# Abstract

Using data from four wearable accelerometer devices, we are able to determine whether or not a user is moving (defined at walking, standing up or sitting down), or stationary (defined as standing or sitting). An initial test of four binary classifiers demonstrated that a K-Nearest Neighbors model produced the overall best results in terms of accuracy, recall, precision, and true and false positive rates. Accurate predictions of movement from a testing set of four human subjects, each with approximately eight hours of measured data, was about 92%.

# Introduction

In the last several years, society has witnessed a shift towards gamification and the so called Quantified Self movement. Humans are increasingly focused on their health, and tracking all aspects of it, one of the simplest metrics being physical activity. In Q3 2016, an estimated 23 million fitness tracking units were shipped worldwide. Are they actually effective at detecting motion? Using a dataset containing the accelerometer measurements of four individuals over an eight-hour period, can we accurately detect movement?

In order to assist a fictional client, I was set with addressing the following four tasks in order to help answer the effectiveness question:

1. Determine whether or not the activity of an individual can be correctly detected from the given accelerometer data.

2. Derive features that exploit the temporal nature of these data. Determine the informativeness of these features and compare with an instantaneous approach (i.e. using only the acceleration measures for the current instant of time).

3. If you are able to reliably detect the current activity, investigate how long it takes to detect a change from one type of activity to another. Try to calculate new features that improve your change-point detection.

4. If only a single accelerometer can be worn, determine which location waist, left thigh, right arm, or right ankle results in the best activity classifier in terms of both classification accuracy and change point detection.

The undertaking of these tasks will be explored in the sections below, and a recommendation for best practices will be presented.

# Method and Results

In order to get a sense of what the data set contains, an initial profiling function was executed. The dataset was then split into testing and training sets to test the models that we plan to build. The training set was built from 250 randomly sampled points per user. The test set contains all points not part of the training set. As we will work with binary classifiers, we have chosen to transform the predicted class variable into a binary “is moving” or “is not moving” value. Moving represents the classes of sitting down, sitting up, and walking, whereas not moving is indicated by sitting or standing.

**Task 1**

To start addressing task 1, determining if movement occurs, we built four separate binary classification models, and selected the best. For use with binary classifiers, we transformed the output column into a binary "is moving"", or "is not moving" value. Additionally, the model is built without the demographics information, in order to focus only on the acceleration data.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | TPR | FPR | Accuracy | Precision | Recall |
| Logistic Regression | .73 | .08 | .84 | .85 | .73 |
| SVM | .68 | .08 | .81 | .86 | .68 |
| Naïve Bayes | .53 | .02 | .79 | .92 | .53 |
| KNN | .82 | .01 | .92 | .99 | .82 |

SVM takes n^2 time, which is not reasonable given the size of our dataset. To be able to use it, I had to perform sampling, which may have skewed the model. Naive Bayes has a decent precision score, but a laughable True Positive Ratio. Logistic Regression was simple to implement, fast, and performed respectably.

Of the three models tested, the KNN model logically provides us with the most accurate prediction of movement. If you think about it, examining the 10 nearest "neighbors" of a data point will give a fairly realistic simulation of movement. With over 90% accuracy, it is fair to say that our model can predict movement from the tracker data.

Given the nature of the data, I am not overly concerned about the true and false positive rates. It is not medically critical that we are accurate in our detections. Also, movement is an event that occurs over time. Our data may not immediately capture the relationship as “moving”, but may take a few time steps. With that being said, accuracy might be a good metric to be concerned with, as the number of targets in our data are fairly small.

**Task 2**

Since we are given time series acceleration data, we can consider examining the change in each user’s physical location over time. For detecting change points, we added a column to our dataset that contains the Euclidean Distance between two consecutive locations for each user. Regardless of the sampling techniques used to capture our sets, the relationship between a current point and its previous timestamp is still maintained using this method.

KNN models are built for each user, and tested using only the newly derived distance feature for prediction on new training and testing sets for each user. The following results were obtained:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | User | TPR | FPR | Accuracy | Precision | Recall |
| KNN | Debora | .83 | .02 | .92 | .97 | .83 |
| Jose Carlos | .80 | .02 | .93 | .94 | .8 |
| Katia | .88 | .01 | .94 | .98 | .88 |
| Wallace | .86 | .03 | .92 | .95 | .86 |

Accuracy for each user is either comparable or greater than the accuracy of the model using the points as standalone, instantaneous reports of activity. We do however see slight decreases in precision, recall, as well as an increase in the false positive rate.This indicates that we may be sacrificing a small amount of our model’s effectiveness, in exchange for only a small gain in one area.

