A Smartphone Based Risk Estimation of Human Activities

Wang Weiqiang
Marquette University

A SMARTPHONE BASED RISK ESTIMATION OF HUMAN ACTIVITIES

By

Weiqiang Wang

A Thesis submitted to the Faculty of Graduate School,
Marquette University,
in Partial Fulfillment of the Requirements for
the Degree of Master of Science

Milwaukee, Wisconsin

August 2012

ABSTRACT

A SMARTPHONE BASED RISK ESTIMATION OF HUMAN ACTIVITIES

Weiqiang Wang

Marquette University, 2012

As handheld devices have been booming up in recent years, the usage of laptops, smartphones, and tablets is increasing exponentially. More important, small handheld devices like smartphones and tablets are becoming more and more popular among people because of their portable and easy-to-use features. They use smartphones and tablets to do business and entertainment which, years ago, could only be accomplished by large personal computers and complicated software programs. Yet, while sensors like the gyroscope, accelerometer, light and magnetic are embedded into smartphones, how to best utilize sensors and improve lives of human beings by those sensors becomes a hot and valuable topic.

In this thesis, we try to monitor and document people`s daily activities by the gyroscope and accelerometer that were built inside the smartphone. And with the help of those data, we calculate how much activity is required or overdone for a subject in case of maintaining a healthy condition. More importantly, based on those data, we built a real time system that could not only judge what basic activity the subject is currently doing, but also estimate simple potential risks that might happen to the subject due to the abnormal data of the activity.

ACKNOWLEGEMENTS

Weiqiang Wang

This thesis was accomplished with continuous support from my Advisor, Thesis committee members, Lab mates, ‘Health Data Collegiate Challenge’ competition members and my family. First, I want to thank Dr. Sheikh Iqbal Ahamed, my Advisor, Director of Ubicomp Lab, for all the suggestions he gave me. He guided me the direction of my research, helped me implement application, and inspired me to confront all the difficulties and troubles during the research and thesis writing. I also want to thank thesis committee members Dr. Kaczmarek Thomas and Dr. Ge Rong for their valuable comments and advice on my thesis. Then, I want to give my thanks to my mates from Ubicomp Lab for their advice and help. I also want to thank members on the competition of ‘Health Data Collegiate Challenge’, for their brilliant ideas and help.

Finally, I am very grateful to my parents for their support and encouragement while I studying, researching, and writing this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHATPER 1: INTRODUCTION

According to statistics, until 2012, there are approximately 6.2 billion mobile subscribers in the world, which take up 87 percent of the whole population [1]. They use their mobile phone for banking, travel, shopping, news, video, sports and blogs and prefer applications for games, social media and maps. Meanwhile, since the sensor technology is growing tremendously, nowadays most smartphones have internal sensors such as light, sound, gyroscope, proximity, accelerometer, orientation and GPS (Global Positioning System). With the success of internal featured sensors, more and more functional apps are brought into smartphone. Yet, most of these apps are location, map pinpoint, voice control, and compass.

In recent years, the expenditures spent on healthcare in United States are almost exceeding 2 trillion [2]. People have various health issues, such as sedentary, lacking exercise, heart disease, high cholesterol, diabetes and so on. Also, in daily life, people have crash accidents that caused by negligence and unaware of risk action. This is very common in elder people, for example, suddenly moving fast might twist ankle, driving too long or unaware of speeding up might lead to car accident, and suddenly stopping while running or vice versa might cause harmful burden to heart.

Based on these ideas, this thesis is aiming at improving people's health and minimizing the risk possibilities that might happen in daily activities. We separate risk into two parts, one is health risk that could occur due to unhealthy habit and longtime improper activity. It is a chronic process and needs relatively longtime monitoring of the subject activities to draw a comprehensive and general conclusion of what change should the subject do to improve healthy condition. Another part of risk is ‘physical risk’. We named it ‘physical risk’ because it happens in real time and it is an opposite idea of health risk. For this part, the ‘risk’ happens quickly, fast and simultaneously when ‘risk behavior’ occur. Therefore, we first built a framework that could be used to easily monitor, get and record sensor data from smartphone. Then we record both accelerometer and gyroscope data in SD card. And we use a list to keep real time data of both gyroscope and accelerometer for a relative short period time. Based on these data, we built algorithm to differentiate basic activities such as walking, resting, running, and driving. After differentiating basic activities, we record the time of each activity for health calculation. Then we could draw a conclusion of how much exercise is needed for the subject to at least maintain current healthy status and not decrease. Also, it reminds the subject what problem might come out if the situation is not change. On the other hand, we built an algorithm that could estimate sudden fluctuation of sensor data, which might indicate a sudden change of activity. At this point, especially for elder people, we tried to estimate running, sharply turning, going backwards, and activities that suddenly change to other ones. And then, it reminds the subject there might be a potential risk due to the activity changing. The goal is prevent dangerous behavior and anticipate it before it happens. This is very helpful for elder people because their reaction to dangerous situations is not as quick as young people and their activity is more easily identified than younger ones whose activities are dramatically changing all the time. At the end, we tried to add more features in order to make the system perfect. We pull some weather data (if the smartphone support internet), and combine those data with risk algorithm to make more accurate judgment on risk estimation.

The outline of this thesis is: Chapter 1 states the generalized introduction of the whole thesis. Chapter 2 lists some background information which includes sensors in smartphone and some definition used in this thesis. Chapter 3 uses two scenarios to conclude the motivation. Chapter 4 states some related work. Chapter 5 illustrates the system and classification algorithms. Chapter 6 is about implementation of risk estimation. Chapter 7 is evaluation and Chapter 8 concludes some summary and future work.

CHAPTER 2: BACKGROUND

2.1 Sensors in Smartphone

Due to the fast advanced technology, sensors become more common in the smartphone than years ago. Light sensor, which is sensitive to the light changes in the smartphone, could be seen by selecting auto adjust display light. Orientation, which estimates which direction the phone is facing and heading to, could be identified when playing some games like teeter (a phone game that use sensor to keep balance). Magnetic sensor shows magnetic changing around the smartphone. Proximity sensor estimates whether the phone is near object or not, for example, if the phone is close to the users` face when picking up a phone, the screen is auto locked. Sound sensor will record decibel change around. For example, it could show the noise level of users working, resting environment. There are also pressure sensor, gravity sensor, rotation, temperature sensor, accelerometer and gyroscope. Among them, accelerometer and gyroscope are primarily used in our system to distinguish activities.

Accelerometer is an MEMS (Micro-electro-mechanical systems) product. It records the acceleration speed at three axis perpendicular to each other. It will return three real time values X, Y, Z.

The more intensive activity, the larger value it gets. And when smartphone is lying still, an accelerometer will only show gravity value along the vertical horizon.

Gyroscope is not a sensor that records the speed of movement, but records the acceleration speed of rotation. It will return three real time values that indicate the acceleration of rolling, whirling, and rotation. It could be used to judge relative location, direction and angle of the smartphone. And it is very popular used after deployed first in the IPhone 4.

2.2 Physical Risks

Physical risk – is a special kind of risk physically happens simultaneously in human activity. For example, twisting, falling, having a car accident and so on.

In this thesis, we divide activities into continuous time segment, which would make it easier to do feature extraction. Because activity within every segment would be seen as a repeated cycle, they have similar crest, trough, and mean value. Then we calculate the average value of each activity in one time segment, in order to find a smooth curve while the activity is going on. We try to find the maximal fluctuation for training data. And set limits of normal activity range. On this point, we assume that subject`s safe activity is within that maximal fluctuation, any incoming sensor data that beyond the limits would be taken as an indicator of potential risk. Because sensor data are collected dynamically, the mean value of data in one segment is changing dynamically. Also, the maximal fluctuation range and limits are changing dynamically. That means if the change between two activities is not quick and fast, there would be no indication of future risk. But if the change between two activities is intensive – short, quick, and violent, such as suddenly running when having rest, there would be a warning of potential risk. Because the limits in one time segment would not have enough time to adjust to new activity and subject`s violent movement would cause the real time incoming sensor data overpass that limits.

In real life it is very difficult to identify and define every kind of risk, especially those occur immediately because of unaware risk action. Therefore, we want to define some basic risks that could make this system running and have some feedback to improve. Here are five categories of physical risk we defined in our system, particularly for elderly people: suddenly speeding up; turning too fast; losing gravity; going backwards; driving. These categories are what we think might cause elder people to fall down, or tissue and muscle damage, such as twisting ankle. For the last category driving, there might be hundreds risk factors in it because of car accident are caused by numerous reasons. Since it is very complicated to identify driving risk by smartphone, we only try to identify the simplest risk such as driving too fast and driving too long.

2.3 Health Risks

Health risks – is a special kind of risk that relates to health condition of human beings. Such as, diabetes, heart disease, sedentary and so on.

METS – is a value that numerated activities. It is a number that could be used to indicate how much activity the subject did and to show relationship with healthy condition. With help of METS, it could relate different activities to different level of health.

In our system, we record different activity data in different file as a long time analysis document. And we calculate how long the subject has been on each activity. Then we transfer those activity data into MET value. From a METS table, we draw conclusion that whether the subject did enough activity to maintain health, or there might be potential risk that would harm subject`s healthy condition. There would be a graph showing that recent activities and MET value subject has received. And subject could see his or her healthy condition as well as related activities.

We also used documented sedentary time to indicate health problem because sedentary is the most popular health issue nowadays.

2.4 Definition and Additional Features

Activity classification: Classify human activities into various categories with the help of data mining.

Physical risk: Kinds of risk that occur in real time and might cause physical damage to the subject.

Physical risk estimation: Anticipate potential physical risk before it really happen.

Health risk: Kinds of risk that would harm subjects' health in future or decrease subjects' health status in a long time.

Health risk estimation: Predict the trend of future health status by current health condition and health related factors.

In our system, in order to make the estimation more accurate, we also put some additional features such as weather information and server side help analysis (if network available). We integrate these features into our system as plus factors that could better help us identify risk situation. As the sensor data is being recorded and analyzed, it raises concern for privacy. In our system, our app can be run in the phone without transferring any data. If the data needs to be uploaded, it is done with the consent of user. All information is recorded in the local SD card.

