Predicting events from physiological data while driving

Lei Wang

A thesis submitted for the degree of Master of Computing at The Australian National University

June 2015
Except where otherwise indicated, this thesis is my own original work.

Lei Wang
3 June 2015
Acknowledgments

I would like to express my deepest appreciation to all those who provided me the possibility to complete this project. A special gratitude I give to my supervisors, Prof. Tom Gedeon, Leana Copeland, and Christopher Chow, whose contribution in stimulating suggestions and encouragement, helped me to coordinate my project especially in experiment conducting.

Furthermore I would also like to acknowledge with much appreciation Xuyanying Zhu, who help me lean the structure of stress model software.

A special thanks goes to my mum and dad, who encourage me to overcome the difficulty in writing report. Thank you to my boyfriend, who cheers me up when I feel frustrated. Last but not least, many thanks go the subjects of the experiment, who spent their time participating in the experiment.
Abstract

This project investigated the feasibility of predicting road events by physiological data while driving. Eleven subjects were asked to drive with driving simulator in virtual city environment for 10 minutes. Galvanic skin responses and electrocardiogram were measured during physiological monitoring. Besides, their virtual world driving performances were recorded. With preprocessing and analysis data collected in the experiment, a road-event-prediction classifier was build by Extreme learning machine algorithm. K-fold cross method was applied to validate this classifier.

From the statistical evaluation of the classification results this study found that subjects’ physiological data have the ability to indicate the occurrences of corresponding road events. The performance of the road-event-prediction classifier not only suffers from inappropriate division of events into several categories, but also suffers from imbalanced class distributions in datasets. Moreover, there is correlation between the occurrence of serious incidents and facial expression change.
Contents

Acknowledgments iv

Abstract v

1 Introduction 1
  1.1 Motivation and Objectives ........................................ 1
  1.2 Literature Survey .................................................. 3
    1.2.1 The Use of Physiological Data While Driving .......... 3
    1.2.2 Previous Work .................................................. 5
  1.3 Hypothesis .......................................................... 5
  1.4 Thesis outline ...................................................... 6

2 Design of building the Road-event-prediction Classifier with GSR and ECG data 7
  2.1 Data Preparation .................................................... 7
  2.2 Train the Classifier with Extreme Learning Machine ........ 8
  2.3 Validate the Classifier with K-fold Cross Validation ........ 9
    2.3.1 Motivation of Validation ..................................... 9
    2.3.2 Cross Validation .............................................. 10
  2.4 Architecture of building the Classifier ......................... 11
  2.5 Summary ............................................................ 12

3 Experiment 13
  3.1 Previous Work ...................................................... 13
  3.2 Experiment Setup ................................................... 14
    3.2.1 Hardware and Software Setup ............................... 15
# Contents

- 3.2.2 Parameter Setting ........................................... 16
- 3.2.3 Survey Questionnaire ...................................... 17
- 3.3 Experiment Conduct ........................................... 18
  - 3.3.1 Experiment Subjects ...................................... 18
  - 3.3.2 Experiment Procedure .................................... 18
- 3.4 Collected Data ................................................ 19
- 3.5 Summary ........................................................ 20

## 4 Data Preprocessing and Analysis ............................... 21
- 4.1 Data need to be Preprocessed ................................ 21
- 4.2 Data Preprocessing ............................................. 22
  - 4.2.1 Select Useful Subjects’ Data .............................. 22
  - 4.2.2 Clean GSR and ECG data ................................ 22
  - 4.2.3 Filter and Normalize GSR and ECG Data ............. 24
  - 4.2.4 Generate Features from GSR and ECG data ......... 26
  - 4.2.5 Label Video Sequence Data ............................... 28
- 4.3 Train and Test the Classifier with Data ..................... 29
- 4.4 Predict Events with the Classifier ........................... 30
- 4.5 Summary ........................................................ 30

## 5 Results and Discussion ........................................ 32
- 5.1 Objectives ..................................................... 32
- 5.2 Statistical Evaluation of Classification results and discussion ... 33
  - 5.2.1 The comparison between prediction results and facial expression change ......................... 39
  - 5.2.2 Summary .................................................... 39

## 6 Conclusion and Future Work .................................. 40
- 6.1 Conclusion .................................................... 40
- 6.2 Future Work ................................................... 42
## Contents

<table>
<thead>
<tr>
<th>Appendix A</th>
<th>Participant Consent Form</th>
<th>44</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix B</td>
<td>Survey Questionnaire</td>
<td>45</td>
</tr>
<tr>
<td>Appendix C</td>
<td>Build the road-event-prediction classifier in Matlab</td>
<td>47</td>
</tr>
<tr>
<td>Appendix D</td>
<td>Independent Study Contract</td>
<td>50</td>
</tr>
<tr>
<td>Bibliography</td>
<td></td>
<td>53</td>
</tr>
</tbody>
</table>
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>The architecture of building the classifier</td>
<td>12</td>
</tr>
<tr>
<td>3.1</td>
<td>Experiment Hardware and software setup</td>
<td>15</td>
</tr>
<tr>
<td>4.1</td>
<td>The GSR and ECG data figure for subject 7</td>
<td>23</td>
</tr>
<tr>
<td>4.2</td>
<td>The raw GSR data figure for subject 4</td>
<td>24</td>
</tr>
<tr>
<td>4.3</td>
<td>The cleaned GSR figure for subject 4</td>
<td>25</td>
</tr>
<tr>
<td>4.4</td>
<td>The comparison between prediction results and facial expression change</td>
<td>31</td>
</tr>
<tr>
<td>A.1</td>
<td>Participant Consent Form sample</td>
<td>44</td>
</tr>
<tr>
<td>B.1</td>
<td>Survey Questionnaire part 1</td>
<td>45</td>
</tr>
<tr>
<td>B.2</td>
<td>Survey Questionnaire part 2</td>
<td>46</td>
</tr>
<tr>
<td>B.3</td>
<td>Survey Questionnaire part 3</td>
<td>46</td>
</tr>
<tr>
<td>C.1</td>
<td>Build the road-event-prediction classifier in Matlab part 1</td>
<td>47</td>
</tr>
<tr>
<td>C.2</td>
<td>Build the road-event-prediction classifier in Matlab part 2</td>
<td>48</td>
</tr>
<tr>
<td>C.3</td>
<td>Build the road-event-prediction classifier in Matlab part 3</td>
<td>49</td>
</tr>
<tr>
<td>D.1</td>
<td>Independent study contract part 1</td>
<td>51</td>
</tr>
<tr>
<td>D.2</td>
<td>Independent study contract part 2</td>
<td>52</td>
</tr>
</tbody>
</table>
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>NeuLog software experiment parameter setting</td>
<td>17</td>
</tr>
<tr>
<td>3.2</td>
<td>ScreenRecorder software experiment parameter setting</td>
<td>17</td>
</tr>
<tr>
<td>4.1</td>
<td>Features generation types and quantity</td>
<td>27</td>
</tr>
<tr>
<td>4.2</td>
<td>Event category and corresponding event tag for seven categories</td>
<td>28</td>
</tr>
<tr>
<td>5.2</td>
<td>Counting table for six categories</td>
<td>34</td>
</tr>
<tr>
<td>5.1</td>
<td>Counting table for seven categories</td>
<td>34</td>
</tr>
<tr>
<td>5.3</td>
<td>Event category and corresponding event tag for four categories</td>
<td>36</td>
</tr>
<tr>
<td>5.4</td>
<td>Counting table for four categories</td>
<td>36</td>
</tr>
<tr>
<td>5.5</td>
<td>Event category and corresponding event tag for three categories</td>
<td>37</td>
</tr>
<tr>
<td>5.6</td>
<td>Counting table for three categories</td>
<td>37</td>
</tr>
<tr>
<td>5.7</td>
<td>Event category and corresponding event tag for two categories</td>
<td>38</td>
</tr>
<tr>
<td>5.8</td>
<td>Counting table for two categories</td>
<td>38</td>
</tr>
</tbody>
</table>
Glossary

**Arb**  Arbitrary analog units. 17

**ECG**  Electrocardiogram. 2

**ELM**  Extreme Learning Machine. 5

**FPS**  Frames Per Second. 17

**GSR**  Galvanic skin response. 2

**HF**  High frequency band. 27

**HZ**  Hertz. 27

**LF**  Low frequency band. 27

**MATLAB**  Matrix laboratory. 22
Chapter 1

Introduction

1.1 Motivation and Objectives

The collection of valuable information on drivers’ behaviors contributes to the development of intelligent vehicular systems that is able to interpret and deal with different situations in traffic [Malta et al., 2008]. It is very vital whether a driver can immediately make precise operation after perceiving dangers and making correct decision in brain. The time between making correct decision in brain and taking precise action is called reaction time [Evans, 2004]. For most driver the reaction time is in the rage of 0.6 to 1 and is also highly depended on the individual and mental state [Green, 2000]. Older drivers have longer reaction time than young driver [Alm and Nilsson, 1995]. Some mental state such as depression and fatigue would increase the reaction time [Burns and Lansdown, 2000].

