Prediction of instantaneous driving safety in emergency scenarios based on connected vehicle basic safety messages

Kai Yu
School of Transport and Logistics, East China Jiao Tong University, Nanchang, China

Liqun Peng
School of Transport and Logistics, East China Jiao Tong University, Nanchang, China and Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Canada

Xue Ding
School of Transport and Logistics, East China Jiao Tong University, Nanchang, China

Fan Zhang
Beijing GOTEC ITS Technology Co., Ltd and Research Institute of Highway, MOT, Beijing, China, and

Minrui Chen
Jiangxi Provincial Highway Network Management Center, Nanchang, China

Abstract

Purpose – Basic safety message (BSM) is a core subset of standard protocols for connected vehicle system to transmit related safety information via vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I). Although some safety prototypes of connected vehicle have been proposed with effective strategies, few of them are fully evaluated in terms of the significance of BSM messages on performance of safety applications when in emergency.

Design/methodology/approach – To address this problem, a data fusion method is proposed to capture the vehicle crash risk by extracting critical information from raw BSMs data, such as driver volition, vehicle speed, hard accelerations and braking. Thereafter, a classification model based on information-entropy and variable precision rough set (VPRS) is used for assessing the instantaneous driving safety by fusing the BSMs data from field test, and predicting the vehicle crash risk level with the driver emergency maneuvers in the next short term.

Findings – The findings and implications are discussed for developing an improved warning and driving assistant system by using BSMs messages.

Originality/value – The findings of this study are relevant to incorporation of alerts, warnings and control assists in V2V applications of connected vehicles. Such applications can help drivers identify situations where surrounding drivers are volatile, and they may avoid dangers by taking defensive actions.

Keywords Collision avoidance system, Connected vehicle, Vehicle crash assessment, Basic safety messages

Paper type Research paper

1. Introduction

Connected vehicles (CVs) technique is advantageous to driving safety solutions by enhancing driver's perception about roadway hazards and informing driver of emergency situations that they cannot immediately recognize. The vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication technology can significantly improve the driving safety on the basis of dissimilated real-time information between connected vehicles, such as vehicle motions, driver manipulate intentions and traffic status. When vehicles share their running information with other vehicles and infrastructures, driving actions can be planned more secure and hazardless in advance.

Although the comprehensive "driver-vehicle-traffic" arrangement data have been integrated and organized by Connected Vehicle BSMs data set, there are still challenges yet to overcome. Most of the currently available Connected Vehicle applications for vehicle crash risk assessment are based on limited BSMs information with partial driving situations. Limited or partial information means that not all the attributes of the BSMs data set are taken into account, and the interactions among them are not investigated comprehensively. To solve the above problems, this paper contributes by identifying the correlation of critical driving risk in emergency cases with the driver maneuvers in period of last short term based on connected vehicle BSMs messages.

In this study, groups of field driving experiments have been conducted to collect BSMs data under potential crash threats in real road traffic. In acquiring real-world driving situation, a more comprehensive dataset is built, which contains driver characteristics, vehicle status, potential crash obstacles, road traffic environment, weather condition and driver behavior. Thereafter, variable precision rough set is used to reveal the relationship among driving safety status, which includes driver/ vehicle characteristics and road environment. Mutual information is applied to evaluate the factors that greatly influence the driving risk level. Finally, a novel similarity measurement is developed for addressing the assessment of driving safety in the classification of categorical field test data.

The remainder of the paper is organized as follows. Section 2 presents a brief overview of the state of the art research on vehicle crash risk assessment. Section 3 introduces a connected vehicle framework for vehicle crash risk warning, including BSMs data description. The illustration of modeling process for near-crash risk assessment and the field test results are respectively presented in Sections 4 and 5. The main conclusions and the future work are discussed finally in Section 6.

2. Related work

The literature reflects substantial increase in research activity with regards to connected vehicles safety applications, covering a wide range topics of how connected vehicles will be adopted and used for collision warning and driver assistance.

2.1 V2X for safety improvement

The applications for driving safety based on sensors have some limitations in pavement assessments, traffic queue estimation, vehicle routing, driving behavior monitoring and warnings, and they exist a large error at curves roads and intersections. But V2X (Vehicle to X) overcomes the limitations of the sensor-based systems to make sure that the vehicles have real and trusted information. Areas of V2X applications for vehicle safety improvement include intersection signals, pavement assessments, traffic queue estimation, vehicle routing travel time estimation, driving behavior monitoring and warnings and vehicle fuel efficiency.

Tian *et al.* (2016) propose a lane speed monitoring (LSM) application based on V2X communication. This application uses BSM which transmitted through dedicated short range communication (DSRC), to estimate the real-time traffic condition at the lane level. Yu *et al.* (2015) propose an integrated cooperative collision warning (CCW) system built around the SAE J2735 BSM over V2V communications to help drivers make right drive decision in different crash scenario, such as forward collision warning (FCW), lane change warning (LCW) and intersection collision warning (ICW). Sadek *et al.* (2017) proposed a transportation safety algorithm based on V2V. The proposed algorithm takes advantage of the information that a host vehicle can get from target vehicle to calculate the time to collision (TTC), and put thresholds for this calculated parameter to make two level collision avoidance, warning and full break according to predefined transportation scenarios. There are also a number of studies using the technology of V2V communication which have focused on investigating driving behaviors. Vehicle motion, such as speed and acceleration, is considered to the key information to describe driving behaviors. Judging the driver's driving style by the speed of the vehicle, this is a better way to address driving problems. Kim and Choi (2013) report thresholds for aggressive and extremely aggressive accelerations in urban driving environments, while De Vlieger *et al.* (2000) did similar work for calm driving, normal driving, and aggressive driving. Simons-Morton *et al.* (2012) advanced the characterization of risky driving by observing the elevated gravitational force (G-force) events that are captured when longitudinal or lateral accelerations exceed certain thresholds. A series of applications based on V2X communication have been designed to ensure the safety of drivers. The previous studies propose ideas for warnings or alerts to drivers using the connected vehicle applications, but they have not fully assimilated the value of information transmitted between connected vehicles. For example, Osman *et al.* (2015) used driving simulator based data. This study fully exploits the BSM data transmitted between vehicles and infrastructure in a real-life connected vehicle deployment. Specifically, this idea extracts useful information about risky events from new sources of data that may be generated in connected vehicle communication.

Most applications use basic safety messages (BSMs) to describe vehicles' position, motion, maneuvering, and instantaneous driving contexts. However, most BSMs describe normal driver behaviors. They do not provide information to drivers when they need to make decisions based on information received through V2X applications. In the real condition, abnormal and extreme driver behaviors determine the safety of drivers. Thus, it is essential to identify danger level from BSMs, and warn drivers to take action through the V2X applications. Our work proposes an innovative way to make real-time risk prediction on the basis of BSMs that may provide warning messages to drivers through V2X applications.

2.2 Vehicle crash risk assessment model

Safety distance (SD) model is one of the most important methods in identification of longitudinal crash risk (Rendon-Velez *et al.*, 2009). On this basis, many scholars began to use the safe distance model on the vehicle collision safety field and proposed corresponding improvement measures. As a key indicator to determinate the probability of longitudinal vehicular collision, longitudinal minimum safety distance (LMSD) got attention widely. So this model for the complex traffic situations is not good enough from the aspect of the

accuracy and the adaptability when adopting the existing safety distance model to determine the vehicle collision risk.

One of the safety distance model which are widely studied could be Time to Collision (TTC) model (Sorstedt et al., 2011; Mobus et al., 2003; Montemerlo et al., 2012). Many researchers consider TTC to inform the driver the driving condition whether is safe or not. Time to collision provides a quick estimate of the severity of a conflict to a driver and driver assistant system, and in its definition assumes that a collision is going to happen. Both SD and TTC models have been extensively applied in many modern developed in-vehicle safety systems based on Information and Communication Technology (Schubert et al., 2010; Kaempchen et al., 2009). Such systems have been expected to support the driver to maintain safe speed and headway in all driving situations by providing timely warning to driver when a critical safety situation emerges. Such systems have been expected to support the driver to maintain safe speed and headway in all driving situations by providing timely warning to driver when a critical safety situation emerges. However, the algorithms based on vehicle kinematic are very susceptible to generate false warnings, especially when driver behavior is ignored in analyzing complex traffic scenarios.