CHAPTER 3: MOTIVATION

Two simple scenarios are presented to illustrate physical risk, health risk and how our system aims at eliminating those risks.

3.1 Scenario 1

Tim, with his smartphone in pant pockets is jogging on the street. Then, it rains, and the phone alarms him for the current speed in rainy weather, which might have danger like slippery roads. Then he suddenly stops, and the phone sounds another alarm that tells him slow down more peacefully. He goes into his car and heads to dinner. After a while, the phone alarms the speed is too fast on the current road and weather situation. And after other few minutes driving, the phone reminds him that he drove a long time and it better to have a rest than continuous, tiring driving.

3.2 Scenario 2

Bob works every day in front of computer, setting more than 8 hours. His smartphone records sedentary time and forwards a warning to him that he needs exercise. He decides go to exercise in order to not let his health condition decrease. After some days, the phone shows him a new chart that how much activity he did and how much he need if he wants maintain health. The phone also shows indexes which indicate how much he needs to do to improve health. Then, Bob clearly knows his future trend health condition and knows what to do to improve.

3.3 Summary

Two scenarios stated above precisely describe how ‘physical risk’ and ‘health risk’ are identified by our system in real life. The ideas and goals of this system are to improve people’s health and help people avoid potential risk that might happen in daily life because of their unaware risk action or negligence of risk factor. It is more like an anticipation of potential risk, and our goal is make that anticipation of physical risk more precise and accurate.

As we can see from scenarios, health risk is much easier to identify because there are many existed health research studies and human activity research studies. Those works would help us to define health risk and find out relationship between health and human activities. But physical risk is a fresh new idea. For now, physical risk in our system only works better for elder people than for young people whose activity is more violent, intensive and hard to anticipate.

CHAPTER 4: RELATED WORK

4.1 Human Activity Classification

There are many research papers focusing on human activities classification. Most of them use external sensor device. In the research of Xi Long, Bin Yin and Ronald M. Aarts [3], they put an external tri-axial accelerometer on subject`s waist to record human activity data. They classify human activity into five categories: walking, running, driving, cycling and sports. They compare Bayes classification with Decision Tree classification to draw a conclusion that Bayes is better in that it has more extensible structure. In the research of Hamed Ketabdar and Matti Lyra [4], they use cell phone instead of external sensor device. They use Gaussian mixture models to classify human activity into four categories: walking, running, resting, and no activity. In the research of Manabu Nii, Kazuki nakai, Takayuki Fujita, and Yutaka Takahashi, they use database to store data collected from MEMS based monitoring system. And then, they use fuzzy rule based approach to define each activity behavior. The activity monitoring system is also based on an external device.

4.2 Risk Estimation

There is no current work that directly relates risk with various human activities, especially physical risk which we introduced into our system. Yet, there are some research papers about risk propensity and health risk corresponding to particular activity.

In the research of Nigel Nicholson, Mark Fenton-O`Creevy, Emma Soane, and Paul Willman [5], they study the relationship between risk propensity and personality. They divide risk taking factors into various classes such as age, sex, job level and so on. It is more like a statistic anticipation of risk propensity based on the subject`s personal information and characteristics.

On the other side, some applications on market have been working on maintaining health by monitoring physical activities. They focus only on one particular disease, such as heart disease or diabetes. And most of them require manual input of activities. But they are not about physical risk or real time human activity classification.

4.3 Summary

There are lots of research studies on human activity classification. Most of them use external sensor device bonded to subject. Others use cellphone to collect activity data. Then by complicated algorithm based on server side, they do classification. Although on algorithm level, they are good at classification, it is difficult to implement on cell phone as a real time monitor system because of their algorithm requires a large amount of activity data. Few research studies focus on health and physical risk relationship with human activity. There are some applications on market aiming at maintaining health by tracking daily activities. But they are not real time monitoring system and cannot anticipate potential physical risk.

Here is a comparing table showing characteristics of system we built and other researches and applications:

	Health Researches	Activity Researches	Application on Market	Our System
Smartphone Based	No	Yes	Yes	Yes
Feature Extraction	No	Yes	Yes	Yes
Activity Classification	No	Yes	Yes	Yes
Health Risk	Yes	No	Yes	Yes
Physical Risk	No	No	No	Yes
Real Time	No	Yes	No	Yes
Risk Prevention	Yes	No	No	Yes

Table 4.1 Comparison of human activity and risk research

CHAPTER 5: ACTIVITY CLASSIFICATION

5.1 Sensor Data Recording

5.1.1 Framework

In Android programming, if we want to operate on sensors or get sensor information from smartphone, there are several steps we have to follow. First, we implement an interface – 'sensorlistener', which is used to implement sensor operation. Then we need register what sensor we want to get information from or monitor the change. After that, we need choose which model to monitor sensor change (normal, fastest, UI, game). At last, we put the operations we want into the body of 'sensorchange' class.

In order to efficiently retrieve data from sensors, we built a library that could provide basic sensor operation on Android. In this library, it could list sensors detail information, like power, frequency, manufacture, and so on. It can also show all those sensor information and real time data on UI. It could record sensor data and write under directory of 'sdcard\Data\'. Each recorded file would be named by system current time. The format of the data file is set as 'time X value Y value Z value', While the format of file name is set as 'month-date, year hour-minutes-seconds'.

All data stored in file would be recorded with an indicator of system current time plus X, Y, Z value of sensors. Each time the application starts up or restarts, there will be a new file created in case file name duplicate or data overwriting. If the application is running all the time without stopping or closing, the data would still be record to one file until any output interruption operation.

In general, there are four methods in this library each responsible for listing sensors, showing sensor information, showing sensor value and recording sensor data. Any other classes in other projects could easily do these four operations by simply importing the library and call methods. Here is the API [6] we built:

Start(SensorManager sm, String sensor, String delay)	Start recording sensor data in the 'sdcard\Data\' directory, as frequency is delay.
ListAllSensor(SensorManager sm, TextView tv)	List all sensor information with name and version in tv.
ShowSensorInfo(SensorManager sm, TextView tv, String sensor)	Show details about sensor including power, accuracy and vendor in tv.
ShowValue(SensorManager sm, TextView tv, String sensor, String delay)	Show real-time value of sensor in tv, as delay frequency

Table 5.1 API of framework we built

Sensor(case insensitive)	accelerometer; orientation; gyroscope; light; temperature; pressure; proximity; magnetic
Delay(case insensitive)	normal; fastest; game; UI

Table 5.2 Sensor and delay explanation

5.1.2 Data Recording

After we built the framework that could help us easily retrieve data from sensors and record data in SD card, we added the library to our project. Then we call methods from library in order to collect sensor data from accelerometer and gyroscope. We set the delay frequency as normal, which means the time interval between two recordings is 200 milliseconds. Then, we kept record of system current time that corresponds to each recording. At last, we write all those information – system time, and value of X, Y, Z of gyroscope and accelerometer in SD card. The files stored in SD card are primarily used for static analysis, which requires a relative long time recording.

On the other hand, in order to keep track of short time change for dynamical analysis, we make several array lists to store information as system time, gyroscope data and accelerometer data. The length of these lists is set to 25 because within that each list would record 5 seconds continuous data, and we assume the subject could complete at least one cycle of activity. For example, no matter whether the subject is walking or running, we assume that in 5 seconds, he or she already finish a set of actions (step left foot – step right foot – step left foot).

Once the application is closed, all data in lists would be deleted. Yet, data in SD card would remain. Because for dynamical analysis, it would cause lots of more work and storage resources waste if we want keep record of every data. We only need a cycle of data for feature extraction. While for static analysis, the more information we get from subject the more accurate result would be. Therefore, we use two ways (mentioned above) to keep track of subject`s activities.

5.2 Activity Estimation

5.2.1 Algorithm

In activity classification, our system sets it up to five categories: rest; walking; fast walking; running; and fast running. From sensor data, driving can only be differentiated from rest by GPS or internet support. Therefore, our system does not include driving as basic activities stated here. In order to build the algorithm, we used smartphone (HTC g11) to record both accelerometer data and gyroscope data as training data. In our system, we set the delay frequency of sensor listener as normal, which is approximately recording every 0.2 second (based on testing result of our model phone). Then, we take 25 data as a cycle (that is 5 seconds) to do feature extraction, which means only the average value during every 5 seconds will be counted to calculated activity classification. In order to eliminate negative number influence, we use integration to the sum of three axis absolutely value. That is $\int_{t_0}^{t_1} |X|+|Y|+|Z|$ (we call result of this integration as energy to describe how much activity subject has done in a period of time), t_1 stands for current recording time and t_0 stands for previous recording time. Here are the steps of algorithm:

Step1. Create five lists e, t with length of 25;

Step2. Add system current time into list t; add the sum of absolute X, Y, and Z value to list e;

Step3. Calculate energy with the formula: energy=e(i)*(t(i+1)-t(i));

With the help of sensor energy, we could calculate how much violent activity the subject is doing despite direction and orientation. Also, we could judge whether the change from one activity to another is smooth enough to avoid potential risk by comparing average energy in a short time segment with real time energy.

5.2.2 Classification

In order to do feature extraction, we calculate the average value of sensor data fluctuations during the time segment choose before. Within those time segment, there are maximal and minimal values of each activity cycle. We try to find out limits of highest maximal value and lowest minimal value as a range of on activity fluctuation. And we define any data over pass those limits would indicate either an activity change or abnormal fluctuation which would refer to a potential risk.

With the help of data mining software tool (WEKA [7]), we build algorithm that classify human activities into five categories:

Step1. Calculate energy (algorithm we mentioned before);

Step2. Calculate average energy value in order to compare with limits

Step3. Check if GPS is available. If no, skip driving classification. If yes, retrieve speed data.

Step4. If the speed data is close to zero, then classify that activity to rest. If not, classify it into driving.

Step5. Compare average energy with multiple limits. Judging which limits the average value within. And then classify activity to corresponding category.