However, for most drivers, even they have detected that dangers, they cannot take immediately actions such as braking or rotating steering wheel. In this situation, if the driver’s reaction time even decreases only 1 second, it would make a big difference on the stopping distance. When a driver is driving in 72 km/h, the stopping distance would decrease 2 meters and the probability of collisions would be decrease to some extent [Davis, 2003]. If a car system can percept the dangerous via drivers’ physiological data, it would be very helpful to develop a safe driving system to help driver control the car when
the serious incident or accident happen.

When some dangers things happen to human, they would feel stressful and some physiological data would well indicate how stressful they are [Pecchinenda, 1996]. Stress can be defined as a body’s method of reacting to a challenge. As for human, stress typically describes a negative condition or a positive condition that can have an impact on a person’s mental and physical [str]. The different road events or different traffic congestion levels can result in different conditions. For example, the events, like rushing to sidewalk or smashing into a tree, result in negative condition. On the other hands, the events, such as driving in a wide, straight and empty highway, lead to positive conditions. For most drivers, they would react differently for such different conditions. In the consequence, different road events would lead to different stress level.

Some physiological data could reflect the individual’s stress level. In previous study, Dong P. Jang’s experiment has suggested that there are two kinds of physiological data, galvanic skin response (GSR) and electrocardiogram (ECG), which can be used as objective measures in monitoring the reaction of non-phobic participants to virtual-world fear driving and flying, which is set by different situations, such as driving in different traffic congestion level and flying in different weather conditions [Jang et al., 2002]. Besides, in Christine and Fatma’s report, they list 21 references which are studying mapping physiological signals to human emotions, which is elicited by a variety of events, like easy or difficult memory task solving, Unpleasant or neutrality film showing, and real life inductions and imagery [Lisetti and Nasoz, 2004].

In summary, this report aims at discussing the correlation between the physiological data of driver and road events that driver meets. A rod-event-
prediction classifier is suggested to map the physiological data and road events. Before building this classifier, an experiment is required to be conducted to collect data. After data preprocessing and analysis, some machine learning techniques are implemented to build and validate the classifier. Based on the classification results, in the report the findings describe the factors, which affect the road-event-prediction classifier performance. Following these findings, conclusions are drawn and a number of recommendations are made to improve this project in the future.

1.2 Literature Survey

1.2.1 The Use of Physiological Data While Driving

Physiological monitoring has widely been used in virtual world and real world driving tasks. In generally, physiological data is regarded as a metric to assess driver’s irritation level [Malta et al., 2008]. As the different road and traffic conditions would also affect driver’s stress level, physiological data is considered for providing feedback about a driver’s state, which can also be collected continuously without interrupting driver’s task performance [Healey and Picard, 2005].

Dong P. Jang analyzed non-phobic participants’ physiological reaction to virtual driving task. During the physiological monitoring galvanic skin response, heart rate, and skin temperature were measured. In this study, it has been found that galvanic skin response and heart rate variability can be used to show arousal of participants, who were exposed to the virtual environment experience. Besides, these two physiological measures generally returned to normal over time. The result of Dong P. Jang’ research suggest that galvanic skin response and heart rate can be used as objective measures
which aim at monitoring the reaction of non-phobic participants to virtual environment [Jang et al., 2002]. Moreover, when it comes to assess the emotional state of participants, it would be helpful to refer to heart rate variability.

Jennifer A. Healey and Rosalind W. Picard have conducted an experiment to collect physiological data during real world driving tasks, and have presented the methods for analyzing physiological data to determine a driver’s relative stress level [Healey and Picard, 2005]. They recorded four kinds of physiological data while drivers followed a set route through open roads in the greater Boston area: electrocardiogram, electromyogram, galvanic skin response, and respiration. After analyzing the results, they found for most drivers their galvanic skin response and heart rate are highly correlated with driver stress level. This finding indicated that physiological data are able to be a metric of driver’s stress level. In the consequence, in the future it is possible to integrate physiological monitoring into car system.

In a multi-modal real world driving data collection and analysis research, physiological data take part in estimating a driver’s irritation level [Malta et al., 2008]. In this research, galvanic skin response is used to represent the physiological state of driver. Physiological data is also used in the research about the management of secondary tasks while real world driving [Collet et al., 2009]. This research found that galvanic skin response and heart rate shows that arousal level increase as a function of dual-task requirements, the in-vehicle conversation eliciting the same strain as the remote conversation. In a research about driving condition recognition, researcher used heart rate variability to recognize different driving condition [Wang et al., 2010]. An approach for real time stress trend detection is presented by using physiological data in wearable computing systems for automotive drivers [Singh et al., 2011]. The physiological the researcher used are galvanic skin response,
photoplethysmogram, and heart rate.

### 1.2.2 Previous Work

Sharma [Sharma and Gedeon 2014] and Xuanying [Zhu 2014] have developed a computational model to evaluate subject’s stress in virtual environments. The main differences between Sharma’s work and Xuanying’s work are shown on the type of physiological data they used to build the stress model, the machine learning algorithm to build the stress classifier, and on the experiment tasks, which is used for collecting physiological data. However, they use same methods to generate statistical features from physiological data. These methods to generate statistical features from physiological data are referred in this project. As Xuanying shows that Extreme Learning Machine (ELM) algorithm is efficient than traditional Artificial Neuron Network algorithm, in this project, we choose Extreme Learning Machine algorithm to build the road-event-prediction classifier. Besides, Lor [Lor 2013] and Le [Le 2013] have set up the hardware and software of driving simulator, so this project would benefit from their work. Additionally, the author have conducted the driving simulator experiment based on Lor and Le’s work before, and the survey questionnaire is also referred to the past experiment survey questionnaire.

### 1.3 Hypothesis

We want to predict potentially hazardous situations by drivers’ GSR and ECG data, so this report would present a method to map GSR and ECG data to hazardous event level. Thus, a road-event-prediction classifier is required to build to predict the events. Before building and validating the classifier, several hypothesis need to be listed.
1) Subjects’ physiological data such as GSR and ECG, which is monitored while they are driving in the virtual world, are able to indicate the occurrences of corresponding road events to some extent.

2) There should be some factors would affect the road-event-prediction classifier performance.

3) There would be correlation between the occurrence of serious incidents and facial expression change.

1.4 Thesis outline

In this report, a design of building the road-event-prediction classifier with GSR and ECG data is presented in chapter 2. According to the design, chapter 3 shows the experiment to collected data for building the classifier, and chapter 4 describes the methods to preprocess and analysis data collected in the experiment. Finally, chapter 5 discusses the classification results with statistical evaluation, and chapter 6 draws the conclusion and makes a number of recommendations for improving this project in the future.
Design of building the
Road-event-prediction Classifier
with GSR and ECG data

To generate the road-event prediction classifier with GSR and ECG data, basic machine learning procedures should be followed with, which includes data preparation, train the classifier, validate the classifier. This chapter describes above three procedures in details so as to guide the experiment for chapter 3 and lay the foundation for the chapter 4 and chapter 5.

2.1 Data Preparation

Data preparation is a procedure to define the three main terms used in training the road-event-prediction classifier, which are the example, the features, and the label. The example is also called the instance. For each example, it might have several features but only have one label. In this project, each example is a 4 seconds interval which is segmented with 50% overlap from time sequence. The features are generated from GSR and ECG data. Since each example is a 4 seconds interval, some statistical measures can be calculated from corresponding GSR and ECG data as features, such as minimum value of this interval, maximum value of this interval, and mean value of this
interval [Sharma and Gedeon, 2014]. Corresponding label for each example is a number which represents a unique road event or a road event category.

