Journal of Safety Research 54 (2015) 29

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Editorial

Letter from the Editors - Fourth international symposium on naturalistic driving research

The *Journal of Safety Research* is pleased to present this collection of papers that were originally presented at the *Fourth International Symposium on Naturalistic Driving Research*. The symposium, hosted by the National Surface Transportation Safety Center for Excellence (NSTSCE) at Virginia Tech, was held in August 2014. From over 40 papers and posters exploring a wide range of naturalistic driving topics, these studies have been selected through our peer-reviewed process to be presented in this special issue.

Although all of the studies included in this special issue use naturalistic driving research methods, the topics explored and analysis methods used vary widely. Studies in this collection can be roughly categorized into three broad groups:

Novice driving:

- Naturalistic teenage driving study — Findings and lessons learn
- Using naturalistic driving data to examine drivers' seatbelt use behavior, comparison between teens and adults
- Personality and crash risk
- Conducting in-depth naturalistic riding study: examples from beginner motorcyclists

Distracted driving:

- Creation of the NEST distracted driving dataset
- Are cellular phone blocking applications effective for novice teen drivers?
- Drivers' visual behavior when using handheld and hands-free cell phones
- Examination of drivers' cell phone use behavior at intersections by using naturalistic driving data

Methodological papers exploring innovative techniques in data extraction and analysis:

- Population distributions of time to collision at brake application during car following from naturalistic driving data
- Evaluation of a video-based measurement of driver heart rate
- Drunk driving detection based on classification of multivariate time series
- Naturalistic drive cycle synthesis for pickup trucks
- Older driver fitness-to-drive evaluation using naturalistic driving data

We hope you find this collection of naturalistic driving research valuable. Through programs like SHRP 2 (see accompanying letter and articles in this issue) naturalistic driving research will become more prevalent in the years to come with the potential of revolutionizing our understanding of motor vehicle safety. However, all research methodologies have limitations, and no single methodology can fully explain the complex causal nature of crashes. The Journal invites all researchers conducting rigorous evidence-based investigations, regardless of the methods used or conclusions made, to consider submitting their studies. These studies add to the understanding of us all. Only through the publishing of findings in peer-reviewed journals and through the subsequent debate on the merits of the research can the field of motor vehicle safety research advance. In this light, the Journal invites thoughtful commentary on this collection of studies.

Thomas W. Planek
Editor-in-Chief

Sergey Sinelnikov
Jonathan Thomas
Kenneth Kolosh
Associate Editors

Kathleen Porretta
Managing Editor

9 June 2015

http://dx.doi.org/10.1016/j.jsr.2015.06.003

Journal of Safety Research 54 (2015) 31

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Editorial

The 4th International Symposium on Naturalistic Driving Research

The Virginia Tech Transportation Institute is proud to have hosted the 4th International Symposium on Naturalistic Driving Research in August of 2014. The papers presented in this special issue are expanded versions of the papers and posters presented at that symposium, and they represent the first dedicated collection of papers in this new area of research. In the past 20 years, we have seen the field of naturalistic driving research expand in incredible fashion. Advances have occurred in all aspects: from vehicles with car trunks and truck cabs filled with analog recording equipment to state-of-the-art miniaturized data collection systems, from a few participants to thousands of participants per study, from manual coding of data using video tape players and spreadsheets to sophisticated data coding and extraction software, and from simple parametric statistical analysis to advanced statistical modeling techniques. Most importantly, naturalistic driving has progressed to the point that the methods, equipment, and data are now available to a wide variety of researchers.

This is what made the 4th Symposium so special: for the first time there were enough researchers doing work in the field that we were able to have a call for papers. By contrast, the three previous symposia were introductory in nature — introducing the methods, equipment, and analysis techniques to a new generation of researchers, with invited papers from those known to be working in the field.

We hope that you find the papers presented in this issue to be useful in your own research, and that you will consider adding the naturalistic driving techniques and data to your research portfolio. Most importantly, we hope that the research highlighted in this issue will provide the impetus to help save lives and improve transportation efficiency worldwide.

Jon Hankey
Senior Associate Director Virginia Tech Transportation Institute, USA

21 June 2015

http://dx.doi.org/10.1016/j.jsr.2015.06.004

Journal of Safety Research 54 (2015) 61–67

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Drunk driving detection based on classification of multivariate time series

Zhenlong Li, * Xue Jin, Xiaohua Zhao

College of Metropolitan Transportation, Beijing University of Technology, Beijing, China

ARTICLE INFO

Article history:
Received 2 December 2014
Received in revised form 23 March 2015
Accepted 23 June 2015
Available online 29 July 2015

KEYWORDS:
Drunk driving detection
Multivariate time series
Bottom-up segmentation
Support vector machine

ABSTRACT

Introduction: This paper addresses the problem of detecting drunk driving based on classification of multivariate time series. *Methods:* First, driving performance measures were collected from a test in a driving simulator located in the Traffic Research Center, Beijing University of Technology. Lateral position and steering angle were used to detect drunk driving. Second, multivariate time series analysis was performed to extract the features. A piecewise linear representation was used to represent multivariate time series. A bottom-up algorithm was then employed to separate multivariate time series. The slope and time interval of each segment were extracted as the features for classification. Third, a support vector machine classifier was used to classify driver's state into two classes (normal or drunk) according to the extracted features. *Results:* The proposed approach achieved an accuracy of 80.0%. *Conclusions and practical applications:* Drunk driving detection based on the analysis of multivariate time series is feasible and effective. The approach has implications for drunk driving detection.

1. Introduction

Drunk driving is to drive a vehicle under the influence of alcohol. Many studies have shown that drunk driving impairs driving abilities. A driver's ability to control the vehicle would be degraded under the intoxicated state. Drunk driving contributes to a great number of vehicle accidents. According to the National Highway Traffic Safety Administration (NHTSA), approximately 10,322 people died in alcohol-related crashes in 2012 in the United States, which accounted for 31% of all traffic deaths. According to a report from the World Health Organization (WHO), about 50%–60% of traffic accidents were caused by drunk driving. About 25% of fatal and injury-causing accidents were related to drunk driving (Babor, Caetano, & Casswell, 2003). In China, alcohol-related accidents accounted for 34.1% of all road accidents (Li, Xie, Nie, & Zhang, 2012). In Europe, 25% of all driver deaths in road accidents were attributed to drunk driving.

