Dynamic Bayesian network approach to evaluate vehicle driving risk based on on-road experiment driving data

YANLI MA¹, SHOUMING QI¹,², LUYANG FAN¹, WEIXIN LU³, CHING-YAO CHAN² and YAPING ZHANG¹

¹School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin 150090, China.
²California PATH, University of California, Berkeley, CA 94804, USA
³School of Electronic Engineering and Computer Science, Peking University, Beijing 100871, China

Corresponding author: YANLI MA (e-mail: mayanli@hit.edu.cn).

This research was supported by National Key Research and Development Program of China (2017YFC0803901), Heilongjiang highway survey and design institute project (2018006), National Natural Science Foundation of China (51108136) and China Scholarship Council Scholarship (201806120277).

ABSTRACT In this work, we utilize a dynamic Bayesian network for an inferential analysis of driving-related risks based on our assessment of real-world driving data. On-road driving tests are carried out to gather and analyze data to identify different evaluation indicators of driving-related risks. These indicators include the distance between vehicles, acceleration, steering entropy, visual distraction duration, visual glance speed, and blink frequency. Moreover, these indicators are processed to build a driving-related risk evaluation model based on the dynamic Bayesian network. The validity of this model is further tested by experimental data. The results show that this model can achieve a reasonable quantitative evaluation of driving-related risks. Vehicle operation-related risks can be further divided into four levels of safety, namely, levels I to IV. The lowest risk is observed at level I, whereas level IV has the highest risk. Among the indicators for risk evaluation, the distance between vehicles is the most sensitive control indicator of vehicle operation-related risks. The research findings provide various methodologies that could be utilized for evaluation and early warning of driving-related risks. Thus, a theoretical foundation to seek solutions for the safety of advanced driving assistance system is formed.

INDEX TERMS On-road Vehicle Driving, Human Factors, Driving Risk Analysis, Data Analysis, Dynamic Bayesian network, Road transportation, Quantitative Evaluation, Sensitive Control Indicator

I. INTRODUCTION
Road traffic crashes are the ninth leading cause of death, which account for 2.2% of all recorded deaths globally [1]. The number of road traffic-related deaths continues to rise, reaching 1.35 million in 2016 [2]. Meanwhile, China’s vehicle ownership and traffic accidents also increase annually. In 2016, China’s vehicle ownership reached 300 million, and the number of traffic fatalities was approximately 63,000 (The number is much higher, per this report [3]). Therefore, monitoring the running status of vehicles can effectively improve road traffic safety. The Advanced Driver Assistance System (ADAS) is a vehicle intelligence system that provides operational support for drivers. The existing ADAS mainly works in the hedging phase. If the abnormal state of the vehicle is found at an early stage by the ADAS, a warning can be given in advance to allow the driver sufficient reaction and operation time. Thus, traffic accidents can potentially be avoided or mitigated.

Guo and Fang [4] have found that the driving-related risks of a driver are correlated to his or her accident rate and personality traits. In addition, Simon et al. [5] have observed that a driver’s inadequate driving experience and the performance of a secondary task while driving contribute to the risk of vehicle collision. Rosey [6] has reported that vehicle lateral movement is closely related to the state of drivers. For instance, driver’s fatigue leads to frequent lane changes. Bargman [7] has compared the impact of performing secondary tasks on traffic safety while driving versus eye glance behavior during pursuit. Between the two, eye glance behavior has been found to have a greater impact. In addition, Lethaus [8] has suggested that eye movements occur earlier than the maneuvers performed by drivers, thus making
driving behaviors predictable. Ma [9] has also considered the indicators of visual variation duration while driving, such as adjusting the speed, acceleration, lateral displacement, steering wheel rotation speed, and lateral position of the vehicle. The evaluation of distraction risks was conducted based on an approach by support vector machine (SVM), Kircher [10] has introduced the driving behavior indicators observed using an eye-tracking algorithm, which aid in improving the accuracy of driving-related risk evaluation. Meanwhile, Xu [11] has analyzed the correlation between traffic flow and risks, assessing driving-related risks under different traffic flows. Gershon [12] has studied the collected driving data of various young people and has analyzed the interactions between the risk factors of driving and other significant factors.

