國立臺灣大學電機資訊學院資訊工程學研究所
碩士論文
Department of Computer Science and Information Engineering
College of Electrical Engineering and Computer Science
National Taiwan University
Master Thesis

使用智慧型手機上加速器辨識交通工具

Transportation Mode Detection with Single Accelerometer on Smart Phones

張育誠
Yu-Chen Chang

指導教授：薛智文 博士 和 許永真 博士
Advisor: Steven Chih-Wen Hsueh, Ph.D. and Jane Yung-Jen Hsu, Ph.D.

中華民國九十九年一月
January, 2010

國立臺灣大學碩士學位論文
口試委員會審定書

使用智慧型手機上加速器辨識交通工具

Transportation Mode Detection with Single Accelerometer on Smart Phones

本論文係張育誠君（學號R96922147）在國立臺灣大學資訊工程學系完成之碩士學位論文，於民國 99 年 1 月 28 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

薛智文
（指導教授）

陳彥仰　　　許永真

系主任　呂育道

Acknowledgments

要謝的人太多了，那就謝天吧！

首先要感謝我的指導教授許永真老師，在這種情況下願意收留我，讓我可以有機會朝著自己的興趣的領域學習研究的方法；除了教導我作研究的方法，在面對各種事情的處理上，也讓我學習到許多做人處世的道理，在未來待人接物上能有所借鏡。感謝薛智文教授指導我嚴謹的實驗方法，必須時時刻刻追問自己，使用的方法是否合理？處理資料的工具是否適宜？解讀數據的手段是否洽當？不能只是因為前人研究採用此種方法，就跟著沿用下去。光是得到結果是不夠的，必須探討各個數字背後的物理意義。諸多有用的建議幫助我反省自己的缺失。感謝台大陳彥仰教授擔任我的口試委員，於口試過程中給我許多重要的建議和改進事項。

感謝婉容學姊長久以來的幫忙，無論是在學術研究上的問題，或是一般事務的處理都給我很多幫助。感謝皓遠在課程上交流機器學習的心得，引領我進入用機器學習的方法來做動作辨識的領域。感謝于晉花時間和我討論加速器和直方圖處理的問題，讓我省去許多閱讀前人研究的時間。感謝翰文願意花時間教我許多知識和如何使用CRF++這項工具，讓我不用從頭開始摸索。感謝文傑、能豪、培堯、人豪、鶴齡、知本、麗珊等眾多學長姐們在事情發生時，鼓

勵和開導我從低潮中走出來。感謝眾多願意幫我收集資料的同學、學長和朋友們，沒有你們的熱心，這個研究就沒辦法完成。

最後，感謝我的父母，無論在物質或精神上，都給我最大的支持，讓我自由選擇做我想做的事，而無後顧之憂，而能順利完成此研究。

Abstract

Learning user mobility from sensor embedded in portable everyday object is a dominant research area in pervasive computing. As a kind of human activity, transportation modes,such as walking, cycling, riding, driving, and etc., can provide more knowledge for mobility understanding. This thesis explores how single low-level acceleromter data from smart phones can be used to recognize high-level properties of user transportation. Our method considers both commonsense constraint of transportation infrastructure and regular user behavior on carrying mobile phone. With de-orientation and relabeling, we constructed the vibration, which was caused by both user action and vehicle motion, layer and extracted discriminable pattern from it for transportation mode inference. Evaluated with 831 user-labeled trails from the daily lives of 17 data collectors over a period of one month, our system got an overall average accuracy of 89% for trail-based analysis and 78% for window-based analysis on 6 kinds of transportation in urban city.

Keywords: Mobile Phone, Accelerometer, Context Sensing, Activity Recognition, Transportation Mode

摘要

從嵌入於攜帶型日常生活物品的感應器中，得知使用者移動行為在普及計算領域中是一個顯學。作為一種人類活動，交通模式，例如走路、騎腳踏車、騎機車、開車等，可以提供知識，用以幫助了解移動行為。這篇碩士論文探索如何用單一、低性能、裝置於手機上的加速器資料來辨識高複雜度使用者交通模式的屬性。我們的方法同時考慮了一般對於交通工具所在基礎建設的常識，以及一般使用者攜帶手機的習慣。藉由消除角度特型和重新標記的技術，我們建立了因使用者行為和交通工具移動導致的震動這層資料，然後從中擷取可辨識的特徵，來進行交通模式辨認。在十七個使用者在至少一個月的日常生活中，從標記的八百三十一筆移動軌跡驗證，整體而言，在辨識都市地區常見的六種交通模式，我們的系統在以整趟移動為樣本的事後分析獲得平均89%的準確度，在以時間視窗為樣本的即時偵測中獲得平均78%的準確度。

關鍵字: 手機、加速器、情境感知、行為辨識、交通模式

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this chapter, we first present the motivation and basic idea for transportation mode detection. Second, several challenges from previous researches are described for the difficulty to solve the problem. Finally, we brief the organization of the entire thesis.

1.1 Motivation

Learning user mobility from sensor embedded in portable everyday object is a dominant research area in pervasive computing. As a kind of human activity, transportation modes, such as walking, cycling, riding, driving, and etc., can provide more knowledge for mobility understanding. In the past few years, more and more researches work on applications supported with this context-sensing technique in various field, including green issues, personal information, geographic applications. For green issue, researchers [12][16] aware people and persuade them into green transportation behaviors in order to lessen their impact on the environment. To provide personal information, transportation mode classification is used to support modeling user behavior and calculate personalized estimates of environmental impact and exposure [22][27]. Also, aided with transportation mode recognition, geographic applications [3][18][21][23][24][40] build closer connections between locality and mobility.

However, the existing approaches based on both GPS/GSM coarse-grained data and wearable sensor-based fine-grained data could be improved in accuracy, range and user load. On one hand, GPS/GSM approaches [22][33][39][40] performs weak on transportation with low speed, such as bicycle. Also, this approach is out of work in the place without radio signal, like metro underground. On the other hand, wearable sensor-based approaches [9][31] provides wider range and better accuracy but extra devices discourage people from using. It's hard to convince people to wear or carry extra device in daily life. Thus, we seek the approach to accurately detect the transportation mode by users' own object.

1.2 Research Objectives

Context sensing on transportation mode is a good motivation and can help researcher to build application on it. To achieve this goal, we want to design and implement the system embedded in object which people bring with them during transportation. By analyzing the data from it, the transportation mode can be real-time inferenced and utilized by context-aware applications.

1.2.1 Transportation Mode Detection by Everyday Object: Mobile phone

On considering what people carry with them during transportation, mobile phone is a good choice since everyone have it. In recent years, the mobile phones with accelerometer graduately becomes prevalent in market where account for 14% of the overall mobile device sales in 2009 and are supposed to make up around 37% by 2012 [29]. Modern people bring the cell phone with them every day so it won't require human effort to bring extra equipments. As coarse-grained data, single accelerometer data is more challenging than GPS/GSM methods. The user behavior will affect the accelerometer reading while GPS/GSM methods are dependent from users. For example, how and where user place acceleromter will influence the orientation of accelerometer. To mining from these coarse-grained data, we first did de-orientation and relabeling. Then, we constructed the vibration, which was caused by both user action and vehicle motion, layer and extracted discriminable pattern. Finally, we built SVM classifiers based on these features for transportation mode inference.

1.2.2 Transportation Transit Recognition

Another goal is recognizing when people tranit the transportation. During the change of transportation, the experimenters tag the transportation mode before or after leaving the vehicle. The action won't be made immediately right on the moment of changing. The advance or the delay depends on the user condition and the transportation circumstance. For example, user on ,who hurries to get out of the crowdy bus when reaching his destination, might label the data a few minutes later. However, transportation transit is also an importation information which not only can seperate two transportation but also an useful context itself.

1.3 Challenges

To detect the transportation mode with single accelerometer on the mobile phone, there are two critical challenges. As a everyday object, the mobile phone is limited to make some strong assumptions in the experiment. Also, the supervised learning performs poorly on the original label and raw data from accelerometer.

1.3.1 Limitation on Everyday Object

Detecting with a everyday object will face many problems caused by the user behavior. Unlike strict-controlled experiments in the laboratory, there are many unpredictable incidents outbreaking in the natural environment. Asking users to change their habits and adapt to our device is not reasonable. Therefore, the common sense of human interaction with machine should be considered for designing the sensing and feedback from the device. Some strong assumptions appearing in the laboratory are not allowed in the real life, such as binding the mobile phone to the body for accelerometer orientation problem. These constraints could not be traded off due to the technical limitations.

1.3.2 Feature Selection and Relabeling

Each transportation is not an instant event but a series of human action and vehicle motion. The fact that people won't be motionless leads to more noisy and confused data to distinguish. Based on previous work, it's hard to recognize by the coarse-grained raw data from sensor, like acceleromter, GPS, GSM, and et al. Also, how much data we should collect for each transportation detection is another problem. The least certain collecting period decides how soon the system can inference and give feedback to user. Only if it is short enough, the system could be made to be real-time. How to select feature and relabel data is a key to transportation detection.

1.4 Thesis Organization

The thesis is organized as follows. In the beginning, we review related work on everyday object and transportation activity classification in Chapter 2. Secondly, the limitations on everyday object and using scenerio are introduced in Chapter 3. Thirdly, we present our transportation classification methodology in Chapter 4. Fourthly, the experiments are conducted to verify the feasibility and evaluate the performance of our approach to transportation classification for trail-based analysis and window-based analysis in Chapter 5. Finally, we conclude the contribution, describe the limitation for deploying in real world, and explore some possible direction for future work in Chapter 6.

Chapter 2

Literature Review

The chapter introduces some related works for activity recognition using everyday object and transportation mode detection. First, we describe how researchs utilize everyday object for activity classigication. Then, we present the current methods for transportation detection.

2.1 Ubiquitous computing

Ubiquitous computing is a research field where information processing has been thoroughly integrated into everyday objects and activities. Researchers on ubiquitous computing share a dream of small, cheap, robust networked sensing and processing devices , distributed at all scales throughout everyday life and generally turned to distinctly common-place ends [36]. We introduce ubiquitous computing researches from the sensor embedded in the environment to using the everyday object itself.

2.1.1 Sensor Embedded in Environment

Researchers image to enhance human life and solve problems by placing small, robust, inexpensive, networked processing devices, distributed at all scales throughout everyday life. Waterbot [2] is a system positioned at a bathroom sink to track the amount of water used in each wash. The system contains flow sensors to detect the amount of water usage. It functions as a platform for experimenting with safety, hygiene and water conservation in a sink. The Playful Toothbrush [8] is an augmented toothbrush that uses assists parents and teachers in motivating kindergarten children to learn proper and thorough toothbrushing skills by linking their brushing actions to a game. The system includes a vision-based motion tracker that recognizes different tooth brushing strokes and a tooth brushing game in which the child cleans a virtual, mirror picture of his/her dirty teeth by physically brushing his/her own teeth. KitchenSense [20] is a kitchen strengthened with sensor networked that uses Common Sense reasoning to simplify control interfaces and augment interaction. A centrally-controlled system develops a shared context across various appliances by combining embedded sensor data together with daily-event knowledge.

2.1.2 Handhelds Device

Recently, the well-spread smart phone integrated with complex sensors are graduately attracting researcher's attention. Frank Siegemund et al. [32] identify the means by which smart objects can make use of handheld devices such as PDAs and mobile phones, and derive the following major roles of handhelds in smart environments: (1)

mobile infrastructure access point; (2) user interface; (3) remote sensor; (4) mobile storage medium; (5) remote resource provider; and (6) weak user identifier. They present concrete applications that illustrate these roles, and describe how handhelds can serve as mobile mediators between computer-augmented everyday artifacts, their users, and background infrastructure services. Christian Frank et al. [11] make a system for monitoring and locating everyday items using mobile phones. They show the design of object location system and provide an algorithm which can be used to search for lost or misplaced items efficiently by selecting the most suitable sensors based on arbitrary domain knowledge. Norbert Gyorbiro et al. [13] develope a novel system that recognizes and records the motional activities of a person using a mobile phone. They measure the intensity of motions by wireless sensors attached to body parts of the user. Sensory data is collected by a mobile application that recognizes prelearnt activities in real-time. Tomas Brezmes et al. [6] implement a real-time classification system for some basic human movements using a conventional mobile phone equipped with an accelerometer. They check the present capacity of conventional mobile phones to execute in real-time all the necessary pattern recognition algorithms to classify the corresponding human movements. No server processing data is involved in this approach, so the human monitoring is completely decentralized and only an additional software will be required to remotely report the human monitoring. SoundSense [25] is a scalable framework for modeling sound events on mobile phones. It uses a combination of supervised and unsupervised learning techniques to classify both general sound types and discover novel sound events specific to individual users. Tong Zhang et al. [38] embed a tri-axial accelerometer in a cellphone, connect to Internet via the wireless channel, and using 1-Class SVM algorithm for the preprocessing, KFD and k-NN algorithm for the classification of fall detection.