A blood alcohol test is a direct method for detecting drunk driving by measuring alcohol concentration in a driver's blood. It can be measured with instruments through blood, urine, or breath. Saab alcokey and a straw-like tube on the driver's seat are used to identify alcohol-intoxicated drivers. The alcokey is a unique alcohol-sensing instrument that was developed by Saab. The alcokey contains a small mouthpiece. The engine will not start if a breath sample from a driver indicates the driver is intoxicated. The straw-like tube, which was developed by Nissan, is installed on the driver's seat. To start the engine, a driver must blow into the straw. The engine will not start if the driver is intoxicated. Leng and Lin (2010) developed a carbon nanotube-based alcohol sensor for measuring blood alcohol concentration. However, the sensor is not mature enough to identify drunk driving.

The indirect method for detecting drunk driving is based on physiological measures, driving behavior, or vehicle states. Physiological measures include electrocardiogram (ECG), EEG or brain wave, facial expressions, eye activity, heart rate, etc. Many studies have shown that physiological measures are different between drunk drivers and normal drivers. Therefore, it is feasible to detect drunk driving through analyzing physiological measures. Kojima et al. developed a noninvasive system to detect drunk driving by physiological measures (Kojima, Maeda, Ogura, et al., 2009). Vehicle-based measures include lateral position, vehicle speed, acceleration, steering wheel movements and brake pedal, and so forth. To improve the detection accuracy, some hybrid methods combining physiological measures with vehicle-based measures were proposed. Compared to physiological measures, the main advantages of vehicle-based measures are (Eskandarian, 2012):

1) No wires and electrodes are required to connect with the drivers.
2) The processing of vehicle-based measures requires less computational power.
3) The collection of vehicle-based measures, such as steering angle, gas pedal, and brake pedal, requires less hardware compared to the collection of physiological signal.
4) The method is more suitable for engineering implementation because of the non-obtrusive characteristic.

Generally, an indirect method for detecting drunk driving consists of three stages: feature selection, feature extraction, and classification.

* Correspond author at: College of Metropolitan Transportation, Beijing University of Technology, Beijing, China.
E-mail address: lzl@bjut.edu.cn (Z. Li).

http://dx.doi.org/10.1016/j.jsr.2015.06.007

Feature selection is a process through which a subset of features is selected from the original features without a feature transformation. Feature extraction is a process that transforms the features from a high-dimensional state space into a low-dimensional one. Classification is a process that determines the class label of a new test sample. The fast Fourier transform (FFT), Gabor wavelets, Gaussian derivatives, Haar-like masks, and the wavelet packet transform (WPT) are the conventional techniques for feature extraction. Artificial neural networks (ANN), K-nearest neighbor (KNN), Bayesian networks (BN), and support vector machine (SVM) are very popular classification methods. Dai, Teng, Bai, and Shen (2010) used a mobile phone with an accelerometer and an orientation sensor to detect drunk driving. Lee et al. (2010) developed three algorithms for detecting drunk driving. One algorithm was to use logistic regression; a second was to use decision trees; a third was to use a support vector machine. Zhao, Zhang, Rong, and Ma (2011) proposed a method to identify drunk driving based on driving behavior, such as acceleration and speed. Robinel and Puzenat (2013) used a multi-layer perceptron (MLP) and support vector machine (SVM) to decide if the driver's blood alcohol concentration (BAC) was over 0.4 g/L. These results show that it is feasible to detect drunk driving based on driving behavior.

Feature selection and feature extraction are key issues for drunk driving detection. Vehicle-based measures, such as speed, steering angle, and lateral position were recorded at each time point. There are interactions and co-movements among these measures. However, the existing approaches fail to extract the feature of interactions and co-movements among a group of vehicle-based measures. A multivariate time series analysis is suitable for modeling and explaining the interactions and co-movements among a group of time series variables. Therefore, our research focused on drunk driving detection using multivariate time series. First, vehicle-based measures were collected using a driving simulator. Lateral position and steering angle were selected as the features for drunk driving detection. Second, a piecewise linear representation (PLR) was used to represent multivariate time series. A bottom-up algorithm was employed to segment multivariate time series. The slope and time interval of each segment were extracted as the features for classification. Third, a support vector machine classifier was used to classify driver's state into two classes (the normal state and the drunk state) according to the feature vector.

The rest of the paper is organized as follows: Section 2 describes the experimental design and data collection. Section 3 presents the methods, which include feature selection, feature extraction, and the classification algorithm. Experimental results are offered in Section 4. Discussions are presented in Section 5. Finally, concluding remarks and directions of ongoing work are given in Section 6.

2. Experimental design and data collection

2.1. Participants

Nagoshi, Wilson, and Rodriguez (1991) showed that driving while intoxicated, male drivers were more impulsive than female drivers. It has been found that accident risk of young drivers was higher than that of older drivers with the same BAC level (Mayhew, Donelson, Beirness, & Simpson, 1986; Zador, 1991). Accordingly, 25 healthy male drivers were recruited by advertisements. The average age was 25 with a standard deviation of 4.1, ranging from 20 to 35 years. All drivers held a valid driver's license for at least 3 years. All drivers had agreed and signed an informed consent before they participated in the experiments. Each driver was paid 400RMB for completing all experimental sessions.

2.2. EQUIPMENT

This study used a fixed-base driving simulator that was located in the Car Simulator Laboratory, Beijing University of Technology, as shown in Fig. 1. The simulator was remolded from a Toyota car. The width of the simulator is about 1.8 m. The simulator is composed of six network computers and some hardware interfaces, such as gas pedal, brake pedal, and steering wheel. A real-time, 3-D virtual scenario is displayed in front of the simulator, as shown in Fig. 2. The scenario provides a 130 degree field of view. The simulator can record driving parameters, such as lateral position, speed, steering wheel angle, and gear state.

