**CMPT 353**

Project Topic: Sensors, Noise, and Walking

Adam Avdic - 301439186

Henry Chen -301422117

Steven DaiChuy - 301422891

[Github link](https://github.com/AdamAvdic/CMPT353-Final-Project)

**Introduction**

In this report, we gathered our dataset from an iPhone utilizing the Physics Toolbox Sensor Suite application developed by Chrystian Vierya. However, a notable challenge arose with this application: it does not incorporate noise filtering when utilizing the linear accelerometer and gyroscope. Consequently, it became imperative for our team to address this limitation by implementing noise-reduction techniques to enhance the reliability of our results. Our primary objective in this study was to investigate the potential applications of walking or step data derived from smartphone accelerometers and gyroscopes. This report is divided into two experiments, each utilizing distinct data collection methods. The first experiment investigates several hypotheses: whether there is a significant correlation between an individual's step frequency and their height, age, or weight; whether walking speed is associated with a person’s gender, activity levels, or type of activity; and whether different individuals exhibit varying step frequencies influenced by their unique characteristics. This was assessed by analyzing the mean step frequencies across participants.

The second experiment focuses on the movement patterns of individuals as they ascend and descend stairs, using data collected from an accelerometer. Here, we sought to determine whether distinct, individual-specific movement patterns could be identified during these activities.

**Data Gathering**

Our data collection process encompassed a diverse cohort of genders, heights, and weights. To capture a dataset, we employed a method involving participants wearing a smartphone strapped to their ankles. Everyone underwent walking sessions with the phone affixed alternately to their right and left legs. Additionally, to enrich our dataset and capture movement patterns, participants navigated staircases while carrying the phone on either their left or right leg. This approach ensured that we captured a wide range of movements and scenarios, enabling us to explore the human gait across different conditions. By varying the placement of the smartphone and incorporating activities such as walking and stair climbing, we aimed to gather robust data that would facilitate an investigation into the relationship between accelerometer readings(acceleration), gyroscope readings (angular velocity), and individual attributes.

Inspired by the work in Maria Yousefian's MSc thesis, where she employed a Butterworth filter to address noise issues in data, we adopted a similar approach to manage the noise in our dataset. This was used by using the SciPy's signal.butter() function, set to a cutoff frequency of 0.1 half-cycles per sample. We then Fourier transform to get the frequency of steps.  The filter coefficients obtained were utilized with SciPy's signal.filtfilt() method to perform zero-phase filtering on both the total linear acceleration and the angular velocity data. The foot's linear movement over time resembles a periodic curve, and consequently, its Fourier transform yielded discrete spikes at frequencies corresponding to those present in the original time-based signal.

**Techniques Used to Analyze Data**

For our initial experiment, we explored whether specific gait characteristics correlate with an individual's height, age, or weight. We employed a multi-faceted approach to analyze the data, beginning with deploying machine learning classifiers as outlined in the flat\_analysis.py script. Machine learning classifiers are invaluable for their ability to categorize and predict outcomes from complex datasets. In this case, we utilized the Gaussian Naive Bayes Classifier, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM), which are renowned for their effectiveness in pattern recognition and predictive analytics. Following the classification phase, we applied regression analysis to delve deeper into the relationships between the variables. This statistical technique enhances our understanding by quantifying how height, age, and weight may influence gait patterns, providing insights that could forecast broader trends or effects. Finally, we conducted an Analysis of Variance (ANOVA) to rigorously assess the differences between the means of step frequencies. ANOVA is particularly useful for identifying the influence of various experimental conditions, enabling us to ascertain the statistical significance of the observed differences. This comprehensive approach not only augmented our understanding of the interplay between physical attributes and gait but also underscored the robustness of combining machine learning with traditional statistical methods in experimental research.

**Results**

**Classification Phase:**

Figure 1.0 displays modest performance by machine learning classifiers on categorizing activity levels, gender, and activity of choice. Activity level classification had the Bayesian and kNN classifiers performing similarly (0.3333 and 0.6111, respectively), while the SVM (0.3333) trailed behind. Gender classification showed a closer match across classifiers, with both the Bayesian and SVM achieving 0.5556 and kNN slightly higher at 0.6111. Activity of choice classification was notably lower, with the Bayesian classifier leading at 0.3889, and both kNN and SVM at 0.2778. These scores suggest a weak association between these attributes and walking speed.

**Regression Phase:**

Figure 1.2 gives us a recap of what's going on in the regression models in Figures 1.3-1.5. For height, it shows a minimal and statistically insignificant inverse correlation with step frequency, with a p-value of 0.2359 and an r-squared value of 0.02691, indicating it explains only 2.7% of the variation in walking pace. For age and weight, both show no significant effect on step frequency, with high p-values (0.6549 for age, 0.7557 for weight) and very low r-squared values (0.00387 for age, 0.001877 for weight), indicating they explain less than 0.4% and 0.2% of the variance, respectively. The flat regression lines further confirm the lack of any meaningful trends. The regression phase does not provide evidence of a significant relationship between step frequency and the subjects' height, age, or weight.

