Application of Random Forest to develop a driving stress model in urban road network

Weiliang Zeng[1], Tomio Miwa[2], Mutsumi Tashiro[3] and Takayuki Morikawa[4]

[1] Postdoctoral Research Fellow, Institution of Materials and Systems for Sustainability, Nagoya University
(Furo-cho, Chikusa, Nagoya, Japan 464-8603)
E-mail: zengweiliang@imass.nagoya-u.ac.jp
[2]Member of JSCE, Associate Professor, Institution of Materials and Systems for Sustainability, Nagoya University
(Furo-cho, Chikusa, Nagoya, Japan 464-8603)
E-mail: miwa@nagoya-u.jp
[3]Member of JSCE, Lecturer, Institute of Innovation for Future Society, Nagoya University
(Furo-cho, Chikusa, Nagoya, Japan 464-8603)
E-mail: mutsumi@civil.nagoya-u.ac.jp
[4]Member of JSCE, Professor, Institute of Innovation for Future Society, Nagoya University
(Furo-cho, Chikusa, Nagoya, Japan 464-8603)
E-mail: morikawa@nagoya-u.jp

Driving performance deteriorates with excess driving stress rises, which may increase vehicle accident likelihood. This study aims to quantify the effect of driving stress by monitoring the heart rate increase with various traffic conditions in a real-world road network. The data collection includes heart rate from electrocardiogram records, vehicle trajectories from GPS, road conditions from video, and vehicle conditions from CAN bus. We propose a machine learning methodology based on Random Forest for the estimation of car driver stress due to different driving events. In contrast to other statistical methods and machine learning methods, Random Forest can handle different types of predictor variables, make a high accurate prediction and give variable importance analysis. Results indicate that average speed, coefficient of covariance of speed, times of brake operation and times of acceleration operation contribute about 78% relative importance to driving stress. Further sensitivity analysis show that low average speed, large speed variance, frequent operations of brake and acceleration will cause high level of driving stress. The driving stress heat map can be applied to a safety-based route guidance system.

Key Words : *driving stress, heart rate, Random Forest, variable importance, probe vehicle*

1. INTRODUCTION

In our highly motorized society, one of the social problems is the considerable numbers of injuries and deaths caused by traffic accidents. Clearly, we will benefit greatly with the improvements in advanced driver assistance system considering driving safety. Many research found that driving stress can have a significant impact on driving safety[1,2]. Drivers with a history of crashes reported significantly higher stress levels while performing common driving tasks[1]. Although moderate level of stress may be beneficial in maintaining driver attention[3], high stress can influence adversely drivers' reactions in critical situations and increase crash likelihood, thus it is one of the most important factors for vehicle accidents among fatigue and aggressive driving[4-6].

As the driving safety becomes a rising concern, many scientists have devoted themselves to finding the connection between driving safety and driving stress. There has been great interest in the use of physiological measures for driving performance monitoring in a new generation of advanced driver assistance systems[7,8]. A growing body of evidence[9-12] has shown that physiological signals can be used to detect continuous changes in heart rate, respiration, skin conductance, and muscle activity, thus they hold the potential to measure the change of driving stress in changing environment. For example, Yamakoshi et al.[13] conducted an experiment on driver's awareness level in monotonous situations and the cardiovascular parameters are used for analysis. It demonstrated that sympathetic activity was increased during monotonous situation whilst vagal tone appeared to be suppressed. Healey and Picard[7] analyzed four types of physiological data (electrocardiogram, electromyogram, skin conductivity, and respiration) during driving tasks to determine a driver's relative stress level. They found that an overall of 97.4% can be achieved by using 5-min intervals data provided by heart rate and skin conductivity. Although different kinds of sensors enable to provide detail physiological signals in different aspects, it is impossible to wear different kinds of sensors simultaneously. For example, wearing many sensors may cause discomfort and dangers to drivers[14]. Among various physiological signals, many studies have validated the effectiveness of heart rate variability (HRV) measures extracted from electrocardiogram (ECG) can be used as an indicator for driver stress[7,14,15]. Thus, it is a worthwhile research topic to develop an efficient driving stress recognition approach by using HRV.