### 2.2 Train the Classifier with Extreme Learning Machine

Basically, machine learning algorithm can be regarded as a box; The input of this box is the features for one example; The output of this black box is the predictive label. When there are a set of examples, it can be obtained a set of predictive labels through the machine learning algorithm. Comparing predictive labels with original labels, it can be found that some predictive labels are same as original labels. That is what we exactly desire: more correct predictive results. Thus, machine learning algorithm would do adjustment on its inner parameter setting to reach this goal. That is the basic concept of training the classifier.

Extreme learning machine is a kind of machine learning algorithm works for single-hidden-layer feedforward neural networks [Huang et al., 2006]. There are three layers in total: inutlayer, hidden layer, and output layer. The number of input neurons in input layer is equal to the number of features for each example. In hidden layer, the number of hidden neurons is selected with suggested number for ELM, such as 400 [Zhu, 2014]. The number of output neurons is equal to the number of event category. One of the advantages of ELM is that it is efficient in sequential learning. As the definition of the examples in this project, we find that training the road-event-prediction classifier is a kind of sequential learning related problem, and choose ELM algorithm to train the classifier would be a good choice.
2.3 Validate the Classifier with K-fold Cross Validation

2.3.1 Motivation of Validation

As our objective is to develop a predictive model which can predict road events based on GSR and ECG physiological data, we must have a way not only to select the best model but also to assess the accuracy, reliability and credibility of this model. More specifically, there are two fundamental problems we need to solve: model selection and performance estimation. Validation techniques are motivated by these two problems. In this chapter we briefly describe five different validation techniques, and present the way to solve those two problems in our project in details.

In model selection problem, almost invariably, all pattern recognition techniques have one or more free parameters, such as the number of hidden units or the number of hidden layers in Artificial Neuron Network. We concern about How we can select the optimal user defined parameters or learning algorithm for a given classification problem. In performance estimation problem, once we have chosen a model, we also have to consider how to estimate its performance. Normally, the performance of a model is measured by true error rate, which is the classifier’s error rate on the entire instances.

If it is possible that we can access to an infinite number of instances, we could choose the model, which provides the lowest error rate on the entire population and we could also estimate its performance by measuring the classifier’s error rate on the entire population, which is equal to the total number of misclassifications on the entire population divide by the size of entire population. However, in practice only we can get a limit set of instances but not
whole instances. If we still implement same method as before to select best model and estimate performance of this model, there would be two fundamental problems. One is that the final model we select will normally overfit the training data especially in a model that has amount of parameters. The other problem is that the error rate we have estimated will be lower than the true error rate [Efron and Gong 1983].

2.3.2 Cross Validation

There are typically two kinds of Cross validation methods: leave-one cross validation and k-fold cross validation. They are all a kind of resampling methods, which repeatedly use different split ways to train several models and test corresponding models [Kohavi et al. 1995]. According to the different split ways, different cross validation methods are named straightaway.

As for leave-one cross validation, assuming we have an initial data in N size, which also means that there are N instances, we would repeatedly train model for N times, and in t time we choose the t-th instance as testing set and the rest is regarded as training set. After N times training, we would get N error rates for all iterations and the overall model performance. The biggest advantage of leave-one-out cross validation is that it make good use of every instance in an initial data. As for drawback, leave-one-our cross validation is really expensive when an initial data has a huge size, that means we need numerous iterations, and sometimes it also has some weird behavior. For example, if there are minority instances in same class, when we select such instance in one iteration as testing set and the rest instances as training data, we might get 100% error rate since this instance cannot be classified correctly by such kind of model which is trained by majority instances in other class. In that case, the overall performance of the model on limit instances would
be not good as real performance of the model on whole instances.

In k-fold cross-validation, which is also called rotation estimation, we partition the initial data into k subsets in equal size. We would repeatedly train classifier in k times, and in t-th time we regard the t-th subset as testing set and the rest as training set. After k times training, we get k error rates for all iterations and the overall performance would be the average of the k error rates. A common choice for k-fold cross validation is to define k equal to 10. K-fold cross-validation is similar to random subsampling, but the former one have a big advantage since all the examples in initial data definitely can be used for both training and testing [Kohavi et al. 1995]. However, considering the attribute of random split, random subsampling cannot guarantee that.

Comparing the advantages and disadvantages of two cross validation methods, we choose k-fold cross validation method to validate the road-event-prediction classifier with setting k equal to 10. The classifier would be the model which have the best correct rate among others. In practical, the input weights and output weights are recorded as the model, which is used to prediction.

2.4 Architecture of building the Classifier

In generally, there are basically three main steps to build the road-event classifier. The first step is to get the examples, the corresponding features, and the corresponding tags ready. In chapter 4, it would introduce an experiment to collect video data and physiological data, which serve for chapter 5.2 Data preprocessing to obtain the examples, the corresponding features, and the corresponding tags. The second step is to build ELM algorithm with considering the attributes of examples’ features and the attributes of examples’
Design of building the Road-event-prediction Classifier with GSR and ECG data

Figure 2.1: The architecture of building the classifier

tags. The final step is to organize ELM algorithm into k-fold cross validation with recording the best model. Besides, according the objective of this project mentioned in chapter 1, we use the classifier to predict on new data and observing the results with comparing the subjects’ facial expression from video data. To sum up, the Figure 2.1 presents the architecture of building the classifier.

2.5 Summary

In this chapter, it illustrates the architecture of building the road-event-prediction classifier, which lead to the following experiment presented in chapter 3. According to the data preparation, the experiment aims at data collection for generating necessary elements of building classifier. This chapter also lay the foundation of data preprocessing in chapter 4 with the definition of the examples, the features, and the tags.
Chapter 3

Experiment

The experiment results of data collection for training and testing a road-event-prediction classifier are presented in this chapter. In chapter 1, it is introduced that some physiological data, such as GSR and ECG, which are highly responsive to the stressful events, can be taken into account for developing the safe driving alert system. To build the road-event-prediction classifier, which has been shown in chapter 2, an experiment on a driving simulator is conducted to collect data. The depth preprocessing and analysis of the collected data is presented in chapter 4. Ethics approval to perform the experiment was received from the ANU Human Research Ethics Committee.

3.1 Previous Work

The experiment introduced in this chapter is built upon previous experiments developed by Lor [Lor, 2013] and Le [Le, 2013]. Benefited from their works, a variety of problems about hardware and software are solved properly. Aiming at building a realistic left-hand driving experiment environment, Lor set up the hardware and software scheme. He has chosen a driving simulator software, named City Car Driving, from several different driving simulator software, within which the car type and traffic conditions can be modified as well. Besides, Lor also has successfully built a left-hand virtual driving environment by using mirroring software to flip the whole driving scene in
City Car Driving. With the aim to compare the participants’ physiological response in different degree of driving simulation, Le proposes a solution to integrate the collection of physiological data into driving simulator experiment developed by Lor. In Le’s work, she presents the details about collecting GSR and ECG data with specific physiological sensors, gives the guidance about setting up an experiment with maximizing the capacity for all equipment, such as driving simulator, physiological sensors, web camera, and projector, and mainly introduces the methods to pre-process the multiple data collected from different equipments.

In this experiment, we still use the hardware and software structure of driving simulator as well as the way to collect GSR and ECG data. However, as the aim of this experiment is aimed at collecting the GSR and ECG data with corresponding virtual world driving scene and is not aimed at comparing the participants’ response in different emulation as Le did, the survey questions, scenario, sequences, and procedure in this experiment are simplified. Besides, considering the difficulty Le meets in her work, which is the synchronization of physiological data and virtual world driving situation video recording data, some minor adjustments in experiment setup are used to reduce this difficulty.