Investigations have been conducted to identify the severity of crashes using the Bayesian inference model and some data mining methodologies such as Decision Tree (DT) and Support Vector Machine (SVM). Bayesian inference model is one of the most practical tools used for analyzing crash risk (Li and Jilkov, 2003; Maybank et al., 1996; Hu et al., 2004). For example, Xu et al. (2015) developed a real-time crash risk model with limited data by using Bayesian meta-analysis and Bayesian inference approach. In his study, the fixed effect meta-analysis, the random effect meta-analysis and the meta-regression were used to formulate the informative priors of the effects of traffic variables on crash risks. Afterward, the Bayesian inference method was used to develop the crash risk models based on the informative priors obtained from Bayesian meta-analyses.

Furthermore, data mining methods have been widely used to identify the factors associated with accident severity (Petrovskaya and Thrun, 2009; Kishimoto and Oguri, 2008). For example, the Multi-Layer Perceptron and Radial Basis Functions neural networks have been applied for evaluating the traffic safety of toll plazas and the impact of electronic toll collection systems on highway safety (Gunnarsson et al., 2006).

However, neural networks models provides the prediction through black box simulation, therefore it cannot be used for supporting the retro design of vehicle collision avoidance system, which helps driver to take effective action before involvement in high risk situation in next step. In the work of Li et al. (2008), they developed vehicle crash prediction models based on Support Vector Machine (SVM) and traditional Negative Binomial (NB) model, respectively. Their evaluation of safety performance functions for vehicle crashes assessment revealed that SVM models performed better than traditional NB models.

In general, these studies present a partial review of the road traffic safety related factors, without investigating the relationships of these elements with other elements. Though some studies have made effort to address these issues by combining multiple elements (e.g. detecting the driving context, analysis of conditions, and proposing actions), their number and genuine contribution were relatively low. Driving can be interpreted as both decision-making process and as complex information processing involving perception, analysis and decision. It is a fact that effective collision avoidance system requires awareness of the actual driving situation, reliable assessment of the risks and rapid decision making about the needed assisting actions, it is also necessary to develop reasoning models, which can integrate the main constituents of a driving situation, i.e. the driver, the vehicle and the environment (including the road) with the generic phases of completing driving, i.e. perception, analysis, decision-making and action.

3. Framework

In this section, we proposed a framework to implement vehicle collision risk prediction and warning system based on the connected vehicle raw BSMs messages and extended driving safety related messages, which are compiled into standard data set that can be transmitted through dedicated short range communication (DSRC) among approached vehicles. As shown in Figure 1, the BSM information is gathered by vehicular on-board unit (OBU), containing the instantaneous vehicle position, vehicle motion and status of driver maneuver. It is synchronically received the BSMs messages of remoted vehicle delivered through the DSRC devices, which is processed to obtain the relative speed and distance to closest objects. These comprehensive BSMs message set are integrated to collision risk assessment model for informing the host vehicle

Figure 1 Schematics of connected vehicle safety applications
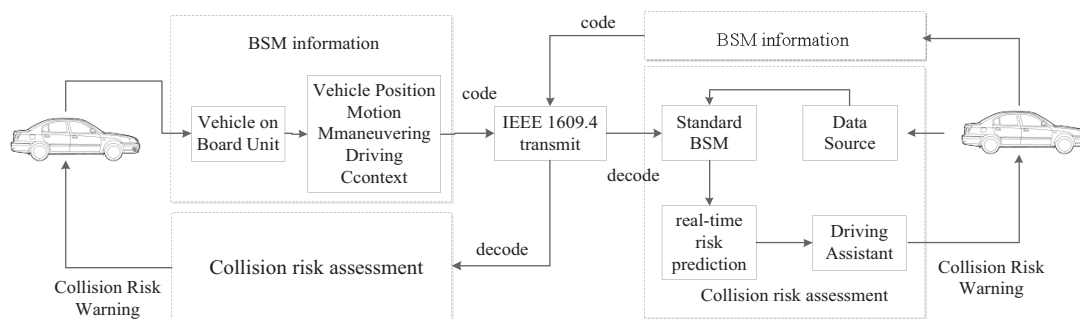
drivers to adjust their maneuver behavior, and also inform the remote vehicle drivers to avoid potential crash risk.

The emerging technique of connected vehicle request an integrated BSM data source for promoting the vehicle safety application, the standard BSM data set is specifically considered the information of vehicle motions status, driver manipulate intentions and traffic environment. These BSM data source provides opportunities and implementations in resolving the real-time vehicle risk assessment.

The MSG_Basic Safety Message is included and defined by the standard of SAE J2735 (Michaels *et al.*, 2010). The security model for this article is designed based on this message set. WAVE devices package vehicle status data according to BSM and periodically broadcast it to surrounding vehicles. The standard specifies that BSM is broadcast once for 10 milliseconds. As is shown in Table I, the basic safety message set consists of two parts. Part I is the required content, mainly including position information (longitude, latitude, altitude, location accuracy), motion information (the transmission state, speed, heading, steering wheel angle, three-axes acceleration plus yaw rate, brake system status, vehicle size (length and width).

MSG_Basic Safety Message Part II is optional and mainly contains vehicle event information including vehicle safety extension, vehicle status and other information, which still need

to be improved and formulated. For example, when a vehicle brake in an emergency, the field of Event Flag can be set to Hard Braking. All BSM communications in this paper will adopt the SAEJ2735 standard. SAEJ2735 is the equivalent of a dictionary in communication, which specifies what the communication program code block represents.

Altogether, the experiment data set included the following six major categories: Vehicle position information (longitude, latitude, relative distance); driver behavior and decision (acceleration, deceleration, steering); road obstacles (time to collision in longitudinal direction); vehicle kinematic status (velocity); road environment (Coefficient of friction between wheel and road, road segment, road slipperiness). Following previous studies, the drivers tended to adopt the rapid braking maneuver to avoid potential crash. Hence, the driving-risk level was represented by the braking process characteristics. Intuitively, the driving risk is higher if the braking maneuver is performed with greater urgency in a near-crash. The above described information is comprehensively collected in our designed field test for the purpose of analyzing the potential relationship among driving risk, driver behavior, vehicle motion and road traffic environment. The clustering braking process characteristics data were investigated to evaluate the involvement of driving risk in a near-crash event (Wang *et al.*, 2015). The distribution of these near-crashes by deceleration is summarized

Table I SAE J2735 Basic safety messages

Variable	Items	Description
BSM Part I		
Message ID	DSRC message ID	The first element in every message, used by the parser to determine how to parse the rest of the message
Message count	message count	A sequence number, incremented with each successive transmission, primarily used to estimate packet error statistics
ID	Temporary ID	A value chosen randomly and held constant for a few minutes to help the receiver correlate a stream from a given sender
Time	DSecond	Current time
Position	Latitude	Geographic latitude
	Longitude	Geographic longitude
	Elevation	Position above or below sea level
	Position Accuracy	Conveys the one-standard-deviation position error along both semi-major and semi-minor axes, and the heading of the semi-major axis
Motion	Transmission state	3 bits encode vehicle transmission
	Speed	13 bits convey unsigned vehicle speed
	Heading	Compass heading of vehicle's motion
	Steering wheel angle	Current position of the steering wheel
	Acceleration set	4ways, i.e. three axes acceleration plus yaw rate
Control	Brake system status	Conveys whether or not braking is active on each of the four wheels, also conveys the status of the following control systems: Traction Control, Anti-Lock Brakes, Stability Control, Brake Boost, and Auxiliary Brakes
Size	Vehicle size	Vehicle length and width
BSM Part II		
Safety extension	Vehicle path history	Optional
	Future vehicle path estimation	Optional
	Hard active braking	Optional
	Coefficient of fric-tion	Optional
Status	Light status	Optional
	Wiper status	Optional
	Vehicle type	Optional

in Table II. The driving-risk level in each near-crash case will be placed in one of the following three groups: low-risk, moderate-risk and high risk.