2.2 Transportation Mode Detection

Transporation mode detection is strong related to traffic issuses and is a research issue for long time. Here, we present some researches which predict the transportation modes that a user takes, such as walking, biking, and taking a bus. Their work can be grouped into three general categories: (1) GPS/GSM localization-based method and (2) wearable sensor-based method (3) applications .

2.2.1 GPS/GSM Localization-Based Method

The first category targets GPS/GSM location trace data. Lin Liao et al. [22] uses hierarchically structured conditional random fields to generate a consistent model of a person's activities and places to extract a person's activities and significant places from traces of GPS data. They demonstrate that the model can be trained cross person and achieves more than 85% accuracy in determining low-level activities and above 90% accuracy in detecting and labeling significant places with data of four different persons, approximately seven days of data per person. Yu Zheng et al. [39][40] propose an approach based on supervised learning to infer people's vibration type modes from their GPS logs. They identify a set of sophisticated features and propose a graph-based post-processing algorithm to further improve the inference performance. Their algorithm considers both the commonsense constraint of real world and typical user behavior based on location in a probabilistic manner. Timothy Sohn et al. [33] explore how coarse-grained GSM data from mobile phones can be used to recognize high-level properties of user mobility, and daily step count. Evaluated by a total of 78 days of GSM logs consisting of 249 walking events and 171 driving events, their algorithm can recognize mobility modes among walking, driving, and stationary correctly 85time, and estimate daily step counts that approximates commercial pedometers. Using GPS and GSM consumes more energy and is limited to the environment where the devices can receive signal from satellites and base stations. On the contrary,we concentrated on the method with the energy-saving accelerometer and extended the detectable range.

2.2.2 Wearable Sensor-Based Method

The second category includes attempts to exploit simple wearable sensor combination or complex mobile sensor platform for transportation mode detection. iLearn [31] classifies human activities using the Apple iPhone's 200Hz three-axis accelerometer and the Nike+iPod Sport Kit. Evaluated with eight students performing four activities in the lab, their results suggest activities including running, walking, bicycling, and sitting can be recognized at accuracies of 97% without any training by an end-user. UbiGreen [12] explore the use of personal ambient displays on mobile phones to give users feedback about sensed and self-reported transportation behaviors. UbiGreen relied on three sources for transportation data: a Mobile Sensing Platform (MSP) [9], the phone' s own GSM cell signals, and the participants themselves. MSP provides inference on board and processes data from many sensors, such as microphone, light photransistor, 550Hz three-axis accelerometer, barometer temperature sensor, humidity/temperature sensor, digital compass, and etc. Both of them require the extra device attached to the user while we focused on using users' mobile phone alone. In addition, our algorithm could work on the accelerometer with less accuracy (25Hz) and less sample rate. It's more scalable to port to other mobile phone with basic acceleromter

2.2.3 Applications

The third category are those using existing transportation detection techniques to support their system. Ecorio [16] track users' mobile carbon footprint and inspire them to reduce and offset it from their cell phone. By helping user figure out his personal contribution to global warming from logging and display the transportation, it aims to presuade people to transfer to transportation with less or no greenhouse gas emission. As a participatory sensing application, PEIR [27] uses location data sampled from everyday mobile phones to calculate personalized estimates of environmental impact and exposure. PEIR system automatically segment location data into trips by mobile handset based GPS location data collection and server-side processing stages for activity classication to determine transportation mode. Then, it lookups of traffic, weather, and other context data needed by the models and uses efficient implementations of established models for environmental impact and exposure calculation. Lin Liao et al. [23] learn and infer a user's daily movements through the community by a hierarchical Markov model . Significant locations such as goals or locations where the user frequently changes mode of transportation are learned from GPS data logs without requiring any manual labeling. Lin Liao's another work [24] learn personal maps customized for each user and infer his daily activities and movements from raw GPS data. In the model, transportation mode is estimated from raw GPS data and is at the lower level for building the middle level goal. Among these applications, transporation mode detection plays a importation role for building higher level model or providing context to analysis and respone to user.

Chapter 3

Design Consideration on Mobile Phone

To fulfill the goal of detecting transportation mode by everyday object mobile phone, some constraints on ubiquitous computing were considered. Ubiquitous computing engages many sensing from environment and computations for transpotation in real time, and better not be aware by user that they are doing so. Except the accuracy and response time, which are tranditional measurements for activity recognition, the potential environment obstacles, the system requirement, such as power, user behaviors, and user feeling, like privacy, are important factors to forge a pratical everyday object in real world. In the following, we describe the limitaions on device user may concern. The detail for implementing the system is mentioned in the next chapter.

3.1 Without Researcher Supervision in Natural Environment

As an everyday object, the device is supposed to use by people in natural environment. Without supervision from researchers, user may arbitrarily operates the device with his imagination. They might not follow the instructions and neglect any warnings from machine. From the experiences by Ling Bao et al. [4], they excutes a seris of user-annotated acceleration experiments on 20 everyday activities and show the result of 95.6The design should prevent users from crashing the device or spoiling the required condition for detection.

The robustness and error-tolerant is an elemental specification for the device. Unlike the experiment in the lab, researchers wouldn't be able to recover and fix the device right away when it is broken or out of order. In our experience, the duration from user's breaking device to researcher's recovery might be several days, sometimes even a week. It might force the experiment to extend to remedy this fixing period. Therefore, the return for repairing should be avoid by preparing problem shooting routine in advance. Even if the user misuse the device, it is supposed to keep partial functioning and log the error for tracking.

3.2 Power-Saving and System Requirement-Reducing

Power consumption is a critical issue for mobile phone since it is not wired and requires people to recharge frequently. As an additional service, turning on transporta-

Table 3.1: Power consumption before and after running the system with acceleromter

	Before	After
Power	760 mW	786 mW
Current	210 mA	207 mA

tion mode detection should not reduce battery life too much and threaten other mobile phone services. Although the high power-consumption hardwares, that were GPS and wifi on the mobile phone, could provide more information like velocity, location, and surrounding buildings for inference ,they were abandoned to save energy.

In the trade-off between energy and sensing data, we chose to detect transportation mode by single accelerometer. We measured the power usage before and after running our program with acceleromter. As shown in Figure 3.1, the result shows a litte more power consumption which is acceptable. Also, similiar results is reported in Mohan’s research [26] on the mobile phone for traffic condition where it shows accelerometer cost almost trivial energy compared to the mobile phone.

Also, the device shouldn’t be special customed and only available to a few people. As a additional service on mobile phones with single acceleromter, the program should be run on most of moblie phone with same level. Thus, we reduce the system requirement to the least standard among mobile phones with single acceleromter when designing the algorithm. The feature selection and relabeling techniques described in the next chapter were done for reducing system requirements.

3.3 Privacy

With personal information from a ubiqutious computing device, context-sensing and context-aware applications can provide appropriate services for people. However, the privacy issues are raised by the personal information. Without extra security promise, people may resist and hesitate to use something which uses and possibly leaks his private data. Take GPS for example, it is a great context-sensing tool which provides location and velocity but involved privacy issues have been discussed and debated for long time [17]. Some approaches by either distributing the privacy [14] or reframing interaction [35] are addressed by previous researches for this challenge.

Another way to avoid the privacy issue is collecting less-sensitive information from users. In the price of being a little less accurate, shifting to other sensors, which people get used to or are not concern with, or dirty the data collecting process with noise is considerable. According to the statement that sensitivity of information refers to the impact of disclosing information [1], the accelerometer is a less-sensitive sensor where most people don't care their acceleration data are revealed. The privacy concern is another reason to support using single accelerometer only and discarding other sensers on mobile phone.

3.4 Using Scenerio

Since we designed to provide more context instead of replacing original functionality, the mibile phone is supposed to maintain its normal services during user's daily life. One critical concern is the position of mobile phone placed by the user might affect the strength of vibration and the orientation of accelerometer. People might put cell phones in any pocket with arbitrary orientation and replace them in different orientation or position after usage. Also, the pocket size, pocket looseness, and other objects in pocket might weaken the strength of vibration by absoring the force. Therefore, we proposed de-orientation to solve orientation and relabeling to lessen the variance of vibration strength in the following section.

Chapter 4

Transportation Mode Detection

In this chapter, we address the processing flow shown in Figure 4.1 for transportation detection. First, we demonstrate the pre-processing technique to de-orientation and relabeling raw data from accelerometer. Then, the approach to cluster pre-processed data into vibration type related to physical vibration without labeling from user is described. Finally, we present the transportation mode classification from the vibration data.

4.1 Pre-Processing

For each sensing from accelerometer, the raw data are 3 independent values for 3-axis. They were processed to solve the orientation problem, filter the noise, and reduce the feature space before converting them into vibration type.

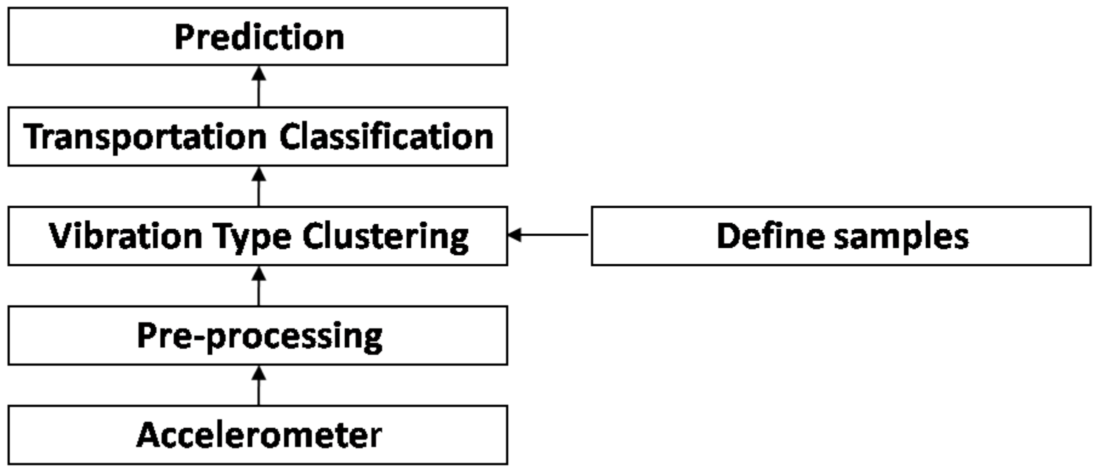

Figure 4.1: Process flow.

4.1.1 De-Orientation

The orientation problem shown in Figure 4.2 is critical for accelerometer embedded in something since the coordinate of accelerometer is different from that of the vehicle, that is human body in this thesis. We would like to know the acceleraion/deceleraion of body instead of the device embedded with accelerometer so the mapping function should be found. Nericell [26] propose a solution to map variables from accelerometer to vehicle with some strong assumptions that are not for general purpose. The system assumes that there exists other known force as strong to compared to gravity and uses eular angle to reorient. The alternative common solution is fixing the device with accelerometer to the vehicle. However, this solution is not pratical for using scenerio and is against our design limitation for everyday object described above.

We observed a phenomenon that the usage and fatigue level of people's legs was based on the vibration of transportation when moving from one place to other place. In other word, the vibration of the transportation makes people unconsciously tense his

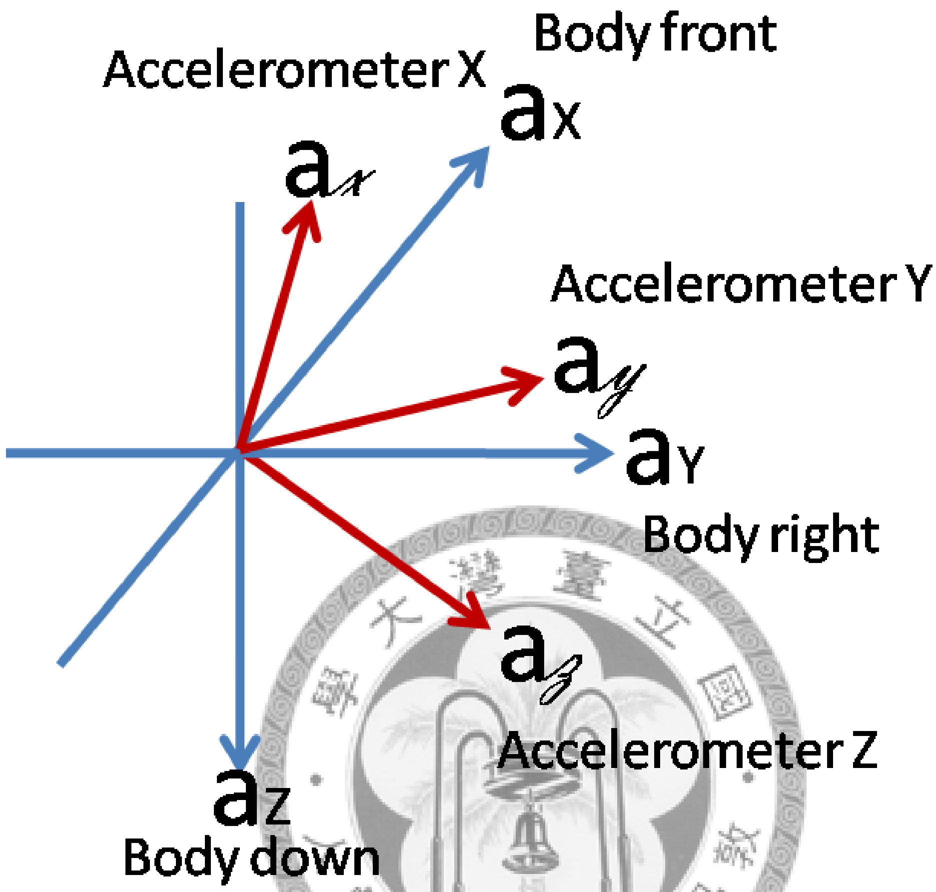

Figure 4.2: Orientation of accelerometer and that of the body.

