DRIVER DRIVING BEHAVIOUR PROFILING USING MACHINE LEARNING TECHNIQUES

LOW JIA MING

SESSION 2016/2017

FACULTY OF COMPUTING AND INFORMATICS
MULTIMEDIA UNIVERSITY

OCTOBER 2016

DRIVER DRIVING BEHAVIOUR PROFILING USING MACHINE LEARNING TECHNIQUES

BY

LOW JIA MING

SESSION 2016/2017

THIS PROJECT REPORT IS PREPARED FOR

FACULTY OF COMPUTING AND INFORMATICS
MULTIMEDIA UNIVERSITY
IN PARTIAL FULFILLMENT

FOR

BACHELOR OF COMPUTER SCIENCE

B.C.S (HONS) SOFTWARE ENGINEERING

FACULTY OF COMPUTING AND INFORMATICS

MULTIMEDIA UNIVERSITY

October 2016

DECLARATION

I hereby declare that the work has been done by myself and no portion of the work contained in this Thesis has been submitted in support of any application for any other degree or qualification on this or any other university or institution of learning.

_____

Low Jia Ming
Faculty of Computing and Informatics
Multimedia University
Date: 24:08:2016

ACKNOWLEDGEMENTS

Thanks guys. I owe you many.

To my parents, my husband, and my daughter.

ABSTRACT

This can be your **Management Summary** or **Abstract**. An abstract or management summary should be not more than one page in length. The abstract should allow the reader or moderator who is unfamiliar with the work to gain a swift and accurate impression of what the project is about, how it arose and what has been achieved.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

The preface in a report is something that comes before the report. This section will typically set up the stage for whatever your report is going to discuss. It may give some background information on the subject.

Normally a preface it will be a three paragraph length answer. The first paragraph should be explaining what you are investigating and why. the second should be the scope of your investigation. the third should be the conclusion that your investigation brought you to.

If your report does not have any preface, you may remove it from your latex.

CHAPTER 1

INTRODUCTION, BACKGROUND STORY, MOTIVATIONS

1.1 Basic Introduction

Currently, the amount of vehicles on the road increases every year in Malaysia. It is because the local brand car is affordable by many low income level household. The manufacturer also provided promotions to attract people to buy car. So that, the number of non-professional driver rapidly increased in Malaysia. Most of the drivers are unskilled and lack of awareness on the traffic safety and vehicle condition. The driver's personal factors have become the main reason of causing the traffic incidents.

According to the general road accident data in Malaysia took from Malaysian Institute of Road Safety Research (MIROS) offcial website shown in Figure 1.1, the Malaysia government put effort on reducing the amount of traffic incidents by introducing the new traffic laws and speed track system. However, the number of cases of road deaths does not drop significantly.

General Road Accident Data in Malaysia (1997 – 2014)

Year	Registered Vehicles	Population	Road Crashes	Road Deaths	Serious Injury	Slight Injury	Index per 10,000 Vehicles	Index per 100,000 Population	Indeks per billion VKT
1997	8,550,469	21,665,600	215,632	6,302	14,105	36,167	7.37	29.1	33.57
1998	9,141,357	22,179,500	211,037	5,740	12,068	37,896	6.28	25.8	28.75
1999	9,929,951	22,711,900	223,166	5,794	10,366	36,777	5.83	25.5	26.79
2000	10,598,804	23,263,600	250,429	6,035	9,790	34,375	5.69	26.0	26.25
2001	11,302,545	23,795,300	265,175	5,849	8,680	35,944	5.17	25.1	23.93
2002	12,068,144	24,526,500	279,711	5,891	8,425	35,236	4.9	25.3	22.71
2003	12,819,248	25,048,300	298,653	6,286	9,040	37,415	4.9	25.1	22.77
2004	13,828,889	25,580,000	326,815	6,228	9,218	38,645	4.52	24.3	21.1
2005	15,026,660	26,130,000	328,264	6,200	9,395	31,417	4.18	23.7	19.58
2006	15,790,732	26,640,000	341,252	6,287	9,253	19,885	3.98	23.6	18.69
2007	16,813,943	27,170,000	363,319	6,282	9,273	18,444	3.74	23.1	17.6
2008	17,971,901	27,730,000	373,071	6,527	8,868	16,879	3.63	23.5	17.65
2009	19,016,782	28,310,000	397,330	6,745	8,849	15,823	3.55	23.8	17.27
2010	20,188,565	28,910,000	414,421	6,872	7,781	13,616	3.4	23.8	16.21
2011	21,401,269	29,000,000	449,040	6,877	6,328	12,365	3.21	23.7	14.68
2012	22,702,221	29,300,000	462,423	6,917	5,868	11,654	3.05	23.6	13.35
2013	23,819,256	29,947,600	477,204	6,915	4,597	8,388	2.90	23.1	12.19
2014	25,101,192	30,300,000	476,196	6,674	4,432	8,598	2.66	22.0	10.64

Figure 1.1: General Road Accident Data in Malaysia (1997 - 2014)

The driver characteristics and the occurence of traffic incident is interrelated.

