Evaluation of Drivers' Driving Behavior in Heavy Traffic Situations from OBD-II Data

Yu-Jen Pan

Department of Supply Chain Management, National Kaohsiung University of Science and Technology, Taiwan.
E-mail: yjpan@nkust.edu.tw

Abstracts: The analysis and discussion of the onboard diagnostic data will help understand the driver's behavioral characteristics and develop a sustainable transportation system. The research content of this paper is to mine the driving behavior data through the vehicle preload equipment, analyze the factors affecting safe driving and establish a prediction model. This study collected data from 50 Taiwanese drivers while operating in heavy traffic. Understand the driver's behavioral characteristics through data analysis, such as calculating the times of the driver's emergency braking is based on the driving speed. From this result, the index of dangerous driving is defined. In addition, the dangerous driving index and other variables were analyzed, and it was found that there was a significant relationship between dangerous driving and vehicle mileage and maximum speed. According to the results, the drive speed increases the driving risk, the mileage reduces the risk probability, and the prediction accuracy is 78.9%. Complete data needs to be collected to evaluate complete driving behavior characteristics for developing sustainable transportation goals.

Keywords: Data mining, Behavior analysis, OBD-II, Behavioral characteristics.

1. INTRODUCTION

The overall trend of networking, cross-domain integration of science and technology, innovative wisdom, and vehicle-based communications services have become an international development trend. According to United Nations statistics, in 1950, only 30% of the world's population lived in cities, and the people of cities living in 2014 had risen to 54%. It has been estimated that more than 70% of the population will live in cities by 2050. Rapid population growth has spawned a series of urban problems that include the social issues that urban dwellers use to derive private transport. In addition, global warming is recognized by the world's facts to mitigate the negative impact of climate change, reduce greenhouse gas emissions, and energy consumption has become the current and future development and governance priorities. With the development and improvement of the Internet, tens of millions of vast amounts of data are constantly rising. Car networking data is one of them. Car networking includes two main elements. The first element is an OBD-II device (On Board Diagnostic). The standard device connected to the vehicle, provided with a Subscribers Identify Module (SIM) card inside that transmits mobility information (e.g., location, speed, acceleration, braking, etc.) as well as information from the vehicle's on-board computer (e.g., engine failures, engine temperature, maintenance needs, etc.). The second element is the cloud-based platform, where mobility data are collected, and the information is analyzed. Figure 1 illustrates the different components framing of car networking. Vehicle's data provided by OBD-II provides a database for analyzing user driving behavior, which records vehicle hardware data, geographic information data, and driver behavior data during user driving.

At present, the OBD-II device can capture a lot of information on the vehicle, as shown in Table **1**. However, due to the different standards of each car factory, the information provided by the vehicle will be different. In addition, manufacturers of OBD-II devices will also develop instruments for other purposes based on their applicability. For example, devices used to detect air pollution will provide more values about vehicle intake and exhaust, engine temperature, etc. More and more researchers are beginning to pay attention to OBD-II data. There are mainly several directions in the analysis and research of OBD-II data, such as using OBD-II data to verify the correctness of related research because OBD-II data is a sensor installed initially to read the vehicle [1-2]. The accuracy of the information is extremely high. In addition, according to the OBD-II data, the vehicle parts life and various fuel consumption rate evaluations can be found, providing manufacturers regarding the spare parts marketing methods [3-6].