leg muscle to keep body balance. For example, people feel more leg vibration riding on the bus than sitting in a car. More stable and comfortable the transportation provides leads to less vibration and fatigue people suffer in the travel. Through this concept, we tried to detect transportation mode by measuring the vibration on legs. In physics, jerk is used to sense stress changes and to adjust their muscle tension [34]. Theoretically, the jerk function is consider to be a vector in the form shown in Equation 4.2. However, we tested four different formulas shown in Equation 4.1, Equation 4.2, Equation 4.3, Equation 4.4 with trail-based evaluation to adapt to accelerometer measurement on the smart phone. The result in Figure 4.3 shows that Equation 4.2 outperforms than others. Therefore, we defined the vibration on user' s leg by summing up the absolute value of jerk, that was the change of three-axis accelerometer, as shown in Equation 4.1 where X was x-axis accelerometer reading, Y was y-axis accelerometer reading, Z was z-axis accelerometer reading, and t was time. In this way, the re-orientation problem due to arbitrary placement of mobile phone was avoided. Also, the inaccurate accelerometer problem is lessened because the orientation of two sequential readings is similar and their subtraction eliminate the error caused by the orientation [28].

$$Vibration = |X_t - X_{t-1}| + |Y_t - Y_{t-1}| + |Z_t - Z_{t-1}| \tag{4.1}$$

$$Vibration = \sqrt{(X_t - X_{t-1})^2 + (Y_t - Y_{t-1})^2 + (Z_t - Z_{t-1})^2} \tag{4.2}$$

$$Vibration = Minimum(|X_t - X_{t-1}|, |Y_t - Y_{t-1}|, |Z_t - Z_{t-1}|) \tag{4.3}$$

$$Vibration = Maximum(|X_t - X_{t-1}|, |Y_t - Y_{t-1}|, |Z_t - Z_{t-1}|) \tag{4.4}$$

4.1.2 Feature Selection and Relabeling

For sensing device, there are two acceleration/deceleration sources to consider: the user and the vehicle. The action taken by user and the motion changed by vehicle both influence the vibration. Since we couldn't seperate them apart, we defined user action

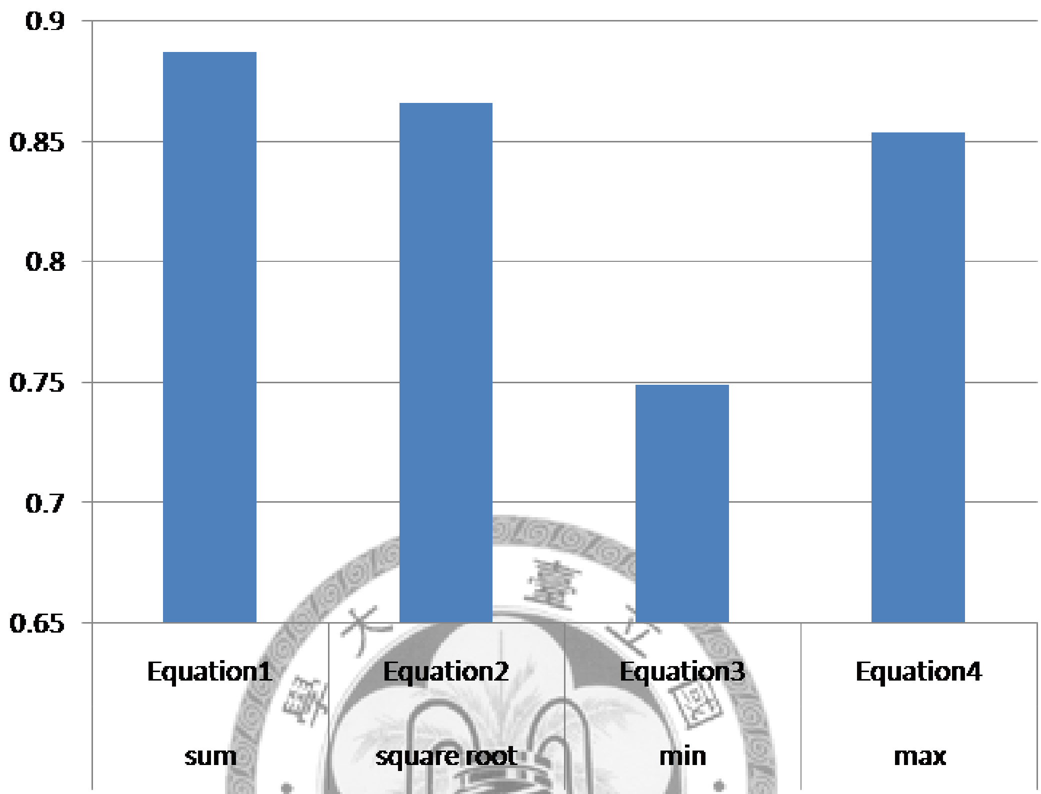

Figure 4.3: The vibration equation comparison.

and vehicle motion as a pair of vibration source shown in Equation 4.5. We selected the feature from this vibration source pair for transportation mode prediction.

$$VibrationSource = f(UserAction, VehicleMotion) \tag{4.5}$$

Figure 4.4 shows a example of vibration for six kinds of transportation mode. It's hard to recognize the transportation mode by measuring the properity of vibration value, such as minimum, maximum , standard error, standard deviation, quartile, mean, meadien, and etc. As shown in Figure 4.5, even using FFT(fast Fourier transform) to transform to frequency domain won't help to extract features. The reason is that multiple transportation shares the same vibration source during some period of time. To be more clear, the vibration was caused by user actions and vehicle motions which compose the travel of the transportation. The same actions appear during different transportation, like sitting on bus and sitting in car. Also, different motions by different kinds of transportation may lead to similar vibration, such as immediately braking in gasoline-powered vehicle: car, bus, and motorcycle. Moreover, It's possible for similar vibrations to exist under different action and different motion, for example waiting for the traffic light during walking and sitting on MRT.

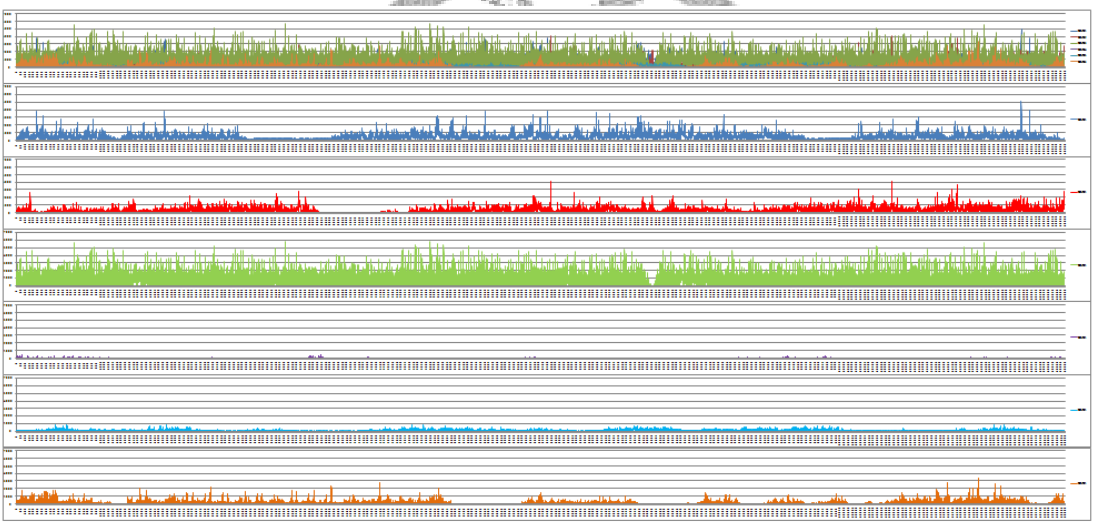

Figure 4.4: Vibration samples of six kinds of transportation mode. From above to below: all, motorcycle, bicycle, walking, MRT, car, bus.

As a result, we first focused to convert the vibration data into the corresponding vibration source. The vibration type was proposed to model different pairs of vibration source. To form the vibration type, we generated histograms by using non-overlapping moving window over the sequential vibration data. Different kinds of window size were tested for minimum duration with discriminability of histogram pattern. Histograms produced with selected minimum window size should have similar graphic contours as those produced with other larger window sizes. In the other word, histograms created from the selected window size and those created from other larger window sizes share the same shape but differ in the scale due to the repeated cycle. Since the sample rate of the chosen accelerometer was low, the unnecessary details were neglected and the major feature, that is distribution of the histogram, is kept. This will reduce computation without general loss of feature. The similar technique is mentioned in Dong et al.’s research [10] that they project feature vectors into an auxiliary space using locality sensitive hashing and to represent a set.

4.2 Vibration Type Clustering

With histogram series from above, we converted each histogram to vibration type without labeling on vibration type from users. We first defined samples and then classified the data by these samples. Two methods for sample definition and two methods for classification with different features were proposed.

4.2.1 Vibration Type Sample Definition

A small amount data were randomly extracted from the whole data set to define samples for convertion in this process. The sample could be retrieved from early data collection, too. The coverage over all users is not required by this process. The sample definition was desinged to extract features and was not involved in the supervised learning process for each user.

Human Definition

For the first method, we observed and divided these histograms into nine kinds of vibration types, shown in the Figure 4.6, by their shape, distribution, and position in the histogram. Some properties like Median, Mean, SD, SE, skewness, and kurtosis [15] were used for helping clustering histogram. Vibration types are not unique in only one transportation mode since they are caused by low-level motions. Some transportation modes share the same vibration types but with different number and order sequence. Some types of histogram pattern represent specific vibration caused by physical action, for example, type 1 in Figure 4.6 is caused by vibration of regularly lifting and putting down the leg. The other vibration types are generated by the traffic vehicles, such as type 7 in Figure 4.6 is due to the slight swing of bus or car. Still others are not clearly identified for the mapping to single physical vibration for multiple factors.

K-Clustering

In order to automatically create samples for the scalability and unsupervision , we processed the small amount data with K-clustering using MDPA(minimum distance of pair assignment) as similarity function. MDPA [7] takes the similarity of the non-overlapping parts of histogram into account as well as that of overlapping parts in linear time.

To compare with the human defined sample, we tested the accuracy for sample defined with the same data and setting. Since the sample defining process used time as random seed to initialize, we repeated 5 times for each cluster amount. Result for trail-based analysis is shown in Figure 4.7. The purple mark and the black line represents the average accuracy and the range from minimum to maximum of accuracy for repeated 5-time test of k-clustering while the red line represents the accuracy of human-defined samples. The accuracy of best machine-defined samples is worse than that of human-defined samples. Moveover, the sample amount created by K-clustering method is much more than human definition for the reason that human definition picks up only a few representative instances while k-clustering cluster all instances into sample groups. To reduce computation time and increase accuracy, we still used human-defined samples in the following experiments.

4.2.2 Vibration Type Classification

With the predefined samples, we classified histogram series into vibration type series. We introduce the two approachs using different features for classification.

Histogram Bin as Feature

First, we treated each histogram bin as a feature and did supervised learning. Among all the classifier we tested, the multiple layer perceptron was chosen for better accuracy. These features made the histograms with similar skelton classified into corresponding class. The big picture of histogram was saved by this method. However, some weak but important cues were ignored because the relative position of histogram bin was not considered. For example in Figure 4.6 , vibration type 6 and vibration type 9 were often confused by this method for two dots in the rightest. Those two dots was important for the meaning that user instantly encountered a strong acceleration or deceleration. The other shortcoming is larger computation time due to the large amount of feature. Therefore, the second feature space was proposed for better performance in the following.

KNN with MDPA as Similarity Function

On our observation of physical motions, when the vibration user took altered from mild to severe, elements whose amount was fixed in histogram moved from left(low acceleration/deceleration change) to right(high acceleration/deceleration change). As a result, MDPA was suitable for measuring difference between two histograms. We used KNN (K-nearest-neighbor) with MDPA as similarity function for classification. Calculating the distances between the judgeing histogram and each histogram in predefined sample, the histogram was categoried to be the vibration type with smallest average distance of that type. MDPA calculated how many steps to move from one histogram to another histogram in linear time. Hence, the computation complexty for each histogram was reduced to O(mn) where m was sample amount and n was histogram bin amount. it was much faster than using histogram bin as feature and training with multilayer perceptron which we terminated the network with limit time before it really converged. Also, the small amount of dots in previous example was solved by weighting these far away dots with more moving steps.

