Article

Risk-Quantification Method for Car-Following Behavior Considering Driving-Style Propensity

Kedong Wang 1,2, Dayi Qu 1,*, Yufeng Yang 1, Shouchen Dai 1 and Tao Wang 1,3

1 School of Mechanical and Automotive Engineering, Qingdao University of Technology, Qingdao 266520, China; 18661831717@163.com (K.W.); yy99718@163.com (Y.Y.); dsc2177@163.com (S.D.); wangtaoq1987@163.com (T.W.)
2 Intelligent Manufacturing Institute, Qingdao Huanghai University, Qingdao 266427, China
3 School of Artificial Intelligence and Big Data, Zibo Vocational Institute, Zibo 255300, China
* Correspondence: dayiqu@qut.edu.cn

Abstract: To systematically study the influence of a propensity for a particular driving style on car-following risk, a safety potential field risk-quantification method that considers driving style is proposed. First, we classify driving behaviors and construct a field-based safety potential car-following model via analogy with intermolecular interactions; second, we establish a risk-quantification model by considering driving style, risk exposure, and risk severity and classify the consequent risk into four levels, high risk, medium risk, low risk, and safe, using the fuzzy C-means algorithm. Finally, we predict the car-following risk using the LightGBM algorithm in real time. The experimental results show that the LightGBM algorithm can recognize up to 86% of medium-high risk levels compared to the Decision Tree and Random Forest Algorithms. It can achieve effective prediction of car-following risk, which provides sufficient warning information to drivers and helps improve the overall safety of vehicle operation.

Keywords: traffic safety; driving style propensity; safety potential field; car following; risk quantification

1. Introduction

Driving style is related to a driver’s intrinsic psychological state and personal characteristics. An accurate assessment of the driving style is crucial for the design of practical driver assistance systems and self-driving vehicle control systems. The concept of “driving style” was first introduced in 1993 [1], and it is defined as a driving pattern that is autonomously chosen or habitual by the driver. Most previous studies have been based on questionnaires. Reason et al. [2] designed the first questionnaire used to study driving behavior, counting the frequency of driver misbehavior and infractions, but did not measure driving style. Motonori et al. [3] designed a driving style questionnaire that specifically classified driving styles, conducted low-speed car-following experiments, and verified its feasibility through modeling. The multidimensional driving style scale proposed by Taubman-Ben-Ari et al. [4] clearly defines the structure and types of driving styles and has been widely used and refined. Cheng et al. [5] constructed an ant colony optimization (ACO) of backpropagation (BP) neural network models for driving style recognition. Murphy et al. [6] proposed categorizing driving styles by analyzing acceleration. To calculate the standard deviation of the acceleration curve within a time window using algorithms, we can find the ratio of the mean acceleration for a specific road type and then classify it using a threshold. The disadvantages of this approach include calibration problems with thresholds and the simultaneous use of only one feature to characterize driving styles. Hei et al. [7] first classified driving styles into three categories according to data collected from driving simulation experiments. Then, according to the classification results, the influence mechanism of the driving style on car-following behavior and lane-changing behavior was analyzed. Finally, models have been constructed based on car-following and lane-changing...
behaviors, which provided a theoretical basis for a more intelligent analysis of driving behavior in the future.

In the early 21st century, scholars proposed the concept of a potential field and applied it to robot-path planning [8]. Inspired by this idea, some scholars have extended potential field theory to the field of traffic flow. Wolf et al. [9] innovatively constructed, based on the potential field theory, the potential field of a vehicle into a wedge shape to realize the mapping between the potential field and traffic behavior. Wang et al. [10] established a unified model of the "driving safety field" based on work by previous researchers, which makes the human–vehicle–road closed-loop system more perfect and can provide reliable driving risk assessment in complex traffic environments. Qu et al. [11] established a molecular follow-up model based on the molecular dynamics analysis of vehicle interaction.

Vehicle safety in driving scenarios is mainly related to its real-time state, which is affected by itself and the surrounding environment, and the vehicle is often in a non-stationary state. If we can predict driving risk and correctly assess the driving risk situation, we can provide drivers with a basis for driving decisions and effectively improve road traffic safety. By analyzing the car-following and braking data of drivers with different driving styles, Liu et al. [12] used the braking time–distance quartile as the threshold value to design a car-following warning model adapted to the needs of drivers with different styles. Ji et al. [13] used the distance-to-collision time as the risk-quantification index, classified it by dividing the threshold value, and, using the random forest algorithm, established a dynamic risk prediction model, which can be used for the design of a dynamic warning system. Based on the field strength theory, Wang [14] used the game theory combination assignment method to determine the impact of human–vehicle–road–environment interaction on risk, expressed it as a field strength value, and calculated the field strength value of the surrounding vehicles on the self-driving vehicle to arrive at a risk assessment.

In summary, most current car-following risk quantification and early warning strategy development only considers the theoretical conflict risk based on theoretical conditions such as vehicle dynamics and does not adequately consider the uncertainty of the car-following behaviors of surrounding drivers with different driving styles, nor that aggressive drivers will increase the car-following risk. To effectively quantify the risk of microscopic car-following behavior in traffic flow, this study first establishes a car-following model based on the safety potential field theory, which systematically portrays the changing trend of the safety potential field of the vehicle under different speed and acceleration values. Simultaneously, it combines driving style and car-following safety, accurately characterizes short-term driving style based on time-series data, constructs different characteristic index groups according to different driving roles, and integrates them into quantitative indicators of car-following risk. Finally, a quantitative model of the integrated car-following risk model, under the safety potential field, is established by considering the real-time risk exposure level and real-time risk severity, and the risk level is categorized. The LightGBM algorithm predicts the risk level in real time under car-following conditions, warns the driver according to the risk level, provides warning information containing qualitative style descriptions, and assists the driver in avoiding risk.

