Speeding Behavior Estimation Using Stop-and-Go Events Based on Velocity Data

Mingxi Chen[1], Zhen Gao[1*], Rongjie Yu[2*], Xuesong Wang[2]

1. School of Software Engineering, Tongji University, Shanghai 201804, China
2. Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai 201804, China

Abstract*— Aggressive driving style has been proved associating with road accidents and nearly a third of fatal crashes in the United States are designated as "speeding-related". We develop a simple approach, by introducing stop-and-go events with speeding behavior estimation. In this paper, drivers are divided into "speeding" and "non-speeding" group. Speed and acceleration data for 5 seconds before stop and 5 seconds after move are chosen to analyze the driving behaviors. We use Mean Decrease Accuracy (MDA) and Mean Decrease Gini (MDG) of Random Forests classifier to select important features of stop-and-go events to discriminate speeding drivers. Experiment results show that drivers in "speeding" group are inclined to use sudden braking before stop and rapid acceleration after moving the car. Mann-Whitney-Wilcoxon Test results further proved the stop-and-go behavior difference between speeding and non-speeding groups. The Model based on Random Forests algorithm had a good performance to classify speeding drivers with AUC 0.94. The potential application of this study is developing a quantitative and real-time method to estimate the security of a driver's speed control and provide feedback to surrounding drivers in future intelligent connected vehicles.

***Keywords*— traffic security, speeding behavior, stop-and-go event, naturalistic driving, Random Forests**

I. INTRODUCTION

With the prevalence of automobiles, driving becomes an important part of daily life. However, it also brings about increased threats to people's life and property safety. The share of social costs of road crashes in high income countries ranges from 1.0% to 4.6% of the GDP with an average of 2.6% [1]. Significant researches have been performed to establish linkage between aggressive driving style and crashes because it has been pointed out as one of the main predictors of road accidents [2],[3]. "Aggressive" drivers are also being found to be more consistent in behaviors and significantly more predictable than "non-aggressive" drivers [4]. Therefore, developing methods to achieve real-time estimation of aggressive driving style and provide warnings to driver herself/himself and surrounding drivers can be meaningful.

There are several different definitions of "aggressive" and "non-aggressive". Hong etc. took questionnaires, self-reports of accidents and speed violations to define aggressive style [5], Doshi etc. took the threshold based on tested drivers' standard deviation of lateral jerk [4]. However, the limitations of these definitions are obvious. Behavioral questionnaires yield responses that are loosely correlated with real world poor driving outcomes [6]. Moreover, it's hard to ensure the rationality of threshold which are defined based on insufficient drivers.

A feasible solution to above problems is finding a more objective and universal way to discriminate drivers. Among many factors related to crashes, speeding is one which cannot be ignored. Nearly a third of fatal crashes in the United States are designated as "speeding-related", which is defined as "either the driver behaviors of exceeding the posted speed limit or driving too fast for conditions" [7]. Therefore, in this paper, we used speeding to assess aggressive drivers which is confirmed by clear legal provision. While speeding behavior cannot be recognized in real-time based on only velocity data without road type information. Therefore, a surrogate index is needed.

In related researches, many events or features are used to assess and estimate aggressive driving style. Significantly higher frequency of large positive and negative jerk is proved to associate with aggressive drivers [8]. Sensor data from a single turn can identify drivers with mean accuracy of 76.9% [9]. Aggressive drivers and violators have higher G-Forces [5]. Excessive speed, hard braking, heavy acceleration and aggressive turns have also been used to build metrics for detecting aggressive driving behaviors [10]. High-risk drivers are more likely to be involved in approaching, near following, and constrained left and right lane changes [10]. However, more different sensor data means more cost for estimation and more acquisition accuracy issues need to be considered.

In this paper, we propose a method with relatively easier realization, lower cost and higher possibility of prevalence to estimate speeding driving style. For simplicity, stop-and-go events are extracted to analyze the driving style in our method. The stop-and-go events can be easily identified with only velocity which could be directly got using vehicle Controller Area Network (CAN). These original data are easily to obtain and have low requirement to acquisition equipment. Then we adopted Random Forests model to find important features which could be used to discriminate speeding and non-speeding drivers, box-plots and Mann-Whitney-Wilcoxon Test give further intuitive and quantitative display for each feature's diverse distribution. We find that these selected stop-and-go features are good candidates for estimating a driver's speeding aggressive style. Lastly, the potential application and limitation of this study

* Co-corresponding authors Email:
gaozhen@tongji.edu.cn; yurongjie@tongji.edu.cn

are discussed.