### 3.2 Experiment Setup

Before experiment conduct, it is necessary to set up the hardware and software for driving simulator and physiological data collection. Survey questions also need to be well designed to serve for preprocessing physiological data in next chapter. Moreover, we also have to take into account the impact of experiment environment on bio-feedback, and choose an appropriate place to tick out the factors, which would divert participants’ attention from exper-
3.2 Experiment Setup

In this experiment, two personal computers are running synchronously. One is to control the hardware and software of physiological data collection and virtual world driving situation recording data collection as Figure 3.1(a) showing; the other is to implement the hardware and software of driving simulator so as to build a realistic left-hand driving experimental environment as Figure 3.1(b) showing.

On the data collection computer, besides mouse and keyboard there are three hardware devices, which need to be well connected. There are two web cameras, GSR and ECG sensors. The software for implementation, each web camera is Amcap, which can broadcast the scene captured by web camera in real-time. For these two web cameras, one is used to broadcast the participants’ facial expression; the other is to broadcast the participants’ virtual driving situation. Therefore, we put one web camera in front of participant,
and put the other web camera in front of the screen, which connects to the computer of implementing driving simulator. We use the proprietary software named NeuLog to implement NeuLog NUL217GSR and ECG sensors, and we set the sampling rate for both GSR and ECG sensors to 5 sampling points per second. In order to reduce the difficulty of synchronization process, which is addressed in Le’s work, we decide use the screen recorder named Screenrecorder to record a monitoring video as Figure 3.1(a) showing, which not only does save the participants’ facial expression recording and virtual world driving situation recording in one video but also can easily trace the corresponding facial expression and virtual world driving situation at a specific GSR and ECG value that the participant behaves.

When we implement driving simulator on another personal computer, for the driving visual simulator part, we use a 1050 x 1680 pixel DELL monitor to display the scene in the perspective of a driver as sitting in a car, which is provided by the software named City Car Driving; for the driving operation simulator part, we use Logitech G25 Racing wheel set to simulate car wheel, gear, brake, accelerator and necessary operation components in a real car. An auto-gear car with normal traffic condition is chosen in our virtual driving environment: City Car Driving.

### 3.2.2 Parameter Setting

There are two software for collecting data. One is NeuLog software that collects GSR and ECG data; The other is ScreenRecorder software that collects participants’ facial expression recording, virtual world driving situation recording, the real time GSR and ECG figure, and the real time digital clock as Figure 3.1(a) showing. A summary about parameter setting for each software is required. Table 3.1 shows the parameter setting for NeuLog software,
§3.2  Experiment Setup

<table>
<thead>
<tr>
<th>Experiment duration</th>
<th>10 minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sampling rate</td>
<td>5 per second</td>
</tr>
<tr>
<td>GSR units of measure</td>
<td>Arb</td>
</tr>
<tr>
<td>ECG units of measure</td>
<td>Arb</td>
</tr>
<tr>
<td>The file suffix</td>
<td>.csv</td>
</tr>
</tbody>
</table>

Table 3.1: NeuLog software experiment parameter setting.

<table>
<thead>
<tr>
<th>Experiment duration</th>
<th>at least 10 minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame rate</td>
<td>25 FPS</td>
</tr>
<tr>
<td>Sound Record</td>
<td>off</td>
</tr>
<tr>
<td>The file suffix</td>
<td>.avi</td>
</tr>
</tbody>
</table>

Table 3.2: ScreenRecorder software experiment parameter setting.

and table 3.2 presents the parameter setting for ScreenRecorder software.

3.2.3  Survey Questionnaire

The pre-experiment and post-experiment questionnaire are designed in our survey. Participants’ basic personal information, like name, age, and gender, is asked in pre-experiment questionnaire. Besides, considering the influence of real life driving experience and virtual driving experience, we also set several questions about it. Participants who have real life driving experience are more familiar with the procedure of driving, such as turning on the car engine, turning off hand break, shifting gear with pressing brake, and switching gear in right moment, so they would more focus on driving rather than spend time on learning basic procedure of driving. We prefer to use the data collected from experienced participants who are familiar with driving to build our road-even-prediction classifier. Post-experiments survey questionnaire asks participants some questions about the virtual driving experience. These feedbacks help us figure out whether participants take same attitude as when they are driving in real life world. We would select the data from those participants for further study. The pre- and post- experiment survey are give
3.3 Experiment Conduct

3.3.1 Experiment Subjects

There are eleven subjects who participant in this experiment. The subjects are chosen at random, and there are 6(54.5%) female and 5(45.5%) male. Their average was 23.18 years with the range from 12 to 35 years of age. Most of them do not have driving simulator experience but have driving experience. In chapter 5.2.1, it introduces one condition of selecting useful data for training and testing road-event-prediction classifier. That condition is depends on whether subjects have real life driving experience.

Basically, each subject has two tasks: file in survey questionnaire and play driving simulator. While subject is playing driving simulator, we record the GSR and ECG data, its facial expression change, and virtual world driving performance. Experiment instructor does not interrupt subject, only if the GSR and ECG sensor lost contact with subject. When the loexperiment instructor should recontact sensor to subject.

3.3.2 Experiment Procedure

Before conducting experiment, the experiment hardware and software need to be setup and placed at appropriate locations in the room. Experiment instructor has to test the experiment hardware and software in order to make every equipment work successfully. Besides, survey questionnaire sheet should be printed. Experiment instructor give briefly expiration of tasks for each subject need to do but do not tell them the purpose of this experiment until at the end of experiment. At the beginning of experience, written consent and pre-experiment questionnaire should be filled in. Let subjects seat on the chair.
The seat is adjusted to a comfortable distance from the driving simulator controller. Then experiment instructor attaches GSR and ECG sensors to subject according to Le’s report. After that, a brief introduction on how to control the vehicle by using driving simulator is given to subjects. Each subject can have a few minutes to drive around the virtual city to practice how to control the driving simulator. When the subject is ready to do experiment and experiment hardware and software are set properly, experiment instructor turn on recording software and let subjects start to experiment. Meanwhile, experiment instructor should observe the real time GSR and ECG figure on the monitor so as to take immediately measures when the temporary loss of GSR data happens due to GSR sensor lost contact with the electrodes. When the temporary loss of GSR data is found from the figure, experiment instructor should immediately attach GSR sensor to the electrodes. After 10 minutes past, GSR and ECG sensor stop recording automatically. Experiment instructor should stop screen recording immediately as well and take off the sensors from subjects’ body. At the end of experiment, subjects need to fill in the post-experiment questionnaire based on just driving simulator experience.

### 3.4 Collected Data

Finally, there are 11 sets of collected data. Each set includes two types data: numerical data and video data. Numerical data includes GSR and ECG data. Participants’ facial expression, virtual world driving situation, real time GSR and ECG graph, and digital clock are contained in video data. The data preprocessing and analysis would be presented in chapter 4.


3.5 Summary

The aim of this experiment is to collect the data for training and testing the road-events-prediction classifier. Firstly, we set up the driving simulator with the integration of physiological data collecting based on Lor and Le’s work, and then change the way to store the multiple data collected from different equipment. Secondly, survey questionary is designed to collect the basic personal information of participants and feedback for virtual driving experience. Meanwhile, the experiment conduct that anticipates the experiment process need to be designed with consideration the aim of our experiment, and emergency measure, such as reconnect the sensors on participants, should be taken to avoid the loss of data. Finally, we make our experiment with eleven participants and collect two data sets for each participant. One is the raw arbitrary GSR and ECG data; the other is the screen recording video data, which not only does save the participants’ facial expression recording and virtual world driving situation recording in one video but also can easily trace the corresponding facial expression and virtual world driving situation at a specific GSR and ECG value that a participant has.
Chapter 4

Data Preprocessing and Analysis

In this chapter, the methods of data preprocessing and analysis are presented in detail. The target data which are required to be processed is described in chapter 4.1. Then the data preprocessing methods are discussed in chapter 4.2. Besides, the implementation of training and testing the road-event-prediction classifier is introduced in chapter 4.3. In the end, the way to compare predicted results on new data with corresponding subjects’ facial expression change is described in chapter 4.4.

4.1 Data need to be Preprocessed

There are two groups of data need to be preprocessed. One is for training and testing the road-event-prediction classifier, so there are two kinds of data: video data and numerical data. Video data is used to label each example with event tag; numerical data includes GSR and ECG data and is used to generate several features for each example. The other group data is for comparison between facial expression change and the prediction of the road-event-prediction classifier with GSR and ECG data collected in last semester. Thus there is only numerical data including GSR and ECG data need to be preprocessed, but there is no need for event annotation. The same methods are applied to preprocess numerical data for both groups. In generally, there are two parts in data preprocessing: numerical data preprocessing and video
Data Preprocessing and Analysis

In numerical data preprocessing, there are two groups of data, which are processed separately but in same methods.