Among the methods of driving-related risks evaluation, Ng [13] has applied clustering analysis and regression analysis to evaluate road accident risks. Using vehicle trajectory data, Li [14] has built a rear-end collision risk model. In addition, Xu [15] has comprehensively utilized the theories of gray clustering, fuzzy consistency, and analytic hierarchy process to propose a new systemic approach to road safety evaluation. Meanwhile, Xiong [16] has described a prediction algorithm for driving-related risks based on the Markov chain. Wang [17] has presented the concept of driving-related risks, which has later been applied to various complex driving scenarios. Cheol [18] has used the Bayesian network to develop a warning system for traffic emergencies, which demonstrates the power of the Bayesian network in probability analysis.

In addition, other traditional techniques should also be considered, such as matched case-control logistic regression [19, 20], artificial intelligence models [21, 22], and Bayesian logistic regression [23]. The existing evaluation methodologies have mainly addressed the status of macroscopic traffic operations with efforts to focus on road safety risk and accident risk evaluations. Certain scholars have dedicated their research to on-road driving studies under different road conditions and certain deterministic driving tasks [24–26]. Among the identified influence factors of risk are drivers’ individual characteristics, eye movement behavior, glance behavior, traffic flow, vehicle maneuver behavior, and operation status. However, the interactions among the influence factors of driving-related risks tend to be neglected.

The Bayesian network is built upon the interactions between variables, and it obtains result distribution by parameter learning and probabilistic reasoning. So far, the Bayesian network has been successfully applied to the risk analysis of aviation system reliability in search and rescue [27] and to the environmental evaluation for operators in the nuclear industry [28, 29]. Wang has applied dynamic Bayesian networks to real-time collision risk assessment of urban expressways [30]. In addition, Guo has studied semi-parametric Bayesian models for evaluating time-variant driving-related risk factors and has used a case-crossover approach to evaluate driver-behavior risks [31].

The assessment of vehicle operating risk is the basis and premise of safety assistance. Appropriate safety assistance can only be carried out after the vehicle risk is correctly assessed. Research on the methods of vehicle safety estimation is also of great significance for improving the core competencies of intelligent transportation vehicles as well as improving the performance of ADAS. In this study, we have applied the dynamic Bayesian network (DBN) to the evaluation of driving-related risks. Data relevant to driving-related risk indicators have been collected in an on-road driving setting. Then, these indicators are utilized to construct a driving-related risk evaluation model based on the Bayesian network. The procedures described in this study for driving-related risk evaluation lay the basis for early assessment and pre-warning of driving-related risks.

II. Experimental Design

A. Participants

On-road driving tests were performed on 24 participants, and the effective sample is 20 (12 males and 8 females) with different ages (Mean = 34.65, SD = 10.4) and years of driving experience (Mean = 7.5, SD = 7.45). This experiment was organized by the School of Transportation Science and Engineering, Harbin Institute of Technology. Recruitment details were posted at different schools, taxi companies, and government agencies to recruit suitable drivers. All participants were physically healthy without prior consumption of stimulating and influential substances, such as alcohol and coffee, 24 hours before the test. The basic information description of the participants is shown in Table 1.
### B. Apparatus
Experimental car was a 2017 Sagitar 1.6T. The experimental equipment included a suite of sensors installed on a passenger car: an eye tracker, acceleration sensors (lateral acceleration and longitudinal acceleration), a steering wheel sensor, a telemeter, a camera, and a driving recorder. The arrangement of the experimental equipment is shown in Figures 1 and 2.

![Figure 1: Layout of sensor inside the vehicle](image)

![Figure 2: Layout of instrument inside the vehicle](image)

### C. Experimental Route
The test was conducted from the intersection of Xinyang Road and Anfa Street to the intersection of Wenchang Street and Yanxing Road in Harbin City, as shown in Figure 3. The road segment from Anfa Street to Xiangzheng Street had an overall length of 3 km on the trunk road. The road segment from Xiangzheng Road to Wenchang Street was the secondary trunk road, which also had an overall length of 3 km. The experiments were conducted during non-peak hours (normally 9:30 am–11:00 am; 2:30 pm–4:00 pm) under favorable weather conditions (sunny, suitable temperature for driving). Road segments with stable traffic flow, ideal road conditions, and little traffic disturbance were chosen for the test.