Fig. 1. The simulator.

2.3. Scenario

A projection simulated the virtual driving scenario, which consisted of the urban straight and curve segments as shown in Fig. 2. The road was designed as bidirectional four lanes. The radii of curvature for the three left-turn curves are 200 m, 500 m, and 800 m and the radii of curvature for the three right-turn curves are 200 m, 500 m, and 800 m.

2.4. Procedure

The procedure was the same as that used in Zhao et al. (2014). To detect drunk driving, subjects were required to perform experiments at four different BAC levels. The first BAC level was 0.00%, which was actually normal driving. The second level was 0.03%, which was regarded as drink driving according to Chinese law. The third level was around 0.06% and the final level was around 0.09%, which was regarded as drunk driving. Each subject drove the simulator at four different BAC levels at intervals of 3, 5, and 7 days, respectively (Zhao et al., 2014).

Fig. 2. The scenario.

To avoid the effect of sleepiness, the experiments were performed after 2 p.m., when the participants were not drowsy according to the sleep cycle. The alcohol dose of each subject was calculated according to Watson's study (1989). The dose for the BAC level was estimated according to Eq. (1) (Watson, 1989),

$$\text{Alcohol dose (g)} = \frac{10 \times \text{BAL} \times \text{TBW}}{0.8} + 10 \times \text{MR} \times (\text{DDP} + \text{TPB}) \times \frac{\text{TBW}}{0.8} \tag{1}$$

where *BAL is the expected BAC level, TBW is the total body water, MR is the metabolic rate (generally 0.015 g/100 mL/h), DDP is the duration of the drinking period, and TPB is the time to peak BAL (generally 0.5 h)*. TBW for men is as follows (Watson, 1989):

$$\text{Men's TBW} = 2.447 - 0.09516 \times \text{Age} + 0.1074 \times \text{Height(cm)} + 0.3362 \times \text{Weight(kg)}. \tag{2}$$

For the alcohol condition, Chinese liquor with an alcohol content of 46% was used to mix into the water. The dose was calculated according to the above equations before each subject drank. The subject did not know how much alcohol he drank. A breath alcohol detector was used to measure drivers' BAC level in the experiment. The simulated driving began when the expected BAC level was achieved. Subjects were asked to sleep well at least three days before experiments. In the first experiment, subjects were instructed to operate the simulator, understand the experimental procedure and the tasks.

Vehicle-based measures include steering angle, brake pedal, gas pedal, speed, acceleration, and lateral position. The data were recorded at a 30 Hz frequency from sensors. Vehicle-based measures on the left-turn curve with the radius of curvature of 200 m with BAC levels of 0.00% and 0.09% were used to differentiate normal and intoxicated states in the paper.

3. Methods

Many studies have indicated that lateral position and steering angle are particularly useful for monitoring and detecting driving status (fatigue/drowsy/drunk/distracted driving) (Dong, Hu, Uchimura, & Murayama, 2011; Eskandarian, Sayed, et al., 2007; Eskandarian, 2012). Therefore, lateral position and steering angle were also used to detect drunk driving in the paper. A multivariate time series analysis was used to extract the features. Below is a brief introduction to the multivariate time series.

3.1. Multivariable time series

Advances in sensor technology have facilitated the generation of a large number of data in the form of time series. A time series is a collection of data obtained over time. Let $x_i(t)$, $(i = 1, \ldots, n;\ t = 1, \ldots, k)$, represent a time series. $x_i(t)$ represents the ith variable or measurement observed at time t. Generally, a matrix can denote a multivariate time series. Each column of the matrix denotes a sample at a given time, and each row represents a measurement or variable (Bankó & Abonyi, 2012).

$$\begin{bmatrix} X_1 \\ X_2 \\ \ldots \\ X_n \end{bmatrix} = \begin{bmatrix} x_1(1),\ x_1(2), \ldots, x_1(k) \\ x_2(1),\ x_2(2), \ldots, x_2(k) \\ \ldots \\ x_n(1),\ x_n(2), \ldots, x_n(k) \end{bmatrix} \tag{3}$$

As mentioned above, the body of the simulator was technologically remolded from a Toyota car. The measures, such as speed, acceleration, and steering angle, were recorded at a frequency of 30 Hz by sensors and computers. Vehicle lateral position and steering angle were selected as the measures to detect drunk driving in this study. Vehicle lateral position and steering angle were recorded sequentially over specified time intervals. A multivariate time series was formed as shown in Fig. 3 and Eq. (4). The multivariate time series can provide the information about the progression of vehicle parameters during the driving.

$$\begin{bmatrix} L \\ \theta \end{bmatrix} = \begin{bmatrix} l(1),\ l(2), \ldots, l(k) \\ \theta(1), \theta(2), \ldots, \theta(k) \end{bmatrix} \tag{4}$$

where, $l(j)$ $(j = 1, \ldots, k)$ represents vehicle lateral position at time j and $\theta(j)$ $(j = 1, \ldots, k)$ denotes steering angle at time j.

3.2. Representation and segmentation of multivariable time series

Time series representation is a major issue in time series analysis, because the efficiency of data mining can be highly affected by the representation model. Many time series representation models have been proposed, such as Fourier Transforms (Agrawal, Faloutsos, & Swami, 1993; Keogh, Chakrabarti, Pazzani, & Mehrotra, 2000), Wavelets (Chan & Fu, 1999) and Piecewise Linear Representation (PLR) (Keogh, Chu, Hart, & Pazzani, 2001). PLR is used to represent the multivariable time series in this paper because it is easy to understand and implement.

