Interaction Analysis with Apple Watch Data

1. Introduction

The project studies a dataset [1] containing data from Apple Watch and labelled sleep states from polysomnography. The original published paper related to the dataset is [2] and the standard citation for PhysioNet is [3].

There are many wearable devices that measure, store and analyze data to give the user feedback about their health. Apple Watch is one of them, while some other popular brands are Galaxy Watches, Oura rings and Fitbit watches. Some of the devices classify the user's sleep stages and predict a sleep score and stress levels. Those can then be used to provide suggestions on how hard you should train/work, or if you should focus on resting and recovering, on a given day. The goal is usually to enhance the user's awareness of their health and promote a healthier lifestyle.

2. Problem Formulation

The objective of this project is to perform data analysis and predict if given subject is sleep or awake based on the heart rate and acceleration data. The second objective is to predict sleep stages, which was also done in [2], [4].

3. Dataset Description

The dataset [1] contains data from 31 subjects, and it was recorded at the University of Michigan from June 2017 to March 2019. The subjects used an Apple Watch for collecting data for 7 – 14 days (2 weeks) and then slept one night (8 hours) in a sleep lab, where their sleep was recorded with polysomnography. The subjects wore Apple Watch during the sleep lab.

The dataset contained following data types:

- Heart rates (bpm).
- Acceleration
- Step count
- Sleep labels

Heart rates were recorded at random intervals of between 5 to 10 minutes. The data was collected during the whole length of the study (7 – 14 days for each subject). The distribution of values is displayed in figure 1, which shows that the data is positively skewed. The total number of observations is 254394.

Further analysis of the data showed that is does not contain significant outliers, minimum value is 30 and maximum is 215. Those values might not be realistic. Maximum heart rate is usually calculated with formula 220 – age, which would make values that are over 200 not possible. It also seems impossible to have a heart rate as low as 30. However, these values are not clear outliers and should not influence the performance of the models and are thus kept in the data.

Heart rates were transformed to a range from 0 to 1 using MinMaxScaler. MinMaxScaler was used to preserve the relationship between values, while scaling them. Standard Scaling assumes normally distributed data, and therefore it was not an option, since the data is not normally distributed.

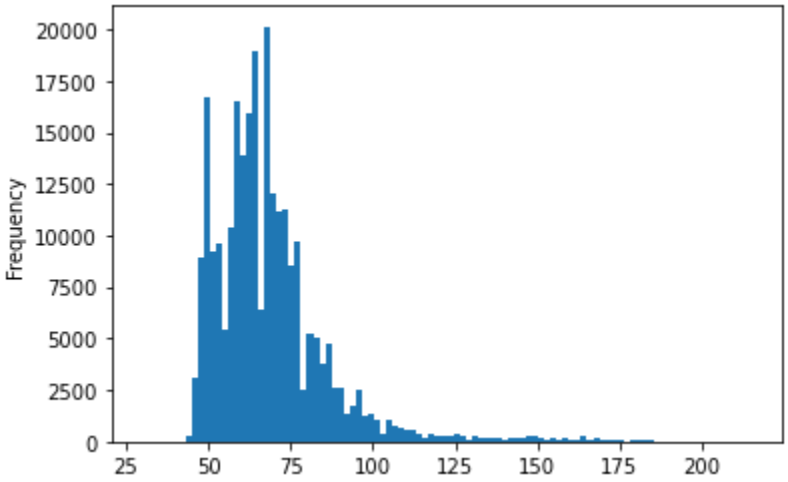

Figure 1: distribution of heart rates.

Steps contained the number of steps, and they were recorded every 10 minutes. Steps were also recorded for the whole duration of the study. Steps were not used in this analysis, since the modelling focused on predicting sleep and steps were obviously 0 when the subject is sleeping. The distirbution of steps is displayed in Figure 2.