![Figure 3: Experimental Route](image)

### D. Experimental procedure
The experimental procedure was as follows:
1. The participants were asked to fill out a personal information form, and the staff provided them with basic instructions before starting the test.
2. The participants were instructed to wear significant test instruments (e.g., eye tracker) and adequate time to familiarize themselves with the test vehicle to perform adaptive driving experiments. This process took approximately half an hour.
3. The driver performed the driving task at the experimental site, whereas the driving video and other data were recorded. The driver then proceeded along the route according to the instructions. If no special or unprecedented situation emerged, then the route could not be changed without authorization. However, depending on the actual road conditions, lane changes could be performed. If the driver perceived that his or her driving was at risk, he or she could bypass the instructions of the guide and perform safe operations. The driving experiment lasted approximately an hour.
4. Upon completion of the driving task, the staff sorted out and analyzed the driving data. Then, they calibrated the driving risk behaviors. During the calibration process, risky driving behaviors were defined.

### III. Data processing
In this section, we describe the data processing steps for the on-road driving data. We aimed at generating effective, reliable, and sensitive evaluation indicators of driving-related risks. Consequently, the study found that risky driving could affect driving behavior and reduce the driver’s ability to control the vehicle, such as the increase in the number of emergency braking and the increase in the standard deviation of the steering wheel angle. When the driver accepted the execution of different difficulty levels, the vehicle operating parameters, such as vehicle trajectory, speed, and longitudinal and lateral acceleration, changed significantly [32, 33].

Considering the significance of vehicle control and driver behavior to driving-related risks, indicators for vehicle operation and driver’s eye movement were collected. The data were used as the judgment variables of vehicle operation risks. The data collected are shown in Table 2.
Table 2 Main indicator data collected by the tests

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Indicator</th>
<th>Unit</th>
<th>Meaning of the parameter</th>
<th>Method of data acquisition</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>Time</td>
<td>s</td>
<td>Start and end time of driving</td>
<td>Driving recorder</td>
</tr>
<tr>
<td>VD</td>
<td>Distance between vehicles</td>
<td>m</td>
<td>Distance from the vehicle in front</td>
<td>Laser ranger finder</td>
</tr>
<tr>
<td>a</td>
<td>Acceleration</td>
<td>G</td>
<td>Vehicle’s acceleration</td>
<td>Acceleration sensor</td>
</tr>
<tr>
<td>ø</td>
<td>Steering angle</td>
<td>°</td>
<td>Rotation angle of the steering wheel</td>
<td>Steering wheel sensor</td>
</tr>
<tr>
<td>BF</td>
<td>Blink frequency</td>
<td>time/s</td>
<td>Blink pattern</td>
<td>Eye tracker</td>
</tr>
<tr>
<td>GS</td>
<td>Glance speed</td>
<td>deg/s</td>
<td>Glance behavior</td>
<td>Eye tracker</td>
</tr>
<tr>
<td>DT</td>
<td>Visual distraction duration</td>
<td>s</td>
<td>Eye-off-road duration</td>
<td></td>
</tr>
</tbody>
</table>

A. VEHICLE MANEUVER INDICATORS

(1) Minimum distance between vehicles

According to the dynamic formula, two vehicles are located at $X_i$ and $X_{i+1}$ before brake deceleration, and the distance between the vehicles is $D_i$. The front vehicle starts to brake, and the distance in which the $i$–$I$-th vehicle brakes is $[v_i(t)\tau]^2/2\alpha_i(t)$. The distance traveled by the $i$-th vehicle during the reaction time $\tau$ is $v_i(\tau)\tau$, and the distance traveled during the braking process is $[v_i(t)]^2/2\alpha_i(t)$. The formula is as follows:

$$v_i(t)\tau + [v_i(t)]^2/2\alpha_i(t) + d_0 = D_i + [v_i(t)]^2/2\alpha_i(t).$$  \hspace{1cm} (1)

The distance between two vehicles is

$$D_i = v_i(t)\tau + [v_i(t)]^2/2\alpha_i(t) - [v_{i+1}(t)]^2/2\alpha_{i+1}(t) + d_0.$$  \hspace{1cm} (2)

The two vehicles are assumed to have the same speed and the same latency for the braking to take effect. The minimum distance between the vehicles is shown in Formula (3). The vehicle braking process is shown in Figure 4.

$$D_{\text{min}} = v_i(t)\tau + d_0.$$  \hspace{1cm} (3)

where $v_i$ is the driving speed of the $i$-th vehicle before braking begins; $\tau$ is the sum of driver’s reaction time and latency for the braking to take effect; $d_0$ is the minimum distance between vehicles after the vehicle has completely stopped.