Time series segmentation can identify change points of the original time series. Accordingly, it divides the original time series into a sequence of discrete segments (Last, Klein, & Kandel, 2000). A time series $T = \{x(t), t = [1, \ldots, k]\}$ is a collection of sequential observations at time $t = 1,2,\ldots,k$. The ith segment of T, $T_i = \{x(a_i), x(a_i + 1),\ldots, x(b_i)\}$, is a set of observations between a and b. q discrete segments, $T_1, T_2, \ldots, T_q$, are separated from time series T by the boundaries $a_1 < a_2 < \ldots < a_q$, such that $a_1 = 1$, $b_q = m$, and $a_i = b_{i-1} + 1$ (Abonyi, Feil, & Nemeth, 2004). Finding the boundaries is the goal of the segmentation. To realize the goal, a cost function $cost(S(a,b))$ which describes the homogeneity or uniformity of each segment was introduced. Generally, the $cost(S(a,b))$ is the distances between the value of the original time series and the value of a simple function, such as a linear function, a polynomial function (Abonyi et al., 2004). Dynamic programming can minimize the segmentation cost of a time series. Dynamic programming, however, is difficult to implement for many real datasets (Himberg, Korpiaho, Mannila, Tikanmaki, & Toivonen, 2001). Therefore, some heuristic optimization algorithms, such as bottom-up, top-down, and sliding window algorithm, were used to segment the time series (Bankó & Abonyi, 2012):

- Bottom-up: Every element of a time series T is regarded as a segment. Calculate the cost of the adjoining segments and merge the adjoining segments into a segment if the cost is minimum.
- Top-down: The whole time series T is regarded as a segment. Split the time series and calculate the cost of every split until the goal is achieved.
- Sliding window: The first element of a time series T is regarded as an initial segment. Calculate the cost of the segment and lengthen the segment until the cost is greater than the given threshold. Take the next element as the start of the next segment and repeat the above process until the segmentation of the whole series is accomplished.

The above segmentation algorithms have their own advantages and disadvantages. It is recognized that the bottom-up algorithm can achieve the best result if the real-time processing is not required (Bankó & Abonyi, 2012).

The bottom-up segmentation algorithm was eventually adopted in this study due to the above mentioned reasons. The algorithm inputted a time series and returned a piecewise linear representation (PLR). For example, a given multivariate time series with a length of 20 s (a sample frequency of 30 Hz) shown in Fig. 3 and Eq. (4) was segmented into 10 segments using the bottom-up segmentation algorithm. That is, the multivariable time series

$$\begin{bmatrix} L \\ \theta \end{bmatrix} = \begin{bmatrix} l(1),\ l(2), \ldots, l(600) \\ \theta(1), \theta(2), \ldots, \theta(600) \end{bmatrix} \tag{5}$$

was divided into 10 segments. Fig. 4 shows the segmentation results. Specifically, 10 non-overlapping segments of lateral position are $S_1(1, 100)$ $S_2(101, 150)$, $S_3(151, 178)$, $S_4(179, 232)$, $S_5(233, 320)$, $S_6(321, 394)$, $S_7(395, 426)$, $S_8(427{,}516)$, $S_9(517{,}560)$, and $S_{10}(561{,}600)$, respectively. Ten non-overlapping segments of steering angle are $S_1(1, 82)$ $S_2(83, 146)$, $S_3(147, 238)$, $S_4(239, 262)$, $S_5(263, 290)$, $S_6(291, 322)$, $S_7(323, 462)$, $S_8(463{,}542)$, $S_9(543{,}552)$, and $S_{10}(553{,}600)$, respectively.

3.3. Feature extraction from the multivariate time series

A piecewise linear representation shown in Eq. (6) was obtained after segmentation of the multivariate time series. Simply stated, the data were divided into p segments.

$$\begin{bmatrix} L \\ \theta \end{bmatrix} = \begin{bmatrix} S_1^l\left(a_1^l, b_1^l\right),\ S_2^l\left(a_2^l, b_2^l\right), \ldots, S_p^l\left(a_p^l, b_p^l\right) \\ S_1^\theta\left(a_1^\theta, b_1^\theta\right),\ S_2^\theta\left(a_2^\theta, b_2^\theta\right), \ldots, S_p^\theta\left(a_p^\theta, b_p^\theta\right) \end{bmatrix} \tag{6}$$

First-order polynomial fitting was implemented on each segment. The slope of each segment was then obtained. The slope and time interval of each segment were extracted as the features for drunk driving detection.

$$\begin{bmatrix} L^F \\ \theta^F \end{bmatrix} = \begin{bmatrix} \left(k_1^l, w_1^l\right),\ \left(k_2^l, w_2^l\right), \ldots, \left(k_p^l, w_p^l\right) \\ \left(k_1^\theta, w_1^\theta\right),\ \left(k_2^\theta, w_2^\theta\right), \ldots, \left(k_p^\theta, w_p^\theta\right) \end{bmatrix} \tag{7}$$

where, k_i^l $(i = 1,2,\ldots,p)$ denotes the slope of ith segment of vehicle lateral position and k_i^θ $(i = 1,2,\ldots,p)$ denotes the slope of ith segment of steering angle. w_i^l $(i = 1,2,\ldots,p)$ denotes the time interval of ith segment of vehicle lateral position and w_i^θ $(i = 1,2,\ldots,p)$ denotes the time interval of ith segment of steering angle. The slope of vehicle lateral position is the change of lateral displacement with respect to time. It is described as a change in lateral displacement divided by a change in time.

$$v = \Delta L/\Delta t \tag{8}$$

The slope of steering angle is the change of angular displacement with respect to time. It is described as a change in angular displacement divided by a change in time.

$$\omega = \Delta\theta/\Delta t \tag{9}$$

Fig. 4 illustrates the principle. For example, feature matrix of the multivariate time series shown in Fig. 4 is as follows

$$\begin{bmatrix} L^F \\ \theta^F \end{bmatrix} = \left[\begin{matrix} (0.0014{,}100), (0.0103{,}50), (-0.0176{,}28), (-0.0065{,}54), (0.0098{,}88), \\ (-0.0000{,}82), (0.0006{,}64), (0.0003{,}92), (-0.0011{,}24), (0.0005{,}28), \end{matrix}\right.$$

$$\left.\begin{matrix} (-0.0005{,}74), (-0.0094{,}32), (0.0045, 90), (-0.0064{,}44), (0.0067{,}40) \\ (-0.0001{,}32)(0.0000{,}140), (-0.0000{,}80), (0.0013{,}10), (-0.0000{,}48) \end{matrix}\right]$$