**ANOVA Phase:**

A screenshot of a computer program

Description automatically generatedThe ANOVA test yields a highly significant p-value of approximately 3.71e-28 (figure 1.0,1.6), which is essentially zero for practical purposes. This extremely low p-value indicates that there are statistically significant differences between the step frequencies of every individual. In other words, the step frequencies across these groups are not all the same and this would support that different individuals have different step frequencies that are possibly affected by their characteristics.

Fig 1.0

Fig 1.2

Fig 1.1

Fig 1.4

Fig 1.3

A graph with a red line

Description automatically generatedA graph with a red line

Description automatically generatedA graph with a red line

Description automatically generatedA graph of a number of bars

Description automatically generated

Fig 1.5

Fig 1.6

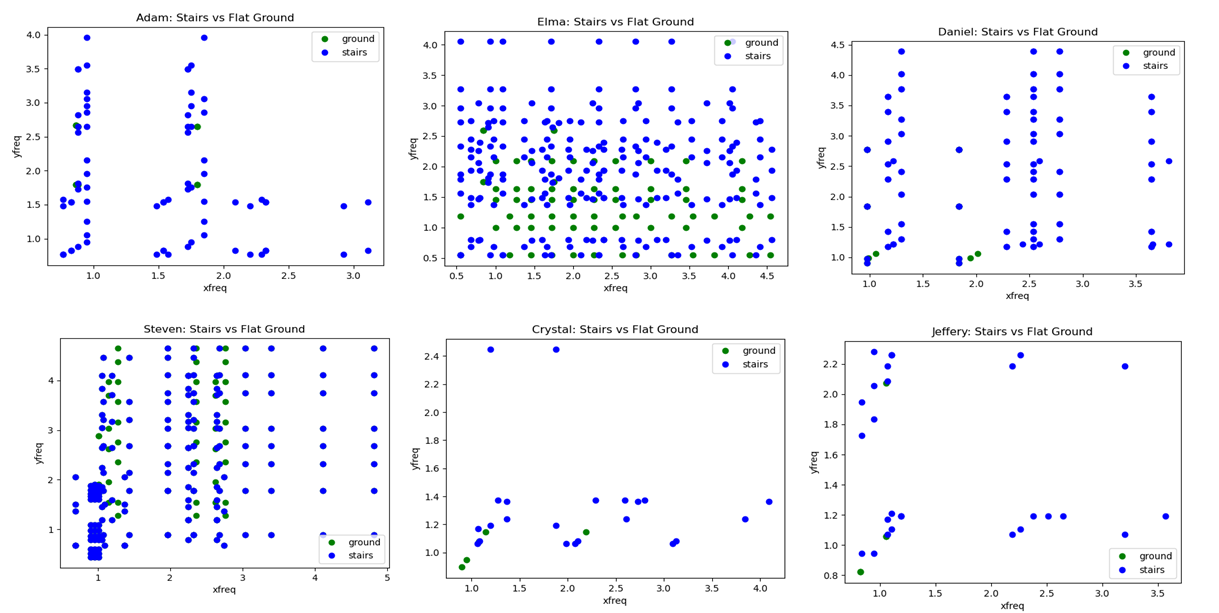
**Conclusion**

In summary, the classification phase revealed that machine learning models provided moderate accuracy in distinguishing between activity levels, gender, and activity of choice, with generally low classification scores indicating a weak link to walking speed. The regression analysis phase suggested no strong or statistically significant relationships between walking speed and the variables of height, age, or weight, as reflected by high p-values and low r-squared values. Meanwhile, the ANOVA phase indicated a statistically significant difference in step frequencies among different individuals, pointing to the potential effects of the grouped variables’ characteristics. Collectively, these findings suggest that while individual attributes may not strongly predict walking speed, there is variability in step frequency among individuals that warrants further investigation.

**Experiment 2: Single-subject data**

We experimented to determine if individuals have unique movement patterns when walking, ascending, or descending stairs. To achieve this, we collected data on each person's stride paths. The methodology involved analyzing the frequency of

acceleration in the x and y directions. This approach was chosen overusing total acceleration, as measuring acceleration in specific directions provides more precise data by focusing on movements within distinct planes. We assumed that the subjects walked in straight lines, allowing us to disregard acceleration in the z-direction as negligible. Data from the eight subjects were visualized in the graphs provided below.