To further reduce the number of accidents, the safety equipment of the vehicle need to be improved as well as the road regulation, but also pay attention on driver behaviour.

1.1.1 Vehicle Operation Data Collection

The behaviour of the driver is hard to be identified. The driver behavior is affected by environment, vehicle condition and the mental or physical state. One way to identify driver behaviour is using the vehicle operation data.

1.1.1 (a) OBD System

The OBD system is also called OBD-II, was proposed in 1996. In 1996, all the cars manufactured in United State (US) were required to equip OBD-II and the cars without OBD-II prohibited to sell in US. The purpose to have OBD-II specifications is to diagnose engine problem. The specifications were being used by the Environment Protection Agency (EPA) and the state of California to meet the emission standards. Since 1996, all the cars in US are required to be equiped with OBD-II to establish the EPA regulation.

The usage of the OBD-II is important for detecting the vehicle exhaustion. If the vehicle is exhaust high level of air-pollution content, Diagnostic Trouble Codes (DTCs) will be geneated by the OBD-II and a Check Engine Light will be displayed on vehicle dashboard. A OBD-II scanning tool can access the DTCs from the Engine Control Unit (ECU).

1.1.1 (b) ELM327

ELM327 is used in the data collection process. The ELM327 is a programed microcontroller. It is an interface to communicate with the On Board Diagnotics port of the vehicle. The ELM327 supports most of OBD-II protocols. ELM327 also contains the bluetooth adapter. The ELM327 needs to be plugged to the OBD-II port that can be found under the vehicle dashboard and above the pedals.

1.1.1 (c) Torque(Lite)

Torque Lite version is a free android application and it is able to be installed in smartphone from Google Play Store. The application will communicate with the ELM327 through the bluetooth connection. The application will collect the data received from the ELM327 and save the data into a .csv file in the smartphone.

1.1.2 Driver Behavior Analysis

Takahiro Wada, Shun'ichi Doi, Keisuke Imai, Hiroshi Kaneko, Naohiko Tsuru, and Kazuyoshi Isaji introduced a performance index for approach and alienation to analyze the drivers' behavior in longitudinal direction. The performance index is calculated based on the relative velocity and distance with the preceding car. The data is collected through a driver simulator. Based on the index, drivers' brake pattern can be analyzed.

1.1.3 Driving Behavior Analysis Based on AdaBoost Algorithms

Shi-Huang Chen, Jeng-Shyang Pan, and Kaixuan Lu proposed a driving behavior analysis based on AdaBoost Algorithms. The proposed method also based on vehicle OBD information. The proposed method colected the vehicle operation data, such as vehicle speed, engine speed, throttle position and engine load. The proposed method calculated the relative ratio of the vehicle speed and the engine speed, the relative ratio of the engine speed and throttle position, and engine load as three characteristics for the modeling usage. Then it uses the AdaBoost Algorithms to classify the driving behavior based on the three characteristics.

1.2 Project Objective

1. To identify the features that contribute to the accuracy of the classification of the driver behavior analysis from the vehicle operation data.

2. To profile the drivers based on the vehicle operation data.

1.3 Research Motivation

This project is designed to capture actual driving behavior data by using sensor. The driver behavior will be analysed based on the vehicle operation data collected from the sensor plugged in the particular vehicle. This project will use the ELM327 as a device that communicates with the vehicle OBD-II interface. The GPS data and vehicle operation data will be collected through the application Torque(Lite) via bluetooth. The performance index for approach and alienation will be used in this projetc as a feature to analyze the drivers' driving behaviour. The driver who exceed the speed limit can be detected by using the vehicle speed and the GPS location information.

1.4 Project Scope

This project focuses on the driver behavior profiling. The vehicle operation data will be collected and pre-processed before execute the analysis. The subjects are 10 adult drivers (2 female and 8 male) who owned the driving license at least two years. Each driver is required to drive the car at least 10 minutes for vehicle operation data collection. The data will be recorded in every second. For each driver, there are at least 600 records in the vehicle operation data file. After using the machine learning technique to classified the drivers, the drivers will be categorized to three classes. The classes are low risk, medium risk and high risk.