![Figure 4. Schematic diagram of the vehicle braking process](image)

Assuming that the driver’s reaction time is 0.8 s, the latency for the braking to take effect is 0.2 s, and the $d_0$ is 3 m. The minimum distances between the two vehicles under different speeds are shown in Table 3.

Table 3 Minimum distance between vehicles under different speeds

<table>
<thead>
<tr>
<th>Speed (km/h)</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum distance (m)</td>
<td>5.78</td>
<td>8.56</td>
<td>11.34</td>
<td>14.11</td>
<td>16.89</td>
<td>19.67</td>
<td>21.44</td>
<td>25.22</td>
</tr>
</tbody>
</table>

The ratio of the minimum distance to the actual distance between the two vehicles during driving $\xi(t) = D_{\text{min}}/D(t)$ is used to assess drivers’ performance in controlling the distance between the vehicles. This indicator is one way of measuring driving-related risks.

(2) Acceleration

The temporal variation characteristics of acceleration during on-road driving are shown in Figure 5. The acceleration percentile of the vehicle is shown in Table 4.

![Figure 5. Temporal variations of acceleration during driving](image)

As shown in Table 4, the probability that the acceleration under $-0.6470$ m/s² or over $0.7226$ m/s² is relatively small (10%). During the driving process, certain driving-related risks exist if the acceleration falls within this range.

(3) Steering entropy

Steering entropy is a measure of the drivers’ ability in operating the steering wheel with smoothness and anticipation of the need to turn the steering wheel. In the relevant literature, steering entropy has been proposed to quantify the drivers’ steering control capacity.

$$SE = \sum_{i=1}^{n} - p_i \log_2(p_i),$$  \hspace{1cm} (4)

where $SE$ is the steering entropy, and $p_i$ is the distribution probability of the prediction errors falling within the error interval.

Figure 6 and Table 5 present the steering entropy for normal driving and risky driving from on-road driving tests on 10 drivers under the age of 30 (We studied the driving behavior of young drivers for the time being, and we...
analyzed data for participants who were under the age of 30.

Except for driver No. 9 (errors resulting from the experimental process), the remaining drivers all showed an apparent increase in the steering entropy (both mean and median.) Thus, the steering entropy was directly proportional to the driving-related risks.

FIGURE 7. Scatter plot of visual deviation duration

As seen from Table 6, the 90th percentile of visual deviation duration was 0.24 s, and the maximum duration was 1.0 s. When the visual deviation duration was less than 1.0 s, the drivers were engaged in collecting the information necessary for the driving task. However, when it was above 1.0 s, the drivers were postulated to exhibit unnecessary eye movements.

(2) Glance speed

The statistics on glance speed during the driving process are shown in Fig. 8. Drivers who came across certain important movements.

(3) Blink frequency

Blink frequency reflects the driver’s level of brain activity and sensitivity to environmental information. The mean blink frequencies under three statuses of the drivers are shown in Table 7. Evidently, dozing off while driving significantly increased driving-related risks.
IV. Modeling

Representation and inference of Bayesian networks have strict mathematical foundations. Based on probability theory, it examines the statistical regularity of interdependence between multiple variables in objective things, which is very suitable for describing complex systems. The multiple and uncertainty relationship between the representation of events and the situation is one of the most effective theoretical models in the field of uncertain knowledge and reasoning. Therefore, the Bayesian network is used to evaluate and predict vehicle operation risks.

A. Bayesian network structure

The risk factors of driving are used as the assessment indicators of driving-related risks. The structure of our proposed Bayesian network model for driving-related risks is shown in Figure 9.

![Figure 9. BN structure of driving risk](image)

As shown in Figure 9, the hidden nodes are driving-related risk (R), vehicle operation risk (OR), and eye movement-characterized risk (ER), all of which are represented by $X_i$. Six nodes, namely, steering rotation entropy (SE), acceleration variation rate ($\xi(t)$), control index of vehicle spacing ($\zeta(t)$), visual deviation duration (DT), glance speed (GS), and blink frequency (BF), are observational nodes of the status, which is represented by $Y_i$.

