Activity Recognition using the WISDM Dataset

Felix Slothower

Prof. Dabdub

Department of Mechanical Engineering
University of California - Irvine
03-20-2020

Abstract

As smart watches are becoming a common accessory among consumers, new opportunities arise in the activity recognition space. Machine Learning tools have the potential to reference both wrist and hip motion simultaneously by combining smart watch and phone accelerometer data. The combination of the two data streams are explored as a potentially major improvement to activity recognition accuracy with improved sensitvity to more nuanced activities such as eating soup versus eating a sandwich. Several features were created in an attempt to flush out distinguishable characteristics between these activities and a broad selection of machine learning models were used for classification. Results were varied however a good foundation for further data processing and model creation was layed out and can hopefully guide fellow data scientists in the Activity Recognition space.

1 Introduction

Activity recognition has many applications in the health and wellness sector. Being able to detect certain conditions quickly can be critical to the effectiveness of the respective treatment measures as is the case with Parkinson's disease [6].

With the advent of Microelectromechanical Systems(MEMS), information about the linear and angular acceleration of an object in discrete time can be obtained using a silicon wafer small enough to fit on the back of a human nail [7]. Inclusion of such sensors has become ubiquitous in mobile electronic devices and has led to an explosion in the amount of accerlation and gyroscopic data available. The raw signals coming from these sensors can be used used by machine learning algorithms to perform activity recogninition, giving abstract and superficial discrete time data meaning. These algorithms can be tailored to detect discrete activities such as a tremor in the hand which might be associated with Parkinson's disease, providing advanced diagnosis as discussed earlier.

The objective of this project is to use the digital signals provided by these MEMS sensors to determine what types of activities are being conducted by their users. Multiple machine learning models are used for the classification of activities including: k-Nearest Neighbors, Random Forests, Support Vector Machines and Linear and Quadtratic Discriminant Analysis.

The dataset being used for this project comes from the WISDM Lab and has been made available under public domain [8]. It is a compilation of raw accelerometer and gyroscopic data recorded from the phone and the watch simultaneously. The dataset has over 45 million datapoints across 51 different subjects each performing 18 different activities. This vast dataset provides more than enough information. However, it being a raw data signal, key statistical and frequency domain features must be extracted in order to improve interpitability.

Table 1: The first four samples of the dataframe in order to see its structure

	1	2	3	4
User	1600	1600	1600	1600
Activity	A	A	A	A
Time	252207.7	252207.7	252207.8	252207.8
Phone Acceleration				
PAX	-0.3647613	-0.8797302	2.0014954	0.4506226
PAY	8.793503	9.768784	11.109070	12.651642
PAZ	1.055084	1.016998	2.619156	0.184555
Phone Gyroscope				
PGX	-0.8532104	-0.8751373	-0.7201691	-0.5716400
PGY	0.29722595	0.01547241	0.38848877	1.22740170
PGZ	0.8901825	0.1622314	-0.2840118	-0.2416687
Watch Acceleration				
WAX	7.091625	4.972757	3.253720	2.801216
WAY	-0.5916671	-0.1583166	-0.1918354	-0.1559224
WAZ	8.195502	6.696732	6.107758	5.997625
Watch Gyroscope				
WGX	0.31494410	0.38738210	0.07099854	0.03797535
WGY	-1.0222765	-0.6185412	-0.2094797	0.2549756
WGZ	-0.3099616	-0.0489718	-0.1959783	-0.1565635

2   Methodology

First the data was compiled into a dataframe that contained roughly 80 percent of all available data. This was the result of making sure all recorded activities had the same number of samples. Increasing the symmetry of the data allowed me to collect all the recorded samples into one dataframe with over three million rows, thus streamlining the feature engineering process. Time data was removed and instead it was simply assumed that samples were recorded at exactly 20Hz. The structure of the dataframe can be seen in the above table with acceleration and gyroscopic data in three axes for both the watch and mobile phone. Some simple graphical analysis is done below to better understand our data.

