CMPT 353: Computational Data Science

Final Project: Model Human Activity From Smartphone Data

*JING WEN*

*301327176*

Introduction

Computational data science is seen as the major area to study human and the society, to be better solve real social problems with a lot of related applications. Mobile phones as daily tools can provide us with a lot of useful information, for example recorded accelerometer data. Human activity recognition is the problem of classifying sequences of accelerometer data recorded by smart phones into known well-defined movement. Here we are classify movement activities: walk, run, go upstairs, go downstairs.

Method

Data Collection

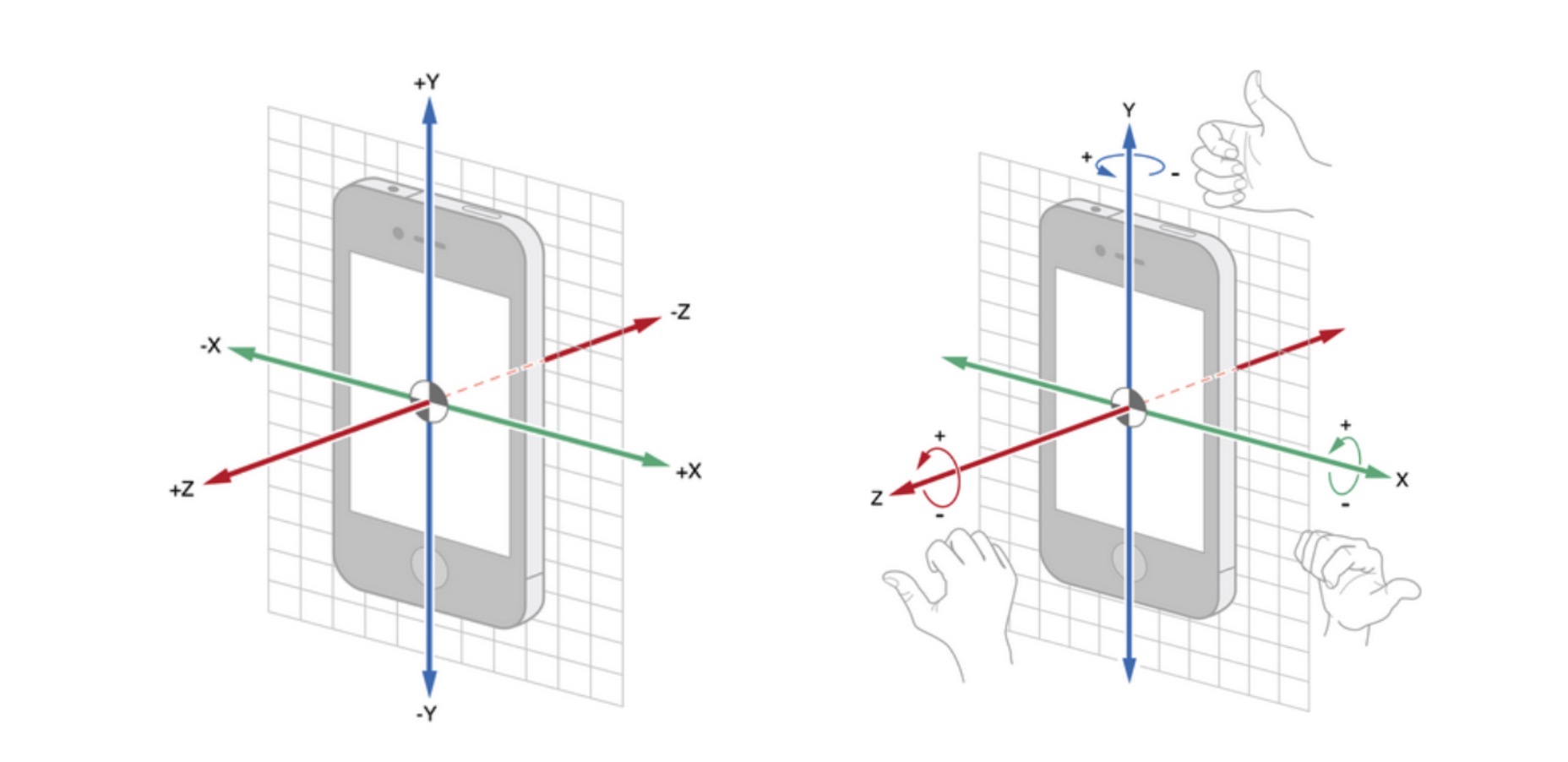
Data was collected by different individuals with the Android Application called PhysicsToolbox Suite by Vieyra Software[1]. Tied the smart phone to each subject’s ankle, hand and pocket while they walked, ran, went upstairs and went downstairs for approximately 30 seconds with their regular walking-speed. Linear acceleration, and angular velocity in the x-, y-, and z- directions was recorded by the smart phone application mentioned at the first.

To collect more comfortable data, we asked the subject stand still and the start and end of the process for around 5 seconds. A random samples were chosen to get fair data, subjects of varying heights, weights, ages were chosen for data collection to be more accurate to understand the relationship between accelerometer data and movement activities.