**Task 3**

Using the model that has been built and executed on the previous two tasks, we want to detect the actual change points for activity. For example, the amount of time steps necessary for the model to accurately notice whether or not an individual has changed from a moving to not moving (or vice versa) state. In order to capture this behavior, we iterate over the dataset, and find when the current class does not equal the previous time step’s class. The amount of “lag” between the model’s prediction changing and the actual class label changing is defined as the time between each change point.

By tweaking the number of neighbors from k = 10 to k = 2, the change point detection became more accurate. This is again logical, as a change in moving or not moving is an instantaneous moment, captured in one recorded time step, which then remains consistent for several more. In detecting a change, we are less concerned with the behavior of the subject over time, but more concerned with a discrete change. To quantify the quality of the change point detection, we constructed a function to detect the difference between the predicted change point and what the model determined was a change point. This first approach gave us a value of 1159032.

With this being said, it might be useful to examine the previous state (moving/not moving) as a feature, and see how that affects our model. By capturing the previous state, our model improved change point detection by about 50%, from 11,590,312 to 5,784,829.

**Task 4**

For examining which sensor location is best, we build four separate KNN models, trained on only the data for that sensor location. Similar to Task 1, we would like to see which location has the best performing model, namely in the area of accuracy. My initial thought is thigh, as our data is capturing the act of getting up and sitting down, where the thigh is flexed, as opposed to only walking.

|  |  |  |  |
| --- | --- | --- | --- |
| Waist | Thigh | Arm | Ankle |
| .876 | .874 | .89 | .855 |

The table above however, shows the accuracy rates for each of the four locations, with arm being the most accurate. Despite my initial thought, the sanity check here seems to be that most modern activity trackers are worn on a user's wrist. If this wasn't one of the best locations, I imagine we would be seeing them placed elsewhere more commonly.

Using our change point magnitude metric defined in task 3, the results are as follows for each location:

|  |  |  |  |
| --- | --- | --- | --- |
| Waist | Thigh | Arm | Ankle |
| 39950685 | 55029693 | 40656184 | 11666236 |

From these metrics, the ankle has been shown to be the best location for change point detection. When considering movement, an ankle flexing or not would give a decent indication of movement. It is unlikely that a user would be moving their ankle without also moving the rest of their body.

# Conclusion and Recommendations

Through a K-Nearest Neighbors model, we were able to provide predictions for acceleration with reasonable accuracy. Exploiting temporal features of this dataset slightly improved accuracy, but demonstrated slight decline in all of the other captured metrics. We found that an initial change point could be detected relatively quickly, but that the model had trouble distinguishing the next. This led to the addition of a new feature, that captures the previous state for each time step, which helped slightly with predictions. Finally, it was noted that the most accurate measurements were found from the accelerometer worn on a user’s wrist.

# Future Work

Although examining movement was interesting, I would really like to extend activity recognition to focus more on pure health. For example, if we focused on a subset of people training for their first 5K race, what sorts of changes to their fitness levels occur? Do we see drops in resting heart rate and weight?

Another question I would like to address would be if wearing a fitness tracker has a positive or negative correlation with someone's general fitness level. From personal experience, I find that trackers are useful for encouraging people to develop healthy habits, but that they may need to first be fostered. This could be achieved by again, surveying a diverse group, and having a simple output column of "wears tracker" or "does not wear tracker".

# Reflection

I found that making weekly entries in the progress report really enabled me to keep the Final Project in my head the entire semester, as opposed to forgetting about it and cramming something in at the last minute. Additionally, the weekly questions aligned with the existing homeworks fairly well, and allowed for double checking of my conceptual understanding.

It might have been interesting to run the same clustering and machine learning techniques on unlabeled data, in order to determine if it was possible to detect individual users. A clustering algorithm such as Peer Pressure clustering, could also have been interesting to find like-minded individuals. Similarly, a larger dataset, such as a sports team, may have provided more insight with more data points.