Driving Behavior Analysis of Multiple Information Fusion Based on SVM

Jeng-Shyang Pan[1], Kaixuan Lu[1], Shi-Huang Chen[2], and Lijun Yan[1]

[1] Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, China
[2] Department of Computer Science and Information Engineering, Shu-Te University, Kaohsiung County, 824, Taiwan, ROC
jengshyangpan@gmail.com, lukaixuan203@sina.com, shchen@stu.edu.tw, yanlijun@126.com

Abstract. With the increase in the number of private cars as well as the non-professional drivers, the current traffic environment is in urgent need of driving assist equipment to timely reminder and to rectify the incorrect driving behavior. To meet this requirement, this paper proposes an innovative algorithm of driving behavior analysis based on support vector machine (SVM) with a variety of driving operation and traffic information. The proposed driving behavior analysis algorithm will mainly monitor driver's driving operation behavior, including steering wheel angle, brake force, and throttle position. To increase the accuracy of driving behavior analysis, the proposed algorithm also takes road conditions, including urban roads, mountain roads, and highways into account. The proposed will make use of SVM to create a driving behavior model in various different road conditions, and then could determine whether the current driving behavior belongs to safe driving. Experimental results show the correctness of the proposed driving behavior analysis algorithm can achieve average 80% accuracy rate in various driving simulations. The proposed algorithm has the potential of applying to real-world driver assistance system.

Keywords: Driving behavior analysis, driver assistance system, SVM.

1 Introduction

Because the continuously increases of the global car ownership, it is unavoidable to raise the traffic density and number of non-professional drivers. This also leads to frequent traffic accidents which have become the first hazard of modern society [1]. Among these traffic accidents, the improper driving behavior habits are an important cause of crashes. Therefore the study of driving behavior analysis has become extremely useful.

Thanks to the modern vehicle manufacturing technology, the safety factors of the vehicle itself caused traffic accidents are smaller and smaller proportion. However, the driver personal factors caused traffic accidents have become the main reason for causing a traffic accident [2]. Toyota's Masahiro MIYAJI, Jiangsu

A. Moonis et al. (Eds.): IEA/AIE 2014, Part I, LNAI 8481, pp. 60–69, 2014.

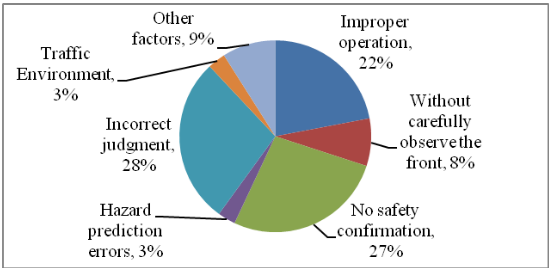

Fig. 1. Distribution of Traffic Accidents Statistics

Universitys Liang Jun [3][4] and other scholars count and analyze the specific reasons of traffic accidents and the results are shown in Fig. 1. From these results, it is observable that the individual factor account for the accidents are more than 90%. In addition to the driver and pedestrian essential training and education, it also needs to monitor and predict driving behavior of the drivers to prevent the traffic accidents and increase the traffic safety.

Many scholars have made a great of contributions in the study of driving behavior analysis. Yoshifumi Kishimoto and Koji Oguri propose a method of modeling driving behavior concerned with certain period of past movements by using AR-HMM (Auto-Regressive Hidden Markov Model) in order to predict stop probability [5]. Antonio Prez, M. Isabel Garca, Manuel Nieto, et.al present Argos, which is a complex and powerfully computerized car to help researchers in the study of car driver behavior [6]. Reza Haghighi Osgouei and Seungmoon Choi address a model-based objective measure for the evaluation of driving skills between different drivers. This metric is based on a stochastic distance between a pair of hidden Markov models (HMMs) each of which is trained for an individual driver [7]. Hong-mao Qin, Zhiqiang Liu and Peng Wang In order to overcome the limitations of single-channel information in the determination of drowsy driving behaviora method was proposed based on multi-channel information fusion [8]. The [9] is proposed for a realtime traffic surveillance system which can help the driver to get more surrounding traffic information.

On the basis of the above study, this paper proposed an algorithm to capture the driver operation and then to analysis the driver behavior and status. The proposed algorithm can assess whether the current driving behavior to make the car in a safe state. If necessary, it will provide the appropriate prompts or operation to make sure the vehicle from a dangerous state and back to a safe state.