3.4. Detection method for drunk driving

The goal of drunk driving detection is to divide the driver's state into two different classes: normal state or drunk state. As shown in the above section, the features have been extracted. Classification can now be performed. In this paper, a support vector machine (SVM) classifier was used. Support vector machine (SVM), which was derived from the statistical learning, was widely used for the classification. The approach has been successfully used for the verification, detection, face recognition, handwritten characters recognition, and image classification (Byun & Lee, 2002). A Support Vector Machine (SVM) classifies the data into the different groups by finding an optimal separating hyperplane. The optimal separating hyperplane can maximize the margin of the separated groups. It often happens that the dataset is not linearly separable in its original space. Therefore, a kernel function is used to map the original data into a higher-dimensional space, in which the data are linearly separable (Cristianini & Shawe-Tayor, 2000). The block diagram of drunk driving detection using a SVM classifier is given in Fig. 5.

4. Results

4.1. The results of statistical analysis

In the paper, three different numbers of segments, 10, 20, and 30, were tested. The features with BAC levels of 0.00% and 0.09% at three different numbers of segments, which include the slopes of lateral position and the slopes of steering angle, have been obtained. IBM SPSS version 20.0 was used for the statistical analysis. The mean, standard deviation, minimum, and maximum were computed. The results are summarized in Fig. 6 and Tables 1–3.

The means of the slopes of vehicle lateral position with BAC level of 0.00% at three different numbers of segments are 0.005, −0.003, and 0.013 m/s, respectively. The means of the slopes of vehicle lateral position with BAC level of 0.09% at three different numbers of segments are 0.003, 0.060, and 0.041 m/s, respectively. The standard deviations of the slopes of vehicle lateral position with BAC level of 0.00% at three different numbers of segments are 0.411, 0.394, and 0.389 m/s, respectively. The standard deviations of the slopes of vehicle lateral position with BAC level of 0.09% at three different numbers of segments are 0.824, 0.900, and 0.889 m/s, respectively. The maximum values with BAC level of 0.00% at three different numbers of segments are 1.550, 1.606, and 1.646 m/s, respectively. The minimum values with BAC level of 0.00% at three different numbers of segments are −1.266, −1.275, and −1.538 m/s, respectively. Accordingly, the ranges with BAC level of 0.00% at three different numbers of segments are 2.816, 2.881, and 3.184 m/s, respectively. The maximum values with BAC level of 0.09% at three different numbers of segments are 3.778, 6.979, and 6.979 m/s, respectively. The minimum values with BAC level of 0.09% at three different numbers of segments are −3.981, −4.072, and −4.072 m/s, respectively. Accordingly, the ranges with BAC level of 0.09% at three different numbers of segments are 7.759, 11.051, and 11.051 m/s, respectively.

Four indicators, the mean and standard deviation of vehicle lateral position, the mean and standard deviation of steering wheel angle, were analyzed. The analysis of variance (ANOVA) was used to determine whether there are any significant difference among the four indicators obtained during normal driving and drunk driving. Table 3 shows the significant difference results. The results show that the standard deviation of the slopes of vehicle lateral position is significantly higher in drunk driving than that in normal driving ($P < 0.05$). The standard deviation of the slopes of steering angle is significantly higher in drunk driving than that in normal driving ($P < 0.01$). Table 3 indicates that there is no significant difference between the means of steering angles during normal driving and drunk driving. Table 3 also indicates that there is significant difference between the means of vehicle lateral position during normal driving and drunk driving when the number of segments is 20. The results imply that lateral position and steering angle are useful for drunk driving detection, which is the foundation for identifying drunk driving.

4.2. The results of drunk driving detection

Our experiments were conducted on a Windows 7 machine with a 2.50 GHz CPU and 4.0 GB of RAM. In these experiments, a radial basis function (RBF) kernel was used. The choices of the kernel parameter σ and the cost parameter C have a great effect on the accuracy of a SVM

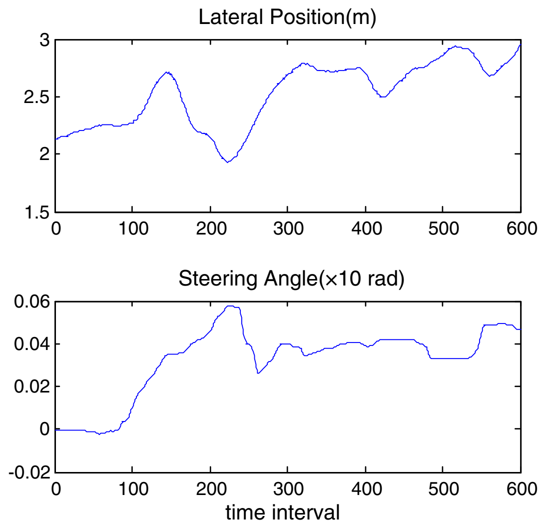


Fig. 3. Progression of lateral position and steering angle signal.

classifier. It is unknown beforehand which C and σ are better for the SVM classifier. Consequently, the methods of parameter search must be employed. Five-fold cross-validation and the grid search were performed to find the optimal (C, σ) in the paper. There were 20 drivers to carry out the whole experiment. Each driver drove the simulator at two states. Therefore, 40 samples were obtained. The samples were randomly separated into two parts (a training set of 25 samples and a test set of 15 samples).