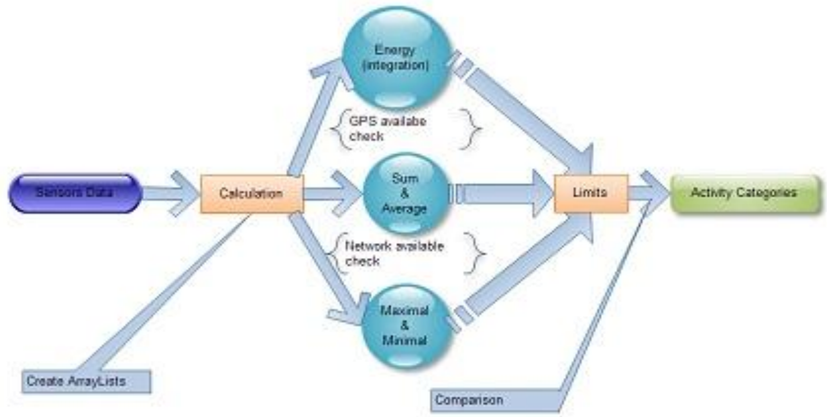


Figure 5.1 Process of activity classification

CHAPTER 6: RISK ESTIMATION

6.1 Physical Risk Estimation

6.1.1 Definition

Although we mentioned before about what physical risk is, we would like to recall the definition again. For the reason that physical risk is a fresh new idea and it is very important to understand it in order to understand our system.

Physical risk: different from mental issue and health problem, it is a special kind risk that could happen in a very near future and cause physical damage to the subject. For example, running fast in rainy day may cause slipping. Slipping has the possibility happen at any time if subject keep running and may cause some damage when she or he fall down.

6.1.2 Risk Category

Because it is very difficult to collect physical risk data, we tried to set safe limits (all data within the limits would be seen as normal activity data) of each activity. As mentioned before, we use array lists to calculate safe limits in order to make sure the limits are changing real time while activity is changing. Whenever activity is changing smooth and fluent, there would be no data overpass that limits. Because the limits is calculated by the average, maximal and minimal value within a segment of time, smooth activity change such as from rest to walking would not let sensor data overpass the limits. However, once the data from the sensor is beyond that limits, we consider it is showing a risk behavior or indicating a potential risk. In our system, we defined five types of 'physical risk': suddenly speeding up, turning too fast, losing gravity, driving speed and time and backwards.

Suddenly speeding up

This is determined when there is a surge of accelerometer energy during walking or rest. Because we assume that whatever activities the subject is doing, there should be a warm-up process. For example, when a subject wants running, he or she should go from rest to walking, then to jogging, and last to running. A sudden change from a low speed activity to a high speed activity is not allowed in our system, such as from rest or walking to fast running. The principle of change from high speed activity to low speed activity is the same. Although the surge of the data might happen due to other intensive activity, we include them all in this type. Therefore, sudden activity changes, like violent movement when in a low energy activity are taken as this particular risk type. On the other hand, the safe limits are first determined at the same time as activity classification and then calculated real time. The algorithm is:

Step1. Calculate accelerometer energy value by integration. And calculate average energy value in a period time;

Step2. With the help of activity classification, we could know what the current activity is by judging average energy;

Step3. Put the real time energy value (which is the last value from list) into the classification algorithm. If it comes out another activity that is more violent than the one we identify in step2, the risk condition is satisfied.

Turning too fast

This is much more like speeding up fast. The difference is for this physical risk type, we use gyroscope to identify sharply increased energy. Because we assume that no matter how subject puts his or her phone in pant pockets, the axis Z is always vertical to his or her leg. And when he or she is turning, the X value of gyroscope would change tremendously. For example, if a subject is circling in a place, the accelerometer energy would show very little difference between that activity and rest. But gyroscope could better identify that activity. So we use gyroscope to estimate if a subject is turning or circling violently, for which we assume would burden and risk elder people`s muscle and tissue health. The algorithm is similar to speeding fast, while a difference is that we use gyroscope data instead of accelerometer.

Losing gravity

This type of risk might relate to falling or jumping down. Because even when a subject is static, sensor still has data that indicates gravity value. Once the energy is far less than gravity value or getting close to zero, it indicates that the subject is losing gravity. In our system, we use this risk type to estimate falling down and warn elder people for jumping. The algorithm is very simple:

Step1. Calculate real time energy value of accelerometer

Step2. Compare that real time value with rest energy limits. If that value drops below rest energy limits and close to zero, then condition is satisfied.

Driving speed and time

As stated before, driving is similar with rest based on data of both gyroscope and accelerometer. Because when the subject is driving, the data of gyroscope and accelerometer looks like rest except a little fluctuation when speeding up and speeding down. Therefore, we use GPS or network as a plus to judge if there is any speed of the subject. If yes, then the activity is in driving type. If not, we classify it as rest. Then we monitor the speed and record the time. As either driving too long or driving too fast is very dangerous. We also calculate safe limits for driving activity, and these limits would be narrower and more restrictive since any little sensor data change indicates a much more violent change on driving. The algorithm is:

Step1. Check if GPS available. If yes, get speed data from GPS.

Step2. Judging from that speed data, identify if current activity is driving. If yes, start recording time.

Step3. Check if network available. If yes, get weather data.

Step4. Compare time data, speed data and weather data with pre-setting limits to see if any risk condition satisfied.

Backwards

This is the most difficult type among these five. Because smartphone may be put in different direction, orientation in pant pockets, and integration from absolute accelerometer data or gyroscope data is very hard to tell whether the subject is forward or backward. We used feature extraction within time segment, and kept the sign in front of sensor data. If there is a number of opposite sign to average value (calculated during time segment), and if the absolute value of that number is larger than the absolute value of any number of the same sign, we conclude that there is a backwards or at least there is a trend of backwards. Here is the algorithm we used to estimate backwards:

Step1. Use lists x, y, z to store accelerometer data X, Y, Z,

Step2. Use lists maxx, maxy, maxz, minx, miny, minz to store the maximal and minimal value of x, y, z,

Step3. By comparing (maxx-minx), (maxy-miny), (maxz-minz), identify along which axis the activity has highest fluctuation,

Step4. After finding the highest fluctuation axis, calculate sum value and absolute sum value,

Step5. Compare the last value of the highest fluctuation list with calculation result. If absolute value of the last value is larger than the average of absolute sum, and the sign of it is opposite to the sign of sum, backwards condition is satisfied.

6.1.3 Integration

After we use activity classification to classify each type of risk, we call those risk algorithms of different risk types to judge whether risk condition is satisfied and which risk condition is satisfied. If the risk condition is satisfied, the phone would warn subject with sound and vibration as a reminder that he or she should watch what is going on. If the risk condition is not satisfied, the estimation algorithm would continue looping when application is running.

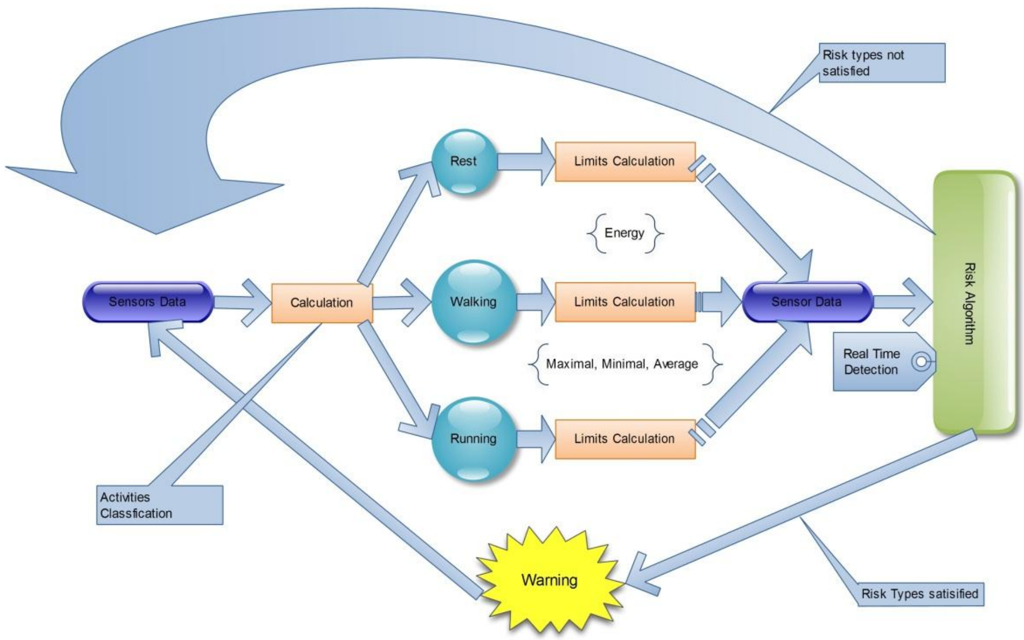


Figure 6.1 Process of physical risk estimation

6.1.4 Additional Features

Our classification of physical risk is not mature. To make it more precise and generalizes, we add weather condition in it. The thought is that, most activities are taken outdoor, which would be influenced a lot by weather condition. For example, running in rainy weather is totally different from running in sunny weather. And if the snow is too strong, it is better to stay indoors, instead of jogging, driving outside. Based on this idea, we first estimate if there is any network available on the smartphone. If no, we would skip the weather feature. If yes, we would go on to the next step. Our system would first find the longitude and latitude of the phone at current location, and then retrieve weather data from Google weather report. Based on the data of temperature, wind speed, and weather condition, we add the weather factors into risk factors.

However, it is very complicated if we want to add all weather features into risk factors. Because classification and definition of risk in our system is not strong enough to differentiate and distinguish various risk situation that could happen. Moreover, many phone users do not have or want internet usage. Therefore, we would take the weather feature as a plus to help risk estimation.