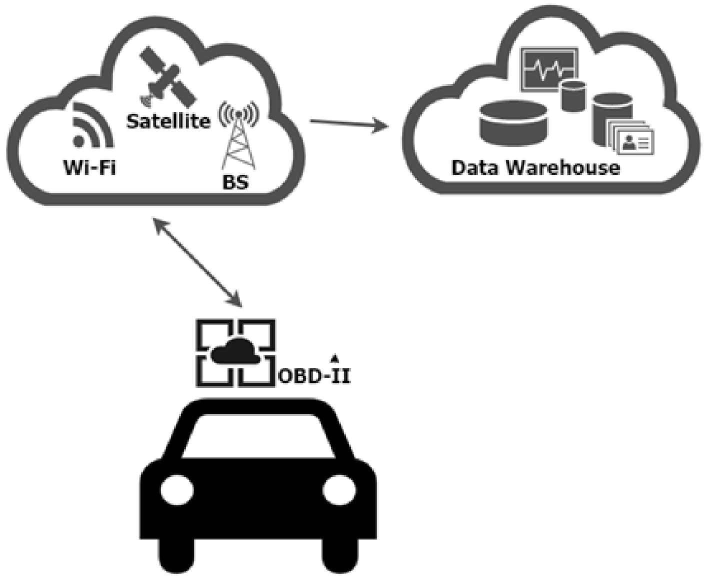

Figure 1. The Diagram of Car Network.

Table 1. Various Data and Units on the OBD-II Device.

Data	Unit
Vehicle speed	km/h
Engine rotations	rpm
Throttle position	%
Engine coolant temperature	°C
MAF air flow rate	grams/sec
Distance travelled	km
Engine fuel rate	L/h
Intake air temperature	°C
Battery voltage	V
Mess air pressure	kPa
Fuel level input	%
Absolute throttle	%
Calculated engine load value	%

Apart from the road's infrastructure and the vehicle's condition, human error is critical in traffic accidents [7]. The significant elements impacting the driving behavior are events that arise during a trip, such as harsh acceleration, sudden brake, sudden lane change, and sharp turns. Therefore, it is crucial and necessary to detect these events. Keeping a check on the driver's driving behavior, monitoring these events, and giving feedback for the observed events can assure safe driving. In this way, potential traffic accidents can be avoided [8].

The emergence of extensive data in the car network brings the possibility of new solutions, and the risk assessment model based on driving behavior is one of them. This study attempts to use OBD-II data to analyze the factors that affect driving behavior and hopes that the results can be the basis for studying the individual driving behavior model. This paper will first explain the concept of data processing. It is mainly based on the analysis target of vehicles driving on congested roads. Next, explain the reasons for using Logis regression to analyze driving behavior and the definition of variables. Finally, in the results section, factors related to dangerous driving are proposed. At the same time, perform classification analysis according to the established prediction model.

2. LITERATURE REVIEW

The purpose of this project is to analyze the relevant factors that affect driving behavior through dynamic information (such as speed, engine speed, engine load, etc.) captured by OBD-II. The current world is facing the petrochemical energy is about to run out, so energy-saving issues born, people began to pay attention to the energy crisis, the relevant research is also mushrooming vigorous development. In the energy-saving driving behavior analysis study, Boriboonsomsinvm scholars and others [9] first proposed a set of assessment methods for energy-saving driving, the process shown in Figure **2**. Boriboonsomsinvm scholars and other people to assess the technique is to consider the state of the vehicle running and the state of travel. The OBD device captures the vehicle's operational status to evaluate fuel consumption and driving conditions, assess emissions, etc. Through these two indicators, the system will give drivers driving advice. However, the energy efficiency indicators developed in the methodology are set as system constants through the existing literature and are not analyzed utilizing data exploration. Then, Wallace scholars and others [10] in 2014 for data collection and analysis of the paper, in the article, through the OBD device to collect the data analysis, Wallace scholars and others, the use of OBD device with GPS, RFID ... Information, containing four drivers of 100 driving trip, hoping to find an expected behavior from the statistical methods to observe the numerical distribution and statistical chart presentation. Similarly, there is no analysis for the individual part of the lack of relevant technology in data exploration.