The probabilities of hidden nodes R, OR, and ER are given below.

\[
P(X_1, X_2, X_3, \ldots, X_t, X_{t+1}, \ldots, X_T | Y_1, Y_2, \ldots, Y_{t-1}, Y_{t+1}, \ldots, Y_T) = \prod_{i=1}^{T} P(Y_i | \text{parent}(Y_i)) \prod_{i=1}^{T} P(Y_i | \text{parent}(Y_i))
\]

(5)

From the above,

\[
P(X_1 | X_2, X_3, Y_1, Y_2, \ldots, Y_T) = \prod_{i=1}^{T} P(Y_i | \text{parent}(Y_i))
\]

(6)

where $i \in [1, 2], j \in [1, 6], m \in [2, 3]$.

The status at a specific time point in the static model is chosen for status transfer. Then, in continuous time, the probability of node transfer can be represented by the probability of the nodes at two time points. Let the previous node status be $X_t$, then the node status after time $T$ is $X'$. $P(X' | \text{parent}(X'))$ is the conditional probability distribution of node in $X'$. Let the transfer probability be $\beta$, then

\[
\beta = P(X' | X_t) = \prod_{i=1}^{T} P(X'_i | \text{parent}(X'_i))
\]

(7)

B. Dynamic Bayesian network model

The dynamic Bayesian network model of driving-related risks is built upon temporal variation characteristics, as shown in Fig. 10.

![Figure 10. DBN for driving risk based on temporal variations](image)

The model for driving-related risk evaluation can be represented by a collection of static Bayesian networks for each time point within time $T$. The hidden sequence from the initial 1 to $t$ is represented by $X_t$, whereas the observation sequence is represented by $Y_t$. At time $t$, the status value of the variable for the $i$-th hidden node is represented by $x_t^i$, and that of the variable for the $j$-th observational node is represented by $y_t^j$.

With all of the observational nodes given, calculating the probability distribution of hidden nodes corresponding to the existing data is the core of dynamic model inference.

\[
P(x_t^1, x_t^2, x_t^3, \ldots, x_t^T, y_t^1, y_t^2, \ldots, y_t^6) = \prod_{i=1}^{T} P(y_t^i | \text{parent}(y_t^i)) \prod_{i=1}^{T} P(y_t^i | \text{parent}(y_t^i))
\]

(8)

According to independence assumption,

\[
P(x_t^1, x_t^2, x_t^3, \ldots, x_t^T, y_t^1, y_t^2, \ldots, y_t^6) = \prod_{i=1}^{T} P(x_t^i | \text{parent}(x_t^i)) \prod_{i=1}^{T} P(x_t^i | \text{parent}(x_t^i))
\]

(9)

From the above formula, the joint probability of risk for time $T$ is derived

\[
P(x_t^1, x_t^2, x_t^3, \ldots, x_t^T, y_t^1, y_t^2, \ldots, y_t^6) = \prod_{i=1}^{T} P(y_t^i | \text{parent}(y_t^i)) \prod_{i=1}^{T} P(y_t^i | \text{parent}(y_t^i))
\]

(10)

\[
P(x_T^1, x_T^2, x_T^3, \ldots, x_T^T, y_T^1, y_T^2, \ldots, y_T^6) = \prod_{i=1}^{T} P(y_T^i | \text{parent}(y_T^i)) \prod_{i=1}^{T} P(y_T^i | \text{parent}(y_T^i))
\]

(11)
Formula (12) can calculate the probability risk at the initial time point, where  \( i \in [0, T], j \in [1, r], k \in [1, 3], n \in [1, T], m \in [2, 3] \).

\[
P(x_i^0) = \frac{P(x_i^0 \vert x_i^1, \ldots, x_i^r)}{\prod_{n} P(x_i^0 \vert \text{parent}(x_i^0))} \quad (12)
\]

V. Results and Analysis

A. Model inference

We selected the threshold based on the percentiles in the statistical analysis of relevant factors. The occurrence probabilities of each observational node during on-road driving are shown in Tables 8 and 9.

Table 8 Occurrence probabilities of vehicle operation indicators under on-road driving

<table>
<thead>
<tr>
<th>Threshold value (( \xi(t) ))</th>
<th>Probability (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0, 0.8)</td>
<td>50%</td>
</tr>
<tr>
<td>(0.8, 0.85)</td>
<td>10%</td>
</tr>
<tr>
<td>(0.85, 1)</td>
<td>15%</td>
</tr>
<tr>
<td>(1, +( \infty ))</td>
<td>25%</td>
</tr>
</tbody>
</table>

Table 9 Occurrence probabilities of eye movement-characterized indicators under on-road driving