Exploratory Data Analysis

In order to become familiar with the data and understand its characteristics I made several plots and in so doing, was able to determine applicable methods for distinguishing the different activities. Figure 1 shows a tiny snippet of the dataset I am working with and is displayed in true Digital Signal Processing format. In this particular example one can notice the clapping signal has some strong outliers which is likely due to the sudden change in velocity once the hands collide to make the clap noise. The walking signal on the other hand is much smoother and does not have the same abrupt behaviour.

In Figure 2 it becomes clear that different activities can have different signatures. Some of the distributions are bimodal while most others are unimodal. This quality is measured later in the Feature Engineering section below.
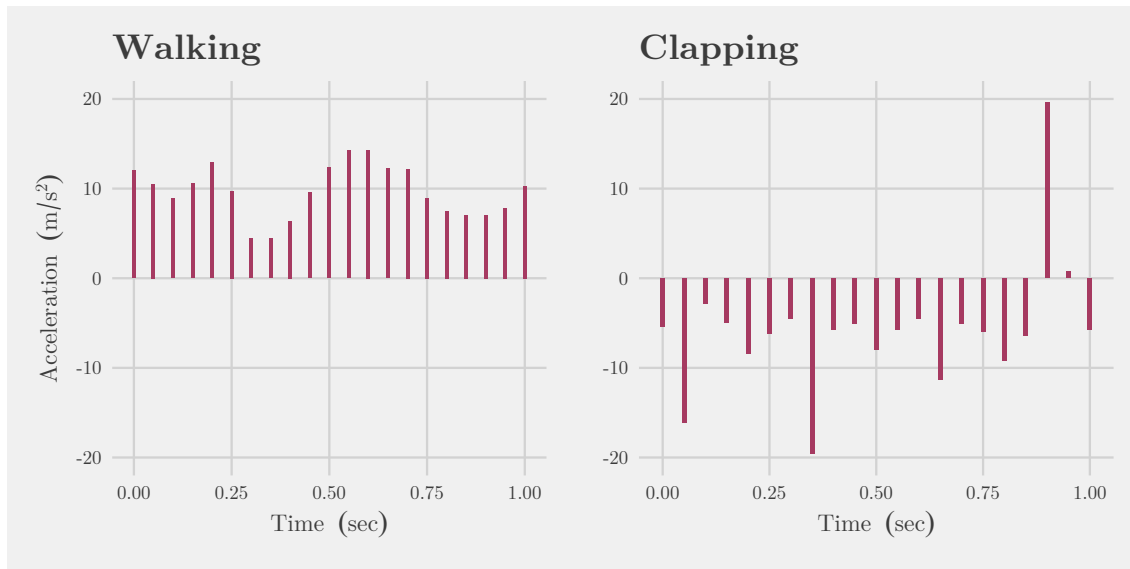

Figure 1: Discrete time signal of two different activities recorded across one second in time. Notice the sudden bursts in the clapping data versus the walking data. This can likely be attributed to the intense deceleration after your hands strike when clapping.
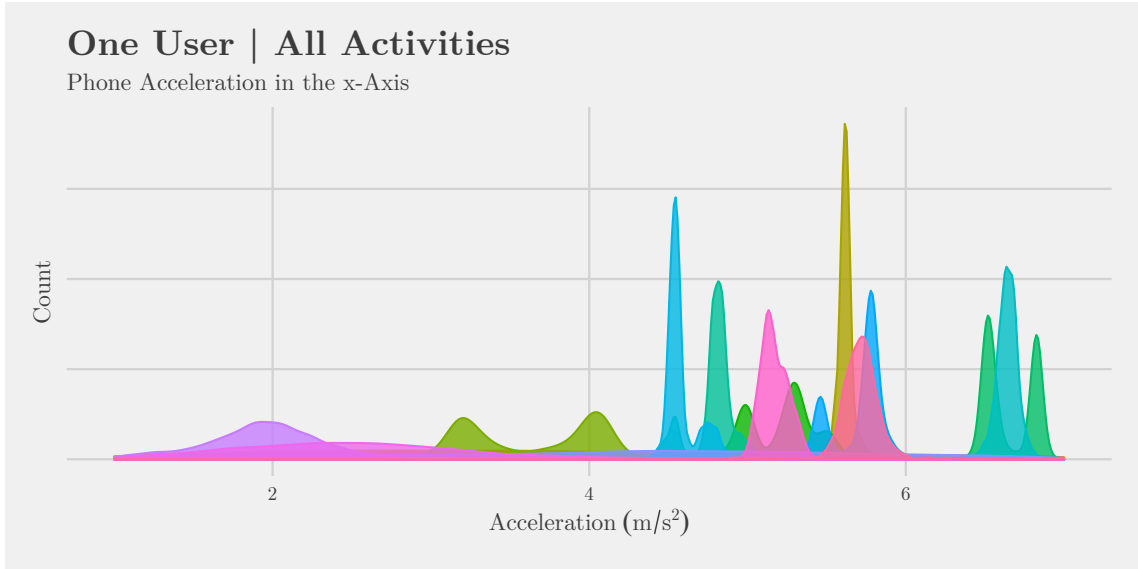
Figure 2: Density plot of all activities for only one user. The various colors represent the different activities.

