A Machine Learning Approach for Highway Intersection Risk Caused by Harmful Lane-Changing Behaviors

Yanping Hao1*; Liangjie Xu2; Bozhao Qi3*; Teng Wang4; and Wei Zhao5

1Dept. of Traffic Engineering, School of Transportation, Wuhan Univ. of Technology, Wuhan 430063, China; State of Key Laboratory of Vehicle of VNH and Safety Technology, Chongqing, China. E-mail: hao_try05@163.com
2Dept. of Traffic Engineering, School of Transportation, Wuhan Univ. of Technology, Wuhan 430063, China. E-mail: laurrie119@163.com
3Dept. of Electrical and Computer Engineering, Univ. of Wisconsin–Madison, Madison, WI 53705. E-mail: bqi2@wisc.edu
4Texas A&M Transportation Institute, San Antonio, TX. E-mail: Teng-Wang@tti.tamu.edu
5Dept. of Traffic Engineering, School of Transportation, Wuhan Univ. of Technology, Wuhan 430063, China. E-mail: zw2253782@hotmail.com
*Co-primary authors

Abstract
Highway intersection-related crashes are suspected to be associated with harmful lane-changing behaviors. To better understand the relationship between them, this study applied an innovative machine learning approach to identify crash risk factors and find solutions to reduce the intersection-related crash frequency and severity caused by harmful lane-changing behaviors. First, a vehicles approach time (VAT) model was developed to define and classify different types of harmful lane-changing behaviors. Second, the real world driving video data was collected and preprocessed to identify the potential crash risk factors of harmful lane-changing behaviors. Finally, an advanced machine learning algorithm, Lasso-LARS, was applied to analyze the relation between intersection-related crash risk factors and lane-changing behaviors. There were no significant differences in the VAT values between the VAT model and the Lasso-LARS regression model. The result shows that both the two models are suitable for the analysis of risk factors of harmful lane-changing behaviors.

INTRODUCTION
Highway intersection-related crashes are suspected to be associated with harmful lane-changing behaviors. Previous research results (Lee, S. E., et al. 2004)
showed that crashes caused by lane-changing behaviors accounted for 4% to 10% of all crashes. This paper will study the crash risk factors due to harmful lane-changing behaviors at signalized intersections. To better understand the relationship between harmful lane-changing behaviors and intersection-related crashes, this study applied an innovative machine learning approach to identify crash risk factors and mitigate the intersection-related crashes caused by harmful lane-changing behaviors. With the rapid development of autonomous vehicle technology (Kang, L. et al. 2018; Qi, B., et al. 2016), this study could provide theoretical support for the identification and collision avoidance of harmful lane-changing behaviors for self-driving vehicles when sharing the road with the conventional vehicles.


In this paper, a Vehicle Approach Time (VAT) model of risk characteristic parameters was developed. The driving video data was trained by the Lasso-LARS algorithm to determine the correlation coefficient. The accuracy of the VAT model was validated by a case study. The correlation of the paired T-test with 95% confidence showed that the correlation reached 99.7%, and indicated the validity of the research results.

**DEFINITION OF HARMFUL LANE-CHANGING BEHAVIORS**

Violation of traffic regulations has a great impact on both of traffic operation and safety. In this research, the allowed and prohibited lane-changing behaviors were defined in Figure 1. Therefore, the harmful lane-changing behaviors defined in this paper are divided into two categories according to the location of harmful lane-changing behaviors: harmful lane-changing happened in the solid-line segment and harmful lane-changing in the non-solid-line segment.
Each type of harmful lane-changing behaviors is defined according to the vehicle's trajectory when the harmful lane-changing behavior takes place. The harmful lane-changing behaviors at the solid-line segment are divided into three types: forced lane-changing, random insertion lane-changing, unsuccessful lane-changing. The harmful lane-changing behaviors at non-solid-line segment are divided into other three types: forced lane-changing, random insertion lane-changing, and continuous lane-changing.

![Figure 1. Allowed and prohibited lane-changing behaviors and range.](image)

**MODELING OF CRASH RISK FACTORS**

**Data Collection**

This study selected two signalized intersections of Youyi Avenue in Qingshan District, Wuhan, China. The survey periods included the peak hours of morning and afternoon on March 1st, 2018 (7:30AM-9:00AM and 5:00PM-6:30PM) and a representative regular hour (9:00AM-10:30AM and 2:30PM-5:00PM). The survey data statistics are shown in Table 1.