Driving stress is usually influenced by internal emotions and external factors. Stress from driving can be classified into two types: trait and state stress[16-20]. Trait stress is associated with those stable vulnerability factors residing within an individual, such as frustration, impatience, anger, and fatigue proneness. State stress is related to specific external situations that are considered challenging to control such as driving in a heavy traffic, driving on a slippery road, limited visibility caused by rain, limited travel time budget, conflicting with pedestrians and cyclists, being put in danger by impatient, ignorant or aggressive drivers, and being caught behind a slow moving vehicle with time urgency. Hill and Boyle found that driving stress increased as the interactions with other drivers increased[3], and the adverse weather conditions (e.g., icy road, heavy rain) and poor visibility conditions (e.g., nighttime driving) may also contribute high levels of stress. In a driving simulation study, it is found that heart rate increased with heightened task demands such as entering a roundabout, and dropped as task demands decreased, for instance, driving on a two-lane highway[21]. Similarly, changes in driving speed and brake operation have been observed when drivers appear to be more stressful[22]. Miller and Boyle found that dimensional differences in road geometry may impact driving stress and vehicle control[15]. A real-world driving test shown that participants' heart rate was highest in the transition segments before tunnel segments. And they also found that stress was higher within a tunnel compared with the open road.

All of the above findings indicate that driving stress have significant relationship with traffic conditions and surrounding environment. However, few studies quantify their correlation and importance. That is, it has not yet determined which variables have the most important contribution to the driving stress and how many variables are needed for driving

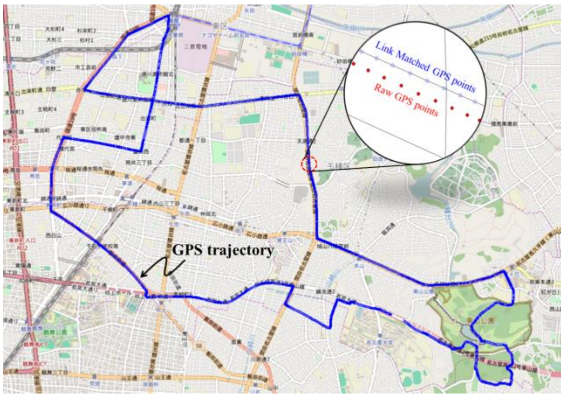
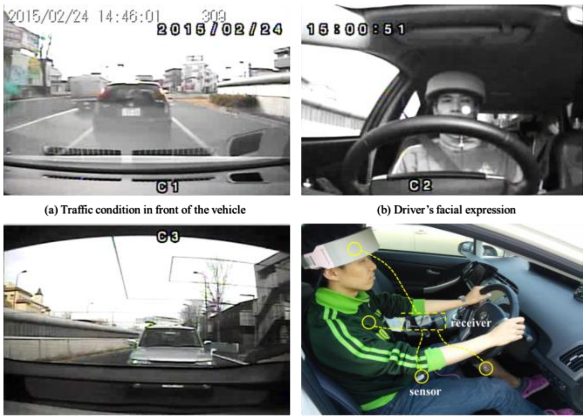


Figure 1 Experimental route in Nagoya city, Japan



(a) Traffic condition in front of the vehicle
(b) Driver's facial expression
(c) Traffic condition in the back of the vehicle
(d) Physiological signal recording during driving

Figure 2 Video data collection

stress estimation. To fill this gap, we aim to quantify the effect of driving stress by monitoring the heart rate increase with various traffic conditions in a real-world road network. Specifically, we propose a machine learning methodology based on Random Forest for the estimation of car driver stress, which enables to handle different types of predictor variables and provide interpretable results. Finally, the relative importance of each variables and the application to network-wide driving stress mapping are given.

2. METHODOLOGY

(1) Data collection

a) Route and GPS

As shown in Figure 1, the experimental route consists of arterial roads, residential streets, and mountain roads. It is about 22.6km and takes about 70 minutes to complete. GPS data contains a vehicle's latitude and longitude position, speed and direction with a fixed sampling frequency of 1s. There are 3 students participating in this experiment. Each student drives along the route 5 times in different dates. Totally, 374,880 GPS records with different traffic conditions in 15 driving experiments are obtained. A map-matching algorithm[23] is used to map the raw GPS records to a digital map, and then the link-based speed and coefficient of variance of speed can be obtained.