While the various activities seem to be somewhat unique when looking at their histrograms, it is important to remember that different users will have the device oriented in different ways with respect to themselves and Earth's acceleration. For example, we can see a high degree of variance between histrograms of standing data produced by twenty different test subjects in Figure 3. These large differences highlight the necessity for orientation agnostic features. That is, information about a particular dataset (or feature) whose value will remain the same, independent of the devices orientation during the activity.

From these plots it is clear that there is a lot of work that has to be done in order to create distinguishable features that our various models can use in order to categorize the eighteen different activities. This is where feature engineering becomes critical.
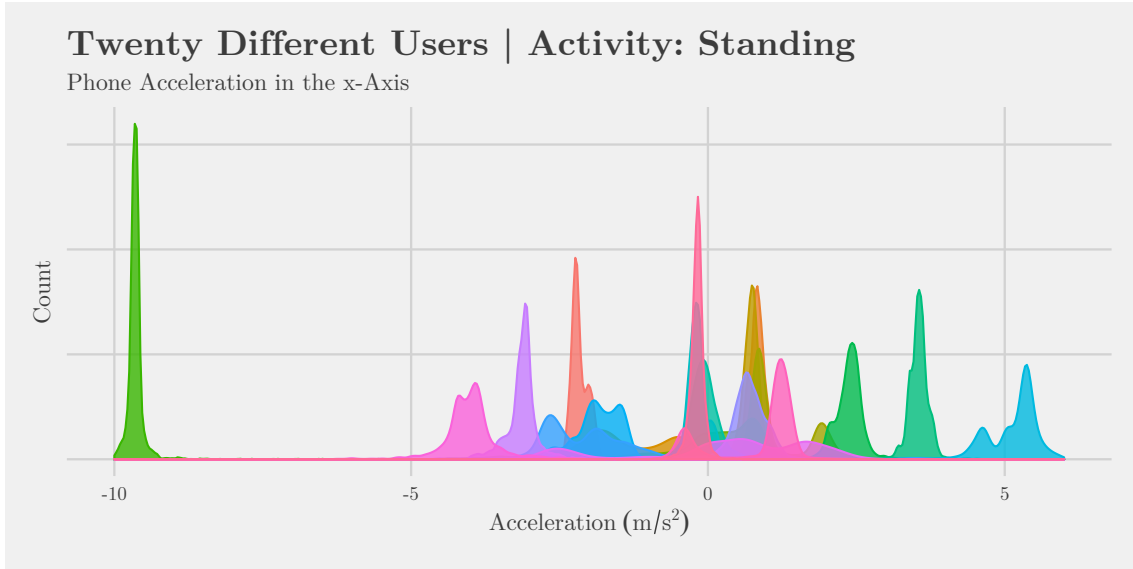
Figure 3: Density plot of all the test subjects' x-axis acceleration data while standing. The various colors represent the different test subjects.