A novel analysis presented in the study will assess the vehicle crash risk with comprehensively considering cross section of the factors that could affect the drivers and their responses. The field driving experiments dataset and real time collected records were built, which was further analyzed and classified into condition attributes set C and decision attribute set D. The condition attribute data set C includes groups of data relating to factors influencing driving safety, while the decision attribute data set D contains groups of data that evaluate driving safety situation. The correlation between condition attributes and decision attribute can be expressed as $D = f(C)$. It must be noted that, the dimension of the dataset D is equal to the data set C. Then, the equation for driving safety under complex "driver-vehicle-environment" situation will be $D = f(C)$, which can be used accurately to explore vehicle crash risk with more related factor.

4. Modeling process

4.1 Variable precision rough set model

The variable precision rough set (VPRS) model allows the existence of error classification rate in a certain degree, so it has the ability to adapt to noise data, and can effectively analyze incomplete or inaccurate information. By setting the precision coefficient or inclusion β, the VPRS model loosens the strict definition of boundary in the standard rough set theory. And the reasoning process of VPRS model is simple and the calculation amount is small, so it can greatly improve the early warning response speed of traffic safety system.

As shown in Figure 2, we collect the basic safety message information and establish the BSM data set. The BSM information applied in this example are mainly composed of four parts: vehicle location, vehicle motion state, driving control behavior, traffic environment. And then the knowledge acquisition, including data preprocessing of the BSM data set, attributes reduction and other steps. The modelling steps are introduced as follows.

4.1.1 Step1: Preprocessing of the basic safety message data set

The rough set theory requires that the data set used is in the form of classification attribute, and the original BSM data is a large number of chaotic time series figures. Therefore, it is necessary to segment the raw BSM data into discrete interval, which is data preprocessing. The preprocessing information is easily expressed in the form of knowledge base, which is processed by the rough set algorithm. And then the information decision table can be formed by classifying and quantifying the conditional attributes. The initial classification of a given condition attribute value is e, $e = \{1, 2, \ldots, n\}$, whose specific value can denote different traffic or vehicle's state. And the output value of the decision attribute can be express as $f = \{1, 2, \ldots, n\}$, whose specific value can represent different driving risk level. To unify the property indexes of the attributes in the BSM dataset, the quantitation

Table II Distributed category of near-crash risk

Driving risk level	Low	Moderate	High
Deceleration when braking m/s²	(−2, 0]	(−5, −2]	(−8, −5]

criterion is proposed in Table III with explicitly considering the factors distribution from BSM and the distribution statistics of crash accidents (Wang *et al.*, 2015). The distributed category of near-crash risk used in paper is shown in Table II, and we can get driving risk level by deceleration.

4.1.2 Step 2: Attributes reduction based on rough set model

According to the quantized data, the decision-making table of driving safety state can be generated. All attributes in the table are divided into two kinds of attributes, the first type is condition attribute and the second type is decision attribute. In this model, the knowledge system of automobile real-time safety situation assessment and prediction is composed of quaternion relationship group $S = \{U, A = C \cup D, V, F\}$. Where, S is a collision event decision table based on BSMs information, and $U = \{x_1, x_2, \ldots, x_n\}$ represents all possible traffic states; A is a quantized set of BSM data including condition attribute C and decision attribute D. Suppose R is an equivalent relation on U, which can be express as U/R, and we can get $U/C = \{a_1, a_2, \ldots, a_{n-1}\}$, $U/D = \{a_n\}$. The attribute in the decision table is expressed as a_j and $a_1 - a_{n-1}$ is divided into conditional attribute, a_n is decision attribute, where $n - 1$ is the number of BSM variables.

In this model, C represents all variables of the BSMs information set, and D represents the judgment of whether there is a collision with other vehicles. The variables of the four parts of BSMs data, including the position movement state, vehicle maneuvering behavior and driving environment, are set as the condition attribute, and the last column of the data column is set as the decision attribute. According to the knowledge acquisition steps shown in Figure 2, judging whether the conditional attributes can be reduced in turn.

4.1.3 Step 3: Rules of export based on variable precision rough set model

The classical rough set model will greatly reduce its ability to predict or classify new data due to overfitting of data. To enhance the anti-interference ability of the rough set model, parameters β are introduced in the variable precision rough set (VPRS), which is supposed as $\beta \in (0.5,1]$. If $X \subseteq U$ is any subset, R is the equivalence relation on U. The improved positive region of d_j to C is defined as equation (1) (Park and Choi, 2015):

$$POS_C(d_j) = \underline{apr}_C(d_j) \tag{1}$$

Then, the lower and upper approximations of d_j can be defined as equation (2):

$$\underline{apr}_C^\beta(d_j) = \cup\{x \in U | P(d_j|C) \geq \beta\} \tag{2}$$

where, $\underline{apr}_C^\beta(d_j)$ can also be recorded as $pos_C(d_j)$. And then we can use it to reduct the attributes as mentioned in step 2. If the attributes can be reduced, the rules will be output. The form of rules is similar to IF—THEN rules. Finally, the judgment rule base of vehicle conflict risk is formed as shown in equation (3):

$$r_{i,j} : des(X_i) \rightarrow des(Y_i), X_i \cap Y_i \neq \varnothing \tag{3}$$

4.1.4 Step 4: risk assessment based on information entropy model

The new BSMs information sequence is updated and acquired in real time, to calculate the information entropy of the BSM
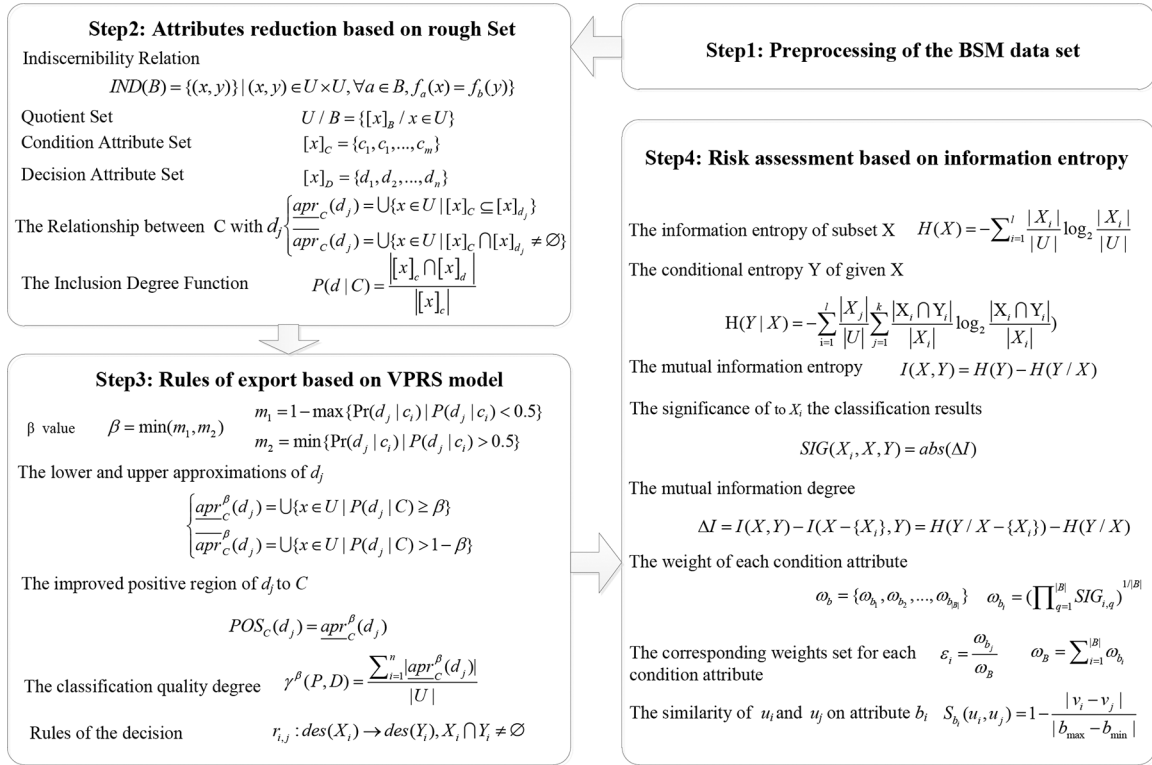
Figure 2 VPRS and information entropy flowchart

Step2: Attributes reduction based on rough Set

Indiscernibility Relation

$$IND(B) = \{(x,y)\} | (x,y) \in U \times U, \forall a \in B, f_a(x) = f_b(y)\}$$