II. Method

A. Data Source

In this study, data are provided by the Shanghai Naturalistic Driving Study (SH-NDS) jointly conducted by Tongji University, General Motors and Virginia Tech Transportation Institute. Five cars involved in this project are equipped with SHRP2 NextGen Data Acquisition Systems which included Controller Area Network (CAN), a radar system, an accelerometer for longitudinal and lateral acceleration, a light meter, a GPS sensor for location, a temperature/humidity sensor, and four synchronized camera views facing inside and outside the vehicle. But only velocity data are used to detect stop-and-go events in our research considering vehicle velocity data have already been widely collected in both industry and academia.

12 drivers are randomly selected by this study, the driver's age ranged from 28 to 61 (mean = 41.2) with an average of 7.3 years of driving experience. Each driver had, on average, 15 trip records and more than 3 hours of driving data.

The purpose of this study is to estimate speeding driving style, since driving data collected on urban expressway are relatively adequate for research and urban expressway is exactly the high-occurrence place of speeding behaviors, following experiments only based on this road type.

B. Features Selection

The speed and acceleration of vehicle are used in this paper. The former is provided by the sensor data which had a sampling rate of 10 Hz. The latter is computed by velocity data using (1) instead of being got from vehicle gyroscopes directly because gyroscopes data have much noise. Fig. 1 shows the obvious diversity in the acquisition accuracy of different vehicle gyroscopes. The black line and the red line represent accelerations of two stop-and-go events collected on different vehicles. When vehicles had come to a stop, their accelerations not only didn't equal 0 m/s^2, but also kept in different levels.

$$\mathrm{a(t)} = \frac{v(t+\Delta t)\text{-}v(t)}{\Delta t} \qquad (1)$$

where a(t) is the derivative of v at time t and Δt is set to 0.1 s.

One situation is detected as a stop-and-go event when the vehicle speed equals 0 m/s for more than 5 seconds. Speed and acceleration for 5 seconds before the vehicle comes to a stop and 5 seconds after the car is moved are chosen to analyze the driving behaviors. In view of traffic congestion may bring about severe behavioral intervention, we only select events whose average speed before stop and after go both exceeded a pre-defined threshold. The threshold we hypothesized is 40 km/h. After processing, a total of 2425 stop-and-go events are detected by using above method. In order to further ensure drivers are not influenced by external factors, we extract 239 out of those 2425 events which have no other vehicles in front of the target vehicle.

For a better characterization of stop-and-go events, 8 different statistical measures including Mean, Absolute Mean, Standard Deviation, Coefficient of Variance, Kurtosis, Skew, Maximum, Minimum are applied to speed and acceleration respectively. These features are used in the further analysis of the relationship between speeding driving style and stop-and-go events.

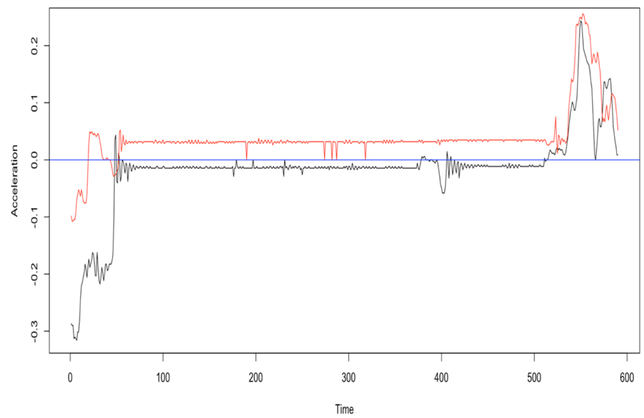


Fig. 1. The diversity of different vehicles' acceleration

C. Speeding Recognition

First, 12 drivers are divided into two groups, "speeding" and "non-speeding" group. To quantity drivers' speeding behaviors, we referred to [8] and defined two speeding rates as (2) and (3) as below.

$$\mathrm{i\_s1} = \frac{i\_x2\_time}{i\_x1\_time} \times 100\% \qquad (2)$$

$$\mathrm{i\_s2} = \frac{i\_x3\_time}{i\_x1\_time} \times 100\% \qquad (3)$$

where i is road type, i has a value of "u" to indicate urban expressway; i_s1 is the speeding rate of x2 on road type i; i_s2 is the speeding rate of x3 on road type i; "i_xj_time" is the duration with speed over "xj" on road type i, where j can be 1, 2 and 3. The different settings of xj are separately based on speed limit and different speeding levels with smooth traffic on specific road type. See Table I for a summary of definitions.

