Conditional Probability of Fitbit Step and Heart Rate Data: A prelude to HMM

Ting Ki Chang

The University of Melbourne

Motivation

In recent years, the use of wearable physical activity tracker such as Apple watch and Fitbit has become the most preferred method of choice in terms of cost-efficiency and comfort. The classification of activity intensity is vital in various settings, for instance, health concern and health-related research purpose. Hidden Markov Model (HMM) has been one of the most popular methods being studied to improve the classification and the essential step of building a precise and accurate model is specification of emission probability distribution, which is hence the study focus.

Model

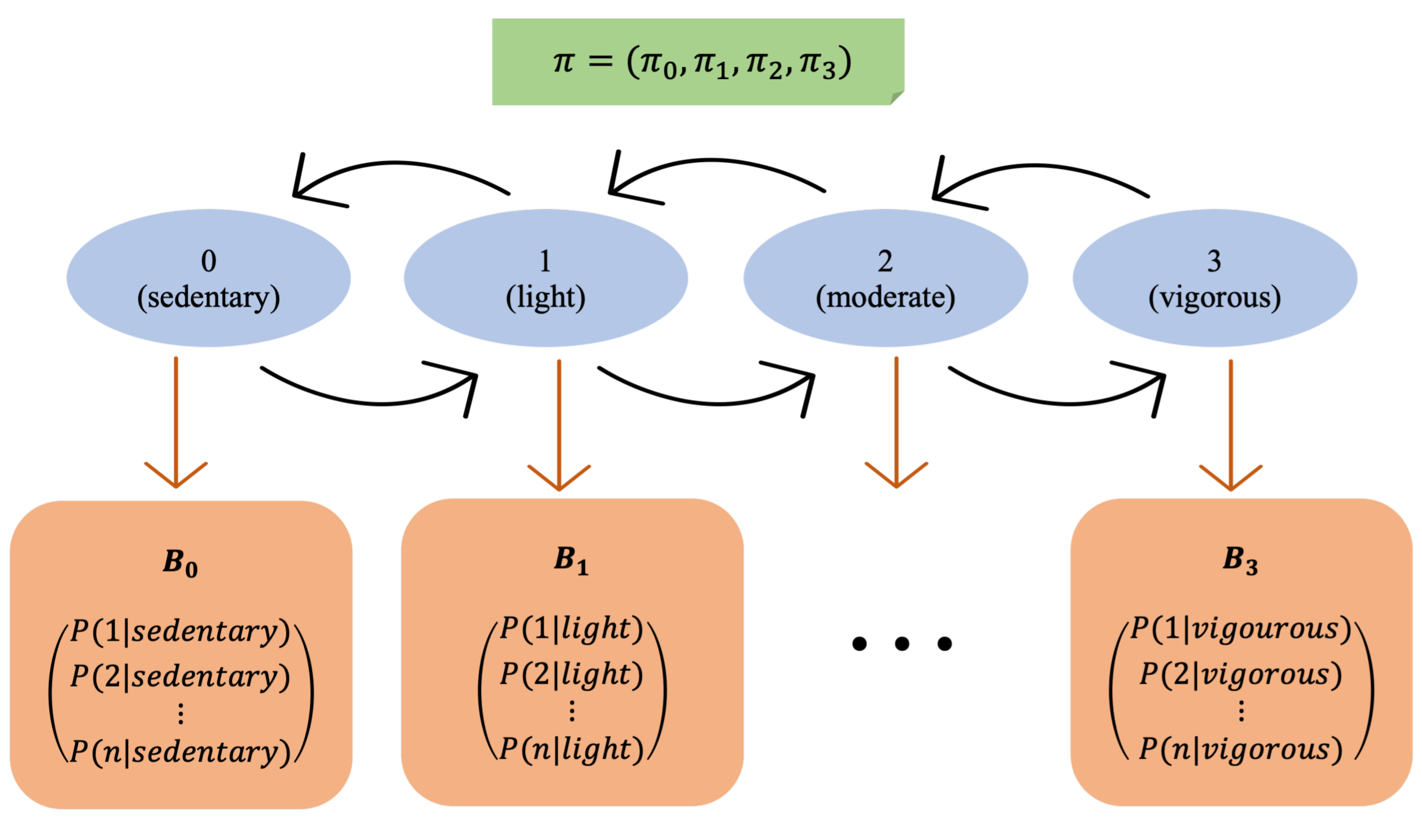

Figure 1. The HMM structure is comprised of a starting distribution π (green part), transition matrix (blue part) and emission matrix (orange part). It should be noted that for the blue part, the arrows do not represent the actual situation but they can point to every state (blue circle) freely depending on the probabilities.

Data of heart rate, number of steps and intensity of exercise (produced by Fitbit) for each minute is analysed for the classification of physical activity. The blue section hence denotes the states of intensity, ranging from 0 (sedentary) to 3 (vigorous). Since we are going to treat Fitbit classification as gold-standard for now, an empirical conditional distribution of step and heart rate under each physical activity (PA) category can be obtained. We fitted several probability distributions to the empirical data and identify the distribution with the best fit as a good candidate to model the emission probability distribution.