The remaining sections are organized as follows. Section 2 brief introduces the SVM theorem. The details of the proposed driving behavior analysis algorithm are presented in Section 3. Section 4 shows experimental results and performance analysis. Finally, Section 5 concludes this paper.

2 Support Vector Machines

In this section we provide a brief review of the theory behind this type of algorithm; Support Vector Machines (SVMs) is developed from the optimal hyper-plane in the case of linearly separable, for more details we refer the reader to [10][11].

Assuming the set of training data $(x_i, y_i), i = 1, \cdots, n$, can be separated into two classes, where $x \in R^d$ is a feature vector and $y \in \{+1, -1\}$ its class label. they can be separated by a hyperplane $H : w \cdot x + b = 0$, and we have no prior knowledge about the data distribution, then the optimal hyperplane is the one which maximizes the margin[11]. The optimal values for w and b can be used to found by solving a constrained minimization problem, can be transform the optimal hyper-plane problem to its dual problem.Lagrange multipliers $\alpha_i\,(i = 1, \cdots, m)$

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{n} \alpha_i{}^* y_i K(x_i, x) + b^*\right) \tag{1}$$

where $\alpha_i{}^*$ and b^* are found by using an SVC learning algorithm [11]. Those x_i with nonzero $\alpha_i{}^*$ are the support vectors. For $K(x, y) = x \cdot y$, this corresponds to constructing an optimal separating hyperplane in the input space R^d.in this paper we use the kernel Radial basis function: $K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right)$, $\gamma > 0$ is an interval relaxation vector.

3 Driving Behavior Analysis System

The proposed driving behavior analysis system consists of driving operation acquisition module, data preprocessing module, the driving operation information fusion module, and SVM classification and recognition modules.

Driving behavior analysis data will be divided into training set and test set. Preprocessing and feature extraction are simultaneously applied to both sets. Classification makes the test samples into the driving model based on SVM algorithm to classify and determine the test sample category. The number of rightly or wrongly classified samples divided by the number of the test set samples is the classification correct rate or error rate.

For example, the following use the driving data in urban traffic road to illustrate the system processes. In the data acquisition module, at first, a good driving data and bad driving data should be collected as training set. Then collecting another data set includes good driving behavior data and bad driving behavior data as a test set using the same method. After data preprocess step, each time slice samples can be regarded as the rate of change the driving operation. We use the training set to establish the driving classification model in the city as a judge model by the SVM theory, and then we use the test se to judgment the accuracy of the model. Finally, judge the merits of driving behavior. Fig. 2 shows the flow of the entire system.

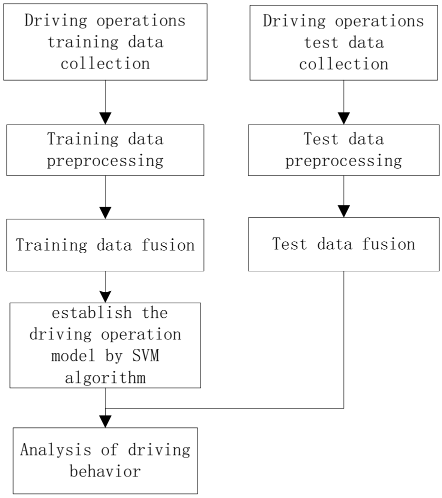

Fig. 2. The flow of the driving behavior analysis

4 Experimental results and analysis

4.1 Experimental Equipment

In this study, driving operation data is simulated through the YOYO driving training machine, produced by Guangzhou Great Gold East Network Technology Co. Driving operation data are read through the USB interface and combined with the simulation environment (city, mountains, and highway) to analysis the driving behavior. Fig. 3 show the data collection equipment and driving environment system.

4.2 Data Collection

In the PC terminal, we used the data collection program (Bus Hound 6.01) real-time recording the driving operation data when driver driving simulation. Each record driving time is about 5-15 minutes and driving environment are selected from the city, high-speed and mountainous road. Each of driving environment corresponds to simulate two driving data set and every set includes good driving behavior and bad driving behavior, respectively for training set and testing set. One of the driving data set as a training set is used to establish city driving model by SVM. The other driving set as the test set is used to test validity about the above driving model established by SVM. The method of collecting driving data is described as follows:

(1) Urban Roads Good driving record: maintain speed below 35 km/h, try to reduce the use of brake, deceleration in advance, try to keep a straight line driving, cornering speeds below 15 km/h, acceleration and braking are operated slowly. Bad driving record: Keep the car at speed driving state, quick step on

Fig. 3. Data collection equipment and driving environment system

and quick release the accelerator brake, Remain uncertain direction and curve driving.