TABLE I. Definitions of X1, X2 and X3

Road type	Urban expressway (i = u)
x1	70
x2	80
x3	90

We hypothesize that if a driver can speed up to 70 km/h then the traffic is smooth enough. Fig. 2 shows 12 drivers' speeding rates of 80 km/h and 90 km/h on urban expressway.

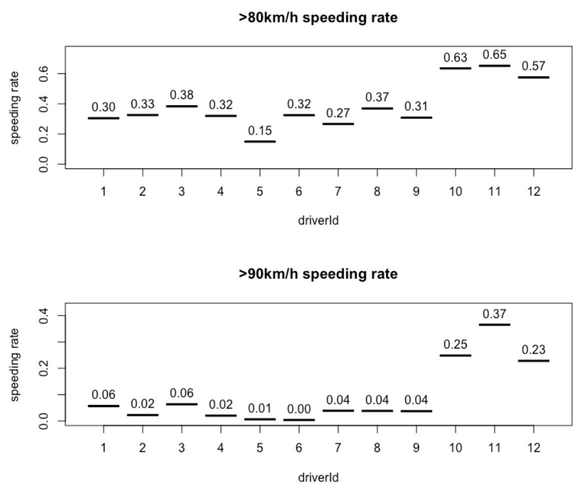


Fig. 2. The speeding rates of 80 km/h and 90 km/h on urban expressway

Drivers whose speeding rates of 80km/h and 90km/h are both significantly higher than others are put into "speeding" group. Others are put into "non-speeding" group.

After comprehensive consideration, driver 10, driver 11 and driver 12 are put into "speeding" group for their significantly longer duration of high speeding rates in Fig. 2. Other drivers are put into "non-speeding" group. And 14.2% of stop-and-go events are taken by drivers in the "speeding" group and labeled by "speeding" accordingly.

For further analysis, we build an input metric which consisted of stop-and-go behavior features including speed and acceleration statistics separately in "before-stop" and "after-go" phases. 80% records of this metric are randomly selected as training dataset and the rest are used as test dataset.

III. EXPERIMENT AND RESULT

The estimation performance of using input metric built above is examined by Random Forests classifier. Three classifiers are built separately using stop based features, go based features and all features. The Area Under the ROC Curve (AUC) is used as evaluation index. The AUC of classifier with all features is 0.94, the AUC of classifier with stop based features and go based features are 0.91 and 0.90, which prove that our proposed method has good practicability.

We use Mean Decrease Accuracy (MDA) and Mean Decrease Gini (MDG) to estimate the importance of features. The larger MDA or MDG is, the greater the feature's importance is. Fig. 3 show the ordering of MDA and MDG respectively.

