A recognition model of driving risk based on Belief Rule-Base methodology

Sun, Chuan; Wu, Chaozhong; Chu, Duanfeng; Lu, Zhenji; Tan, Jian; Wang, Jianyu

DOI
10.1142/S0218001418500374

Publication date
2018

Document Version
Final published version

Published in
International Journal of Pattern Recognition and Artificial Intelligence

International Journal of Pattern Recognition
and Artificial Intelligence
Vol. 32, No. 11 (2018) 1850037 (23 pages)

DOI: 10.1142/S0218001418500374

A Recognition Model of Driving Risk Based on Belief Rule-Base Methodology

Chuan Sun[*,§], Chaozhong Wu[†,¶], Duanfeng Chu[†,‖], Zhenji Lu[‡,**],
Jian Tan[*,††] and Jianyu Wang[*,‡‡]

[*]*College of Transportation, Huanggang Normal University*
Huanggang 438000, P. R. China

[†]*Intelligent Transport Systems Research Center*
Wuhan University of Technology, Wuhan 430063, P. R. China

[‡]*Faculty of Mechanical, Maritime and Material Engineering*
Delft University of Technology, Mekelweg 2
2628 CD Delft, The Netherlands
[§]*sunchuan33@foxmail.com*
[¶]*wucz@whut.edu.cn*
[‖]*chudf@whut.edu.cn*
[**]*z.lu@tudelft.nl*
[††]*2529011569@qq.com*
[‡‡]*191253016@qq.com*

Received 24 January 2018
Accepted 28 March 2018
Published 29 May 2018

This paper aims to recognize driving risks in individual vehicles online based on a data-driven methodology. Existing advanced driver assistance systems (ADAS) have difficulties in effectively processing multi-source heterogeneous driving data. Furthermore, parameters adopted for evaluating the driving risk are limited in these systems. The approach of data-driven modeling is investigated in this study for utilizing the accumulation of on-road driving data. A recognition model of driving risk based on belief rule-base (BRB) methodology is built, predicting driving safety as a function of driver characteristics, vehicle state and road environment conditions. The BRB model was calibrated and validated using on-road data from 30 drivers. The test results show that the recognition accuracy of our proposed model can reach about 90% in all situations with three levels (none, medium, large) of driving risks. Furthermore, the proposed simplified model, which provides real-time operation, is implemented in a vehicle driving simulator as a reference for future ADAS and belongs to research on artificial intelligence (AI) in the automotive field.

Keywords: Driving data; vehicle driving risk; data-driven; belief rule-base; ADAS.

1. Introduction

With the rapid growth of vehicle volume worldwide, the number of road traffic accidents is increasing correspondingly. Driving risk identification is one of the key

[§]Corresponding author.

technologies to improve driving safety. As vehicle state estimation and prediction are largely improved in recent studies, the precise evaluation of vehicle behavior can be achieved through these techniques.[30] However, the complex models of vehicle dynamics, which are based on inference of accurate formulae and modeling, have difficulties dealing with the uncertainties in complicated conditions.[24] Similarly, advanced driver assistance systems (ADAS) only utilize vehicle-mounted sensors to detect the vehicle states as well as information about road traffic (obstacles, lane marking, etc.) to identify risks during driving. The current methods to identify driving risks are mostly based on simple factors (safety distance, lane departure, etc.). Even though these traditional technologies on driving risk identification have reduced the possibility of accidents to certain extent, they neglect the complex effects on driving safety brought by driver, vehicle and road environment interaction. The lack of consideration of all factors from driver, vehicle and road environment limits the usability of the ADAS under complicated driving conditions.[19,31]

With the increasing intelligence levels of vehicles as well as the unceasing accumulation of all kinds of massive online/offline driving data, it has become possible for data-driven methods to improve the ADAS. By utilizing the system's online or offline data, the data-driven methods can generate a variety of data-based functions including forecasting, evaluating, adjusting, monitoring, diagnosing, decision-making and optimizing.[29] Data-driven modeling refers to establishing the mathematical connections between leading variables and instrumental variables by mining useful information from the controlled system's input/output data for modeling. The data-driven models include artificial neural network, support vector machine, fuzzy logic, expert system, etc.[6]

While driving in complex traffic, a number of factors such as a driver, a vehicle and road environment are involved in a closed-loop system of driving risk recognition. If the analysis comes from the accurate mathematical modeling, the identification of driving risks will become particularly sophisticated. As a matter of fact, the ADAS merely need (intuitively) qualitative instructions of risk identification to carry out corresponding alarms. Instead, the system requires the driving risk recognition model to be able to make real-time responses.[5] Therefore, in order to utilize all kinds of information with uncertainties to realize quick and effective recognition of driving risks, this paper establishes a recognition model of driving risk based on belief rule-base (BRB) methodology that can take multi-source heterogeneous driving data into consideration. The accuracy and applicability of the proposed model are verified and evaluated by utilizing real world test data. Finally, the model is implemented in a vehicle driving simulator as a reference for future ADAS.

The driving risk recognition model discussed in the paper belongs to the field of pattern recognition (PR). Its application scenario is the road environment of unpiloted driving in the future, which provides preliminary basis for the highly intelligent driving in the future, and belongs to research on artificial intelligence (AI) in the automotive field. This paper is organized as follows: Section 2 presents a review of previous research on driving behavior and BRB methodology; in Sec. 3, the data collection and pre-analysis are described; Sections 4 and 5 detail the establishment, test and implementation of the model; Section 6 gives some concluding remarks and discusses possible improvements.

2. Literature Review

This section provides an overlook of the research development in the field of driving behavior and BRB methodology.

The necessity for an objective method to understand daily driving behavior which derives from the driving style has been emphasized in many studies. Aarts and Van Schagen[1] highlighted the importance of correlation between vehicle speed on road and traffic safety. According to their research, the collision risk and its severity will rise with increasing speed. Miyajima *et al.*[16] used longitudinal, lateral acceleration and velocity signals captured in driving recorders for the driving risk analysis. Bonsal *et al.*[7] explored the modeling of various driving styles, especially in urban traffic, based on several driving parameters. Another interesting technique was presented by Macadam *et al.*,[15] in which the driving behavior was classified under five different categories using both range and range rates of longitudinal closures. Othman *et al.*[18] conducted a study on driver behavior and obtained data from a driving simulator using a predetermined computer simulated driving course. In order to extract relevant information from the raw data set, the authors used a linear prediction analysis technique to extract relevant features that could best describe the driver operation behavior. Raksincharoensak *et al.*[20] used a combined driver behavior model based on a state transition feature for modeling naturalistic driving behavior in traffic scenarios. Lin *et al.*[14] categorized the driver characteristics based on either driver's operational behavior or the driver behavior characteristics. They discussed applications of the identification of the driver behavior characteristics to the intelligent driver advisory system the driver safety warning system and the vehicle dynamics control system. These studies advanced the field and some of the technologies developed have been commercialized.