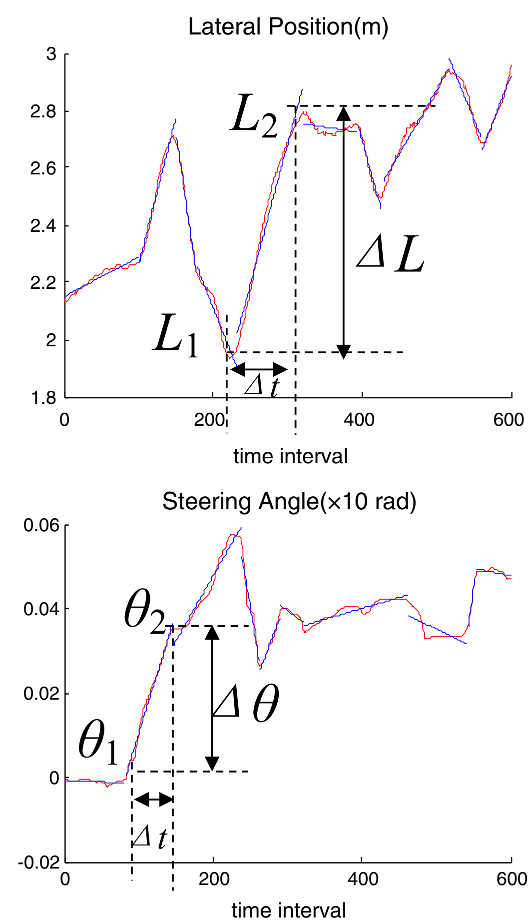


Fig. 4. The segmentation results.

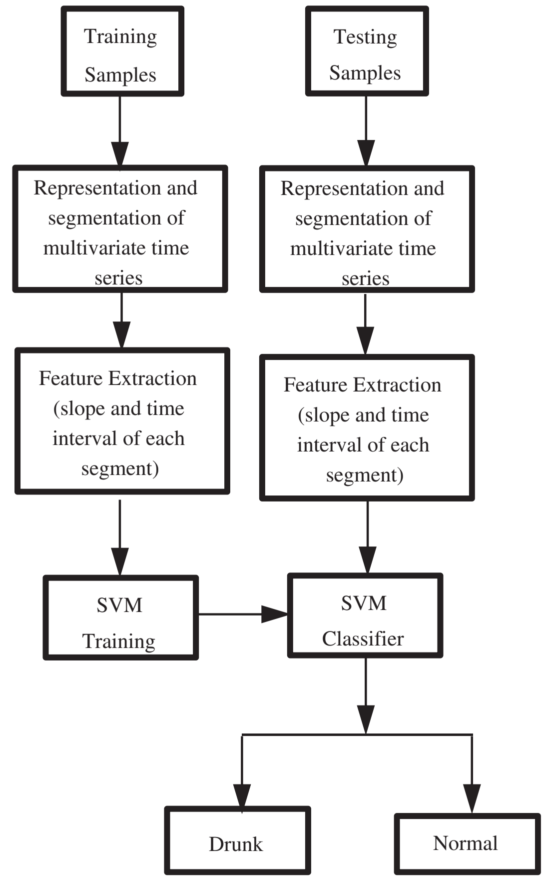


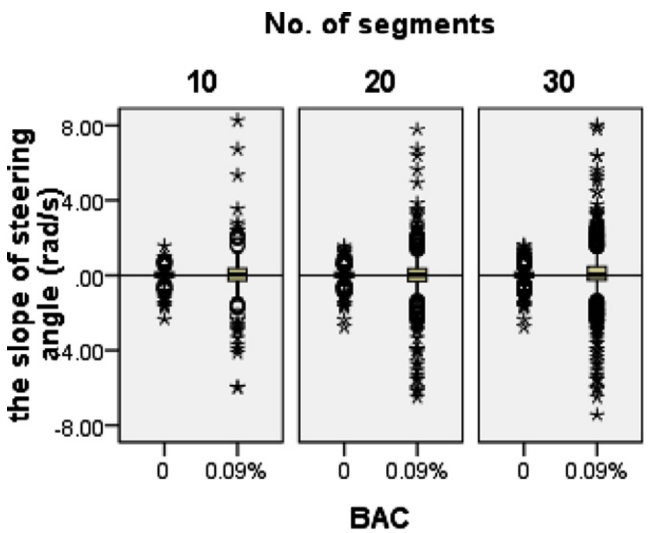
Fig. 5. Block diagram of drunk driving detection.

The models were fitted on the training set (including the choice of the parameter (C, σ)) and the performance was tested on the test set. The slope and time interval of each segment of lateral position and steering angle were selected as a feature vector for the SVM classifier. Three different numbers of segments, 10, 20, and 30, were tested in the paper. For each number of segments, the classifier was trained and tested. The optimal (C, σ) for different numbers of segments are shown in Table 4.

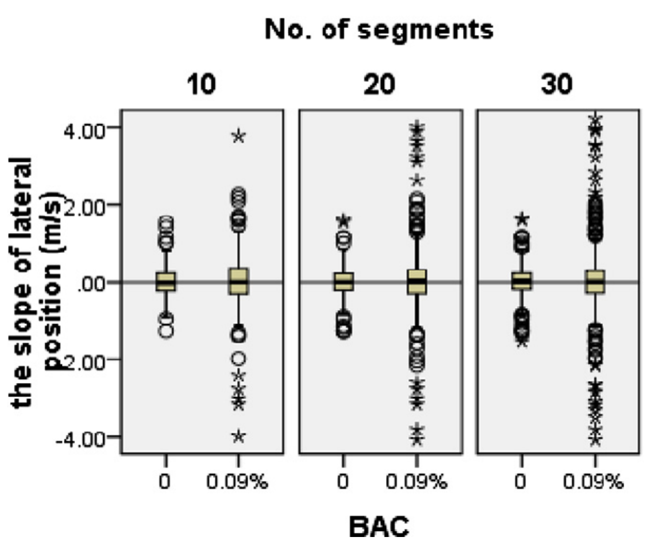
Detection outcomes were labeled either as drunk (Positive) or normal (Negative). The classifier performance was evaluated with three different measures: true positive rate (TPR), false positive rate (FPR), and accuracy (ACC). TPR, FPR, and ACC of the classifiers were calculated. Table 5 presents the performance of the SVM classifier when different numbers of segments are considered.