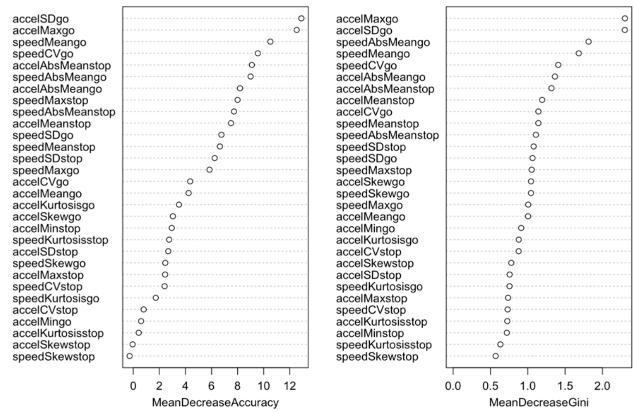


Fig. 3. The ordering of MDA and MDG

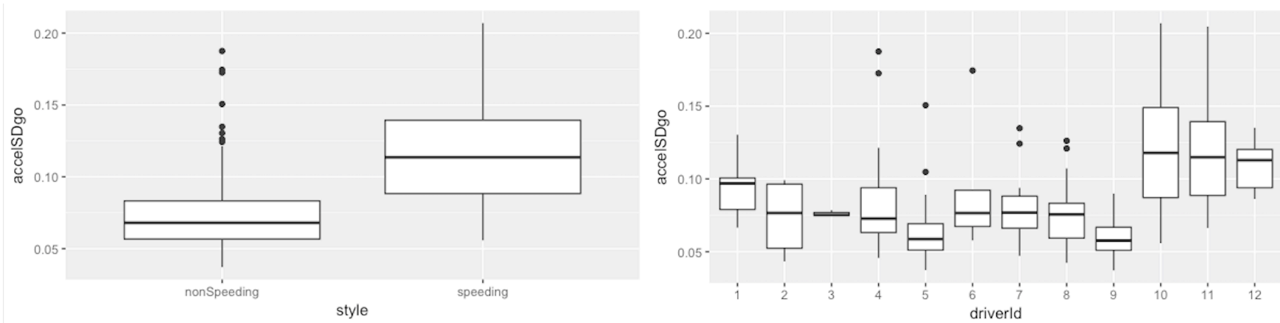
After comprehensive consideration, 3 go based features (accelSDGo, accelMaxGo, speedMeanGo) and 3 stop based features (accelAbsMeanStop, speedMaxStop, speedMeanStop) are selected as important features for further analysis, the descriptions are shown in Table II.

Figure 4 shows the distribution differences of same feature between two different groups in which driver 10, 11 and 12 are labeled "speeding" and others "non-speeding".

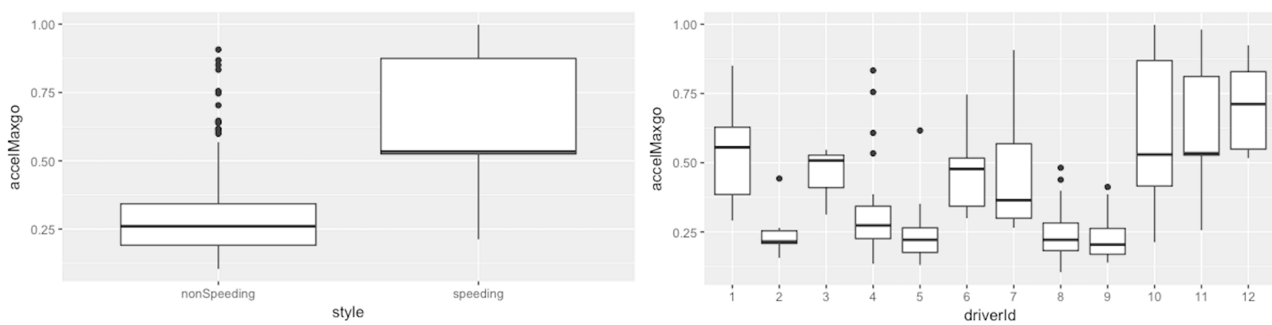
Shapiro-Wilk test shows that most features in each group do not follow a Gaussian distribution ($p<0.05$). Therefore, we use Mann-Whitney-Wilcoxon Test to examine the distribution similarity of features. The test results (Table III) show that these 6 important features' distributions are significantly different ($p<0.05$). Drivers in the speeding group are associated with greater standard deviation of acceleration, maximum of acceleration and greater average speed for 5 seconds after go, also greater absolute mean of speed, greater maximum of and greater mean of speed for 5 seconds before stop.