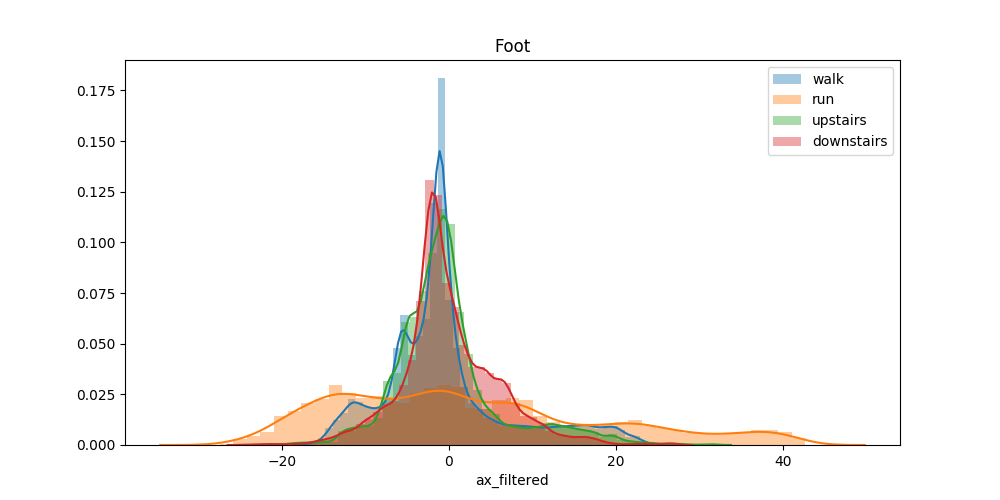
Each csv data and collected with the name format ACTIVITY-POSITION-TIME, e.g. walk-hand-2019-11-2717.02.36.csv, which will be further used to extract, transform and load.

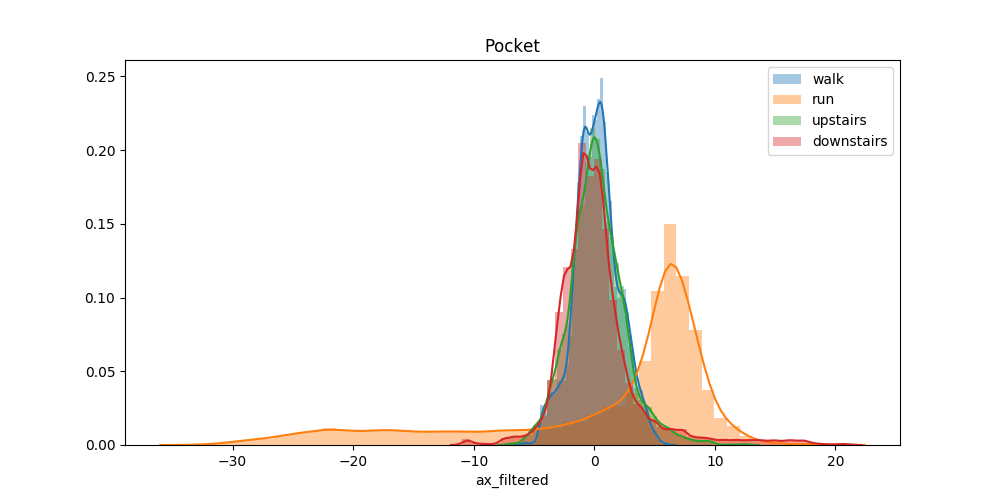
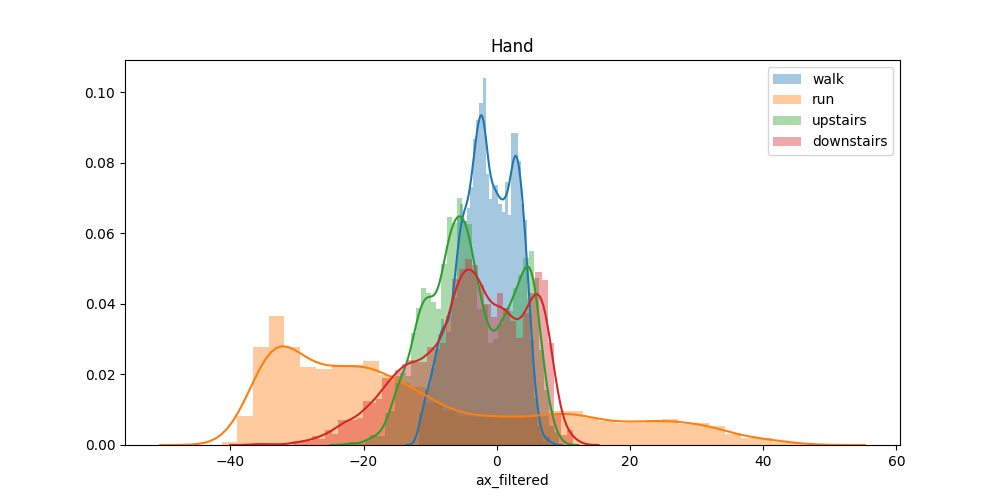
Data ETL (extract, transform, load )

Varies Python library were used to process the original data including numpy and pandas. There are 36 csv files, each file has 6 columns: ax, ay, az, wx, wy, wz. The first three are linear acceleration and the last three are angular velocity in x, y, z directions. The reason why we include angular velocity is that we want to combine and use mobile phone gyroscope data to better predict the movement activity. As shown in the fig1, the mobile phone gyroscope can also be used to conduct the direct of the movement.



Next step is to join the same category data by their file name. There are 3 files collected under same category(e.g. walk-hand-timestamp) and consider the 5 seconds dummy data in the beginning and the end, data were joint with delete those 5 seconds dummy data. The data was then further transformed by filtering with a low pass Butterworth filter to eliminate high frequency noise present from the sensor readings. This was accomplished through the use of SciPy’s signal.butter() method, with an order of 3 and a cutoff frequency of 0.05 half-cycles per sample. The coefficients returned from this function were then used in SciPy’s signal.filtfilt() method. Column added for those filtered data, e.g. ax\_filtered.

After processed the data, to be better evaluate we plotted the following distribution graph: 



Figures show single factor ax for different activities and different position we tie to the subjects.

As shown from the graphs, use single factor e.g. ax to distinguished different types of movement activities is not realistic. Most of the data are concentrated around 0.

In the following steps, all the factors will be used to model human activities.