Feature Engineering

When applying the feature engineering, sequential windows of the dataframe were examined and various statistics and functions were applied to them, transforming a sequence of rows into a single row of many different values. These values will be the predictors in our models. The windowing process is visualized in Figure 4 and a custom function was written in R for this particular task.
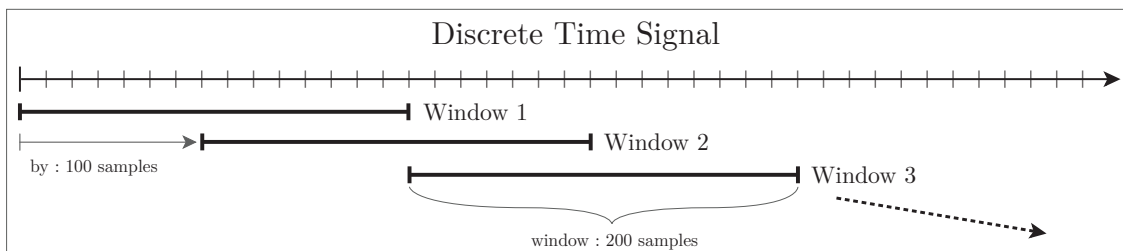

Figure 4: Diagram explaining the windowing process applied to the raw dataframe during the feature engineering process. A series of functions are applied to each window and the resultant values become the relevant features used by our models.

A total of eight different features were used and applied to all twelve signals listed in Table 1. Some features would return a value for each respective signal and others would return a

5

subset of values, one for each sensor for example. The eight features were mean, variance, axis correlation, integral, extrema, max frequency, spectral energy and the dip test. The use of some of these features was inspired by other peoples' work in the area [12] [4] and others, like the dip test, were the result of my own curiosity. The eight features are explained in closer detail below.

Mean - This feature is not orientation agnostic but it is too powerful to ignore since accuracy drops when excluded as a predictor. While different people will wear devices in different orientations there seems to be some regularity in orientation. For example, a mobile phone worn in someone's pant pocket is typically not oriented with the screen parallel to Earth's surface. It is much more likely that the phone screen is parallel to their leg in that scenario. This consistency is likely why the mean feature has some importance as a predictor.

Variance - A metric that helps explain how spread out the numbers are in a dataset, the variance is calculated as the averaged sum of the squared difference between each observation and the mean of all the observations. The square root of the variance gives the standard deviation and the two should not be used simultaneously as predictors for this reason. Highly correlated predictors can potentially reduce a model's accuracy and should be avoided.

Axis Correlation - There are many ways to calculate correlation and in this case the Kendall's correlation coefficient was used since it is less sensitive to outliers [3]. Correlation between axes is helpful for distinguishing activities that occur primarily in one dimension like walking forward versus multiple dimensions, such as climbing a flight of stairs [11].

Integral - The integration of the the acceleration yields velocity and the integration of velocity yields position. Unfortunately this cannot be directly applied to each component acceleration with the expectation of understanding the net point-to-point distance that the user traveled since the angle of the device is variable and must be taken into consideration. These calculations can become considerably complex and so only the double integral of each axis acceleration was computed. These values can still be helpful in understanding the distance the device traveled in each of its local axes.

Extrema - The extrema are calculated by subtracting the maximum and minimum observations from the mean and selecting the largest absolute value from the two. The extrema is made relative to the mean in an attempt to make the feature orientation agnostic and seeks to identify sudden jerks in the sensors that might be indicative of specific activities such as clapping as seen in right-hand plot of Figure 1.

Max Frequency - The maximum frequency of the signal is found by first performing a Fourier Transform to the 200 sample window. This converts the signal from a time domain to a frequency domain. The frequency with the most strength is then selected and used as

a feature for that particular window. A demonstration of this is provided in Figure 5. In this particular case it is quite clear that there is a strong frequency of approximately 12 Hz throughout the digital signal as illustrated in the right-hand plot.

Spectral Energy - The spectral energy is another feature found by transforming the signal from the time domain to the frequency domain. It is the squared sum of the spectral coefficients and is normalized by the window length [4]. The spectral energy tries to explain the amount of energy put into the sensor and can be helpful in detecting high intensity activities [9].