(2) Highway Good driving record: maintain speed between 80 km/h and 100 km/h, Keep the car was going along a straight line with almost no brake operation, No big turning in high-speed. Bad driving record: speed is too high and instability, curve driving, quick step on and quick release the accelerator brake, often braking.

(3) Mountain road Good record: maintain speed below 40 km/h, turning speed below 20 km/h. Remained constant speed. Braking slowly and downshift in downhill. Bad driving record: speed is too high and instability, quick step on and quick release the accelerator brake, curve driving Finally obtain the following 12 groups data: Training set: good city driving, bad city driving, good mountain driving, bad mountain driving, good highway driving, bad highway driving Testing set: good city driving, bad city driving, good mountain driving, bad mountain driving, good highway driving, bad highway driveing

4.3 Data Preprocessing

he data record format is shown in Fig. 4.

(1) Extract data packet from the "data" column of the Fig. 4. The packet is converted from hex to decimal. These data include steering, brakes, and accelerator information. The first column is the steering wheel angle information, expressed by the number from 0 to 255. The second column is the turn left and turn right information: Digital 1 represents the first circle turn to the left; 0 represents the second circle turn to the left; 2 represents the first circle turn to the right; 3 represents the second circle turn to the right. In order to facilitate the data analysis, the steering angle information will be converted into a continuous angle data from -510 to 510, negative numbers indicate turn left, positive number indicate turn right, Where each number represents 1.4118 degree angle. The

Device	Phase	Data		Description	Cmd.Phase.Ofs(rep)
13.1	IN	00 02 80 80	00 19 10 55	.......U	1.1.0(90)
13.1	IN	00 02 80 88	00 19 10 4e	.......N	91.1.0
13.1	IN	00 02 80 9e	00 19 10 0e	....	92.1.0
13.1	IN	00 02 80 b5	00 19 10 1e	....	93.1.0
13.1	IN	00 02 80 ff	00 19 10 e3	....	94.1.0(48)
13.1	IN	00 02 80 ff	00 19 00 bf	....	142.1.0(24)
13.1	IN	00 02 80 ff	08 19 00 be	....	166.1.0(16)
13.1	IN	00 02 80 ff	00 19 00 bf	....	182.1.0(35)
13.1	IN	00 02 80 ff	00 19 02 57	.......W	217.1.0(79)
13.1	IN	00 02 80 ce	00 19 02 56	V	296.1.0
13.0	CTL	21 09 00 03	00 00 08 00	SET REPORT	297.1.0(2)
13.0	OUT	cc 00 00 00	00 00 00 00	....	297.2.0
13.0	CTL	21 09 00 03	00 00 08 00	SET REPORT	299.1.0
13.0	OUT	00 cc 00 00	00 00 00 00	....	299.2.0
13.0	CTL	21 09 00 03	00 00 08 00	SET REPORT	300.1.0(2)
13.0	OUT	00 0a 00 00	00 00 00 00	....	300.2.0
13.1	IN	00 02 80 cc	00 19 02 07	....	302.1.0

Fig. 4. Schematic of collection data text

third column is the brake and throttle Information, according the same method as changes of steering angle information, the throttle data is converted from -128 to 0, and the brake data is converted from 0 to 128.

(2) In the data record, the "Phase" column data indicates the data packet input or output state, "IN" represents the driving operation from driver training machine input information, "OUT, CTL" represents the pc control information output, we only need to extract the information that driver training machine input which the "IN" correspond to data packet from the "data" column.

(3) The time information processing, the "Cmd.Phase.Ofs (rep)" column represents the time series, where figures in brackets indicate the time of the operation remain unchanged. We will restore the driving operation information of each time slice and finally composite the driving operation record of the continuous time slice. The driving operation data of each time slice is a sample, including Attribute 1: steering wheel angle; Attribute 2: Brake throttle Information; and sample label, 1 represents a good driving, -1 represents a bad driving. According to the above preprocessing methods, we get a series of graph about city driving record data. Fig. 5 shows the city good driving record steering wheel data graph.

4.4 SVM Parameters Setting

In this paper, SVM algorithm with RBF kernel function is applied to convert the actually problem into a high dimensional space, where C and $\gamma$are punished coefficient parameters and intervals which are necessary two parameters of RBF. Its value directly will affect the classification accuracy. Here we adopt a cross-validation-based "grid search" method [12][13] to select the values of C and γ, that is selected the training samples divided into v parts, the part of v-1 as a training samples, the remaining as the test sample to determine model parameters, used to verify classification accuracy of the results classified by the v-1 part of the data, and constantly change the $C$and γ to obtain higher sample classification accuracy.

