An Innovative Way to Define and Exploit the Sleep Quality

Group 3: Litian Zhou, Bangyao Zhao, Chenyi Yu, Ningyuan Wang, Qingzhi Liu

April 27, 2020

Abstract

High quality sleep is the basis of a healthy lifestyle. Now there are many different ways to measure sleep quality, but it is often assessed via subjective methods with low accuracy. In this study, we defined two novel model-based sleep quality scores based on E4 data. The data were collected by Empatica E4 wristbands from 25 college students, including 3-axis Acceleration (ACC), Electrodermal Activity (EDA), Temperature (TEMP) and Heart Rate (HR). The models also considered the gender, exercise and working duration of each subject. We constructed an automated and reproducible data preprocessing pipeline, dealing with data import, missingness, batch effect, and feature extraction. Then, we fitted a logistic regression model and random forest model to give each 5-minute epoch two different sleep quality scores. Finally, we compare two models in terms of prediction accuracy, ROC curve and rationality among different sub-groups. In conclusion, Empatica E4 is capable of evaluating people's sleeping status and shed light on how to improve sleeping quality.

Key words: sleep quality, E4 Empatica, wearable device, logistic regression, random forest

Introduction

Sleep is a vital component of people's health, though it is often neglected. It enables our body to repair and recharge. Getting adequate and quality sleep can protect people's physical and mental health, and thus improve one's quality of life. In this study, the main data are collected by the Empatica E4 wristband. It is a wearable device that enables people to collect real-time physiological data in daily life. Key measurements include 3-axis Acceleration (ACC), Electrodermal Activity (EDA), Temperature (TEMP) and Heart Rate (HR). EDA measures an individual's nervous system arousal caused by stress or excitement. ACC measures one's physical activities and captures one's motion-based activities. Also, HR measures heart rates and TEMP measures one's body temperature. Hence, these four features are expected to be lower and more stable during the sleep period than during daytime. Sleeping quality generally refers to how well a person sleeps, which has not been rigorously defined and is hard to be quantified. In this study, the goal is to first define two novel model-based sleep quality scores and then find out which features are associated with higher sleep quality scores. The project follows the flowchart below:

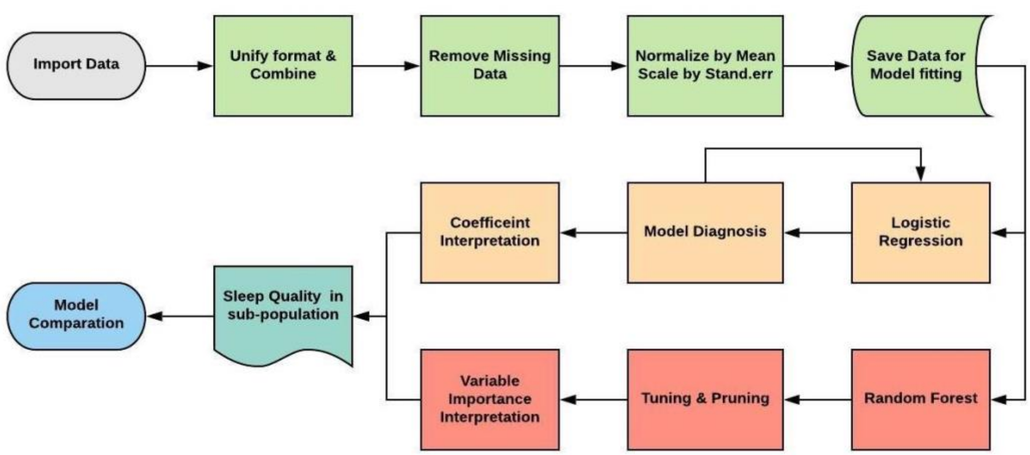

Figure 1. The workflow of the entire project. Green: data preprocessing; Yellow: Logistic regression; Red: random forest; Cyan & Blue: Model Interpretation and Application.

