer display and the other without. Both of the intersections are on the major arterials and bicycles and pedestrians have little impact on the vehicle flow. Only through movements data was considered in this paper. Signal phases of the intersections are four and yellow interval is 3 s. Secondly, an approach based on video was used to collect driver behavior data in yellow interval. The video camera was placed at a high point contributing to collect related data. Thirdly, reference lines were marked with 5 m clearance in each intersection approach to acquire accurate location and speed of the vehicles [7]. Field observation was made from 7:00 a.m. to 10:00 a.m. on two weekdays at the intersections.

After the collected data being processed, the distance from the vehicles to the stop line, the speed, and drivers’ decision when the signal turned yellow at the signalized intersection were collected. The scatter diagrams of the collected data at signalized intersections without and with countdown timer are shown in Figures 2 and 3, respectively. Drivers’ decision is divided into two categories: cross and stop. It depends on several factors of which the most important are the distance to stop line, vehicle speed, and the installation situation of countdown timer of the intersections.

It is obvious that the distance to stop line is the most significant factor of driver behaviors in yellow interval from the collected data. Drivers will choose to cross the intersection when the distance is less than 10 m, whether there is countdown timer or not. With the increase of distance, the number of vehicles which choose to pass through intersections with countdown timer increases. When the distance
2.2.2. Binary Logistic Regression Model. The previous studies [12–14] have shown that drivers’ choice behaviors are resulting from the joint efforts of various factors during the yellow phase at the signalized intersection, which presents a nonlinear relation. Based on the actual data, this paper builds the drivers’ choice model by means of analyzing statistics, introducing a dummy variable and generating logistic regression model. The dependent variable is driver’s choice behavior at yellow signal’s period which is represented by \( y \), where \( y = 1 \) represents that the vehicle chooses to stop and \( y = 0 \) means that the vehicle is driving into the intersection. Due to the limitations of data collecting means, only three main factors are taken into consideration including \( d \), the distance from the vehicles to the stop line when the yellow signal’s turns on, and \( v \) representing the vehicle speed and \( t \) which means the installation of countdown timer display. The regression model can be represented as follows:

\[
\text{Logit} (P) = \ln \left( \frac{P}{1-P} \right) = \beta_0 + \beta_1 d + \beta_2 v + \beta_3 t, \quad (1)
\]

where \( P \) represents the probability of a driver’s choice of stopping and \( \beta_0 \) represents the constant of the model, where \( d \) represents the distance from the vehicle to the stop line and \( v \) is the vehicle’s speed at the moment of yellow signal’s turning on. Yet \( t \) is a dummy variable [6] of which \( t = 1 \) means there is a countdown timer at the signalized intersection while \( t = 0 \) indicates there is no countdown timer at the intersection. \( \beta_1 \sim \beta_3 \) are the regression coefficients of each factor.

The probability of vehicle stopping at the moment the yellow signal is being on at the signalized intersection can be calculated by

\[
P (y = 1) = \frac{e^{\beta_0 + \beta_1 d + \beta_2 v + \beta_3 t}}{1 + e^{\beta_0 + \beta_1 d + \beta_2 v + \beta_3 t}}. \quad (2)
\]

Similarly, the probability of vehicle crossing through can be calculated as

\[
P (y = 0) = \frac{1}{1 + e^{\beta_0 + \beta_1 d + \beta_2 v + \beta_3 t}}. \quad (3)
\]

The logistic regression analysis of collected data could be observed by employing SPSS 16.0 software. Table 1 shows the determination of the regression coefficients for each factor in

<table>
<thead>
<tr>
<th>Step 1a</th>
<th>( B )</th>
<th>S.E.</th>
<th>Wald</th>
<th>df</th>
<th>Sig.</th>
<th>( \exp(B) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \nu )</td>
<td>-0.050</td>
<td>0.019</td>
<td>6.885</td>
<td>1</td>
<td>0.009</td>
<td>0.952</td>
</tr>
<tr>
<td>( d )</td>
<td>0.151</td>
<td>0.028</td>
<td>29.759</td>
<td>1</td>
<td>0.000</td>
<td>1.163</td>
</tr>
<tr>
<td>( td )</td>
<td>-0.082</td>
<td>0.020</td>
<td>16.361</td>
<td>1</td>
<td>0.000</td>
<td>0.921</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.552</td>
<td>0.445</td>
<td>12.173</td>
<td>1</td>
<td>0.000</td>
<td>0.212</td>
</tr>
</tbody>
</table>

\( \nu \) and \( d \) and \( td \).
(1). It is determined that $\beta_0 = -1.552$, $\beta_1 = 0.151$, and $\beta_3 = -0.082$ with statistical significance ($P < 0.001$). $\beta_2 = -0.050$ is determined with significance $P < 0.05$. These factors all comply with the validity check under the confidence level of 5%, suggesting that they all significantly influence drivers’ decision-making.

