Machine Learning Engineer Nanodegree

Capstone Proposal

Predicting heart rate dynamics and training load from cycling data

Ruud Goorden

ruud.goorden@essent.nl

November 5th, 2020

Abstract

In the sport of cycling coaches / riders can benefit from predicting heart rate dynamics before a ride takes place. This can help to estimate training load and determine whether certain heart rate dynamics are desired. Also this information could be used for planning purposes further downstream when multiple rides are simulated. Moreover predicted heart rate dynamics could ex-post be compared to realized heart rate dynamics which might give further insight in the rider's physical condition. Therefore the goal of this project is to construct an individual level model for heart rate dynamics and training load calculation by using cycling power data and relevant other data which can be pre-planned. Keywords: Supervised learning, cycling, heart rate, power

Domain background

In competitive cycling powermeters are important instruments that are used to guide training and performance progression. By using a powermeter the actual mechanical work can be monitored by the coach / rider. Also (external) training load metrics can be derived from powermeter data.

Before powermeters were commercialized training with power was only accessible for professional cyclists. Non-professionals had other ways to setup training programs and training guidance. One of those ways was to use heart rate in training prescriptions. Heart rate is a metric which follows power development, meaning is it a lagged variable. When riding on the bike and sprinting power goes up immediately while heart rate has to ''catch up'', because of an increase of oxygen demand in the working muscles. For that reason heart rate was often not a good way for prescribing intermittent based work bouts.

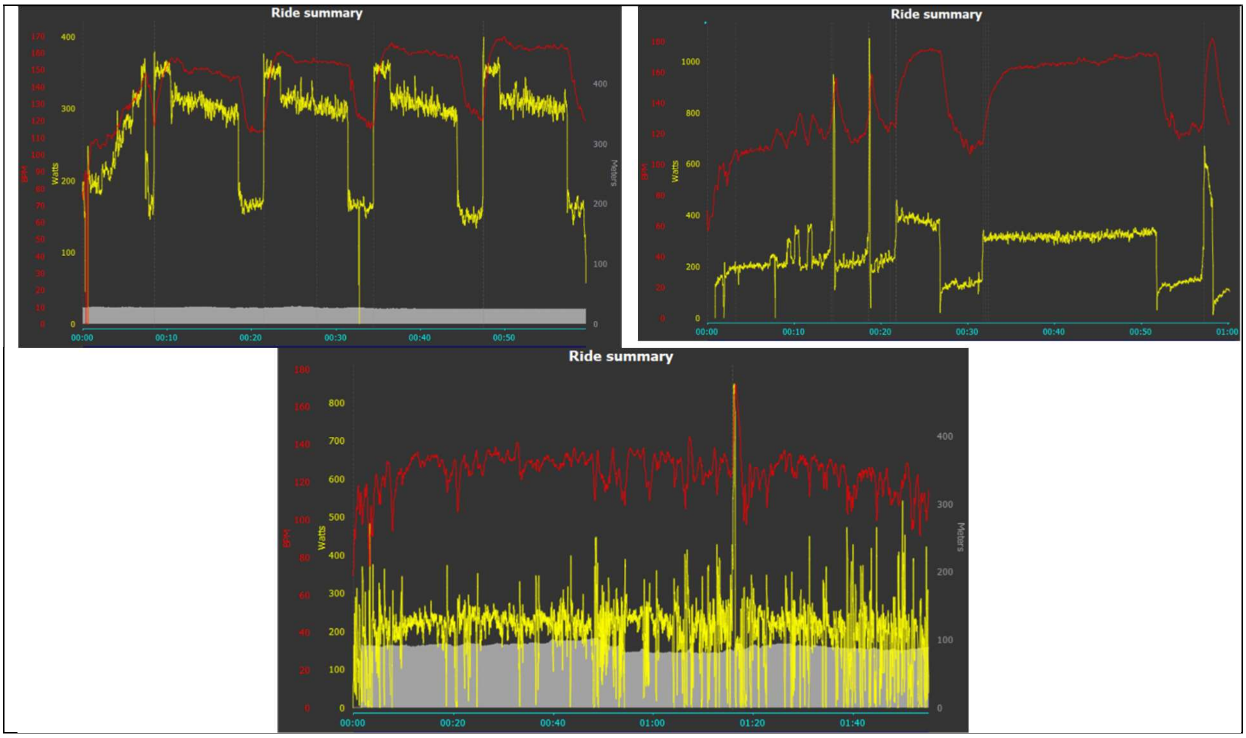
However, heart rate of an individual varies with the intensity of their exercise in a nonlinear manner, and understanding the relationship between heart rate response and exercise intensity is a primary interest for riders. In addition to being an indicator of exercise intensity, heart rate is also related to other physiological responses such as oxygen consumption, energy expenditure, and cardiac output. Therefore, it is useful to also take into account heart rate responses during training and racing to base training strategies on.