Bayes Network (BN) is one of the most popular methods used for risk assessment in the transportation field.[32] BN builds the relationships among factors by conditional probability tables (CPTs). The assessment is then carried out by combining CPT with prior values of all variables on the basis of Bayes theory. However, one limitation for BN is that the size of the CPT grows exponentially with the number of variables, which makes it very difficult to apply in real time.[13] One possible solution is that the CPT can be decomposed and calculated separately if the variables are independent.[21] The theory of belief functions, also referred to as evidence theory, is a general framework for reasoning with uncertainty, with the combination of other frameworks such as probability, possibility and imprecise probability theories.[4] BRB is another method that has a different theoretical foundation compared to BN.[26,27] BRB makes inferences by combining all the activated BRBs based on evidence theory, which is nonlinear.[8] BRBs and CPTs in BN are similar methods. However, the evidence theory makes it possible to obtain BRBs by combining various experts' knowledge, so that the BRBs can be more reliable. Furthermore, the factors that have impacts on risks can be divided into groups, and assessed separately in an iterative way. This reduces the dimensions of BRBs substantially. BRB theory has been successfully applied to the accident analysis[28] and technique selection for ship emission reduction[25] in the maritime domain. BN and evidence reasoning were also integrated to carry out quantitative risk assessment.[22]

For the modeling of complicated decision-making problems with uncertain quantitative information and qualitative knowledge, and for the purpose of risk recognition and evaluation on target model, in Refs. 2, 3, 9 and 12 the author has adopted algorithms such as Bayesian network, neural network, support vector machine, grey theory, and rough set to construct risk evaluation model, and achieved certain effects. The major research fields include military security, network security, and oil and gas engineering. However, effective weight allocation is lacking in input index and training rule of these methods, resulting in low data service efficiency. Some methods have to be realized with large amounts of computation and complicated reasoning process. Considering advanced driver assistance system demands high instantaneity, while the calculation capacity of vehicle device is quite limited, this paper plans to build a vehicle driving risk recognition model based on driving data. The recognition model will recognize the driving risks in the vehicle driving process with novel method of belief rule-base which is convenient for plug-and-play on vehicle device. Inference method of confidence rule base is developed based on D-S evidence theory, decision-making theory, fuzzy theory, and traditional IF-THEN rule base and equipped with the capability of modeling incomplete, and fuzzy data with probabilistic uncertainty, subjectivity/objectivity, and nonlinearity. The inference method is suitable for building evaluation rules on vehicle driving risk recognition and the relevant knowledge representation method and is able to realize data input and reading in the vehicle device effectively. It can be applied to advanced driver assistance system in practice.

3. Data Collection and Preanalysis

3.1. *Experimental design*

Thirty professional taxi drivers with an average age of 46.3 (SD = 8.2) were recruited to participate in this field driving test. Each participant was required to drive the test vehicle along the G70 (Han-Shi) Expressway from Wuhan to Suizhou in China for more than 2.5 h to complete a test trip. In each single test day, only one participant would drive through the test trip. So 30 participants did the experiment in 30 days. However, we needed to confirm that the weather was sunny and traffic condition was moderate every single test day, so the experiment lasted for about two months in total. The length of the selected road is about 300 km with two-way four lanes, and the width of each lane is 3.75 m. The vehicle test was instrumented with an inertial navigation system (INS) and a Mobileye C2-270 system. Three additional cameras were installed on the windscreen of the vehicle (Resolution ratio is 800*640. Video sample rate is 30 fps), which, respectively, recorded the front road environment, the facial expression of the driver and the operation behavior of the driver. The driving data, such as longitude and latitude, speed, accelerations, azimuth, headway to lead vehicles, lane position, were collected via INS and Mobileye together.

3.2. *Screening driving data for safety-critical events*

The recorded data include no collisions or other accidents, but several critical events were encountered along with many periods without apparent risks. Manually, analyzing the road and driver video data of all 75 h collected data would need an immense effort. Hence, the video reviewing method of safety-critical events was adopted from previous research.[23] The key capture has been done to the defined safety-critical events by setting a certain screening threshold, which is to save the events at the moment when the acceleration absolute value is greater than $1.99\,\mathrm{m/s^2}$. The capture process of these safety-critical events has been done by manual handling. The time range of emergency in safety-critical events has been classified. Other parameters, such as road environment and the operational behavior of driver, were extracted from watching videos and vehicle-mounted data synchronously, which is the primary method of constructing a BRB model extracting multi-source heterogeneous driving parameter in a complicated traffic environment.

Safety-critical events of the test vehicle on the expressway mainly occurred in the following driving scenes: lane-changing, car following, overtaking/overtaken. It was found, after reviewing and analyzing the videos, that the primary potential accident type of the instrumented vehicle is a vehicle to vehicle collision. It is unlikely to have other accident types like rollover or road departure. In conclusion, the risk identification of driving behavior analyzed in this paper is mainly done based on the risky events of vehicle's possible collisions on expressways.

3.3. *Driving parameter quantization for collision risk*

Too many input variables will result in redundancy of making driving risk identification rules when constructing the recognition model of driving risk on the basis of BRB. This could overload the calculation of the subsequent process of evidence reasoning and optimization. However, it will not be sufficient to describe the relationships between input and output if only few input variables are taken into account. Therefore, deciding what to quantify on these multi-source heterogeneous driving data shall be done according to the driving data characteristics and testing conditions as shown in Table 1. Specific values used to define the categories of driving parameter and the boundaries of the categories will be defined later in Table 4.

Table 1. Driving parameter quantization for collision risk.

Category	Parameter/Description	Quantization		Selection
Driver	Gender	Male	M	√
		Female	F	
	Age	Small	YS	√
		Medium	YM	
		Large	YL	
	Driving experience in years	Short	DS	√
		Medium	DM	
		Long	DL	
Vehicle state	Brake Performance	Normal		—
	Speed (km/h)	Small	VS	√
		Medium	VM	
		Large	VL	
	Acceleration (vector) (absolute value, m/s^2)	Small	AS	√
		Medium	AM	
		Large	AL	
	Time headway (s)	Small	TS	√
		Medium	TM	
		Large	TL	
Driving condition	Road condition	Dry asphalt		—
	Lane position of vehicle	Running	DL	√
		Overtaking	OL	
		Ramp	RL	

3.4. *Driving risk state calibration*

The driving parameter selection provides a driving assessment index to the data-driven model. While, a large amount of historical data is needed as a data-driven source, one part of the historical data can be used as a training sample data to train the recognition model. The second part can be used as testing sample data to check the accuracy rate of the trained recognition model. Before training the data, driving risk state needs to be classified by a subjective estimate method which has been widely used in the research of transportation.[11] The driving data collected including three-channel camera videos and real-time vehicle data provide reliable evidence for experts' subjective evaluation.

With the warning characteristic of the existing ADAS as a reference, the experts classified the risk states of driving events into three levels, namely, None (Low), Medium and Large.[5] The data was sampled by reviewing the (near) safety-critical events in the videos, which helped us quickly gather enough data for all the levels of driving risk. The experts determined the risk states of all the samples. We randomly selected 16.7 points of time in each participant's data on average, 500 (16.7*30) points of time in total as historical data set which is shown in Fig. 1 and Table 2. Each sampled data was the driving data at a selected momentary time point. In addition, the assessment was made by jointly analyzing objective information, such as speed, acceleration, time headway and the video of the driver's face, excluding obvious inaccurate labeling of the sample data.

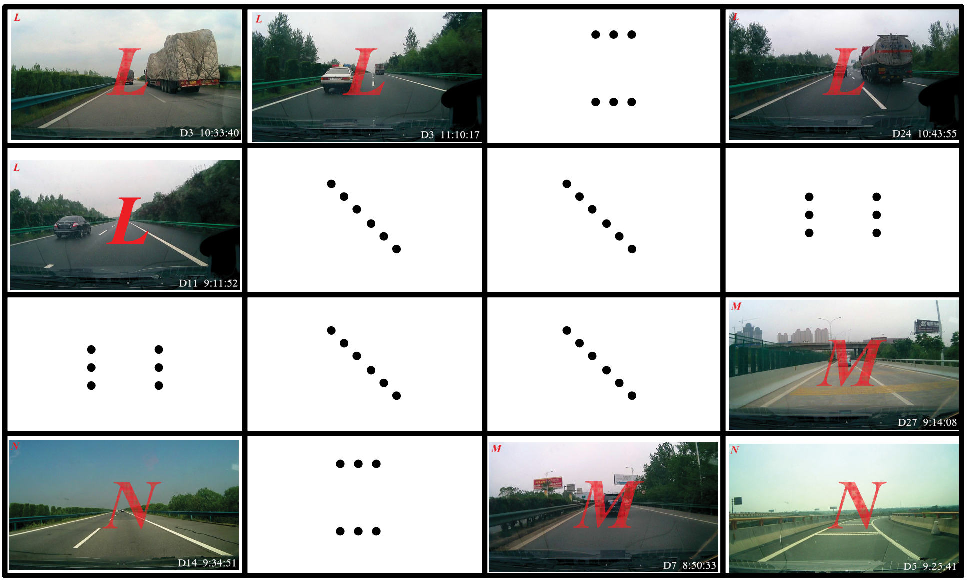


Fig. 1. 500 Fragments of driving risk state for Calibration.