b) CAN bus

CAN bus is a control network for the vehicle electronic equipment. It was designed as a communication medium of control units in vehicles. A CAN bus connects actuators with sensors enabling to detect the health of vehicle from various vehicle-related information. The driving parameters such as brake operation, acceleration operation, steering wheel reversal rat and yaw velocity can be extracted second-by-second. In this study, an on-board diagnostics (OBD) device was installed in the experimental car (Toyota Prius) for CAN data collection.

c) Video

As shown in Figure 2, one video camera was installed on the dashboard to capture the front scenes of the road condition along the route. One video camera was mounted on the backrest of the back seat to capture the road condition in the back of the video. And another video camera was installed on the windshield to monitor the driver's behavior. The road conditions and driver's behavior such as congestion, conflicting with pedestrians, signalization, and lane changing can be obtained after post processing.

d) Physiological signal

As shown in Figure 3(a), we use the multi-channel telemetry system (Web-7000) developed by NIHON KONDEN Co.Ltd., Japan, for physiological signal acquisition. In this study, we acquired only the R-wave from ECG for driving stress analysis. To ensure the accuracy of R-wave occurrence time estimates, the ECG signal is acquired using 1000 Hz sampling rate.

e) Data processing

We derive the heart rate from the R-wave of ECG. The R-wave time instants can be detected by using the QRS complex (a combination of Q, R, and S waves in ECG) developed by Pan-Tompkins algorithm[24]. A typical QRS complex detection algorithm consists of a preprocessing part followed by a decision rule[26]. The preprocessing includes bandpass filtering of the ECG to reduce power line noise and baseline wander, squaring of the data samples and moving average filtering to smooth close-by peaks. The decision rules include amplitude threshold and expected time between adjacent R-waves. As shown in Figure 3(b), R is a point corresponding to the peak

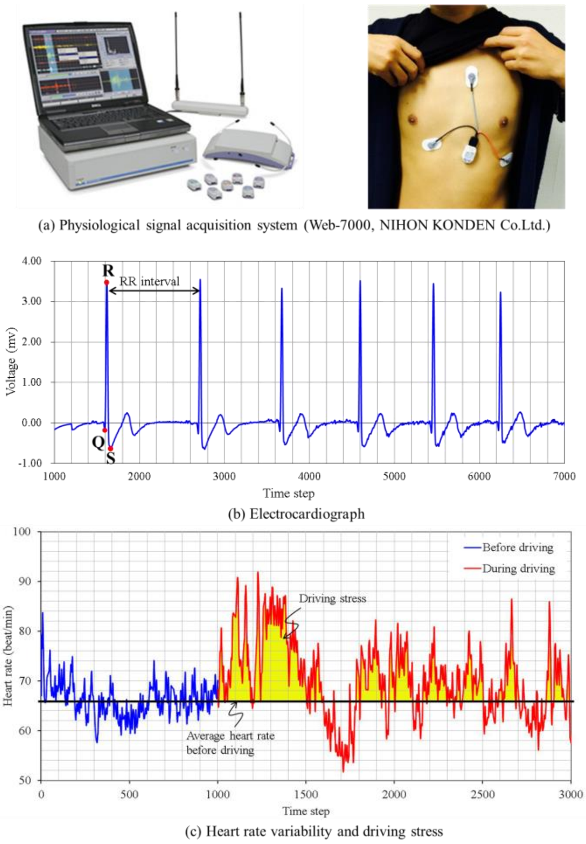


Figure 3 Physiological signal detection

of the QRS complex of the ECG wave. We use the RR interval (RRI) to represent the heart beat interval. The heart rate can be presented by 60,000/RRI with the unit of beat/min as shown in Figure 3(c). To synchronize all the data, the original RRI time series needs to be re-sampled and interpolated because RRI series is an irregularly time-sampled signal. The RRI series is converted into equidistantly sampled form by using cubic spline interpolation prior to HR analysis[25]. Finally, data from GPS, CAN, ECG and video are synchronized second-by-second.

As shown in Figure 3(c), the driving stress is defined as the heart rate increase comparing to the average heart rate before driving. Specifically, the instantaneous driving stress can be expressed as follows:

$$DS = \max(0, HR_{drive} - HR_{calm}) \quad (1)$$

where DS denotes the instantaneous driving stress driving stress, HR_{drive} denotes the instantaneous heart rate during driving, and HR_{calm} denotes the mean of heart rate in a calm condition before driving.