TABLE II. THE DESCRIPTION OF EACH FEATURE

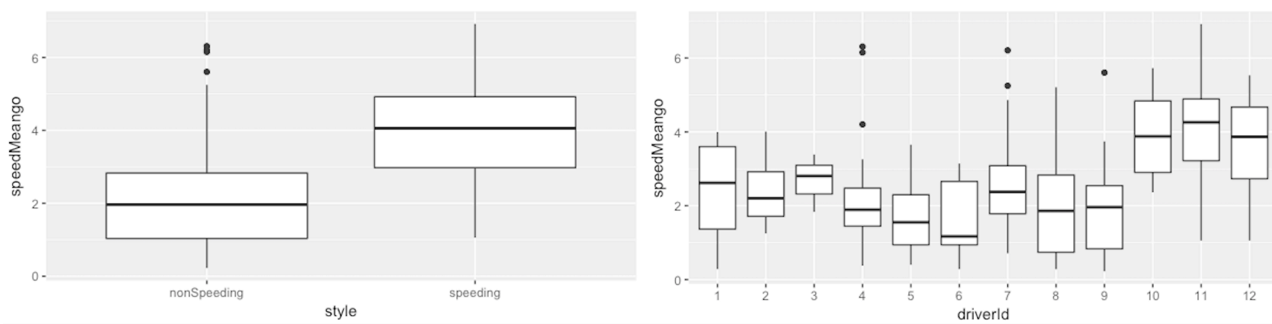
Feature name	Description	Summary			
		Min	Max	Mean	SD
accelSDGo	The standard deviation of acceleration after go	0.037	0.207	0.078	0.031
accelMaxGo	The maximum of acceleration after go	0.104	0.998	0.345	0.202
speedMeanGo	The mean of vehicle speed after go	0.224	6.923	2.357	1.472
accelAbsMeanStop	The absolute mean of acceleration before stop	0.028	0.292	0.099	0.053
speedMaxStop	The maximum of speed before stop	0.368	13.633	4.566	2.903
speedMeanStop	The mean of speed before stop	0.203	6.968	2.446	1.458



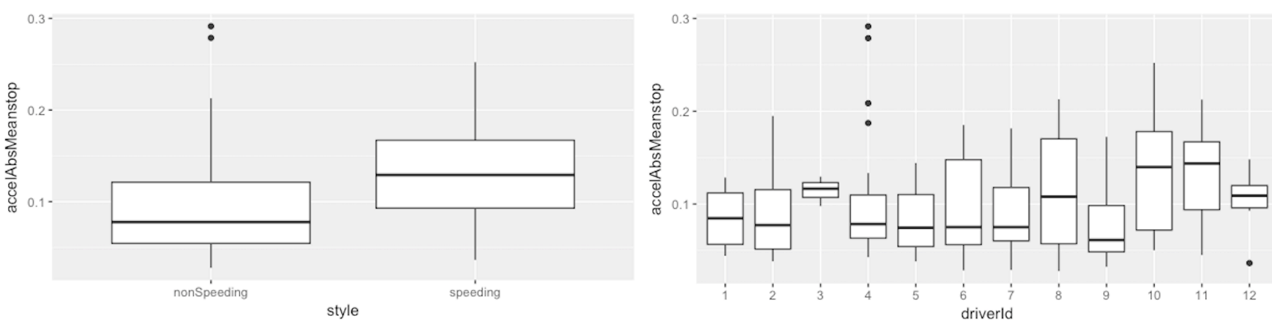
(a) The box-plot of accelSDGo for each group and each driver



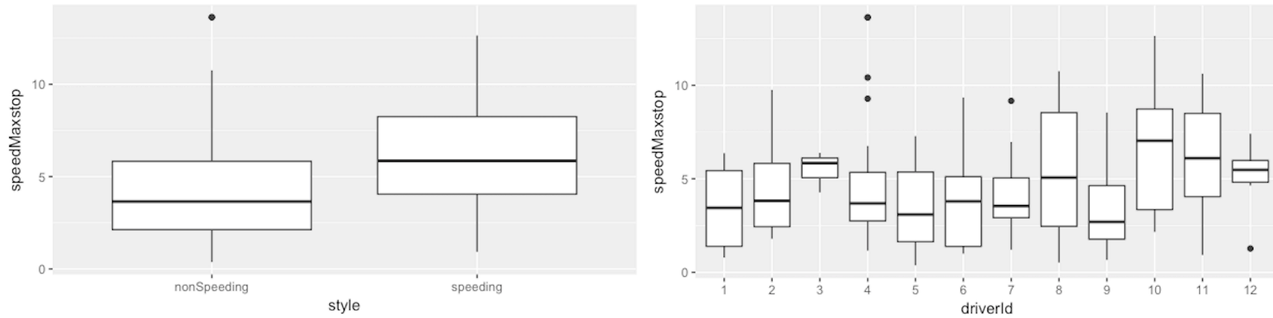
(b) The box-plot of accelMaxGo for each group and each driver