2. Driving Style Propensity

2.1. Driving Behaviors

Driving style can reflect the driver’s operating habits, and accurately identifying driving style is crucial for better understanding human driver behavior. In application scenarios that require real-time prediction of the risk of following a driver, a good driving style characterization can help improve the accuracy of risk assessment. Driving styles, inherently complex and variable, are significantly influenced by diverse driving purposes and environments. In scenarios involving multiple vehicles, understanding the driving styles of nearby vehicles enables drivers to make informed decisions, enhancing road safety.

According to Zhao [15], “driving propensity” captures a driver’s unique characteristics in the short term and its influence on safety. The outward driving performance embodied in
the operation of the vehicle by the driver is the driving behavior. The process of generating driving behavior is that the driver, in the process of driving, analyzes information about the surrounding traffic environment and the road environment, makes a decision, and changes the direction and speed of the vehicle by active operation to generate a specific type of driving behavior. The driving behavior reflects the specific characteristics of a vehicle on the road.

When car-following, the driver must observe the driving behaviors of both the leading and following vehicles in the lane and control the horizontal and longitudinal movements of the vehicle in time to ensure the safety of car-following behaviors.

2.2. Selection of Indicators

Utilizing the highD dataset, collected from German highways in 2017 and 2018, our study focused on transforming driver behavior data into stylistic propensity semantic labels. The dataset captures a wide array of traffic scenarios with four and six-lane highways featuring 3.75 m-wide lanes. Aerial surveillance via drones equipped with high-resolution cameras recorded vehicle movements over a 420 m stretch, yielding a comprehensive dataset of 110,516 vehicles, including 89,139 cars and 21,377 trucks, over a total distance of 44,050 km. The average observation time was 13.6 seconds per vehicle, providing a substantial sample for analysis. These data underwent rigorous statistical analysis, as detailed in Table 1.

Table 1. Vehicle operating characteristics.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Speed (m/s)</th>
<th>Acceleration (m/s²)</th>
<th>Headway (s)</th>
<th>Crash Time (s)</th>
<th>Headspace (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>28.0427</td>
<td>0.2356</td>
<td>1.9048</td>
<td>48.7392</td>
<td>52.5850</td>
</tr>
<tr>
<td>Variance</td>
<td>47.2258</td>
<td>0.0577</td>
<td>1.4839</td>
<td>1590.8703</td>
<td>1329.2565</td>
</tr>
</tbody>
</table>

The red line segment in Figure 1 shows the median of the statistics and the average longitudinal speed of the vehicles is 28.04 m/s, which indicates the pronounced acceleration and deceleration behaviors. The data show an average headway time of 1.90 s and an average spacing of 52.58 m, suggesting that vehicles generally maintain a medium following distance. However, a notable amount of data points toward shorter following distances. Additionally, the average time to collision stands at 48.74 seconds, further illustrating the driving behavior patterns observed in the dataset.

During car-following, a sudden deceleration by the lead vehicle necessitates immediate reaction from the following vehicles to prevent collisions. Timely deceleration is crucial to avoid potential conflicts; thus, the acceleration and deceleration patterns of the lead vehicle are key characteristic indicators for analysis. It is important to note the significant
impact that sharp deceleration of the lead vehicle can have on the following vehicle, which requires careful calibration of style propensity.

The following vehicle must effectively control its speed and maintain a safe distance from the leading vehicle, with speed serving as a critical metric for identifying the driving style of the following driver. According to [16], drivers adjust their speed based on a comfort zone, with those inclined to higher speeds at risk of maintaining such speeds, increasing the potential for conflict. The risk and severity of collisions escalate with the speed of the following vehicle.

Based on previous studies [15], the selected feature indicators and their thresholds in this study are shown in Tables 2 and 3, and the driving style propensity scores for both leading and following vehicles are shown in Figures 2 and 3, with darker colors indicating larger values and higher driving style propensity scores.

### Table 2. Characterization indicator thresholds of the leading car.

<table>
<thead>
<tr>
<th>Characterization</th>
<th>Statuses</th>
<th>Thresholds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed (m/s)</td>
<td>Slow</td>
<td>&lt;=24.72</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>(24.72, 36.53)</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>&gt;36.53</td>
</tr>
<tr>
<td>Acceleration (m/s²)</td>
<td>Decelerate</td>
<td>&lt;=-0.13</td>
</tr>
<tr>
<td></td>
<td>Decelerate slowly</td>
<td>(-0.13, 0.00)</td>
</tr>
<tr>
<td></td>
<td>Non-acute deceleration</td>
<td>&gt;0.00</td>
</tr>
</tbody>
</table>

### Table 3. Characterization indicator thresholds of the following car.

<table>
<thead>
<tr>
<th>Characterization</th>
<th>Statuses</th>
<th>Thresholds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed (m/s)</td>
<td>Slow</td>
<td>&lt;=23.09</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>(23.09, 34.08)</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>&gt;34.08</td>
</tr>
<tr>
<td>Acceleration (m/s²)</td>
<td>Decelerate</td>
<td>&lt;=-0.13</td>
</tr>
<tr>
<td></td>
<td>Decelerate slowly</td>
<td>(-0.13, -0.02)</td>
</tr>
<tr>
<td></td>
<td>Uniform acceleration/ deceleration</td>
<td>(-0.02, 0.06)</td>
</tr>
<tr>
<td></td>
<td>Slower acceleration</td>
<td>(0.06, 0.17)</td>
</tr>
<tr>
<td></td>
<td>Accelerate rapidly</td>
<td>&gt;0.17</td>
</tr>
<tr>
<td>Headspace (m)</td>
<td>Short</td>
<td>&lt;=28.91</td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>(28.91, 79.46)</td>
</tr>
<tr>
<td></td>
<td>Long</td>
<td>&gt;79.46</td>
</tr>
</tbody>
</table>

**Figure 2.** Leading car style propensity score setting.
They form a road line field, constraining vehicles to maintain their position in the middle of the lane. Assuming the leftmost road marks the location for the axis $y$, with reference to the class of Gaussian function to build the road line potential field model, the formula is as follows

$$
\Phi_L = \sum_{i=1}^{n} A_i \exp \left[ -\frac{(x - x_i)^2}{2\theta^2} \right] \frac{x - x_i}{|x - x_i|}
$$

(1)

where $A_i$, for all types of road marking potential field strength coefficients, can determine the maximum threshold of the road line field; $A_1$ represents the dotted gray line; $A_2$ represents the double yellow line; and $A_1 << A_2$; $x_i$ is the horizontal coordinate of the $i$th road line; and $\theta$ for the determination of the road field strength of the speed of the lift.