<table>
<thead>
<tr>
<th>Threshold value (( [a] ))</th>
<th>Probability (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0, 0.5)</td>
<td>80%</td>
</tr>
<tr>
<td>(0.5, 0.6)</td>
<td>15%</td>
</tr>
<tr>
<td>(0.6, 0.7)</td>
<td>13%</td>
</tr>
<tr>
<td>(0.7, +( \infty ))</td>
<td>2%</td>
</tr>
</tbody>
</table>

Table 10 Conditional probabilities of inferred risk at each observational node

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Threshold value (( \xi(t) ))</th>
<th>Probability (%)</th>
<th>Indicator</th>
<th>Threshold value (([a]))</th>
<th>Probability (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DT</td>
<td>(0, 0.2)</td>
<td>85%</td>
<td>GS</td>
<td>(440, 1000, 2500, +( \infty ))</td>
<td>95%</td>
</tr>
<tr>
<td></td>
<td>(0.2, 0.4)</td>
<td>13.99%</td>
<td></td>
<td>(440)</td>
<td>4.89%</td>
</tr>
<tr>
<td></td>
<td>(0.4, 1)</td>
<td>1%</td>
<td></td>
<td>(1000)</td>
<td>0.1%</td>
</tr>
<tr>
<td></td>
<td>(1, +( \infty ))</td>
<td>0.01%</td>
<td>BF</td>
<td>(0.8, 0.9)</td>
<td>4%</td>
</tr>
<tr>
<td></td>
<td>(0.2, 0.5)</td>
<td>16%</td>
<td></td>
<td>(0.44)</td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>(0.5, 1)</td>
<td>50%</td>
<td></td>
<td>(2000)</td>
<td>30%</td>
</tr>
</tbody>
</table>

Netica is a Bayesian network analysis software. It is mainly used for system risk analysis and simulation modeling for system failure. The driving-related risk estimate was obtained from the Bayesian network model using Netica. The conditional probabilities of hidden nodes, vehicle operation risk, eye movement-characterized risk, and the final risk value were calculated. Figure 11 shows the workflow of driving-related risk evaluation. Y indicates risk, whereas N means non-risk. Risk levels were divided based on risk values.

**FIGURE 11. Bayesian network model developed by Netica**

Based on the China National Emergency Response Plan for Public Emergencies, the risk levels were classified into low risk, moderate risk, fairly high risk, and high risk according to the degree of hazard, urgency, and development potential of public emergencies. Depending on the occurrence probability of driving-related risks and the actual operation of the vehicle, the percentage method was used to distribute the probabilities of the four types of risks evenly:

- Level I (low risk, 0–25%), Level II (moderate risk, 25–50%), Level III (fairly high risk, 50–75%), and Level IV (high risk, 75–100%).

B. Sensitivity analysis

To assess the influence of each observational node, a sensitivity analysis was conducted on each risk. After determining the occurrence probability of risks for each node, the risk factors were ranked in decreasing order of sensitivity, as shown in Table 11.

In terms of mutual information (which could reflect the degree of influence between factors), VD, BF, \( \alpha \), SE, DT, and GS declined successively. The degree of influence of each node on a specific risk declined. The distance between vehicles was the most important factor and was the most sensitive among driving-related risks.
To verify whether the data was accurate and whether the analysis function was consistent with the actual situation, the general sensitivity analysis method was utilized. This method is the general algorithm for changing the parameter value to find evident changes. If the factors of the same degree (such as a 10% increase in risk) changed based on the original probability of occurrence, then sensitivity could be judged by observing the influence of each edge probability on the final node.

As shown in the Figure 12, the sensitivity of VD, BF, and $\alpha$ was successively reduced. This conclusion was consistent with the conclusion of Netica sensitivity analysis, proving that the sensitivity analysis made by Netica was reasonable.

### Table 11 Sensitivity analysis

<table>
<thead>
<tr>
<th>Node</th>
<th>Mutual info</th>
<th>Percent</th>
<th>Variance of beliefs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk</td>
<td>0.968</td>
<td>100</td>
<td>0.2389</td>
</tr>
<tr>
<td>OR</td>
<td>0.159</td>
<td>16.400</td>
<td>0.0516</td>
</tr>
<tr>
<td>ER</td>
<td>0.117</td>
<td>12.100</td>
<td>0.0369</td>
</tr>
<tr>
<td>VD</td>
<td>0.029</td>
<td>3.030</td>
<td>0.0095</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.006</td>
<td>0.604</td>
<td>0.0019</td>
</tr>
<tr>
<td>SE</td>
<td>0.002</td>
<td>0.197</td>
<td>0.0006</td>
</tr>
<tr>
<td>DT</td>
<td>0.001</td>
<td>0.128</td>
<td>0.0004</td>
</tr>
<tr>
<td>GS</td>
<td>0.000</td>
<td>0.015</td>
<td>0.0001</td>
</tr>
<tr>
<td>BF</td>
<td>0.014</td>
<td>1.400</td>
<td>0.0043</td>
</tr>
</tbody>
</table>