Quotient Set $\qquad\qquad U / B = \{[x]_B / x \in U\}$

Condition Attribute Set $\qquad [x]_C = \{c_1, c_1, ..., c_m\}$

Decision Attribute Set $\qquad [x]_D = \{d_1, d_2, ..., d_n\}$

The Relationship between C with d_j $\begin{cases} \underline{apr}_C(d_j) = \bigcup\{x \in U \,|\, [x]_C \subseteq [x]_{d_j}\} \\ \overline{apr}_C(d_j) = \bigcup\{x \in U \,|\, [x]_C \cap [x]_{d_j} \neq \varnothing\} \end{cases}$

The Inclusion Degree Function $\qquad P(d|C) = \dfrac{\left|[x]_c \cap [x]_d\right|}{\left|[x]_c\right|}$

Step1: Preprocessing of the BSM data set

Step3: Rules of export based on VPRS model

β value $\qquad \beta = \min(m_1, m_2) \qquad \begin{matrix} m_1 = 1 - \max\{\Pr(d_j | c_i) | P(d_j | c_i) < 0.5\} \\ m_2 = \min\{\Pr(d_j | c_i) | P(d_j | c_i) > 0.5\} \end{matrix}$

The lower and upper approximations of d_j

$$\begin{cases} \underline{apr}_C^\beta(d_j) = \bigcup\{x \in U \,|\, P(d_j | C) \geq \beta\} \\ \overline{apr}_C^\beta(d_j) = \bigcup\{x \in U \,|\, P(d_j | C) > 1 - \beta\} \end{cases}$$

The improved positive region of d_j to C

$$POS_C(d_j) = \underline{apr}_C^\beta(d_j)$$

The classification quality degree $\quad \gamma^\beta(P, D) = \dfrac{\sum_{i=1}^n |apr_C^\beta(d_j)|}{|U|}$

Rules of the decision $\qquad r_{i,j} : des(X_i) \rightarrow des(Y_i), X_i \cap Y_i \neq \varnothing$

Step4: Risk assessment based on information entropy

The information entropy of subset X $\quad H(X) = -\sum_{i=1}^l \dfrac{|X_i|}{|U|} \log_2 \dfrac{|X_i|}{|U|}$

The conditional entropy Y of given X

$$H(Y|X) = -\sum_{i=1}^l \dfrac{|X_i|}{|U|} \sum_{j=1}^k \dfrac{|X_i \cap Y_i|}{|X_i|} \log_2 \dfrac{|X_i \cap Y_i|}{|X_i|})$$

The mutual information entropy $\qquad I(X,Y) = H(Y) - H(Y/X)$

The significance of to X_i the classification results

$$SIG(X_i, X, Y) = abs(\Delta I)$$

The mutual information degree

$$\Delta I = I(X,Y) - I(X - \{X_i\}, Y) = H(Y / X - \{X_i\}) - H(Y / X)$$

The weight of each condition attribute

$$\omega_b = \{\omega_{b_1}, \omega_{b_2}, ..., \omega_{b_{|B|}}\} \qquad \omega_{b_i} = (\prod_{q=1}^{|B|} SIG_{i,q})^{1/|B|}$$

The corresponding weights set for each condition attribute $\quad \varepsilon_i = \dfrac{\omega_{b_i}}{\omega_B} \qquad \omega_B = \sum_{i=1}^{|B|} \omega_{b_i}$

The similarity of u_i and u_j on attribute b_i $\quad S_{b_i}(u_i, u_j) = 1 - \dfrac{|v_i - v_j|}{|b_{max} - b_{min}|}$

Table III Quantitation of attribute

Attributes	Type	Description	
Driver behavior and action			
Steering	Boolean	0: No; 1: Yes	Further categorized into:
Acc pedal	Boolean	0: No; 1: Yes	1: Keep constant;
Brake switch	Boolean	0: No; 1: Yes	2: Acceleration;
Turn indicator	Boolean	0: No; 1: Yes	3: Deceleration;
			4: Steering
Road obstacles			
Vehicular distance with obstacles in longitudinal direction	Consecutive	Evaluated by TTC (time to collision, seconds) and quantified into three levels: 1: >5; 2: 2.1-5; 3: 0-2	
Vehicle kinematic status			
Velocity	Consecutive	Evaluated by km/h, quantified into four levels: 1: 0-40; 2: 41-50; 3: 51-60; 4: >60	
Deceleration	Consecutive	Evaluated by m/s^2, quantified into three levels: 1: >−2; 2: [−5, −2.1]; 3: <−5	
Steering wheel angle	Consecutive	Evaluated by °, quantified into three levels: 1: <20; 2: [20, 100]	
Weather condition			
Weather	Qualitative	1: Sunny; 2: Cloudy; 3: Rainy & Snowy	
Road slipperiness	Consecutive	Evaluated by coefficient of friction between tyre and road surface, quantified into three levels: 1: 0.7-1; 2: 0.4-0.69; 3: 0-0.39	

information, and then the similarity is calculated with the rules in the knowledge base. Considering the comprehensive similarity of two events on all conditional attributes, the information entropy can be used to express the similarity between two events. Calculate the similarity between each rule in the rule base and the new BSM information sequence, store the result in set S, and proceed to the next step after all the calculation is completed. The maximum value in the set S is selected and the rule of the event base corresponding to the maximum value is determined. The decision column result of this rule is obtained to correspond to the rule decision, and the decision result of the risk assessment of inter-vehicle conflict is obtained and output.

4.2 Illustrative example

This section further studies how to determine β values in a dynamic environment, and uses VPRS model and information entropy model to demonstrate the calculation and processing of data samples. The selected data samples in this section are from part of the BSM data from our real car test. The specific description can be seen in Section 5.1, which can prove the feasibility of the proposed model to use the BSM data to evaluate the driving safety state.

By taking considering of braking deceleration, crash danger can be quantified into three levels according to Table II, which corresponds to low, moderate and high risk, respectively. In all of 43 groups of sample data, which were collected as a decision table (DT), some events were in high crash risk level when the test drivers take braking deceleration greater than $5 \, \text{m/s}^2$, others were in low risk condition when lower than $2 \, \text{m/s}^2$ and yet others were in moderate risk condition. The dataset which can be shown in Table IV contains braking deceleration values (present the emergency in these near crash scenarios) and its influencing factors (a_1: driver behavior and action, a_2: steering wheel angle, a_3: acceleration, a_4 vehicle velocity, a_5: vehicle performance, a_6: headway to the preceding vehicle, a_7: minimum headway to the approaching vehicles in the neighbor lane, a_8: coefficient of friction between wheel and road surface), and all the eight condition attributes can be obtained in BSM data set. In this DT, some of the attributes values are consecutive variables, which need to be quantified before we investigate the correlation of the crash risk with the influencing factors and identify the risk level based on the reduct of this DT. Table IV shows the DT after the quantification.

To illustrate the application of the proposed algorithm, four steps for calculating the reduced DT and assessing vehicle crash risk in near crash scenarios is given as follows.

4.2.1 Step 1: obtaining the value of precision parameter β

In the DT shown in Table IV, $C = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8\}$ is the condition attributes set, and $D = \{d\}$ is the decision attribute set, which represents crash risk level. Based on the values of the condition attributes, a family of indiscernibility class U/C determined by C can be expressed as $X_i = \{[u_t]/u_t \in U, t \in [1,43]\}$ and a set of indiscernibility class U/D according to D can be expressed as $X_i = \{[u_t]/u_t \in U, t \in [1,43]\}$. For example, $X_3 = \{u_3, u_4\}$, $Y_3 = \{u_{10}, u_{18}, u_{29}, u_{38}, u_{43}\}$, *et al*. Totally, we can define 36 equivalence classes determined by condition attribute set C, and 3 indiscernibility classes based on decision attribute set D.

