Tracking Physical Activity State: An implementation of Hidden Markov Model using information from wrist-worn devices

Quoc Anh Nguyen (supervised by Prof. Agus Salim)

The University of Melbourne

Motivation

Recently, increasing awareness of health risks from inactive lifestyles encouraged the necessity to monitor people's levels of physical activity (PA). As a result, there has been ongoing competition in the wearable activity tracker market in both prices and technologies. In general, state-of-the-art methods are applied for the analyses of wrist movement patterns to capture the PA state of a person. However, those approaches are not only cost-effective but also overestimate step counts from non-exercise activities with a high level of wrist movement such as folding laundry and playing video games. In this project, we aim to optimize the PA estimators from systematic biases, with the use of data on heart rates and step counts.

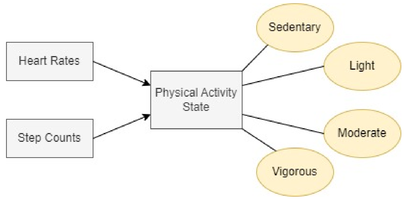

Figure 1. Diagram: The main goal for the project

Discussing potential methods

The dataset contains records about heart rates, step counts, and Fitbit classification for a total of 227,587 consecutive time points between 6 AM and midnight. With this unsupervised classification problem, we first attempt to fit a K-Means Clustering, which is a very common algorithm for this type of problem:

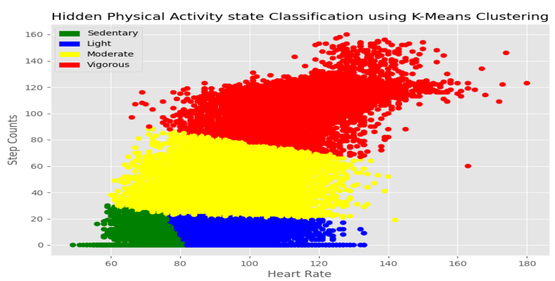

Figure 2. Classification using K-Means Clustering with heart rates and step counts as predictors

In a total of 227,587 time points, K-Means produced 162,389 identical results to Fitbit (71%), which is not a bad performance. However, one disadvantage of this algorithm is that they treat each time point as independent, which completely ignores the temporal correlation between the steps and HRs data. For example, a person with the previous PA state classified as 'Vigorous' will have different statistics compared to one who has a prior 'Sedentary' activity state.

For this reason, we introduce the Hidden Markov Model (HMM), which is better at utilizing this correlation and thus improves our prediction.

The Hidden Markov Model

For this project, our Hidden Markov Model will try to predict the hidden physical activity states $(q_1, q_2 ... q_T)$ using observations (heart rates and step counts) $(o_{11}, o_{21} ... o_{1T}, o_{2T})$. The model is specified by the initial probability distribution over states π, the transition probability matrix A, and the emission probabilities.

Finally, the HMM needs to satisfy two assumptions:

- **Markov Property**: the probability of a particular state depends only on the previous state
$$P(q_i|q_1 ... q_{i-1}) = P(q_i|q_{i-1})$$
- **Output Independence**: The observations depend only on the current state, not on any other states or observations: $P(o_{1i}, o_{2i}|q_1 ... q_T, o_{11} ... o_{1T}, o_{21} ... o_{2T}) = P(o_{1i}, o_{2i}|q_i)$

Modelling

Now, we will start implementing a Hidden Markov Model using both the heart rates and step counts data. The modelling part is divided into 2 steps:

1. **HMM Training using the Baum-Welch algorithm:**
In order to estimate the transition probability matrix A and the emission probabilities, we will use the Baum-Welch algorithm. The algorithm is *iterative*, which starts with initial probabilities for the state transition and the emission, then computes better estimations based on the previous ones. It terminates when the estimates are close to convergence, or we reach the maximum limit of iterations.
In this case, we fit a Poisson distribution for the heart rates and step counts in the dataset, and start estimating parameters for the model:

Table 1. Estimate for the emission probabilities (HMM with heart rates and step counts)

Predictors	Sedentary	Light	Moderate	Vigorous
Heart Rates	Poisson(77.07)	Poisson(84.99)	Poisson(90.29)	Poisson(118.26)
Step Counts	0	Poisson(11.60)	Poisson(37.33)	Poisson(111.07)

Table 2. Estimate for the transition probability matrix (HMM with heart rates and step counts)

	Sedentary	Light	Moderate	Vigorous
Sedentary	0.775	0.108	0.049	0.068
Light	0.637	0.097	0.178	0.088
Moderate	0.304	0.544	0.000	0.152
Vigorous	0.226	0.174	0.318	0.282

2. **Decode the hidden PA-state sequence:**
In this step, we run the Viterbi algorithm to determine the most probable sequence of physical activity states, given the heart rates and step counts in the dataset, and with the Hidden Markov Model with estimated parameters.
The result will then be compared to the Fitbit classification, which will serve as the truth values to evaluate our model performance.

Two other Hidden Markov Models that only consider the step counts, or the heart rates will also be implemented. Thus, we could see the benefits of minimizing systematic biases from wrist-worn devices by including additional observation values.

Interpretation of results

Now, we evaluate the performance of each model based on the number of similar outcomes and matching rate with the Fitbit classification:

Table 3. Comparision of the 4 models with Fitbit classification

	Number of matches	Matching rate
K-Means	162,389	71%
HMM with HRs only	163,368	72%
HMM with steps only	203,935	89%
HMM with HRs and steps	205,525	90%

All Hidden Markov Models show a greater number of identical results and a higher degree of matching, compared to the K-Means algorithm, which highlights the significance of the temporal correlation between HRs and step counts when building such predictive models. Among the HMMs, it is not surprising that the model using both the heart rates and step counts data has the highest matching rate (90%). The HMM which uses only the step counts gets quite close, with an 89% matching rate. Meanwhile, the model utilizing solely the heart rates data has a matching rate of 72%, which is only slightly higher than the K-Means.

Next, we look at the confusion matrix from the Hidden Markov Model with both heart rates and step counts:

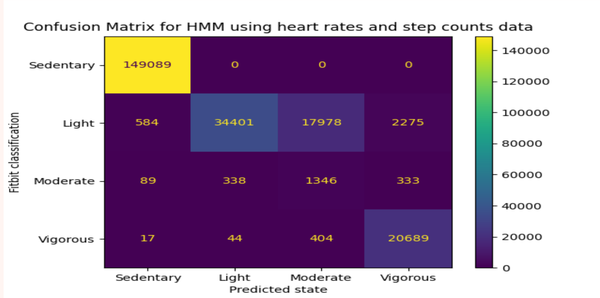

Figure 3. Confusion Matrix for the Hidden Markov Model with heart rates and step counts data

Even though our model has a 90% matching rate compared to Fitbit classification, only about 60% of the light and moderate activities labeled by Fitbit classification align with our predictive model. This is not unexpected since they are 'middle' physical activity states, which are very difficult to predict correctly. In reality, people also often mistake light activity for moderate activity, and vice versa.

References

[1] Sean R Eddy.
What is a hidden markov model?
2004.

[2] Daniel Jurafsky and James Martin.
Speech and Language Processing.
2021.

[3] Vitali Witowski, Ronja Foraita, Yannis Pitsiladis, Iris Pigeot, and Norman Wirsik.
Using hidden markov models to improve quantifying physical activity in accelerometer data – a simulation study.
PLOS ONE, 2014.

quocanhn@student.unimelb.edu.au