4.3 Transportation Mode Classification

From the vibration type clustering, we converted the data into a series of vibration type. In the following, we described two methods with different feature space for supervised learning on these vibration type data with user labeled transportation mode.

4.3.1 Vibration Type Occurance Probability as Feautre

From the sequential vibration type series, we observed that the occurrence frequency of each vibration type was different for each transportation mode. Some vibration type has more chance to appear in some transportation mode while the other vibration type rarely exists in the same transportation mode. Some transportation mode has its own dominant vibration type which is seldom found in other transportation mode. We examined the occurrence probability of vibration type in each trail to distinguish different transportation modes. Figure 4.8 shows the CDF (cumulative distribution function) of probability distribution of vibration type occurrence in each trail of six transportation modes for nine vibration types. If the CDF grows fast to top, most trails of this transportation have lower occurrence probability of this vibration type. On the contrary, if the CDF converges slowly to top, most trails of this transportation have higher occurrence probability of this vibration type. For example, a point with x=70%, y=28.6% in the CDF of walking of vibration type 1 (shown in the green line in the left-top graph in Figure 4.8) means that occurrence probability of vibration type 1 for the least 28.6% of walking trails is less than or equal to 70%. The vibration type is hard to separate two transportation modes well when their CDFs are close to each other. For example, the CDFs of motorcycle and bicycle of vibration type 7 (red line and deep blue line shown in the left-most and fourth-from top graph in Figure 4.8) are twisted and are hard to distinguish with this vibration type.

Based on this finding, we used occurrence probability of each vibration type in the specific window size of transportation trail as feature to build classifier. Small amount of feature is computation-saving which is important for inference on the mobile phone. There are nine features and each feature is dependent on other features because the sum of all occurrence probability of each vibration type is 1. We filtered the algorithms which assumed the independence between feature and another feature, like naive bayes [30]. Each feature space is from 0% to 100% with interval unit 1% and is linear because each feature is occurrence probability of each vibration type. Therefore, we selected three common linear classifiers: SVM with linear kernel, logistic regression, and multilayer perceptron to compare with. We learned models and classified transportation mode over our extracted features the Weka machine-learning toolkit [37].

4.3.2 Vibration Type as Feautre

We observed another phenomenon that the vibration series was meaningful in their sequential order. During different transportations, user' physical motion causes different vibration combinations. We assumed 10 seconds as a unit for each vibration combination made up of 5 vibration type instance. Using CRF(Conditional random field), we treated each label of transportation mode as the hidden nodes and the group of vibration type as the observations. Figure 4.9 shows the dependency structure between the hidden nodes and observations. Each segment was represented by 5 features, $X_i^1, X_i^2, X_i^3, X_i^4, X_i^5$, which were vibration type data. The assumption was that the transportation mode of segment i(Y_i) depends on the label of current segment($X_i^1, X_i^2, X_i^3, X_i^4, X_i^5$).

CRF uses feature functions to calculate the conditional probability distribution. For feature functions, we used only the uni-gram functions between the label and the feature of one segment. For each combination of a hidden node and its 5 features, a set of feature functions is defined for every transportation mode as Equation 4.6.

$$f_{a,k}(X_i, Y_i) = \begin{cases} x, & \text{if } Y_i = a \text{ and } X_i^k = x \\ 0, & \text{otherwise.} \end{cases} \tag{4.6}$$

where $a \in TM \cup P, k = 1, 2, ..., 5$.

In this research, we use an open source API CRF++ [19] for implementation. In training phase, given the training data $D = (D_1, D_2, ..., D_N)$ where $D_i = (A_i, X_i)$, the learning criteria was to find the weight vector w that maximized the log-likelihood of the training data. In prediction phase, given the observation of segments $X = X_1, X_2, ..., X_N$, we got the label sequence $Y = Y_1, Y_2, ..., Y_N$ with the maximum conditional probability as the output. Then the label sequence in the specific window size of transportation trail voted for the transportation mode inference in this duration.

4.4 Three Combinations

For this two-layer approach: vibration type clustering and transportation classification to infer modes of transportation, there were two approaches for each layer. There are three combinations shown in Table 4.1 by picking up one approach from each layer. The first combination was multilayer perceptron using histogram bin as feature in first layer and SVM using vibration type occurance probability as feautre in second layer. The second combination was KNN using MDPA as similarity function in first layer and SVM using vibration type occurance probability as feautre in second layer. The third combination was KNN using MDPA as similarity function in first layer and CRF using vibration type as feautre in second layer. Among three combinations, the second combination was implemented on mobile phone for real time transportation detection due to better performance than others.

Table 4.1: Three combinations for transportation mode prediction

Combination	Vibration type clustering	Transportation classification
MLP+SVM	multilayer perceptron using histogram bin as feature	SVM using vibration type occurance probability as feautre
KNN+SVM	KNN using MDPA as similarity function	SVM using vibration type occurance probability as feautre
KNN+CRF	KNN using MDPA as similarity function	CRF using vibration type as feautre

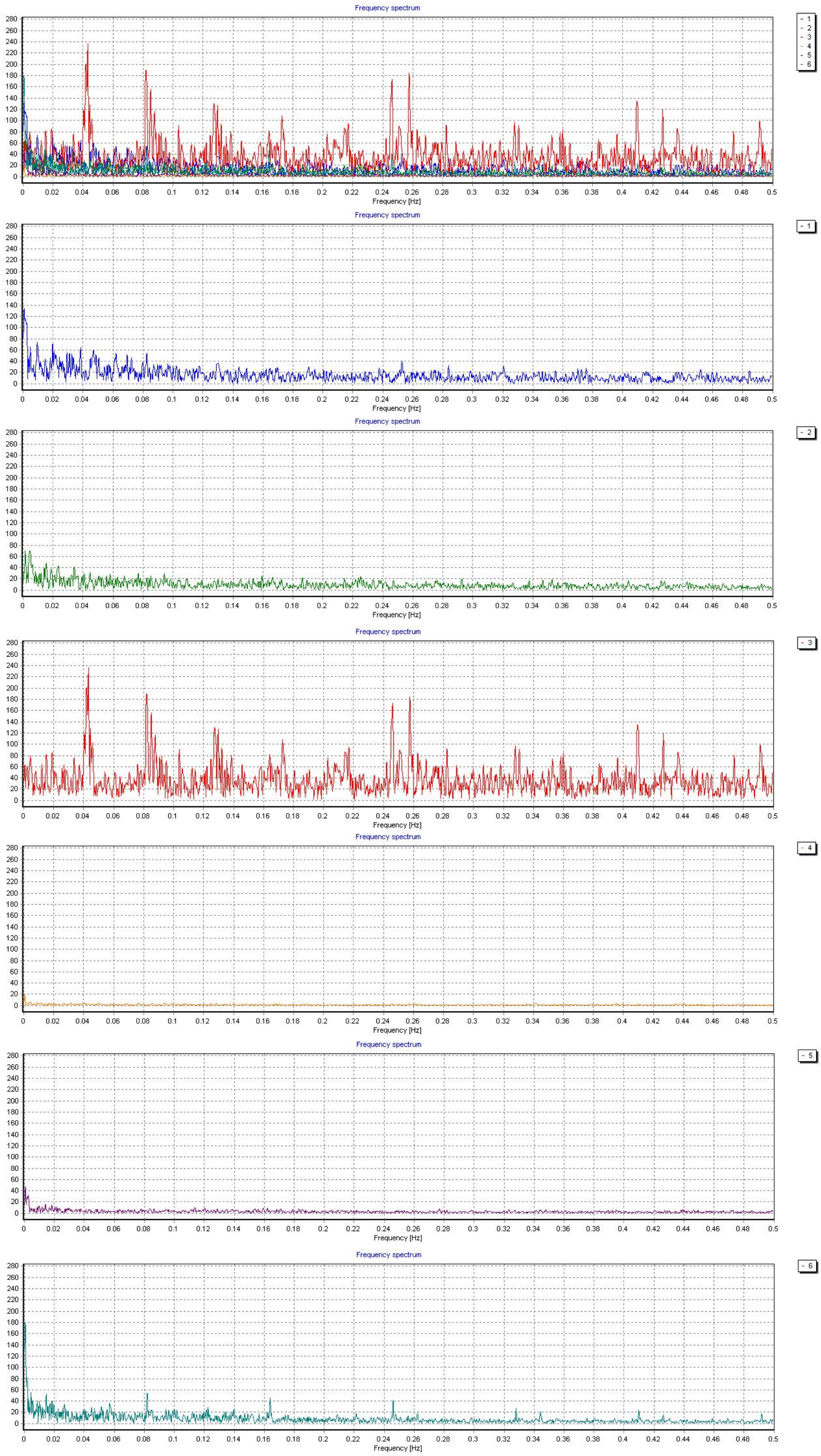

Figure 4.5: Vibration in frequency domain samples of six kinds of transportation mode. From above to below: all, motorcycle, bicycle, walking, MRT, car, bus.

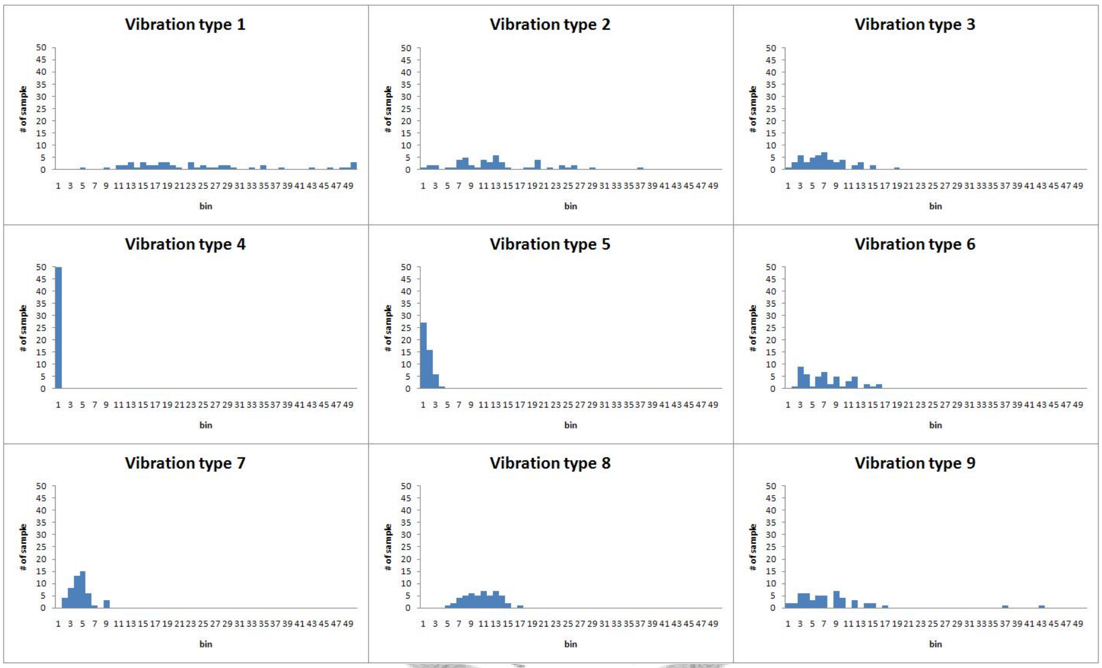

Figure 4.6: Human defined nine vibration type patterns. Each figure is a histogram where x axis is bin and y axis is number of sample..

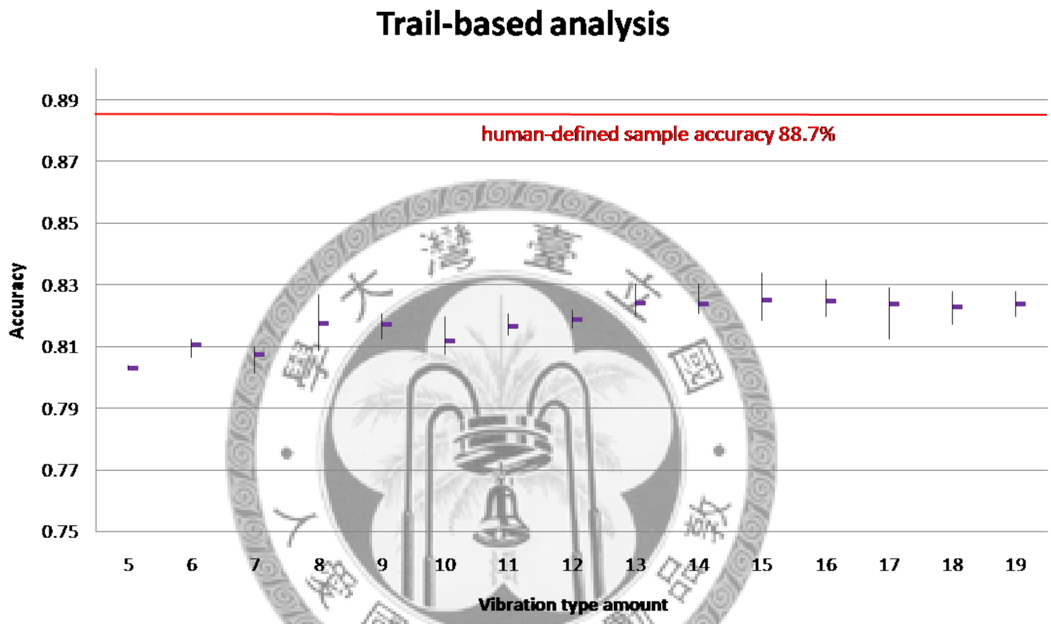

Figure 4.7: The trail-based analysis accuracy on different cluster amount. The purple mark and the black line represents the average accuracy and the range from minimum to maximum of accuracy for repeated 5-time test of k-cluster. The red line represents the accuracy of human-defined samples.