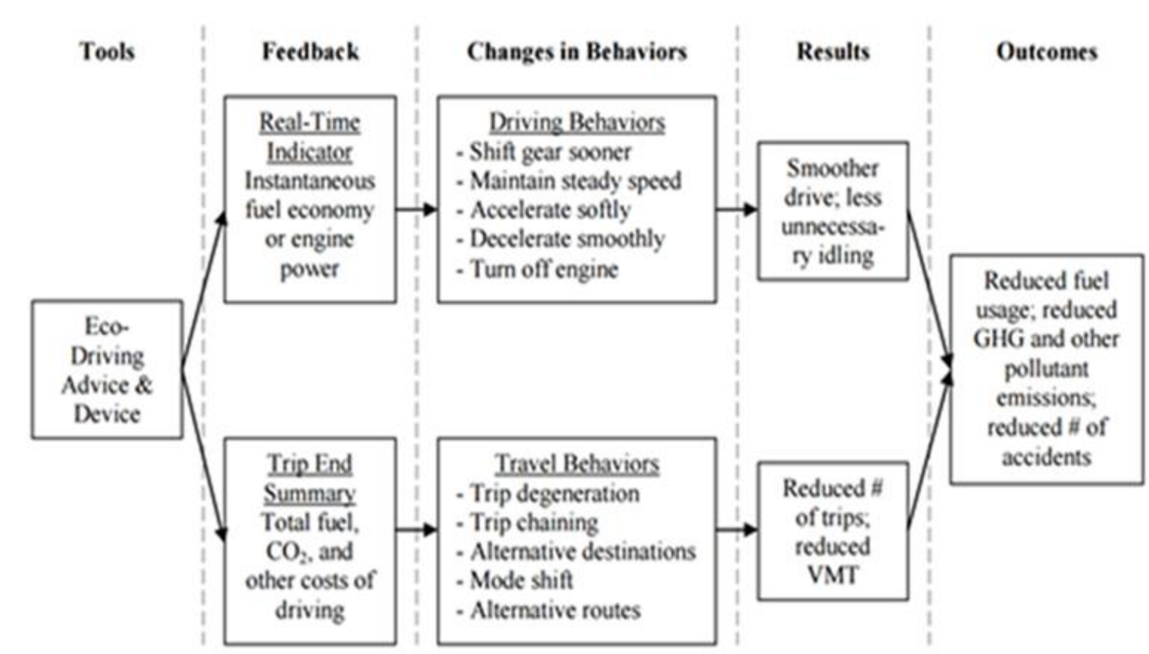

Figure 2. Boriboonsomsinvm Scholars and Others Proposed Energy-Saving Driving Ssessment Method.

Driving risk is another critical issue in driving behavior analysis. Finding risky driving behavior through examination is a difficult challenge. Compared to safe driving, what can say risk-driven data collection is not accessible, so very lacking. So how to define the risk of existing driving behavior is an important issue. Zaldivar et al. [11] proposed a detection system for the detection of accidents on the road. This system combines the OBD-II interface, mobile phone, Internet, emergency ambulance notification system, as shown in Figure **3**. They used the timely vehicle diagnosis of information to determine whether the vehicle was abnormal and early warning to avoid accidents. Or the failure has occurred, you can directly send personnel to rescue through the ambulance notification system. Unfortunately, the system did not conduct a driving behavior analysis for risk driving. Risk driving in recent years in Taiwan has more and more attention, often see the news reports, "Road Sambo" new term birth, more and more people pay attention to the driver's body function with age and not suitable for driving. The Vlahodimitrakou and other scholars [12] for the elderly driving, conducting behavior analysis, put forward an assessment method. Hope to find the elderly driving behavior patterns. In addition, for the usual common speeding, Kumar scholars and others [13] use the visual presentation to limit the driver speeding. Demonstrate the driver by presenting different dashboards at different speed limits. Hsu et al. (2015) presents a framework for the design and development of a multi-sensor selection optimization mechanism for a driver assistance simulation system. The selection algorithm and multiple-vehicle driving examples are successfully implemented in a virtual driving simulator [14].

Figure 3. Zaldivar Scholars Put Forward the Application Situation.