Table 2. Driving data of 500 fragments.

No.	ID	Speed	Acceleration	Headway	Brake	Road	Lane	Risk State
1	D3	83 kph	0.21 g	2.5 s	Normal	Dry Asphalt	Overtaking	L
2	D3	95 kph	0.13 g	1.8 s	Normal	Dry Asphalt	Running	L
3	D8	98 kph	0 g	0 s	Normal	Dry Asphalt	Running	N
4	D15	92 kph	0 g	0 s	Normal	Dry Asphalt	Running	N
5	D9	65 kph	0 g	2.5 s	Normal	Dry Asphalt	Ramp	M
...	...	...	...	...	...	...	...	...
496	D21	86 kph	−0.26 g	0.5 s	Normal	Dry Asphalt	Overtaking	L
497	D20	71 kph	0.32 g	0 s	Normal	Dry Asphalt	Overtaking	M
498	D14	108 kph	−0.11 g	1.8 s	Normal	Dry Asphalt	Running	L
499	D7	98 kph	−0.08 g	1.6 s	Normal	Dry Asphalt	Overtaking	L
500	D5	65 kph	−0.02 g	0 s	Normal	Dry Asphalt	Ramp	N

Finally, there were 116 samples of risk state N, 160 samples of risk state M, and 224 samples of risk state L. Out of the 500 samples of driving events, 300 were taken as the training sample data of the recognition model leaving the other 200 as testing sample data. Note the time headway being 0 s indicates no other vehicles are detected in front of the subject vehicle.

4. Methodology

Based on Dempster–Shafer theory of evidence, decision theory and fuzzy set theory, Yang *et al.* proposed a new methodology for building a hybrid rule-base using a belief structure and for inference in the rule-based system using the evidential reasoning (ER) approach. The methodology is referred to as a generic rule-base inference methodology using the ER approach: RIME.[8,25–28]

4.1. *Belief rule-base*

In the RIMER approach, a belief IF-THEN rule, for example the kth rule R_k, is expressed as follows:

$$\begin{aligned}
&R_k\text{: } If\ x_1\ is\ A_1^k \wedge x_2\ is\ A_2^k \wedge \cdots \wedge x_{T_k} \text{ is } A_{T_k}^k,\\
&\text{Then } \{(D_1, \beta_{1k}), (D_2, \beta_{2k}), \ldots, (D_N, \beta_{Nk})\}, \quad \sum_{i=1}^{N} \beta_{ik} \le 1, \qquad (1)\\
&\text{with a rule weight } \theta_k \text{ and attribute weight } \delta_1, \delta_2, \ldots, \delta_{T_k},
\end{aligned}$$

where A_i^k is the referential value of the ith antecedent attribute and T_k the number of antecedent attributes used in the kth rule. $\beta_{ik}(i = 1, 2, \ldots, N)$ is the belief degree to which D_i is believed to be the consequent if $(x_1, x_2, \ldots, x_{Tk}) = (A_1^k, A_2^k, \ldots, A_{Tk}^k)$. L is the number of all rules in the rule-base. If $\Sigma_{i=1}^N \beta_{ik} =1$, the kth rule is complete; otherwise, it is incomplete. Note that $\Sigma_{i=1}^N \beta_{ik} = 0$ denotes total ignorance about the output given the input in the kth rule. Rule (1) is also referred to as a belief rule. It is further supposed that T is the total number of antecedent attributes used in the rule base.

Let

$$\begin{aligned}
&X = (x_1, x_2, \ldots, x_{T_k}), \quad A^k = (A_1^k, A_2^k, \ldots, A_{T_k}^k), \quad D = (D_1, D_2, \ldots, D_N),\\
&\beta^k = (\beta_{1k}, \beta_{2k}, \ldots, \beta_{Nk}), \quad \text{and} \quad \delta = (\delta_1, \delta_2, \ldots, \delta_T).
\end{aligned}$$

X is referred to as an input vector to the kth rule; A^k is a packet antecedent, A_i^k $(i = 1, 2, \ldots, T_k)$ is the ith referential values of the packet antecedent A^k; D is the consequent vector; β^k is the vector of the belief degrees; and δ is the attribute weights of all the T antecedent attributes in the rule base.

It is not difficult to see the difference between a traditional IF-THEN rule and a belief IF-THEN rule. In the traditional rule, the consequence is either 100% true or 100% false. Such a rule base has limited capacity in representing knowledge in a real world. The belief structure in the belief rule base provides better flexibility in representing the knowledge of different structures and complexity, such as continuous and uncertain relationships between antecedents and consequents.

4.2. *Inference with BRB using ER approach*

Given an input to the system, $X = (x_i, i = 1, 2, \ldots, T)$, T is the total number of antecedents in the rule base; $x_i(i = 1, 2, \ldots, T)$ is the ith attribute which can be one of the following types: continuous, discrete, symbolic and ordered symbolic.

Before the start of an inference process, the matching degree of an input to each referential value in the antecedents of a rule needs to be determined, so that an activation weight for each rule can be generated. This is equivalent to transforming an input into a distribution on referential values using belief degrees. It can be accomplished using different techniques such as the rule or utility-based equivalence transformation techniques.

Using the notations provided above, the activation weight of the kth rule w_k is calculated as

$$w_k = \theta_k \prod_{i=1}^{T_k} (\alpha_{ik})^{\overline{\delta}_i} \Big/ \sum_{j=1}^{L} \theta_j \prod_{l=1}^{T_k} (\alpha_{lj})^{\overline{\delta}_l}, \tag{2}$$

where $\overline{\delta}_i = \frac{\delta_i}{\max_{i=1,\dots,T_k}\{\delta_i\}}$, $\alpha_{ik}(i = 1, \dots, T_k)$ is the individual matching degree to where the input x_i matches the ith referential value A_i^k of the packet antecedent A^k in the kth rule, $\alpha_{ik} \geq 0$, and $\Sigma_{i=1}^{Tk}\alpha_{ik} \leq 1$. $\alpha_k = \prod_{i=1}^{T_k} (\alpha_{ik})^{\overline{\delta}_i}$ is called the combined matching degree.

Having determined the activation weight of each rule in the rule base, the ER approach can be directly applied to combine the rules and generate final conclusions. Suppose the outcome of the combination yields the following:

$$Y(X) = \{(D_j, \beta_j), \quad j = 1, \dots, N\}. \tag{3}$$

Equation (3) means that if the input is given by X, then the consequent is D_1 to a degree of β_1, D_2 to a degree of $\beta_2, \dots,$ and D_N to a degree of β_N. Using the analytical format of the ER algorithm, the combined belief degree β_j in D_j can be generated as follows:

$$\beta_j = \frac{u\left[\prod_{k=1}^{L}\left(w_k\beta_{jk} + 1 - w_k\sum_{j=1}^{N}\beta_{jk}\right) - \prod_{k=1}^{L}\left(1 - w_k\sum_{j=1}^{N}\beta_{jk}\right)\right]}{1 - u\left[\prod_{k=1}^{L}(1 - w_k)\right]}, \tag{4}$$

where $u = \left[\sum_{j=1}^{N} \prod_{k=1}^{L} \left(w_k\beta_{jk} + 1 - w_k\sum_{j=1}^{N}\beta_{jk}\right) - (N-1)\prod_{k=1}^{L}\left(1 - w_k\sum_{j=1}^{N}\beta_{jk}\right)\right]^{-1}$ and w_k is as given in Eq. (2).