3.1.2. Road Boundary Potential Field

The potential field generated by road boundaries exerts a significant influence on vehicles in motion, intensifying as the distance to the boundary decreases. This mechanism effectively deters vehicles from approaching too closely, thereby mitigating the risk of collisions with other vehicles should be avoided. That is, the safety potential field is based on the assumption of a risk field, which describes the risk faced by vehicles in the process of traveling.

The vehicle’s traveling process is constrained by the road line and road boundary line, and collisions with other vehicles should be avoided. That is, the safety potential field consists of the superposition of three types of risk fields, namely, the road line potential field, road boundary field, and vehicle interaction potential field, which are represented by $\Phi_L$, $\Phi_R$, and $\Phi_V$. 

3.1.1. Road Line Potential Field

As shown in Figure 4, there are two types of road markings in the road traffic scenario: the dashed gray line indicates the lane demarcation line, which allows vehicles to cross the line to change roads, and the double yellow line regulates the direction of vehicle travel, which prohibits crossing over to change lanes. They form a road line field, constraining vehicles to maintain their position in the middle of the lane. Assuming the leftmost road marks the location for the axis $y$, with reference to the class of Gaussian function to build the road line potential field model, the formula is as follows

$$
\Phi_L = \sum_{i=1}^{n} A_i \exp \left[ -\frac{(x - x_i)^2}{2\theta^2} \right] \frac{x - x_i}{|x - x_i|}
$$

(1)

where $A_i$, for all types of road marking potential field strength coefficients, can determine the maximum threshold of the road line field; $A_1$ represents the dotted gray line; $A_2$ represents the double yellow line; and $A_1 << A_2$; $x_i$ is the horizontal coordinate of the $i$th road line; and $\theta$ for the determination of the road field strength of the speed of the lift.

3. Risk Analysis and Modeling Based on Security Potential Field

3.1. Security Potential Field Modeling

From a physics perspective, a field measures interactions between entities based on their relative positions, affecting the strength of their interactions. Analogous to potential field theory, microscopic driving behavior can be regarded as the influence of the potential field of the surrounding environment on the target vehicle. This means that even without physical contact, vehicles adjust their movement in sync with others, akin to navigating a potential field to achieve a balance between following distance, safety, and efficiency. Every traffic element influencing vehicle movement serves as a field source, which is ultimately superimposed on the safety potential field. The safety potential field is based on the assumption of a risk field, which describes the risk faced by vehicles in the process of traveling.

Figure 3. Following car style propensity score setting.

3.1.1. Road Line Potential Field

As shown in Figure 4, there are two types of road markings in the road traffic scenario: the dashed gray line indicates the lane demarcation line, which allows vehicles to cross the line to change roads, and the double yellow line regulates the direction of vehicle travel, which prohibits crossing over to change lanes. They form a road line field, constraining vehicles to maintain their position in the middle of the lane. Assuming the leftmost road marks the location for the axis $y$, with reference to the class of Gaussian function to build the road line potential field model, the formula is as follows

$$
\Phi_L = \sum_{i=1}^{n} A_i \exp \left[ -\frac{(x - x_i)^2}{2\theta^2} \right] \frac{x - x_i}{|x - x_i|}
$$

(1)

where $A_i$, for all types of road marking potential field strength coefficients, can determine the maximum threshold of the road line field; $A_1$ represents the dotted gray line; $A_2$ represents the double yellow line; and $A_1 << A_2$; $x_i$ is the horizontal coordinate of the $i$th road line; and $\theta$ for the determination of the road field strength of the speed of the lift.

3.1.2. Road Boundary Potential Field

The potential field generated by road boundaries exerts a significant influence on vehicles in motion, intensifying as the distance to the boundary decreases. This mechanism effectively deters vehicles from approaching too closely, thereby mitigating the risk of collisions with other vehicles should be avoided. That is, the safety potential field is based on the assumption of a risk field, which describes the risk faced by vehicles in the process of traveling.

3.1.1. Road Line Potential Field

As shown in Figure 4, there are two types of road markings in the road traffic scenario: the dashed gray line indicates the lane demarcation line, which allows vehicles to cross the line to change roads, and the double yellow line regulates the direction of vehicle travel, which prohibits crossing over to change lanes. They form a road line field, constraining vehicles to maintain their position in the middle of the lane. Assuming the leftmost road marks the location for the axis $y$, with reference to the class of Gaussian function to build the road line potential field model, the formula is as follows

$$
\Phi_L = \sum_{i=1}^{n} A_i \exp \left[ -\frac{(x - x_i)^2}{2\theta^2} \right] \frac{x - x_i}{|x - x_i|}
$$

(1)

where $A_i$, for all types of road marking potential field strength coefficients, can determine the maximum threshold of the road line field; $A_1$ represents the dotted gray line; $A_2$ represents the double yellow line; and $A_1 << A_2$; $x_i$ is the horizontal coordinate of the $i$th road line; and $\theta$ for the determination of the road field strength of the speed of the lift.

3. Risk Analysis and Modeling Based on Security Potential Field

3.1. Security Potential Field Modeling

From a physics perspective, a field measures interactions between entities based on their relative positions, affecting the strength of their interactions. Analogous to potential field theory, microscopic driving behavior can be regarded as the influence of the potential field of the surrounding environment on the target vehicle. This means that even without physical contact, vehicles adjust their movement in sync with others, akin to navigating a potential field to achieve a balance between following distance, safety, and efficiency. Every traffic element influencing vehicle movement serves as a field source, which is ultimately superimposed on the safety potential field. The safety potential field is based on the assumption of a risk field, which describes the risk faced by vehicles in the process of traveling.