3. METHODS

This paper presents the analysis of the driving behavior of the vehicle body information obtained by the vehicle diagnostic system (OBD-II), focusing on exploring dangerous driving behavior. The data used in this study is the daily driving conditions of 54 cars recorded using OBD-II for three months. The research structure shown in Figure 4 and the OBD-II data are analyzed statistically. First of all, for the definition of dangerous driving, first through the relevant research literature and industry development direction to collect relevant information for pre-processing operations. Next, study the possible factors related to dangerous driving, then check the dangerous driving and the relationship between the relevant factors.

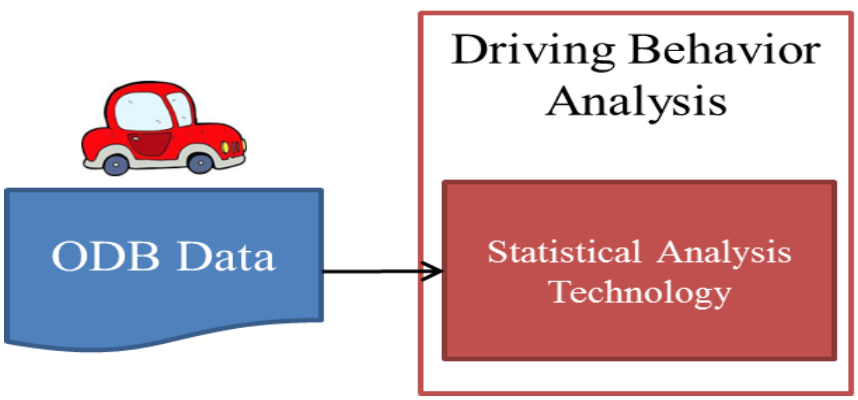

Figure 4. Driving Behavior Analysis Architecture.

3.1. Data Preprocessing

In Feature Construction, we will obtain records of time, moving vehicle speed, engine speed, engine air volume, engine temperature, engine load, fuel consumption, etc., from OBD-II equipment. In addition, we also use GPS data, and most of the records are numerical data. First, we use the moving window method to extract feature values (average mileage, maximum speed, etc.). These feature values are calculated using statistical techniques and are called statistical features. At the same time, we will also refer to the opinions of experts to calculate the eigenvalues. This type of feature is called an expert-based feature.

Due to the different types of vehicles, the data characteristics of each car are different. Therefore, there is a time delay in the data record; each value is not synchronous. So we use a fixed time interval as a sliding window to retrieve information. As shown in Figure 5, for example, we use 180 seconds as a sliding window. For each value in the moving pane, we calculate the average value, maximum value, and minimum value as the characteristic value of each moving pane and find distinct meaningful values from the calculation results.

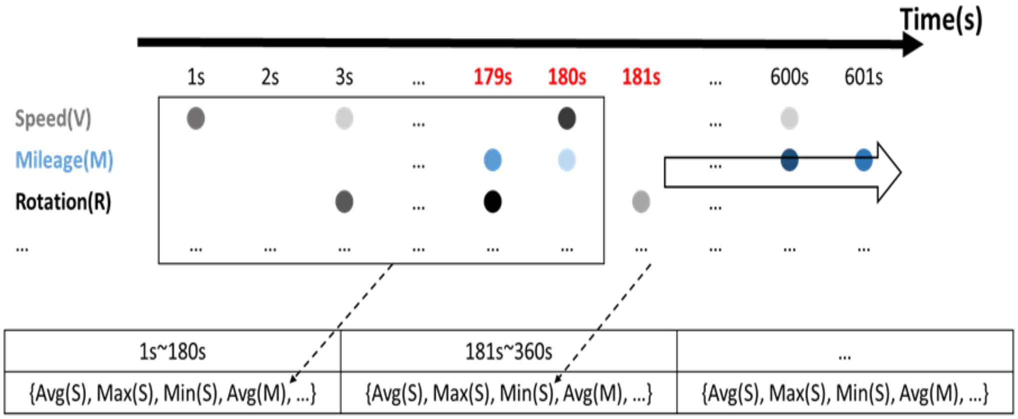

Figure 5. Use the Concept of A Sliding Window to Retrieve Data.