Therefore, the binary logistic regression model of driver behavior in yellow interval after calibration can be expressed with the following formula:

$$\text{Logit}(P) = \ln\left(\frac{P}{1-P}\right) = -1.552 + 0.151d - 0.050v - 0.082td.$$  \hspace{1cm} (4)

According to the variables in Table 1, drivers’ decisions to stop behind stop line without and with countdown timer can be represented using the following two functions:

Without countdown timer:

$$\text{Logit}(P) = -1.552 + 0.151d - 0.050v.$$  \hspace{1cm} (5)

With countdown timer:

$$\text{Logit}(P) = -1.552 + 0.069d - 0.050v.$$  \hspace{1cm} (6)

The probability of vehicle stopping in yellow interval at signalized intersections can be calculated by (6). Therefore, the probability of stopping at signalized intersections without and with countdown timer display can be calculated as in (7) and (8), respectively:

Without countdown timer:

$$P(y = 1) = \frac{e^{-1.552+0.151d-0.050v-0.082td}}{1 + e^{-1.552+0.151d-0.050v-0.082td}}.$$  \hspace{1cm} (7)

With countdown timer:

$$P(y = 1) = \frac{e^{-1.552+0.069d-0.050v}}{1 + e^{-1.552+0.069d-0.050v}}.$$  \hspace{1cm} (8)

Vehicles’ stopping probabilities based on vehicle speed and distance to stop line at intersections without and with countdown timer are plotted in Figures 4 and 5. We can conclude that the probability of stopping at intersections without countdown timer is higher than that with countdown timer in yellow interval when the vehicles are at the same distance and speed.

2.2.3. Fuzzy Decision Model. Firstly, we classify the influencing factors according to the principle of fuzzy decision tree and then set the membership functions [5, 15, 16] of each category based on the survey data and subsequently obfuscate the collected data [16]. Therefore, fuzzy rules could be established and finally the decision tree of driver behavior combined with fuzzy information entropy principle could be got.

In fact, the data is randomly divided into two groups of which one is for model generation and the other is for model prediction. First of all, the data is preprocessed and the nonnormal data is excluded. Obfuscating the data, the paper sets up three attributes, namely, installing countdown timer or not and the speed of the vehicle and its distance away from the stop line. The fuzzy sets are characterized by membership functions as in (9). A fuzzy set can be defined as (10):

$$\mu_A(x) : X \rightarrow [0, 1],$$  \hspace{1cm} (9)

$$A = \{(x, \mu_A(x)) \mid x \in X\}.$$  \hspace{1cm} (10)

Membership function is represented by idealized straight line. Based on different purposes, there are usually three kinds of membership functions, that is, triangular membership function, trapezoidal membership function, and Gaussian membership function. The paper adopts the triangle membership function according to the survey results because of its easy implementation and mathematical simplicity.
Mathematical Problems in Engineering

The main idea of fuzzy logic is using a value to represent the truth on the closed interval, where the classical true value is represented by 1 and the classical false value is represented by 0. The varying of truth degrees is indicated by values in (0, 1).

The vehicle speed is divided into three categories: low speed, medium speed, and high speed in the development process. Fuzzy variable is uncertain which means it is not a fixed number to describe low, medium, and high but an interval. The concerned speed domain is 55 km/h and it is divided into 11 subsets of which each interval is 5 km/h. The interval defined to the concerned distance domain is 45 m and is divided into 9 subsets of which each interval is 5 m. The interval defined to describe close distance is (0, 20), (20, 35) for medium distance, and (20, 35) for far distance.

The fuzzy subsets and memberships of vehicle speed and distance to stop line at the moment yellow light starts are listed in Figures 6 and 7, respectively, formulating the triangular membership function.