The vehicle’s traveling process is constrained by the road line and road boundary line, and collisions with other vehicles should be avoided. That is, the safety potential field consists of the superposition of three types of risk fields, namely, the road line potential field, road boundary field, and vehicle interaction potential field, which are represented by $\Phi_L$, $\Phi_R$, and $\Phi_V$. 

3.1.1. Road Line Potential Field

As shown in Figure 4, there are two types of road markings in the road traffic scenario: the dashed gray line indicates the lane demarcation line, which allows vehicles to cross the line to change roads, and the double yellow line regulates the direction of vehicle travel, which prohibits crossing over to change lanes. They form a road line field, constraining vehicles to maintain their position in the middle of the lane. Assuming the leftmost road marks the location for the axis $y$, with reference to the class of Gaussian function to build the road line potential field model, the formula is as follows

$$
\Phi_L = \sum_{i=1}^{n} A_i \exp \left[ -\frac{(x - x_i)^2}{2\theta^2} \right] \frac{x - x_i}{|x - x_i|}
$$

(1)

where $A_i$, for all types of road marking potential field strength coefficients, can determine the maximum threshold of the road line field; $A_1$ represents the dotted gray line; $A_2$ represents the double yellow line; and $A_1 << A_2$; $x_i$ is the horizontal coordinate of the $i$th road line; and $\theta$ for the determination of the road field strength of the speed of the lift.

3.1.2. Road Boundary Potential Field

The potential field generated by road boundaries exerts a significant influence on vehicles in motion, intensifying as the distance to the boundary decreases. This mechanism effectively deters vehicles from approaching too closely, thereby mitigating the risk of collisions with other vehicles should be avoided. That is, the safety potential field is based on the assumption of a risk field, which describes the risk faced by vehicles in the process of traveling.

The vehicle’s traveling process is constrained by the road line and road boundary line, and collisions with other vehicles should be avoided. That is, the safety potential field consists of the superposition of three types of risk fields, namely, the road line potential field, road boundary field, and vehicle interaction potential field, which are represented by $\Phi_L$, $\Phi_R$, and $\Phi_V$. 

3.1.1. Road Line Potential Field

As shown in Figure 4, there are two types of road markings in the road traffic scenario: the dashed gray line indicates the lane demarcation line, which allows vehicles to cross the line to change roads, and the double yellow line regulates the direction of vehicle travel, which prohibits crossing over to change lanes. They form a road line field, constraining vehicles to maintain their position in the middle of the lane. Assuming the leftmost road marks the location for the axis $y$, with reference to the class of Gaussian function to build the road line potential field model, the formula is as follows

$$
\Phi_L = \sum_{i=1}^{n} A_i \exp \left[ -\frac{(x - x_i)^2}{2\theta^2} \right] \frac{x - x_i}{|x - x_i|}
$$

(1)

where $A_i$, for all types of road marking potential field strength coefficients, can determine the maximum threshold of the road line field; $A_1$ represents the dotted gray line; $A_2$ represents the double yellow line; and $A_1 << A_2$; $x_i$ is the horizontal coordinate of the $i$th road line; and $\theta$ for the determination of the road field strength of the speed of the lift.

3.1.2. Road Boundary Potential Field

The potential field generated by road boundaries exerts a significant influence on vehicles in motion, intensifying as the distance to the boundary decreases. This mechanism effectively deters vehicles from approaching too closely, thereby mitigating the risk of collisions with other vehicles should be avoided. That is, the safety potential field is based on the assumption of a risk field, which describes the risk faced by vehicles in the process of traveling.

The vehicle’s traveling process is constrained by the road line and road boundary line, and collisions with other vehicles should be avoided. That is, the safety potential field consists of the superposition of three types of risk fields, namely, the road line potential field, road boundary field, and vehicle interaction potential field, which are represented by $\Phi_L$, $\Phi_R$, and $\Phi_V$.
collisions. Generally, there are two road boundaries, and the road boundary potential field can be expressed as

$$\Phi_R = \sum_{j=1}^{2} 2^\rho \left( \frac{1}{|x - x_j|} \right)^2 \frac{x - x_j}{|x - x_j|}$$

(2)

where $j$ is the road boundary line, provided that $j = 1, 2$ are the left and right boundaries, respectively; $x_j$ is the horizontal coordinate of the road boundary line; and $\rho$ is the position gain parameter.

Figure 4. Road traffic scene.

According to reference [17], we can set $A_1$, $A_2$, $\theta$ and $\rho$, respectively, as $2$, $8$, $1.22$, and $3$. The distributions of the road line potential field and the road boundary potential field are shown in Figure 5. The superimposed 3D map of the safety potential field is constructed as shown in Figure 6, and different colors are used to correspond to different vertical coordinate values.

Figure 5. Road lines and road boundary potential fields.

Figure 6. Schematic diagram of potential field superposition.
3.1.3. Vehicle-Actuated Potential Field

The vehicle interaction potential field originates from other vehicles on the road, which can be analogized to molecules, and the potential field method is used to study car-following behavior. Scholars have developed various mathematical models to describe the interactions between molecules within microphysical systems, among which the Lennard–Jones potential model is extensively used due to its simple structural form. This model conceptualizes the interactions between molecules as a combination of attractive (gravitational) and repulsive forces [18]. Given that both attractive and repulsive forces are inversely related to the powers of the distance between molecules, its mathematical expression is

\[
\varphi(r) = \frac{A}{r^\alpha} - \frac{B}{r^\beta}
\]

(3)

where \(A\) and \(B\) are the intermediate variables of the potential function, and \(\frac{A}{r^\alpha}\) and \(\frac{B}{r^\beta}\) are the repulsive and gravitational terms, respectively, where the gravitational force on the molecule is close to zero when \(r\) is too large.