3.2. Driving Behavior Analysis

We know from relevant literature that the number of kilometers a vehicle travels will affect the probability of traffic accidents. Define an approximate dangerous driving standard from the data processing process. When the number of sudden braking times per hour is more than five times, it is considered dangerous driving. Other related descriptions show in Table 2.

We use the variables presented in Table 2 to build a predictive model for judging whether driving behavior is dangerous. Dangerous driving behavior defines the number of emergency braking per hour with more than four times. Figure **6** shows an example where we analyze the driver's driving behavior. The driver braked four times in 55 minutes, so the driver was classified as having dangerous driving behavior. In this way, all drivers are analyzed to determine driving with dangerous driving behaviors.

Table 2. Summary of Analytical Data.

Analyze parameters			
Mileage (km)	**Number of emergency brakes**	**Maximum speed (km/h)**	**Driving time (hours)**
(i) Sample average speed range <= 70 km/h (ii) Dangerous driving behavior definition: the number of emergency braking per hour with more than 5 times (iii) Emergency Brake Definition: ➢ the speed dropped from 35 to 10 below in 4 seconds (low speed) ➢ the speed down more than 40 in 3 seconds (high speed)			

Figure 6. Analyze the Frequency of Emergency Braking by the Driver.

Since the dependent variable is a 0-1 categorical variable, this study uses a classification method to model and predict the data. Commonly used classification methods include Logistic regression, decision tree, neural network, and so on. This study uses Logistic regression to model the data. The reasons are as follows: (1) Logistic regression can show the relationship between the dependent and independent variables through the regression equation. (2) Logistic regression enables variable selection. (3) Logistic regression has a small amount of calculation and a faster calculation speed.

We use Logistic regression to explore the relationship between approximate dangerous driving and related factors:

$$\ln\left(\frac{Prob}{1-Prob}\right) = \beta_0 + \beta_{mlg}avmlg + f(x) \qquad (1)$$

In the above formula, Prob represents the probability of dangerous driving, avmlg represents the number of kilometers the vehicle drives per day, and f(x) represents factors that consider various people and vehicles, including driver's gender, age, vehicle speed, etc. We can use this formula to estimate the mileage and increase the probability of dangerous driving.

4. RESULTS

This study uses logistic regression analysis to analyze the pre-processed data. The relationship between dangerous driving behavior and related factors is shown in Table **3**. Table **3** shows that mileage has significant significance for dangerous driving. Its parameter is negative, indicating that the probability of danger is lower at a higher speed (70 km/h) as the mileage increases. This result shows that under long-distance driving and higher speeds, drivers are less likely to drive dangerously. However, when the vehicle is traveling in a short distance, and at a higher speed, the driver's probability of dangerous driving will increase. Figure **7** shows the judgment of dangerous driving by mileage and vehicle max-speed. Since the vehicle is driving on a congested road, when the vehicle's speed is higher, the frequency of the driver's emergency braking is higher. According to our definition of dangerous driving behavior (Table 2): In a short drive, the more emergency braking, the higher the probability of dangerous driving behavior.

Returned by Logistic analysis:

$$ln\left(\frac{P}{1-P}\right) = (-3.93 - 0.275mileage + 0.096maxspeed) \qquad (2)$$

$$odd = \left(\frac{P}{1-P}\right) = e^{(-3.93-.275mileage+0.096maxspeed)} \qquad (3)$$

In the above formula, P represents the probability of dangerous driving behavior, and mileage means the average daily mileage of the vehicle. Max speed represents the maximum speed of the average daily.

Table 3. Variables in the Equation.