Dip Test - There are many different tests for multi-modality in a distribution but the Hartigan Dip Test offers a novel approach which does not yield false positives when the distribution is heavily skewed [5]. The Hartigan dip test produces a value between 0 and 1 where a high value gives a strong indication for deviation from unimodality [1]. This feature ended up being the least important when predicting the various activities and was eventually removed in order to improve accuracy.
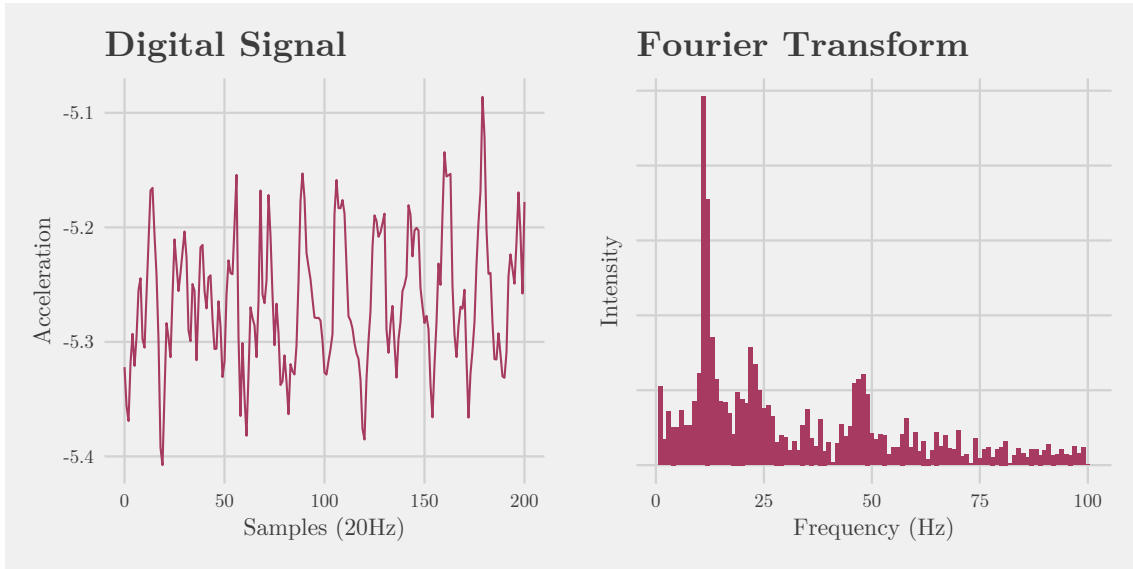

Figure 5: Example of a fourier transform applied to a ten second window of someone brushing their teeth.

Not all features were equally useful and as it turns out the Hartigan Dip Test was substantially less important than all the other features. Now that we have the final feature dataset it can be applied to our various models and their accuracies can be compared.

Modeling

Six different models were used for this project with varying success. The six models include Quadratic Discriminant Analysis, k-Nearest Neighbors, Random Forest and three Support Vectors Machines, each using a different kernel: Linear, Polynomial and Radial. As testing progressed it became clear that the Random Forest was performing much better than the others in predicting all eighteen activities. However, all six models performed well when only distinguishing between two activities.

Accuracies were validated using the k-Fold Cross Validation technique. A special function was used for splitting the data into k folds so that subsets of users were collected for each fold rather than subsets of individual data points. By doing this I was able to test the versatility of the models when faced with data from new users not previously incorporated. Had this not been implemented the accuracies would have likely been much higher.

3 Results and Discussion

In Table 2 you will find the accuracies of the six outlined models when applied to the full dataset and asked to classify all eighteen activities. As we can see, the Random Forest model performed the best but was only able to reach an accuracy of around 70%. This accuracy is not high enoough for our model to be used in a real world aplication and would need further improvement. Some improvements are discussed in the conclusion section below.

Table 2: Classification of all activities

	Accuracy
QDA	53.96%
KNN	54.89%
SVM-Linear	62.78%
SVM-Polynomial	59.67%
SVM-Radial	65.77%
RF	70.47%

In Table 3, classification between various sets of only two different activities is performed and here we see the accuracies have improved significantly. Noticing the differences between someone's watch accelerometer while writing versus clapping is much easier than distinguishing between different eating activities, as shown in Table 4. The four activities classified in Table 4 were eating soup, eating chips and eating pasta and eating a sandwich. Differences in the digital signals of these four activities are likely very nuanced and hard to seperate into different categories. In this particular case, finding features that can highlight the differences between such activities might be a task better suited for a deep learning neural network.