4.3. *BRB training*

The initial belief rules and knowledge representation parameters including rule weights, attribute weights and consequent belief degrees in a BRB can be given by the domain experts or randomly generated. Hence, the rules may not be 100% accurate. An initial BRB can be trained using historical data to improve its ability for representing the clinical domain knowledge.

The aim of BRB training is to find a set of parameters $(\beta_{ik}, \theta_k, \delta_i)$ of a BRB that can help it accurately represent the domain specific knowledge. The training process is implemented by minimizing the discrepancy between BRB results and sampled data. Assuming there are M cases in a training sample, and the input–output pairs of the M cases are $(X_m, \widehat{Y}_m)(m = 1, 2, \dots, M)$, the process of learning from these M datasets can be illustrated as Fig. 2, where Y_m is produced by the BRB system; the

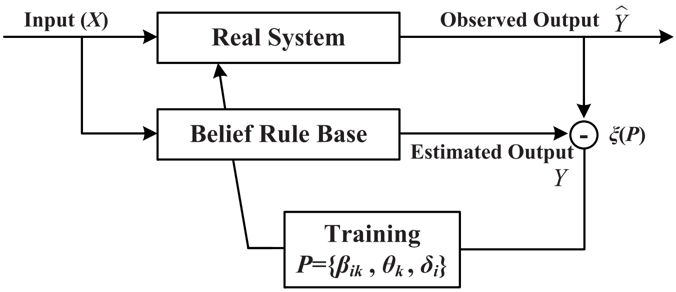


Fig. 2. Training process.

real output $\widehat{Y}_m$ is observed by experts or acquired by instruments; and $\xi(P)$ represents the difference between the real output and the system output. In the BRB optimization model, the objective function is to minimize $\xi(P)$, and the constraints define what the knowledge representation parameters of a BRB system should follow. As a result of the training process, there will be a new set of $(\beta_{ik}, \theta_k, \delta_i)$ for BRB.

5. Model

By establishing the applicable BRB system for the driving risk identification, nonlinear relations are expected to be described between the driving behavior characteristic, vehicle states, road environment and the driving risk levels. For the established BRB system, the input of X is the driving data related to vehicle collision risk, the output of Y is the risk level of vehicle driving.

5.1. *Input and output*

The inputs X of the model have been selected according to Table 1 (subject data, vehicle data and environment). As for the output indexes, we take as Ref. 5 existing ADAS for collision mitigation that is divided into two parts: early warning and active control. The warning signals given by auditory and visual interfaces are normally categorized into several levels depending on the severity of the potential hazards. Therefore, the output parameter Y was divided into None (Low), Medium and Large according to the level of driving risk (see Table 3).

The reference values (classification boundaries) also need to be determined for all the inputs X. Instead of using strict mathematical derivation, the reference values were determined on the basis of the previous research on driving data characteristic distribution.[10,17,23] For example, the rate data for near-crashes are fairly evenly distributed among the four-time headway categories of < 1 s, 1–1.99 s, 2–2.99 s, and > 3 s. Other driving indexes data are categorized reasonably with similar rules (see Table 4).

Table 3. Input X and Output Y classification.

Driving Behavior X_D	Semantic Value		Vehicle X_V	Semantic Value	
Gender x_1	Male	M	Speed x_4	Small	VS
	Female	F		Medium	VM
				Large	VL
Age x_2	Small	YS	Accelerated Speed x_5	Small	AS
	Medium	YM		Medium	AM
	Large	YL		Large	AL
Driving years x_3	Small	DS	Time Headway x_6	Small	TS
	Medium	DM		Medium	TM
	Large	DL		Large	TL
Road X_R	Semantic Value		Level of Driving Risk Y		
Lane position x_7	Running	DL	Level y	None	N
	Overtaking	OL		Medium	M
	Ramp	RL		Large	L

Table 4. Reference values of input X and output Y.

Input X	Quantization	Reference Value	Input X	Quantization	Reference Value
x_1	M	1	x_5	AM	3
x_1	F	2	x_5	AL	5
x_2	YS	25	x_6	TS	1
x_2	YM	45	x_6	TM	2
x_2	YL	55	x_6	TL	3
x_3	DS	10	x_7	DL	1
x_3	DM	20	x_7	OL	2
x_3	DL	30	x_7	RL	3
x_4	VS	50	Output Y	Quantization	Reference Value
x_4	VM	80	y	N	0
x_4	VL	110	y	M	1
x_5	AS	1	y	L	2

5.2. *Constructing the initial BRB system*

Based on the input of the driving risk recognition model, a double-layered BRB system is established, see Fig. 3. The first layer system is composed of three BRB subsystems: (1) three driver factors are utilized as the first input of BRB subsystems to judge the driver's status; (2) three vehicle factors are taken as the second input of BRB subsystems to judge the vehicle's status; (3) one road environment factor is the third input of BRB subsystems to judge the road environment's status. The second layer system utilizes the driver's status, vehicle's status, road environment's status data from the first layer system as the input, and gives the final driving risk recognition output. Since the bottom layer's antecedent input has been divided into various states (discrete values), there is no need to handle input transformation to continuous variables.

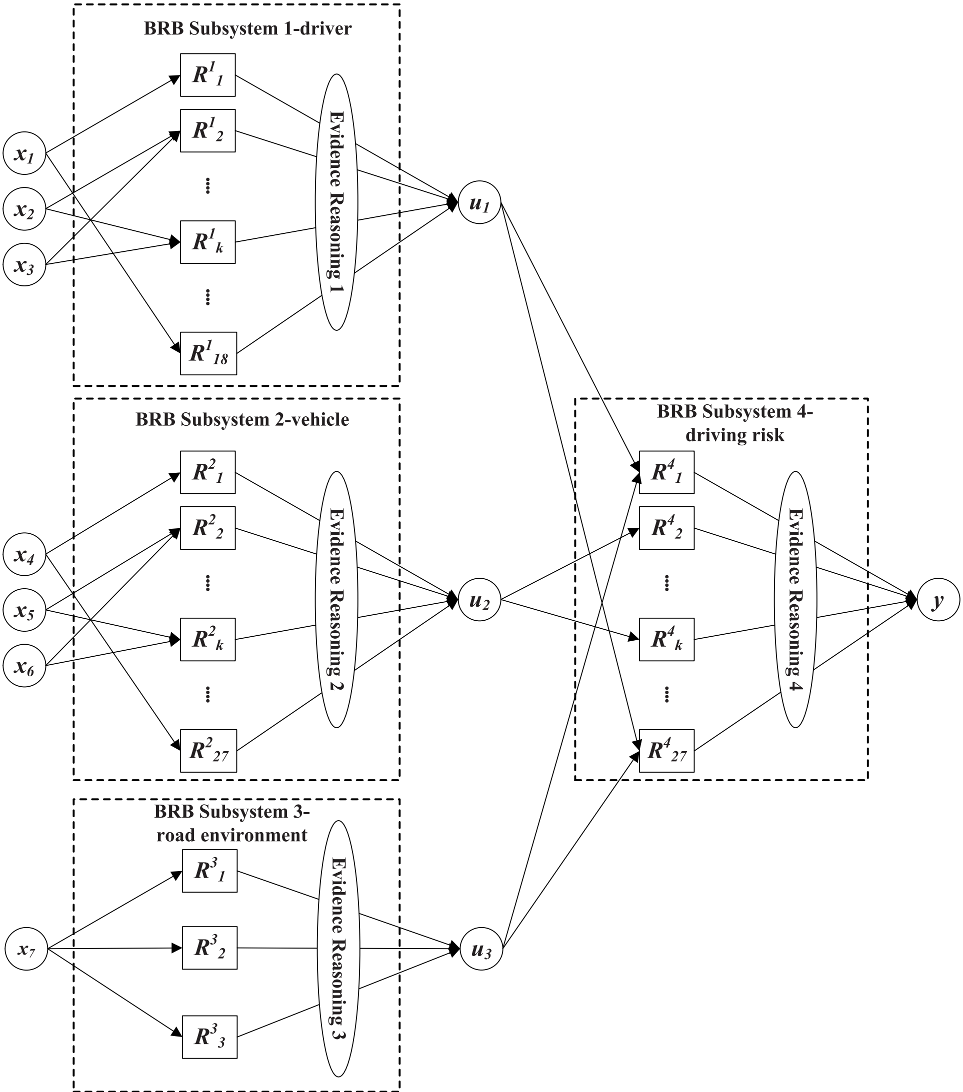


Fig. 3. BRB system of driving risk recognition model.

where x_1–x_7 are system antecedent inputs; y is system output; and u_1–u_3 are introduced as the middle factors to evaluate the driver, the vehicle and the road environment.