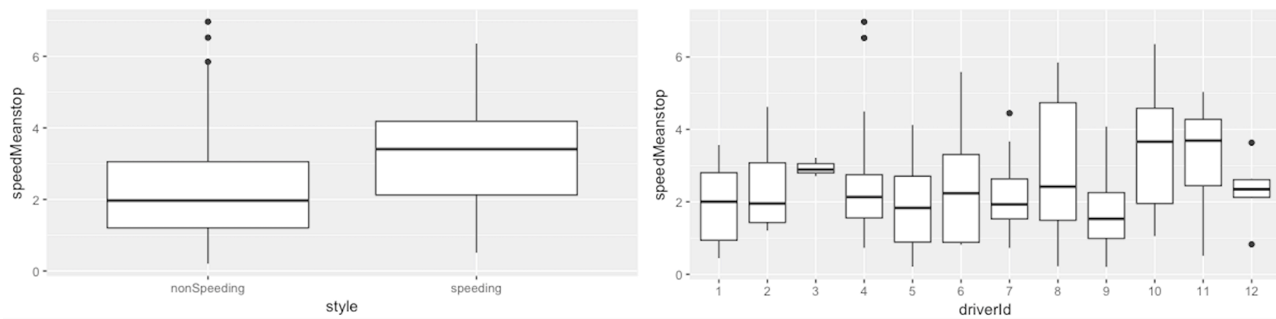
(c) The box-plot of speedMeanGo for each group and each driver



(d) The box-plot of accelAbsMeanStop for each group and each driver



(e) The box-plot of speedMaxStop for each group and each driver



(f) The box-plot of speedMeanStop for each group and each driver

Fig. 4 Box-plot of important features

TABLE III. RESULTS OF MANN-WHITNEY-WILCOXON TEST WITH VARYING FEATURES IN THE ANALYSIS

Feature name	p value of Mann-Whitney-Wilcoxon Test
accelSDGo	3.903e-12
accelMaxGo	1.546e-13
speedMeanGo	2.88e-10
accelAbsMeanStop	0.0005651
speedMaxStop	0.0005132
speedMeanStop	0.0005238

IV. DISCUSSION

The purpose of this paper is to investigate whether drivers' speeding behaviors can be estimated from stop-and-go events. The results illustrate that accelSDGo, accelMaxGo, speedMeanGo, accelAbsMeanStop, speedMaxStop and speedMeanStop are important features to discriminate drivers into "speeding" and "non-speeding" group. Then the box-plot of each important feature shows the obvious diversity of feature distribution in an intuitive way. Mann-Whitney-Wilcoxon Test results further prove that there are significant differences between the driving behaviors of the two groups, which strengthen the evidence linking stop-and-go behaviors with speeding driving style. By comparison, the differences of go based features between "speeding" drivers and "non-speeding" drivers are more obvious. We can reflect these findings into actual operations during drivers' driving period. The results of accelSDGo, accelMaxGo and speedMeanGo (Fig. 4 a-c) show that drivers in speeding group are inclined to use higher and more discrete acceleration and speed after they begin moving the car; The results of accelAbsMeanstop, speedMaxStop and speedMeanStop (Fig. 4 d-f) show that speeding drivers prefer to use shorter time to decelerate before stop with higher acceleration and speed. These findings are consistent with the conclusions of existing studies [5],[11].

The potential application of this study is developing a quantitative and real-time way which can use only several stop-and-go events to estimate the security of a driver's speed control and provide feedback to the driver and warnings to surrounding drivers. All data used in this paper are velocity data which are easily to obtain. There are still several limitations in the current study. First, only the driving data on urban expressway are used. It is meaningful to investigate other road types' driving data such as highway and city road. Secondly, only features based on speed and acceleration are used to build connection between stop-and-go events with speeding behaviors, data of gas pedal, brake pedal, jerk or other features can also be used to do more comprehensive analysis. Lastly, only 12 drivers involved in this study, it is necessary to collect more drivers' data for analysis.