Molecules close to each other present a state in which they are neither infinitely close nor infinitely far away; that is, the combined intermolecular force is zero, which is known as the equilibrium distance between molecules \(r_0\). At this point, the molecular potential energy was minimized and specified as \(\epsilon\); then

\[
\left. \frac{d\varphi(r_0)}{dr} \right|_{r=r_0} = 0
\]

(4)

\[
\varphi(r_0) = -\epsilon
\]

(5)

Joining the two equations gives

\[
r_0^{\alpha-\beta} = \frac{\alpha A}{\beta B}
\]

(6)

The Lennard–Jones potential function expression is

\[
\varphi(r) = \frac{\epsilon}{\alpha-\beta} \left[ \beta \left( \frac{r_0}{r} \right)^\alpha - \alpha \left( \frac{r_0}{r} \right)^\beta \right]
\]

(7)

Drawing an analogy between vehicles traveling on the road and molecules within a potential field, there exists an equilibrium state between leading and following vehicles, characterized by a necessary safe following distance. The mathematical model is

\[
X_r = (\tau - \omega \Delta v)v_n
\]

(8)

where \(\tau\) and \(\omega\) are the coefficients to be determined; \(\Delta v\) is the speed difference between the leading and following vehicles; and \(v_n\) is the speed of the following vehicle.

When the distance between the leading and following vehicles is smaller than the required safety distance, the following vehicle tends to move away from the leading vehicle; conversely, the following vehicle tends to move closer to the leading vehicle. The traveling vehicle will not be too close or too far from the leading vehicle to ensure the dynamic balance of the following vehicle. Based on this, the vehicle interaction potential function is constructed as

\[
\varphi_n(r) = \frac{\epsilon}{\alpha-\beta} \left[ \beta \left( \frac{X_r}{L} \right)^\alpha - \alpha \left( \frac{X_r}{L} \right)^\beta \right]
\]

(9)

where \(L\) is the actual distance between the leading and following vehicles.
Let the mass of the vehicle be \( m \). Then, the longitudinal acceleration of the following vehicle under the action of the interacting potential field is

\[
a_n = \frac{\alpha \beta e}{m(\alpha + \beta)} \left( \frac{X^\beta}{L^\beta + 1} - \frac{X^\alpha}{L^\alpha + 1} \right) X^r \beta L^\beta + 1 - X^r \alpha L^\alpha + 1
\]

(10)

Within the potential field, the intensity of the potential at any point is negatively correlated with the distance to the target vehicle. Typically, the intensity of the potential field decreases as the distance increases. The coordinates of the vehicle’s center of mass are set as \((x_0, y_0)\), and the distance \( s \) between any point \((x, y)\) and the target vehicle can be represented as

\[
|s| = \sqrt{(x - x_0)^2 - (y - y_0)^2}
\]

(11)

Vehicles exhibiting various motion states exert distinct impacts on the distribution of their potential field; notably, velocity and acceleration play crucial roles in shaping the potential field’s distribution. To ensure that the following vehicle does not collide with the leading vehicle during braking, it should be ensured that the distance between the leading and following vehicles during braking is not less than the minimum safety distance \( s_0 \). Therefore, the range of the repulsive field distribution is corrected as

\[
|s'| = s_0 \left[ s_0 \left( \frac{x - x_0}{\exp(\gamma X^r)} \right) \right]^2
\]

(12)

where \( \gamma \) is the coefficient to be determined.

Modeling the vehicle action potential field considering the state of motion,

\[
\Phi_V = \varphi_n(r) \cdot \exp(\cos \theta) \cdot \frac{s'}{|s'|}
\]

(13)

where \( \theta \) is the angle between the line from the surrounding vehicles to the center of mass of the target vehicle and the direction of travel of the target vehicle (in the clockwise direction).

3.2. Safety Potential Field Following Spatial Model

The safety potential \( \Phi \) shown in Figure 7 consists of the lane line potential field \( \Phi_L \), the road boundary potential field \( \Phi_R \), and the vehicle interaction potential field \( \Phi_V \). As shown in Figure 7, it can reflect the real-time risk faced by the vehicle, and different colors indicate the degree of risk of the vehicle, from blue to red indicates that the value of risk of the vehicle increases. The comprehensive risk field model of the traveling vehicle is

\[
\Phi = \Phi_L + \Phi_R + \Phi_V
\]

(14)

The microscopic motion behavior of the vehicle is determined by analyzing the total potential field distribution of its surroundings. With the safety of the vehicle traveling as the primary goal, the following vehicle should keep traveling at a low risk. As shown in Figure 8, with the vehicle’s center of mass as the origin, a coordinate axis is established, the red curve indicates the vehicle potential field and the green curve indicates the lane potential field. Car-following behavior can be viewed as the motion under the combined effect of various potential fields. Therefore, the vehicle’s longitudinal acceleration based on the safety potential field can be expressed as

\[
a = \frac{\partial \Phi_L}{\partial x} + \frac{\partial \Phi_R}{\partial x} + \frac{\partial \Phi_V}{\partial y}
\]

(15)
**Figure 7.** Comprehensive risk field for driving.

**Figure 8.** Schematic diagram of the safety potential field-based follow-through.

### 4. Risk Quantification and Behavioral Modeling

#### 4.1. Stop Distance Index

The Stopping Sight Distance (SSD) is the shortest distance required for a vehicle to brake and stop when it encounters an obstacle in front of it while traveling [19]. The expression is as follows

\[
SSD = \frac{v^2}{3291.84 \times (a / g \pm \theta)} + 0.077vt_f
\]

where \( v \) is the vehicle speed (m/s); \( a \) is the deceleration speed (m/s\(^2\)), which is set according to the vehicle type, and the deceleration speed of a passenger car is set to 3.4 m/s\(^2\) [20]; \( g \) is the gravitational acceleration; \( \theta \) is the gradient of the road; and \( t_f \) is the driver’s reaction time(s), which is usually 2.5 s [21].