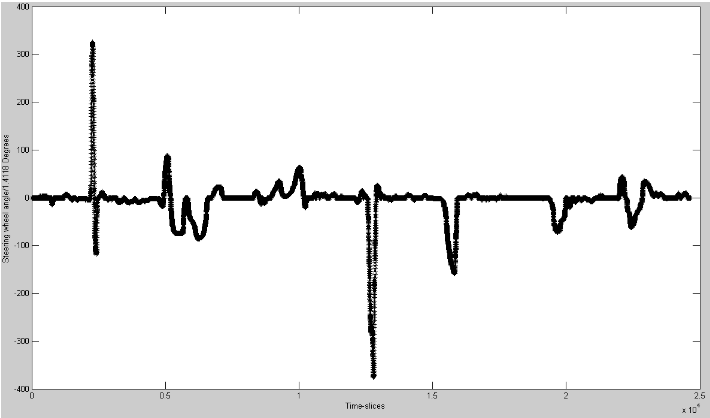

Fig. 5. The city good driving record data of steering wheel graph

Experience has shown, C and γ searched follow the exponential mode of growth ($C = 2^{-10}, 2^{-9}, \cdots, 2^{10}, \gamma = 2^{-10}, 2^{-9}, \cdots, 2^{10}$) is better way to quickly determine the parameters C and γ, where C and γ are independent growth.

4.5 Experimental Results

Use the processed data, select the good and bad driving data under different road as the SVM training and test sets, such as selecting good city driving and bad city driving in training set as training, good city driving in test set as good driving record test, bad city driving in test set as bad driving record test, using the method of parameters set on the section 4.4 described to find out the optimal parameters C and that the most suitable for the city, high-speed and mountain road.

The sample sets are obtained from the section 4.3 Data preprocessing, calculated the rate of change value of steering wheel angle, brake, and throttle to compose the new sample set to create the SVM driving model and test results are as shown in Table 1 SVM-based information fusion (steering wheel, brake throttle, road conditions) driving behavior data analysis, Table 2 SVM-based single information (steering wheel) driving behavior analysis, table 3 SVM-based single information (brake throttle) driving behavior analysis, table 4, 5, 6 are the result of cross-validation, respectively correspondence the SVM-based driving model is established in one of the three roads, and tested by driving data on the other road.

4.6 Experimental Analysis

From the data in Tables 1 and 2, one can clearly see where the testing process in the modeling, SVM comprehensive test accuracy rate can reach 80%, where

Table 1. SVM-based information fusion (steering wheel, brake throttle, road conditions) driving behavior data analysis

	Good driving behavior correct rate	Bad driving behavior correct rate	Comprehensive assessment
City road ($C = 1, \ = 2$)	95.5208%	68.7341%	82.12745%
Highway road ($C = 1, \ = 4$)	97.2902%	77.0102%	87.1502%
Mountain road ($C = 1, \ = 4$)	97.1227%	60.0445%	78.5836%

Table 2. SVM-based single information (steering wheel) driving behavior analysis

	Good driving behavior correct rate	Bad driving behavior correct rate	Comprehensive assessment
City road ($C = 1, \ = 2$)	95.7117%	60.882%	78.29685%
Highway road ($C = 1, \ = 4$)	93.722%	71.2691%	82.49555%
Mountain road ($C = 1, \ = 4$)	97.9687%	41.2564%	69.61255%

Table 3. SVM-based single information (brake throttle) driving behavior analysis

	Good driving behavior correct rate	Bad driving behavior correct rate	Comprehensive assessment
City road ($C = 1, \ = 2$)	99.9881%	2.17372%	51.08091%
Highway road ($C = 1, \ = 4$)	99.9403%	2.7111%	51.3257%
Mountain road ($C = 1, \ = 4$)	99.9739%	5.59072%	52.78231%

Table 4. The result of cross-validation SVM-based driving model is established in mountain

Mountain road driving training model ($C = 1, \ = 4$)	Good driving behavior correct rate	bad driving behavior correct rate	Comprehensive assessment
City road driving test data	95.4695%	62.3225%	78.896%
Highway road driving test data	97.8128%	62.7329%	80.27285%

Table 5. The result of cross-validation SVM-based driving model is established in highway

Highway driving training model (C = 1, = 4)	Good driving behavior correct rate	bad driving behavior correct rate	Comprehensive assessment
City road driving test data	88.764%	70.4949%	79.62945%
Mountain road driving test data	92.5274%	54.7351%	73.63125%