According to Fig. 3, the BRB system is composed of four BRB subsystems. Also three driver factors x_1–x_3 are the antecedent input of the first BRB subsystem, and driver status u_1 is the subsystem output. Three vehicle factors x_4–x_6 are the antecedent input of the second BRB subsystem, vehicle status u_2 is the subsystem output. One road environment factor x_7 is the antecedent input of the third BRB subsystem, road environment status u_3 is the subsystem. The first three layer subsystems' outputs form the inputs for the second layer BRB subsystem. The final identification findings for the driving risk level are given according to these antecedent inputs, utilizing the fourth BRB subsystem. In the second layer BRB structure, because it is a basic belief structure lead by the first layer BRB structure, input transformation of the second layer BRB subsystem is also not necessary.

The establishment of initial belief rules employed the following four methods: expert knowledge, credible historical data, previous driving risk identification rules, and random selection rules. In this paper, according to the expert knowledge and the distribution characteristic of real vehicle test statistics, the initial belief rules of four subsystems are established. The belief rule of the final subsystem (BRB subsystem 4) is

$$\begin{aligned}&R_k^4\text{: } If\ u_1\ is\ B_{1,k} \wedge u_2\ is\ B_{2,k} \wedge u_3 \text{ is } B_{3,k}\\ &\text{Then } \{(N, \beta_{1,k}^4), (M, \beta_{2,k}^4), (L, \beta_{3,k}^4)\},\end{aligned}$$

where $B_{1,k}$, $B_{2,k}$, $B_{3,k}$ are respectively the semantic values of middle factor U and the specific data is not set; N, M, L are semantic values of the system output showing low, middle and high level of driving risk status; $\beta_{1,k}^4$, $\beta_{2,k}^4$, $\beta_{3,k}^4$ are respectively belief degree to which N, M or L is believed to be the consequent if (u_1, u_2, u_3) = ($B_{1,k}$, $B_{2,k}$, $B_{3,k}$); and in total we have 27 rules ($3 \times 3 \times 3$).

The initial belief rule of the final subsystem (BRB subsystem 4) is shown in Table 5.

As the initial BRB system is established on the basis of expert knowledge and historical data with respective subjective and objective indeterminacy, the initial BRB system is imprecise and needs to be optimally trained.

5.3. *Training of the initial BRB system*

The training of the initial BRB system is conducted on the basis of a subset of the sample data. A total of 300 samples were chosen randomly from 500 historical data samples as training sets, which leaves the remaining 200 samples as test samples (see Fig. 1). All these data cover every categories of input and output data. The learning process is implemented in MATLAB and outlined in the following seven steps.

(1) Set initial parameters;
(2) Transform the input;
(3) Calculate rule activation weight;
(4) Combine activated rules;
(5) Estimate driving risk;
(6) Calculate the driving risk difference between the observed and estimated;
(7) Find a new set of parameters P to minimize the difference defined in Eq. (7).

Table 5. Initial belief rules of BRB subsystem 4.

	Input U(Middle Factor)						
	Attribute Weights	1	1	1	Output Y(Belief Degree)		
Rule Number	Rule Weight	u_1	u_2	u_3	N	M	L
1	1	S	S	S	1	0	0
2	1	S	S	M	0.8	0.1	0.1
3	1	S	S	L	0.6	0.2	0.2
4	1	S	M	S	0.6	0.2	0.2
5	1	S	M	M	0.7	0.1	0.2
6	1	S	M	L	0.5	0.2	0.3
7	1	S	L	S	0.4	0.2	0.4
8	1	S	L	M	0.2	0.2	0.6
9	1	S	L	L	0	0.2	0.8
10	1	M	S	S	0.9	0.1	0
11	1	M	S	M	0.7	0.2	0.1
12	1	M	S	L	0.5	0.3	0.2
13	1	M	M	S	0.5	0.2	0.3
14	1	M	M	M	0.6	0.1	0.3
15	1	M	M	L	0.4	0.2	0.4
16	1	M	L	S	0.2	0.2	0.6
17	1	M	L	M	0.1	0.2	0.7
18	1	M	L	L	0	0.1	0.9
19	1	L	S	S	0.8	0.1	0.1
20	1	L	S	M	0.5	0.2	0.3
21	1	L	S	L	0.4	0.3	0.3
22	1	L	M	S	0.4	0.2	0.4
23	1	L	M	M	0	0.4	0.6
24	1	L	M	L	0	0.3	0.7
25	1	L	L	S	0.1	0.1	0.8
26	1	L	L	M	0	0.1	0.9
27	1	L	L	L	0	0	1

Having obtained the outcome shown in Eq. (3), the estimated the level of driving risk is calculated as follows:

$$\text{Risk}(Y) = D_1\beta_1 + D_2\beta_2 + D_3\beta_3. \tag{5}$$

Hence the Risk(Y) is a continuous quantity between 0 and 2. Discretizing the Risk(Y) makes it easier to compare the driving risk difference between the observed and the estimated ones. The ADAS also often need discrete qualitative instructions of risk identification to provide corresponding alarms. The level of driving risk is discretizing as

$$\text{Estimated\_Risk}(Y) = \begin{cases} 0 & 0 \leq \text{Risk}(Y) \leq 0.5 \\ 1 & 0.5 < \text{Risk}(Y) \leq 1.5 \\ 2 & 1.5 < \text{Risk}(Y) \leq 2. \end{cases} \tag{6}$$

According to the above-mentioned steps, we put the 300 sets of training samples into the initial BRB system, then we get the estimated values of corresponding

Table 6. Identification accuracy rate based on initial BRB (Training samples).

Driving Risk Level	N (0)	M (1)	L (2)
Accuracy rate	83.6%	51.4%	50.8%

training samples. By comparing the estimated values with the real sample output, the accuracy rate of driving risk identification can be obtained (see Table 6). We can see that the initial BRB is not perfect because of the subjective indeterminacy of expertise. The accuracy of identifying the middle and high-level driving risk status is only about 50%, leading to not-so-high overall detection accuracy. This is because the BRB built on the basis of expertise and historical data cannot model the relationship between input and output accurately. However, the correct identification of middle and high-level driving risk status is essential in the early warning strategy of ADAS.

Therefore, the training sample needs to be optimized to improve the accuracy of identification.

Table 7. Trained belief rules of BRB subsystem 4.