Data Description and Preprocessing

In the original dataset, there are 25 participants and 28 variables (See Appendix Table1). Demographic variables such as sex and age are included. There are 14 female and 11 male participants between the age 22 and 35 with a median age at 23 years old. There are 12 participants having vigorous exercise at least 30 minutes per day. In the original data, sleep status and E4 data have been processed and summarized by each 5-minute epoch. Sleep is a binary variable, representing whether a subject is sleeping or not. Mean, standard deviation, first quantile, median, and third quantile are calculated for each 5-minute epoch of ACC, EDA, TEAM and HR in each participant. Most participants have had the device on for over 37 hours.

The original dataset is downloaded from the course website in the format of xls. There are three steps used for data preprocessing. First, we unify the table format and coding scheme, which contains multiple types of labeling due to separate data submission. Then, we remove the missing data which only appear in one subject. Last, we normalize the all-numeric raw values by their means and scale by their standard errors, which alleviates the batch effect problem. The normalized data are saved on the GitHub page (Zhou, 2020), so other group members have access to the same processed data for further analysis.

Methods

The preprocessed data are used for modeling, where outcome is Sleep (a binary variable) and candidate covariates include demographic variables, daily diary-related variables and E4-related variables, which are described above. We randomly choose 70% participants for the training data and 30% participants for the test data. Two machine learning algorithms, logistic regression and random forest, are used for modeling. The reason to choose these two algorithms is that our previous project has shown that both of them have a good performance on our E4 data (including 5 participants). We use different procedures to select the best logistic regression model and the best random forest model using the training data. First, in logistic regression both forward and backward variable selection are used. Based on Akaike information criterion (AIC), a certain method of variable selection is chosen for the best logistic model. In this model, Cook's distance and residual standard error of fitted values are used to discover influential points and outliers. Second, in terms of the random forest algorithm, we use the training data to tune the parameters based on 5-fold cross validation. After model fitting, AUC-ROC curves are used to compare their performance based on the test data.

We use the whole dataset (instead of training data) to fit the final logistic model with the selected covariates in the best logistic model above, and the final random forest model using the parameters tuned in the best random forest model above. We propose two approaches to define the sleep quality score at each epoch: 1) the estimated probability of sleep from the final logistic model and 2) the estimated probability of sleep from the final random forest model. The sleep quality score ranges from 0 to 1, where a higher score corresponds to a deeper sleep state at a 5-minute epoch. After getting the sleep quality scores, we compare the scores in different subgroups defined by Sex and VE by using box plots and t tests.

Results

Important Covariates in Logistic Regression and Random Forest

As the backward selection model (AIC = 8282.7) has a smaller AIC than the forward selection model (AIC = 8286.4), the backward selection model is selected to be the final logistic model. Final Features are ACC_MEAN, ACC_SD, ACC_Q2, ACC_Q3, TEMP_MEAN, TEMP_Q1, TEMP_Q3, HR_SD, HR_Q1, HR_Q2, EDA_MEAN, EDA_SD, EDA_Q1, EDA_Q2.

We plot the Cook's distance and residual standard error of fitted values to discover influential points and outliers (See Appendix Figure 1). There are 3 points with extremely high influence and 52 outliers. Considering their fairly small amount compared to our total sample, we delete those points in order to make the logistic model more accurate and robust.

Variables	Log Odds Ratio	95% CI	p-value
(Intercept)	-2.01077	[-2.1, -1.921]	$< .001$
ACC_MEAN	-0.63288	[-0.897, -0.369]	$< .001$
ACC_SD	-0.58740	[-0.72, -0.455]	$< .001$
ACC_Q2	0.76595	[0.554, 0.977]	$< .001$
ACC_Q3	-0.27382	[-0.524, -0.023]	0.0322
TEMP_MEAN	-0.27601	[-0.374, -0.178]	$< .001$
TEMP_Q1	4.70325	[3.742, 5.664]	$< .001$
TEMP_Q3	-3.61769	[-4.554, -2.681]	$< .001$
HR_SD	-0.30014	[-0.407, -0.194]	$< .001$
HR_Q1	-1.24601	[-1.653, -0.839]	$< .001$
HR_Q2	-0.63994	[-1.084, -0.195]	0.0048
EDA_MEAN	5.19059	[2.603, 7.778]	$< .001$
EDA_SD	-0.81768	[-1.184, -0.451]	$< .001$
EDA_Q1	-3.31920	[-5.006, -1.633]	$< .001$
EDA_Q2	-1.13672	[-2.587, 0.313]	0.124