6.2 Health Risk Estimation

6.2.1 Definition

In the part of health risk, we are aiming at evaluating subject's health condition from daily activities and forward suggestions that whether the subject is within safe activity range to maintain health or beyond that range. In order to reach this goal, we use MET value as a numerical way to measure activity. "The metabolic equivalent of task (MET), or simply metabolic equivalent, is a physiological measure expressing the energy cost of physical activities and is defined as the ratio of metabolic rate (and therefore the rate of energy consumption) during a specific physical activity to a reference metabolic rate" [8]

6.2.2 Method

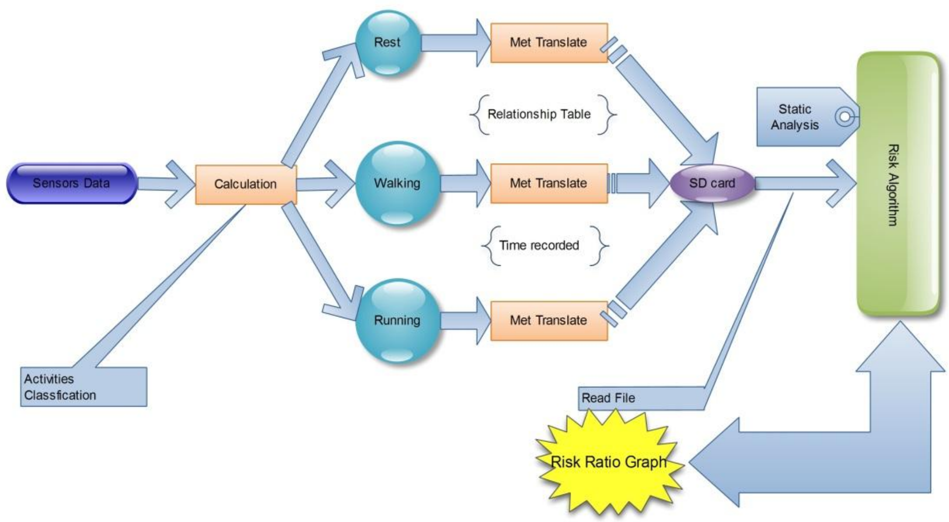


Figure 6.2 Process of health risk estimation

Every time we collect activity data from smartphone, we first classify activities. Then we transfer those classified activities into MET value by a specific MET-Relation table (from health science research). The table shows 600 kinds of activities and related MET value. Here is a table showing part of the relation table:

METS	Category	Examples
8.5	Bicycling	Bicycling, BMX or mountain

3.0	Conditioning Exercise	Bicycling, stationary, 50 watts, very light effort
4.8	Dancing	Ballet or modern, twist, jazz, tap, jitterbug
2.5	Fishing and Hunting	Hunting, duck, wading
6.0	Home Activities	Moving furniture, household items, carrying boxes
1.3	Miscellaneous	Sitting and reading, book, newspaper, etc.
12.0	Sports	Boxing, in ring, general
4.0	Volunteer Activities	Walk, run play with children, moderate, only active periods

Table 6.1 MET-Activity relationship

For our system, we only choose running, waking and rest, because our activity algorithm is not powerful enough to distinguish so many activities. We record daily running and walking activities and transfer them into MET value, then we calculate how many MET values the subject has acquired for one week. After that, we compare the MET value with a MET-Disease table to draw conclusions whether the subject did enough activity or not. Below is a chart showing relationship between activities and premature death (taken from U.S. Guidelines 2008):

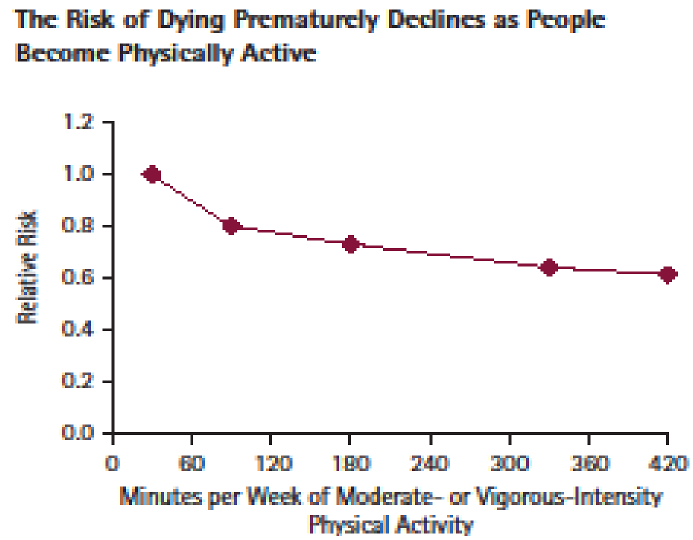


Figure 6.3 Relationship between premature death and activity (US guideline 2008)

On the other hand, we keep record of how long a subject is at rest. And we take the rest time as sedentary time (we assume that the phone user would turn off phone or close application when he or she is sleep). By counting the sedentary time, our system would draw a simple graph that indicate risk ratio of a subject based on his or her sedentary time. For example, the relationship table between sedentary and obesity is as below:

Time (Hours)	**Risk**
0	Lowest Risk
2	23% increase risk
4	46% increase risk
6	69% increase risk

Table 6.2 Sedentary and obesity relationship (US guideline 2008)

6.2.3 Additional Features

In our system, the automatically health risk estimation could only analyze on basic activities like rest, walking, running and driving. It is limited by the algorithm that we used to classify activities. In our system, we integrate the definition of other complex activities into those classified basic activities instead of building more categories.

In order to make the analysis results more precise and in case the subject does not carry phone, we try to add the manual input as another way to help adjust auto estimation error. We built a manual input UI to allow subject to choose what activities he or she did (from that 600 activities table) and input how long he or she did on that. Then we calculate that data, transfer into MET value and use them as an additional help to our automatically risk estimation. A subject can either choose to input his or her activities manually or let the application record or classify into basic activities automatically.

6.3 Summary

In this chapter, we introduce how our system works and divide it into four parts. First part is about how to record sensor data and the library we built to simplify retrieving and recording sensor data. In the second part we include the algorithms that distinguish basic human activities, and show some graphs of data collected from both accelerometer and gyroscope. The third part and fourth part are about the risk estimation. We try to illustrate the risk idea by separating it into physical risk and health risk. As to physical risk, we built a real time alarm system that estimates risk all the time while application is running. The system would warn five types risk derived from three to four basic activities.

On the other side, as to health risk, we built algorithms that translate basic activities into the MET value which is used as an indicator of health condition and relative risk ratio of health problems. In general, our system records sensor data in the SD card, classify human activity, and for each activity calculate potential risk that might happen to subject or harm subject`s health in the future.

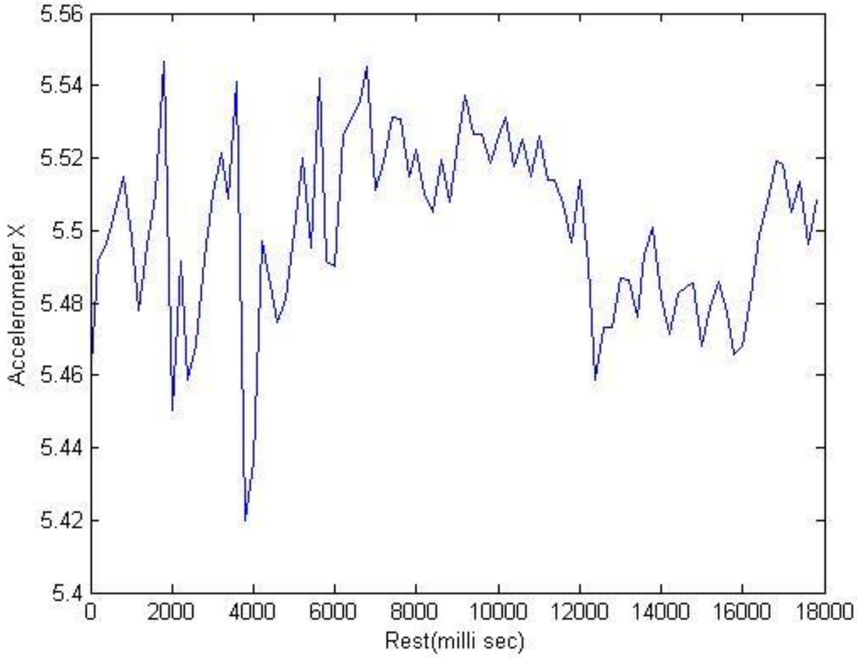
CHAPTER 7: EVALUATION

7.1 Experimental Setup

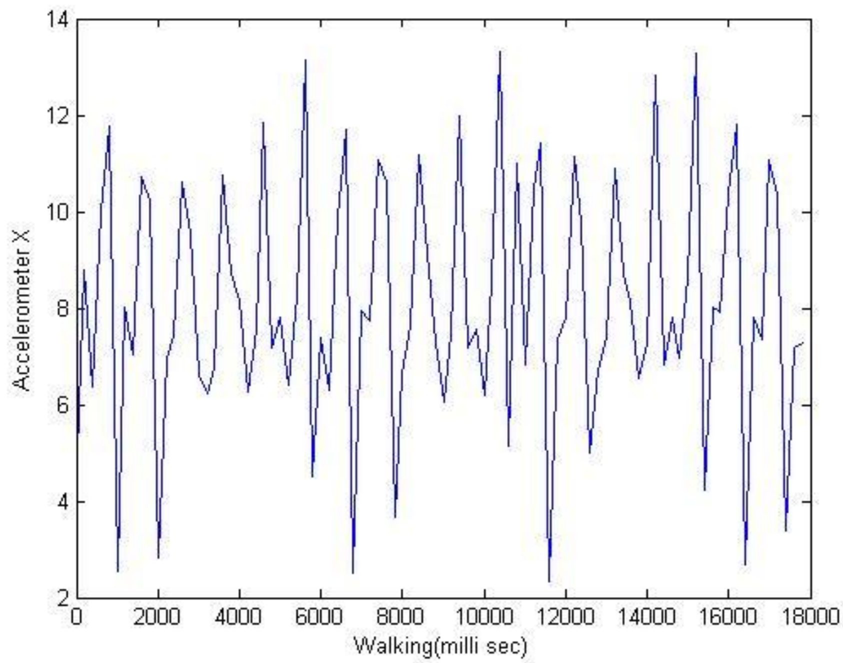
We found five different subjects for our system`s implementation. The model phone we use was HTC G11. Each subject was required to carry his or her phone in pants pocket. And we asked each subject to do six different activities: rest, walking, fast walking, running, fast running and driving. Subjects were asked to do at least twenty minutes on each activity. In the activity like rest, walking or running the subject was asked to imitate physical risk behaviors that we defined in our system, for example, going backwards, suddenly speeding up, and so on. Moreover, the subject was asked to carry the phone for a week in order to monitor his or her daily activities.