CHAPTER 2

RELATED WORK, LITERATURE REVIEW

2.1 Related Work

There are various literatures proposed to introduce the driving behaviour analysis methods.

2.1.1 Driver Behaviour Analysis through Speech Emotional Understanding

This paper was proposed by N. Kamaruddin and A. Wahab. The scholars analyzed the driver behavior state (DBS) based on the emotion of the driver when the driver was driving. The emotion of the driver can be detected through speech. The scholars used the Berlin dataset and NAW dataset as training set. The Berlin dataset and NAW dataset are standard dataset for speech emotion recognition and have been used by many researchers. The proposed method used the Generic Self-organizing Fuzzy Neural Network (GensoFNN) as a classifier for identification purpose.

2.1.1 (a) Generic Self-organizing Fuzzy Neural Network (GensoFNN)

Fuzzy neural network is the combination of neural network and fuzzy inference system. It contains the interpretability of fuzzy inference system and the adaptability of neural network. GensoFNN is a classifier that can generate consistent result as it is able to do self-clustering process of the training data.

2.1.1 (b) Driver Behavior Analysis Method

This proposed method collected data from 11 adults (3 female and 8 male). The drivers have at least two years of driving experience and drive at least 10 hours per week. The drivers are required to do four designed actions. The actions are:

1. driver talking through the mobile phone while driving.
2. driver feeling sleepy.
3. driver laught while driving
4. driver in the initial driving exercise where the driver is in neural state of emotion.

A microphone is embeded in the vehicle to collect the speech of the driver while driving. The experiments to analyzed the three DBS of talking, laughing, and sleepy were conducted. The scholar use the Berlin dataset and NAW dataset to related the three DBS with the angry, happy, and sad emotion.

2.1.1 (c) Conclusion

The results were showed that the sleepy DBS can be recognized as sad emotion consistently. However, the talking and laughing DBS gave mixed results. It means that more work need to be conducted for better classification of the two DBS. The accuracy of sleepy emotion detection is up to 65% using the proposed speech emotion recognition system.

2.1.2 Driver Behaviour Analysis and Route Recognition by Hidden Markov Models

THis paper was proposed by Amardeep Sathyanarayana, Pınar Boyraz, and John H.L. Hansen.

2.2 Literature Review

2.2.1 Driving Behavior Analysis Based on Vehicle OBD Information and AdaBoost Algorithms

This paper was proposed by Shi-Huang Chen, Jeng-Shyang Pan, and Kaixuan Lu. The scholars analyzed the drivers' behavior based on the on board diagnostic (OBD) information and using the AdaBoost Alogorithms to create the driving behav-

ior classification.Finally, the experimental results show the correctness of the proposed driving behavior analysis method has 99.8% accuracy rate in various driving simulations.

2.2.1 (a) AdaBoost Algorithms

AdaBoost is a classification machine learning algorithm. The AdaBoost algorithms is to form a strong classifier by combining the a large number of weak classifier. There are three different types of AdaBoost algorithms. They are Gentle AdaBoost, Modest AdaBoost and Real AdaBoost.

2.2.1 (b) Driving Behaviour Analysis Method

This proposed method used the OBD-II ssytem in the vehicle and EZ-SCAN5 as the OBD-II to Bluetooth adapter to collect the vehicle operation information. OBD-II is proposed in 1996 to replace the OBD-I system. The OBD-II system is implemented in every vehicle uncer the Environmental Protection Agency (EPA) regulation in USA since 1996. When teh air-pollution contents exhausted by the vehicle exceed teh minimum level,the OBD-II system of the particular vehicle will generate the Diagnostic Trouble Code (DTC) message and Check Engine light will display on the vehicle dashboard. In order to communicate with the OBD-II system of the vehicle, EZ-SCAN5 as a OBD-II to Bluetooth adapter is required. The EZ-SCAN5 support most of the OBD-II communication protocols, such as SAE J1850 PWM, SAE J1850 VPW, ISO 9141-2, ISO 14230-4 KWP, and ISO 15765-4 CAN. If the communication protocol of the OBD-II system is not supported by the adapter, the vehicle operation information will not be able to retrieve.

This proposed method collected vehicle speed, engine speed(RPM), throttle position and engine load as the vehicle operation information. According to the engine characteristic curve, the proposed method developed two criteria for the data collection.

1. The normal vehicle condition data

   The relative ratio of the vehicle speed and the engine speed is remained in a range that between 0.9 and 1.3. The result was tested in the same gear. The relative ratio of the engine speed and throttle valve is remained in a range that between 0.9 and 1.3. The engine load is remained between 20% and 50%.