Table 1. Logistic Regression Summary

In our final model, one unit increase in HR_SD results in -0.30014 change in log odds ratio of sleeping (95% CI [-0.407, -0.194], $p < .001$), which indicates a lower heart rate is important to sleep, which is expected. The log odds ratios for TEMP_Q1 (4.70, 95%CI [3.742, 5.664], $p < .001$) and TEMP_Q3 (-3.62, 95%CI [-4.554, -2.681], $p < .001$) are in different directions, which can be explained by the absence of TEMP_SD, and this different direction implies when sleeping, people tend to have less variation in temperature. EDA_SD (-0.818, 95%CI [-1.184, -0.451], $p < .001$), EDA_Q1 (-3.319, 95% CI [-5.006,-1.633], $p < .001$) and EDA_Q3 (-1.137, 95% CI [-2.587, 0.313], $p = 0.124$) indicate low and stable EDA during sleeping.

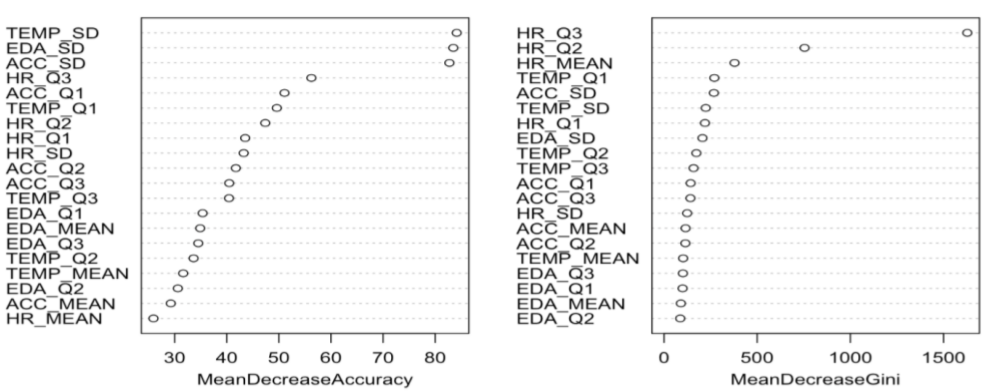

Figure 2. Important features in random forest

Combining **Table 1.** and **Figure 2.**, we find that the top five significant variables ranked by p-value in the logistic regression are TEMP_Q1, ACC_SD, TEMP_Q3, ACC_Q2 and HR_Q1. In the random forest model, the top five important variables ranked by mean decrease accuracy and mean decrease Gini are TEMP_SD, EDA_SD, ACC_SD, HR_Q2, and HR_Q3. In general, all selected covariates in both two models are either quartiles (i.e. Q1, Q2 or Q3) or standard deviation (SD), and it indicates the above two types of quantities are more robust than the others. In terms of types of health data, body temperature (TEMP) and acceleration (ACC) are more important in logistic regression, while heart rate (HR) is more important in random forest model.

Performance Evaluation

To verify the efficiency of the model, the ROC curve is used to examine the sensitivity and FPR. The data is divided into the training data (70% of the full data) and the testing data (30% of the full data). Both the two models are trained by the same training data, and both the two ROC curves are plotted by using the same testing data.

Figure 2. shows ROC curves and AUC for two models. Both models provide high AUC which indicates both models are good at predicting correct classes of outcome SLEEP. Relatively, the random forest model has a higher AUC at 0.967.

