Machine Learning and Heart Rate Zones

# Project Report

Tom Springett

10/8/20

#### Project Definition

To improve physical fitness, there has been a surge of interest in high intensity training. This involves short bursts of all-out exercise followed by brief periods of rest. The ‘on’ duration can be as short as 30 seconds. These pulsed sessions have been found to have the maximum benefit for the shortest workout periods. These high intensity sessions (HIIT) not only benefits elite athletes but people at all levels of fitness and age. With the proper guidance, 90-year old’s have been shown to benefit from HIIT.

Heart rate monitors are used to determine training intensity. The problem is determining what is high intensity. There are various calculations such as 220 – age for maximum heart rate. And then high intensity may be 70-90% of maximum. But these are guidelines and have not been calculated vigorously nor applied consistently. Another method is to go to a lab with an electrocardiogram and do a stress test. But this option is expensive and not widely available. A new technique is needed.

This project will investigate machine learning techniques for determining high intensity heart rates. First, data will be extracted from an Apple watch. It will then be be loaded and filtered into a pandas dataframe. The data that will be analyzed is beats per minute or BPM. This one-dimensional data will be visualized with histograms. Various industry benchmarks will be applied to see how well they match up with the distribution.

Next modeling will be done to understand how well clustering results compare to the industry benchmarks. The models selected will focus on clustering and unsupervised learning. These models will be implemented and their hyperparameters optimized. Their performance will be evaluated based on silhouette scores. The silhouette score will be used to evaluate the models.

The process will be to load and filter the data, select a sample, optimize hyper-paramters and select a model based on performance metrics. techniques.

These algorithms could then be embedded into heart rate monitors and apps.

#### Analyze

Explore the Data

It is now time to explore the data. Summary statistics and graphical analysis will all be used to understand the dataset.

##### Hardware Limits, Sample Size, Maximum HR

The heart rate data was obtained from an Apple Watch and exported from Healthkit. Do to the sensitivity and availability of health data, the dataset is collected from only one white 63 year old male. The watch sensor's valid hardware limits are from 30 bpm to 210 bpm. Values outside of this range were filtered. Do to the large database with over 349,000 data points, a random sample of 10,000 from this population was generated. This did not significantly affect the distribution or results but did improve execution time.

Maximum heart can be calculated from age. The formula is 220-age. There is no comparable algorithm for minimum heart rate. In general, a lower resting heart rate is indicative of better conditioning.

|  |  |  |
| --- | --- | --- |
| **Age** | **62** | **Years** |
| **max\_hr (220- Age)** | **158** | **BPM** |
| **sample\_size** | **10000** | **Count** |
| **hardware\_upper\_limit** | **210** | **BPM** |
| **hardware\_lower\_limit** | **30** | **BPM** |

##### Descriptive Statistics

The summary below shows that there were 349,781 heat beat measurements. The minimum was 5 and the maximum is 215. These values are outside of the sensor's valid range, which is 30 to 210. Consequently, they will be filtered out. The mean at 78 bpm is significantly above the median at 69 bpm, suggesting there the distribution has a high end tail. The standard distribution of 24 is also high further suggesting a wide range of values.

**count 349781.0**

**mean 78.0**

**std 24.0**

**min 5.0**

**25% 61.0**

**50% 69.0**

**75% 86.0**

**max 200.0**

##### Sampling

For performance metrics, this analysis relies heavily on the silhouette score, which is computationaly intensive. In order to reduce execution times, a sample was drawn from the original population. The sample size was 10,000. Descriptive statistics were again generated on this sample and the distibution is virtually unchanged. The results are presented below:

**BPM**

**count 10000.0**

**mean 77.0**

**std 23.0**

**min 38.0**

**25% 61.0**

**50% 69.0**

**75% 86.0**

**max 200.0**

The next step was for graphical analysis. Since the data is 1D, histograms are used. The graph in the upper right is the raw data. The red reference line is drawn at 158 BPM. This calculated maximum heart rate agrees well with the data.

The histogram on the upper right is for the 5 zones recommended by the Polar ™ sensor company. The distribution looks continuous with a tail on the high end. The zones have been superimposed on the raw distribution. The high intensity zone is zone 5, which is 90-100% of the maximum heart rate.

The bottom two graphs are for the Mayo Clinic and the CDC target zones. They only define 2 zones, moderate and intense. The CDC and Mayo limits are more conservative than the Polar zones. The Polar zones are more appropriate for atheletes as it is common to achieve the calculated maximum heart rate indicated by the red reference line.