V. CONCLUSION

The main purpose of this paper is to examine whether stop-and-go events can be used to estimate drivers' speeding behaviors. We first extract driving data on urban expressway and divide drivers into "speeding" and "non-speeding" group by their behavior of using excessive speed. Then we detect stop-and-go events and select only "no car before" events in order to exclude the possibility of influence caused by traffic congestion. 8 statistical methods are used to characterize stop-and-go events and composed input metrics. The AUC of Random Forests classifier based on all features is 0.94, the AUC of classifier with stop based features and go based features are 0.91 and 0.90 separately. Mean Decrease Accuracy and Mean Decrease Gini are used to evaluate the importance of each feature. After comprehensive consideration, we choose 3 stop based features and 3 go based features. The box plots give intuitive display of the diversity distribution range of each feature between speeding and non-speeding group. Then the Mann-Whitney-Wilcoxon Test results further strengthen the evidence that accelSDGo, accelMaxGo, speedMeanGo, accelAbsMeanStop, speedMaxStop and speedMeanStop have good performance in estimating a driver's speeding behaviors. The results indicate that drivers with speeding behaviors prefer to use sudden braking before the vehicle comes to a stop and rapid acceleration after they move the car. The potential application of this study is developing a quantitative and real-time way which can use only several stop-and-go events to estimate the security of a driver's speed control and provide feedback to the driver and warnings to surrounding drivers in intelligent connected vehicles.

ACKNOWLEDGMENT

This study is jointly sponsored by the Tongji Education Reform Project, Chinese National Natural Science Foundation (51522810), the Science and Technology Commission of Shanghai Municipality, China (18DZ1200200), the Chinese National Engineering Laboratory for integrated Optimization of Road Traffic and Safety Analysis Technologies, and the Chinese 111 Project (B17032).

REFERENCES

[1] Wijnen, Wim, and H. Stipdonk, "Social costs of road crashes: An international analysis," *Accident Anal. & Prevention*, vol. 94, pp. 97-106, 2016.

[2] Bogdan-Ganea, Smaranda Raluca, and D. Herrero-Fernández, "Aggressive thinking on the road. The mediation effect of aggressive thinking in the relationship between driving anger and aggression in Romanian drivers," *Transportation Research Part F Traffic Psychology & Behaviour*, vol. 55(55), pp. 153-166, 2018.

[3] Dahlen, E. R., Edwards, B. D., Tubré, T., Zyphur, M. J., & Warren, C. R, "Taking a look behind the wheel: an investigation into the personality predictors of aggressive driving," *Accident Analysis & Prevention*, vol. 45(1), pp. 1-9, 2012.

[4] A. Doshi and M. Trivedi, "Examining the impact of driving style on the predictability and responsiveness of the driver: Real-world and simulator analysis," in *Proc. IEEE Intelligent Vehicles Symposium (IV)*, pp. 232-237, 2010.

[5] Jin-Hyuk Hong, Ben Margines, and Anind K. Dey, "A smartphone-based sensing platform to model aggressive driving behaviors," in *Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI '14)*, pp. 4047-4056, 2014.

[6] Af Wåhlberg, Anders E, L. Dorn, and J. Freeman. "Commentary on the rebuttal by de Winter and Dodou concerning "The Driver Behaviour Questionnaire as a predictor of accidents: A meta-analysis"," *Physica Scripta*, vol. 42(6), pp. 749-760, 2012.

[7] Fitzpatrick, Cole D., S. Rakasi, and M. A. K. Jr, "An investigation of the speeding-related crash designation through crash narrative reviews sampled via logistic regression," *Accid Anal Prev*, vol. 98, pp. 57-63, 2017.

[8] Feng, Fred, et al, "Can vehicle longitudinal jerk be used to identify aggressive drivers? An examination using naturalistic driving data," *Accid Anal Prev*, vol. 104, pp. 125-136, 2017.

[9] Hallac, David, et al, "Driver identification using automobile sensor data from a single turn," in *Proc. IEEE/Int. Conf. Intell. Transportation Syst.*, pp. 953-958, 2016.

[10] Z. Chen, J. Yu, Y. Zhu, Y. Chen, and M. Li, "D3: Abnormal driving behaviors detection and identification using smartphone sensors," in *Proc. IEEE/Int. Conf. Sensing, Commun., and NETWORKING*, pp. 524-532, 2015.

[11] S. Jafarnejad, G. Castignani, T. Engel, "Towards a real-time driver identification mechanism based on driving sensing data", *20th International Conference on Intelligent Transportation Systems (ITSC)*, 2017.