**Data Gathering**

As we can see above the data collected, we got the best results from data collected by Elma, Henry, Adam, and Daniel, compared to Crystal, Kelly, and Jeffrey, their data seems to be off, this could be based on the location of the data collection, time of the data collected, the path of the data collection, and data collection device, as some of these testing were not done in the same location with the same path and same time, as well as some testing was done on different devices, such as iPhones and Android devices.

**Analysis of Data**

We plotted the frequencies of the x acceleration and the y acceleration against each other based on the flat ground and stairs. From the data we collected, we can see a trend that typically there is a higher x and y frequency when the subject was on the stairs, as seen in the plots of Elma, Steven, Adam, and Daniel, there is a significantly larger amount of stairs points compared to the ground points, we used machine learning classifications to see if the x and y frequencies can be used to distinguish between walking on flat ground and walking on stairs. We notice that data collected from Daniel and Adam have very similar results, this could be based on the fact that a person’s build has a factor in the frequencies of x and y, from their data we can see that they take significantly fewer steps for both the flat ground walk and the stairs compared to Elma, Steven and Henry, similarly, we can see that Henry and Steven's data yield a similar result, once again this is due to their builds being similar.

**Results**

Below we have the results of the 3 classifiers we decided to collect for our machine learning model. As you can see all the subjects' classifier values are close to 1 and above the 0.5 threshold, this can give us the answer that the model we trained works well and yields the results we wanted, because we do expect the gaits from walking on the flat ground compared to walking on stairs to be different.

A screenshot of a computer

Description automatically generated

**Conclusions**

The experiment successfully demonstrated that unique movement patterns can be detected in individuals when walking on flat ground versus stairs by analyzing the frequency of acceleration in the x and y directions. The results indicated that the machine learning model was effective in classifying the type of movement based on these frequencies, with a high level of confidence as indicated by classifier values above the 0.5 threshold. This suggests that such a model could potentially be used in applications that require the distinction of movement patterns, such as in health monitoring or in enhancing the responsiveness of devices to human movement. Additionally, the findings support the notion that an individual's build plays a significant role in their movement patterns, which could be a consideration for further personalized analysis in future studies.

**Limitations and Final Conclusions**

The report is segmented into two main experiments, each aimed at exploring different aspects of human movement patterns through specific methodologies. The first experiment delves into various hypotheses: analyzing whether there's a significant correlation between an individual's step frequency and their height, age, or weight; investigating the association between walking speed and factors such as a person’s gender, activity levels, or chosen activity; and exploring variations in step frequencies among individuals potentially influenced by their distinct characteristics. This involved calculating the mean step frequencies of all participants.

The second experiment concentrates on studying the movement patterns of individuals ascending and descending stairs, with data gathered using accelerometers. The objective was to ascertain whether unique, individual-specific movement patterns could be discerned during these actions.

However, the study encountered significant limitations that impacted the validity of the results. The dataset was not sufficiently large to provide definitive conclusions, and the machine learning models used only achieved moderate accuracy in classification tasks. Regression analysis did not reveal strong links between walking speed and variables such as height, age, or weight, indicating a lack of data depth and breadth. Additionally, the method of data collection—attaching phones to subjects' legs—led to inconsistencies due to physical movements, further affecting the accuracy of the findings.

In retrospect, a more rigorous approach to data collection and an expanded dataset might have yielded clearer insights. Enhancements in the precision of the devices used, along with increasing the diversity and number of participants, could have bolstered the study's outcomes. Future research should focus on refining these areas, using advanced statistical techniques and machine learning models for a more nuanced analysis of walking speed and its associated factors.

**Accomplishment Statements**

**Henry Chen**

-Collected comprehensive data on individuals' height, age, weight, activity levels, and preferred activities using standardized measurement tools and surveys

-Collected bilateral walking data on various individuals by attaching sensors to both left and right ankles and recording their strides on flat ground

-Applied Bayesian, kNN, and SVM classifiers to training and testing datasets for predicting activity-related categories, achieving measurable accuracy

-Executed linear regressions to correlate step frequency with height, age, and weight, producing detailed plots and statistics

**Adam Avdic**

-With gyroscope data stored in CSV format, I performed ETL (Extract, Transform, Load) operations using Pandas Data Frames to analyze the data in a suitable format.

-Wrote code to analyze and plot characteristic frequencies for both total acceleration and x and y components of the acceleration.

-Applied ANOVA testing techniques to compare means of step frequencies.

-Performed data collection for walking data.

-Summarized the analysis methods for experiment 1.

**Steven Dai Chuy**

-Collection of Data by asking for subjects’ information about height, weight, gender, activity level

-Wrote code about the results of the data analysis of the subjects' walking and stairs data

-Summarized the analysis methods for experiment 2

-Brainstormed ideas with the group for different methods of classifiers