Drivers’ decision can be divided into two cases that stop behind stop line and cross through intersection which can be calculated by

\[ F(E) = \frac{M(C) \cap D_i}{M(C)} \cdot \log \left( \frac{M(C) \cap D_i}{M(C)} \right) \]  \hspace{1cm} (12)

We use \( M(C \cap T_j) \) to represent the sum of membership degrees where the value of attribute \( B^j \) is \( T_j \) and \( M(C^j) \) indicates the base of fuzzy subset \( C_j \) which is the sum of all membership degrees. The steps of building decision tree using the theory of fuzzy information entropy are as follows.

**Step 1.** According to the collected data, calculate the fuzzy information entropies of each attribute and fuzzy subset. The fuzzy information entropy computation results of fuzzy subset \( C_j \) are shown in Table 5.

---

**Table 2:** Fuzzy subsets and membership function for vehicle-speed.

<table>
<thead>
<tr>
<th>Fuzzy subsets</th>
<th>Membership function</th>
</tr>
</thead>
</table>
| Low speed     | \( f_v = \begin{cases} 
1.0 & (v \leq 10) \\
1.5 - \frac{1}{20}v & (10 < v \leq 30) \\
0 & (v > 30) \end{cases} \) |
| Medium speed  | \( f_v = \begin{cases} 
0 & (v \leq 10) \\
-0.67 + \frac{1}{15}v & (10 < v \leq 25) \\
2.67 - \frac{1}{15}v & (25 < v \leq 40) \\
0 & (v > 40) \end{cases} \) |
| High speed    | \( f_v = \begin{cases} 
0 & (v \leq 30) \\
-2 + \frac{1}{15}v & (30 < v \leq 45) \\
1 & (v > 45) \end{cases} \) |

**Table 3:** Fuzzy subsets and membership function for vehicle-distance.

<table>
<thead>
<tr>
<th>Fuzzy subsets</th>
<th>Membership function</th>
</tr>
</thead>
</table>
| Close distance| \( f_d = \begin{cases} 
1.0 & (d \leq 5) \\
1.33 - \frac{1}{15}d & (5 < d \leq 20) \\
0 & (d > 20) \end{cases} \) |
| Medium distance| \( f_d = \begin{cases} 
0 & (d \leq 5) \\
-0.5 + \frac{1}{10}d & (5 < d \leq 15) \\
2.5 - \frac{1}{10}d & (15 < d \leq 25) \\
0 & (d > 25) \end{cases} \) |
| Far distance  | \( f_d = \begin{cases} 
0 & (d \leq 20) \\
-1.33 + \frac{1}{15}d & (20 < d \leq 35) \\
1 & (d > 35) \end{cases} \) |

---

The fuzzy subsets and memberships of which each interval is 5 m. The interval defined to describe close distance is (0, 20), (5, 25) for medium distance, and (20, 35) for far distance.

Table 4. According to the collected data using membership functions, partly shown in Table 5.

\[
FE(C, B^j) = \sum_{j=1}^{n} \frac{M(C \cap T_j)}{M(C)} \cdot \log \left( \frac{M(C \cap T_j)}{M(C)} \right)
\]  \hspace{1cm} (11)

\[
FE(C, B^j) = \sum_{j=1}^{n} \frac{M(C \cap T_j)}{M(C)} \cdot \log \left( \frac{M(C \cap T_j)}{M(C)} \right)
\]  \hspace{1cm} (11)

The varying of truth degrees is indicated by values in (0, 1).

The fuzzy subsets and memberships of which each interval is 5 m. The interval defined to describe close distance is (0, 20), (5, 25) for medium distance, and (20, 35) for far distance.

Table 4. According to the collected data using membership functions, partly shown in Table 5.
Then the fuzzy information entropy of every attribute related to the data set can be calculated. The results are as follows:

\[
\begin{align*}
    &\text{FE}(C, B^1) = 0.8680, \\
    &\text{FE}(C, B^2) = 0.8905, \\
    &\text{FE}(C, B^3) = 0.8250.
\end{align*}
\]  

(13)

**Step 2.** Choose the attribute, the distance to stop line \(B_3\) which has the minimum fuzzy information entropy as the root node of the whole decision tree.