7.2 Activity Classification

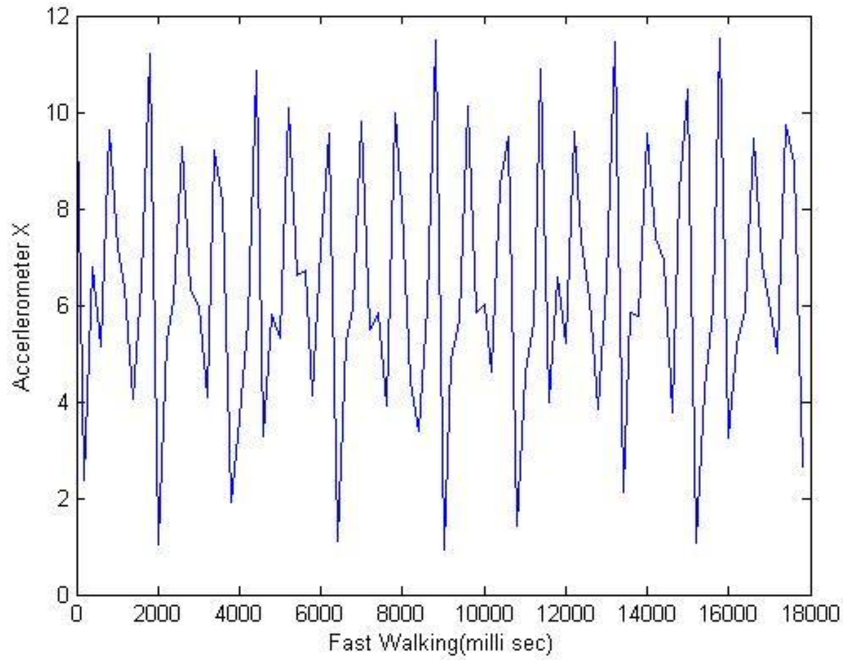
We collect training data of five activities: rest, walking, fast walking, running, fast running:


Accelerometer X
Rest(milli sec)
5.56
5.54
5.52
5.5
5.48
5.46
5.44
5.42
5.4
0
2000
4000
6000
8000
10000
12000
14000
16000
18000

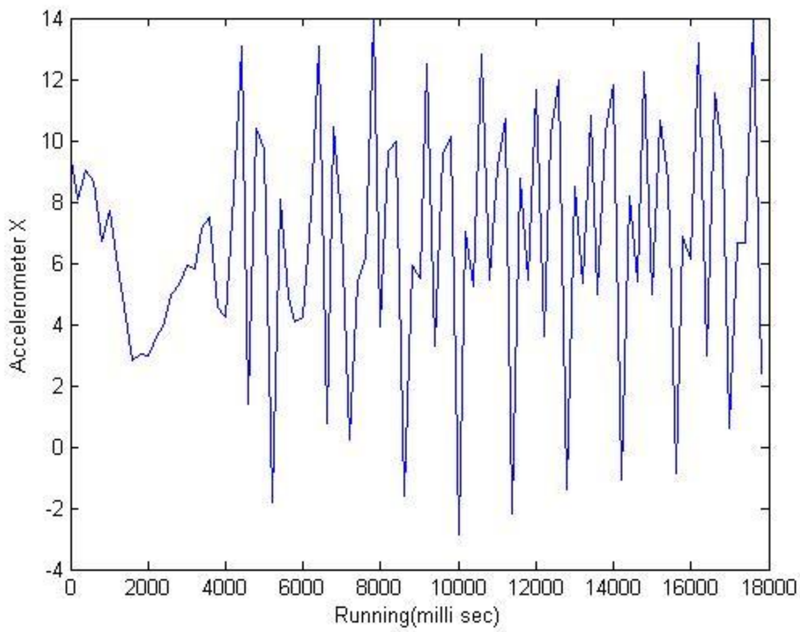
(a)


Accelerometer X
Walking(milli sec)
14
12
10
8
6
4
2
0
2000
4000
6000
8000
10000
12000
14000
16000
18000

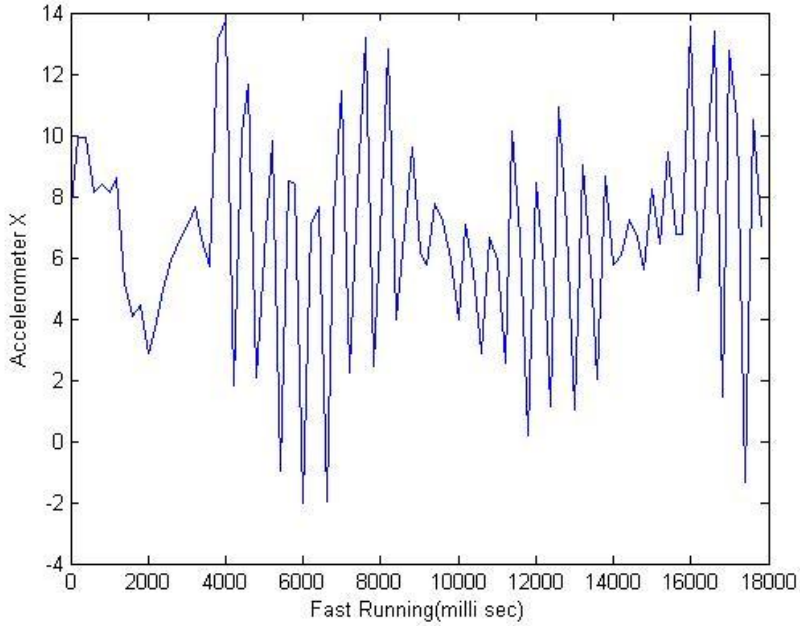
(b)


Accerlerometer X
Fast Walking(milli sec)
0
2
4
6
8
10
12
0
2000
4000
6000
8000
10000
12000
14000
16000
18000

(c)


Accelerometer X
Running(milli sec)
-4
-2
0
2
4
6
8
10
12
14
0
2000
4000
6000
8000
10000
12000
14000
16000
18000

(d)

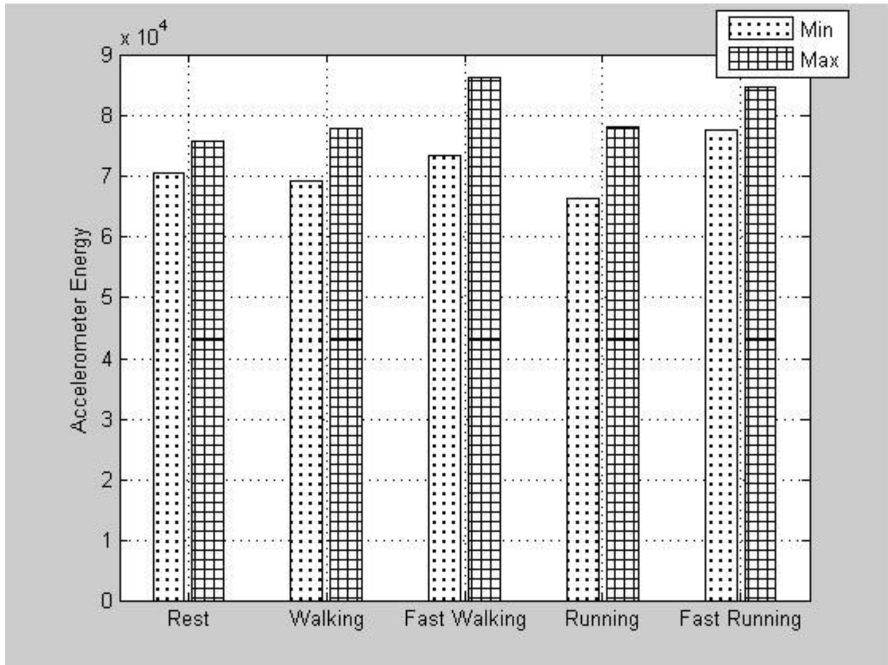


(e)

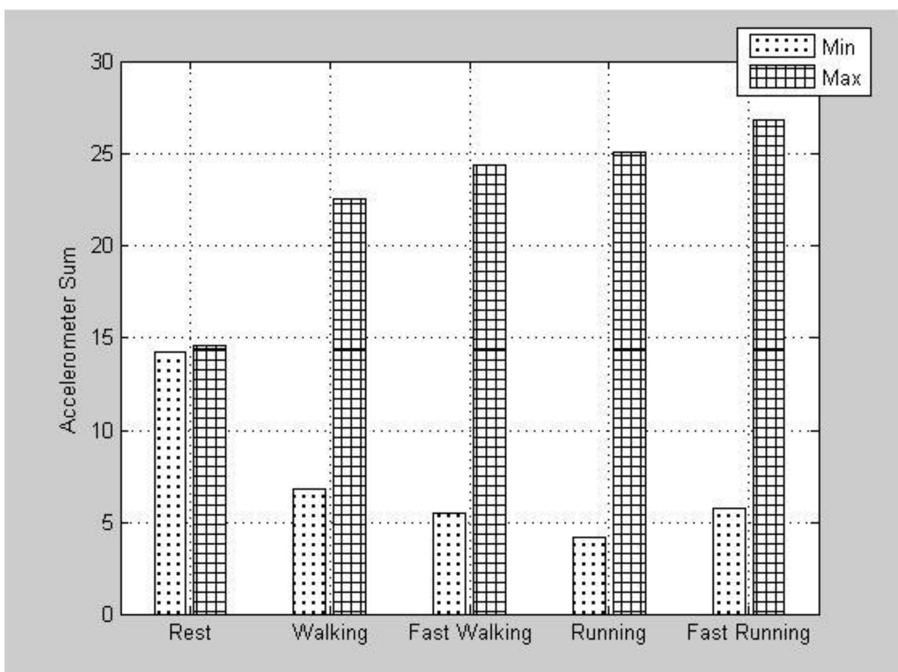
Figure 7.1 X value of accelerometer with different activities (a) Rest (b) Walking (c) Fast walking (d) Running (e) Fast running

From these five graphs we could see that maximal values are different for each activity; the ranges of fluctuation are different – the more violent activity is the bigger range of fluctuation it has. Yet, when we compare walking with fast walking, we find that they are very similar. And difference is only on the limit of positive and negative value.