4.2 Data Preprocessing

4.2.1 Select Useful Subjects’ Data

At the beginning of data preprocessing, it is necessary to review the GSR and ECG figures for all subjects. We draw above figures in Matrix laboratory (MATLAB). After observing all figures, we find that only some subjects’ GSR data are keeping at maximum as Figure 4.1 showing. The reason behind this phenomenon is that the temperature of finger skin is too low, or that these subjects have excessive sweating on their fingers. As extremely sweating subjects’ GSR data do not fluctuate for a long period of time, we decide drop such subjects’ data. Among 11 subject’s data, we find there are 3 extremely sweating subjects. In the end, we keep 8 subjects’ data for further data preprocessing.

4.2.2 Clean GSR and ECG data

After obtaining the useful subjects’ data, we need to clean their GSR and ECG data. This process aims at dealing with the missing GSR data issue. In chapter 4, it has mentioned that experiment instructor should continually observe subjects’ real time GSR and ECG figure in order to take immediately measures when GSR or ECG data drop to 0. This phenomenon dues to that GSR or ECG sensors cables lost contact with electrodes.

In this experiment, it only happens on GSR data when subjects try to avoid incidents or accidents with extreme arm movements on steering wheel. With extreme arm movements on steering wheel, the GSR sensors’ cables sometime
tangle and stuck into the gap of steering wheel, but the GSR sensor electrodes are still on subjects’ finger. Thus, GSR sensors cables lost contact with electrodes and the real time GSR figure shows that GSR line drops dramatically to 0 as Figure 4.2 showing. Luckily, there are only short period of loss GSR data time since experiment instructor has taken immediately measures to connect GSR sensor cables to GSR sensor electrodes. Therefore, in this sector we only consider how to deal with missing GSR data.

The solution of dealing with missing GSR data is based on statistical method. Basically, there are three steps in this method. Firstly, the positions of missing GSR are required to be located precisely. We can roughly locate the missing GSR data positions by observing their figures plotted by MATLAB and find the precise positions by observing the data matrix loaded

Subject 7: The Arbitrary[Arb] GSR data

Subject 7: The Arbitrary[Arb] ECG data

Figure 4.1: The GSR and ECG data figure for subject 7
in MATLAB. The missing GSR data area would include a period of time that GSR data drops to 0 and increases to the pervious height. Secondly, the two sides of the neighbors of missing GSR data area are also required to be located properly. Finally, the missing GSR data is replaced with the mean of the two sides of the neighbors of missing GSR data.

Take subject 4 GSR data for example, we has seen the raw GSR data in Figure 4.2 and after processing this GSR data with above method in MATLAB the Figure 4.3 shows the processed results for subject 4 GSR data.

### 4.2.3 Filter and Normalize GSR and ECG Data

The aim of filtering GSR and ECG data is to eliminate the noisy. In this research, we only consider eliminating high frequency noisy that is created by sensors and cables. Nandita had discussed that the low frequency band of ECG would provide more useful information to indicate stress [Sharma and Gedeon, 2012]. How to eliminate the low frequency noisy but leave the useful information is not a easy problem. Limited in the time, we only consider how to eliminate high frequency noisy with a practical low-pass filter.
The low-pass filter implemented in this project is a simple moving average low-pass filter. In mathematically, this low-pass filter is a procedure to replace each original data with the mean value in sequentially. The mean value is calculated among this original data and several previous neighbors of this original data. Therefore, it would eliminate the extreme change to some extent. The formula of a simple moving average low-pass filter is presented in (4.1), where \( w \) is called window size and \( X_n \) is the current data. In this project, window size is set to 4.

\[
X_n = \frac{\sum_{i=0}^{w-1} X_{n-i}}{w} \tag{4.1}
\]

After filtering GSR and ECG data, normalization is required in order to eliminate the individual differences among subjects. It is obviously that different subjects have different attributes on their physiological data. Some subjects are more likely to stay calm but the others might be easy excited. These differences are also reflected on their physiological data. Thus, considering these individual differences, we apply normalization on GSR and ECG data. For each subject, we implement the formula (4.2) on both of GSR and ECG data.
to do normalization. In the formula, $X_n$ is the current data; $X_{\text{min}}$ and $X_{\text{max}}$ are the minimum data and maximum data among this subject’s data respectively. In the end, for all subjects their GSR and ECG data are both in same range of $[-1, 1]$.

$$X_n = \frac{2(X_n - X_{\text{min}})}{X_{\text{max}} - X_{\text{min}}} - 1$$  \hspace{1cm} (4.2)

### 4.2.4 Generate Features from GSR and ECG data

Based on the definitions of the examples and the features that are described in chapter 3.1, we generate features from GSR and ECG data for the road-event-prediction classifier. As mentioned in chapter 3.1, in this project, each example is a 4 seconds interval, which is segmented with 50% overlap sequentially. The features are generated from GSR and ECG data in each example. Thus, there are basically three main steps to generate features from GSR and ECG data.

Firstly, for each subject, the GSR and ECG data are required to be segmented into several examples with the rule mentioned above. In this project, as the sampling rate for both of GSR and ECG sensor is set to 5 sampling points per second and the experiment duration for each subject is 10 minutes, there are 3001 points in each GSR and ECG data including the base point. Since each example is a 4 seconds interval, there are 20 points in each example. The overlap is 50%, so the overlap segment is a 10 points segment. For each example, it is defined with 20 points from GSR data and 20 points from ECG data. The right neighbor of one example share 10 same points from GSR data and 10 same points from ECG data. In the end, there are 299 examples for each subject. Since there are 8 useful subjects’ data, there are 2392 examples in total.
Secondly, for each example, we generate 27 features in total. In each example, for GSR data, the gradient GSR data, and ECG data, each of them would be used to generate 8 statistical features: minimum, maximum, median, interquartile range, mean, standard deviation, variance, root mean square. Those features can be calculated with inbuilt function in MATLAB. Thus, there are 24 features for each example now. For ECG data, another three features are generated: energy ration, low frequency band \(\text{LF}\) [Sharma and Gedeon, 2012], and high frequency band \(\text{HF}\). LF is in the range of \([0.05\text{HZ}, 0.15\text{HZ}]\); HF is in the range of \([0.16\text{HZ}, 0.4\text{HZ}]\). Energy ration is the result from the total energy in LF divided by the total energy in HF. In the Table 4.1 we make a summary about feature generation type.

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Features</th>
<th>Target Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum, Maximum, Median, Interquartile range, Standard deviation, Mean, Root mean square</td>
<td>GSR</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>Gradient of GSR</td>
</tr>
<tr>
<td>Energy Ratio</td>
<td></td>
<td>ECG</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LF</td>
<td>ECG</td>
</tr>
<tr>
<td></td>
<td>HF</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1: Features generation types and quantity

Finally, we form a feature matrix with all examples for all subjects, where the size of feature matrix is 2391 by 27. Then the same normalization method described in formula (5.2) is applied to each feature in this matrix in order to eliminate the individual difference for features. The reason bind this normalization aims at limiting each feature in same range of \([-1, -1]\) and serving for further classifier training stage.
4.2.5 Label Video Sequence Data

The most important part in data preprocessing is video sequence data labeling. This procedure aims at labeling each example with proper event tag, which is a number. There are three main steps to achieve this goal: define event tags, synchronize video data and physiological data, and identify event in video data.

Firstly, we need to define every event tag. We refer to Le’s work to define event category but rearrange the event tag according to the degree of serious incident. In the Table 4.2, we show the definition of event tags.

<table>
<thead>
<tr>
<th>Tag</th>
<th>Event category</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Normal driving</td>
</tr>
<tr>
<td>1</td>
<td>Near hit stationary object</td>
</tr>
<tr>
<td>2</td>
<td>Hit stationary object</td>
</tr>
<tr>
<td>3</td>
<td>Near hit car</td>
</tr>
<tr>
<td>4</td>
<td>Hit car</td>
</tr>
<tr>
<td>5</td>
<td>Near hit pedestrian</td>
</tr>
<tr>
<td>6</td>
<td>Hit pedestrian</td>
</tr>
</tbody>
</table>

Table 4.2: Event category and corresponding event tag for seven categories.