A confusion matrix is provided in Table 5 along with an activity key code in Table 6. This confusion matrix shows good performance for the classification of many activities but falls short when classifying the four different eating activities as well as drinking from a cup.

Table 3: Classification of Writing vs. Clapping

	Accuracy
QDA	94.12%
KNN	92.1%
SVM-Linear	93.35%
SVM-Polynomial	95.64%
SVM-Radial	96.47%
RF	97.24%

Table 4: Classification of the four types of Eating

	Accuracy
QDA	43.72%
KNN	44.47%
SVM-Linear	53.43%
SVM-Polynomial	51.09%
SVM-Radial	54.81%
RF	56.51%

Finally, by using the importance tool in the Random Forest library 'randomForest', we can get a nice visual representation of the varying importance for our predictors, as in Figure 6, and perhaps glean information about which types of predictors perform the best. The importance is obtained by measuring the reduction in impurity after using a given predictor at any node across all trees. The first thing to notice, and what I found most surprising, was that the mean for all three axes of the watch accelerometer were the three most important predictors. This clearly points to a certain consistency in the orientation of the watch across users which would make sense considering there aren't many ways a watch can be worn. One additional explanation for this could be the fact that only one smartwatch model was used across all 51 users. Using the same watch model means that the accelerometers orientation with respect to the users wrist will be consistent across all users. Conversely, the mobile phones used for the dataset varied between three different models and each model is likely to have their accelerometer oriented in a different fashion resulting in less consistent orientation and thereby less important mean axial acceleration predictors.
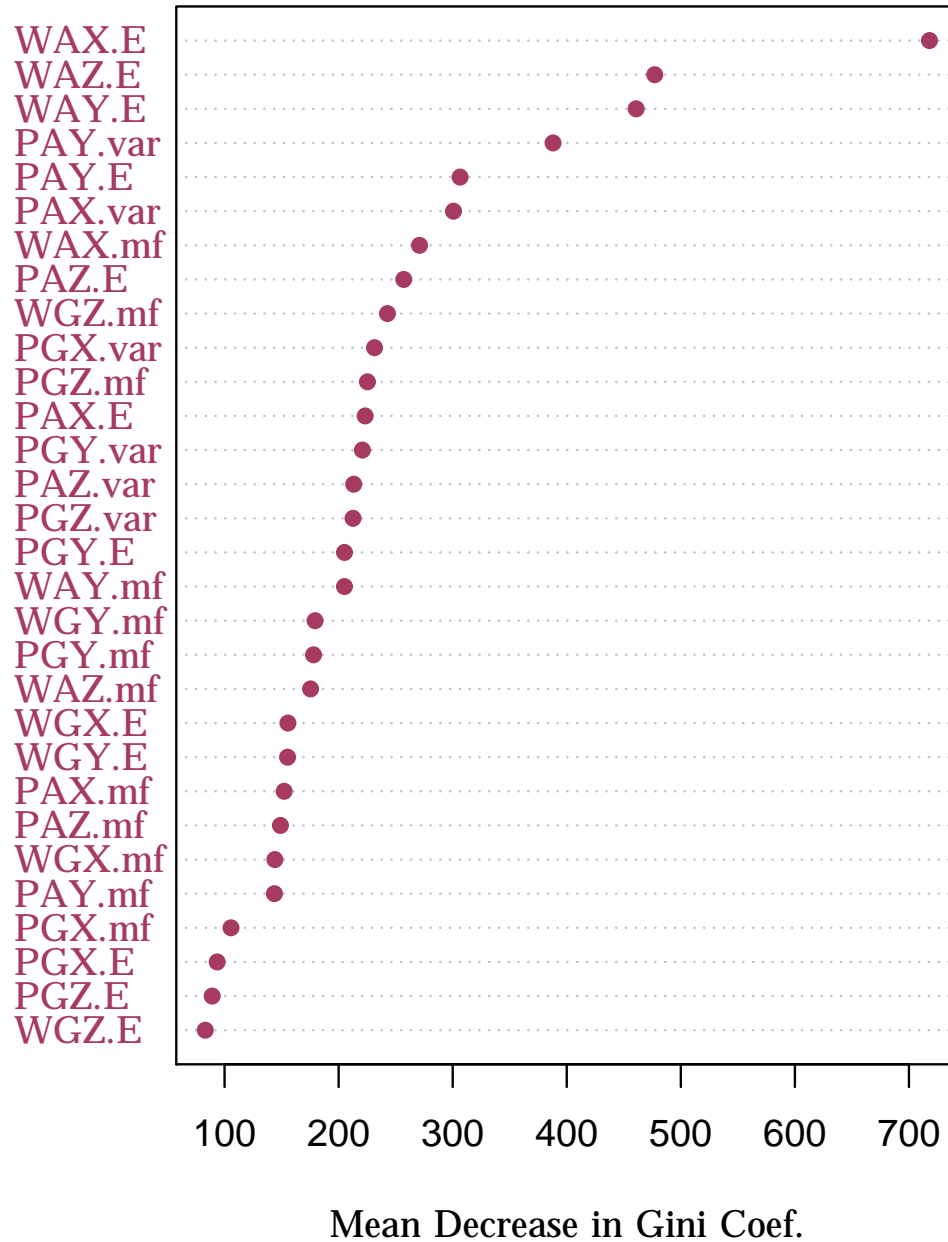
Figure 6: This plot helps visualize the importance of the different predictors. Interestingly the mean watch acceleration in the x-axis is the the most important. This indicates that orientation agnostic data might not be so critical for the watch. This would make sense since there are not many different ways a watch can be worn.