Because driving stress is influenced by both internal and external factors, it is difficult to predict it in a short-time interval. There may be an increase in driving stress due to anticipatory, monitoring, and planning effects before a stressor is observed. For example, drivers may feel nervous before entering a congestion area though he/she was still driving on a free-flow roadway. In addition, the physiological effect of a stressor may occur slightly after the stimulus and may take several seconds or several minutes to recover[27)]. To alleviate this problem, we aggregate the driving stress in a link-based level. That is, we estimate the heart beat increase for each road segment, instead of the heart beat increase second-by-second. Since the lengths of the road segments are different, we normalize the heart beat increase in 1km by dividing the road length.

$$LDS_j = \frac{\sum_{i=1}^{n} DS_{ij}}{L_j} \tag{2}$$

where LDS_j denotes the link-based driving stress of link j, DS_{ij} denotes the driving stress in the i^{th} second in link j, n denotes the number of time interval in link j. Because the time interval for each sample is 1s, $\sum_{i=1}^{n} DS_{ij}$ represent the total heart beat increase in a link.

(2) Driving stress modelling

Driving stress is an effective measure for safety evaluation. However, it is unpractical to collect the physiological signal by requiring drivers wearing the detector in a road network every day. Thus, it is necessary to develop a driving stress model by using other easy-to-measure variables (e.g., speed, acceleration, traffic condition) which can be obtained from probe vehicles or video camera. In this section, we propose a tree-based ensemble model to predict the link-based driving stress by using the relevant variables obtained from GPS, CAN bus, and video data.

In recent years, tree-based ensemble model is popular in solving prediction and classification problems because it not only achieve strong predictive performance but also enables to identify which predictor variables are the most important to make these predictions[28-30)]. These advantages make the tree-based ensemble models good candidates in solving regression problems.

a) Regression tree

A single regression tree model partitions the feature space into a number of regions and fits a simple model for each region. The region space is first partition into two regions, and then each region is further partitions into two more regions. To decide the best splitting point, a greedy algorithm is implemented to maximize the information gain at a tree node (31). The split chosen at each tree node is selected from the set $\arg\max_s IG(D,s)$ where $IG(D,s)$ is the information gain when a split s is applied to dataset D. The information gain is the difference between the parent node impurity and the weighted sum of the children node impurities, which is defined as the following equation.

$$IG(D,s) = I(D) - \frac{N_{left}}{N} I(D_{left}) - \frac{N_{right}}{N} I(D_{right}) \tag{3}$$

where D, D_{left} and D_{right} are the datasets for the parent node, the left-side child node and the right-side child node, respectively; N, N_{left} and N_{right} are the corresponding size of the datasets.

$$I(d) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \mu)^2 \tag{4}$$

$$\mu = \frac{1}{n} \sum_{i=1}^{n} y_i \tag{5}$$

where y_i is the label for an instance in dataset d, n is the sample size of dataset d, and μ is the mean of the instances.

b) Random Forest

A single regression tree usually suffers from high variance, which makes them very unstable. Ensemble methods provide an efficient way to improve the accuracy over a single regression tree. A basic ensemble method is called Bagging[32)], in which trees are generated on random bootstrap samples from the original dataset. In Random Forest[33)], two powerful machine learning techniques are combined: Bagging and random features selection[34)]. For a given training dataset with sample size n, Bagging generates k new training set, each with sample size n, by sampling with replacement. Then, k models are trained by using the k training set and combined through averaging. The second technique in Random Forest is random feature selection. Instead of using all features (predictor variables) as input for each splitting node, it only uses a random subset of features. Thus, it enforces diversity between base models.

The idea in Random Forest is to improve the forecasting performance through variance reduction (Eq.(4)) of Bagging by reducing the correlation between the trees, without increasing the variance too much[35)]. The variance of a random forest is determined by the variance of individual tree (σ^2), correlation between trees (ρ) and the number (M) of trees as follows.

$$VAR = \rho\sigma^2 + \frac{1-\rho}{M}\sigma^2 \tag{6}$$

To optimize the Random Forest model, Liaw and Wiener[36)] suggested growing forests with good number of trees (M) until increasing tree numbers do not improve the accuracy. Breiman[33)] suggested to try different numbers of feature selections by setting $v = \frac{p}{6}, \frac{p}{3}$ and $\frac{2p}{3}$.