Table IV Decision table after quantification process with data from experiment

U	a₁	a₂	a₃	a₄	a₅	a₆	a₇	a₈	d
u_1	1	1	1	2	1	2	1	1	1
u_2	1	1	1	2	1	2	3	1	1
u_3	1	1	1	3	1	2	2	1	1
u_4	1	1	1	3	1	2	2	1	1
u_5	1	1	1	3	1	2	1	1	1
u_6	1	1	1	3	1	2	2	2	1
u_7	1	1	1	3	1	2	2	2	2
u_8	1	1	1	3	1	2	2	2	2
u_9	2	1	1	2	1	2	2	1	2
u_{10}	2	1	1	3	1	3	2	1	3
u_{11}	2	1	1	1	1	2	1	1	1
u_{12}	2	1	1	1	1	3	1	1	2
u_{13}	3	1	1	1	1	2	2	1	1
u_{14}	3	1	1	2	1	2	3	1	1
u_{15}	3	1	1	3	1	3	2	1	2
u_{16}	3	1	1	2	1	2	3	2	2
u_{17}	4	1	1	1	1	3	1	1	1
u_{18}	4	1	1	2	1	2	3	1	3
u_{19}	1	2	2	2	1	3	1	1	2
u_{20}	1	2	2	2	1	2	1	1	1
u_{21}	1	2	2	2	1	2	1	1	1
u_{22}	1	2	2	2	1	2	2	1	1
u_{23}	1	2	2	2	1	2	2	1	1
u_{24}	1	2	2	2	1	2	2	1	2
u_{25}	2	2	2	2	1	1	3	1	1
u_{26}	2	2	2	2	1	2	3	1	2
u_{27}	2	2	2	1	1	2	1	1	2
u_{28}	2	2	2	2	1	2	2	2	2
u_{29}	2	2	2	3	1	3	2	1	3
u_{30}	3	2	2	2	1	2	3	1	1
u_{31}	4	2	2	3	1	2	2	1	2
u_{32}	4	2	2	2	1	2	3	1	2
u_{33}	1	1	3	2	1	1	1	1	1
u_{34}	1	1	3	2	1	2	1	1	1
u_{35}	1	1	3	2	1	3	1	1	2
u_{36}	2	1	3	1	1	2	1	1	1
u_{37}	2	1	3	1	1	2	1	1	2
u_{38}	2	1	3	3	1	2	1	1	3
u_{39}	3	1	3	2	1	2	3	1	1
u_{40}	3	1	3	3	1	2	3	2	2
u_{41}	4	1	3	2	1	2	2	1	1
u_{42}	4	1	3	3	1	2	2	1	1
u_{43}	4	1	3	2	1	2	3	1	3

Then, the inclusion degrees of each conditional indiscernibility X_i to each determinal indiscernibility Y_j is respectively calculated according to Figure 2 and expressed as $P(Y_j \mid X_i)$, e.g. $P(Y_1 \mid X_1)$, $P(Y_2 \mid X_5) = 2/3$, $P(Y_3 \mid X_1) = 0$. With these inclusion degrees, the least upper bound value on β between (0.5, 1] is calculated according to Figure 2. Taking an example:

$$m_{1Y_1} = 1 - \max\{P(Y_1|X_k)|P(Y_1|X_k) < 0.5\} = 0.67$$

$$m_{2Y_1} = 1 - \min\{P(Y_1|X_k)|P(Y_1|X_k) > 0.5\} = 0.67$$

Thus, $\beta_{Y_1} = 0.67$. We calculate β_{Y_2} and β_{Y_3} by the same way, and get the results as follows:

$$\beta_{Y_2} = 0.67, \beta_{Y_3} = 1$$

Then, the percentage of effective sorting decision information D based on C is calculated with the β value on the range of (0.5, 0.67] and (0.67, 1] respectively. The quality of classification is:

$$\gamma^{\beta \in (0.5, 0.67]}(X, Y) = 0.95$$

$$\gamma^{\beta \in (0.5, 0.67]}(X, Y) = 0.81$$

Obviously, when β is equal to 0.67, it meets the requirement of the most quality of classification. Therefore, set β as 0.67 according to two propositions presented in Section 4.2.

4.2.2 Step 2: β reduct of decision table

The choice of the β reduct attribute set is another procedure, which focuses on the identification of the effective attributes that could be really related with the decision attributes, and used to form a set of deduction rules for decision making. According to the two properties, the β reduct attribute set B, where $B \subseteq C$, should satisfy that $\gamma^{\beta}(C, d) = \gamma^{\beta}(B, d)$. The reduct set B is showed in Table V.

Table V Description of rules obtained from β-reduced DT

| | If the attributes takes the values as | | | | | Then | |
Rules	a_1	a_4	a_6	a_7	a_8	d	Support
1	1	2	1	1	1	1	1
2	1	2	2	1	1	1	4
3	1	2	2	2	1	1	2
4	1	2	2	2	1	2	1
5	1	2	2	3	1	1	1
6	1	2	3	1	1	2	2
7	1	3	2	1	1	1	1
8	1	3	2	2	1	1	2
9	1	3	2	2	2	1	1
10	1	3	2	2	2	2	2
11	2	1	2	1	1	1	2
12	2	1	2	1	1	2	2
13	2	3	2	1	1	3	1
14	2	1	3	1	1	2	1
15	2	2	1	3	1	1	1
16	2	2	2	2	1	2	1
17	2	2	2	3	1	2	1
18	2	2	2	2	2	2	1
19	2	3	3	2	1	3	2
20	3	1	2	2	1	1	1
21	3	2	2	3	2	2	1
22	3	2	2	3	1	1	3
23	3	3	3	2	1	2	1
24	3	3	2	3	2	2	1
25	4	1	3	1	1	1	1
26	4	2	2	3	1	3	2
27	4	2	2	3	1	2	1
28	4	2	2	2	1	1	1
29	4	3	2	2	1	2	1
30	4	3	2	2	1	1	1

In Table V, the β - reduct produces 30 rules. Noted that the value of support means the number of which the rule is observed. The results of β -reduct show that, compared with the driver behavior and action, vehicle velocity, headway between vehicle and obstacles, and road slipperiness, the participants age and gender have a little influence on driving safety. It should be also noted that, the condition attribute a_5 (vehicle performance) is a constant in Table III, which in turn, reflect the less influence of vehicle performance on driving safety during the process of attributes reduct. However, vehicle performance is of importance attribute on decision of driving safety status. The results of attributes reduct in Table V only present the crash risk evaluation rules under condition of good vehicle performance. In addition, variable a_1 has steering action, such as $a_1=4$ means steering. So a_1 and a_2 is kind of correlated, and we can conclude the reason that variable a_2 is reduced is that a_1 may be significant, but there is no significance founded in our test because of potential correlation between a_1 and a_2.

4.2.3 Step 3: evaluating weights of condition attributes on decision-making

The importance of the variables can be obtained based on β-reduced DT by using mutual information entropy method, which can be used to quantify the influence of potential risk factors on driving risk level. Groups of indiscernibility classes U/B_{subset} and U/D determined by subsets of B and D can be expressed as $U/B - \{a_i\} = \{X_i^{B-\{a_i\}}/i = 1, 2, \ldots, |B - \{a_i\}|\}$.

The conditional entropy of D given different conditions B are calculated according to Figure 2 as follows:

$$H(D/B) = 0.3317$$
$$H(D/B - \{a_1\}) = 0.7816$$
$$H(D/B - \{a_4\}) = 0.4531$$
$$H(D/B - \{a_6\}) = 0.5387$$
$$H(D/B - \{a_7\}) = 0.4442$$
$$H(D/B - \{a_8\}) = 0.456$$

Then, the significance of each attribute in $\{a_1, a_2, a_3, a_7, a_8\}$ to the classification results can be evaluated according to Figure 2 as follows:

$$SIG = \begin{bmatrix} 1 & 3.71 & 2.17 & 3.99 & 3.62 \\ 0.27 & 1 & 0.59 & 1.08 & 0.98 \\ 0.46 & 1.71 & 1 & 1.84 & 1.67 \\ 0.25 & 0.93 & 0.54 & 1 & 0.91 \\ 0.28 & 1.02 & 0.6 & 1.1 & 1 \end{bmatrix}$$

The corresponding weight for each condition attribute is calculated according to Figure 2 as follows:

$$\omega_i = \begin{bmatrix} 2.5902 & 0.6989 & 1.1917 & 0.6477 & 0.7156 \end{bmatrix}$$

After normalization, the attribute weights are

$$\omega' = \begin{bmatrix} 0.4432 & 0.1196 & 0.2039 & 0.1108 & 0.1225 \end{bmatrix}$$

4.2.4 Step 4: decision-making on attributes weighted similarity

Actually, if the decision table is constructed with information of the driver braking deceleration under limited near-crash scenarios, the decision table cannot cover the complete cases in the field test situation. In such situations, there will be some

unseen cases which will not match the reduced classification rules extracted from the DT. This makes the classification more ambiguous. For example, the driving safety situation is evaluated with a group of detected information (the left-turn indicator is on and driver is taking steering action, vehicle velocity is 47.3 km/h, the headway to preceding vehicle is 1.25 s, minimum headway to the approaching vehicle in the target lane is 5.13 s, road adhesion coefficient is 0.7. After

quantification according to Table III, the driving safety related condition attributes set in current situation is $u_j = \{4\ 3\ 2\ 1\ 1\}$, which does not match any rules in β - reduced DT shown in Table V. Then, attributes weighted similarity measurement can be used to classify the evaluated driving safety level by searching the rule with the maximum similarity with u_j.