	B	Sig.	Exp(B)	95% C.I. for EXP(B)	
				Lower	Upper
Mileage	-0.275	0.000	0.759	0.668	0.863
Maxspeed	0.096	0.000	1.101	1.050	1.154
Constant	-3.930	0.001	0.020		
Mileage	-0.275	0.000	0.759	0.668	0.863

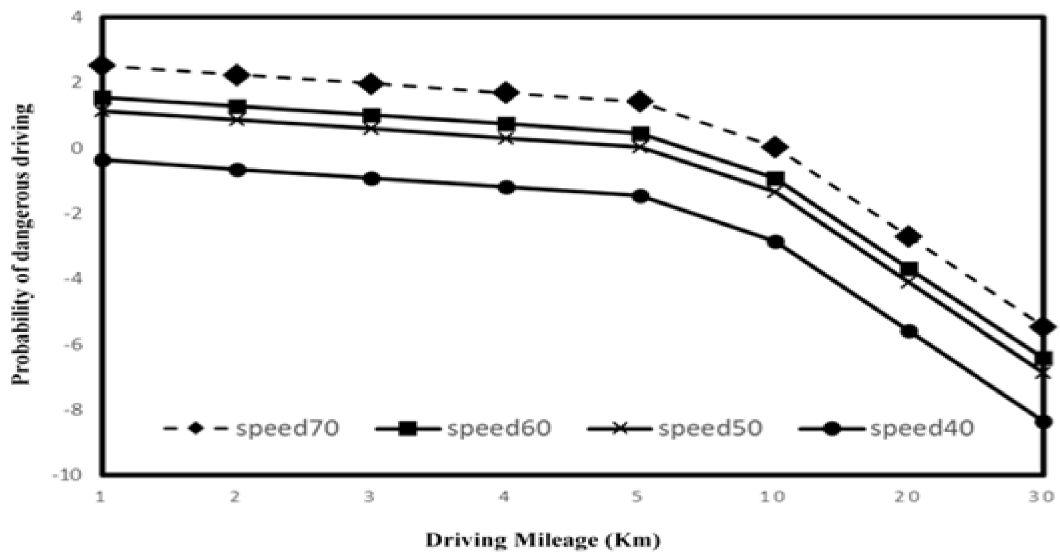

Figure 7. The Judgment of Dangerous Driving.

From the cross-tabulation of predicted classification correctness in Table **4**, the original 58 will not be dangerous driving observations. According to Rogers regression analysis prediction, there has 44 classified in the absence of hazardous driving (correct), 14 Classifieds of dangerous driving (errors), the original 56 counts of dangerous driving. Predicted according to the Rogers regression analysis, 10 were classified as not hazardous driving (error), 46 were Dangerous classified driving (correct).

Table 4. Predictive Accuracy of Logistic Regression.

The number of observations		**The number of predictions**		
		Dangerous driving		**Correction**
		0	**1**	
Dangerous driving	0	44	14	75.9%
	1	10	46	82.1%
Summary percentage	78.9%			
a. Split value = .500				

5. CONCLUSIONS

This paper proposes a driving behavior pattern analysis concept, using data mining technology to analyze OBD-II data. In this method, the system can identify relevant factors and analyze the driving behavior pattern of the user, find the driving behavior pattern, and give corresponding driving style suggestions. According to the model results, the maximum speed increased driving risk. Driving mileage will reduce the risk of going probability, analysis, and prediction of the accuracy rate is 78.9%. However, the OBD-II data was collected in cities with congested traffic. Its properties are short distance and low speed. However, we believe that different driving situations should have different analysis results. For example, does it have the same result in a suburban environment even if it is a short distance driving? Therefore, we will explore driving behavior in different situations to establish a complete driver's behavior analysis model on safe driving.