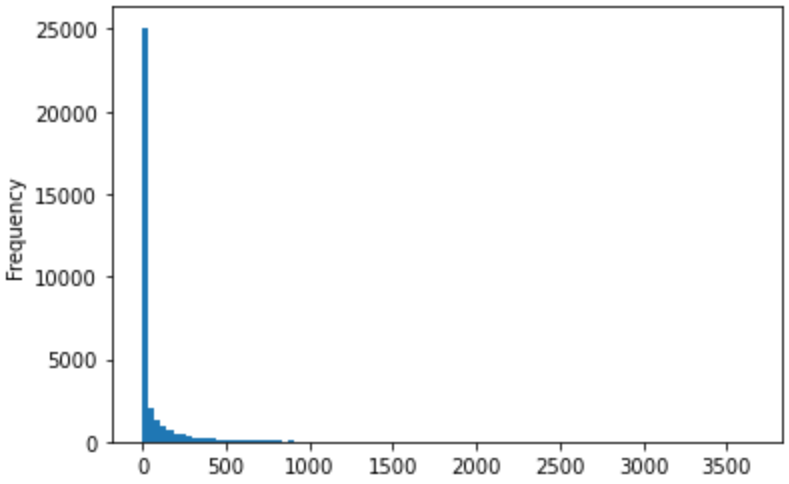

Figure 2: Distribution of steps after removing clear outliers.

Sleep labels were recorded using polysomnography (PSG). The sleep stages were categorized in following way: 0 – wake, 1 – N1, 2 – N2, 3 – N3, 4 – N4 and 5 – REM, according to [1] and [2]. However, the data contained values from –1 to 5 and had in total 7 categories instead of 6. Category –1 is assumed to also reflect the stage of the subject being awake. Based on Figure 4, it probably corresponds to the initial setup during PSG, since it occurs only in the beginning.

Sleep labels were recorded every 30 seconds during the PSG and the dataset contained 27180 sleep label observations in total. No missing values were present.

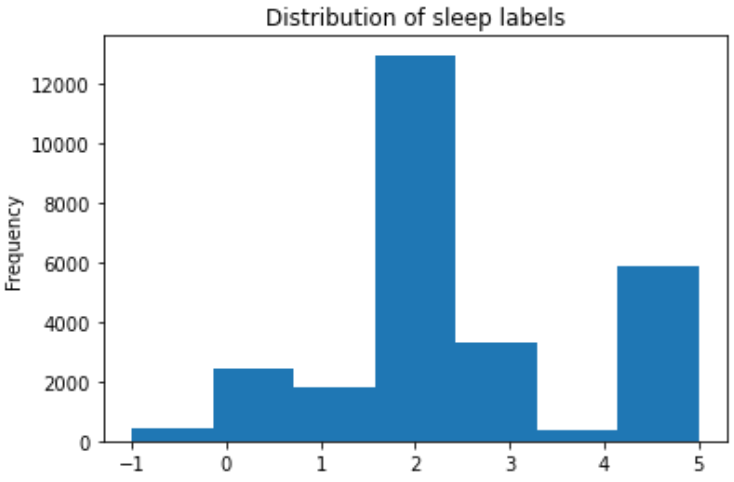

Figure 3: Distribution of sleep labels

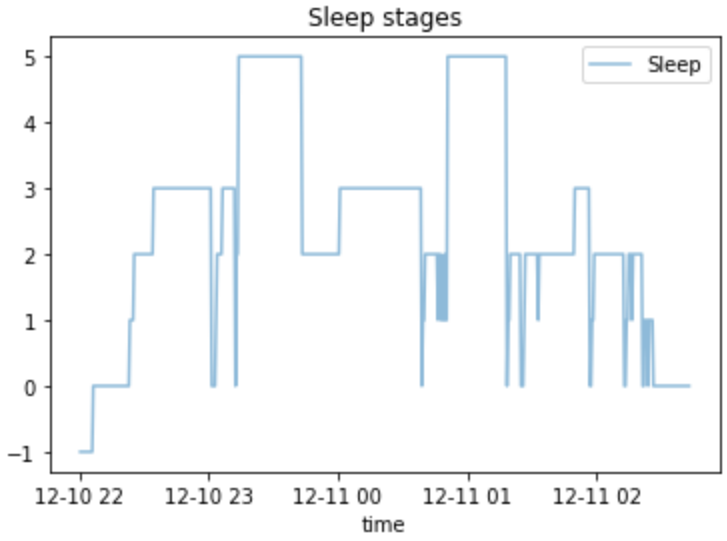

Figure 4. Example of sleep stages as a function of time.