	Input U(Middle Factor)						
	Attribute Weights	0.967	1	0.927	Output Y(Belief Degree)		
Rule Number	Rule Weight	u_1	u_2	u_3	N	M	L
1	1	S	S	S	1	0	0
2	0.86	S	S	M	0.9	0	0.1
3	0.99	S	S	L	0.5	0.3	0.2
4	1	S	M	S	0.6	0.2	0.2
5	1	S	M	M	0.7	0.1	0.2
6	0.56	S	M	L	0.4	0.3	0.3
7	0.87	S	L	S	0.3	0.2	0.5
8	0.97	S	L	M	0.1	0.2	0.7
9	1	S	L	L	0	0.2	0.8
10	1	M	S	S	1	0	0
11	1	M	S	M	0.7	0.2	0.1
12	1	M	S	L	0.5	0.3	0.2
13	0.87	M	M	S	0.4	0.3	0.3
14	0.82	M	M	M	0	0.5	0.5
15	0.98	M	M	L	0.3	0.3	0.4
16	0.79	M	L	S	0.1	0.2	0.7
17	1	M	L	M	0.1	0.2	0.7
18	1	M	L	L	0	0.1	0.9
19	0.97	L	S	S	0.7	0.2	0.1
20	0.85	L	S	M	0.4	0.2	0.4
21	0.79	L	S	L	0.3	0.4	0.3
22	0.91	L	M	S	0.3	0.3	0.4
23	1	L	M	M	0	0.4	0.6
24	0.91	L	M	L	0	0.2	0.8
25	0.97	L	L	S	0	0	1
26	0.99	L	L	M	0	0	1
27	1	L	L	L	0	0	1

The mean value of the quadratic sum of difference between the real sample output True_Risk(Y) and the estimated output Estimated_Risk(Y) can be expressed as

$$\xi(P) = \frac{1}{300}\sum_{i=1}^{300}[\text{True\_Risk}(Y)_i - \text{Estimated\_Risk}(Y)_i]^2. \tag{7}$$

The objective of the learning process is to find a set of parameters P, so that the difference between the observed and the estimated driving risk is minimized. This leads to the minimal $\xi(P)$, the constraint condition including: $0\leq \theta_k \leq 1$, $0\leq \delta_k \leq 1$, $0\leq \beta_j, k \leq 1$ and $\sum_{j=1}^{3} \beta_{j,k} = 1$, $k = 1, 2, \ldots, 27$.

The Fmincon function of the MATLAB optimization toolbox is used to realize the above processes. Then, the optimized belief rules of BRB can be composed by its optimization results. The results of optimized belief rules are shown in Table 7.

After confirming the optimized BRB system, the preceding steps were repeated: input 300 sets of training samples into the optimized BRB to get corresponding estimated values of training samples. The results show that the optimized BRB has greatly improved the accuracy on the driving risk identification. The accuracy of identifying all three levels of driving risk has reached 95%. Therefore, the reliability of the BRB model can be evaluated in the remaining test data as described below.

6. Test and Implement

6.1. *Model test*

A total of 200 sets of data were used to test the identification accuracy of the optimized BRB. The differences between the results of testing sample output and estimated output are as shown in Fig. 4, and the test accuracy is shown in Table 8.

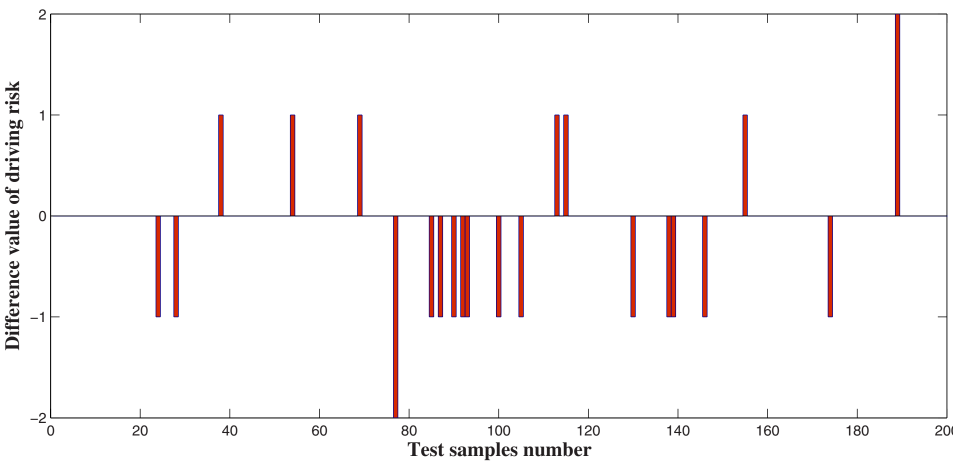


Fig. 4. Difference value between testing sample output and estimated output.

Table 8. Identification accuracy rate based on optimized BRB (Test samples).

Driving Risk Level	N (0)	M (1)	L(2)	Average
Accuracy amount	39/43	49/55	90/102	178/200
Accuracy rate	90.7%	89.1%	88.2%	89%
Missing amount	3/43(M) 1/43(L)	4/55(N) 2/55(L)	3/102(N) 9/102(M)	22/200
Missing rate	7.0%(M) 2.3%(L)	7.3%(N) 3.6%(L)	2.9%(N) 8.8%(M)	11%

According to the test results, a more accurate nonlinear mapping relationship between input and output has been established by the optimized BRB. The difference in values of driving risks represents the absolute difference between testing sample output and the estimated output. For example, if a testing sample output is 0, and the corresponding estimated output is 2, then the driving risk difference is 2. The higher the absolute value is, the higher the error is. There are two samples with the absolute difference of 2 and 20 samples with the absolute difference of 1 in 200 sets of test data. Moreover, the identification accuracy of the optimized BRB can reach almost 90% in three levels (none, medium, large) driving risk situations.

In the testing environment (Intel(R) Core(TM) i5-5200U CPU @ 2.20 GHz, RAM: 8 GB), time needed to give the estimated values by BRB is 0.875 ms at every turn. At present, the sampling frequency of existing various vehicle sensors generally does not exceed 50 Hz, so the time interval of our BRB outputting the estimated value is far shorter than the generic signal sampling intervals in ADAS. Certainly, this model can be applied in real time as part of an early warning strategy in ADAS.

6.2. *Model implementation*

The driving risk recognition model (the optimized BRB system), which has reliable accuracy and low computational cost, was implemented in a driving simulator to test its safety and real time performance. We used a programmable and fully interactive virtual reality driving simulator powered by a programmable software engine. It includes three independently configurable driving displays that provide a wide driver field-of-view, a full-sized steering wheel with advanced dynamics based feedback, and advanced vehicle dynamics software modeling (see Fig. 5).

While setting its external-connected procedures, the recognition model was written in MATLAB environment, and the driving parameters generated from the simulator in real-time were introduced to the BRB system. Then the recognition model outputs the level of driving risk in real-time. Finally, according to the risk status (N, M, L), the warning interface was set and feedback given to the driver (see Fig. 6).

This simulator test was to discuss the possible implementation of model on ADAS, so only one test driver and one test scenario were selected. In order to implement the driving risk recognition system, a 32-year old male with six years of driving experience, was chosen to be the subject of the test. A typical expressway driving environment is adopted as the road environment for the test. The driver's

Fig. 5. Driving simulator.

Fig. 6. Human–computer interaction.

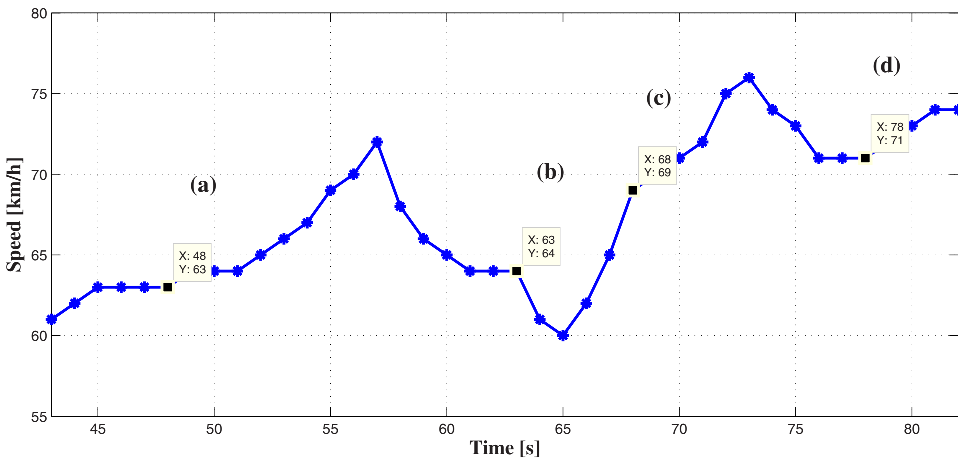
parameters are set into the driving simulator system beforehand, namely, $x_1 = 1$, $x_2 = 32$, $x_3 = 6$. The other parameters of x_4, x_5, x_6, x_7 are input into the driving simulator directly as parameters from the model.