<table>
<thead>
<tr>
<th>Intersection</th>
<th>Direction of entry</th>
<th>Traffic parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Left</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Q</td>
</tr>
<tr>
<td>No. 1</td>
<td>South</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>No. 2</td>
<td>East</td>
<td>122</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
</tr>
</tbody>
</table>

Note: No. 1 intersection indicates Youyi Avenue - Gongye Road; No. 2 intersection indicates Youyi Avenue - Jiansheyi Road. Q indicates traffic volume, pcu (Passenger Car Unit); R indicates the number of harmful lane-changing behaviors, pcu.
Data Preprocessing

The collected vehicle data of harmful lane-changing behaviors is analyzed and summarized in Table 2.

Table 2. Traffic Parameter Statistics of Harmful Lane-Changing Behaviors

<table>
<thead>
<tr>
<th>Number</th>
<th>Harmful lane-changing area</th>
<th>Harmful lane-changing behaviors type</th>
<th>Steering angle (°)</th>
<th>Speed of the front vehicle (Km/h)</th>
<th>Speed of the rear vehicle (Km/h)</th>
<th>Speed difference (Km/h)</th>
<th>Headway (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Solid-line segment</td>
<td>A-3 16-20</td>
<td>.36</td>
<td>19.51-25</td>
<td>21.43-26</td>
<td>3.06-5.</td>
<td>2.01-3.</td>
</tr>
<tr>
<td>A</td>
<td>Solid-line segment</td>
<td>B-3 24-36</td>
<td>.19</td>
<td>25.71-33</td>
<td>28.91-34</td>
<td>3.16-6.</td>
<td>2.31-2.</td>
</tr>
</tbody>
</table>

As shown in Table 2, the traffic parameters and crash risk factors were collected when the harmful lane-changing behavior occurred. There are correlations between different factors. Therefore, a Vehicle Approach Time (VAT) model was considered and developed to analyze the crash risk factors.

The operational status of the road traffic is constantly changing due to the influence of driver, vehicle, road and environment. The driver human factors were not incorporated into this study due to limitations of the investigation method. In this research, the influencing factors included the time when the harmful lane-changing behavior occurred, the model of the lane-changing vehicle, the steering angle, the duration of the lane-changing, the speed of the target vehicle (before change and after change), and the headway between the front and rear vehicles.

Vehicle Approach Time (VAT) Model

This study defined VAT as the vehicle approach time of the target vehicle \((i)\) with the preceding vehicle \((i+1)\) or the following vehicle \((i-1)\) which indicates time between the traffic conflict occurred moment and the possible collision moment due to
harmful lane-changing behaviors at a signalized intersection. When the target vehicle is changing lane, two vehicles will collide if the vehicle in the merging lane does not slow down or speed up, as shown in Figure 2.

![Vehicle Approach Time (VAT) model diagram](image)

**Figure 2. Vehicle Approach Time (VAT) model diagram**

The traditional TTC model is suitable for the following vehicle collision time calculation. Based on the foundation of TTC model, the proposed VAT model added the steering angle variable to evaluate the crash risk factors of the harmful lane-changing behaviors. The VAT model equation is shown below:

$$ VAT_i = \frac{|x_{i-1}(t) - x_i(t)| - l_i}{v_{i-1} - \cos \theta_i \times v_i} \quad \text{or} \quad VAT_i = \frac{|x_i(t) - x_{i+1}(t)| - l_{i+1}}{\cos \theta_i \times v_i - v_{i+1}} $$

(1)

Where:
- $VAT_i$ — the vehicle approaching time between two vehicles;
- $x_i(t)$ — the position of the $i^{th}$ vehicle on the road (m);
- $v_i(t)$ — the time mean speed of the $i^{th}$ vehicle from conflict occurred moment to the possible collision moment (m/s);
- $\theta_i$ — the steering angle of the $i^{th}$ vehicle (target vehicle) (degree);
- $l_i$ — the length of the preceding vehicle (m).

The relationship between headway and spacing is:

$$ d_i = x_i(t) - x_{i-1}(t) = v_{i-1} \cdot h_{i-1} \quad \text{or} \quad d_i = x_{i+1}(t) - x_i(t) = v_i \cdot h_i $$

(2)

Where:
- $d_i$ — spacing (m);
- $h_i$ — headway (t).

