Machine Learning and Heart Rate Zones

Tom Springett

9/1/20

Domain Background

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

Problem Statement

High intensity training has become an increasingly popular method to improve athletic performance. Heart rate zones are used to measure training intensity. However, these zones are not rigorously defined. Health organizations, such as the CDC, define 2 zones, a target heart rate zone and a high intensity target zone. Heart rate manufactures (Polar™ for example) will use up to 5 zones. And for any given zone, say high intensity, there is no rigorous or standard definition.

This project will seek to understand, customize and improve high intensity zone definitions with machine learning techniques.

Datasets and Inputs

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.

Solution Statement

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.

Benchmark Model

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.

Evaluation Metrics

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 alternative method is wearing a chest strap with an embedded electrocardiogram or ECG. The ECG measures the electrical signal in a beating heart and determines heart rate. This technique offers quicker response times, important for intensity workouts, and higher accuracy but is less convenient than the wrist mounted optical devices.

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". Polar. 8/30/17.
"How Apple Watch Measures Your Heart Rate". Apple.3/24/2020.