Driving Behavior Analysis Based on Vehicle OBD Information and AdaBoost Algorithms

Shi-Huang Chen, Jeng-Shyang Pan, and Kaixuan Lu

***Abstract*—This paper proposes a novel driving behavior analysis method based on the vehicle on board diagnostic (OBD) information and AdaBoost algorithms. The proposed method collects vehicle operation information, including vehicle speed, engine RPM, throttle position, and calculated engine load, via OBD interface. Then the proposed method makes use of AdaBoost algorithms to create a driving behavior classification model, and finally could determine whether the current driving behavior belongs to safe driving or not. Experimental results show the correctness of the proposed driving behavior analysis method can achieve average of 99.8% accuracy rate in various driving simulations. The proposed method has the potential of applying to real-world driver assistance system.**

***Index Terms*—Driving behavior analysis, driver assistance system, AdaBoost algorithm, on board diagnostic (OBD)**

I. Introduction

Currently, with the economy developing, the amount of the vehicles increases every year. As the same time, the amount of non-professional drivers increases rapidly. Since most novice drivers are unskilled, unfamiliar with the vehicle condition and in weak awareness of traffic safety, the drivers' personal factors have become the main reasons of traffic accidents. The driving auxiliary equipment is urgently needed to remind drivers the vehicle information in time and correct improper driving behavior.

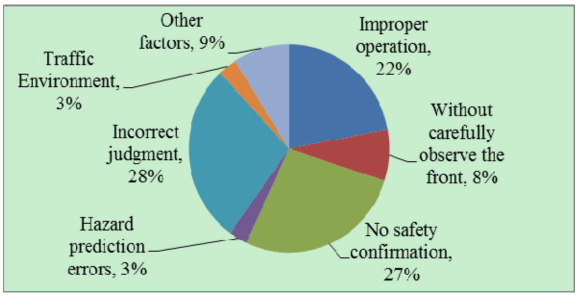


Fig. 1. Distribution of Traffic Accidents Statistics [1]-[2].

Fig. 1 shows the distribution of traffic accidents statistics [1]-[2]. According to the statistical analysis data shown in Fig. 1, the most accidents are caused by the personal factors of driver. They include improper operation, incorrect judgment, and etc [1]-[2]. To further reduce the accident and to protect the driver/pedestrian safety, it should not only continuously improve the vehicle safety equipment as well as road conditions, but also should pay attention to the driver as the research object.

Various literatures have introduced driving behavior analysis methods based on computer vision image processing, multi-sensor fusion, and etc. Yoshifumi Kishimoto and Koji Oguri proposed a HMM (hidden Markov model)-based method to estimate the possibility of the driver braking from past driving period [3]. Antonio Prez, M. Isabel Garca, Manuel Nieto, and other scholars realized a system called "Argos". The Argos is an advanced in-vehicle data recorder which can help other researchers to better study the driver behavior analysis [4]. Reza Haghighi Osgouei and Seungmoon Choi completed the driving behavior model for different drivers based on objective principles. Their objective principle is determined by a random distance of HMM model trained by each driver [5]. Cuong Tran, Anup Doshi, and Mohan Manubhai Trivedi make use of video optical changes and HMM to complete the driver's foot movement estimated algorithm [6]-[7]. The scholars of University of California extract the video images from outside of the vehicle to analyze the driving behavior by the positional relationship with other vehicle, and then judge the vehicle is in safe state or danger state [8]. Shan Bao and Linda Ng Boyle completed intersection driving experiments in different age drivers and established a different age driver rearview mirror observation model [9]. Toshikazu Akita and others proposed the driver behavior modeling method based on the fusion system, which is mainly recognition and analysis the car following state, accurate recognition rate is higher [10].

According to the above literatures, current research methods of driving behavior analysis include the driving data collection, driving modeling algorithms, and applications. Driving data collection includes automotive video capture, car-mounted sensors, and the on board diagnostic (OBD). In terms of driving behavior modeling algorithms, there are HMM, support vector machine (SVM), decision trees, and other principles. The main application of the driving behavior analysis is the identification of driving lethargy or the driver's actions forecast. In this paper, a novel driving behavior analysis method based on the vehicle OBD and AdaBoost algorithms is proposed. This proposed method collects the vehicle operation information, including vehicle speed, engine RPM, throttle position, and calculated engine load,

Manuscript received Jan. 8, 2015, revised Feb. 10, 2015. This work was supported in part by the Ministry of Science and Technology, Taiwan under Grant NSC 101-2221-E-366-002-MY3.