Table 5: Confusion Matrix of the Random Forest model applied to all Activities

	A	B	C	D	E	F	G	H	I	J	K	L	M	O	P	Q	R	class.error
A	1222	0	18	0	0	0	0	0	0	0	0	0	28	2	0	0	1	0.0385523
B	7	1264	7	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0.0125000
C	20	40	1097	0	0	0	0	0	0	0	0	0	54	5	0	0	0	0.0978618
D	3	3	23	1140	4	18	5	11	31	12	6	15	17	2	0	16	6	0.1310976
E	0	0	4	6	1201	6	4	6	11	5	5	10	16	3	0	2	1	0.0617188
F	0	0	2	23	5	1186	2	1	12	5	1	3	2	0	0	5	1	0.0496795
G	0	0	1	3	7	0	1258	1	12	5	3	4	7	1	1	1	8	0.0411585
H	0	0	1	4	5	0	18	1179	20	43	12	14	9	0	0	6	1	0.1013720
I	0	0	1	20	12	3	2	27	1135	30	37	30	10	1	0	3	1	0.1349085
J	0	0	0	7	4	5	0	35	29	1112	7	25	7	1	0	13	3	0.1089744
K	0	0	1	21	5	5	2	16	46	15	1158	29	5	0	0	9	0	0.1173780
L	0	0	5	21	6	5	5	27	67	47	52	998	7	0	0	8	0	0.2003205
M	0	4	35	0	1	0	0	0	0	0	0	1	1254	14	3	0	0	0.0442073
O	1	3	1	0	2	0	0	1	1	0	0	0	15	1254	2	0	0	0.0203125
P	0	0	3	0	1	0	0	1	0	0	0	2	16	35	1253	0	1	0.0449695
Q	0	0	1	10	2	0	2	2	3	8	2	3	10	16	21	1196	4	0.0656250
R	5	2	9	0	0	0	6	1	1	1	0	1	9	5	3	0	1310	0.0317812

Table 6: Activity Key Code

Activity	Code
Walking	A
Jogging	B
Stairs	C
Sitting	D
Standing	E
Typing	F
Brushing Teeth	G
Eating Soup	H
Eating Chips	I
Eating Pasta	J
Drinking from Cup	K
Eating Sandwich	L
Kicking (Soccer Ball)	M
Playing Catch/Tennis Ball	O
Dribbling (Basketball)	P
Writing	Q
Clapping	R
Folding Clothes	S

4   Conclusion

In this project we have covered all aspects of the machine learning process from exploratory data analysis to feature generation, model making, tuning and analysis of results. The initial ideas about creating orientation agnostic features turned out to be less critical than intially forecasted. With the added support of consistently oriented devices our models would likely become far more accurate.