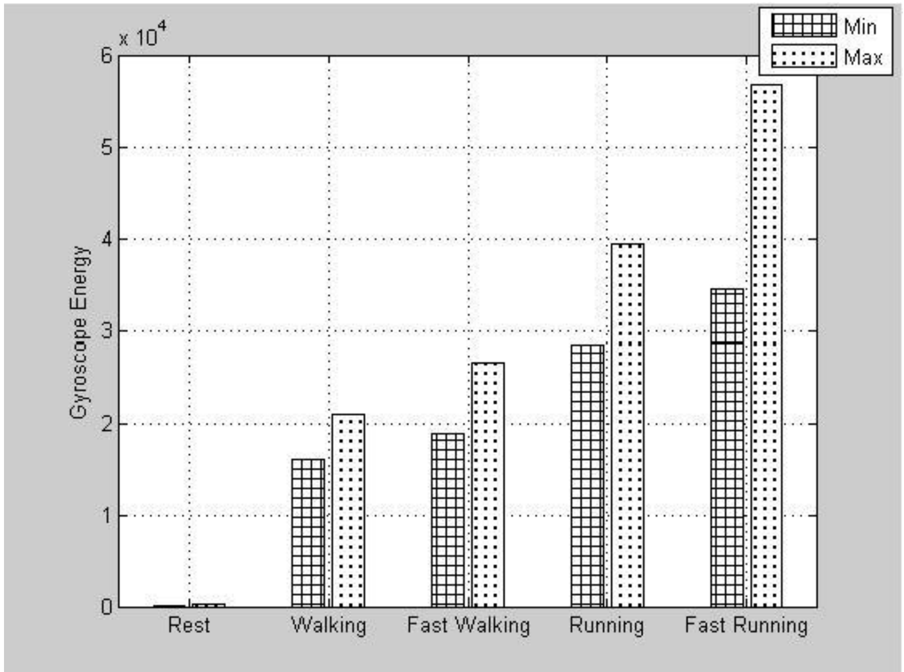
Then, we collect gyroscope data of these five activities in order to make a comparison with the accelerometer. Before the comparison, we calculate the sum of absolute value on X, Y, and Z axis. And we calculate the integration of real time sensor value. At last, we calculate the minimal and maximal value of both the sum and integration result.


Min
Max
x 10^4
9
8
7
6
5
4
3
2
1
0
Accelerometer Energy
Rest
Walking
Fast Walking
Running
Fast Running

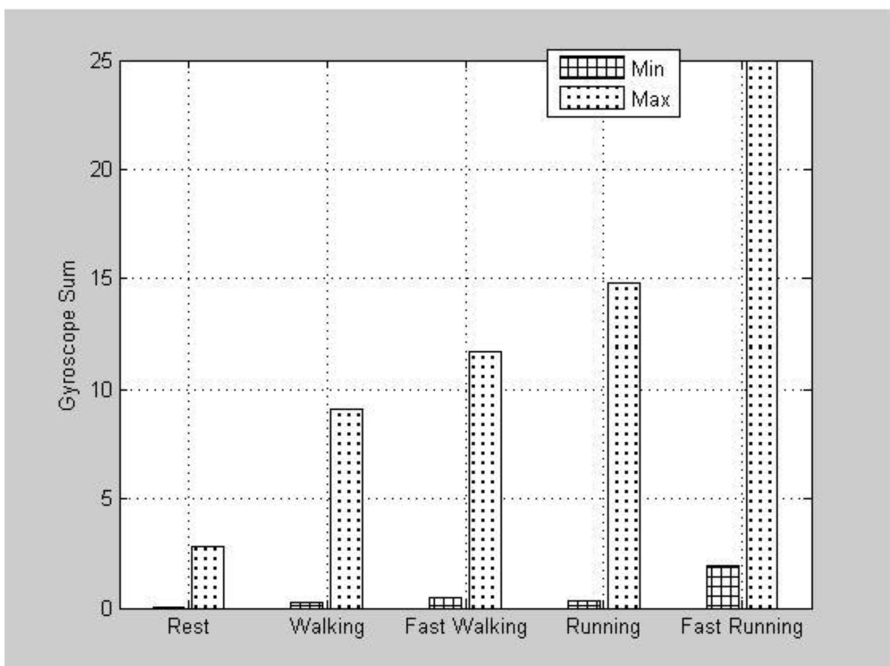
(a)


Min
Max
30
25
20
15
10
5
0
Accelerometer Sum
Rest
Walking
Fast Walking
Running
Fast Running

(b)



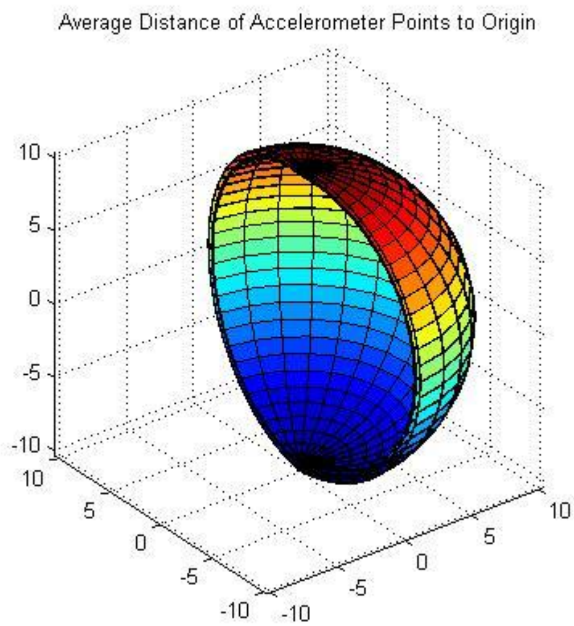
(c)



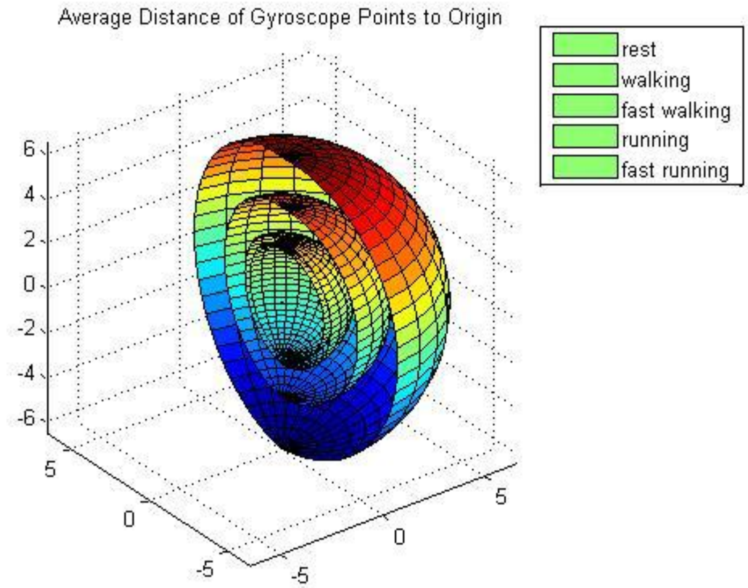
(d)

Figure 7.2 Activities comparison under different value (a) Accelerometer integration (b) Accelerometer sum (c) Gyroscope integration (d) Gyroscope sum

Judging from the graph above, we could see that there is an overlapping between the previous activity maximal value and the next minimal value in accelerometer. But in gyroscope, it seems there is a better classification at the integration value. Because data in the accelerometer and gyroscope both have X, Y, Z three values, we draw a three-dimensional graph to compare the difference among different activities.



(a)

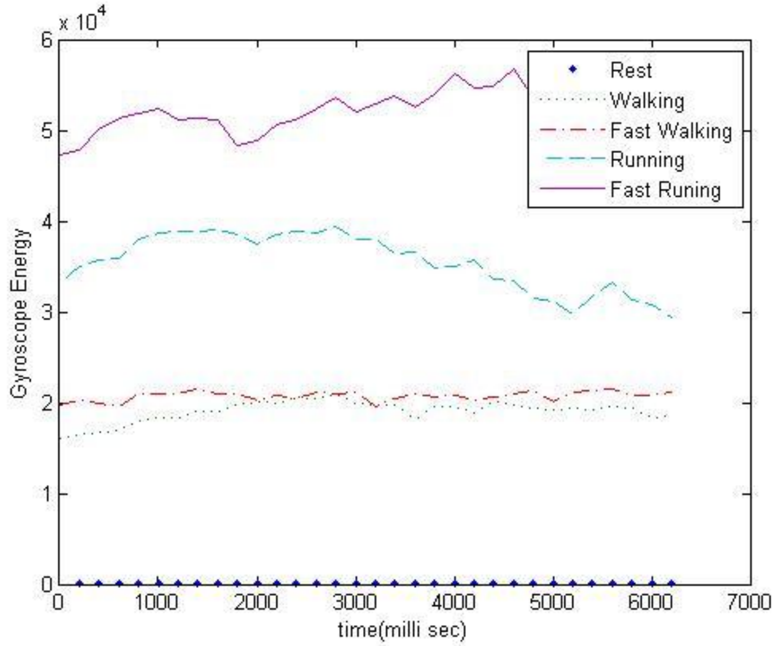


(b)