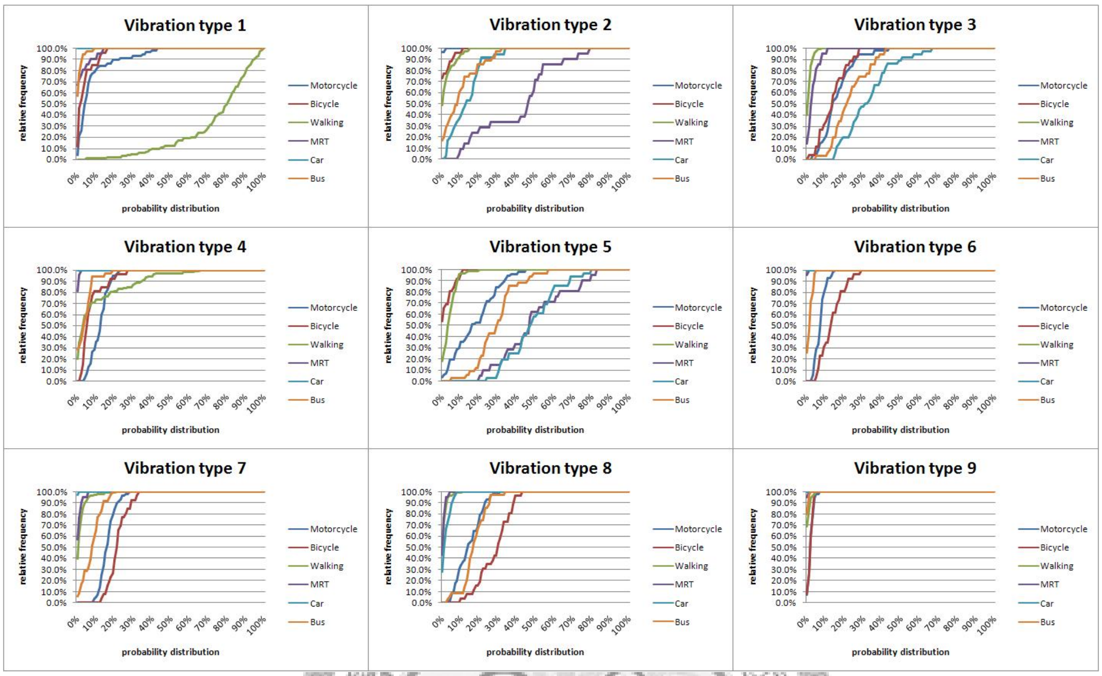

Figure 4.8: CDF (cumulative distribution function) of probability distribution of vibration type occurrence in each trail of six transportation modes for nine vibration types.

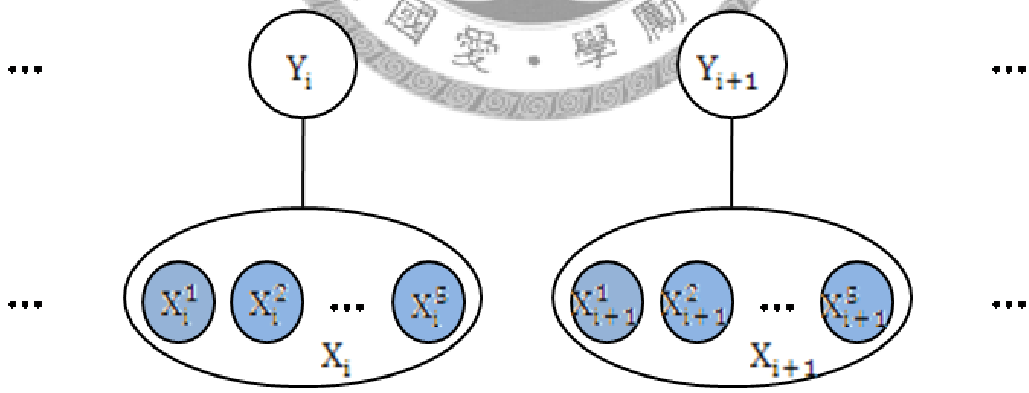

Figure 4.9: The CRF model. The assumption was that the transportation mode of segment i depended on the 5 features of observation of segment i.

Chapter 5

Experiment Design and Result

In this chapter, we present the design and results of our experiment. First, we compared four formulas for defining vibration equation. Second, we introduce the collecting device, progross for subjects to collect, and data description. Third, the evaluation for post-hoc and real-time detection from mixed data from all subjects is shown. Finally, we evaluated the data from each individual subject as basic unit for cross validation.

5.1 Data Collection

To explore the feasibility of our approach for classifying transportation activities, we collected data from 17 subjects in 4 experiments performing several of six activities in their daily life: walking, bicycle, bus, car, motorcycle, and MRT. Noticeably, user was told to record the data during his daily transportation and was not required to perform all kinds of transportation activities. Also, user was allowed to stop logging for privacy or any other personal reason. Some people may place the mobile phone somewhere and forgot to take with them or misplace them in the bags. The data were abandoned in these situations when they reported. Each travel from one place to another place was divided into several trails where each trail was labeled to be only one kind of six transportation mode. The duration of each trail varied from 5 minutes to hours due to the different routine the user take. We trimmed several seconds off the start and the end in consideration of the time the user interacted with the interface to log.

5.1.1 Collecting Device

Among all kinds of smart phone with acceleromter, we chose HTC diamond mobile phone shown in Figure 5.1 to implement our application for its weak accelerometer performance. Unlike other smart phones, such as Apple I-phone(100Hz) and Nokia N95(100Hz), the highest sampling frequency of accelerometer in HTC diamond is only 25Hz. This sample rate is little more than sufficient compared to the 20 Hz frequency required to assess daily physical activity [5].

Table 5.1: Collected six kinds of transportation data

Processor	528 MHz
Weight	110 g (with battery)
Dimensions	102 mm (Length) 51 mm (Width) 11.35 mm (Tall)
Storage	4 GB
Battery	900 mAh
Battery time	330 minutes(Talk) 285 hours(Standby)

The hardware specification of HTC diamond in shown in Table 5.1. The compu-

Figure 5.1: The HTC diamond mobile phone.

tation power is strong enough for most classifiers with small amount of feature and the storage is large enough for logging in a long period. The battery is moderate like ordinal mobile phone and is required to recharge regularly. The weight and the volume is unobtrusive for user to carry in daily life. The software was written in C-sharp language for the HTC diamond C-sharp accelerometer SDK.

5.1.2 Data Description

There were 4 experiments held from June 16 2009 to July 16 2009, from September 14 2009 to October 18 2009 , from October 7 2009 to November 7 2009, and from October 19 2009 to November 23 2009 respectively for collecting data. Except the first experiment whose duration was two month, the rest 3 experiments were lasting for one month. The duration of last 3 experimemt was decided by the observation from the first one (See the Data Observation subsection). The subjects were asked to carry the mobile phone with them in the daily life for two month(one month). Additionally, they were not required to begain and end in the same time that some of them may start (and end) early or later. The subject data is shown in Table 5.2.

Table 5.2: Subject data in 4 experiments

Experiment	Date	Amount	Job	Sex
1	Jun. 16 2009 - Jul. 16 2009	8	6 students 2 office workers	7 male 1 female
2	Sep. 14 2009 - Oct. 18 2009	4	4 office workers	4 males
3	Oct. 7 2009 - Nov. 7 2009	1	1 student	1 male
4	Oct. 19 2009 - Nov. 23 2009	4	4 students	3 males 1 female

Take one daily routine for example, one subject walked from home to bus stop, took bus to the bus stop near school, and walked from bus stop to the department building on her way to school. There were three trails collected in the traveling. To lessen their burden, the subjects were only required to log trail more than 5 minutes from one transportation mode to another transportation mode. They labeled the transportation mode by pressing the corresponding button on the HTC diamond screen, and then put the HTC diamond in their front trouser pocket in the beginning and the end of the routine. The back trouser pocket wasn't used due to the sitting posture which might squeeze and damage the mobile phone. We did not constrain what types of pockets were worn or the orientation of the smart phone. Some data loss caused by user’s forgetting or other reasons to label when he finished movement were discarded.

We collected total 1125 trail data from 4 experiments where 294 trail data were not used due to short duration. The detail of collected valid data is shown in the Table 5.3. As expectation, the duration of car is longest and that of walking is shortest for the reason that people walk to near place and drive to far place. Also, motorcycle, bicycle, MRT, and bus meet the common sense which considers them as transportation for short/middle range.

Table 5.3: Collected six kinds of transportation data

Class	Transportation mode	Event amount(trail)	Average duration(minute, second)
A	Motorcycle	222	16m19s
B	Bicycle	68	16m36s
C	Walking	220	12m37s
D	MRT	110	19m32s
E	Car	60	30m22s
F	Bus	151	23m51s

5.1.3 Data Observation

The duration of experiment for each subject to collect sufficient data is a major problem. Peoeple gradually lose freshness and lower their willing to attend the experiment with the time passing. Generally, the daily routine from home to working place is repeated for each subject in one week. However, there are many conditions affecting the acceleromter reading for consideration, such as whether, on/off peak, multiple choice, and riding with a passenger.

If it rains, the traffic vehicles on road, like motorcycle, bicycle, car, and bus, will be slowed down. On peak, the public transportation system, like bus and MRT, is more crowded and more vibrations caused by actions are recorded by accelerometer. On the other hand, subject usually have a seat and relative static values are logged when riding off peak. Occasionally, there are multiple transportation choices for subject to consider. For example, one subject in our experiment could take a bus or MRT from home to school. For the vehicle whose weight is light and relatively close to that of human, like bike and motorcycle, riding with a passenger increases mass to manage, causes the center of gravity shifted rearward and higher, and influences acceleration/deceleration. These variances make the g-sensor reading much different from that of riding alone.

As a result, we examined the data from the first experiment and verified the difference between one-month data and two-month data. The trail-based analysis and window-based analysis were done in Figure 5.2. The accuracy, weighted recall, weighted precision (not using average recall and average precision for different length of each class data set) of them shows almost no difference between one-month data and two-month data. Therefore, the experiments held in the following were designed for one-month long.

5.2 Trail-Based Evaluation

From our observation on data collected, most subjects performed only 1 or 2 transportation activity except walking in daily routine. Under this unbalanced condition, we replaced subject-based evaluation with trail-based evaluation. Trails from all subjects were mixed for cross validation where the instances from the same trail were put into the training set or the testing set in the same time. The subject-based evaluation was still done in the next section for reference.

5.2.1 Trail-based Analysis

To evaluate the feasibility of our method, we first verified the trail-based learning by treating each trail as an instance. That is, the device have to collect data from the whole trail and process post-hoc before answer the transportation mode. Using the labeled trails of activity, we trained our classifier and evaluated it using a 10-fold cross validation method over the entire data set for three combinations. This produced models which worked well across all collected accelerometer data. Table 5.4, Table 5.5, and Table 5.6 show the confusion matrices, accuracy, (true positive+ true negative) / (true positive+ true negative + false positive + false negative), precision, true positive / (true positive + false positive), and recall, true positive / (true positive + false negative) for three combinations. The values along the diagonal indicate the classifiers' performance for predicting and matching the ground truth events. Precision is the percentage of predicted events that are correct. A high precision number indicates less false positives. Recall is the percentage of ground truth events that were correctly identified. A high recall number indicates that many ground truth events were hit. Accuracy represents the percentage of predictions that are correct and our overall accuracy for three methods are 87%, 89%, and 84%. As shown in Figure 5.3, the KNN with MDPA for first layer and SVM with occurance probability as feature for second layer outperforms the other methods on accuracy, average recall, and average precision.

In this combination, our classification scheme performed very well for walking and MRT, correctly detecting most trails of themselves (recall 98%, 92%) and raising few spurious events (precision 98%, 94%). Within a walking activity, there are two motions: regular swings of legs causing strong vibration in spite of walking speed or standing to wait for the traffic light. In our observation, these motions makes walking feature strongly different from the other classes. The acceleration and brake of MRT is highly automatically controlled and performed regularly so it also shows strongly pattern to distinguish. For two-wheel drives, motorcycle and bicycle are inverse to each other since their motions ,such as dodging and turning, and driving scenerio are similar. The motorcycle detection(recall 94%,precision 95%) surpass bicycle detection (recall 81%,precision 74%)for more data collected (222 trails vs 68 trails). Also, for the other two transportation mode sharing similar behaviors, the prediction of bus (recall 82%,precision 76%) defeats that of car (recall 53%,precision 68%)for bigger data set(151 trails vs 60 trails).