Evaluation of variable importance is one significant advantage of Random Forest model, which indicates the contribution of a variable to the output prediction when all other variables are present in the model. The variable importance in a random forests regression model is measured by decrease in node impurity. In the case of regression tree, the node impurity is measured by sample variance (Eq.(4)). The decrease in node impurity is given by information gain (Eq.(3)). The importance of variable X_k is measured by adding up the weighted decreases in node impurities for all nodes where X_k is used for splitting.

$$Imp(X_k) = \frac{1}{M}\sum_T \sum_{D \in T} \frac{N_D}{N} IG(D,s) \tag{7}$$

where N_D is the number of data points at node D, N is the total sample size, T is the regression tree.

3. RESULTS

(1) Variable importance

As shown in Table 1, we collected 24 predictor variables from GPS, CAN bus and video for driving stress analysis. Predictor variables usually have different influences on the output. Variable importance is measured by weighted decreases in node impurities. Figure 4 gives the variable ranking in the predictor set based on their importance in producing accurate predictions. The relative importance of each individual variable is scaled so that the sum of them for all the input variables equals to 1. The average speed contributes the most to the link-based driving stress. A possible reason is that the average speed is one of the most important variables reflecting the general traffic condition and driving behavior. Lower average speed may indicate the unfavorite driving condition such as reverse weather and traffic congestion, while higher average speed usually indicates a smooth or free-flow traffic condition. The second important factor variable is COV of speed, which directly indicates the fluctuation of the traffic condition. A higher COV of speed may indicate an unstable driving state such as following an imprudent leading vehicle or moving jam. Brake and acceleration also ranked in front of most of the variables. This is expected, as brake and acceleration frequency directly reflects the stressor when the drivers deal with a complex driving situation such as conflict with other vehicles or pedestrians. The top four most important variables altogether had approximately 78% contributions in modelling driving stress. It suggests that it is possible to estimate the driving stress by using limited variables such as average speed, COV of speed, brake and acceleration frequency. This finding is important because collecting ECG and a mass of variables is unpractical for real-world implementation. It indicates that it is possible to estimate the driving stress in a large-scale network by developing a simple random forest model with these four variables collect from probe vehicle.

Table 1 Input variables

Variables	Note
COV of Speed	Coefficient of variance of speed in a link
Average Speed	Average speed in a link
Brake	Frequency of brake operation per km
Acceleration	Frequency of acceleration operation per km
Wheel angle	Accumulated wheel angle per km
Divide	Diverging point exists
Merge	Merging point exists
Parking vehicle	Parking vehicle exists
Pedestrian	Number of conflicting pedestrians
Congestion	Congestion occurs
Sidewalk	Sidewalk exists
Maintenance	Road maintenance
Opposite Vehicle	Conflict with opposite vehicle
Number of Lane	Number of lane
Signalization	Signalized control exists
Lane Change	Frequency of lane changing behavior
Temperal Stop	Frequency of stopping behavior
Intersection	Passing through intersection
Crosswalk	Passing through crosswalk
Bus Lane	Bus lane exists
Curve	Passing through a curve link
T_road	Passing through road with T junction
Right-turn lane	Driving on the right-turn lane
Left-turn lane	Driving on the left-turn lane

