So you want to build a ML system

Jan Macháček

February 12, 2018

Abstract

Amongst other things, a machine learning system should be built! It can solve anything, there are so many wonderful frameworks, so many conference talks that show just how easy it is to throw together just a few line of Python, and hey presto!–a machine that can recognise digits in images, recognise hot dog and *not* hot dog... it even runs on a mobile phone. This essay is about the difficulties that are lurking in the execution and running of a robust & mature machine learning system.

1 What can ML do?

The exercise analysis system sets out to be a computerized personal for resistance or strength exercise. Exercise systems collect heart rate data as the baseline indication of physical exertion. Depending on the chosen sport, the systems typically collect GPS and accelerometer data. Geolocation, acceleration, and heart rate are fairly comprehensive source of data for outdoors sports: think running or cycling. For cycling, geolocation sampled at, say, 50 Hz reliably establishes the route ridden, but is also the source of data for speed and acceleration; with underlying elevation data & the weight of the rider and the bicycle, it is possible to estimate power output. With the heart rate, it is possible to get a fairly accurate view of the activity in Figure 1.

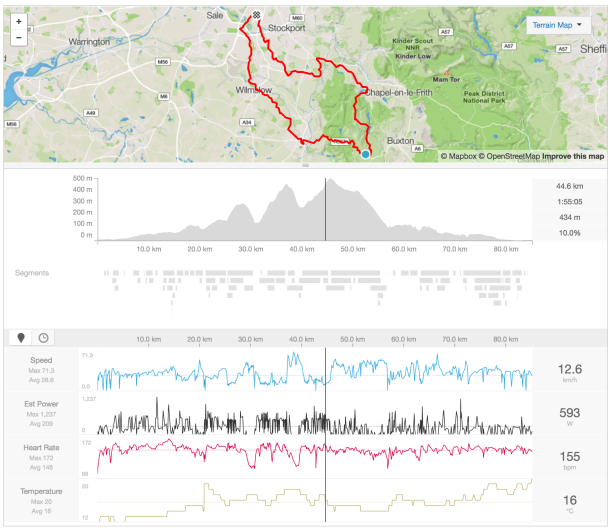

Figure 1: An afternoon ride

Everything becomes much more difficult indoors; and then even more difficult when not in a swimming pool, on a rowing machine, or on a spinning bike; when the user turns to resistance exercise. Some exercise machines now come with Bluetooth connectivity that advertise the details of the activity. Turning to the basics, bodyweight exercises do not come with any machine attached; similarly, free weights do not have any beacons, and even if they did, they would not tell us which exercise the user is performing. The geolocation coordinates, if at all available, will show that the user is not moving much; the heart rate lags behind the exertion, so it'll only increase by the time the user is done with a set of a particular exercise (assuming the exercise is done with appropriate intensity).

The only thing left are sensors that the user is wearing; the baseline sensor is the accelerometer in a smart watch. Figure 2 shows two time windows that contain exercise.

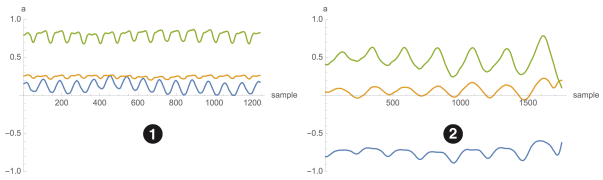

Figure 2: Two exercises

References

[1] Alfie Kohn. *Why Incentive Plans Cannot Work*. `https://hbr.org/1993/09/why-incentive-plans-cannot-work`.