The results show that we are able to distinguish between different transportation states with high accuracy without having to equip a person with any other additional devices or sensors. The precision and recall numbers show that this type of scheme could be used in a person’s daily life, to give an accurate logging of transportation activity. In the following section, we will discuss how to extend trail-based analysis to window-based analysis for real-time detection in the daily life.

Table 5.4: Trail-based Analysis confusion matrix for MLP + SVM combination

		Predicted results(trail)							
		A	B	C	D	E	F		
Ground truth	A	207	10	2	0	0	3	0.93	Recall
	B	11	45	3	0	0	9	0.66	
	C	1	0	215	2	0	2	0.98	
	D	0	0	1	105	1	3	0.95	
	E	0	0	2	1	30	27	0.5	
	F	3	5	1	3	16	123	0.81	
		0.93	0.75	0.96	0.95	0.64	0.74	Accuracy	
		Precision						0.87	

Table 5.5: Trail-based Analysis confusion matrix for KNN + SVM combination

		Predicted results(trail)							
		A	B	C	D	E	F		
Ground truth	A	209	10	1	0	0	2	0.94	Recall
	B	5	55	2	0	2	4	0.81	
	C	0	2	216	2	0	0	0.98	
	D	0	0	1	101	0	8	0.92	
	E	0	1	0	2	32	25	0.53	
	F	5	6	0	3	13	124	0.82	
		0.95	0.74	0.98	0.94	0.68	0.76	Accuracy	
		Precision						0.89	

Table 5.6: Trail-based Analysis confusion matrix for KNN + CRF combination

		Predicted results(trail)							
		A	B	C	D	E	F		
Ground truth	A	201	15	2	4	0	0	0.91	Recall
	B	9	44	3	2	1	9	0.65	
	C	1	1	214	2	1	1	0.97	
	D	0	0	1	101	2	6	0.92	
	E	1	0	0	1	20	38	0.33	
	F	4	3	0	8	19	117	0.77	
		0.93	0.70	0.97	0.86	0.47	0.68	Accuracy	
		Precision						0.84	

5.2.2 Window-based Analysis

Unlike the post-travel analysis, feedback from the mobile phone in real time should be as soon as possible. Extending to real-time detection, we sampled the instances from an overlapping moving window with 5 minutes of window size and 30 seconds of each moving over all trails. That is, the system made a judgment on which transportation mode the user was after 5 minutes from turning on the system and re-judged every 30 seconds. The system was supposed to be independent from the place or routine so the bias of testing with the model trained by the instances from the same trail shall be prevented. Therefore, we used trail-based 10-fold cross validation where the instances from the same trail were put into the training set or the testing set in the same time. As a comparison, the same criteria as the trail-based analysis for three combinations were applied here shown in Table5.7, Table5.8, and Table5.9. The overall window-based performance is worse than the trail-based analysis with accuracy declining from 87% to 78%, 89% to 78%, and 84% to 75%. Like the result in trail-based analysis, the KNN with MDPA for first layer and SVM with occurance probability as feature for second layer defeats the other methods on accuracy and average recall shown in Figure 5.4.

The decreasing of right prediction of motorcycle and walking is small (recall 94% to 93%, 98% to 94%) and the increasing fallacious events is not much (precision 95% to 92%, 98% to 91%). The diminishing of correctly identifying taking on the MRT is a few (recall 92% to 89%) and the enlarged false positive is higher (precision 94% to 82%). The recognition of bus performed much worse than trail-based analysis (recall 82% to 76%) and raising much more false positives (precision 76% to 65%). The inference of car became worse (recall 53% to 42%) but remained the precision (preci-

Table 5.7: Window-based Analysis confusion matrix for MLP + SVM combination

		Predicted results(5-minute overlapping window)							
		A	B	C	D	E	F		
Ground truth	A	4269	179	67	9	1	178	0.91	Recall
	B	263	706	43	24	17	417	0.48	
	C	27	5	2868	82	7	104	0.93	
	D	0	0	72	2745	27	171	0.91	
	E	0	13	1	112	1341	1439	0.46	
	F	143	152	21	264	778	4042	0.75	
		0.91	0.67	0.93	0.85	0.62	0.64	Accuracy	
		Precision						0.78	

Table 5.8: Window-based Analysis confusion matrix for KNN + SVM combination

		Predicted results(5-minute overlapping window)							
		A	B	C	D	E	F		
Ground truth	A	4353	146	107	2	2	93	0.93	Recall
	B	234	788	42	32	15	359	0.54	
	C	61	10	2895	39	8	80	0.94	
	D	0	0	114	2674	26	201	0.89	
	E	2	12	2	154	1208	1525	0.42	
	F	107	301	14	361	489	4128	0.76	
		0.92	0.63	0.91	0.82	0.69	0.65	Accuracy	
		Precision						0.78	

Table 5.9: Window-based Analysis confusion matrix for KNN + CRF combination

		Predicted results(5-minute overlapping window)							
		A	B	C	D	E	F		
Ground truth	A	4215	254	78	3	8	145	0.90	Recall
	B	353	641	42	23	55	356	0.44	
	C	83	27	2816	76	6	85	0.91	
	D	0	10	98	2565	113	229	0.85	
	E	30	6	0	115	1505	1250	0.52	
	F	148	247	9	229	1143	3624	0.67	
		0.87	0.54	0.93	0.85	0.53	0.64	Accuracy	
		Precision						0.75	

sion 88% to 65%) Among all transportation guessing, the decay (recall 81% to 54%, precision 74% to 63%) of bicycle was worst. On our observation, the higher dropping of the recall and precision numbers of transportation was related to the larger variance on vibration type series of each trail. The other possible reason was that the window size is smaller than the pattern cycle of the transportation.

Compared to the trail-based analysis, the result is not good enough. However, the prediction is made every 30 seconds and the voting of the results from several predictions can be used to raise accuracy. With the historical records of inference, the window-based system could be used in a real scenario.

5.2.3 Different Window Size

To estimate the degeneration of recognition from infinite window size, that was the whole trail, to small window size, we evaluated data with different moving window size. The accuracy, recall, and precision of inference with different window sizes are shown in the Figure 5.5. The instances of each classification were extracted from an overlapping moving window with 1 to 5 minutes of window size and 30 seconds of each moving over all trails. Larger window size is not tested since people won't tolerate larger activating computation time on real time detection. An additional 30-second non-overlapping moving window size was tested for shorter duration. All classification generated the model by the KNN with MDPA for first layer and SVM with occurance probability as feature for second layer and tested it in trail-based 10-fold cross validation.

The overall accuracy is over 701.5 minutes and the recall of MRT drops dramatically from 2.5 minutes. The precision of motorcycle, car, and walking decreases heavily from 1.5 minutes and the precision of bike, bus, MRT decreases heavily from 3 minutes. Some classes have local optimum in recall or precision curve, for example class motorcycle has local optimum in 3.5 minutes. This observation could be used to build classifiers for each transportation mode with different window sizes. The result also presents the inference performance before reaching the minimum activating time, which is the window size, on real-time detection.

5.3 Subject-Based Evaluation

For reference, we trained and tested classifiers using subject-based cross validation in two configurations : within-person models and cross-person models. However, the result may be biased for the imbalanced data collected by each subject.

5.3.1 Single User Cross Validation

For each of the 17 people, we chose KNN with MDPA as first layer and vibration type occurance probability as second layer to build classifier for 10-fold cross-validation on his own data. As shown in Table 5.10, the average accuracy of each subject for trail-based analysis and window-based analysis are 95% and 93%. Without doubt, the accuracy of those subjects with only one mode is definitely one hundred percent and meaningless. However, the accuracy of subjects with equal or more than 2 modes is not related to amount of mode and is shown in Figure 5.6. The accuracy doesn't decrease when the transportation mode amount increasing. It is more related to the usage and habit of individual subject himself.

Table 5.10: Single User Cross Validation

Subject	Mode amount	Trail-based accuracy	Window-based accuracy
1	5	98%	96%
2	1	100%	100%
3	4	100%	96%
4	1	100%	100%
5	4	95%	92%
6	1	100%	100%
7	2	100%	96%
8	4	93%	76%
9	3	95%	88%
10	2	96%	99%
11	1	100%	100%
12	4	80%	85%
13	2	67%	72%
14	2	100%	97%
15	4	99%	96%
16	4	85%	96%
17	1	100%	100%
Average	2.65	95%	93%

5.3.2 Leave-One-User-Out Validation

We also tested data from subject with models built by other people. For each of the 17 people, we trained a classifier, which used KNN with MDPA as first layer and vibration type occurance probability as second layer, using data from other 16 people and testing it on his own data. As shown in Table 5.11, the accuracy for trail-based analysis and window-based analysis are 74% and 63%. One reason for low accuracy is individual difference among subjects that the accelerometer data is affected by their habits for transportation. For example, one motorcycle rider is used to not brake but slide while the other rides fast and brake often. The other reason is the imbalanced data collection from each transportation mode. For the scarce transportation type, like car and bike , if the major providing person is chosen for testing, the test data set may be much larger than the train data set.

Table 5.11: Leave-One-User-Out Validation

Trail-based accuracy	Window-based accuracy
74%	63%

5.4 Transportation Transit

During user's transferring from one transportation mode to another transportation mode, the window-based analysis was easily confused by mixed data of two modes in the moving window. With the large window size of window-based analysis, it took too much time for the pattern of new transportation mode to dominate the moving window. This caused a fuzzy period between the transferring of two trails. To remedy this deficiency, a fast detecting method was come up to detect the accurate time stamp of change of transportation.

Since the change of transportation was between walking and the other transportation mode, additional short time interval walking detector was used. It was responsible for predicting whether user state was walking or non-walking. To verify how soon we can guess the change of transportation without loss of accuracy, we evaluate the recall and specificity of walking detection with different moving window size shown in the Figure 5.7. The overlapping moving window data were tested by 10-fold cross validation and the same classifier as window-based analysis. The recall and the specificity of walking remained 82% and 90% even when the window size dropped to thirty seconds. It was high enough to detect and record when the user changed the transportation.

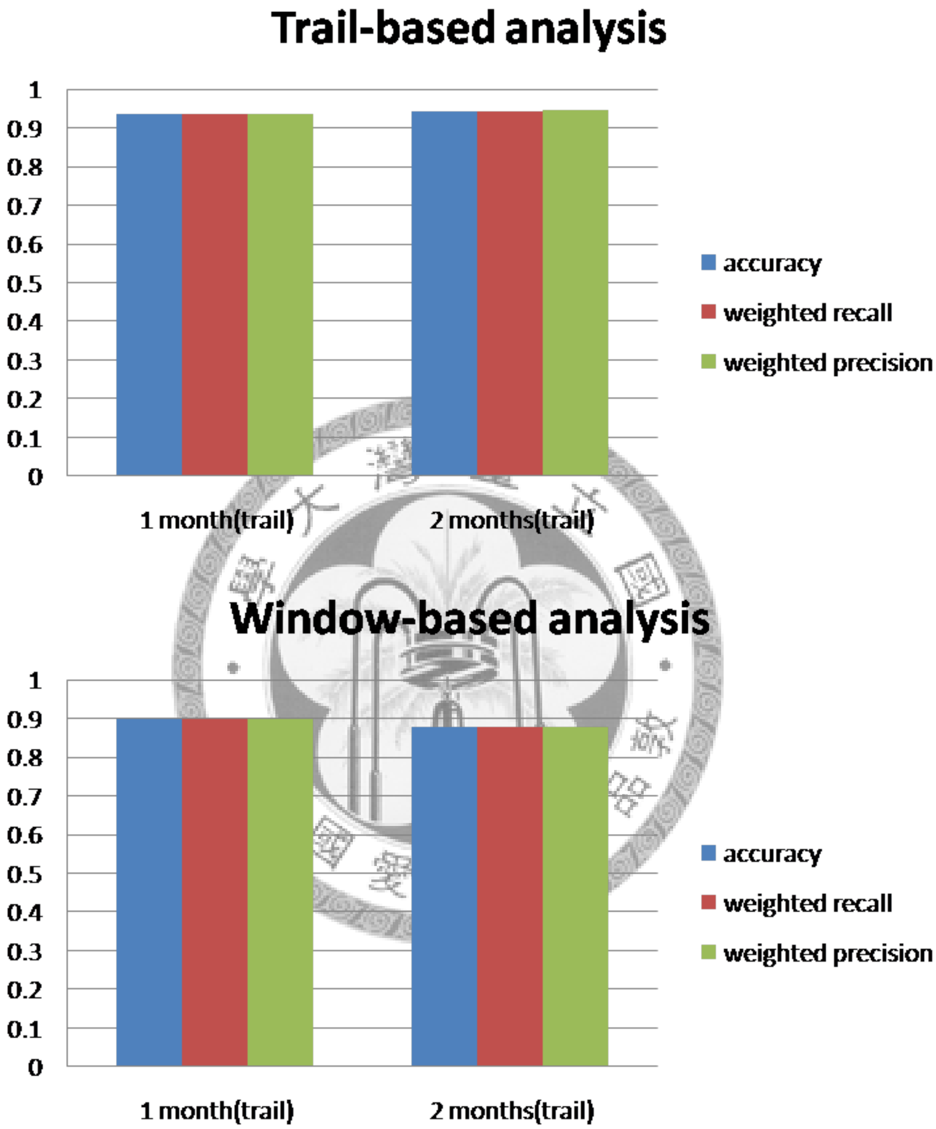

Figure 5.2: Trail-based analysis comparison between 1-month-long data and 2-month-long data.