According to Figure 2, the weighted similarity of u_j and u_i are respectively calculated as follows:

$$S(u_1, u_j) = 0.07, S(u_2, u_j) = 0.09, S(u_3, u_j) = 0.07, S(u_4, u_j) = 0.07, S(u_5, u_j) = 0.06, S(u_6, u_j) = 0.11,$$
$$S(u_7, u_j) = 0.08, S(u_8, u_j) = 0.07, S(u_9, u_j) = 0.05, S(u_{10}, u_j) = 0.05, S(u_{11}, u_j) = 0.11, S(u_{12}, u_j) = 0.11,$$
$$S(u_{13}, u_j) = 0.11, S(u_{14}, u_j) = 0.13, S(u_{15}, u_j) = 0.07, S(u_{16}, u_j) = 0.10, S(u_{17}, u_j) = 0.09, S(u_{18}, u_j) = 0.09,$$
$$S(u_{19}, u_j) = 0.12, S(u_{20}, u_j) = 0.13, S(u_{21}, u_j) = 0.11, S(u_{22}, u_j) = 0.12, S(u_{23}, u_j) = 0.15, S(u_{24}, u_j) = 0.10,$$
$$S(u_{25}, u_j) = 0.19, S(u_{26}, u_j) = 0.15, S(u_{27}, u_j) = 0.15, S(u_{28}, u_j) = 0.16, S(u_{29}, u_j) = 0.16, S(u_{30}, u_j) = 0.16.$$

Based on the similarity results, the class of the sample u_i can be assessed using decision in u_j that maximizes $S(u_i, u_j)$:

$$\max_{30 \geq i \geq 1} S(u_i, u_j) = S(u_{25}, u_j)$$

In this case, the existed crash risk with the situation μ_j is low. The result also explains that although the headway of subject vehicle with the preceding vehicle is less than 2 s, which is thought to be a dangerous situation by safety distance model, the driving status is evaluated to be safe. Because the participant driver is steering to the target lane, on which, the longitudinal vehicular headway is satisfied for lane change.

5. Results and discussion

5.1 Experimental conditions

The field driving test was conducted using experimental vehicle equipped with on-board units (OBU) devices, which provided BSM messages according to the SAE J2735 standard. The field test was set over a route in the central part of the city of Wuhan, China. On the left of Figure 3 is the area of one test route (solid black) where the data were collected. However, the vehicle positions described in longitudinal and lateral degree could not

be directly used for evaluating the vehicles motion status in the original tested data collected. The route selected for the driving test is almost similar with the urban city traffic conditions in China, i.e. city ring road and expressway (usually low traffic volume and may have congestion). The experiments were carried out from 7:30 a.m. to 9:30 a.m. and 17:00 p.m. to 19:00 p.m. Within these time frames, the traffic flow is denser and traffic crash is more frequent, also as shown in the right part of Figure 3(a)-(f) shown in the Figure 3 are the real road scene of some near-crash events.

In this study, the driver's high deceleration behavior was considered to be a crash risk related event. The samples from field test were extracted in this study. The sum of driving time and range was approximately 51 h and over 978.9 km, respectively. In our experiments, the near-crash events were initially identified by detecting unusual vehicle kinematic data from BSMs data set. The BSM data was marked under real traffic environment, when the vehicle deceleration reached a threshold value (longitudinal: $-1\ m/s^2$, lateral: $-0.5\ m/s^2$) or TTC (time to collision) between the test vehicle and preceding vehicle is less than 3 s, the OBU recorded the vehicle state (i.e. speed, brake signal, steering signal and three-axis acceleration), the TTC with approaching vehicles in the longitudinal

Figure 3 The area of one on-road test
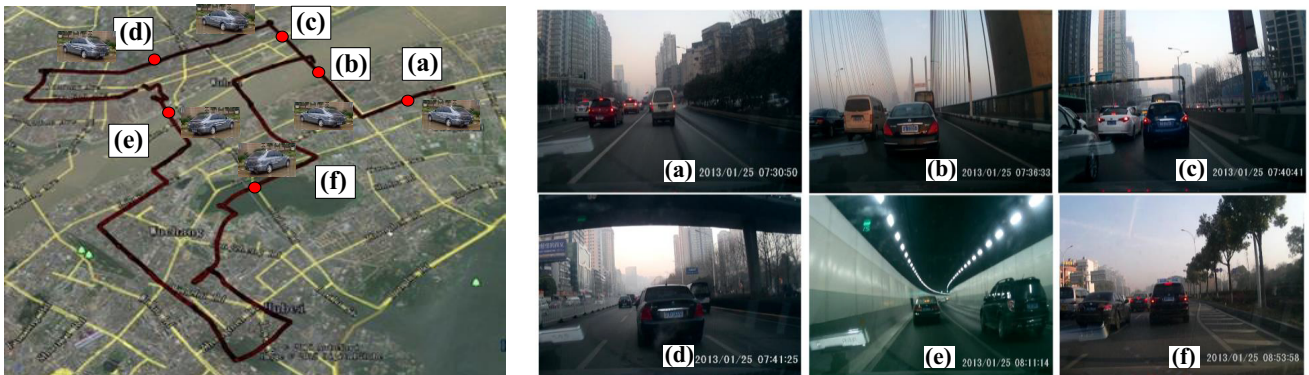
direction, while a video device synchronically recorded the extreme events happening at the time. Note that, it is very necessary to review the recorded video data to decide whether an event triggered by kinematic thresholds was actually safety critical. If not, such an event was not defined as near-crash and was deleted from the data set. For our experimental data, the recorded cases were checked manually.

5.2 Distribution of driving behavior

Figures 4 and 5 show the comparison of the velocity and acceleration of the vehicle in both the longitudinal and lateral directions, respectively. Distributions of variables seemed reasonable in terms of magnitude and spatial characteristics. Data sampled at a high frequency, 10 Hz, yielded deeper insights into instantaneous driving behaviors. This study used various data visualization tools to show the extent of instantaneous driving volatility, including distributions of longitudinal and lateral acceleration, speed-based distributions, three-dimensional distributions of longitudinal acceleration-lateral acceleration-speed, and driving volatility on different road types. This paper provides data visualization details. Then, extreme driving events will be identified in accordance with special rules.