Table 6. The result of cross-validation SVM-based driving model is established in city

City road train driver model (C = 1, = 2)	Good driving behavior correct rate	bad driving behavior correct rate	Comprehensive assessment
Highway driving test data	98.8507%	62.1706%	80.51065%
Mountain road driving test data	97.9687%	45.4524%	71.71055%

the good driving behavior recognition rate is relatively higher, while the identification of bad driving behavior is lower, that mainly caused by the following points:

(1) Since in this test we used the test and training data set which composed with multiple time-slice sample set include the steering wheel angle gradient and brake throttle gradient to establish the driving model in different road. The driving behavior analysis system will judge the each of the time slice sample in real time. While there are so many good driving behavior sample in the bad driving data set, such as Uniform motion in a straight, slow start and so on. Thats the reason why the identification of bad driving behavior is lower.

(2) we can clearly see from Tables 1 and 2, the behavior analysis based on multi-information fusion driving is better than based on the single information in judgment the bad driving behavior. It is indicate that the single drive operation is difficult to correctly reflect the current driver's driving behavior, in real-time driving behavior analysis, the driving behavior analysis system need to combine multiple driving operation information to make driving behavior analysis to get more accurate results.

(3) Tables 4, 5, and 6 show that there are some major differences between the driving models which established in the different road based on SVM, resulting in lower correct rates in the driving behavior judgment. It is indicate that we need to create different driving behavior model to determine the correct analysis according to different road conditions.

5 Conclusion

With the real-time driving behavior analysis becomes more and more important, this paper utilizes a number of important driving operation data (brakes, throttle, steering wheel angle and road conditions) to comprehensive analysis the driving behavior. Driving behavior analysis model based on SVM can effectively achieve the correct judgment on driving behavior analysis; timely corrective driver's driving improperly. In future work We need to further study on combination with other data analysis methods to achieve more accurate and rapid analysis of driving behavior. Such as the method of Discrete Hidden Markov Model (DHMM) applied to driver behavior analysis [14] etc.

References

1. Leonard, E.: Traffic safety. Science Serving Society (2004)
2. Yan, X., Zhang, H., Wu, C., Jie, M., Hu, L.: Research progress and prospect of road traffic driving behavior. Traffic Information and Safety (1), 45–51 (2013)
3. Miyaji, M., Danno, M., Oguri, K.: Analysis of driver behavior based on traffic incidents for driver monitor systems. In: IEEE Intelligent Vehicles Symposium, pp. 930–935 (2008)
4. Liang, J., Cheng, X., Chen, X.: The research of car rear-end warning model based on mas and behavior. In: IEEE Power Electronics and Intelligent Transportation System, pp. 305–309 (2008)
5. Yoshifumi, K., Koji, O.: A modeling method for predicting driving behavior concerning with driver's past movements. In: IEEE International Conference on Vehicular Electronics and Safety, pp. 132–136. IEEE (2008)
6. Prez, A., Garcia, M.I., Nieto, M., Pedraza, J.L., Rodrguez, S., Zamorano, J.: Argos: an advanced in-vehicle data recorder on a massively sensorized vehicle for car driver behavior experimentation. IEEE Transactions on Intelligent Transportation Systems 11(2), 463–473 (2010)
7. Osgouei, R.H., Choi, S.: Evaluation of driving skills using an hmm-based distance measure. In: IEEE Haptic Audio Visual Environments and Games, pp. 50–55 (2012)
8. Qin, H., Liu, Z., Wang, P.: Research on drowsy driving behavior based on multichannel information fusion. China Safety Science Journal 2, 020 (2011)
9. Huang, D.Y., Chen, C.H., Hu, W.C., Yi, S.C., Lin, Y.F.: Feature-based vehicle flow analysis and measurement for a real-time traffic surveillance system. Journal of Information Hiding and Multimedia Signal Processing 3(3), 279–294 (2012)
10. Nello, C., John, S.: An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press (2000)
11. Vapnik, V.: Statistical learning theory (1998)
12. Sathiya, K., ChihJen, L.: Asymptotic behaviors of support vector machines with gaussian kernel. Neural Computation 15(7), 1667–1689 (2003)
13. Joachims, Thorsten: Making large scale svm learning practical (1999)
14. Pan, S.T., Hong, T.P.: Robust speech recognition by dhmm with a codebook trained by genetic algorithm. Journal of Information Hiding and Multimedia Signal Processing 3(4), 306–319 (2012)