Then, the equation (1) can be expressed as:
\[ VAT_i = \frac{v_{i-1} \cdot h_{i-1} - l_i}{|v_{i-1} - \cos \theta \times v_i|} \] or \[ VAT_i = \frac{v_i \cdot h_i - l_{i+1}}{|\cos \theta \times v_i - v_{i+1}|} \] (3)

Through the onsite traffic observation and in-house video investigation, it is noticed that when the target vehicle started lane-changing behavior, other vehicles in the target (merging) lane would brake. Hence the driver reaction time should be considered. If the calculated VAT value is less than the driver reaction time (as shown in Case A of Figure 2), the VAT can be directly calculated by equation (3). If the calculated VAT value is greater than the driver reaction time (as shown in Case B of Figure 2), the VAT needs to be calculated according to the headway and speed of following vehicle after applying brake. In this calculation, the speeds and accelerations of the front and rear vehicles should be considered. Therefore, the time mean speed is used in the model instead of the instantaneous speed to improve the accuracy of the model. Based on the calculated VAT value, the severity of different harmful lane-changing behaviors can be quantified for analysis. A smaller VAT value means the greater probability of collision which could lead to higher risk and severer crashes.

CRASH RISK FACTORS OF HARMFUL LANE-CHANGING BEHAVIORS BASED ON LASSO-LARS ALGORITHM

Variable Standardization

In this study, the data was trained and analyzed by using the Lasso-LARS (Least Angle Regression) algorithm to determine the correlation coefficient of each factors of harmful lane-changing behaviors. After preprocessing, the model input data contains 10 independent variables and 1 dependent variable (Table 3). The dependent variable is the VAT (numeric variable) which is calculated using the collision time model and represented by the vector \( Y \).

| Table 3. Description of the Independent Variables |
|---------------------------------|----------|-----------|-----------|
| Category                        | Independent variable | Types   | Value     | Symbol |
| Vehicle Information             | Model     | C*       | 1-5       | X₁      |
|                                | Movement  | C*       | 1-3       | X₂      |
|                                | Duration  | N*       | Numerical value | X₃ |
|                                | Change time  | C*       | 1-6       | X₄      |
|                                | Distance from parking line | N* | Numerical value | X₅ |
|                                | Speed (before change) | N* | Numerical value | X₆ |
| Environment Information        | Headway from the front of the vehicle (before change) | N* | Numerical value | X₇ |
|                                | Headway from the vehicle | N* | Numerical value | X₈ |

© ASCE
Lasso-LARS Algorithm

Considering there are correlations between crash risk factors in the event of harmful lane-changing behaviors, the Lasso-LARS algorithm can be used to solve the high dimensional feature problems with correlation of different factors (Zhao, W., et al. 2018) as shown in Equation 4 below.

\[ Y = X \theta \]  

(4)

Where: \( Y \) is the vector of \( m \times 1 \), \( X \) is the matrix of \( m \times n \), \( \theta \) is the vector of \( n \times 1 \). \( m \) is the number of samples and \( n \) is the feature dimension.

Consider matrix \( X \) as \( n \) vectors of \( m \times 1 \), which is \( X_i (i = 1,2,\ldots,m) \), and in the \( X \) variable \( X_i (i = 1,2,\ldots,m) \) of \( Y \), select the variable \( X_k \) which is the closest to the target \( Y \) (when the cosine distance is the largest), and use \( X_k \) to approximate \( Y \), then:

\[ \bar{Y} = X_k \theta_k \]  

(5)

Where:

\[ \theta_k = \frac{X_k^T Y}{\|X_k\|_2} \]  

(6)

Where, \( \bar{Y} \) is the projection of \( Y \) on \( X_k \). Defined residuals: \( Y_{yes} = Y - \bar{Y} \), \( Y_{yes} \) and \( X_k \) are orthogonal. Taking \( Y_{yes} \) as the new dependent variable. \( X_i \{i = 1,2,\ldots,k-1,k+1,\ldots,n\} \) is the new set of independent variables after \( X_k \) is removed. Repeat the projection and take the residual as the new dependent variable for the model. Stop the iteration until the residual is 0 or running out of all the independent variables. The argument matrix is defined as:

<table>
<thead>
<tr>
<th>Feature</th>
<th>Type</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed of the front vehicle (before change)</td>
<td>N*</td>
<td>Numerical value</td>
</tr>
<tr>
<td>Speed of the vehicle rear (before change)</td>
<td>N*</td>
<td>Numerical value</td>
</tr>
</tbody>
</table>