The test scenario was as follows: The tested driver was driving on one side of the closed highway, where the density of traffic flow was medium, and the target vehicle had occupied the overtaking lane for quite a long time. In front of the target vehicle, there was a truck that had also occupied the overtaking lane for quite a long time. The target vehicle intended to overtake the truck by illegally cutting through the right side of the truck's lane. The speed, acceleration and time headway of target vehicle were recorded, and the level of the driving risk status was calculated in real time (see Fig. 7).

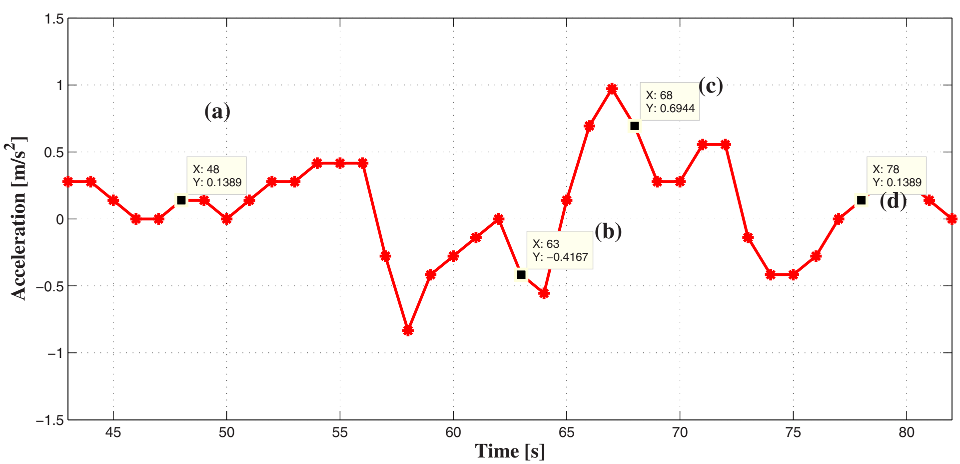
The segments of the whole process are as follows: (a) $\boxed{0{:}48}$ the target vehicle (tested) cruised in the overtaking lane, when a truck was cruising in front of the target vehicle. (The target vehicle's state was 63 km/h, 0.14 m/s^2, and driving risk was N); (b) $\boxed{1{:}03}$ as the target vehicle found that it was too close to the truck, it braked and intended to overtake the truck by illegally changing lane to the running lane. (The target vehicle's state was 64 km/h, -0.42 m/s^2 and the driving risk was L);

(c) [1:08] the target vehicle completed the changing-way and sped up to overtake the truck. (The target vehicle's state was 69 km/h, 0.69 m/s^2, and the driving risk was L); (d) [1:18] the target vehicle was overtaking steadily. (The target vehicle's state was 71 km/h, 0.14 m/s^2, and the driving risk was M). The whole process lasted for about 30 s.

This proves that the driving risk recognition model based on BRB has rather feasibly estimated the level of driving risk status during the whole process of illegally overtaking. Even though the parameters of the model as well as the settings of the driving scenario in the simulator were quite simple, this simulation offers the possibility of applying our model to improve ADAS.

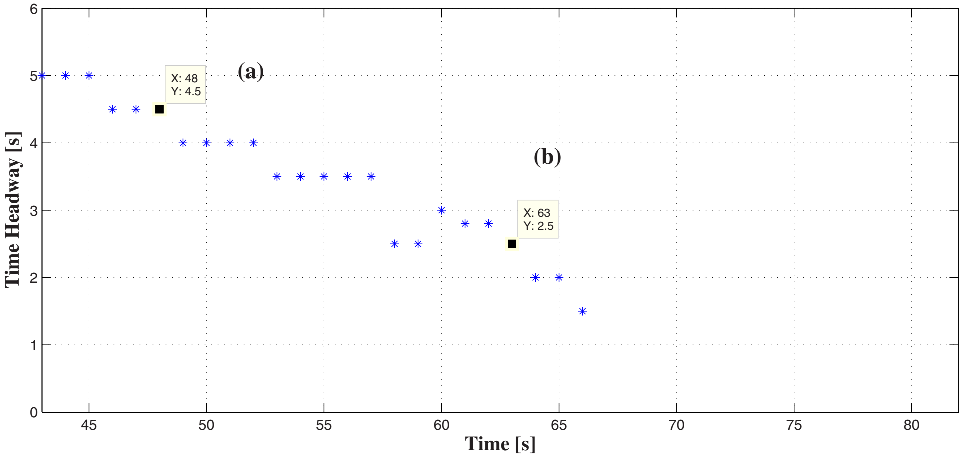


(a) Speed

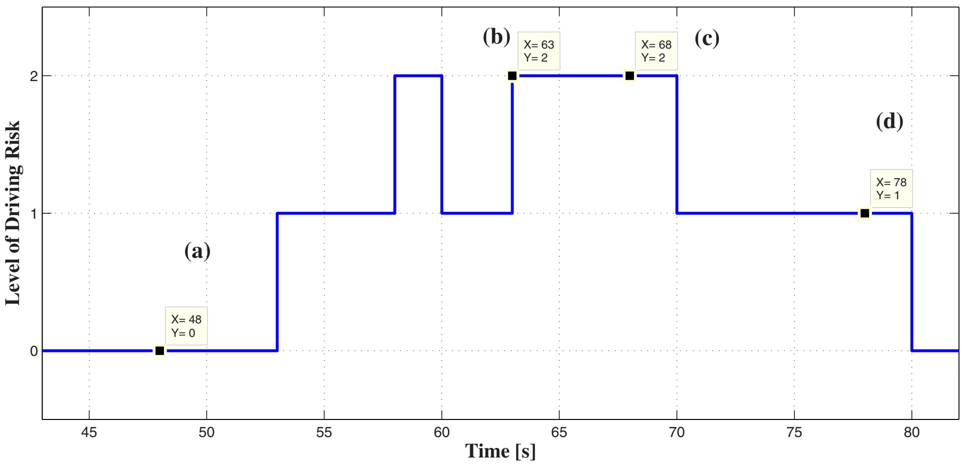


(b) Forward acceleration

Fig. 7. Time-domain plot of driving data in test scenario.



(c) Time headway



(d) Level of driving risk

Fig. 7. (*Continued*)

7. Conclusions and Recommendations

This paper adopts the approach of data-driven modeling, utilizing the accumulation of on-road driving data, to comprehensively consider the influence on driving safety brought by driver, vehicle and road environment, and establish a recognition model of driving risk based on a BRB methodology. The results show that the model has high accuracy of up to about 90% for the three-level driving risk recognition. The model with reliable accuracy and fast computation time can meet the needs of practical systems. Finally, the application and test of our model were carried out to provide a certain basis for improving the ADAS in a driving simulator.

Compared to nonlinear modeling methods, such as neural network, BRB system uses not only the objective sampled, but also subjective information provided by the experts. Moreover, its adjustable parameters have a clear physical meaning. Therefore, it is closer to practical application, and it is also easier for drivers to accept and take part in the entire modeling process of the BRB. However, the on-road vehicle driving data utilized in this paper was collected in certain driving conditions which may affect the usability of the model in all conditions in the real world. Therefore, a further assessment involving a wider range of drivers and test scenarios can be conducted for model implementation as a follow-up study.

Acknowledgments

This study is sponsored by the National Undergraduate Training Program for Innovation and Entrepreneurship (201710514011); the National Key Technology R&D Program of China (2014BAG01B03); Hubei Provincial Natural Science Foundation of China (2018CFC863); the National Natural Science Foundation of China (51675390); and Truck Merging Support-a Step towards Autonomous Driving (HTSM Automotive No. 12831). The authors would like to thank the China Scholarship Council.