Method

Heart rate and steps are discrete variables so probability mass functions are considered. Poisson, negative binomial and geometric distributions, which are typically utilised to model counts, are included in the test [1].

Given that the number of steps is 0 during sedentary activity, it is assumed that $P(no. of steps = 0 | sedentary) = 1$. As shown in Figure 2, small steps can even reach another state, inferring that there is no other range of values. Then it is sensible to assume that under sedentary activity, the expected number of step under any distributions is 0.

Finding the set of parameters that maximise the likelihood function, maximum likelihood estimation is hence used along with Bayesian information criterion (BIC) [2]:

$$BIC(Model) = -2log(\hat{L}) + (\#estimators) * log(size).$$

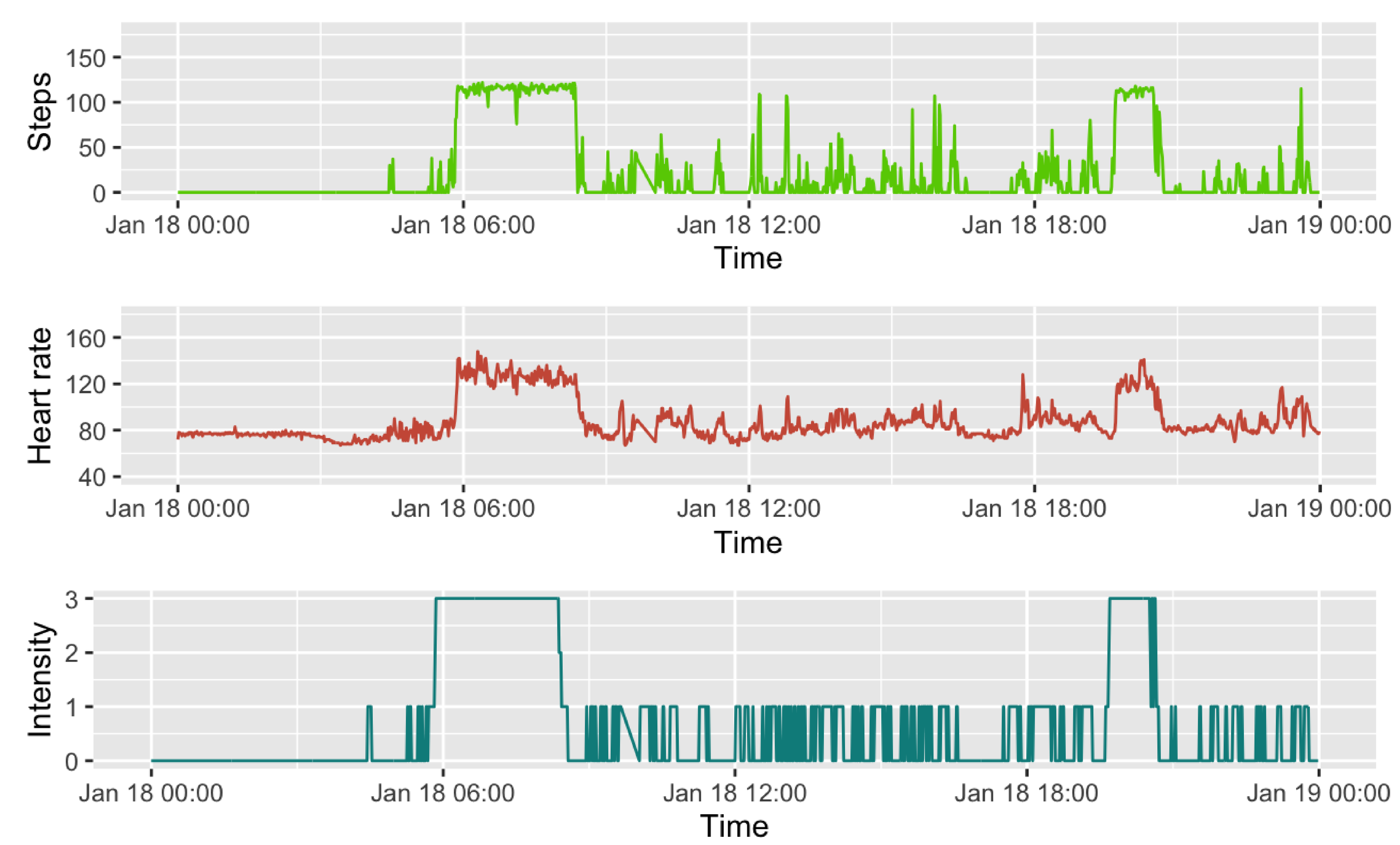

Figure 2. Time series plots of a random day.