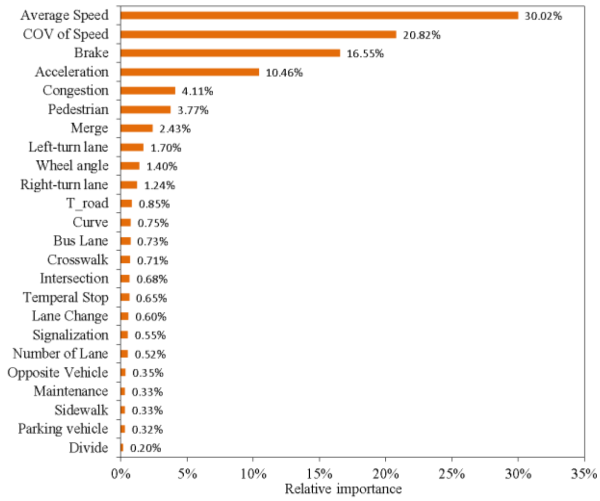


Figure 4 Variable importance

(2) Variable sensitivity

To reveal how the selected variables influence the driving stress, this study proposes the following simple methodology for variable sensitivity analysis. After training the Random Forest structure with a large set of input variables, we calculate an average value for each input variable. Then, holding all variables at their average values but one each time, vary the one input over its entire range and analyze the variability produced in the outputs. Here, sensitivities of the top 4 variables are examined. The average values for average speed, COV of speed, brake and acceleration are set to 9 m/s, 0.25, 41 times/km, and 71 times/km, respectively. As shown in Figure 5(a), the driving stress keeps a relatively high level when the average speed is less than 6 m/s, and the fluctuation of driving stress is also large. This reflects the unstable psychological changes in a low speed situation. When the average speed increases from 6 m/s to 12 m/s, the driving stress gradually reduces to a low level. And the driving stress keeps a relatively stable level when the average speed is larger than 12 m/s. The trend of impacts of average speed on driving stress is similar under different settings of COV of speed. Figure 5 (b) shows a higher COV of speed results in higher driving stress when the average speed is set to 10m/s and 15m/s. However, the increasing trend of driving stress with increasing COV of speed is not significant when the average speed is small (e.g., 5 m/s), but the driving stress keeps a high level. A possible reason is that drivers always feel uncomfortable and nervous in poor traffic situations. When the average speed is larger than 10 m/s, driving stress increases gradually as the COV of speed increases. Figure 5 (c) shows the impact of brake frequency on driving dress. As expected, higher frequency of brake results in higher level of driving stress, especially in low speed situation. Figure 5 (d) shows the relationship between driving stress and acceleration frequency. In lower average speed situation, the trend is not significant but the driving stress keeps a high level. In higher average speed situation, the driving stress increases as the acceleration operation frequency increases. Figure 5 (c) and (d) indicates that the high frequency of stop-and-go behavior in congested condition significantly results in high level of driving stress.

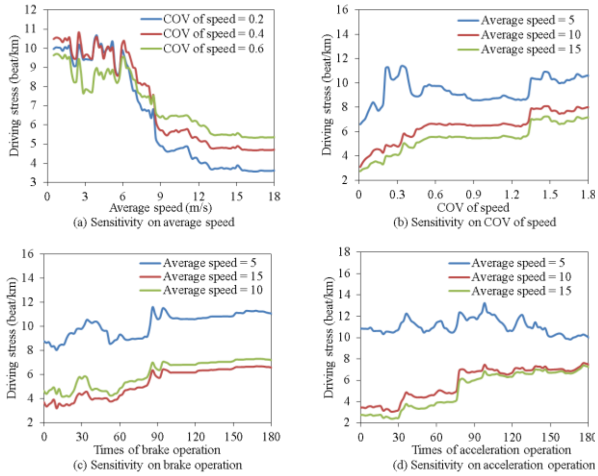


Figure 5 Sensitivity analysis on important variables

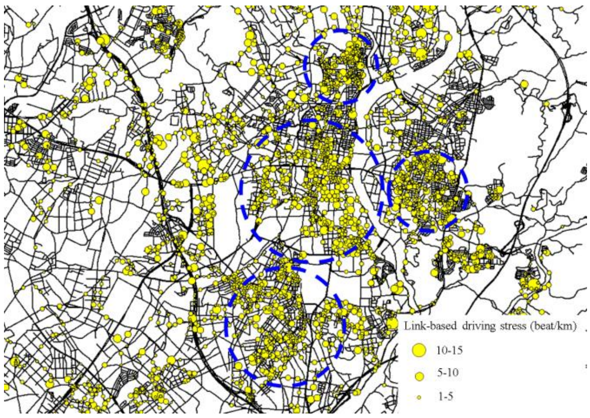


Figure 6 Driving stress mapping in a network-wide scale

4. APPLICATION

A promising application of driving stress estimation is to develop a route guidance system considering traffic safety. Driving stress can be seen as the surrogate index for safety evaluation. Because the variable importance analysis indicates that the top 4 important variables (average speed, COV of speed, brake and acceleration frequency) contribute 78% importance to the proposed model, we can draw the heat map of driving stress distribution by using these variables collected from probe vehicle data. Here, we collect large-scale GPS and CAN bus data from 153 probe vehicles in Toyota city, Japan in 10 months. More than 70,000 trips with millions of second-by-second data records are available for analysis. The GPS trajectories cover the whole road network with 4072 nodes and 12,877 links. Figure 6 shows the driving stress mapping in the whole network. It indicates that most of the high driving stress occurs in the center of the city. It is expected, as the traffic conflict and congestion are more frequent in a dense population area, which may cause higher level of driving stress. The heat map of driving stress helps to find a comfortable route avoiding the location with potentially high driving stress, which is especially beneficial to older drivers and beginners.

