Predicting Sleep Using Consumer Wearable Sensing Devices

Miguel A. Garcia

Stanford University
EIGHT Lab

Motivation

It is increasingly becoming more common for patients to present data acquired from consumer wearables to clinicians. However...

- The efforts undertaken to validate the accuracy and reliability of consumer wearable health devices have been minimal.
- Sleep researchers have yet to define standard metrics for validation.
- Gold-standard wearable device companies keep information about sensors' raw data, accuracy, and algorithms hidden as trade secrets.

Using machine learning, what can we infer about sleep trackers, and how can we improve on the current models?

Data

This dataset, which was provided by the Emergent Innovative Global Health Technologies (EIGHT) lab at Stanford, used a Basis Peak watch to track user health and sleep habits for several study subjects over the course of months, with user information being collected by the second.

There are 7 features provided with each training example, and labels for wakefulness, rem sleep, light sleep, and deep sleep with each. PCA was performed over the entire moth of January 2016, and while no feature was discarded, it was interesting to see that the first 2 principal components were heart rate and galvanic skin response (gsr), with variance ratios of 0.47 and 0.14 respectively.

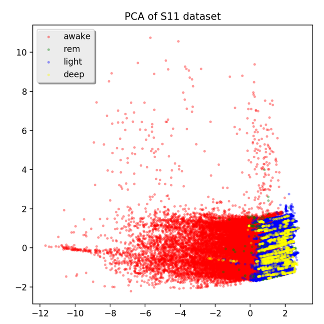

Figure 1: Figure caption

Binary models

LOGISTIC REGRESSION

We learn the parameters θ_i of this model by maximizing

$$l(\theta) = \sum_{i=1}^{m} y^{(i)} \log(h(x^{(i)})) + (1-y^{(i)}) \log(1-h(x^{(i)})) \quad (1)$$

SUPPORT VECTOR MACHINES

Given training vectors $x_i \in \mathbb{R}^p, i = 1 \dots n$ and a vector $y \in \{1,-1\}^n$, SVM's can solve the binary classification problem

$$\max_{\alpha} W(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} y^{(i)} y^{(j)} \alpha_i \alpha_j K(x^{(i)}, x^{(j)}) \quad (2)$$

s.t. $0 \leq \alpha_i \leq C, i = 1, \dots, m, \Sigma_{i=1}^{m} \alpha_i y^{(i)} = 0$

The following kernels were used for K:
linear : $\langle x, x' \rangle$, *polynomial* : $(\gamma \langle x, x' \rangle + r)^d$

Multi-class models

K-NEAREST NEIGHBORS

Assign class j to $x^{(i)}$ that maximizes the following:

$$P(y^{(i)} = j | x^{(i)}) = \frac{1}{k} \sum_{i \in N} 1\{y^{(i)} = j\} \quad (3)$$

SOFTMAX REGRESSION

We learn the parameters θ_i of this model by maximizing

$$l(\theta) = \sum_{i=1}^{m} \log \prod_{l=1}^{k} \left(\frac{e^{\theta_l^T x^{(i)}}}{\sum_{j=1}^{k} e^{\theta_j^T x^{(i)}}} \right)^{1\{y^{(i)} = l\}} \quad (4)$$

SUPPORT VECTOR MACHINES

SVMs can be used to build a multi-class predictor for k classes by using the following approach:

- **One vs. rest** k SVMs trained

Results

TRAINING = 28709, DEV = 12303, TEST = 18479

Classifier	Train	Dev	Test
Logistic Reg. (L2 penalty)	6.99%	7.32%	8.86%
SVM (linear)	6.83%	7.19%	9.15%
SVM (d = 2)	9.17%	9.68%	11.5%
SVM (d = 3)	6.61%	7.09%	8.09%

Table 1: Binary classifier error (hold-out cross validation)

Classifier	Train	Dev	Test
K-Nearest Neighbors	9.03%	13.43%	20.66%
Softmax Reg.	17.29%	17.25%	15.03%
Linear SVM	17.31%	16.88%	15.42%
SVM (d=2)	19.18%	19.03%	17.28%
SVM (d=3)	17.13%	16.91%	14.94%

Table 2: Multi-class classifier error (hold-out cross validation)

	'awake'	'asleep'
'awake'	12826	1029
'asleep'	466	4158

Table 3: Actual vs predicted for binary SVM(d=3))

	'awake'	'rem'	'light'	'deep'
'awake'	13032	3	820	0
'rem'	189	1	680	2
'light'	468	8	2685	4
'deep'	60	0	527	0

Table 4: Actual vs predicted for multi-class SVM(d=3)

Discussion

Although at first glance the algorithms perform well, our interpretation of the results changes once we realize the data is skewed towards wakefulness. Given only 7 initial features, our binary models perform acceptably. On the other hand, the multi-class models markedly fail to acceptably predict rem and deep sleep. In a different perspective, our results expose how easy it is for tracking devices to present information that is reasonable at face-value, but misleading.

Future Work

These results could be used to:

- Firstly, build a dataset such that the accuracy of predicting wakefulness does not come at the cost of predicting rem and deep sleep.
- Investigate whether datasets across different people can be integrated by building a mixture of Gaussians model.
- Update the refined predictors with features from sleep studies that collect smartphone usage data.
- Develop consumer devices that improve users' quantity and quality of sleep.

References

[1] de Zambotti M, Godino JG, Baker FC, Cheung J, Patrick K, Colrain IM. *The boom in wearable technology: Cause for alarm or just what is needed to better understand sleep? Sleep. 2016;In Press:1761-1762. doi:10.5665/sleep.6108.*

[2] Marino M, Li Y, Rueschman MN, Winkelman JW, Ellenbogen JM, Solet JM, Dulin H, Berkman LF, Buxton OM. Measuring Sleep: Accuracy, Sensitivity, and Specificity of Wrist Actigraphy Compared to Polysomnography. Sleep. 2013;36(11):1747-1755. doi:10.5665/sleep.3142.

Acknowledgements

I would like to thank Dr. Katarzyna Wac, PhD, for her guidance and support, and in general our colleagues at the EIGHT lab for their work in developing health solutions through technology.