Secondly, video data and numerical physiological data are required to be synchronized. This procedure has to match start time of video data with start time of physiological data. Considering the delay of physiological sensor, the true physiological data, which is reflect the corresponding state showing in video data, is a little bit behind the current video data. Thus, when we observing the video data, after we find the GSR and ECG figure start to show the data, we record the digital clock time as the start time. This start time is
regarded as the base point in GSR and ECG data. Besides, each example in video sequence data is defined by a 2 second interval with 0% overlap. By this mean, the delay of philological data problem can be relaxed. Finally, we build a table in excel, where the first column is the timeline of physiological data, and the second column is the timeline of the digital clock time line showing in video data. As playing the video, we observe the subject’ driving performance in virtual world, make a judgment on the type of event category, and fill the tag number in the third column in the excel table. This procedure is important, because how correct you label the event would lead to how reliable the classifier will be.

4.3 Train and Test the Classifier with Data

Before train and test the classifier, two kinds of data would be prepared: feature matrix and corresponding target label vector. In this project, we have generated the feature matrix, where each row represents each example and each column represents each feature. The size of the feature matrix we obtained in this project is 2392 by 27. That means there are 2392 examples and 27 features for each example. The corresponding target label vector is obtained in previous section, whose size is 2392 by 1. When these two kinds of data are ready, we put them as input of ELM function, which is provided as an open source in MATLAB. Besides, we set the hidden neuron number as 400 and choose sigmod function as active function. This ELM function in MATLAB would provide the prediction results. We modify this function in order to obtain the input weight and output weight. The input weight and output weight would be the necessary model parameters. The road-event-classifier would use these specific model parameters. K-fold cross validation is the main structure of training and testing the classifier. The detailed code for training and testing the classifier is presented in Appendix.
4.4 Predict Events with the Classifier

As obtained the road-event-prediction classifier, now we can use this classifier to predict the events on new data, which was collected in the last semester. We use same data preprocessing method to acquire the final feature matrix. There are 30 subject’s data in total, but there are only 26 useful subjects’ data. The sampling rate of GSR and ECG data is 10HZ, for each subject the experiment time is 5 minutes. Then use the classifier to predict the event tags. Based on the predicted event tags, we play the video and observe corresponding subjects’ facial expression, to see whether subjects’ facial expression change when the predicted serious incident occurs. It would be the easiest way to make a comparison between predicted events and subjects’ facial expression recorded in the video that when we open a video player to play the video, we could put the bar chart, which is plotted based on the predict result, under the bottom of play bar and adjust the bar chart length as same as the play bar.

The Figure 4.4 is the screenshot about how to make a comparison based on human observation.

4.5 Summary

Data preprocessing methods have been discussed in this chapter. The road-event-prediction classifier is trained with ELM algorithm and is tested by k-fold cross validation approach. The way to compare the predicted results on new data with subjects’ facial expression change is presented.
Figure 4.4: The comparison between prediction results and facial expression change
Chapter 5

Results and Discussion

This chapter shows the road-event-prediction results based on the classifier built and validated with the methods described in chapter 3 with the data obtained from the experiment presented in chapter 2. Additionally, this chapter also investigates the factors that affect the performance of road-event-prediction classifier. The relationship between road event occurrence and subject’ facial expression change is also discussed.

5.1 Objectives

This project aims to build a road-event-prediction classifier by using two kinds of physiological data: GSR data and ECG data collected from young students while they are simulating driving in city with normal traffic condition. As in chapter 1.3, three hypotheses have been proposed to support building a road-event-prediction classifier. Thus, the result analysis focuses on answer the following questions that are corresponds to these three hypotheses:

1) Could subjects’ physiological data such as GSR and ECG, which is monitored while they are driving in the virtual world, indicate the occurrences of corresponding road events?

2) What kinds of factors would affect the road-event-prediction classifier performance?
3) Is there any correlation between the occurrence of serious incidents and facial expression change?

The first and second questions would be discussed in chapter 6.2 with statistical evaluation. The last question would be discussed in chapter 6.3 with human observation.

### 5.2 Statistical Evaluation of Classification results and discussion

Before answering the first question listed in chapter 6.1, the classification results should be presented, because the better classification results obtained by the road-event-prediction classifier trained with GSR and ECG datasets implies the higher correlations between physiological data and road events. The classification results are measured with accuracy (correct rate), recall, precise, and F1-score, which are calculated from the counting table.

The classification results of predicting seven categories as the Table 4.2 showing reach to 67.2% in average accuracy, 91.5% in average recall, 75.8% in average precise, and 82.9% in average F1-score which are calculated from the counting Table 5.1. It seems like the performance of this road-event-prediction classifier is not bad since about two-thirds of prediction are correct. However, after observing the counting tables carefully, it has been found that a certain correct driving event category prediction contributes much more than another correct driving event predictions. Therefore, even though the rest of correct predictions for another event categories are really bad, the average accuracy can reach a considerable number.
Results and Discussion

<table>
<thead>
<tr>
<th>True value</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction</td>
<td>78</td>
<td>73</td>
<td>42</td>
<td>14</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>results</td>
<td>64</td>
<td>143</td>
<td>4</td>
<td>62</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>45</td>
<td>65</td>
<td>0</td>
<td>71</td>
<td>23</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
<td>29</td>
<td>23</td>
<td>22</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5.2: Counting table for six categories.

<table>
<thead>
<tr>
<th>True value</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction</td>
<td>0</td>
<td>1456</td>
<td>124</td>
<td>110</td>
<td>178</td>
<td>48</td>
<td>4</td>
</tr>
<tr>
<td>results</td>
<td>30</td>
<td>22</td>
<td>18</td>
<td>23</td>
<td>4</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>36</td>
<td>24</td>
<td>32</td>
<td>18</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>63</td>
<td>30</td>
<td>26</td>
<td>87</td>
<td>8</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5.1: Counting table for seven categories.

Take the counting Table 5.1 for example, as this counting table is built with seven categories and there are majority of examples in normal driving category, we try to use the rest of six event categories data to build a classifier. In the end, we obtain a bad classifier results: 39.3% in average accuracy, 39.0% in average recall, 37.0% in average precise, and 38.0% in average F1-score, which are calculated from the counting Table 5.2.

It shows that imbalanced datasets would affect the classifier performance. This issue occurs when the number of examples, which represent same class, is much lower than the ones of the other classes [Mazurowski et al., 2008]. In the driving simulator experiment, there are many examples that repre-
sent normal driving period, but there are a bit of examples that represent hit pedestrian. Thus, the classifier trained with imbalanced datasets can well distinguish the class who has amount of examples but cannot well distinguish the class who has a bit of examples. Even the classifier can have high average accuracy or F1 score for all classes but it cannot reach high accuracy for every class. As the consequence, in the future how to deal with imbalanced data as for building a road-event-prediction classifier needs to be concerned. If we were not considering the impact of imbalanced distribution and only estimate classifier performance with average measures, the evaluation results would be highly misleading.

In addition to considering the impact of imbalanced datasets on the classifier performance, we also want to see whether the different division of events can affect the classifier performance since some events share similar attributes. As the consequence, we compare the different division of events classification results. We considering some events are related with same objects, so we divide all events into four categories: normal driving, other objects related, car related, and pedestrian related. For each category, we define it with several sub-event categories, which is referred to the Table 4.2. Thus, we have the event category and corresponding event tag for four categories as the Table 5.3 showing. What’s more, based on the driving experience there are many incident more than accident. Incident is kind of event which does not lead to terrible crush, such as near hit car or near hit pedestrian. Accident is not like incident, and it brings serious results, such as hit pedestrian. Therefore, we also can divide all events into three categories: nothing, incident, and accident. Furthermore, considering how serious of an events, we also can divide all events into two categories: minor incident, and serious incident.
Results and Discussion

<table>
<thead>
<tr>
<th>Tag</th>
<th>Event category</th>
<th>Sub-Event category</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Normal driving</td>
<td>Normal driving</td>
</tr>
<tr>
<td>1</td>
<td>Other objects</td>
<td>Near hit stationary object</td>
</tr>
<tr>
<td></td>
<td>related</td>
<td>Hit stationary object</td>
</tr>
<tr>
<td>2</td>
<td>Car related</td>
<td>Near hit car</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hit car</td>
</tr>
<tr>
<td>3</td>
<td>Pedestrian related</td>
<td>Near hit pedestrian</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hit pedestrian</td>
</tr>
</tbody>
</table>

Table 5.3: Event category and corresponding event tag for four categories.