The evidence variable of a certain situation was input into the Bayesian network, whereas the known state of the evidence variable was used to solve the posterior probability problem of other nodes. In Netica, if the evidence variable was deterministic, setting its state to 100% and updating the probability of the entire network would allow researchers to observe the probability change of the relevant node. As shown in Figure 13, assuming that the driver has poor speed control, the risk of malicious acceleration and deceleration significantly increases from 60.5% to 74.4%.

### FIGURE 12. Risk sensitivity analysis

#### C. Posterior probability inference

(1) Risk prediction

(2) Causality inference

Another important application of the Bayesian network is the systematic cause diagnosis. The Bayesian network can perform a two-way inference, which is not only able to calculate the probability of system risk under the joint fault condition of each node but can also calculate the posterior probability of each node under system fault conditions. These conditions can easily find the most likely cause of system failure, thus making an intuitive and convenient analysis.

Assuming that a risk must occur, the state probability was 100%. Figure 14 shows that after the evidence is input through the automatic update function of Netica, the safety of the vehicle operation and the safety of the eye movement representation significantly decreases. The most evident change among the indicators was the observed range of the vehicle distance control index, which was greater than 1 with a significant increase from 25% to 30%. This suggested that, in the absence of other evidence, the most likely cause of the risk was the reduction in vehicle spacing.

### FIGURE 13. Changes in acceleration lead to changes in risk

### FIGURE 14. Changes in risk status lead to network changes

### VI. Discussions and Limitations

#### A. Effectiveness determination of DBN model

A period of risky driving under on-road driving conditions was compared with that of normal driving, with both periods lasting for 10 s. The results of the indicator values and
driving-related risk inference are shown in Tables 12 and 13, respectively.

### Table 12 Indicator of risk driving in the time period

<table>
<thead>
<tr>
<th>Vehiciles</th>
<th>Acceleration deviation (m/s²)</th>
<th>Steering deviation entropy (Deg/s)</th>
<th>Glance speed frequency probability (Time/s)</th>
<th>Blink risk probability (%)</th>
<th>Risk level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (s)</td>
<td>Control indicator</td>
<td>Duration (ms)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2732</td>
<td>0.0728</td>
<td>0.3663</td>
<td>0.28</td>
<td>595.70</td>
<td>1</td>
</tr>
<tr>
<td>0.4791</td>
<td>0.0839</td>
<td>0.3663</td>
<td>0.28</td>
<td>595.70</td>
<td>1</td>
</tr>
</tbody>
</table>

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/
16(a). The separate figures for the four types of risks are shown in Fig. 16 (b–e).

**FIGURE 16. ROC curves for four risk states**

Figure 16(a) shows that all the points are concentrated in the upper left portion, wherein the experimental accuracy appears fitting, indicating that the model learning effect is also sound. At the same time, the four types of risks were analyzed. Additionally, the identifications of risks II and III were better, whereas the effect of risk IV was worse. In the future, the Bayesian network could be deduced by adding more risk indicators for a better learning of the model.

**C. Limitations**

1. At present, the parameters determination method of Bayesian network is mainly based on expert systems and parameter learning, which more or less deviates from the actual situation.
2. The network topology determined based on expert knowledge is relatively simple. The determination of the topology of relatively complex systems and its structural learning are subject to further study.
3. Given the limitations of test conditions and equipment, the selection factors of variables and prediction models also have limitations. Thus, comprehensive consideration should be added in future research, such as driver’s personality, vehicle type, and deviation from the center of the line. At the same time, more samples should be collected for reinforcement learning, which improves the accuracy and precision of the model.
4. The division of driving risk levels needs to be more objective, and future research should combine the established ANN model to carry out the risk division of clustering.