Table VI shows the descriptive statistics of selected variables in the final datasets. Based on the error-checked descriptive statistics and the distributions, the data seemed to be of reasonably good quality. The experiment was carried out on urban roads with complex road conditions, and the operational data of vehicles in different traffic environments can be obtained. The running trajectory of the vehicle during the morning and afternoon experiments was basically the same. In

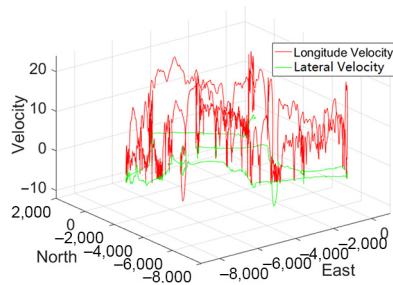**Figure 4** Longitudinal and lateral velocity

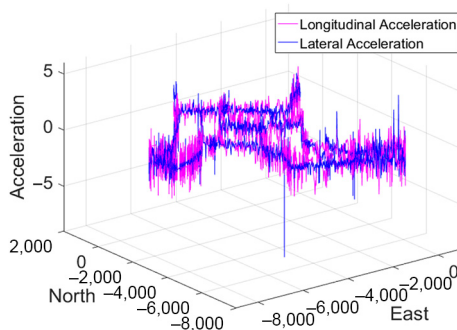**Figure 5** Longitudinal and lateral acceleration

Figure 4, the green line indicates lateral velocity, and the red line indicates longitudinal velocity. It can be seen that the lateral velocity changes are relatively stable, and the lateral velocity has a small variance value, indicating that the lateral velocity value distribution is concentrated, and is not like longitudinal velocity. The range of variation is large. In Figure 5, the blue line represents lateral acceleration, and the purple line represents longitudinal acceleration. There are several large decelerations events in the lateral acceleration of the vehicle in the longitudinal direction. Unlike Figure 4, the acceleration values in both directions are smaller and the distribution is more concentrated, but the difference between the minimum and maximum values is larger. It indicates that there is a sudden deceleration and rapid acceleration event in the running of the vehicle, but generally it will keep running smoothly. After statistics, the number of events with a forward acceleration of more than $3\,m/s^2$ or less than $-3\,m/s^2$ is 5, and the number of events with a lateral acceleration of more than $3\,m/s^2$ or less than $-3\,m/s^2$ is 4; in the pm data, the number of events with a forward acceleration of more than $3\,m/s^2$ or less than $-3\,m/s^2$ is also 5, and the lateral acceleration of more than $3\,m/s^2$ or less than $-3\,m/s^2$ is 4. The number of incidents was also four, accounting for less than one thousandth of the total number of incidents.

Previous studies indicate that extreme driving events (e.g. hard braking or acceleration) are associated with comprehensive "road-driver-vehicle" conditions, such as obstacles on roads, poor pavements, slippery road surface, sharp curves, and sensitive acceleration or braking systems, which can also be reasons for extreme driving events (Liu and Khattak, 2016). Therefore, we focused on the analysis and assessment of driving safety in near crash scenarios. Driving risk is identified as a potential threat that could cause vehicle crashes. Usually, the consequence of driving risk for a driver in hihe/sher normal state is mainly reflected by rapid evasive maneuvers (i.e. emergency braking and/or steering operation), which have been employed by many studies on naturalistic driving to identify near-crashes situations (Bareket *et al.*, 2003; Dingus *et al.*, 2005; Wang *et al.*, 2015). Near crash implies that the driver performs a rapid evasive maneuver (i.e. emergency braking and/or steering operation) that did not result in real crash.

5.3 Evaluation of driving safety

In this work, we take vehicle longitudinal emergency cases as examples to explicitly evaluate the crash risk of the test vehicle and preceding vehicle in near-crash scenarios, as described in Section 4.2. This subset is a representative sample, in which the experimental vehicle recorded all the parameters in Table III when the vehicle deceleration reached a threshold of $-1.5\,m/s^2$ or TTC was less than 3 s, the immediate data and previous sampling points were both recorded. So we extract a group of BSM data set samples in the near crash scenarios and predict driving risk level before 0.5 s by integrating different attributes as inputs. A Total of 678 groups of sample data were collected, and randomly divided into two subsets: 628 for learning and the other 50 for testing.

The actual driver acceleration/deceleration are represented with circle point and classified into three scopes m/s^2: $(-2,1.8]$, $(-5, -2]$ and $(-6,-5]$, which respectively indicate three crash

Table VI Statistical calculation of experimental data

Variable	Mean	Var	Std	Minimum	Maximum
Longitudinal velocity (*m*/s)	4.6644	29.6814	5.4481	−9.594	22.634
Lateral velocity (*m*/s)	0.0561	0.5259	0.7252	−10.602	4.173
Longitudinal acceleration (*m*/s^2)	−0.0096	0.3723	0.6102	−5.641	5.246
Lateral acceleration (*m*/s^2)	0.004	0.1599	0.3998	−3.893	6.619

risk levels. The fitted results are represented with solid dot, predicting the extent of driver's acceleration/deceleration in next short term. The longitudinal headway between vehicle on changing lane and approaching vehicles in neighbor lane is also evaluated by TTC, which is usually widely accecpted as binomial judgement for assessing vehicle crash risk by setting a threshold (Weng *et al.*, 2014) . When the TTC between the test vehicle and preceding vehicle was less than 2 s, these scenarios are viewed as risk situation.

Furthermore, the impact of "driver-vehicle-road" arrangement on the driving safety has been investigated by using different combination of the attributes vector $C = \{ c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9 \}$ as inputs, where c_1 presents driver, c_2 presents the wheel steering angle, c_3 presents the vehicle acceleration, c_4 presents vehicle velocity, c_5 presents TTC in occupied lane, c_6 presents TTC in neighbor lane, c_7 presents road segment type, c_8 presents traffic congestion, c_9 presents road slipperiness. In Figure 6, we illustrated the prediction results by respectively using rough set reduct attributes vector $\{c_1, c_4, c_5, c_6, c_9\}$, selected attributes vector $\{c_1, c_5, c_6\}$ and TTC $\{c_5\}$ as inputs. Consequently, we achieve 88.68 per cent of correct prediction before the driver take the harsh deceleration in consequent short term when using the selected attributes vector as input, while only 82.13 per cent of vehicle crash risk has been accurately predicted by using TTC as input. It indicates the significance of driver behavior and decision on the impact of safety driving, for example, when driver approaches

to preceding vehicle, they may take both acceleration and steering for lane changing if they confirm the safe longitudinal headway on both occupied lane and target lane. Although longitudinal headway of the vehicle with preceding vehicle in occupied lane gradually decreases, the host vehicle increases the lateral displacement by lateral movement and finally avoids the collision risk. We further examined the prediction performance by using reduct attributes as input according to our proposed model, then we can achieve the higher accuracy as 94.34 per cent. Although the attribute a_4 and a_8 have no direct relativity with vehicle crash risk, when comprehensively consider all attributes above, the prediction has been improved, which testify the over speeding behavior and road snippiness effectively characterize the potential vehicle crash risk. The prediction performance have been further compared by using all the attributes in vector C as inputs, as shown in Figure 7, the accuracy is 92.45 per cent, which testify the attributes c_2, c_3 and c_8 having insignificant impact on driving safety.

6. Conclusions

Connected vehicles are a relatively new and emerging area of research activity in intelligent transportation systems, with strong interest from a wide audience that includes government agencies, auto makers, practitioners and researchers who are interested in implementing connected vehicles. The findings of this study are relevant to incorporation of alerts, warnings and control assists in V2V applications of connected vehicles. Such

Figure 6 Results of vehicle collision risk prediction based on reduct attributes, selected attributes and TTC
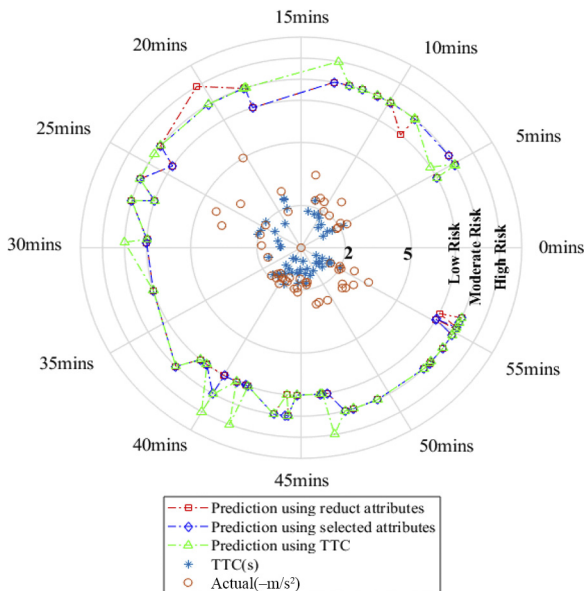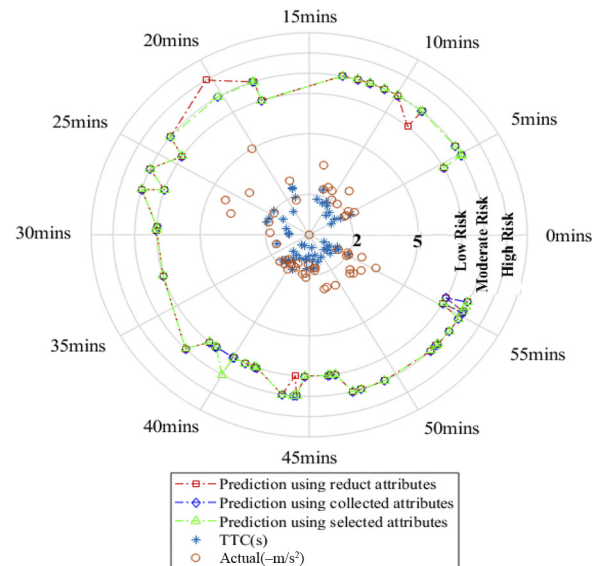

Figure 7 Results of vehicle collision risk prediction based on reduct attributes, collected attributes and selected attributes

applications can help drivers identify situations where surrounding drivers are volatile, and they may avoid dangers by taking defensive actions.