Heart rate dynamics & training load

Heart rate dynamics during exercise are influenced by a lot of factors, which can broadly be defined by factors occurring during (inside) the ride and factors outside of the ride. Moreover these factors can be further divided by factors influenced by the rider him/herself and from environmental factors. The riders physiology in essence also plays an overarching role in these dynamics.

During the ride there are several factors. A very important factor is the type and makeup of the workout itself. A long endurance ride at relatively slow pace shows different dynamics than a ride with short work bouts above the riders threshold. These specific workout related factors can be summarized by work bout intensity, duration of the work bout, recovery period intensity, recovery period duration, number of intervals or series duration, number of interval bout series, between series-recovery duration and the intensity [1]. In Figure 1 three examples of heart rate dynamics and workout type can be seen.

Figure 1*



*red = Heart rate (beats per minute), yellow = power(watt), grey = altitude

Also other factors play a role during a ride. Environmental variables have an influence like i.e. temperature, humidity and altitude (hypoxic conditions), inclination and surface (e.g. cobbles vs tarmac). Moreover nutritional variables in the refueling strategy play a role. Examples are hydration, food / carb ingestion, caffeine ingestion etc. In addition during a ride mechanical factors like cadence (better said angular velocity) for generating power plays a role. Riding at a lower cadence tends to put more stress on the muscular system instead of the cardio-pulmonary system which tends to lower heart rate for the same power output. Riding up a mountain will recruit muscles from in the upper body which need oxygen and therefor heart rate can increase. Also changing breathing rate for the same workload can alter heart rate.

As said next to factors inside the ride also factors outside of the ride play an important role. Some of these are also environmentally related and nutritionally related like during a ride. Dynamics can differ for example for a well fueled rider versus a rider which is not well-fueled. Also the riders diet (lots of carbs / fat) can have an influence. The degree to which the rider is adapted to these diets (factors in general) also play a role. Moreover the training status of the rider plays an important role. Heart rate response after a period of intense training which is possibly accompanied by fatigue which gives different dynamics than compared to when the rider is ''fresh''.

Mental aspects can also play a role in the dynamics. Arousal of the athlete before a certain interval starts can have an influence, but also when the rider experiences stress (from other life factors) this might alter the heart rate dynamics during the ride. Finally certain type of medication (e.g. beta blockers) can influence heart rate dynamics.

Training load for a session can be expressed by several heart rate related metrics. TRIMP (TRaining IMPulse) [2] and eTRIMP [3] are often used for this purpose. Both metrics are related to internal load. TRIMP is based on the intensity of the exercise as calculated by the product of the average percentage of heart rate reserve and the duration of exercise. eTRIMP is calculated as the product of the cumulated training duration (in minutes) for 5 heart rate zones multiplied by a coefficient relative to each zone.

Solution

The goal of this project is to estimate heart rate response to the internal load imposed during a ride. A coach / rider can get an insight in predicted heart rate dynamics and evaluate the internal training load *before* the ride takes place. This can aid in training planning strategies. Also -in a later project- these estimated could be used to compare it to actual realized dynamics and load.

Datasets and inputs

As described there are multiple ways heart rate dynamics are affected during a session. However, for this project we will at first stick to session data simply due to the fact we don't have any other recorded data from outside of the sessions.

Datasets are ride files (.csv) from actual performed training and racing rides from two riders over at least a year period. Rider 1 has approximately 1200 second by second ride files and rider 2 200. Each of these files contain timestamped data (second by second) which have been registered by a cycling computer. Data files contain: power (watt), heart rate (beats per minute), cadence (rate per minute), temperature (Celsius), speed (km/h), altitude (hm) and GPS (latitude, longitude) features.

In the exploration phase we consider which features next to power we will include. Since the main idea is to be able to load data which is predefined by the coach / rider we want to keep it as simple as possible. Since power is a strong predictor of heart rate and coaches are used to prescribe workouts by power we include power and possibly 1 or two other features which are known / planned by the coach / rider in advance. Generating a file which contains -next to powerdata- also other data like altitude, etc. is a separate project and could be described as a workout builder program. For testing purposes we make use of files with simulated data and some files which were not used for model building.

In literature modelling of heart rate dynamics is often focused on dynamic system approaches using stochastic optimization techniques [4,5,6] done in controlled (lab) environments. Also these models focus mostly on steady state exercise intensities. Lefever, et al. [7] use time-variant and time-invariant models and use power and inclination as predictor variables in a multivariable dynamic transfer function. Also this was the first attempt in the field. A R2 of 0.88 (+-0.08) was reported using a time variant approach. Mazzoleni et al [8] used a nonlinear dynamic model to predict heart rate from power and cadence data. They report an R2 of 0.90 (+-0.05). The first study which attempts to use machine learning to predict heart rate response is from Hilmkil et al [9] in which they used sensor data from 15 professional road cyclists in road conditions. They also i.e. used power and cadence, but included heart rate 30 second prior to predict heart rate at time t. From the LSTM architecture they got a mean RMSE of 5.62.