The Stop Distance Index (SDI) is chosen as the discriminant index to fully consider the sudden risk during the following process. The SDI between the leading and following vehicles calculated from the SSD of the leading and following vehicles can be used to determine whether the following event is safe. An SDI greater than 0 indicates that the following vehicle can stop safely when the leading vehicle stops suddenly, whereas an SDI less than 0 indicates that the following vehicle is unable to avoid a collision with the leading vehicle; therefore, the smaller the SDI, the greater the likelihood of a collision. By
combining the SDI, which represents the theoretical likelihood of collision, and the effect of driver heterogeneity, a quantitative indicator of collision risk can be obtained.

$$SDI_t(e) = \begin{cases} s_t(e) + SSD^e_t(e) - SSD'^e_t(e), & \text{if } e \text{ exists} \\ \infty, & \text{if } e \text{ doesn’t exist} \end{cases} \quad (17)$$

where $e$ denotes the interaction event between $k_S$ and each neighboring vehicle $(k_F, k_R)$, and $e = 1$ denotes the interaction of $k_S$ and $k_F$. $SDI_t(e)$ and $s_t(e)$ denote the Stop Distance Index and vehicle spacing between the leading and following vehicles in the interaction event $e$ at time $t$. The definition of vehicle spacing is shown in Figure 9, which must be obtained based on the longitudinal distance between the leading and following vehicles and the difference between the lengths of the vehicle bodies.

![Figure 9. Car-following state.](image)

The SDI quantifies the theoretical likelihood of a collision based on current speed and distance, with surrounding vehicles’ varying driving styles introducing differing degrees of collision risk. A higher driving style propensity score for the lead vehicle indicates an increased likelihood of sudden decelerations and erratic lateral movements, thereby heightening the risk of a collision. As such, the driving style propensities of both the lead and following vehicles are factored into the collision risk assessment to account for the variability introduced by driver behavior. Given that the style propensity scores of the lead and following vehicles derive from different metrics, they are normalized to a uniform scale ranging from 0 to 1, ensuring comparability. This normalization allows for the driving styles of both vehicles to be evaluated within the same dimensional framework, facilitating a comprehensive analysis of their impact on collision probability.

$$N_t(e) = \frac{SDI_t(e)}{1 + DS(e)} \quad (18)$$

where $N_t(e)$ denotes a composite influence factor that considers both the theoretical collision likelihood and the additional collision likelihood due to driver heterogeneity in the interaction event $e$ at moment $t$, and $DS(e)$ is the driving style propensity score of the surrounding vehicles in the interaction event $e$ after the normalization process of the corresponding interaction roles (as the leading or the following vehicle of the target vehicle).
4.2. Real-Time Risk Exposure Level

The likelihood of a collision between the target vehicle and surrounding vehicles during the follow-through can be measured by the Real-time Risk Exposure Level (RREL), which is expressed as

\[
RREL_t(e) = \begin{cases} 
\exp\left(-\frac{N_t(e)}{\sigma}\right), & SDI_t(e) < 0, SDI_t(e) \geq 0 
\end{cases}
\]  

(19)

where \(RREL_t(e)\) is the real-time risk exposure of interaction event \(e\) at time \(t\), and \(\sigma\) is the shape parameter of the exponential decay function EDF. When \(SDI < 0\), the following vehicle cannot avoid a collision with the leading vehicle with a likelihood of 1. When \(SDI \geq 0\), \(N_t(e)\) is input into the EDF [16], which transforms the collision likelihood into a real-time risk exposure level with a numerical value restricted between 0 and 1, where the EDF is \(y = \exp\left(-\frac{z}{\sigma}\right)\), and \(\sigma = 2\). The smaller the \(SDI\), the larger the RREL, and the larger the DS, the more threatening the driver’s style is to the target vehicle, and the larger the corresponding RREL. In addition, if the interaction event \(e\) does not exist, \(SDI_t(e)\) and \(N_t(e)\) are infinite, and the real-time risk exposure level \(RREL_t(e)\) of this interaction event is almost zero.

4.3. Real-Time Risk Severity Level

The Real-time Risk Severity Level (RRSL) serves as a metric for evaluating the potential conflict risk’s severity between the target vehicle and its surrounding counterparts during the follow-through phase. According to [22], the energy loss during a collision is directly proportional to the square of the velocity difference between the vehicles involved. Consequently, the square of the speed differential among the conflicting vehicles offers a direct proportional to the square of the velocity difference between the vehicles involved.

\[
\Delta V_t(e) = \begin{cases} 
\left(V_t^f(e) - V_t^p(e)\right), & if e exists \\
0, & if e doesn’t exist 
\end{cases}
\]

(20)

\[
RRSL_t(e) = \begin{cases} 
\exp\left(-\frac{1}{\Delta V_t(e)}\right), & if e exists \\
0, & if e doesn’t exist 
\end{cases}
\]

(21)

where \(\Delta V_t(e)\) denotes the square of the speed difference between the following and leading vehicles in the interaction event \(e\) at time \(t\); \(V_t^f(e)\) denotes the speed of the following vehicle in the interaction event \(e\) at time \(t\); and \(V_t^p(e)\) denotes the speed of the leading vehicle in the interaction event \(e\) at time \(t\). The real-time collision severity index \(RRSL_t(e)\) is limited to the range of 0–1 by the exponential decay function. The higher the square of the velocity difference, the higher the real-time collision severity. In addition, if the interaction event \(e\) does not exist, \(\Delta V_t(e)\) is zero, and \(RRSL_t(e)\) is zero.

4.4. Car-Following Risk Index Considering Driving Style Propensity

The following collision is a system failure, denoted by \(P(k_S)\), which is the top event. The safety failure \(\gamma(e)\) between the following vehicle and surrounding vehicles, as an intermediate event, and the interaction safety failure between the following vehicle \(k_S\) and surrounding vehicles \((k_F, k_P)\) are denoted by \(\gamma(1)\) and \(\gamma(2)\), respectively. \(RREL\) and \(RRSL\) are two important factors that cause safety failure; they are set as the failure factors and are used as the bottom event. \(RREL\) and \(RRSL\) together affect the interaction safety failure of the vehicles. Using the \(RREL\) and \(RRSL\) together affects the vehicle interaction safety fault using the “with” operation:

\[
\gamma(e) = RREL(e) \times RRSL(e)
\]

(22)
where $\gamma(e)$ denotes the probability of a safety failure of the interaction event $e$. If an interaction event $e$ does not exist, its failure probability is 0.