5. CONCLUSION AND FUTURE WORK

This study presents a framework for driving stress detection and prediction in a real-world network. The driving stress is quantified by using the heart beat increase in a link-based level. Various data collected from GPS, CAN, and video are used as the predictor variables for estimating the driving stress. The Random Forest, a superior machine learning method, is introduced to model the car driver stress caused by different driving events. The benefits of the proposed method are mainly from two aspects. First, Random Forest provides a method to measure the relative importance of each predictor variable. Second, the prediction accuracy can be improved by using the ensemble techniques such as Bagging and random feature selection.

Based on the results of the driving experiment, key findings are summarized as follows:

(1)Average speed has the most important impact on driving stress. Lower average speed usually results in higher driving stress in urban road network, because drivers may feel nervous in congested traffic condition.

(2)The top four most important variables, i.e., average speed, COV of speed, brake and acceleration frequency, altogether had approximately 78% contributions in modelling driving stress. It suggests that it is possible to estimate the driving stress by using limited variables collected from probe vehicle.

(3)The driving stress heat map shows that most of the high driving stress occurs in the center of the city. A possible reason is that the traffic conflict and congestion are more frequent in a dense population area.

The driving stress model will be incorporated to a route guidance system in the future study. Not only the travel time, but also the driving stress for safety improvement will be considered for the route planning in the next generation navigation system. It will be beneficial to older drivers, beginners, and those drivers who are unfamiliar with the road network.