Shi-Huang Chen is with the Department of Computer Science and Information Engineering, Shu-Te University, Kaohsiung City, 824, Taiwan. (phone: +886-7-6158000 ext. 4607; Fax: +886-7-6152559; e-mail: shchen@stu.edu.tw).

Jeng-Shyang Pan is with the Fujian University of Technology and the Harbin Institute of Technology Shenzhen Graduate School, China (e-mail: jengshyangpan@gmail.com)

Kaixuan Lu was with the Harbin Institute of Technology Shenzhen Graduate School, China (e-mail: lukaixuan203@sina.com)

from the OBD interface. Then the proposed method makes use of AdaBoost algorithms to create a driving behavior classification model, and finally could determine whether the current driving behavior belongs to safe driving or not. Experimental results show the correctness of the proposed driving behavior analysis method can achieve average of 99.8% accuracy rate in various driving simulations.

The remaining sections of this paper are organized as follows. Section II brief introduces the AdaBoost theorem. The detail of the proposed driving behavior analysis method is presented in Section III. Section IV shows experimental results. Finally, Section V concludes this paper.

II. The basic principle of AdaBoost

AdaBoost is a classical classification machine learning algorithm. The basic principle of AdaBoost algorithm is to use a large number of weak classifiers combined together by a certain method form a strong classifier. The strong classifier has a strong ability of classification [11]-[13]. Strong classifier generated as follows:

Assuming given a two-classification training data set:

$$T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\} \tag{1}$$

where each sample composed by the instance and flags. Instance $x_i \in \mathrm{X} \in R^n$, flags $y_i \in Y=\{-1,+1\}$, X is the instance space, Y is flags set. AdaBoost use the follow algorithm to generate the strong classifier.

A. Basic AdaBoost Algorithm

Input: training data set $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$,

where $x_i \in \mathrm{X} \in R^n$, $y_i \in Y=\{-1,+1\}$; weak learn algorithm.

Output: Strong classifier $G(x)$.

(1) Initialization of the weight value distribution of the training data

$$D_1=(w_{11},\cdots,w_{1i},\cdots,w_{1N}),\ w_{1i}=\frac{1}{N},\ i=1,2,\cdots,N \tag{2}$$

(2) $m=1,2,\cdots,M$ (m is the times of train.)

(a) Using training data set has the weight distribution D_m to learn, get the basic classification

$$G_m(x): X \to \{-1,+1\} \tag{3}$$

(b) The classification error rate of $G_m(x)$ is calculated on the training data

$$e_m=P(G_m(x)\neq y_i)=\sum_{i=1}^{N} w_{mi} I(G_m(x)\neq y_i) \tag{4}$$

(c) Calculation the coefficient of $G_m(x)$, the logarithm is the natural logarithm

$$\alpha_m=\frac{1}{2}\log\frac{1-e_m}{e_m} \tag{5}$$

(d) Update the weight value distribution of the training data

$$D_m=(w_{m+1,1},\cdots,w_{m+1,i},\cdots,w_{m+1,N}) \tag{6}$$

$$w_{m+1,i}=\frac{w_{mi}}{Z_m}\exp(-\alpha_m y_i G_m(x_i)) \tag{7}$$

where, Z_m is normalization factor which could make the D_m become a probability distributions.

$$Z_m=\sum_{i=1}^{N} w_{mi}\exp(-\alpha_m y_i G_m(x_i)) \tag{8}$$

(3) Build a linear combination of basic classifiers

$$f(x)=\sum_{m=1}^{M}\alpha_m G_m(x) \tag{9}$$

The final classification can be expressed as

$$G(x)=sign(f(x))=sign\left(\sum_{m=1}^{M}\alpha_m G_m(x)\right) \tag{10}$$

III. Driving behavior analysis method

The proposed driving behavior analysis method consists of driving operation acquisition module, data preprocessing module, and AdaBoost classification 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 AdaBoost algorithms 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.