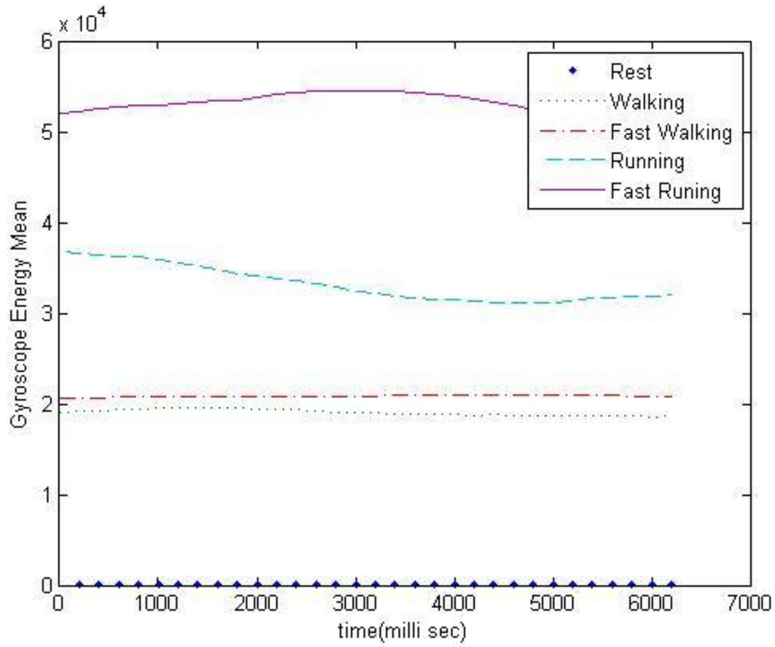
Figure 7.3 Gyroscope data and Accelerometer data on three-dimensional space of different activities (a) Accelerometer data (b) Gyroscope data

It is obviously that from this data we collected from our model phone for one subject, the gyroscope is better to identify different activities than the accelerometer. Therefore, we choose the gyroscope as the primary sensor used to identify and classify activities.

Below is a comparison of gyroscope energy (integration) before and after the average calculation:



(a)



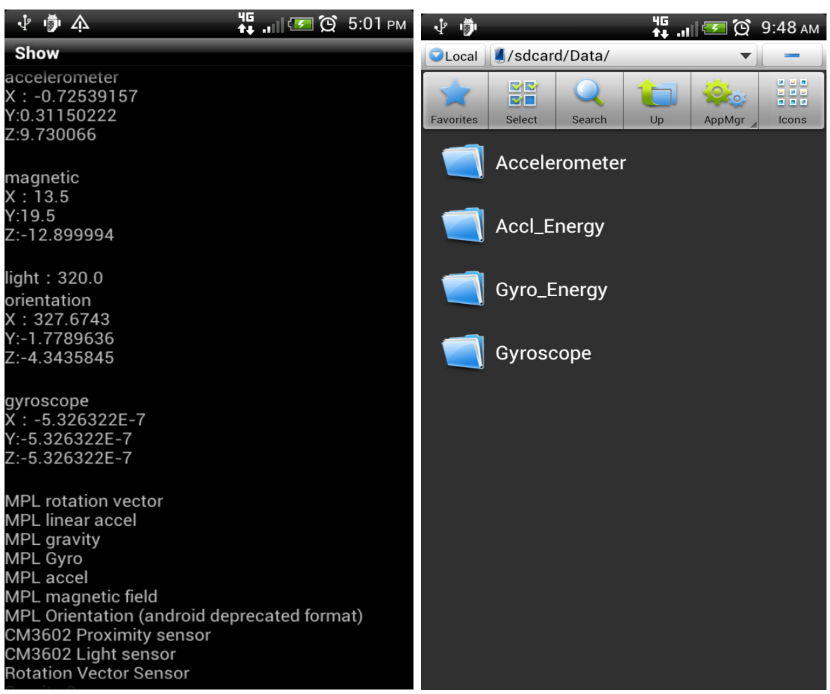
(b)

Figure 7.4 Gyroscope energy before and after average calculation (a) before (b) after

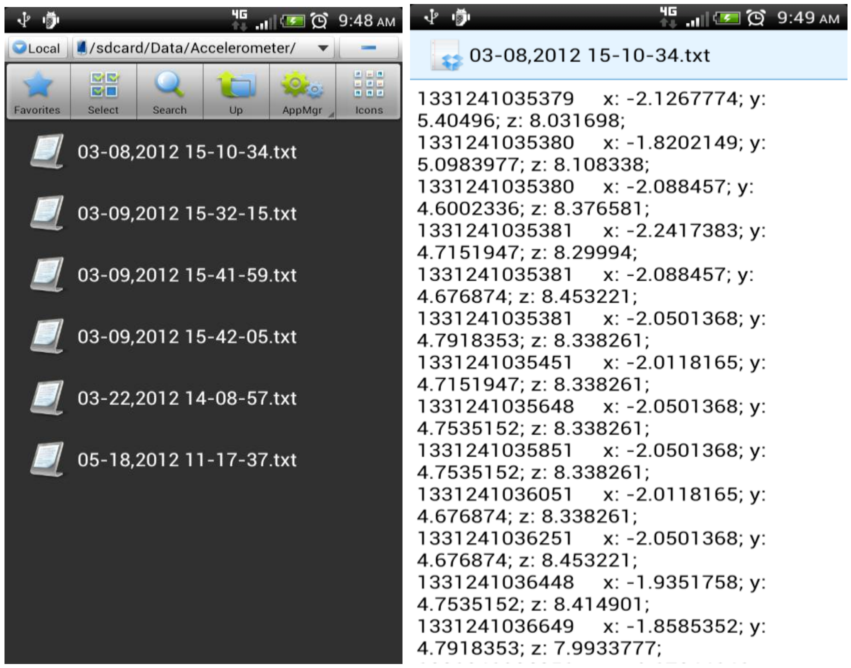
Before calculating the average energy, there are some cross and overlapping between activity walking and fast walking. After calculating average energy, the activity line becomes straighter without any cross or overlapping. The difference between walking and fast walking is intuitive and obvious.

7.3 Framework

In order to test the framework function, we create a new project and import the library we built. We implement all methods stated in the API. Below are screenshots of test results:



(a), (b)



(c), (d)

Figure 7.5 Test results of framework function (a) sensor list and information (b) file folder created (c) data file name (d) data recorded in file

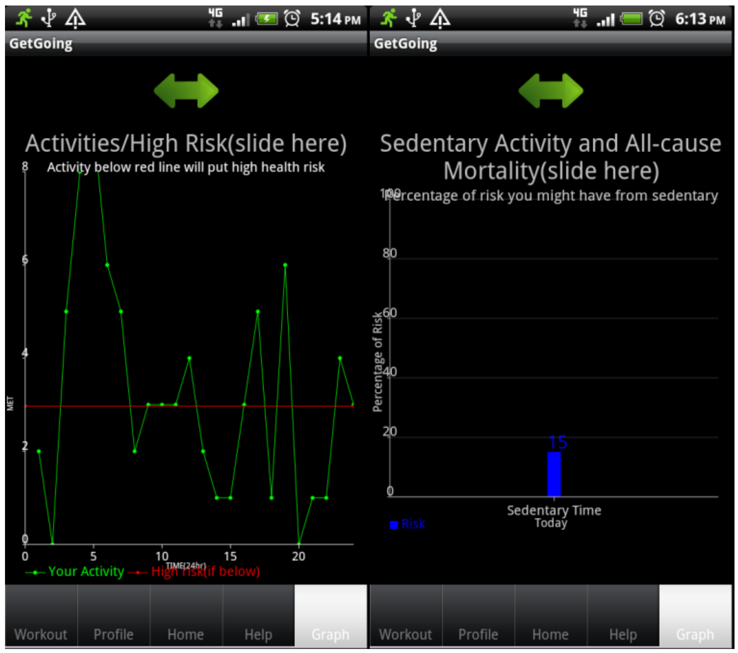
All results come out as we expected. Our framework works functionally.

7.4 Physical Risk Estimation

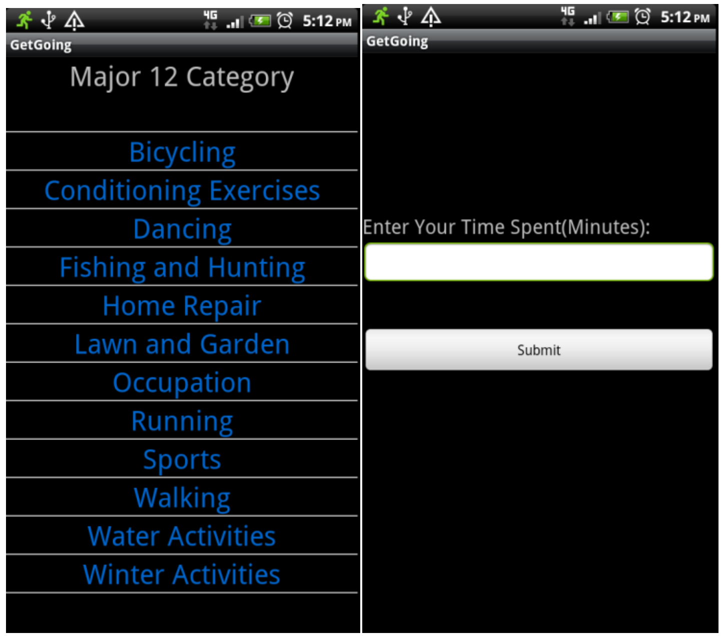
We did some tests for five risk types. The moving backwards risk is not very accurate sometimes. Exception is that when the subject is forcing to kick back, it is recognized by our system as a backwards. And we found that these five risk definitions are much more suitable for elder people whose action and movement is smooth and peaceful. Because some violent activity, which changes direction, speed, power and intense level sharply, may lead the algorithm wrongly to identify activity as risk one. This is very common in younger people who are more active. And their activity change is much more difficult to anticipate. Therefore, this physical risk estimation system is only suitable for elder people use until we find out how to define risks in young people`s activities and algorithm to identify those risks.