My hypothesis that variable orientation might hurt the model making process was somewhat misfounded and the mean predictors ended up being some of the most important predictors created. One advantage of the mean predictors is their ability to remove noise. Taking the mean across a signal will quiet any of the noise that might appear either by the surrounding electronics or due to vibrations or temperature fluctuations effecting the mechanical behaviour of the sensor [2].

It should come as no surprise that the Random Forest performed the best when asked to classify all the different activities. Random Forests are very versatile when higher dimension datasets are used, like the one created for this project. Because random subsets of the predictors are used when creating each new tree, the forest becomes decoupled thus reducing the variance in averaged accuracy across trees and improving reliability.

Other models might have better success with detecting the more nuanced activities such as the consumption of different foods. A deep learning neural network for example, would be able to create it's own instructions for detecting these activities and has been quiet successful in other studies [13] [10].

Acknowledgements

References

[1]   T.A. (https://stats.stackexchange.com/users/80090/t-a). *Interpretation of Hartigans' dip test*. Cross Validated. URL:https://stats.stackexchange.com/q/157568 (version: 2015-06-18). eprint: `https://stats.stackexchange.com/q/157568`. URL: `https://stats.stackexchange.com/q/157568`.

[2]   Izudin Cemer. *Noise Measurement. FierceElectronics*. 1 Feb. 2011.

[3]   Christophe Croux and Catherine Dehon. "Influence functions of the Spearman and Kendall correlation measures". In: *Statistical Methods & Applications* 19.4 (May 2010), pp. 497–515. DOI: `10.1007/s10260-010-0142-z`.

[4]   Davide Figo et al. "Preprocessing techniques for context recognition from accelerometer data". In: *Personal and Ubiquitous Computing* 14.7 (2010), pp. 645–662. DOI: `10.1007/s00779-010-0293-9`.

[5]   J. A. Hartigan and P. M. Hartigan. "The Dip Test of Unimodality". In: *The Annals of Statistics* 13.1 (Mar. 1985), pp. 70–84. DOI: `10.1214/aos/1176346577`.

[6]   R Hauser. "Early pharmacologic treatment in Parkinson's disease. The American Journal of Managed Care". In: 16 (Mar. 2010), pp. 100–107.

[7]   Jacopo Iannacci. "Introduction to MEMS and RF-MEMS: From the early days of microsystems to modern RF-MEMS passives". In: *RF-MEMS Technology for High-Performance Passives*. 2053-2563. IOP Publishing, 2017, 1-1 to 1–39. ISBN: 978-0-7503-1545-6. DOI: `10.1088/978-0-7503-1545-6ch1`. URL: `http://dx.doi.org/10.1088/978-0-7503-1545-6ch1`.

[8]   Jennifer Kwapisz, Gary Weiss, and Samuel Moore. "Activity Recognition using Cell Phone Accelerometers". In: *Proceedings of the Fourth International Workshop on Knowledge Discovery from Sensor Data (at KDD-10)*. Washington DC, 2010.

[9]   B Nham, K Siangliulue, and S Yeung. "Predicting mode of transport from iphone accelerometer data". In: *CS* 229 (2008).

[10]   Schalk Pienaar, Reza Wilhelm, and null Malekian. "Human activity recognition using LSTM-RNN deep neural network architecture". In: (2019. 2019).

[11]   N Ravi et al. "Activity recognition from accelerometer data". In: (2005), pp. 1541–1546.

[12]   Kishor Walse et al. "A study of human activity recognition using AdaBoost classifiers on WISDM dataset". In: *The Institute of Integrative Omics and Applied Biotechnology Journal* 7 (2016), pp. 68–76.

[13]   Li Xue et al. *Understanding and Improving Deep Neural Network for Activity Recognition*. 2018. arXiv: `1805.07020 [cs.CV]`.