In this paper, we demonstrated how the Connected Vehicle BSM data can be linked with the real-time crash risk in certain emergency situation through the proposed machine learning method, which can be trained and validated using BSM data. By aggregating a large sample of BSM data it will be possible to learn typical driver behavior for a road network. Location specific models of driver behavior are an important resource for future safety systems. Safety systems that are able to compare observed driver behavior against a database of expected behaviors for a specific location will be better equipped to detect abnormal activity. Observing actions outside of the expected range of normal behaviors may be a strong signal that a high risk situation is developing.

Notwithstanding the extensive contribution of the study to the growing body of literature and expert information on driving safety, it must be noted that, there are some limitation in our conducted field driving test. In our current database, the influence of BSMs on the driving risk was not fully addressed. Only longitudinal driving safety situation assessment has been processed and evaluated. The time-duration of the current experiment was not very long enough to collect data under all conditions. Despite such limitations, the proposed method quantify the driving risk in near-crash event and to analyze the associated risk-factors, this can be extrapolated to specific studies on more complex scenarios. These scenarios can be constructed by the basic scenarios studied in this paper.

References

Bareket, Z., Fancher, P.S., Peng, H., Lee, K. and Assaf, C.A. (2003), "Methodology for assessing adaptive cruise control behavior", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 4 No. 3, pp. 123-131.

Dingus, T.A. Klauer, S.G. Neale, V.L. Petersen, A. Lee, S.E. Sudweeks, J. Perez, M.A. Hankey, J. Ramsey, D. Gupta, S. and Bucher, C. (2005), "The 100-Car naturalistic driving study, phase II-Results of the100-Car field experiment", National Highway Traffic Safety Administration (DOT HS 810593), Department of Transportation.

Gunnarsson, J., Svensson, L., Bengtsson, F. and Danielsson, L. (2006), "Joint driver volition classification and tracking of vehicles", *Proceeding of Nonlinear Statistical Signal Processing Workshop*, Cambridge, pp. 95-98.

Hu, W., Xiao, X., Xie, D., Tan, T. and Maybank, S. (2004), "Traffic accident prediction using 3d model based vehicle tracking", *IEEE Transactions on Vehicular Technology*, Vol. 53 No. 3, pp. 677-694.

Kaempchen, N., Schiele, B. and Dietmayer, K. (2009), "Situation assessment of an autonomous emergency brake for arbitrary vehicle-to-Vehicle collision scenarios", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 10 No. 4, pp. 678-687.

Kim, E. and Choi, E. (2013), "Estimates of critical values of aggressive acceleration from a viewpoint of fuel consumption and emissions", Presented at the *Transportation Research Board 92nd Annual Meeting*, Washington, DC.

Kishimoto, Y. and Oguri, K. (2008), "A modeling method for predicting driving behavior concerning with driver's past movementa", *IEEE International Conference on Vehicular Electronics and Safety*, Columbus, pp. 132-136.

Li, X.R. and Jilkov, V. (2003), "Survey of maneuvering target tracking. Part I. Dynamic models", *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 39 No. 4, pp. 1333-1364.

Li, X., Lord, D., Zhang, Y. and Xie, Y. (2008), "Predicting motor vehicle crashes using support vector machine models", *Accident Analysis & Prevention*, Vol. 40 No. 4, pp. 1611-1618.

Liu, J. and Khattak, A.J. (2016), "Delivering improved alerts, warnings, and control assistance using basic safety messages transmitted between connected vehicles", *Transportation Research Part C: Emerging Technologies*, Vol. 68, pp. 83-100.

Maybank, S.J., Worrall, A.D. and Sullivan, G.D. (1996), "Filter for car tracking based on acceleration and steering angle", Proceeding of British Machine Vision Conference, Edinburgh, pp. 615-624.

Michaels, C. Kelley, D. Sumner, R. Chriss, S. and Suz, D. (2010), "DSRC implementation guide a guide to users of SAE J2735 message sets over DSRC", *Communication*.

Mobus, R., Baotic, M. and Morari, M. (2003), "Multi-object adaptive cruise control", *HSCC'03 Proceedings of the 6th international conference on Hybrid systems: computation and control*, pp. 359-374.

Montemerlo, M., Becker, J. and Junior, B.S. (2012), "The Stanford entry in the urban challenge", *Journal of Field Robotics*, Vol. 25 No. 9, pp. 569-597.

Osman, O.A., Codjoe, J. and Ishak, S. (2015), "Impact of time-to-collision information on driving behavior in connected vehicle environments using a driving simulator test bed", *Journal of Traffic and Logistics Engineering*, Vol. 3 No. 1, pp. 18-24.

Park, I.K. and Choi, G.S. (2015), "A variable-precision information-entropy rough set approach for job searching", *Information Systems*, Vol. 48, pp. 279-288.

Petrovskaya, A. and Thrun, S. (2009), "Model based vehicle detection and tracking for autonomous urban driving", *Autonomous Robots*, Vol. 25, pp. 123-139.

Rendon-Velez, E., Horvath, I. and Opiyo, E.Z. (2009), "Progress with situation assessment and risk prediction in advanced driver assistance systems: a survey", *Proceedings of the 16th ITS World Congress*, pp. 21-25.

Sadek, A., Abdullah, B. and Anis, W.R. (2017), "Safety improvement in vehiclar communication systems", Presented at the *2017 IEEE 12th International Conference on Computer Engineering and Systems (ICCES)*.

Schubert, R., Schulze, K. and Wanielik, G. (2010), "Situation assessment for automatic lane-change for arbitrary maneuvers", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 11 No. 3, pp. 607-616.

Simons-Morton, B.G., Zhang, Z., Jackson, J.C. and Albert, P.S. (2012), "Do elevated gravitational-force events while driving predict crashes and near crashes", *American Journal of Epidemiology*, Vol. 175 No. 10, pp. 1075-1079.

Sorstedt, J., Svensson, L., Sandblom, F. and Hammarstrand, L. (2011), "A new vehicle motion model for improved predictions and situation assessment", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 12 No. 4, pp. 1209-1219.

Tian, D., Li, W. and Wu, G. (2016), "Evaluating the effectiveness of V2V-based lane speed monitoring application: a simulation study", Presented at the *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*.

Vlieger, I.D., Keukeleere, D.D. and Kretzschmar, J.G. (2000), "Environmental effects of driving behaviour and congestion related to passenger cars", *Atmospheric Environment*, Vol. 34 No. 4, pp. 4649-4655.

Wang, J., Zheng, Y., Li, X., Yu, C., Kodaka, K. and Li, K. (2015), "Driving risk assessment using near-crash database through data mining of tree-based model", *Accident Analysis & Prevention*, Vol. 84, pp. 54-64.

Weng, J., Meng, Q. and Yan, X. (2014), "Analysis of work zone rear-end crash risk for different vehicle-following patterns", *Accident Analysis & Prevention*, Vol. 72, pp. 449-457.

Xu, C., Wang, W., Liu, P. and Li, Z. (2015), "Calibration of crash risk models on freeways with limited real-time traffic data using Bayesian meta-analysis and Bayesian inference approach", *Accident Analysis and Prevention*, Vol. 85, pp. 207-218.

Yu, J., Kim, Y. and Mun, C. (2015), "Demo: development of an integrated cooperative collision warning system based on established standards", *IEEE Vehicular Networking Conference (VNC)*, Kyoto, pp. 173-174.

Corresponding author

Liqun Peng can be contacted at: lq.peng@ecjtu.jx.cn