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 procedure. After data preprocess step, each time slice samples can be regarded as the rate of change the driving operation. This paper uses the training set to establish the driving classification model by the AdaBoost algorithms, and then the proposed method could use the test set to judgment the accuracy of the model. Finally, the proposed method judges the merits of driving behavior. Fig. 2 shows the flowchart of the entire proposed method.

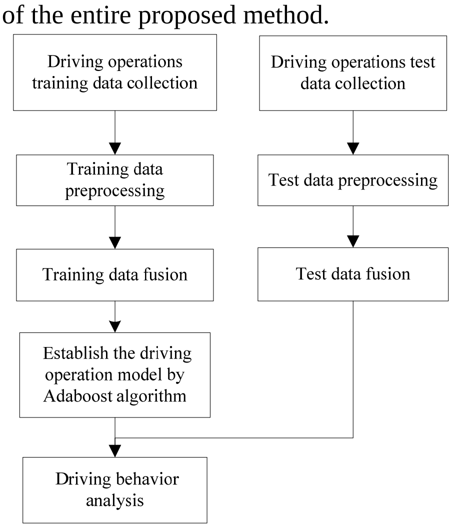


Fig. 2. The flowchart of the proposed driving behavior analysis method.

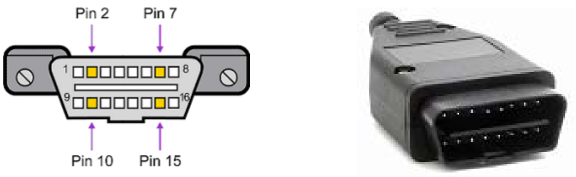
A. OBD System

Current OBD system, or called OBD-II, was proposed in 1996 to replace the deficient former system, namely OBD-I. The specifications of OBD-II were recognized by the Environmental Protection Agency (EPA), US, and California Air Resources Board for air-pollution control. Since 1996, all vehicles are required to be equipped with OBD-II under EPA regulation in the USA. When vehicle exhausts higher level of air-pollution contents, OBD-II will generate Diagnostic Trouble Code (DTC) messages and Check Engine light will display. Some symbols of Check Engine light are shown in

Fig. 3. Meanwhile, OBD-II will save this DTC in the memory inside ECU. Thus the DTC messages can be retrievable through an OBD-II scan tool [14]-[16].

Fig.3. Variety of icons of Check Engine light.

The main features of OBD-II are (1) unified J1962 16-pin socket and data link connector (DLC) (as shown in Fig. 4); (2) unified DTC and meanings; (3) storage and display DTC; (4) contains vehicle record capability; and (5) auto-clear or reset function for the DTC [18]. Hence, the advantage of OBD-II is its standardization. In other words, just one set of OBD-II scan tool is able to perform the diagnosis and can scan against variety of vehicles which equipped with OBD-II system.



(a) (b)

Fig. 4. (a) J1962 OBD-II 16-pin socket, (b) OBD-II DLC.

The OBD-II socket has 16 pins and is usually installed below the driving panel. Among these 16 pins, nine of them have fixed functions and the rest pins are left to the discretion of the vehicle manufacturer [16].

B. OBD Signal Generator

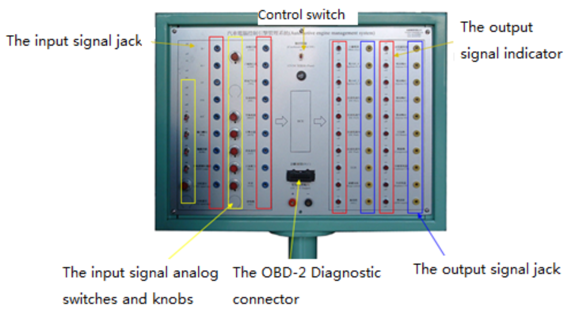


Fig. 5. The control panel of OBD-II signal generator.