<table>
<thead>
<tr>
<th>True tag</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction results</td>
<td>0</td>
<td>1443</td>
<td>188</td>
<td>203</td>
</tr>
<tr>
<td>1</td>
<td>90</td>
<td>136</td>
<td>74</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>58</td>
<td>73</td>
<td>113</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5.4: Counting table for four categories.

The classification results of predicting four categories as the Table 5.3 showing reach to 70.7% in average accuracy, 90.7% in average recall, 78.5% in average precise, and 83.7% in average F1-score. From seven categories to four categories there is an increase in the average accuracy of 3.5%, in the average recall of 3.5%.
§5.2  Statistical Evaluation of Classification results and discussion

<table>
<thead>
<tr>
<th>Tag</th>
<th>Event category</th>
<th>Sub-Event category</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Nothing</td>
<td>Normal driving</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Near hit stationary object</td>
</tr>
<tr>
<td>1</td>
<td>Incident</td>
<td>Hit stationary object</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Near hit car</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Near hit pedestrian</td>
</tr>
<tr>
<td>2</td>
<td>Accident</td>
<td>Hit car</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hit pedestrian</td>
</tr>
</tbody>
</table>

Table 5.5: Event category and corresponding event tag for three categories.

<table>
<thead>
<tr>
<th>True tag</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction</td>
<td>1644</td>
<td>338</td>
<td>53</td>
</tr>
<tr>
<td>results</td>
<td>139</td>
<td>169</td>
<td>19</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>15</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 5.6: Counting table for three categories.

The classification results of predicting three categories as the Table 5.5 showing reach to 76.1% in average accuracy, 91.8% in average recall, 80.8% in average precise, and 85.9% in average F1-score. From four categories to three categories there is an increase in the average accuracy of 5.4%.
Results and Discussion

<table>
<thead>
<tr>
<th>Tag</th>
<th>Event category</th>
<th>Sub-Event category</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Minor incident</td>
<td>Normal driving</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Near hit stationary object</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hit stationary object</td>
</tr>
<tr>
<td>1</td>
<td>serious incident</td>
<td>Near hit car</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hit car</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Near hit pedestrian</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hit pedestrian</td>
</tr>
</tbody>
</table>

Table 5.7: Event category and corresponding event tag for two categories.

<table>
<thead>
<tr>
<th>True tag</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction</td>
<td>1867</td>
<td>295</td>
</tr>
<tr>
<td>results</td>
<td>122</td>
<td>108</td>
</tr>
</tbody>
</table>

Table 5.8: Counting table for two categories.

The classification results of predicting two categories as the Table 5.7 showing reach to 82.6% in average accuracy, 93.9% in average recall, 86.4% in average precise, and 90.0% in average F1-score. From three categories to two categories there is an increase in the average accuracy of 6.5%.

On one hand, the increase of the average accuracy indicates that it is much more practical to distinguish general classes where some events are sharing similar features and are considered belonging to same class than to distinguish every event precisely. On the other hand, to some extent the GSR and ECG datasets can only provide limited features to describe a unique event. Thus, it leads to that it is hard to distinguish every event precisely and that the classifier performance is not good enough. In other words, the division
of events into several categories is caused by that the GSR and ECG datasets can only provide limited features to describe a unique event leads to that it is hard to distinguish every event precisely and that the classifier performance is not good enough, but it is practical to distinguish general classes where some events are sharing similar features and are considered belonging to same class. The classifier gets better performance with proper division of events into several different categories.

### 5.2.1 The comparison between prediction results and facial expression change

When we make a comparison as the way described in chapter 4.4, it has been found that for most subjects they have reactions on their face when a serious incident happens, which is predicted by the two categories classifier for distinguishing minor incident and serious incident. Minor incident category includes three events: Normal driving, near hit stationary object, and hit stationary object; Serious incident category includes four events: Near hit car, hit car, near hit pedestrian, and hit pedestrian. To some extent, there is actually correlation between the occurrence of serious incidents and facial expression change.

### 5.2.2 Summary

To sum up, from the statistical evaluation of the classification results it can be concluded that subjects’ physiological data have the ability to indicate the occurrences of corresponding road events. The performance of the road-event-prediction classifier not only suffers from inappropriate division of events into several categories, but also suffers from imbalanced class distributions in datasets. There is correlation between the occurrence of serious incidents and facial expression change.
Conclusion and Future Work

This chapter reviews the main findings and results of the research. Besides, it draws a main conclusion and links to the future work heading for making up the limitations of this research.

6.1 Conclusion

The aim of this report is to build a road-events-prediction classifier by using physiological data collected from young participants while they are using a driving simulator. The classifier should tell the periods of normal driving from the periods of different kinds of events while participants are driving in virtual world, such as hit a pedestrian, a car or some stationary object, the nearly-hitting events should also be included.

A driving simulator experiment has been designed and conducted to collect 11 young participants’ physiological datasets and video datasets while they undertake a task of using simulator to mock driving in the city. Physiological datasets include galvanic skin response and electrocardiographs data; video datasets contain facial expression recording, the virtual world driving performance recording, and a real time clock. The physiological datasets are used to generate 27 features for every 4 seconds interval with 50% overlap; the video datasets are observed by human to annotate proper event name for
every 2 seconds interval without overlap. The most important work before training the classifier is events annotation, which also the most time consuming works. Since ELM algorithm is much more efficient than ANN, ELM algorithm is used to train the classifier. K-fold cross validation is applied to evaluate classifier performance so as to make use of all datasets.

After analyzing and observing the classification results, the author found that the classifier performance not only suffers from inappropriate division of events into several categories, but also suffers from imbalanced class distributions in datasets. The division of events into several categories is caused by that the GSR and ECG datasets can only provide limited features to describe a unique event leads to that it is hard to distinguish every event precisely and that the classifier performance is not good enough, but it is practical to distinguish general classes where some events share similar features and are considered belonging to same class, such as all events including normal driving can be subsumed under three broad categories: nothing, incident, and accident. Nothing category includes normal driving and near hit stationary object; Incident category includes hit stationary object, near hit car, and near hit pedestrian; Accident category includes hit car and hit pedestrian. The classifier gets better performance with proper division of events into several different categories. In other words, Subjects’ physiological data such as GSR and ECG, which is monitored while they are driving in the virtual world, can better indicate the occurrences of corresponding general road events category; if it was required to better indicate each specific road event for development of a safe driving alert system based on monitoring driver’s physiological data in the future, the lack of features to describe each unique road event should be considered.

Imbalanced class distributions affect the classifier performance. This issue
occurs when the number of examples, which represent same class, is much lower than the ones of the other classes. In the driving simulator experiment, there are many examples that represent normal driving period, but there are only a bit of examples that represent hitting pedestrian. Thus, the classifier trained with imbalanced datasets can well distinguish the class who has amount of examples but cannot well distinguish the class who has a bit of examples. Even the classifier can have high average accuracy or F1 score for all classes but it cannot reach high accuracy for every class. As the consequence, in the future how to deal with imbalanced data as for building a road-event-prediction classifier has to be concerned.

Besides, most subjects have reactions on their face when a serious incident happens, which is predicted by the two categories classifier for distinguishing minor incident and serious incident. Minor incident category includes three events: Normal driving, near hit stationary object, and hit stationary object; Serious incident category includes four events: Near hit car, hit car, near hit pedestrian, and hit pedestrian.

To sum up, the research presented in this report contributes to reveal the relationship between drivers’ physiological data monitored while they are driving and corresponding road events which they meet while driving, and lays foundation of a road-event-prediction classifier building for the further development of a safe driving alert system based on monitoring driver’s physiological data.