**Step 3.** Select the element whose membership degree is greater than the threshold value \(\alpha\) decided by membership functions to constitute a new fuzzy subclass set which generates a child node of the current node from the corresponding fuzzy set of the current attribute description for each node.

**Step 4.** The fuzzy decision tree is complete if all the elements in their fuzzy subclass sets possess the same class attribute or all the nodes from the root node to the current branch are set to the leaf nodes for each child node of the current node. Otherwise, choose the attribute unused which possesses the minimal information entropy as the child node and set the child node as the current node and return to Step 3.

We can get the fuzzy decision tree based on the above steps in Figure 8. From Figure 8, part of the fuzzy rules can be easily concluded as follows.

**Rule 1.** If \(B_3\) attaches to \(M_1(\mu_{12} > \alpha)\) and \(D\) attaches to \(D_1\), the vehicle will cross the intersection.

**Rule 2.** If \(B_3\) attaches to \(M_2(\mu_{22} > \alpha)\) and \(B^1\) attaches to \(K_2(\mu_{12} = 1)\), \(D\) will attach to \(D_1\) which means that the vehicle will cross the intersection.
Table 6: Prediction of logistic regression model.

<table>
<thead>
<tr>
<th>Observed</th>
<th>Predicted</th>
<th>Percentage correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>125</td>
<td>0 (cross) 12</td>
</tr>
<tr>
<td>1</td>
<td>27</td>
<td>1 (stop) 36</td>
</tr>
<tr>
<td>Overall percentage</td>
<td>80.5</td>
<td></td>
</tr>
</tbody>
</table>

Table 7: Prediction of fuzzy decision tree model.

<table>
<thead>
<tr>
<th>Observed</th>
<th>Predicted</th>
<th>Percentage correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>122</td>
<td>0 (cross) 15</td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>1 (stop) 43</td>
</tr>
<tr>
<td>Overall percentage</td>
<td>82.5</td>
<td></td>
</tr>
</tbody>
</table>

Rule 3. If $B^3$ attaches to $M_2(\mu_{22} > \alpha)$, $B^1$ attaches to $K_1(\mu_{11} = 1)$, and $B^2$ attaches to $T_1(\mu_{11} > \alpha)$, $D$ will attach to $D_2$ which means that the vehicle will stop behind stop line at the intersection.

3. Comparison of Two Models

To summarize, the prediction accuracy of the two models above is compared. The prediction result of logistic regression model is shown in Table 6 and fuzzy decision tree model is shown in Table 7 in view of the test sample.

From the two result tables of two models, it could be easily concluded that the prediction accuracy of the fuzzy decision tree model is two percent higher than that of the logistic regression model which suggests the fuzzy decision tree model can better reflect driver’s decision-making behavior during the yellow interval on the whole. However, the logistic regression model is better in predicting vehicles’ crossing through behavior at the intersections than fuzzy decision tree model which is more accurate in predicting drivers’ stopping decision. Both of the two models perform badly in predicting stopping behavior of vehicles which fully suggests that stopping behavior is more discrete.

4. Conclusions

This paper builds the drivers’ choice behavior models by using logistic regression and fuzzy decision method during the yellow interval based on the collected data. A dummy variable is introduced as well as two basic factors, namely, the distance of vehicles’ position to the stop line and the speed taken into consideration to build the logistic regression model. The probability of vehicles’ stopping at the intersection without countdown timer display is greater than that of the intersection with countdown timer display. By building the decision tree model, it is found that distance from the vehicle to the stop line has the greatest influence on drivers’ decision-making. Besides, the prediction accuracy of fuzzy decision tree model is higher than that of logistic regression model after comparing two models. The study on drivers’ behavior model enables engineers’ better understanding about drivers’ decision in yellow interval so as to make a better optimization in the signal timing of signalized intersection and take measures to reduce the influence of dilemma zone.

In this paper, only leading vehicles approaching the intersections in yellow phase are researched in order to get rid of the influence of car following and other interference from other vehicles. Because of the limitations, it is advisable to take car following behavior and more factors into consideration with more field data collection efforts in the future research.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research has been jointly supported by National High Technology Research and Development Plan Project (Grant no. 2014BAG03B03), China Postdoctoral Science Foundation Special Funding (Grant nos. 2012T50300 and 2013T60331), and Science and Technology Development Projects of Jilin Province, China (Grant no. 20140520134JH).
References