![Chart

Description automatically generated]()

##### Pre-Processing and Data Exploration Summary

The dataset has been loaded, columns of interest selected, data filtered and a smaller sample generated to speed execution. Summary statistics and graphical analysis have also been completed. This is a simple 1D dataset with no features or labels.

The most significant take-away, highlighted by the distribution, is that there does not appear to be any clusters. The is no clear separation in any of the data. However, the analysis will continue. The models, particularly a Gaussian Mixture Model, may be able to tease apart at least a second distribution.

Also noted the maximum heart rate calculation is validated by the data.

##### Training and Test Datasets

Note also that training and test datasets have not been generated. This is because it is un-supervised learning. There are no labels for model predictions to be compared against. Instead a 'score', explained more below, will be used to evaluate the model with the best fit.

#### Implementation

* Document Pre and Post Processing

#### Results

#### Conclusions

*Implement*

One of the major challenges for this project is finding a heart rate dataset. For this exploratory study, the data will be exported from the authors own Apple Series 4 watch. The dataset was collected over 4 years with over 350,000 datapoints across various activities (walking, biking, cross-country skiing, rowing). Unfortunately, the data export tool does not include these activities.

Obtaining data from additional people is complicated by the need to follow health care privacy laws. In addition, the export utilities were also found to be either restrictive in the type of data exported or buggy with frequent crashes. If the results from this study are promising, there will be an effort to work through these issues and expand the data collection. With more participants, more features, such as age, weight, sex, race and activity could be added.

However, even with its limitations, the dataset should be able to determine clusters of high intensity heart rate zones. A search will also be done to see if this personal dataset can be augmented by open source heart rate datasets.

#### Implement

Unsupervised learning will be used to predict heart rates in the high intensity zone. Along with other models, Kmeans clustering will be used for training, testing and predicting. The upper and lower bounds for each zone will be customized to an individual. This should allow for a more accurate training regime. Although outside of the scope of this project, if successful, this algorithm could be imbedded in an IOS high intensity training (HIIT) app.

#### Results

The clustering definitions will be compared to the standard maximum (220 – Age) and target heart rate zone definitions such as 77% to 95% of maximum heart rate.

#### Conclusion

Both the KMeans score and silhouette score will be used to determine the optimum number of clusters. A comparison will be made to the standard algorithms to determine if machine learning offers improved heart rate zone definitions.

#### Project Design

Heart rates will be exported from an Apple Watch. The HealthKit data will be converted from the XML output to a cvs file. The file will be filtered for relevant information. A Jupyter notebook will be opened and the files uploaded to a git repository. The .cvs file will be converted into a pandas dataframe. Outliers will be removed. Next a histogram will be plotted along with relevant descriptive statistics. Then training, validation and testing datasets will be generated. Various models will be explored, and the best candidate will be down selected. The chosen model will be trained, and predictions made. The results for a high intensity zone will be compared to published algorithms.

#### Appendix

#### Background on Heart Rate Sensors

The Apple watch uses an optical method, photoplethysmography (PPG), to measure heart rate. A green LED is focused on the skin and the reflected light is measured. When the heart pulses, more blood flows, and more green light is absorbed. The heart rate is determined from this absorption. The valid range is 30-210 bpm.

#### References:

“Scared of High Intensity Interval Training? A Hear Monitor Can Make it Fun and Easy”. New York Times, 5/29/18.

“Really, Really Short Workouts”. New York Times, 3/15/19.

“Preventing Muscle Loss as We Age”. New York Times. 9/22/18.

“Feel the beat of Heart Rate Training”. Harvard Health Publishing. 12/2017.

[“What is Optical Heart Rate Tracking”.](https://www.polar.com/blog/optical-heart-rate-tracking-polar/) Polar. 8/30/17.

[“Polar H10 Heart Rate Sensor System”,](https://www.polar.com/sites/default/files/static/science/white-papers/polar-h10-heart-rate-sensor-white-paper.pdf) Polar. 11/11/19.

“[How Apple Watch Measures Your Heart Rate](https://support.apple.com/en-us/HT204666)”. Apple.3/24/2020.

[“Target Heart Rate and Maximum Heart Rate”.](https://www.cdc.gov/physicalactivity/basics/measuring/heartrate.htm) CDC.

[“Intensity Exercise: How to measure it”.](https://www.mayoclinic.org/healthy-lifestyle/fitness/in-depth/exercise-intensity/art-20046887) Mayo Clinic.

[“Gaussian Mixture Model”](https://www.youtube.com/watch?v=lLt9H6RFO6A). YouTube Video.