This paper uses a customized OBD-II signal generator to complete and simulate the OBD-II vehicle information data collection. The control panel of the OBD-II signal generator is shown in Fig. 5. The left and right sides of the control panel are OBD data input control terminal and output control terminal, respectively. The input control terminal has knob input and jack input methods. The knob input is manual adjustment mode. One can control multiple data changes on the OBD-II signal generator, including the commonly 7 kinds of OBD data. Jack input is the external analog signal input. The outputs control terminal has indicator lights and jack output method. There are 15 kinds of engine running data. Indicator lights will show the operation status of engine components, e.g., fuel injections, ignition coils, and etc. The jack output can achieve data output displayed on other devices, such as an oscilloscope. The middle of the control panel contains interlocking switches and OBD-II data link connector. The interlocking switches is designed to control the synchronization between the throttle value and the engine RPM. When it is in the open state, the accelerator pedal is available to control the throttle and engine RPM at the same time. When it is in the close state, one can control the throttle and engine RPM individually. The OBD-II data link connector achieves the OBD data output.

Because the OBD-II interface is not a standard equipment of general computer, an OBD-II to Bluetooth adapter was fabricated to accommodate with the PC to acquire the vehicle status. This paper selects EZ-SCAN5 as the OBD-II to Bluetooth adapter. EZ-SCAN5 supports most of OBD-II communication protocols including SAE J1850 PWM, SAE J1850 VPW, ISO 9141-2, ISO 14230-4 KWP, and ISO 15765-4 CAN. The OBD-II messages will be decoded and transmitted to OBD-II diagnosis encoder via EZ-SCAN5.

C. The Decode of OBD Signal

The OBD-II vehicle information data collected in this paper include the vehicle speed, engine speed (RPM), throttle position, and engine load. These data could reflect the results of the driving behavior parameter data. One can decode these OBD data via OBD Mode01 operation with appropriate PIDs [18]. The PID stands for Parameter ID. The PIDs of vehicle speed, engine speed (RPM), throttle position, and engine load are defined in SAE J1979 [17] and given in Table 1.

Table 1. PIDs definitions of vehicle speed, engine speed (RPM), throttle value, and engine load

PID	Data Bytes	Description	Units	Formula#
OD	1	Vehicle speed	km/h	A
OC	2	Engine speed	rpm	((A*256)+B)/4
11	1	Throttle position	%	A*100/255
04	1	Engine load value	%	A*100/255

#A and B are represented the returned data bytes.

Engine speed is one of the most important parameter in the engine output data. The best driving operation is not just the high vehicle speed and low engine speed, also the driver's operation can make the car through the low efficiency area of the engine in the shortest time and keep the vehicle speed and engine speed running in smooth. Vehicle speed is a direct response to the driver's operating. It can directly reflect the current driving state whether speeding or maintain safe. The throttle position is adjusted to send the number of fuel into engine. Suitable throttle opening in order to ensure the efficient operation of the engine. Too high or too low value of throttle will cause incomplete combustion of fuel and air pollution. The engine load is an indicator of a driver operating environment. The climbing or running overload will cause the engine load to too large. It is easy to damage the engine and therefore is not conducive to safe driving.

This paper develops two criteria for data collection based on the engine characteristic curve.

(1) The normal vehicle condition data:

The relative ratio of the vehicle speed and engine speed is maintained at between 0.9 and 1.3 (test in the same gear); the relative ratio of the engine speed and throttle valve is maintained at between 0.9 and 1.3; the engine load is maintained between 20% and 50%.

ISBN: 978-988-19253-2-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

(2) The bad vehicle condition data:

The relative ratio of the vehicle speed and engine speed is maintained more than 1.3 or less than 0.9; the relative ratio of the engine speed and throttle valve is maintained more than 1.3 or less than 0.9; the engine load is maintained greater than 50% or less than 20%.

According to the above data collection procedure, a total of 1000 sets of vehicle information simulation data, including normal and bad driver condition, are collected in this paper. Some collected vehicle data are listed in the Table 2.

Table 2. The vehicle OBD-II information data (partial) used in this paper.

Time	The vehicle speed km/h	Engine speed r/min	Throttle %	Engine load %
1	3	625	3.9	16.1
2	20	679	10.2	27.5
3	25	812	26.7	27.5
4	28	914	36.5	27.5
5	28	937	38.8	27.5
6	33	937	38.8	27.5
7	37	1011	43.5	27.5
8	39	1070	46.3	27.5
9	46	1171	51.8	27.5
10	49	1292	57.6	28.2
11	58	1441	63.1	27.5
12	65	1558	67.1	27.5
13	66	1558	67.1	28.2
14	68	1871	73.3	28.2
15	70	2082	75.3	28.2
16	73	2496	82.7	28.2
17	77	2878	86.3	28.2

D. Vehicle OBD Information Data Preprocessing

This paper uses three characteristics, i.e., the relative ratio of the vehicle speed and engine speed, the relative ratio of throttle position and engine speed, and the engine load to analyze whether the current state of the driving behavior is safe or dangerous. First, the proposed method needs to compute the change rate of vehicle speed, engine speed and throttle position by the Equation (11), where $t_2 - t_1 = 1$.

$$D(t) = \frac{data(t_2) - data(t_1)}{t_2 - t_1} \tag{11}$$

A part of results are given in Table 3.