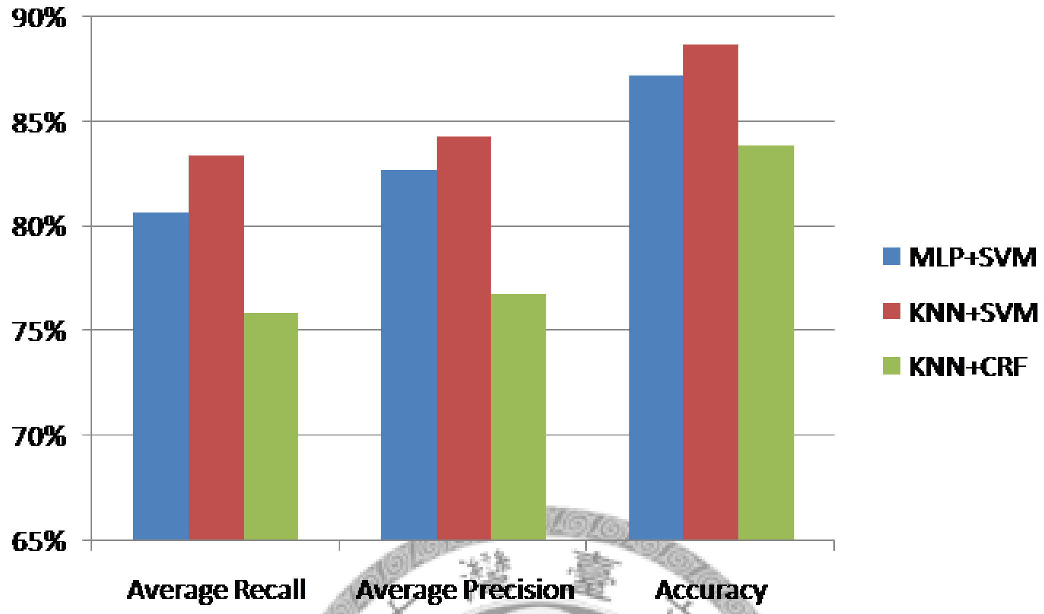

Figure 5.3: Comparison of three methods for trail-based analysis.

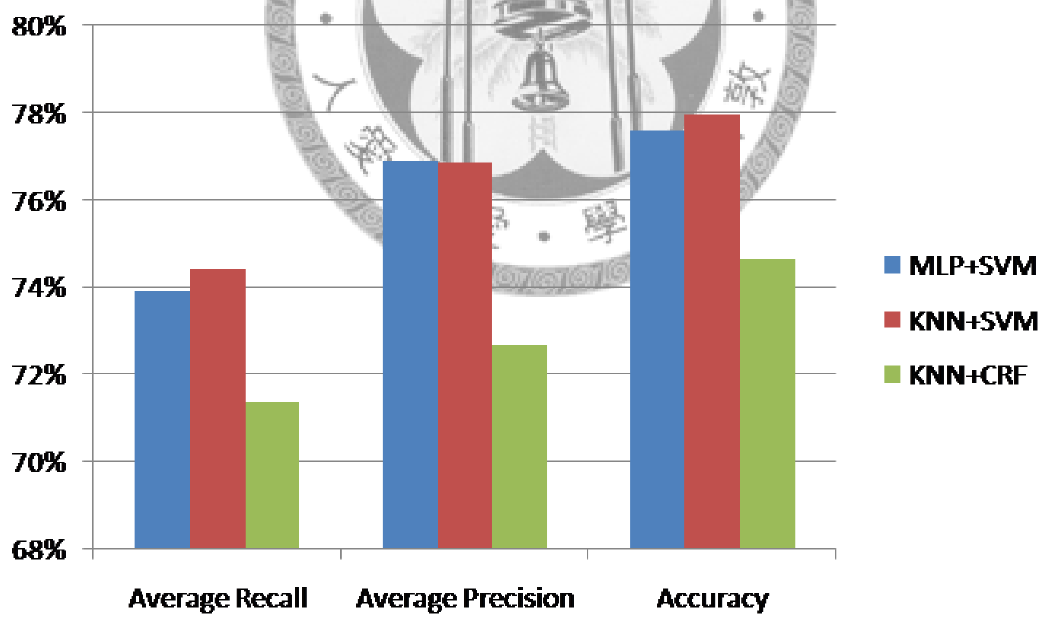

Figure 5.4: Comparison of three methods for window-based analysis.

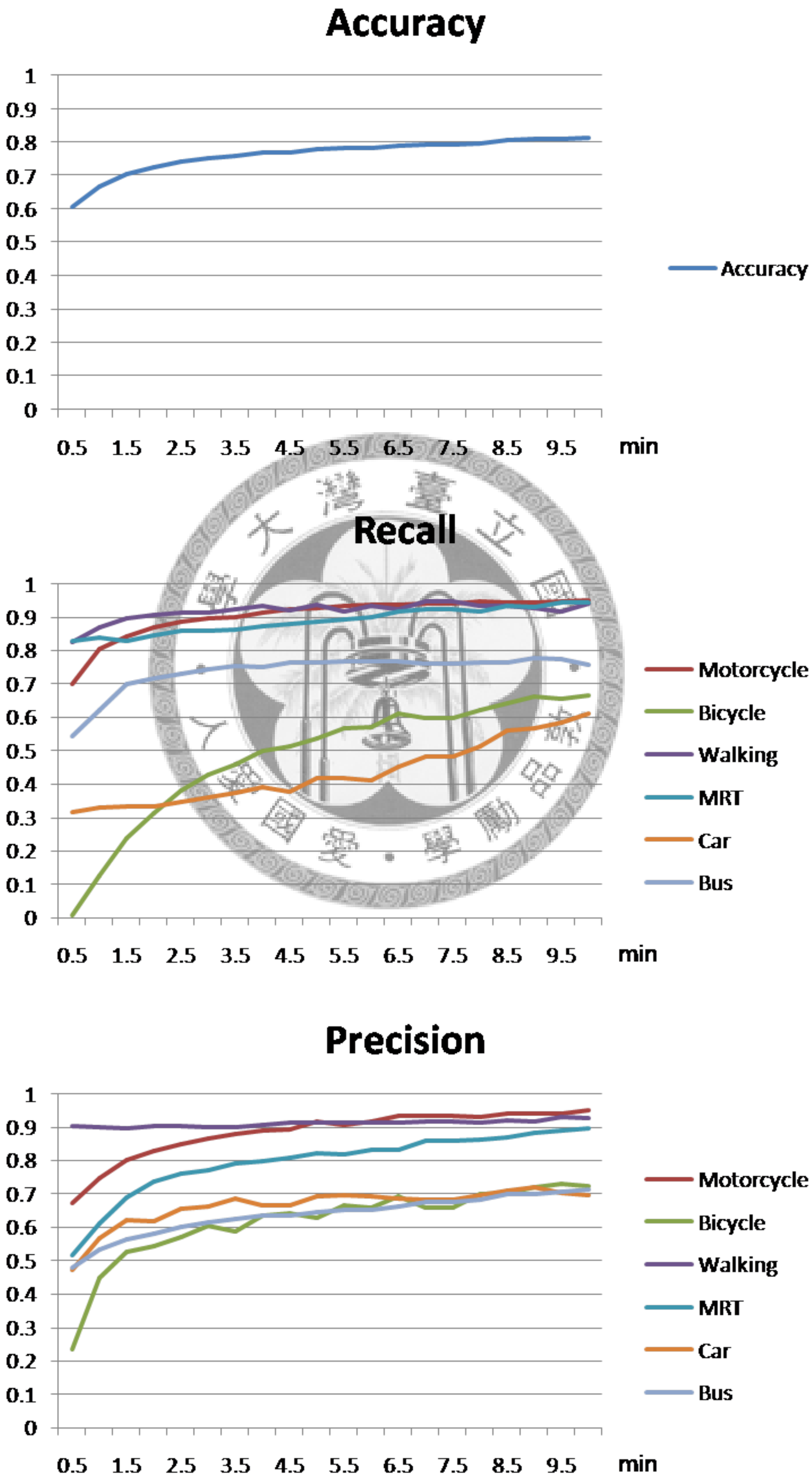

Figure 5.5: The accuracy, recall, and precision of inference with different window sizes.

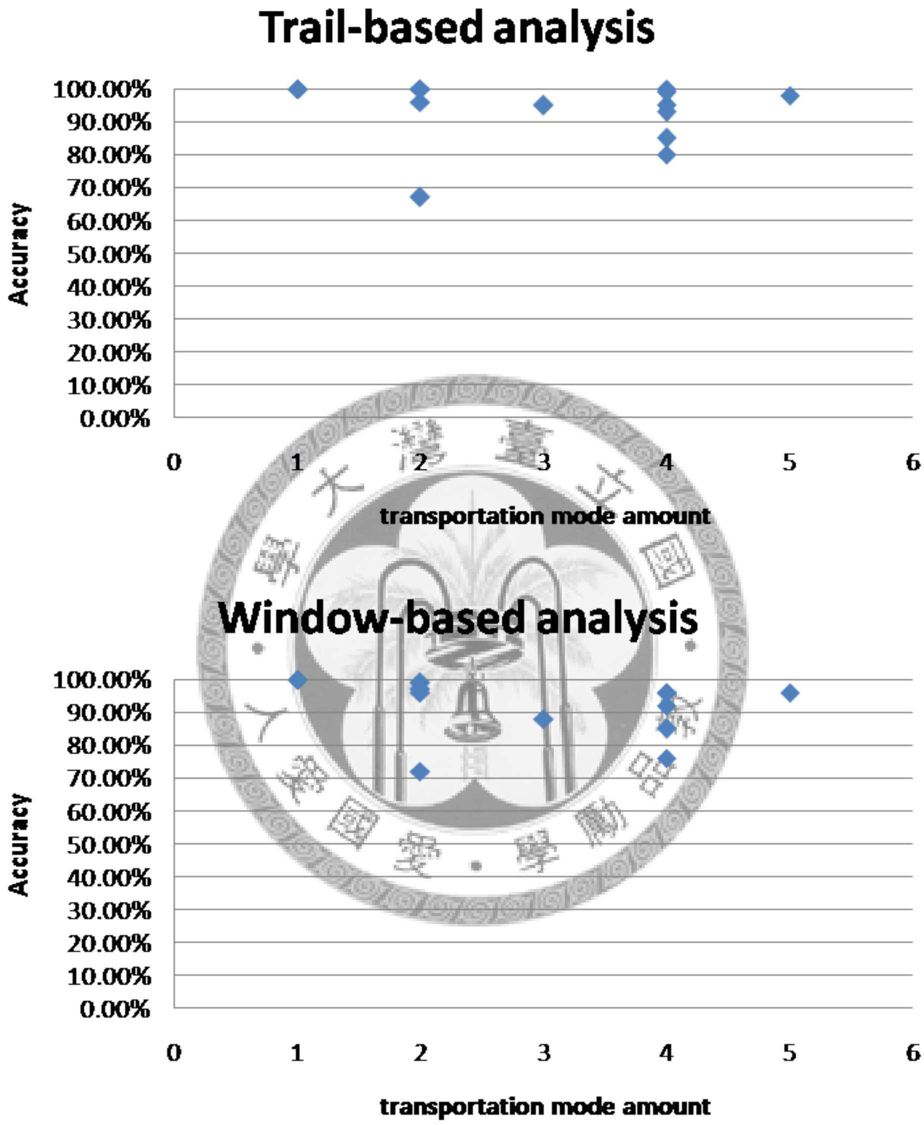

Figure 5.6: The accuracy and transportation mode amount.