2. The bad vehicle condition data

   The relative ratio of the vehicle speed and the engine speed is out of the range that between 0.9 and 1.3. The result was tested in the same gear. The relative ratio of the engine speed and throttle valve is out of the range that between 0.9 and 1.3. The engine load is out of the range that between 20% and 50%.

2.2.1 (c) Vehicle OBD information Data Preprocessing

The proposed method used three characteristics. The three characteristics are the relative ratio of the vehicle speed and engine speed, the relative ratio of throttle position and engine speed, and engine load. Using the characteristics to analyzed the current state of the driving behavior whether the driver is in safe state or dangerous state.The proposed method needed to compute the change rate of vehicle speed, engine speed and throttle position in the first step. The calculation shown in Equation(2.1), where $t_2 - t_1 = 1$.

$$D(t) = \frac{data(t_2) - data(t_1)}{t_2 - t_1} \tag{2.1}$$

The next step is to calculate the relative ratio of the vehicle speed and engine speed, and the relative ratio of the throttle position and engine speed. The calculation shown in the Equation(2.2) and (2.3).

$$R_{cz}(t) = \frac{cs(t)}{220} \div \frac{zs(t)}{8000} \tag{2.2}$$

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

speed a time t. The 8000 is the value of maximum engine speed of the vehicle.

$$R_{jz}(t) = \frac{jq'(t)}{max(jq')} \div \frac{zs'(t)}{max(zs'(t))} \tag{2.3}$$

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

Based on the two computed features and engine load, this three features were combined to determine the vehicle data is normal or bad driving behavior.

2.2.1 (d) Experimental Results

The proposed method used the tookit-GML-AdaBoost-matlab of MATLAB to execute data preprocessing and driving behavior modeling. The preprocessed data is tested on three diferent types of the AdaBoost algorithms. They are Gentle AdaBoost, Modest AdaBoost and Real AdaBoost. The Real AdaBoost are better than other AdaBoost algorithms with the highest accuracy rate, 99.8

CHAPTER 3

DUMMY CHAPTER

3.1 Research Methodology and Design

In this chapter, you can define your research methodlogy and how you studied the problem and what you used – materials, subjects and equipment. Describe on how you performed the research – methods and procedure.

You can include your pseudocode or algorithm in here.

Algorithm 1 MyTestingProcedure

```
1: procedure MYTESTINGPROCEDURE(G,p)
2:     if p ≥ G then
3:         p ← G
4:     else
5:         if p + k ≤ G then
6:             G ← p + k
7:         end if
8:     end if
9: end procedure
```

$$X = \sum_{i=1}^{n} (x^2 - x)$$

$$Y = \left[\prod_{i=1}^{n} (X|x_n) \right]^{\frac{1}{2}}$$

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant

morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Test 3

Figure 3.1: Let's see. What have we got here?

CHAPTER 4

CONCLUSION

4.1 Introduction

In this last chapter, you may outline the success of your project when compared to the objectives that were set. You may suggest further work for your research area.

4.2 Conclusion

A good final year report should summarise the most important findings and conclude. Always make explanations complete. Avoid speculation that cannot be tested in the foreseeable future. Discuss possible reasons for expected or unexpected findings.

APPENDIX A

MANUALS, TECHNICAL SPECIFICATIONS, DOCUMENTATIONS, EXAMPLE SCENARIOS

You may want to include appendix in your report. Appendix such as manuals, technical specification, or documentations. You should **NOT** include all your source codes as appendix. Generally source code should be included in CD/DVD and **NOT** in your report.

APPENDIX B

APPENDIX 2: WHAT IS APPENDIX

Appendix is included in your report as it is information that is not essential to explain your findings, but that supports your analysis (especially repetitive or lengthy information), validates your conclusions or pursues a related point should be placed in an appendix (plural appendices). Sometimes excerpts from this supporting information (i.e. part of the data set) will be placed in the body of the report but the complete set of information (i.e. all of the data set) will be included in the appendix. Examples of information that could be included in an appendix include figures/tables/charts/graphs of results, statistics, questionnaires, transcripts of interviews, pictures, lengthy derivations of equations, maps, drawings, letters, specification or data sheets, computer program information.

There is no limit to what can be placed in the appendix providing it is relevant and reference is made to it in the report. The appendix is not a catch net for all the semi-interesting or related information you have gathered through your research for your report: the information included in the appendix must bear directly relate to the research problem or the report's purpose. It must be a useful tool for the reader

NOTES

1. This is a footnote, or rather an endnote. Note that footnotes/endnotes are not encouraged in scientific and engineering disciplines.

2. don’t you agree?