Acceleration was measured for every subject for the duration of the PSG. Acceleration was measured in x, y and z axis several times a second. The total number of values was 51819120. For some subjects there was also recorded accelerations outside of PSG and those were removed during preprocessing. Around 25% of the data was removed during this process.

A total amount of acceleration was computed using formula:

$$total = \sqrt[2]{x^2 + y^2 + z^2}$$

The goal was to reduce the number of variables by calculating the total amount of movement, but it resulted in tight distribution around 1 and was discarded.

Accelerations contained a lot of outliers as can be seen from figure 5. The distributions of acceleration in x, y and z axis were not symmetrical and therefore the use of Interquartile Range (IQR) or Z-score methods might not provide best results. The outliers were then removed using percentile based trimming and lower 0.1% and upper 99.9% of values were removed from x, y and z (around 0.33% of values). The number of remaining rows after the processes was 38913809.

Finally, the acceleration features were scaled to range from 0 to 1 using MinMax scaler. All of the timestamps were saved as number of seconds since the PSG started. For this analysis a timestamp "2023-12-10 22:00:00" was selected to represent the start time.

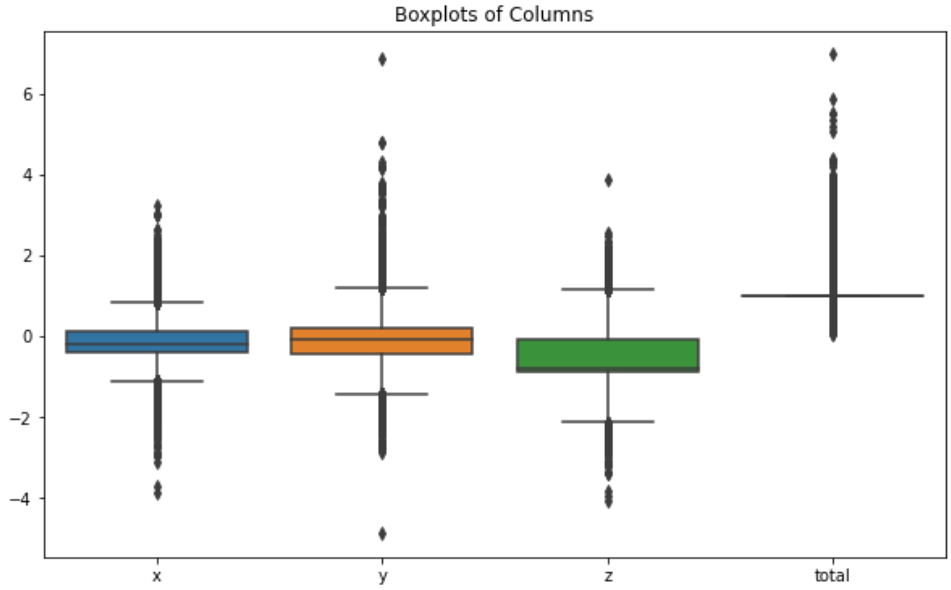

Figure 5. Boxplots of accelerations.