Of studies reported in literature our project has most resemblance with Hilmkil et al., because they also used sensor data from the field. However we will use less predictor variables and also heart rate will not be a predictor variable since that is what we want to predict in advance. The evaluation metric is the RMSE (Root Mean Square error). We will calculate this metric for each of the sessions in the train, test and validation sets and overall. Also we will calculate RMSE given realized TRIMP scores on aggregate ride level.

Project Design

To build the solution data is gathered from the two riders. Both riders have given consent to (anonymously) use their data. This data is already available in a cycling analytics tool called Golden Cheetah, but it first needs to be extracted from the tool. Because the tool has the option of bulk downloading the ride files in .csv data can easily extracted.

After the data extraction we will prepare the data: explore and clean the data. There will be erroneous data in it since sensor signals might drop from time to time, some data might not be consistently collected, some rides are maybe a few minutes, and possible extreme values might be recorded, etc. These steps will in general also hold for the data preparation part of the script in the pipeline before it goes into the model to return results.

When the data preparation steps are done we will start with generating features for the time series ride data. Since we are trying to model on a ride level basis features will also be extracted on a ride level basis. Part of these features will be derived from patterns in the data and part will come from features as found in literature. Standardization and normalization will be applied if appropriate.

Possibly we will also add features not strictly related to the ride itself, but from previous rides. Our goal is to first start with a relatively simple model and see how difficult it will be to get better results in some iterations moving from classical regression techniques to tree methods (e.g. LightGBM / XGBoost) and neural nets (MLP). Given the time series nature of our data and depending on progress (and timing) we will try other methods (like LSTMs) which are also more computationally intensive given the type of data we have (some rides have over 25K data points). Hyperparameter tuning will be part of the model building process.

We will evaluate the models based on the RMSE distributions of the rides in our validation data after we split the files into train, test and validate sets. Note that we will not split the time series data itself, but the ride files. Also we will go visually through some rides to get a feeling on the dynamics of the predictions. In addition we will evaluate the total aggregate heart rate load metrics (TRIMP and eTRIMP) based on RMSE. In essence our goal is to get a model with similar features which can explain both riders data well.

After these steps which occur mostly in Jupyter Notebooks we will build the model in AWS. In terms of functionality we have a pipeline for one rider in which a planned workout (.csv file) can be loaded which will return a graph (visual) with the heart rate response and an overall (aggregate) heart rate training load metrics. Since each model is rider specific for the project we have a model for one rider.

Literature

1) Science and Application of High Intensity Interval Training, Laursen P, Buchleit M, 2019, Human Kinetics, ISBN 978-1-4925-5212-3

2) Thomas W. Calvert, Eric W. Banister, Margaret V. Savage, Tim Bach, A Systems Model of the Effects of Training on Physical Performance, IEEE Transactions on Systems, Man, and Cybernetics, volume SMC-6, issue 2, 1976, pages 94–102

3) Edwards S (1993) High performance training and racing. In: Edwards S, editor. The heart rate monitor book. Sacramento: Feet Fleet Press 113-123

4) Zakynthinaki MS, Stirling JR (2008) Stochastic optimization for the calculation of the time dependency of the physiological demand during exercise and recovery. Comput Phys Commun 179(12):888–894

5) Stirling JR, Zakynthinaki MS, Refoyo I, Sampedro J (2008) A model of heart rate kinetics in response to exercise. J Nonlinear Math Phys 15(3):426–436

6) Mazzoleni MJ, Battaglini CL, Mann BP (2015) Modeling heart rate dynamics in response to changes in exercise intensity. In: Proceedings of the ASME 2015 International Design Engineering Technical Conferences and Computers in Engineering Conference, DETC2015–47587

7) Lefever J, Berckmans D, Aerts J (2014) Time-variant modelling of heart rate responses to exercise intensity during road cycling. Eur J Sport Sci 14(S1):S406–S412

8) Modelling and predicting heart rate dynamics across a broad range of transient exercise intensities during cycling, Sports engineering, 19, 117–127(2016)

9) Towards Machine Learning on data from Professional Cyclists (2018), Hilmki Al, Ivarsson O, Johansson M, Kuylenstierna D, van Erp T, Accepted for the 12th World Congress on Performance Analysis of Sports, Opatija, Croatia, 2018, https://arxiv.org/abs/1808.00198