Notes: C* is Categorical variable; N* is Numerical variable.
\[
X = \begin{pmatrix}
  x_{i1} x_{i2} x_{i3} \cdots x_{ij} \\
  x_{21} x_{22} x_{23} \cdots x_{2j} \\
  x_{31} x_{32} x_{33} \cdots x_{3j} \\
  \vdots \\
  x_{11} x_{12} x_{13} \cdots x_{1j}
\end{pmatrix} \quad (i = 1, 2, \ldots, 270, \ j = 1, 2, \ldots, 10)
\]  

(7)

Where, \( Y \) is the set of corresponding dependent variables. The formula for solving the Lasso-LARS problem is shown in Equation 8.

\[
Y_j = \theta_{i1}x_{i1} + \theta_{2}x_{i2} + \cdots + \theta_j x_{ij} (i = 1, 2, \ldots, 270, \ j = 1, 2, \ldots, 10)
\]

(8)

Where, \( \theta_j \) is the coefficient

**Analysis of Training Results**

Next, the independent variable vector \( X \) from Table 3 and the dependent variable \( Y \) (VAT) were substituted into the Lasso-LARS algorithm. The model training result and the correlation coefficient matrix of the independent variables and the dependent variable were calculated using R package. The results are shown below in Figure 3.

![Figure 3. Model training results](image)

The following observations are taken from the training results in Figure 3: 1). From left to right, each polyline represents the change of the coefficient of each variable. 2). Vertical lines with numbers indicate the steps of adding variable. 3). Some of the polylines have a coefficient of 0 on the left, and the coefficient start to change when a certain vertical line is reached, indicating that the variable is added. 4). Different variables have different effects on the VAT of harmful lane-changing behaviors (Zhao, W., et al. 2015). The matrix of correlation coefficients of each variable are as follows:
\[ \theta^T = \begin{bmatrix} 0 & 5.569663 \times 10^{-3} & 6.983032 \times 10^{-3} & 1.948074 \times 10^{-1} \\ 8.637113 \times 10^{-1} & 2.103403 \times 10^{-1} & 3.903212 \times 10^{-1} \\ -9.738479 \times 10^{-2} & 7.713337 \times 10^{-1} & 8.381305 \times 10^{-2} \end{bmatrix} \]

According to the training results and the correlation coefficient matrix, the correlation coefficients of different variables are different. The VAT of each lane-changing behavior could be calculated using the correlation coefficient matrix and Equation 9.

**CASE STUDY VERIFICATION**

**Data of Youyi Avenue and Jiansheyi Road Intersection**

The frequency of harmful lane-changing behaviors at signalized intersections is significantly different for various VAT ranges. Total 20 signal lights circles video logs are used for analysis. The preliminary results are shown in Table 4 and Figure 4.

<table>
<thead>
<tr>
<th>VAT Ranges (s)</th>
<th>0.0</th>
<th>0.5</th>
<th>1.0</th>
<th>1.5</th>
<th>2.0</th>
<th>2.5</th>
<th>3.0</th>
<th>3.5</th>
<th>4.0</th>
<th>4.5</th>
<th>&gt;5.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number (pcu)</td>
<td>5</td>
<td>8</td>
<td>17</td>
<td>46</td>
<td>73</td>
<td>47</td>
<td>18</td>
<td>8</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

**Figure 4. Frequency distribution of harmful lane-changing behaviors in different VAT ranges**

Figure 4 shows that the frequency of the harmful lane-changing behaviors versus VAT follow a bell shaped distribution. A MATLAB test was conducted to check whether the data fits with the normal distribution. It is found that the VAT curve follows a normal distribution with a confidence of 0.95. The frequency of harmful lane-changing behaviors reaches the maximum value when the VAT is about 2.5s. There was no accident happened during data collection.
Risk Analysis of Different Types of Harmful Lane-changing Behaviors

Based on the definition and classification of harmful lane-changing behaviors in previous sections, the harmful lane-changing behaviors at Youyi Avenue and Jiansheyi Road intersections were summarized in Table 5.