REFERENCES

1) Hill, J. D., & Boyle, L. N. (2007). Driver stress as influenced by driving maneuvers and roadway conditions. Transportation Research Part F: Traffic Psychology and Behaviour, 10(3), 177-186.
2) Rigas, G., Katsis, C. D., Bougia, P., & Fotiadis, D. I. (2008). A reasoning-based framework for car driver’s stress prediction. In Control and Automation, 2008 16th Mediterranean Conference on (pp. 627-632). IEEE.
3) Matthews, G., Sparkes, T. J., & Bygrave, H. M. (1996). Attentional overload, stress, and simulate driving performance. Human Performance, 9(1), 77-101.
4) Simon, F., & Corbett, C. (1996). Road traffic offending, stress, age, and accident history among male and female drivers. Ergonomics, 39(5), 757-780.
5) Smart, R. G., Cannon, E., Howard, A., Frise, P., & Mann, R. E. (2005). Can we design cars to prevent road rage? International journal of vehicle information and communication systems, 1(1-2), 44-55.
6) Hartley, L. R., & El Hassani, J. (1994). Stress, violations and accidents. Applied Ergonomics, 25(4), 221-230.
7) Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. IEEE Transactions on intelligent transportation systems, 6(2), 156-166.
8) Hajek, W., Gaponova, I., Fleischer, K. H., & Krems, J. (2013). Workload-adaptive cruise control-A new generation of advanced driver assistance systems. Transportation research part F: traffic psychology and behaviour, 20, 108-120.
9) Kanamori, R., Ando, A., Yamamoto, T. and Morikawa, T. (2016). Preliminary study on driving stress with multiple physiological indicators in driving experiment, Presented at 3rd IEEE International Conference on Biomedical and Health Informatics, Las Vegas, USA, 24-27.
10) Healey, J., Seger, J., & Picard, R. (1999). Quantifying driver stress: Developing a system for collecting and processing bio-metric signals in natural situations. Biomedical sciences instrumentation, 35, 193-198.
11) Mehler, B., Reimer, B., Coughlin, J., & Dusek, J. (2009). Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. Transportation Research Record: Journal of the Transportation Research Board, (2138), 6-12.
12) Belyusar, D., Mehler, B., Solovey, E., & Reimer, B. (2015). Impact of Repeated Exposure to a Multilevel Working Memory Task on Physiological Arousal and Driving Performance. Transportation Research Record: Journal of the Transportation Research Board, (2518), 46-53.
13) Yamakoshi, T., Rolfe, P., Yamakoshi, Y., & Hirose, H. (2009). A novel physiological index for Driver’s Activation State derived from simulated monotonous driving studies. Transportation research part C: emerging technologies, 17(1), 69-80.
14) Wang, J. S., Lin, C. W., & Yang, Y. T. C. (2013). A k-nearest-neighbor classifier with heart rate variability feature-based transformation algorithm for driving stress recognition. Neurocomputing, 116, 136-143.
15) Miller, E. E., & Boyle, L. N. (2015). Driver Behavior in Road Tunnels: Association with Driver Stress and Performance. Transportation Research Record: Journal of the Transportation Research Board, (2518), 60-67.
16) Matthews, G., Dorn, L., & Glendon, A. I. (1991). Personality correlates of driver stress. Personality and Individual Differences, 12(6), 535-549.
17) Kontogiannis, T. (2006). Patterns of driver stress and coping strategies in a Greek sample and their relationship to aberrant behaviors and traffic accidents. Accident Analysis & Prevention, 38(5), 913-924.
18) Rowden, P., Matthews, G., Watson, B., & Biggs, H. (2011). The relative impact of work-related stress, life stress and driving environment stress on driving outcomes. Accident Analysis & Prevention, 43(4), 1332-1340.
19) Oz, B., Ozkan, T., & Lajunen, T. (2010). Professional and non-professional drivers’ stress reactions and risky driving. Transportation research part F: traffic psychology and behaviour, 13(1), 32-40.
20) Desmond, P. A., & Matthews, G. (2009). Individual differences in stress and fatigue in two field studies of driving. Transportation Research Part F: Traffic Psychology and Behaviour, 12(4), 265-276.
21) Brookhuis, K. A., and D. De Waard. Assessment of Drivers’ Workload: Performance and Subjective and Physiological Indexes. In Stress, Workload, and Fatigue (P. A. Hancock and P. A. Desmond, eds.), Lawrence Erlbaum Associates, Mahwah, N.J., 2001, pp. 321–333.
22) Apparies, R. J., Riniolo, T. C., & Porges, S. W. (1998). A psychophysiological investigation of the effects of driving longer-combination vehicles. Ergonomics, 41(5), 581-592.
23) Miwa, T., Kiuchi, D., Yamamoto, T., & Morikawa, T. (2012). Development of map matching algorithm for low frequency probe data. Transportation Research Part C: Emerging Technologies, 22, 132-145.
24) Pan, J., & Tompkins, W. J. (1985). A real-time QRS detection algorithm. IEEE transactions on biomedical engineering, (3), 230-236.
25) Daskalov, I., & Christov, I. (1997). Improvement of resolution in measurement of electrocardiogram RR intervals by interpolation. Medical engineering & physics, 19(4), 375-379.
26) Tarvainen, M. P., Niskanen, J. P., Lipponen, J. A., Ranta-Aho, P. O., & Karjalainen, P. A. (2014). Kubios HRV-heart rate variability analysis software. Computer methods and programs in biomedicine, 113(1), 210-220.
27) R. A. Sternbach, Principles of Psychophysiology. New York, NY: Academic,1966.
28) Lin, C. W., Wang, J. S., & Chung, P. C. (2010). Mining physiological conditions from heart rate variability analysis.

IEEE Computational Intelligence Magazine, 5(1), 50-58.
29) Zhang, Y., & Haghani, A. (2015). A gradient boosting method to improve travel time prediction. Transportation Research Part C: Emerging Technologies, 58, 308-324.
30) Saha, D., Alluri, P., & Gan, A. (2015). A Random Forests approach to prioritize Highway Safety Manual (HSM) variables for data collection. Journal of Advanced Transportation.
31) Marsland, S. (2009) Machine learning: An algorithmic perspective. CRC Press.
32) Breiman, L. (1996). Bagging predictors. Machine learning, 24(2), 123-140.
33) Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.
34) Ho, T. K. (1998). The random subspace method for constructing decision forests. IEEE transactions on pattern analysis and machine intelligence, 20(8), 832-844.
35) Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J., Tibshirani, R. (2009). The Elements of Statistical Learning. Springer.
36) Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. R news, 2(3), 18-22.

(Received July 1, 2016)