Table 3. The vehicle OBD-II data change rate

Time	Vehicle Speed change rate	Engine speed change rate	Throttle change rate	Engine load %
1	17	54	6.3	16.1
2	5	133	6.5	27.5
3	3	98	10.2	27.5
4	0	23	2.3	27.5
5	5	0	0	27.5
6	4	74	4.7	27.5
7	2	59	2.8	27.5
8	7	1	5.5	27.5
9	3	121	5.8	27.5
10	9	149	5.5	28.2
11	7	117	4	27.5
12	1	0	0	27.5
13	2	1558	6.2	28.2
14	2	211	2	28.2
15	3	414	8	28.2
16	4	382	3.7	28.2
17	77	2878	86.3	28.2

Because there are larger difference between the values of vehicle speed, engine speed, and throttle opening. This paper should compute the relative ratio of the vehicle speed and engine speed, the relative ratio of throttle position and engine speed. Calculation method shows in Equations (12) and (13):

$$R_{cz}(t) = \frac{cs(t)}{220} \div \frac{zs(t)}{8000} \tag{12}$$

where $R_{cz}(t)$ is the relative ratio of the vehicle speed and engine speed, $cs(t)$ is the vehicle speed at time t with maximum speed of 220, $zs(t)$ is the engine speed at time t with maximum engine speed of 8000.

$$R_{jz}(t) = \frac{jq'(t)}{\max(jq')} \div \frac{zs'(t)}{\max(zs')} \tag{13}$$

where $R_{jz}(t)$ is the relative ratio of throttle position and engine speed, $jq'(t)$ is the change rate of throttle position at time t, $zs'(t)$ is the change rate of engine speed at time t, $\max(jq')$ and $\max(zs')$ are the maximum change rate values of throttle position and engine speed, respectively.

Finally, the proposed method combines these three feature values which are the relative ratio of the vehicle speed and engine speed, the relative ratio of throttle position and engine speed, and engine load to determine whether the vehicle data is normal or bad driving condition. The proposed method adds positive and negative labels to each data sample. OBD information data sample is stored as following format:

$S_{OBD}(t)$ = {the vehicle speed and engine speed, the relative ratio of throttle position and engine speed, engine load, the label}

IV. Experimental results

This paper uses the toolkit-GML-AdaBoost-matlab [19] of MATLAB to complete the data preprocessing and driving behavior modeling. After pretreatment, a set of experimental data sample is shown in Table 4. It follows from Table 4 that there are three characteristics and one label. The number of iterations of AdaBoost algorithm is obtained by the three different ways. They are Gentle AdaBoost, Modest AdaBoost, and Real AdaBoostas. The results of Real AdaBoost are better than those of Modest AdaBoost and Gentle AdaBoost. The accuracy rates of driving behavior analysis based on these three AdaBoost algorithms are given in Table 5.

Table 4. A set of the vehicle OBD information data samples

Time	The relative ratio of the vehicle speed and engine speed	The relative ratio of throttle valve and engine speed	Engine load %	label
1	1.0232	1.4397	27.5	-1
2	1.0866	0.1213	27.5	-1
3	1.7368	1	26.7	-1
4	1.4383	1	26.3	-1
5	0.9017	1	14.1	-1
6	1.0866	1	27.5	1
7	0.9172	1	26.3	1
8	1.1193	1	26.3	1
9	1.2383	1	28.2	1
10	1.0676	1	26.3	1

In contrast experiments, this paper also completes training the identification function based on SVM algorithm. Under the optimal parameters c = 0.25, g = 4 condition, the recognition accuracy rate is 94.7475%. The result is poorer than the results of AdaBoost algorithms.