The probability that the target vehicle fails to perform follow-through safely is referred to as the follow-through risk index (CFR—Car-Following Risks). When a safety failure event occurs with any vehicle, it is determined that the system is faulty. Therefore, the probability of failing to perform the follow-through safely is obtained from each safety interaction failure event through the “or” operation:

$$P(k_s) = 1 - \prod_{e=1}^{4} [1 - \gamma(e)]$$  \hspace{1cm} (23)

The smaller the CFR during car-following, the safer the car-following condition.

5. Experimental Analysis

5.1. Data Processing

The following criteria were used for car-following behavior, which is an event in which the driver did not change lanes unintentionally during data collection and no lane change occurred for car-following event extraction:

1. The lane number of the following vehicle remained constant throughout the detection section to ensure that the extracted vehicle did not change lanes.
2. The duration of the follow-through event was determined to be 5 s according to international standards [23].

Model samples were derived from segments of the highD real vehicle trajectory dataset, which encompasses both four-lane and six-lane divided highways. From this dataset, a total of 632 following events were identified. Of these, data from 505 events were allocated for model training purposes, while the remaining 127 events were utilized to test the model’s recognition capabilities.

5.2. Risk Level Classification

To apply CFR to car-following risk prediction, a risk class classification is required. The Fuzzy C-Means (FCM) clustering method determines each sample’s degree of membership through iterative calculations, addressing the ambiguity in sample categorization to achieve objective classification. This membership degree enhances the precision in sample classification, offering a more nuanced and adaptable clustering outcome than the K-means algorithm. Consequently, FCM was selected for delineating risk states. Post-clustering, classification thresholds are established by averaging the minimum and maximum CFR values across adjacent categories.

The implementation of the FCM clustering algorithm was carried out in Python. Drawing on findings from prior studies [24], setting the number of categories to four yields a more logical categorization. Inputting the number of clusters and real-time CFR data for 632 vehicle groups undergoing the heeling process into the FCM algorithm facilitated the derivation of thresholds for the four defined heeling risk categories, detailed in Table 4. The identified clustering centers were 0.3514, 0.5765, 0.7624, and 0.9418, respectively.

Table 4. Risk classification.

<table>
<thead>
<tr>
<th>Classification</th>
<th>Thresholds</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Risk (IV)</td>
<td>$CFR &gt; 0.8569$</td>
</tr>
<tr>
<td>Medium Risk (III)</td>
<td>$0.6741 &lt; CFR \leq 0.8569$</td>
</tr>
<tr>
<td>Low risk (II)</td>
<td>$0.4635 &lt; CFR \leq 0.6741$</td>
</tr>
<tr>
<td>Safe (I)</td>
<td>$CFR &lt; 0.4635$</td>
</tr>
</tbody>
</table>
Level I represents the safest category, characterized by a CFR below 0.4635. Vehicles in this category are deemed to have minimal risk, emphasizing the importance of maintaining a safe following distance.

Level II is designated for low-risk vehicles, with a CFR ranging from 0.4635 to 0.6741. Drivers in this category should be particularly mindful of the spacing between their vehicle and others on the road, ensuring adequate distance is maintained to prevent potential collisions.

Level III corresponds to a medium-risk status, where the CFR falls between 0.6741 and 0.8569. In this category, heightened vigilance is recommended, with drivers advised to closely monitor the driving behaviors of vehicles both ahead and behind and to make prompt adjustments to their driving as necessary to enhance safety.

Level IV is indicative of highly unsafe following practices, with a CFR exceeding 0.8569. Vehicles in this category are at significant risk, necessitating a state of readiness to respond effectively to unforeseen events and situations on the road.

5.3. LightGBM-Based Risk Prediction for Car Following

5.3.1. Selection of Characteristic Indicators

LightGBM was chosen to predict the risk level of car following. As a gradient enhancement algorithm, the LightGBM algorithm adopts a leaf growth strategy with depth limitation to improve the accuracy of the algorithm while avoiding the problem of overfitting.

Car-following involves real-time interactions with adjacent vehicles; hence, the motion characteristics (E) extracted for risk assessment encompass not only the dynamics of the following vehicle but also those of the leading and following vehicles (the interaction traits of nearby vehicles). The predictive outcome (F) spans four categories: safe, low risk, medium risk, and high risk, with the Collision Frequency Rate (CFR) being computed and sorted based on predefined thresholds to serve as the classification labels. In contrast to the feature selection in the behavioral recognition module, risk prediction attributes account for the target vehicle’s relative lateral distance and velocity in relation to nearby vehicles, as well as the driving style tendencies of those vehicles concerning the assessed risk level. A set of eight attributes outlined in Table 5 were identified for the purpose of car-following risk prediction.

<table>
<thead>
<tr>
<th>Category</th>
<th>Characteristics</th>
<th>Characteristic Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characteristics of following vehicles</td>
<td>Vehicle speed</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Vehicle acceleration</td>
<td>1</td>
</tr>
<tr>
<td>Characteristics of interaction with vehicles in front and behind</td>
<td>Relative longitudinal distance between the following vehicle and the vehicle in front of and behind it</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Longitudinal speed of the following vehicle relative to the leading and following vehicles</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Driving style propensity of leading and following vehicles</td>
<td>2</td>
</tr>
</tbody>
</table>

5.3.2. Analysis of Risk Prediction Results

Classification models are assessed using metrics such as accuracy, precision, recall, and F1-score. Accuracy, the proportion of correctly classified predictions out of the total predictions, is a widely utilized metric for evaluating the overall predictive performance on a global scale. However, these metrics encounter limitations when applied to unbalanced datasets, diminishing their relevance for assessing accuracy in scenarios characterized by imbalanced risk levels.
The failure to accurately predict medium-high-risk conditions as such carries significant safety implications and should be minimized. From a safety perspective, erring on the side of caution by classifying a low-risk scenario as medium–high risk is preferable to misidentifying actual medium–high-risk conditions as lower risk. Consequently, the objective shifts toward maximizing the recall of medium–high risk categories, even if it means sacrificing some precision. The emphasis is thus on enhancing recall for medium- and high-risk states while maintaining high precision.