REFERENCES

[1] Ribeiro V, Rodrigues J, Aguiar A. Mining Geographic Data for Fuel Consumption Estimation, In 2013 16th International IEEE Conference on Intelligent Transportation Systems-(ITSC) 2013; 124-129. https://doi.org/10.1109/ITSC.2013.6728221

[2] Wallace B, Rockwood M, Goubran R, Knoefel F, Marshall S, Porter M. Measurement of Vehicle Acceleration in Studies of Older Drivers from GPS Position and OBDII Velocity Sensors, In 2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA) 2015; 391-396. https://doi.org/10.1109/MeMeA.2015.7145234

[3] Ahmed Q, Iqbal A, Taj I, Ahmed K. Gasoline Engine Intake Manifold Leakage Diagnosis/Prognosis Using Hidden Markov model, Iternational Journal of Innovative Computing, Information and Control (IJICIC) 2012; 8(7): 4661-4674.

[4] Jhou JS, Chen SH, Tsay WD, Lai MC. The Implementation of OBD-II Vehicle Diagnosis System Integrated with Cloud Computation Technology, In 2013 IEEE International Conference on Robot, Vision and Signal Processing (RVSP) 2013; 9-12. https://doi.ieeecomputersociety.org/10.1109/RVSP.2013.55

[5] Lin, J.; Chen, S. C.; Shih, Y. T.; Chen, S. H. A Study on Remote On-Line Diagnostic System for Vehicles by Integrating the Technology of OBD, GPS, and 3G, World Academy of Science, Engineering and Technology 2008; 56: 435-441.

[6] Siegel J, Bhattacharyya R, Deshpande A, Sarma S. Vehicular Engine Oil Service Life Characterization Using On-Board Diagnostic (OBD) Sensor Data, In 2014 IEEE SENSORS 2014; 1722-1725. https://doi.org/10.1109/ICSENS.2014.6985355

[7] Saiprasert C, Pholprasit T, Thajchayapong S. Detection of Driving Events Using Sensory Data on Smartphone, International Journal of Intelligent Transportation Systems Research 2017; 15(1): 17-28. https://doi.org/10.1007/s13177-015-0116-5

[8] Chen C, Zhao X, Zhang Y, Rong J, Liu X. A Graphical Modeling Method for Individual Driving Behavior and Its Application in Driving Safety Analysis Using GPS Data, Transportation Research Part F: Traffic Psychology and Behaviour 2019; 63: 118-134. https://doi.org/10.1016/j.trf.2019.03.017

[9] Boriboonsomsin K, Vu A, Barth M. Eco-Driving: Pilot Evaluation of Driving Behavior Changes Among US Drivers. University of California Transportation Center 2010.

[10] Wallace B, Goubran R, Knoefel F, Marshall S, Porter M, Harlow M, Puli A. Automation of the Validation, Anonymization, and Augmentation of Big Data from a Multi-year Driving Study. In 2015 IEEE International Congress on Big Data 2015; 608-614. https://doi.org/10.1109/BigDataCongress.2015.93

[11] Zaldivar J, Calafate CT, Cano JC, Manzoni P. Providing Accident Detection in Vehicular Networks Through OBD-II Devices and Android-Based Smartphones. In 2011 IEEE 36th Conference on Local Computer Networks (LCN) 2011;813-819. https://doi.org/10.1109/LCN.2011.6115556

[12] Vlahodimitrakou Z, Charlton JL, Langford J, Koppel S, Di Stefano M, Macdonald W, Smith GA. Development and evaluation of a Driving Observation Schedule (DOS) to Study Everyday Driving Performance of Older Drivers. Accident Analysis & Prevention 2013; 61: 253-260. https://doi.org/10.1016/j.aap.2013.03.027

[13] Kumar M, Kim T. Dynamic Speedometer: Dashboard Redesign to Discourage Drivers from Speeding. In CHI'05 extended abstracts on Human factors in computing systems 2005; 1573-1576. https://doi.org/10.1145/1056808.1056969

[14] Hsu CM, Lian FL, Hsieh YC, Tseng SP. Multi-sensor selection optimization and driver warning decision for dynamical virtual driving simulator. Journal of the Chinese Institute of Engineers 2016; 39(3): 303-314. https://doi.org/10.1080/03088839.2015.1112470

DOI: https://doi.org/10.15379/ijmst.v10i2.1214