**VII. Conclusions**

1. The on-road driving test was conducted to obtain data on the following indicators of driving-related risks: control indicator distance between vehicles, acceleration, steering entropy, visual deviation duration, glance speed, and blink frequency. The occurrence probability of the driving risks was analyzed based on these indicators. The thresholds of each indicator for each risk level were also determined.
2. A dynamic Bayesian network model was developed to assess driving-related risks. The results showed that the proposed model could quantitatively describe the driving-
related risks under the influence of different dynamic factors. The dynamic Bayesian network model was used to calculate and infer driving-related risks via Netica in a spectrum of safe to unsafe driving at four different risk levels. Sensitivity analysis indicated that the distance between vehicles was most sensitive among the driving-related risks.

(3) The periods of normal and risk driving were compared under on-road driving conditions. The dynamic Bayesian network model was further verified for its capability in evaluating driving-related risks. Constructing the ANN model to gather the dynamic Bayesian network risk derivation results showed that the learning effect was fitting but was also subject to the selection of indicators and other factors. Alternatively, the prediction of risk levels II and III were more accurate.

ACKNOWLEDGMENT

The authors had the gratitude to the reviewers and appreciated the valuable comments from the doctoral students of Institute of Traffic Engineering and colleagues at UC Berkeley PATH.

REFERENCES

YANLI MA received the B.S. degree (in 1997) and M.S. degree (in 2002) in Traffic Engineering from Harbin Institute of Technology, Harbin, China and got the Ph.D. degree in Road & Railway Engineering from Harbin Institute of Technology, Harbin, China, in 2007.

From 1997 to 2007, she was a Research Assistant in Department of Traffic Engineering, Harbin Institute of Technology, China. She was an Associate Professor since 2007. Her research interests include Road Traffic Safety Theory and Technology, Driver Behavior and Human-computer Interaction, Driving Distraction Characteristic and Secondary Tasks. She is a member of American Society of Civil Engineering, and a member of Easter Asia Society for Transportation Studies.


CHING-YAO CHAN received the Ph.D. degree in mechanical engineering from the University of California at Berkeley (UC Berkeley), Berkeley, CA, USA, in 1988.

He is a Research Engineer with the California Partners for Advanced Transportation Technology (PATH), UC Berkeley, which he joined in 1994. Since 2009, he has been serving as the Program Leader with the Transportation Safety Research Area, California PATH. His research interests include the development of Driver-assistance Systems, Evaluation of Sensing, Wireless Communication Technologies to Vehicular Safety Systems, and Highway Network Safety Assessment. He possesses in-depth knowledge in specialty areas of safety systems and technology use for transportation applications.

SHOUMING QI received the B.S. degrees in Civil Engineering from Ludong University, Yantai, china, in 2013 and the master degree in Transportation Engineering from Kunming University of Science and Technology, Kunming, China, in 2016. He is currently pursuing a Ph.D degree in Traffic Engineering at Harbin Institute of Technology, Harbin, China since Sep. 2016, and he is still a visiting scholar at UC Berkeley PATH since Sep. 2018.

His research interests include Human-machine driving in Autonomous vehicle, machine learning, Characteristics of Driver's Performance, Characteristics of Driving Distraction, Traffic Congestion Problems. He won the China National Master Scholarship in 2015.

LUYANG FAN received the B.S. degree in transportation engineering from Harbin Institute of Technology, Harbin, China, in 2016, and the master’s degree in transportation engineering from Harbin Institute of Technology, Harbin, China, in 2018. She is currently working as an engineer in Chengdu. Her main research focus on Driver Behavior and Traffic Safety.

WEIXIN LU is working towards the Bachelor degree in Computer Science from Peking University, Beijing, China. He is an undergraduate in School of Electronic Engineering and Computer Science, Peking University. His research interests include Graph Database, Graph Embedding and Reinforcement Learning.

YAPING ZHANG received the B.S. degree in Photogrammetry and Remote Sensing from Wuhan University, China, in 1988; the M.S. degree in Road & Railway Engineering from Hunan University, China, in 1999 and the Ph.D. degree in Road & Railway Engineering from Harbin Institute of Technology, China in 2005.

From 1988 to 2000, he was a Lecturer in Changsha Transportation College. From 2001 to 2004, he was an Associate Researcher in Changsha University of Science & Technology. He is a professor in Harbin Institute of Technology, which he joined in 2005. His research interest includes Traffic Design & Planning, Transportation Safety, Traffic Simulation, Logistic Engineering, 3S Technology.