Table 5. AdaBoost algorithm accuracy rates under different iterations

Iteration time	5	10	15	20
Gentle AdaBoost	99.39%	100%	100%	100%
Modest AdaBoost	92.12%	97.98%	98.99%	100%
Real AdaBoost	100%	100%	100%	99.8%

V. Conclusion

A novel driving behavior analysis method using vehicle on board diagnostic (OBD) information and AdaBoost algorithms is described in this paper. The proposed driving behavior analysis method utilizes OBD interface to collect a number of critical driving operation data, i.e., vehicle speed, engine speed (RPM), throttle position, and calculated engine load. Then it uses AdaBoost algorithms to comprehensive the analysis of driving behavior. Experimental results show the correctness of the proposed driving behavior analysis method can achieve average of 99.8% accuracy rate in various driving simulations. Using the real AdaBoost algorithm, the proposed method could obtain 100% accuracy rate under 15 iterations. The proposed method has the potential of applying to real-world driver assistance system.

References

[1] Miyaji, M., Danno, M., Oguri, K.: Analysis of driver behavior based on traffic incidents for driver monitor systems. In: Intelligent Vehicles Symposium, IEEE(2008) pp. 930-935.

[2] Liang, J., Cheng, X., Chen, X.: The research of car rear-end warning model based on mas and behavior. In: Power Electronics and Intelligent Transportation System, IEEE (2008), pp. 305-309 .

[3] Kishimoto Y, Oguri K. A modeling method for predicting driving behavior concerning with driver's past movements, IEEE International Conference on Vehicular Electronics and Safety (IEEE ICVES 2008), pp: 132-136.

[4] Perez A, Garcia M I, Nieto M, et al. Argos: an advanced in-vehicle data recorder on a massively sensorized vehicle for car driver behavior experimentation, IEEE Transactions on Intelligent Transportation Systems, 2010, 11(2): pp.463-473.

[5] Osgouei R H, Choi S. Evaluation of driving skills using an hmm-based distance measure. 2012 IEEE International Workshop on Haptic Audio Visual Environments and Games (HAVE 2012), pp. 50-55.

[6] Tran C, Doshi A, Trivedi M M, Modeling and prediction of driver behavior by foot gesture analysis. Computer Vision and Image Understanding, 2012, 116(3), pp.435-445.

[7] Tran C, Doshi A, Trivedi M M, Pedal error prediction by driver foot gesture analysis: A vision-based inquiry. 2011 IEEE Intelligent Vehicles Symposium, 2011, pp. 577-582.

[8] Sivaraman S, Trivedi M M. Active learning based robust monocular vehicle detection for on-road safety systems. 2009 IEEE Intelligent Vehicles Symposium, 2009, pp. 399-404.

[9] Bao S, Boyle L N. Age-related differences in visual scanning at median-divided highway intersections in rural areas. Accident Analysis & Prevention, 2009, 41(1), pp. 146-152.

[10] Akita T, Suzuki T, Hayakawa S, et al. Analysis and synthesis of driving behavior based on mode segmentation, 2008 IEEE International Conference on Control, Automation and Systems, pp. 2884 - 2889.

[11] Wu B, Ai H, Huang C, et al. Fast rotation invariant multi-view face detection based on real AdaBoost., 2004 the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, pp.79-84

[12] Hang Li, "Statistical Learning Method." Tsinghua University Press, 2012.

[13] Parviz, M., Moin, M. S.: Boosting approach for score level fusion in multimodal biometrics based on AUC maximization, Journal of Information Hiding and Multimedia Signal Processing, 2(1), 2011, pp.51-59.

[14] *Diagnostic Trouble Code Definitions Equivalent to ISO/DIS 15031-6*, SAE Standard J2012, 2002.

[15] *E/E Diagnostic Test Modes — Equivalent to ISO/DIS 15031-5*, SAE Standard J1979, 2002.

[16] Diagnostic Connector Equivalent to ISO/DIS 15031-3, SAE Standard J1962, 2002.

[17] SAE J1979/ISO 15031-5, E/E Diagnostic Test Modes, SAE International. Revised: 2014-08-11.

[18] SAE International. On-Board Diagnostics for Light and Medium Duty Vehicles Standards Manual. Pennsylvania, 2003

[19] http://www.codeforge.com/article/239875 (GML AdaBoost Matlab Toolbox Manual)