### 6.2 Future Work

This project is limited in width of finding sufficient or significant features in other physiological or physical datasets to describe a unique event and is
limited in depth of exploring specific approaches to cope with imbalanced datasets when building a road-event-prediction classifier. In addition, the lack of an integrated event definition leads to hesitation of labeling event tags. Sometimes, personal subjective judgment would affect the event annotations results. The last but not the least, the driving simulator used in this experiment is limited in simulating real world driving.

As the consequence, limitations above should be concerned in the future work. In the light of solving the first limitations, it is recommended that the combination of other physiological data like brain activity or physical data such as eye gaze could be adopted to provide more sufficient features to describe a unique event and to help draw more clear demarcation line between two closer events. For the limitation of building the classifier with imbalanced datasets, related literature survey is needed in order to review the main issues of this problem and find some proper approaches to deal with imbalance. The investigation of traffic related research is also necessary for the further study so as to get an integrated event definition and eliminate the hesitation of event annotation. Besides, the consistency about labelling all examples with event tags based on one specific person’s subjective judgment also contributes to classifier performance. Finally, if it is possible to improve the driving simulator performance in hardware and software, the experiment would obtain more meaningful data than the data collected in this project because such kind of data are highly correlated with that in real driving.
Appendix A

Participant Consent Form

WRITTEN CONSENT for Participants
Investigation of natural computer interfaces

I have read and understood the Information Sheet you have given me about the research project, and I have had any questions and concerns about the project (listed here)

addressed to my satisfaction. I agree to participate in the project. YES ☑ NO ☐

I agree to this interview being audio-recorded YES ☑ NO ☐

I agree to be identified in the following way within research outputs:

Full name YES ☑ NO ☐
Pseudonym YES ☑ NO ☐
No attribution YES ☑ NO ☐

Signature: __________________________

Figure A.1: Participant Consent Form sample
Appendix B

Survey Questionnaire

Volunteer Participants Info:

Name (optional):
Gender: Male/Female
Age: 16–24 / 24–40 / 40–60

Driving Experience:
   How many km do you drive each year (roughly)?
   How many years/months have you been driving?
   How often do you drive?

Driving Simulator Experience:
   How many times have you used a driving simulator?
   How often have you used a driving simulator?

Road Accidents:
   Have you ever been involved in road accidents as the driver? Yes / No
   If yes, please estimate how many in total:

Ethics issues:
Dear volunteer participants, when you are playing with the driving simulator, we will record your facial expression and screen view, the ask you some questions after you finish using the driving simulator for our research.

Do you agree for us to record your facial expressions and screen view? Yes/No

Figure B.1: Survey Questionnaire part 1
Training:
Since our driving simulator has some bugs we haven’t solved yet, it might not be easy to control when you want to turn around or accelerate your car. Please be gentle to your steering wheel and accelerator.

Pretend its for real:
Now, please imagine that you are sitting in a real car. The people walking around on the street you see on screen are real; the cars moving around your car are also real. We don’t want to have any car accidents. Now, please drive, following the traffic rules you are used to in Canberra …

Record Time:
Start: End:

Figure B.2: Survey Questionnaire part 2

Questionnaire:
1. Do you think this driving simulator made you feel like driving a real car? If not, please tell us where made you feel it was not like driving a real car?

2. When a car accident happened, did you feel scared?

3. When a car accident started to happened, did you take immediate action to avoid it? Did you succeed?

Figure B.3: Survey Questionnaire part 3
Appendix C

Build the road-event-prediction classifier in Matlab

```matlab
%% Set training set and testing set into one data matrix
data = getTrainingAndTestingData();
% Change the label category definition
classes = data(:,1);
classes = classes * -1.0; classes(classes == -1) = 0;
classes(classes == -2) = 0; classes(classes == -3) = 1;
classes(classes == -4) = 1; classes(classes == -5) = 1;
classes(classes == -6) = 1;
% Change the class number into the required range for ELM
classes = classes + 1;
dataFeatures = data(:,2:size(data,2));
data_ELM = [classes, dataFeatures];
% Select the classes which is larger than 'classesThreshold'
classesThreshold = 0;
data_ELM = data_ELM(data_ELM(:,1) > classesThreshold,:);
% initializes the CP object with whole instances target labels
cp = classperf(data_ELM(:,1) - classesThreshold);
```

Figure C.1: Build the road-event-prediction classifier in Matlab part 1.
% Set parameters for Extreme Learning machine
Elm_Type = 1; % 1 means classification; 0 means regression
NumberOfHiddenNeurons = 350; % the number of hidden Neurons
ActivationFunction = 'sig'; % the activation function for training/testing/predicting neuron network model.

% Set parameters for k-fold validation
N = length(data_ELM); % The number of whole instances without separation for train and test
indices = crossvalind('Kfold',N,10);
bestTestingAccuracy = 0;
for i = 1:10

% There are 15 examples as testing set
test = (indices == i); train = ~test;
% so there would 15 classifier result for testing set
[TrainingTime, TestingTime, ...
  TrainingAccuracy, TestingAccuracy, ...
  training_label_index_actual, testing_label_index_actual, ...
  InputWeight, OutputWeight, BiasofHiddenNeurons] = ELM(data_ELM(train,:),
  , data_ELM(test,:),...

  Elm_Type,
  NumberOfHiddenNeurons
  ,

  ActivationFunction);

% trace the best model which have the highest testing accuracy
if bestTestingAccuracy < TestingAccuracy
    bestTestingAccuracy = TestingAccuracy;
    bestInputWeight = InputWeight;
    bestOutputWeight = OutputWeight;
    bestBiasofHiddenNeurons = BiasofHiddenNeurons;
end
% updates the CP object with the current classification results
classperf(cp,testing_label_index_actual,test);
end
score = cp.CorrectRate % queries for the correct classification rate
CountingMatrix = cp.CountingMatrix ;

Figure C.2: Build the road-event-prediction classifier in Matlab part 2.
%% Predicting with the best model which have the best testing accuracy
bestInputWeight = InputWeight;
bestOutputWeight = OutputWeight;
bestBiasofHiddenNeurons = BiasofHiddenNeurons;

PredictingData = getPredictingData();

[PredictingTime,...
predicting_label_index_actual] = Predicting_ELM( PredictingData, ActivationFunction,...
                                          InputWeight, OutputWeight
                                          ,...
                                          BiasofHiddenNeurons);

predicting_labels = reshape(predicting_label_index_actual,[299,26]);

Figure C.3: Build the road-event-prediction classifier in Matlab part 3.
Appendix D

Independent Study Contract

Project title:
Predicting events from physiological data while driving

Learning objectives:
1) Experience with GSR and ECG data
2) Experience with pre processing GSR and ECG data
3) Experience with AI prediction

Project description:
1) Brief literature survey of use of physiological data while driving
2) Data processing of video labels, GSR and ECG
3) Train one or more classifiers using above data
4) Compare predictions on new data with human evaluation of new data
5) Statistical evaluation of results
6) Write report
Figure D.1: Independent study contract part 1
### ASSESSMENT (as per course’s project rules web page, with the differences noted below):

<table>
<thead>
<tr>
<th>Assessed project components</th>
<th>% of mark</th>
<th>Due date</th>
<th>Evaluated by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Report: name style...</td>
<td>45%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(e.g. research report, software description...)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Artifact: name kind...</td>
<td>45%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(e.g. software, user interface, robot...)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presentation</td>
<td>(10%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### MEETING DATES (IF KNOWN):

- Weekly

### STUDENT DECLARATION: I agree to fulfill the above defined contract:

Student Signature: [Signature]
Date: [Date]

### SECTION B (Supervisor):

I am willing to supervise and support this project. I have checked the student’s academic record and believe this student can complete the project.

Supervisor Signature: [Signature]
Date: [Date]

### REQUIRED DEPARTMENT RESOURCES:

- Access to datasets
- Access to driving simulator

### SECTION C (Course coordinator approval)

Course Coordinator Signature: [Signature]
Date: [Date]

### SECTION D (Projects coordinator approval)

Projects Coordinator Signature: [Signature]
Date: [Date]

---

Research School of Computer Science
Form updated Jun-12

---

Figure D.2: Independent study contract part 2


GREEN, M., 2000. "how long does it take to stop?" methodological analysis