Parameter Estimation

1. MLE of poisson distribution: $\lambda = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$
2. MLE of negative binomial distribution: $p = \frac{r}{\bar{x}+r} = \frac{size}{size+\mu}$, where r is the number of success.
3. MLE of geometric distribution: $\theta = \frac{1}{(\bar{x})}$

Results

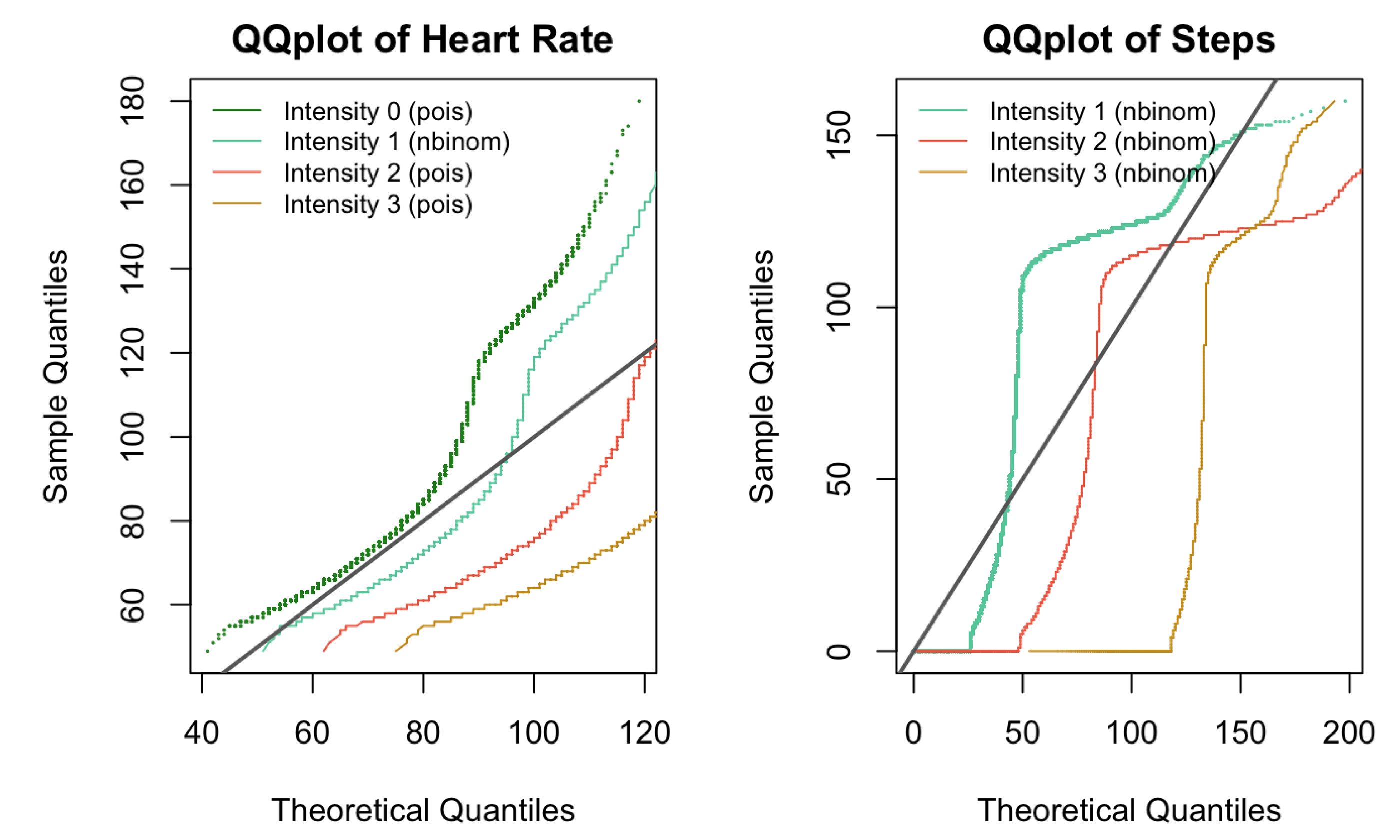

Figure 3. QQplots of heart rate and steps.

	Intensity0	Intensity1	Intensity2	Intensity3
Heart rate	$\lambda = 76.9235$ (0.0227)	size = 625.1252 (30.5771) $\mu = 86.8563$ (0.0423)	$\lambda = 103.9582$ (0.2222)	$\lambda = 120.6617$ (0.0756)
Steps	NA	size = 625.1252 (30.5771) $\mu = 86.8563$ (0.0423)	size = 625.1252 (30.5771) $\mu = 86.8563$ (0.0423)	size = 625.1252 (30.5771) $\mu = 86.8563$ (0.0423)

Table 1. Parameter estimates from the best model under each category with standard error (in parentheses).

The distributions in parentheses has the lowest BIC comparing to the other two. Nevertheless, as illustrated in Figure 3 and 4, these three distributions are not the best fit.

Acknowledgements

Many thanks to Dr. Agus Salim for valuable guidance.

Reference

[1] Witowski, V., Foraita, R., Pitsiladis, Y., Pigeot, I., Wirsik, N. (2014). Using hidden markov models to improve quantifying physical activity in accelerometer data–a simulation study. PloS one, 9(12), e114089.

[2] Wit, E., Heuvel, E. V. D., Romeijn, J. W. (2012). 'All models are wrong...': an introduction to model uncertainty. Statistica Neerlandica, 66(3), 217-236.