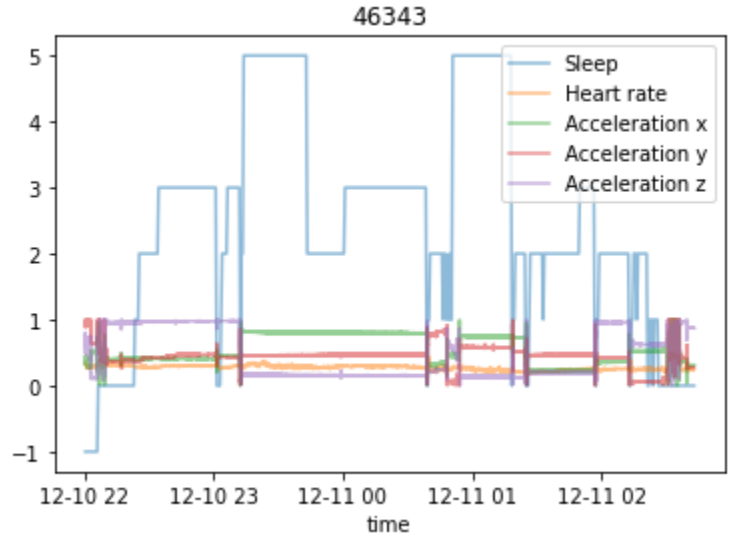

Figure 6. Sleep stages, heart_rates and acceleration for subject id 46343 during PSG.

4. Methods

The first modelling option was to predict if a user is awake or asleep, based on consecutive heart rate observations. Heart rates were labelled with sleep stages (here awake or sleep) based on their timestamps to create a dataset for predicting sleep stages. The data was grouped based on subject id and 6 consecutive observations were used for predicting the sleep stage. Observations that were outside of PSG were excluded.

The resulting training data contained around 10 000 observations (array of 6 consecutive observations) contained to class awake and 120 000 observations for sleep. The test set contained around 2700 awake observations and 31 000 sleep observations. The train and test split was 20% and it was used for all methods.

The second method was to predict sleep stages based on heart rates. For this method the heart rates were also labelled with sleep stages, but here we used all 6 stages described in section 3. For the analysis, 6 consecutive observations were used to create one sequence with a label.

A third method was to predict sleep stages based on accelerations. The data was constructed using similar methods as in method 1 and 2, but for accelerations and 20 consecutive observations were selected.

5. Results

For the first modelling method a MLP Classifier was trained using the data, but failed to predict whether the the subject was awake or sleep. A Random Forest Classifier and support vector machine was also trained with the same data, but they also failed. All models seemed to predict that nearly all of the observations belong to the sleep category.

The second model also failed to predict sleep stages reliably. The accuracy score for the MLP classifier was 0.544.

The third model also failed to predict sleep stages or wether user is awake or sleep reliably. This model was also very slow, due to the excessive amount of data.

Conclusion and Discussion

As a conclusion, predicting sleep stages based on heart rates or acceleration was not successful in this study. This is most likely due to poor model selection and input data construction. The

preprocessing steps were also more time consuming than expected and that resulted in less time for the modelling part.

For the future, the input data formulation is an important step that could be improved. The amount of data should also be limited for the acceleration analysis and a method for sampling a feasible number of observations should be done. These would result in better input data for the models and faster training times.

The model selection could also be improved. A Long Short-Term-Memory (LSTM) model or Recurrent Neural Network (RNN) could be better modelling options for the sequential data that is used to predict the sleep stage.

References

[1] Walch, O. (2019). Motion and heart rate from a wrist-worn wearable and labeled sleep from polysomnography (version 1.0.0). PhysioNet. https://doi.org/10.13026/hmhs-py35.

[2] Olivia Walch, Yitong Huang, Daniel Forger, Cathy Goldstein, Sleep stage prediction with raw acceleration and photoplethysmography heart rate data derived from a consumer wearable device, *Sleep*, Volume 42, Issue 12, December 2019, zsz180, https://doi.org/10.1093/sleep/zsz180

[3] Goldberger, A., et al. "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215–e220." (2000).

[4] de Zambotti M, Baker FC, Willoughby AR, Godino JG, Wing D, Patrick K, Colrain IM. Measures of sleep and cardiac functioning during sleep using a multi-sensory commercially-available wristband in adolescents. Physiol Behav. 2016 May 1;158:143-9. doi: 10.1016/j.physbeh.2016.03.006. Epub 2016 Mar 9. PMID: 26969518; PMCID: PMC5498752.

[5] Feifei Wang & Szilvia Boros (2021) The effect of physical activity on sleep quality: a systematic review, European Journal of Physiotherapy, 23:1, 11-18, DOI: 10.1080/21679169.2019.1623314