5. Discussions

From Tables 1–2 and Fig. 6, we can see that alcohol affects the slopes of vehicle lateral position and the slopes of steering angle. The slopes of vehicle lateral position and the slopes of steering angle seem to increase with the blood alcohol concentration. The range also tends to increase with the blood alcohol concentration. Standard deviations of the slopes of vehicle lateral position and the slopes of steering angle during drunk driving are greater than that during normal driving. The ranges of the slopes of vehicle lateral position and the slopes of steering angle during drunk driving are larger than that during normal driving. It implies that vehicle lateral position and steering angle during drunk driving are more variability and more spread out. These indicate that the ability of judgment and operation are impaired to some extent by alcohol, which corresponds with the previous research. The number of segments has some effect on the slopes; however the effect seems to be small.



a) The boxplot of the slopes of steering angle



b) The boxplot of the slopes of lateral position

Fig. 6. The boxplot. (6-a) The boxplot of the slopes of steering angle. (6-b) The boxplot of the slopes of lateral position.

An accuracy of 80% was achieved using a SVM classifier in the paper. The accuracy of the algorithms proposed by Lee et al. was 73% to 86%, which was achieved with approximately 8 min of driving performance measures. Driving performance measures were mean speed, standard deviation of speed, and standard deviation of lane position in that study. Driving performance measures used in the study are the slopes of steering angle and the slopes of vehicle lateral position. Approximately 25 s of driving performance measures were used to detect drunk driving. Driving scenarios in Lee's study were composed of an urban segment, a rural segment, and a freeway segment, while driving scenarios were composed of the urban straight and curve segments in this study. The detection results of this study are encouraging. It can be seen that the classifiers achieved the satisfactory results. Drunk driving detection based on the classification of a multivariate time series is feasible and effective. A piecewise linear representation of the multivariate time series is succinct because a large number of time series data are represented by a few line segments. At the same time, the line segments can represent the important information of multivariate time series. The piecewise linear representation can extract the feature of interactions and co-movements between lateral position and steering angle. The PLR requires the smaller storage, while the PLR-based computation is more efficient. A succinct and representative PLR is desirable for drunk driving. From Table 5, we observe that the number of segments has a significant effect on the results and the best value for the number of segments is 20. In this case, the SVM classifier achieved the highest accuracy (80.0%). It seems that the piecewise linear representation with a smaller or larger number of segments cannot accurately represent the primary feature of the multivariate time series. The number of line segments must be determined. Developing an adaptive method for deciding the number of line segments is highly necessary, which is also a topic worthy of further investigation.

Table 1
The descriptive statistics of the slopes of lateral position.

BAC (%)	*0.00*			*0.09*		
No. of segments	10	20	30	10	20	30
Mean (m/s)	0.005	−0.003	0.013	0.003	0.060	0.041
SD (m/s)	0.411	0.394	0.389	0.824	0.900	0.889
Min (m/s)	−1.266	−1.275	−1.538	−3.981	−4.072	−4.072
Max (m/s)	1.550	1.606	1.646	3.778	6.979	6.979
Kurtosis	1.542	1.725	1.900	6.825	13.849	12.007
Skewness	0.359	0.013	−0.082	−0.580	1.336	1.076

Table 2
The descriptive statistics of the slopes of steering angle.

BAC (%)	*0.00*			*0.09*		
No. of segments	10	20	30	10	20	30
Mean (rad/s)	−0.017	−0.016	−0.004	−0.022	−0.073	0.009
SD (rad/s)	0.449	0.443	0.436	1.604	1.739	1.705
Min (rad/s)	−2.334	−2.766	−2.766	−9.103	−11.325	−11.325
Max (rad/s)	1.542	1.542	1.586	8.255	10.427	10.427
Kurtosis	6.277	7.346	6.126	11.399	10.988	10.099
Skewness	−1.556	−1.240	−0.884	−0.414	−0.357	−0.064

6. Conclusion

This paper addresses the problem of drunk driving detection based on a classification of the multivariate time series. The study shows that alcohol affects the slopes of vehicle lateral position and the slopes of steering angle. The slopes of vehicle lateral position and the slopes of steering angle seem to increase with the blood alcohol concentration. Accordingly, lateral position and steering angle are selected for drunk driving detection. Lateral position and steering angle were recorded at each time point during driving. A multivariate time series item was formed. A bottom-up segmentation algorithm was used to segment multivariate time series. A piecewise linear representation of the multivariate time series is succinct because a large number of time series data are represented by a few line segments. At the same time, the line segments can represent the important information of multivariate time series. A succinct and representative PLR is desirable for drunk driving detection. The slope and time interval of each segment of vehicle lateral position and steering angle were extracted as the feature vector. A SVM classifier was used to classify driver's state into two classes. The experimental results demonstrated that drunk driving detection based on the classification of the multivariate time series is feasible and effective. The number of segments has a significant effect on the detection accuracy. An effective mechanism of the number of segments on the accuracy warrants further studies. Future research is to develop an adaptive method for determining the number of line segments in order to improve the detection accuracy.

A limitation of the study is that the classification algorithm was only tested on urban curves and only differentiated between normal state

Table 3
The significant difference results.

Features	The slopes of lateral position						The slopes of steering angle					
No. of segments	10		20		30		10		20		30	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
P-value	0.940	0.040	0.016	0.015	0.383	0.017	0.740	0.001	0.546	0.000	0.740	0.000

(BAC level of 0.00%) and drunk state (BAC level of 0.09%). Another limitation is the relatively small sample size. However, the proposed approach to detecting drunk driving is feasible. Future studies will employ a larger sample size.

Table 4
The optimal (C, σ).

Parameters	No. of segments		
	10	20	30
C	0.0039	0.0039	0.0039
σ	4.00	5.66	8.00

Table 5
The performance of SVM classifier.