An evaluation of Decision Tree, Random Forest, and LightGBM algorithms, as illustrated in Figure 10, compares the precision and recall rates across three heeling risk-prediction models. The Decision Tree algorithm exhibits lower performance metrics, whereas LightGBM outperforms with the highest overall precision and recall rates.

In Figure 11, the recalls for different vehicle risk propensities are shown in different color depths, with darker colors indicating larger values. The Decision Tree algorithm achieves a recall of 0.695 for medium-risk classifications, indicating that 69.5% of medium-risk samples are correctly identified, with a high-risk classification recall of 0.575, resulting in an average medium and high-risk recall of 0.635. The Random Forest algorithm records a medium-risk recall of 0.686 and a high-risk recall of 0.806, averaging a medium–high risk recall of 0.746. LightGBM demonstrates superior recall rates of 0.916 for medium–risk and 0.814 for high-risk classifications, achieving an average medium–high risk recall of 0.865. This performance underscores LightGBM’s superior effectiveness in identifying medium–high risk conditions accurately.

Table 6 summarizes the metrics of LightGBM, and it can be seen that the performance is better and more suitable for car-following risk level prediction.

![Figure 10. Comparison of forecast results.](image)

![Figure 11. Categorical recall.](image)

(a) Decision Tree. (b) Random Forests. (c) LightGBM.
Table 6. LightGBM model evaluation.

<table>
<thead>
<tr>
<th>Driving Behavior</th>
<th>Risk Level</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car Follow</td>
<td>High Risk</td>
<td>0.93</td>
<td>0.81</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>Medium Risk</td>
<td>0.92</td>
<td>0.92</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>Low Risk</td>
<td>0.88</td>
<td>0.94</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>Safe</td>
<td>0.92</td>
<td>0.84</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>Overall Performance</td>
<td>0.91</td>
<td>0.88</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Figure 12 displays the significance of various features within the risk level prediction model, highlighting the pivotal role of interaction attributes in determining risk. The most critical feature is the relative speed to the lead vehicle, underscoring its impact on risk assessment. Equally significant are the relative distances to the vehicles ahead and behind, indicating their substantial influence on risk evaluation. The driving style propensity of the driver ranks as the second and third most influential factors, respectively, showcasing its considerable effect in modeling car-following behavior. The notable impact of driver style propensity on additional crash likelihood emphasizes its relevance in car-following scenarios. This analysis of feature importance not only aids in identifying key risk influencers but also serves as a valuable resource for enhancing risk-management strategies and informing driver education programs.

Figure 12. Importance of features.

6. Conclusions

(1) Potential field theory is applied to the traffic system, analogous to establishing the vehicle interaction potential field function based on intermolecular interactions. This approach led to the development of a safe potential field car-following model that integrates lane, road boundary, and vehicle potential fields. By introducing acceleration as a variable in the vehicle potential field, its fluctuations directly influence the field’s distribution, enabling the model to reflect the safety risks encountered during driving. This provides a foundational framework for promoting safer vehicle operations.

(2) Based on the driver’s continuous time series, two categories of indicators were built to extract the short-term driving style inclination during the following process of interacting vehicles in different roles. The short-term driving style inclination features were integrated into the quantification indicators of the following risk, combined with real-time risk exposure and real-time risk severity. A quantification model of the following risk considering collision potential and severity, as well as the additional collision probability of the surrounding vehicle drivers’ driving style inclination, is established. According to fuzzy c-means clustering, the threshold for risk level
division is determined by dividing the following risk into four levels: safe, low risk, medium risk, and high risk.

(3) Leveraging highD dataset insights and focusing on the interaction risks with surrounding vehicles, the LightGBM algorithm was applied for real-time prediction of follow-through vehicle risk. Achieving a recognition accuracy of over 86% for medium to high-risk identifications, this application facilitates the scientific quantification of behavioral risks in the following scenarios, taking into account driving style propensities. The outcomes offer nuanced guidance for enhancing follow-warning systems and the overall efficacy of vehicle driving-assistance technologies.

(4) The exploration of risks associated with vehicle following is a burgeoning area of interest, dealing with multi-vehicle interactions within complex driving contexts. This study provides a comparative and systematic assessment of follow-through risks by modeling behaviors within a safety potential field, acknowledging driver diversity, and developing risk indicators. Utilizing the highD dataset, the research faced limitations due to the specific conditions under which data were collected—weekday hours from 8 am to 7 pm, in clear and calm weather. While this study accounts for the behavior of both the ego vehicle and its surroundings, external factors like weather and road types also play a crucial role in driving behavior, marking areas for future enhancement. Future research aims to employ a more comprehensive dataset enriched with environmental details and extended observation periods for a deeper analysis.

Author Contributions: Conceptualization, K.W. and D.Q.; methodology, K.W.; software, Y.Y.; validation, K.W., D.Q. and T.W.; investigation, S.D.; writing—original draft preparation, K.W.; writing—review and editing, D.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China [grant number: 52272311]. “Modeling of driving safety potential field and resolving longitudinal & lateral dimensions steady-state response mechanism of vehicle clusters based on molecular dynamics in connected environment” and Qingdao West Coast New District College President’s Fund: [grant number: GXXZJ202301]. “Carbon Fiber Fully Wound with Plastic Lining for Automotive Use High Pressure Gas Cylinder (Type IV Cylinder) R&D”.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Data are available on request owing to restrictions such as privacy or ethics.

Conflicts of Interest: This paper is original and does not contain any content that violates laws and regulations or infringes on the rights of others.

References
7. Hei, K. Research on Vehicle Interactive Behavior Characteristics and Models Based on Driving Style; Qingdao University of Technology: Qingdao, China, 2021.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.