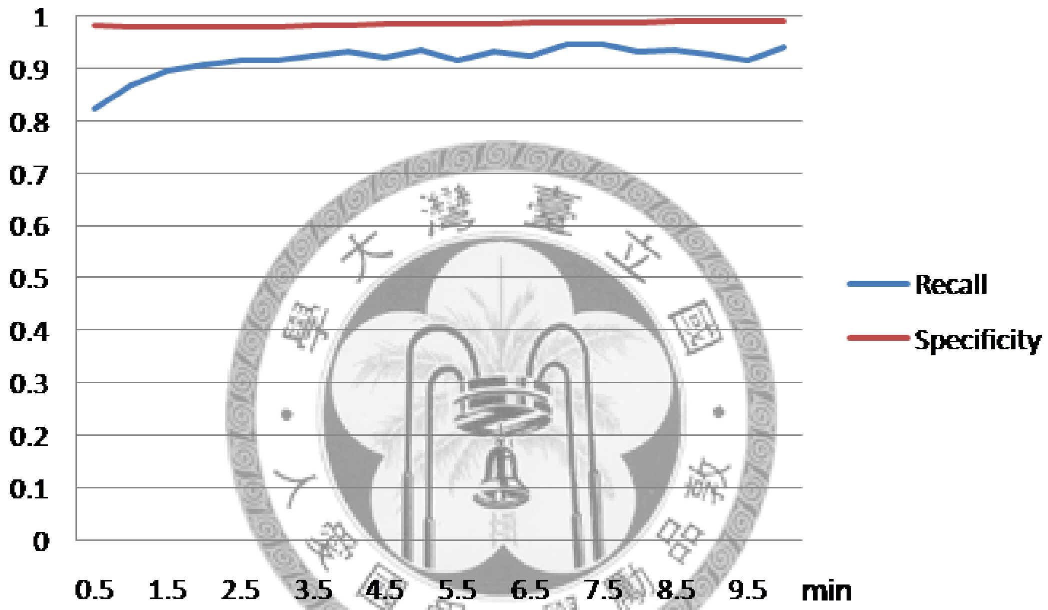

Figure 5.7: The recall and specificity of walking with different moving window size for transportation transit detection.

Chapter 6

Conclusion

The thesis explores the work aiming to infer transportation modes using mobile phone with accelerometer. For user to use in daily life, we designed the system to meet some limitation, including no researcher supervision, power-saving, system requirement reducing, and privacy issue. Our method considers both commonsense constraint of transportation infrastructure and regular user behavior on carrying mobile phone. In the pre-processing, we introduces the de-orientation to solve the orientation problem for mobile phone place. Without labeling from user, we built the samples by two different methods, researcher's handy labeling with histogram properity and machine's clustering, for the vibration, which was caused by both user action and vehicle motion, Then, using two different feature selection and relabeling techniques, multi-layer perceptron using histogram bin as feature and KNN with MDPA as similarity function, we constructed layer and extracted discriminable pattern from it for transportation mode inference. Finally, two different feature selections and classifiers, SVM using vibration type occurance probability as feautre and CRF using vibration type as feautre, were used for the transportation mode classification.

We held 4 experiments on 8 people for 2 months and 9 people for 1 month for retrieving the accelerometer data. Among multiple approach combinations, the combination KNN using MDPA as similarity function in first layer and SVM using vibration type occurance probability as feautre in second layer outperformed than others. Evaluated by trail-based 10-fold cross validation, we got the 89% and 78% accuracy for post-hoc trail-based analysis and the real-time window-based analysis respectively. To understand how short for initial time of real time detection, we evaluated data with different moving window size. The accuracy remains 71% when the initial time decreases to 1.5 minutes. For the transportation transit, another detector based on walking model was built and evaluated.

6.1 Summary of Contribution

Our approach was relabeling the pre-processed data to select new features for physical vibration. Then we did supervised learning on the vibration type with subjects' labeled transportation modes. Using the accelerometer logs collected by 8 people for 2 months and 9 people for 1 month, we evaluated our approach via a set of experiments. To adapt to different scenarios, we analyzed the post-hoc trail-based processing and real-time window-based processing respectively. Based on two-layer relabeling approach and SVM inference model, we inference motorcycle, bicycle, walking , MRT, car, and bus with accuracy 89% for trail-based analysis and 78% for window-based analysis respectively. Furthermore, to better understand the transportation transferring, another detector based on walking model was built and evaluated. Without general loss of accuracy (recall 82%), the detecting time shrank to 30 seconds. With above results, we believe our approach can be used in real daily life.

6.2 Limitation

For now, our system is far from perfect. There are two limitations for user to use in real life. First, the mobile carried by user should be placed in the pocket and shouldn't be placed in the bag or package. The system still works when other objects are put in the pocket with mobile phone. But larger space of bag or package largely increase the difficulty to recognize the vibration absorption by other objects, like a jacket. Second, the longer initial time is required for resl time detection. Even we lower the standard for accuracy to 71%, the initial time is still 1.5 minutes.

6.3 Future Work

There are many places needed to improve in the future. Our current system focuses primarily on accuracy of vibration type pattern for transportation classification. The property of sequential relation between vibration types is not sutdied right now. It is a great indicator and provides more content on time series relation. Understanding time series relation will not only increase the accuracy and the response time but also bring us more action detail during transportation, such as waiting for the traffic light.

Bibliography

[1] S. S. Al-Fedaghi. How to calculate the information privacy. In *The Third Annual Conference on Privacy, Security and Trust*, 2005.

[2] E. Arroyo, L. Bonanni, and T. Selker. Waterbot: exploring feedback and persuasive techniques at the sink. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 631–639, New York, NY, USA, 2005. ACM.

[3] D. Ashbrook and T. Starner. Using gps to learn significant locations and predict movement across multiple users. *Personal Ubiquitous Comput.*, 7(5):275–286, 2003.

[4] L. Bao and S. S. Intille. Activity recognition from user-annotated acceleration data. *Pervasive 2004*, pages 1–17, April 2004.

[5] C. V. Bouten, K. T. Koekkoek, M. Verduin, R. Kodde, and J. D. Janssen. A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *IEEE Trans Biomed Eng*, 44(3):136–147, March 1997.

[6] T. Brezmes, J.-L. Gorricho, and J. Cotrina. Activity recognition from accelerometer data on a mobile phone. In *IWANN '09: Proceedings of the 10th International Work-Conference on Artificial Neural Networks*, pages 796–799, Berlin, Heidelberg, 2009. Springer-Verlag.

[7] S.-H. Cha and S. N. Srihari. On measuring the distance between histograms. *Pattern Recognition*, 35(6):1355–1370, June 2002.

[8] Y.-C. Chang, J.-L. Lo, C.-J. Huang, N.-Y. Hsu, H.-H. Chu, H.-Y. Wang, P.-Y. Chi, and Y.-L. Hsieh. Playful toothbrush: ubicomp technology for teaching tooth brushing to kindergarten children. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 363–372, New York, NY, USA, 2008. ACM.

[9] T. Choudhury, G. Borriello, S. Consolvo, D. Haehnel, B. Harrison, B. Hemingway, J. Hightower, P. P. Klasnja, K. Koscher, A. LaMarca, J. A. Landay, L. LeGrand, J. Lester, A. Rahimi, A. Rea, and D. Wyatt. The mobile sensing platform: An embedded activity recognition system. *IEEE Pervasive Computing*, 7(2):32–41, 2008.

[10] W. Dong, Z. Wang, M. Charikar, and K. Li. Efficiently matching sets of features with random histograms. In *MM '08: Proceeding of the 16th ACM international conference on Multimedia*, pages 179–188, New York, NY, USA, 2008. ACM.

[11] C. Frank, P. Bolliger, F. Mattern, and W. Kellerer. The sensor internet at work: Locating everyday items using mobile phones. *Pervasive Mob. Comput.*, 4(3):421–447, 2008.

[12] J. Froehlich, T. Dillahunt, P. Klasnja, J. Mankoff, S. Consolvo, B. Harrison, and J. A. Landay. Ubigreen: investigating a mobile tool for tracking and supporting green transportation habits. In *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*, pages 1043–1052, New York, NY, USA, 2009. ACM.

[13] N. Györbíró, A. Fábián, and G. Hományi. An activity recognition system for mobile phones. *Mob. Netw. Appl.*, 14(1):82–91, 2009.

[14] B. Hoh, M. Gruteser, R. Herring, J. Ban, D. Work, J.-C. Herrera, A. M. Bayen, M. Annavaram, and Q. Jacobson. Virtual trip lines for distributed privacy-preserving traffic monitoring. In *MobiSys '08: Proceeding of the 6th international conference on Mobile systems, applications, and services*, pages 15–28, New York, NY, USA, 2008. ACM.

[15] D. N. Joanes and C. A. Gill. Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(1):183–189, 1998.

[16] J. Kao, R. Lam, G. Pong, T. Talukdar, and J. Wong. Ecorio: Track your mobile carbon footprint. http://www.ecorio.org/.

[17] W. Karim. The privacy implications of personal locators: Why you should think twice before voluntarily availing yourself to gps monitoring. *Washington University Journal of Law & Policy*, 14:485–515, 2004.

[18] J. Krumm and E. Horvitz. Predestination: Inferring destinations from partial trajectories. In *In Ubicomp*, pages 243–260, 2006.

[19] T. Kudo. Crf++: Yet another crf toolkit, December 2007.

[20] C.-H. J. Lee, L. Bonanni, J. H. Espinosa, H. Lieberman, and T. Selker. Augmenting kitchen appliances with a shared context using knowledge about daily events. In *IUI '06: Proceedings of the 11th international conference on Intelligent user interfaces*, pages 348–350, New York, NY, USA, 2006. ACM.

[21] L. Liao, D. Fox, and H. Kautz. Location-based activity recognition using relational markov networks. In *In: Proceedings of the Nineteenth International Conference on Artificial Intelligence, IJCAI'05*, 2005.

[22] L. Liao, D. Fox, and H. Kautz. Extracting places and activities from gps traces using hierarchical conditional random fields. *Int. J. Rob. Res.*, 26(1):119–134, 2007.

[23] L. Liao, D. J. Patterson, D. Fox, and H. Kautz. Learning and inferring transportation routines. *Artif. Intell.*, 171(5-6):311–331, 2007.

[24] L. I. N. Liao, D. J. Patterson, D. Fox, and H. Kautz. Building personal maps from gps data. *Annals of the New York Academy of Sciences*, pages 249–265, December 2006.

[25] H. Lu, W. Pan, N. D. Lane, T. Choudhury, and A. T. Campbell. Soundsense: scalable sound sensing for people-centric applications on mobile phones. In *MobiSys '09:*

Proceedings of the 7th international conference on Mobile systems, applications, and services, pages 165–178, New York, NY, USA, 2009. ACM.

[26] P. Mohan, V. N. Padmanabhan, and R. Ramjee. Nericell: rich monitoring of road and traffic conditions using mobile smartphones. In *SenSys '08: Proceedings of the 6th ACM conference on Embedded network sensor systems*, pages 323–336, New York, NY, USA, 2008. ACM.

[27] M. Mun, S. Reddy, K. Shilton, N. Yau, J. Burke, D. Estrin, M. Hansen, E. Howard, R. West, and P. Boda. Peir, the personal environmental impact report, as a platform for participatory sensing systems research. In *MobiSys '09: Proceedings of the 7th international conference on Mobile systems, applications, and services*, pages 55–68, New York, NY, USA, 2009. ACM.

[28] M. Park and Y. Gao. Error analysis and stochastic modeling of low-cost mems accelerometer. *J. Intell. Robotics Syst.*, 46(1):27–41, 2006.

[29] G. research. Pc vendors eye smartphone market: Report.

[30] I. Rish. An empirical study of the naive bayes classifier. In *IJCAI-01 workshop on "Empirical Methods in AI"*.

[31] T. S. Saponas, J. Lester, J. E. Froehlich, J. Fogarty, and J. A. Landay. ilearn on the iphone: Real-time human activity classification on commodity mobile phones. Technical Report UW-CSE-08-04-02, Human-Computer Interaction and Design, Department of Computer Science, University of Washington, 2008.

[32] F. Siegemund, C. Floerkemeier, and H. Vogt. The value of handhelds in smart environments. *Personal Ubiquitous Comput.*, 9(2):69–80, 2005.

[33] T. Sohn, A. Varshavsky, A. Lamarca, M. Chen, T. Choudhury, I. Smith, S. Consolvo, J. Hightower, W. Griswold, and E. de Lara. Mobility detection using everyday gsm traces. pages 212–224. 2006.

[34] J. C. Sprott. Some simple chaotic jerk functions. *American Journal of Physics*, 65:537–543, June 1997.

[35] E. Troshynski, C. Lee, and P. Dourish. Accountabilities of presence: reframing location-based systems. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 487–496, New York, NY, USA, 2008. ACM.

[36] M. Weiser. The computer for the 21st century. pages 933–940, 1995.

[37] I. Witten and E. Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, March 2005.

[38] T. Zhang, J. Wang, P. Liu, and J. Hou. Fall detection by embedding an accelerometer in cellphone and using kfd algorithm. *IJCSNS International Journal of Computer Science and Network Security*, 6(10):277–284, 2006.

[39] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma. Understanding mobility based on gps data. In *UbiComp '08: Proceedings of the 10th international conference on Ubiquitous computing*, pages 312–321, New York, NY, USA, 2008. ACM.

[40] Y. Zheng, L. Liu, L. Wang, and X. Xie. Learning transportation mode from raw gps data for geographic applications on the web. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 247–256, New York, NY, USA, 2008. ACM.