Performance measures	No. of segments		
	10	20	30
TPR	50.5%	87.5%	75.0%
FPR	14.3%	28.6%	28.6%
ACC	66.7%	80.0%	73.30%

References

Abonyi, J., Feil, B., Nemeth, S., & Avra, P. Principal component analysis based time series segmentation: A new sensor fusion algorithm. Preprint, 2004.

Agrawal, R., Faloutsos, C., & Swami, A. (1993). Efficient similarity search in sequence databases. *Proceedings of the 4th Conference on Foundations of Data Organization and Algorithms*.

Babor, T., Caetano, R., & Casswell, S. (2003). *Alcohol: No ordinary commodity research and policy*. Oxford: Oxford University Press. London.

Bankó, Z., & Abonyi, J. (2012). Correlation based dynamic time warping of multivariate time series. *Expert Systems with Applications*, *39*(17), 12814–12823.

Byun, H., & Lee, S. W. (2002). Applications of support vector machines for pattern recognition: a survey. *Proc. 1st Int. Workshop, SVM Pattern Recog. With Support Vector Mach., Niagara Falls, ON, Canada* (pp. 213–236).

Chan, K., & Fu, W. (1999). Efficient time series matching by wavelets. *Proceedings of the 15th IEEE International Conference on Data, Engineering* (pp. 126–133).

Cristianini, N., & Shawe-Tayor, J. (2000). *An introduction to support vector machines*. Cambridge: The press syndicate of the university of Cambridge.

Dai, J. P., Teng, J., Bai, X. L., et al. (2010, March). Mobile phone based drunk driving detection pervasive computing technologies for healthcare 2010. *4th International IEEE Conference* (pp. 1–8).

Dong, Y. C., Hu, Z. C., Uchimura, K., & Murayama, N. (2011). Driver inattention monitoring system for intelligent vehicles: A review. *IEEE Transactions on Intelligent Transportation Systems*, *12*(2), 596–614.

Eskandarian, A. (Ed.). (2012). *Handbook of intelligent vehicles drowsy and fatigued driving problem significance and detection based on driver control functions*. Springer-Verlag London Ltd.

Eskandarian, A., Sayed, R., et al. (2007). *Advanced driver fatigue research*. FMCSA-RRR-07-001.

Himberg, J., Korpiaho, K., Mannila, H., Tikanmaki, J., & Toivonen, H. (2001). Time series segmentation for context recognition in mobile devices. *ICDM* (pp. 203–210).

Keogh, E., Chakrabarti, K., Pazzani, M., & Mehrotra (2000). Dimensionality reduction for fast similarity search in large time series databases. *Journal of Knowledge and Information Systems*, 263–286.

Keogh, E., Chu, S., Hart, D., & Pazzani, M. (2001). An online algorithm for segmenting time series. *Proc. of IEEE International Conf. on Data Mining* (pp. 289–296). Los Alamitos: IEEE Computer Society Press.

Kojima, S., Maeda, S., Ogura, Y., et al. (2009). Non-invasive biological sensor system for detection of drunk driving. *Information Technology & Applications in Biomedicine. 9th International Conference*, 1–4.

Last, M., Klein, Y., & Kandel, A. (2000). Knowledge discovery in time series databases. *IEEE Transactions on Systems, Man, and Cybernetics*, *31*(1), 160–169.

Lee, John D., Fiorentino, D., Reyes, M. L., Brown, T. L., Ahmad, O., Fell, J. C., et al. (2010). *Assessing the feasibility of vehicle-based sensors to detect alcohol impairment*. DOT HS 811–358 Washington, DC: National Highway Traffic Safety Administration.

Leng, H. J., & Lin, Y. Z. (2010, June). Design and experimental study of a CNT sensor for measuring alcohol content with short response delay. *IEEE Sensors Journal*, *10*(6), 1091–1097.

Li, Y., Xie, D., Nie, G., & Zhang, J. (2012). The drink driving situation in China. *Traffic Injury Prevention*, *13*(2), 101–108.

Mayhew, D. R., Donelson, A. C., Beirness, D. J., & Simpson, H. M. (1986). Youth, alcohol and relative risk of crash involvement. *Accident Analysis and Prevention*, *18*(4), 273–287.

Nagoshi, C. T., Wilson, J. R., & Rodriguez, L. A. (1991). Impulsivity, sensation seeking, and behavioral and emotional responses to alcohol. *Alcoholism*, *15*(4), 661–667.

Robinel, A., & Puzenat, D. (2013, April). Multi-user blood alcohol content estimation in a realistic simulator using artificial neural networks and support vector machines. *Proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning* (pp. 24–26) (Bruges (Belgium)).

Watson, P. E. (1989). *Total body water and blood alcohol levels: Updating the fundamentals. Human Metabolism of Alcohol*, *Vol. 1*. (pp. 41–58), 41–58.

Zador, P. L. (1991). Alcohol-related relative risk of fatal driver injuries in relation to driver age and sex. *Journal of Studies on Alcohol*, *52*(4), 302–310.

Zhao, X. H., Zhang, X. J., Rong, J., & Ma, J. M. (2011). Identifying method of drunk driving based on driving behavior. *International Journal Of Computational Intelligence Systems*, *3–4*, 361–369.

Zhao, X. H., Zhang, X. J., & Rong, J. (2014). Study of the effects of alcohol on drivers and driving performance on straight road. *Mathematical Problems in Engineering*, *2014*, 9. http://dx.doi.org/10.1155/2014/607652.

Zhenlong Li is an associate professor in the College of Metropolitan Transportation of Beijing University of Technology. He received his Ph.D. from Chinese Academy of Sciences. His primary research interest is in traffic control and driving behavior.

Xue Jin is a postgraduate student at Beijing University of Technology.

Xiaohua Zhao is an associate professor in the College of Metropolitan Transportation of Beijing University of Technology. She received her Ph.D. from Beijing University of Technology. Her primary research interest is in driving behavior.