References

1. L. Aarts and I. Van Schagen, Driving speed and the risk of road crashes: A review, *Accident Anal. Prev.* **38**(2) (2006) 215–224.
2. A. Alhussan and K. El Hindi, Selectively fine-tuning Bayesian network learning algorithm, *Int. J. Pattern. Recognit. Artif. Intell.* **30**(8) (2016) 1651005.
3. G. Alimjan *et al.*, A new technique for remote sensing image classification based on combinatorial algorithm of SVM and KNN, *Int. J. Pattern. Recognit. Artif. Intell.* **32**(7) (2018) 1859012.
4. O. Basir and X. Yuan, Engine fault diagnosis based on multi-sensor information fusion using Dempster–Shafer evidence theory, *Inform. Fusion* **8**(4) (2007) 379–386.
5. K. Bengler *et al.*, Three decades of driver assistance systems: Review and future perspectives, *IEEE Intel. Transp. Syst.* **6**(4) (2014) 6–22.
6. B. Biggio, G. Fumera and F. Roli, Pattern recognition systems under attack: Design issues and research challenges, *Int. J. Pattern. Recognit. Artif. Intell.* **28**(7) (2014) 1460002.
7. P. Bonsall, R. Liu and W. Young, Modelling safety-related driving behaviour — impact of parameter values, *Transport Res A-Pol* **39**(5) (2005) 425–444.
8. K. S. Chin *et al.*, An evidential-reasoning-interval-based method for new product design assessment, *IEEE Trans. Eng. Manag.* **56**(1) (2009) 142–156.
9. X. Cui *et al.*, Tire defects classification with multi-contrast convolutional neural networks, *Int. J. Pattern. Recognit. Artif. Intell.* **32**(4) (2018) 1850011.
10. T. A. Dingus *et al.*, The 100-car naturalistic driving study: Phase ii — results of the 100-car field experiment, National Highway and Traffic Safety Administration, DOT HS 810 593, Washington, DC (2006).
11. A. Eskandarian and A. Mortazavi, Evaluation of a smart algorithm for commercial vehicle driver drowsiness detection, in *Intelli. Vehicles Symp.* (IEEE, 2007), pp. 553–559.

12. A. Lasisi *et al.*, Extracting information in agricultural data using fuzzy-rough sets hybridization and clonal selection theory inspired algorithms, *Int. J. Pattern. Recognit. Artif. Intell.* **30**(9) (2016) 1660008.
13. K. X. Li *et al.*, Bayesian network with quantitative input for maritime risk analysis, *Transportmetrica A.* **10**(2) (2014) 89–118.
14. N. Lin *et al.*, An overview on study of identification of driver behavior characteristics for automotive control, *Math. Probl. Eng.* **2014** (2014) 569109.
15. C. MacAdam *et al.*, Using neural networks to identify driving style and headway control behavior of drivers, *Vehicle Syst. Dyn.* **29**(S1) (1998) 143–160.
16. C. Miyajima *et al.*, Driver risk evaluation based on acceleration, deceleration, and steering behavior, in *IEEE Int. Conf. Acoustics, Speech and Signal Processing, 2011/ICASSP* (IEEE, 2011), pp. 1829–1832.
17. N. Mohajer *et al.*, Vehicle motion simulators, a key step towards road vehicle dynamics improvement, *Vehicle Syst. Dyn.* **53**(8) (2015) 1204–1226.
18. M. R. Othman *et al.*, Modeling driver operation behavior by linear prediction analysis and auto associative neural network, in *IEEE Int. Conf. In Systems, Man and Cybernetics, 2009/SMC* (IEEE, 2009), pp. 649–653.
19. A. Prioletti *et al.*, Part-based pedestrian detection and feature-based tracking for driver assistance: Real-time, robust algorithms, and evaluation, *IEEE Trans. Intell. Transp.* **14**(3) (2013) 1346–1359.
20. P. Raksincharoensak *et al.*, Integrated driver modelling considering state transition feature for individual adaptation of driver assistance systems, *Vehicle Syst. Dyn.* **48**(S1) (2010) 55–71.
21. X. Wang *et al.*, A two-stage fuzzy-AHP model for risk assessment of implementing green initiatives in the fashion supply chain, *Int. J. Prod. Econ.* **135**(2) (2012) 595–606.
22. Y. M. Wang *et al.*, The evidential reasoning approach for multiple attribute decision analysis using interval belief degrees, *Eur. J. Oper. Res.* **175**(1) (2006) 35–66.
23. K. F. Wu and P. P. Jovanis, Screening naturalistic driving study data for safety-critical events, *Transport. Res. Rec.* **2386** (2013) 137–146.
24. L. Xiong *et al.*, Vehicle dynamics control of four in-wheel motor drive electric vehicle using gain scheduling based on tyre cornering stiffness estimation, *Vehicle Syst. Dyn.* **50**(6) (2012) 831–846.
25. Z. Yang, S. Bonsall and J. Wang, Fuzzy rule-based Bayesian reasoning approach for prioritization of failures in FMEA, *IEEE Trans. Reliab.* **57**(3) (2008) 517–528.
26. J. B. Yang *et al.*, Belief rule-base inference methodology using the evidential reasoning approach-RIMER, *IEEE Trans. Syst. Man. Cybem. Sust.* **36**(2) (2006) 266–285.
27. Z. L. Yang *et al.*, Use of fuzzy evidential reasoning in maritime security assessment, *Risk Anal.* **29**(1) (2009) 95–120.
28. Z. L. Yang *et al.*, Selection of techniques for reducing shipping NOx and SOx emissions, *Trans. Res. D-Tr. E.* **17**(6) (2012) 478–486.
29. S. Yin *et al.*, A review on basic data-driven approaches for industrial process monitoring, *IEEE Trans. Ind. Electron.* **61**(11) (2014) 6418–6428.
30. X. Zhang, Robust gain-scheduling energy-to-peak control of vehicle lateral dynamics stabilization, *Vehicle Syst. Dyn.* **52**(3) (2014) 309–340.
31. R. Zhang *et al.*, Vehicle detection method for intelligent vehicle at night time based on video and laser information, *Int. J. Pattern. Recognit. Artif. Intell.* **2**(4) (2018) 1850009.
32. L. Zhao, X. Wang and Y. Qian, Analysis of factors that influence hazardous material transportation accidents based on Bayesian networks: a case study in China, *Saf. Sci.* **50**(4) (2012) 1049–1055.

Chuan Sun received his Ph.D. degree in Vehicle Engineering from Wuhan University of Technology, China, in 2017. He has been working as an Assistant Professor in Huanggang Normal University since 2017. His research interests include Vehicle Safety and Driving Behavior.

Chaozhong Wu is currently a Professor at Wuhan University of Technology, China. He received his Ph.D. degree from Wuhan University of Technology, China, in 2002. His research interests include Vehicle Safety.

Duanfeng Chu is currently an Associate Professor at Wuhan University of Technology, China. He received his Ph. D. degree from Wuhan University of Technology, China, in 2010. His research interests include Vehicle Safety and Vehicle Dynamics.

Zhenji Lu joined the Department of Biomechanical Engineering at TU Delft in August 2014 as a Marie Curie Fellow in the HF Auto project, He is studying human behavior in highly automated driving during transient manoeuvres such as merging, splitting, platoon entry, platoon exit, authority transitions between manual and highly automated driving, and unexpected situations, such as sensor and computer failure.

Jian Tan was born in January 1998 and raised in Enshi city, Hebei province of China. She is currently working towards her Bachelor Degree in Vehicle Engineering with Huanggang Normal University. Her research is in the areas of Vehicle Safety.

Jianyu Wang is currently a Professor at Huanggang Normal University, China. He received his Master degree from Central China Normal University, China, in 2002. His research interests include Automobile Electronic Technology.