Table 5. Distribution of Different Types of Harmful Lane-Changing Behavior

<table>
<thead>
<tr>
<th>Subclass number*</th>
<th>Types</th>
<th>Number</th>
<th>Percentage</th>
<th>VAT ranges</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-1</td>
<td>Forced to change lanes</td>
<td>196</td>
<td>24.75%</td>
<td>0.0-0.7</td>
</tr>
<tr>
<td>A-2</td>
<td>Randomly change lanes</td>
<td>120</td>
<td>15.15%</td>
<td>0.6-2.2</td>
</tr>
<tr>
<td>A-3</td>
<td>Unsuccessful change lanes</td>
<td>101</td>
<td>12.75%</td>
<td>&gt;4.5</td>
</tr>
<tr>
<td>B-1</td>
<td>Forced to change lanes</td>
<td>183</td>
<td>23.11%</td>
<td>2.2-5.4</td>
</tr>
<tr>
<td>B-2</td>
<td>Randomly change lanes</td>
<td>108</td>
<td>13.64%</td>
<td>0.7-1.5</td>
</tr>
<tr>
<td>B-3</td>
<td>Continuously change lane</td>
<td>84</td>
<td>10.61%</td>
<td>1.0-2.5</td>
</tr>
</tbody>
</table>

Table 5 shows different harmful lane-changing behavior associates with different VTA range and risk level. Some interpretations based on the data are made below: (1) The harmful lane-changing behaviors at the signalized intersection mostly occurred in the solid line segment of the signalized intersection approaches and accounted for 52.65% of all harmful lane-changing behaviors. At the same time, the risk level of harmful lane-changing behaviors in the solid line was high as indicated with small VAT values. (2) Forced lane-changing accounted for 47.85%. The risk of forced lane-changing behavior was the highest in the solid line segment. (3) There were 28.79% of all harmful lane-changing behaviors were randomly inserted and changed lanes at the signalized intersections. (4) Continuous lane-changing behavior occurred mostly in non-solid line segments. It was more harmful when the traffic flow at the signalized intersection was large or the speed of the following vehicle was high.

Analysis of Results

The VAT value of the harmful lane-changing behavior is calculated from the data training model. The calculated value from the VAT model and from the Lasso-LARS algorithm were paired for a T-test by using the SPSS statistical analysis software with a confidence of 0.95. Test results are shown in Table 6 and Table 7.

Table 6. Paired Sample Statistics

<table>
<thead>
<tr>
<th></th>
<th>Average</th>
<th>Number of cases</th>
<th>Standard deviation</th>
<th>Average standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAT model</td>
<td>2.64</td>
<td>270</td>
<td>4.913</td>
<td>0.299</td>
</tr>
<tr>
<td>Lasso-LARS algorithm</td>
<td>2.62</td>
<td>270</td>
<td>5.015</td>
<td>0.305</td>
</tr>
</tbody>
</table>
Table 7. Paired Sample Test

<table>
<thead>
<tr>
<th>VAT model &amp; Lasso-LARS algorithm</th>
<th>Average standard deviation</th>
<th>Pairing difference Average standard deviation</th>
<th>95% confidence interval Min</th>
<th>Max</th>
<th>t</th>
<th>Degree of freedom</th>
<th>Significant (two-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.021</td>
<td>0.416</td>
<td>0.025</td>
<td>-0.029</td>
<td>0.71</td>
<td>0.827</td>
<td>0.269</td>
</tr>
</tbody>
</table>

Based on the test results, the VAT calculated value and the observed value are basically consistent with 95% confidence. This indicates the applicability and effectiveness of the vehicle approach time (VAT) model.

CONCLUSION

This study investigated the harmful lane-changing behaviors at signalized intersections based on the video data survey at two intersections in Wuhan, China. The VAT model for the lane-changing behavior was constructed. The comparison analysis found that there was no significant differences between the VAT calculated results and Lasso-LARS regression results as they were within the 95% confidence interval. The accuracy of the improved VAT model was verified. Results from this research can provide theoretical support for the safety analysis of harmful lane-changing behaviors at signalized intersections.

ACKNOWLEDGMENT

This research project is supported by the State of Key Laboratory of Vehicle of VNH and Safety Technology through 2019 annual public foundation awards, research topic No. 12. The project is also supported by the innovation fund project of Inner Mongolia University of Science and Technology, research on the coupling mechanism between productive services and leading industries in Baotou city through project No. 2016QDW-B05. Yanping Hao and Bozhao Qi contributed equally to this work.

REFERENCES