7.5 Health Risk Estimation

Here are some screenshots of health risk estimation system:



(a), (b)



(c), (d)

Figure 7.6 Health risk estimation screenshots (a) activity curve and high risk line (b) sedentary time and related risk ratio (c) manual input activity category (d) manual input time of activity

In order to solve the problem of lacking both training data and testing data on the algorithm design. We built a server to receive data uploaded from user side. If phone user is willing to upload his or her activity data for us to use to improve the algorithm, he or she could choose upload file to our server by selecting the upload button. And by collecting a large amount data from different phone and different users, it would greatly help us in the improvement of the algorithm.

7.6 Discussion

We tested our system on five different subjects. Results show that the real time activities` classification could accurately distinguish basic activities like rest, walking and running. Our system could also roughly estimate health risk by converting those identified activities into MET value. For the physical risk, because our system takes every dramatically activity change as a risk behavior, occasionally it mistake the subjects` violent normal activity as risk. However, our system requires smooth and slow activity change as well as non-violent activities. In general, our system can better identify and detect the five physical risk categories if the subject does activity smoothly and slowly.

CHAPTER 8: CONCLUSION AND FUTURE WORK

8.1 Conclusion

We built risk monitor system for application in the smartphone in order to improve people`s health and minimize the potential risk that could happen in future. To reach this goal, we tried to classify basic human activities with the help of accelerometer and gyroscope inside the smartphone. Once activity classification is done, we set up an algorithm to calculate several types of risk for each activity category. When the risk condition is satisfied, our system would alarm phone users for the future risk. Below are features our system has reached so far:

1. Smartphone based application.

   Unlike other research that using external sensors to monitor human activities, we built this system that only depends on sensors inside smartphone. The subject has to carry phone in their pant pockets after starting up the application. Because every experiment and research we have done so far is based on the assumption that the phone is in pant pockets. If not, the algorithm has to be changed and all criteria have to be modified.

2. Combination of dynamic and static.

   Most healthcare applications on the market are static. They cannot monitor real time activity and all calculation is based on users` manual input. Some other activity applications could classify activity, but they do not have manual settings for input and activity choice.

3. Risk definition.

In our system, we put forward two risk ideas. One is health risk that relates to people's health condition. Another is physical risk that relates to real time potential risk which could happen in the very near future because of improper activity.

4. Sensor utilization.

Our system not only depends on accelerometer for analysis, but also based on the gyroscope. We try to use both gyroscope and accelerometer data together to make a more accurate classification and make a clear definition on risk types.

5. Combination of multiple facilities on smartphone.

Our system is not limited to gyroscope and accelerometer data. Yet, it adds GPS and internet features that could better help system calculate driving speed and retrieve weather information. This information is important factor which could influence risk judging criteria.

8.2 Future Work

As mentioned before, all data collecting and system testing work are done on the model HTC G11. And we made the rule that subject should put smartphone in pant pockets. Actually, there might be a little difference between different phones. Which means the way sensor gathering data, the data type, or facility inside smartphone is different from each other. Therefore, in order to make the application universal, we need a calibration that could transform different sensor data into a standard unique type. To which, the algorithm would easily apply. On the other hand, some people, especially women do not like keep their phone in pant pocket (or they do not have pockets) would also makes this system hard to practice. So, a pre-location algorithm is needed. It is used to estimate where phone is; for example, the phone is in T-shirt pockets, bag, or something else. Only after knowing the phone`s location, can we apply different algorithm to monitor activities and estimate risk.

Moreover, the weather data we used as an additional factor for risk estimating is simply a judging criterion. For example, when system estimates it’s rainy, the potential risk range would narrow down to walking. Because we assume that running in rain might cause unanticipated accident. In the future, we need to create more detail risk definition and more risk categories, including various kinds of weather condition influence.

Also, for the application, we need more function to make it mature. Right now we only have real time risk warning and longtime health indication. We still need to add more functions such as accident processing guide, emergency instruction, SD card occupancy limitation, daily health suggestion, and primary doctor advice with subject who is providing information recorded by sensors.

At last, since our system is the first step on risk of human activity, especially on the fact that it precisely anticipates potential risk by classified activity, we need large amounts of testing data. We need feedback of how the application is doing with daily activity and how accuracy it really is. We need a large number of test subjects with different smartphones to examine whether the algorithm is good or not. Most importantly, we need to acquire as many risk definition ideas as possible from practice of daily life. Because we cannot imagine all possible unanticipated risks that might happen in various situations, it is crucial to get feedback from application users.

BIBLIOGRAPH

[1] EMMA OKONJI, *Nigeria: Global Mobile Subscription to Hit 9bn in 2017,* July 2012

[2] Centers for Medicare and Medicaid Services, Office of the Actuary, National Health Statistics Group, *National Health Care Expenditures Data*, January 2012

[3] Xi Long, Bin Yin, Ronald M.Aarts, *Single-Accelerometer-Based Daily Physical Activity Classification*, 31st Annual international Conference of the IEEE EMBS, Minneapolis, Minnesota, USA, September 2-6, 2009, pp. 6107-6110

[4] Hamed Ketabdar, Matti Lyra, *System and Methodology ofr Using Mobile Phones in Live Remote Monitoring of Physical Activities*, IEEE International Symposium on Technology and Society, 2010, pp.350-356

[5] Nigel Nicholson, Mark Fenton-O`Creevy, Emma Soane, Paul Willman, *Risk Propensity and Personality*, http://www.london.edu/facultyandresearch/research/docs/risk.ps.pdf

[6]API: Application Programming Interface: http://en.wikipedia.org/wiki/API

[7]WEKA: Open Source Data Mining Software: http://www.cs.waikato.ac.nz/ml/weka/

[8] Ainsworth, Barbara E., Haskell, William L., Herrmann, Stephen D., Meckes, Nathanael, Bassett, David R., Tudor-Locke, Catrine, Greer, Jennifer L., Vezina, Jesse et al.,*2011 Compendium of Physical Activities, Medicine & Science in Sports & Exercise* 43 (8): 1575–81

[9] Rakesh Agrawal, Ramakrishman Srikant, *Fast Algorithms for Mining Association Rules*, VLDB '94 Proceedings of the 20th International Conference on Very Large Data Bases, 1994, pp. 487-499

[10] Hari Thiruvengada, Soundararajan Srinivasan, Aca Gacic, *Design and Implementation of an Automated Human Activity Monitoring Application for Wearable Devices*, IEEE 2008, pp. 2252-2258

[11] Carlijn V.C. Bouten, Karel T.M. Koekkoek, Maarten Verduin, Rens Kodde, Jan D. Janssen, *A Triaxial Accelerometer and Portable Data Processing Unit for the Assessment of Daily Physical Activity*, IEEE Transactions on Biomedical Engineering, Vol. 44, No. 3, 1997, pp.136-147

[12] Naresh Iyer, Piero P. Bonissone, *Automated Risk Classification and outlier Detection*, Proceedings of the 2007 IEEE Symposium on Computational Intelligence in Multicriteria Decision Making , 2007, pp. 272-279

[13] Tang Yiping, Zheng Zhiying, Gu Hui, Lu Huiqiang, Wang Wei, *Elder Abnormal Activity Detection by Data Mining*, SICE Annual Conference in Sapporo, August 4-6, 2004, Vol. 1, pp.837-840

[14] Manuel J. Sueiro Abad, Ivan Sanchez-Iglesias, Alejandra Moncayo de Tella, *Evaluating Risk Propensity Using an Objective Instrument*, The Spanish Journal of Psychology 2011, Vol. 14 No. 1, pp. 392-410

[15] Manabu Nii, Kazuki Nakai, Takayuki Fujita, Yutaka Takahashi, *Action Estimation from Human Activity Monitoring Data using Soft Computing Approach*, Third International Conference on Emerging Trends in Engineering and Technology, IEEE 2010, pp. 434-439

[16] J. R. Kwapisz, G.M. Weiss, S. A. Moore, *Activity Recognition Using Cell Phone Accelerometers*, ACM SIGKDD, December, 2010, Vol. 12, pp. 74-82

[17] C.W. Han, S.J. Kang, N.S. Kim, *Implementation of HMM-Based Human Activity Recognition Using Single Traxial Accelerometer*, IEICE Transactions Fundamentals, Vol. E93-A, No. 7, 2010, pp. 121-125

[18] Xindong Wu, Vipin Kumar, J.Ross Quinlan, Joydeep Ghosh, Qiang yang, Hiroshi Motoda, Geoffrey J. Mclachlan, Angus Ng, Bing Liu, Philip S. Yu, Zhihua Zhou, Michael Steinbach, David J. hand, Dan Steinberg, *Top 10 Algorithms in data mining*, Knowledge and Information Systems, December, 2007, Vol. 14, pp. 1-37

[19] S. J. Preece, J.Y. Goulermas, L.P.J. Keeney, D. Howard, *A Comparison of Feature Extraction Methods for the Classification of Dynamic Activities from Accelerometer Data*, IEEE Transactions on Biomedical Engineering, 2009,Vol. 56, No. 3, pp. 871-879

[20] Mridul Khan, Sheikh Iqbal Ahamed, Miftahur Rahman, Roger O. Smith, *A Feature Extraction Method for Realtime human Activity Recognition on Cell Phones*, Third International Symposium on Quality of Life Technology, June, 2011, Toronto, Canada.

[21] Jr-Min Fan, Ruey-Shan Guo, Shi-Chung Chang, Jian-Huei Lee, *Abnormal Trend Detection of Sequence-Disordered Data Using EWMA method*, IEEE Advanced Semiconductor Manufacturing Conference, 1996, pp. 169-174

[22] James B. Kampert, Steven N. Blair, Carolyn E. Barlow, Harold W. Kohl, *Physical Activity, Physical Fitness, and All-Cause and Cancer Mortality: A Prospective Study of Men and Women*, Annals of epidemiology, September, 1996, 6(5): 452-7

APPENDIX

All data, files, and source code of application are available at:
http://www.mscs.mu.edu/~wwang02/Thesis/