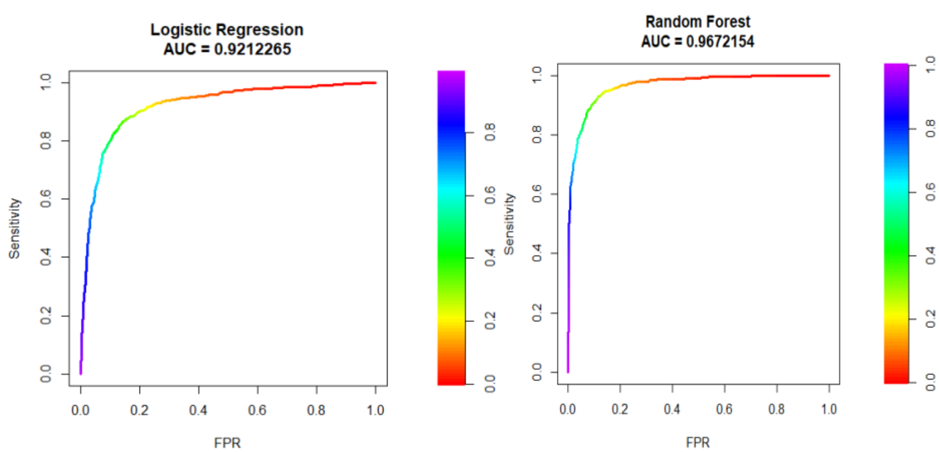

Figure 2. ROC-AUC of Two Models

In the logistic regression, the ROC curve determines the best cutoff is 0.353, which is to predict an epoch as sleep = 1 when sleep quality score > 0.353, and sleep = 0 otherwise. The result prediction accuracy (# of accurate prediction/total number) is 86.024%. In the random forest model, the optimal model ends with a parameter combination of *mtry* = 2 (i.e. the number of random covariates considered at each split) and *ntree* = 500. The accuracy of the model is 92%.

Sleep Quality Findings

Here are two interesting findings about sleep quality in different sub-groups.

A. *Sleep Quality vs. Vigorous Exercise*

By looking at sleep quality score between VE=1 (i.e. having vigorous exercise at least 30 minutes for one day prior to sleep) group and VE=0 group in **Figure 3.**, we find that the participants who had daily vigorous exercise have a higher sleep quality compared to the participants who did not have exercise.

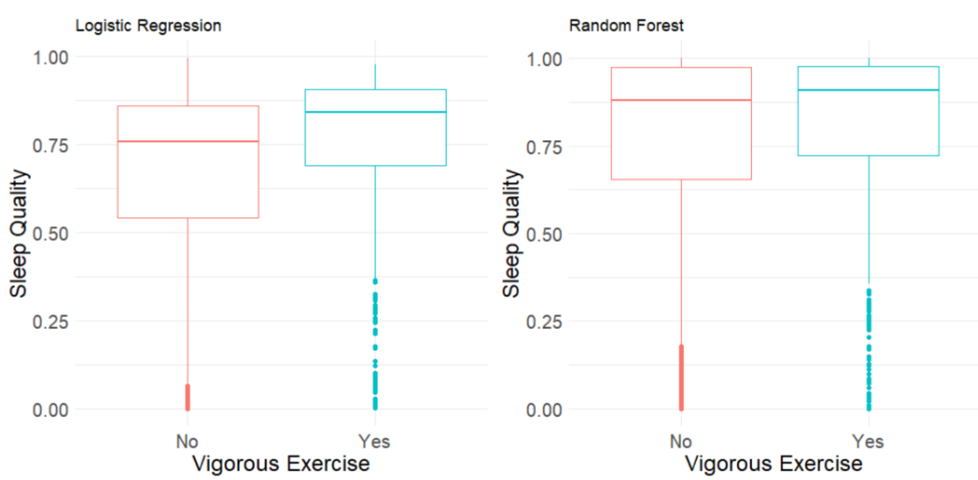

Figure 3. Boxplot of Sleep Quality vs. Vigorous Exercise

We perform t-tests (one-sided) to verify the fact that the sleep quality of the group with vigorous exercise is significantly larger than the group without vigorous exercise ($p < 0.001$). The differences are 0.083 and 0.015 by logistic and random forest model respectively.

B. Sleep Quality vs. Sex

In terms of sleep quality and sex , with the boxplot of sleep quality between male and female in **Figure 4.**, we find that male's sleep quality tends to be a slight better than female group (one sided t-test, $p < 0.001$ in logistic model, $p = 0.0395$ in random forest Model). The differences are 0.073 and 0.006 by logistic and random forest model respectively. Besides, the inter-quartile of male's sleep quality score is narrower than the female group, and it indicates that male's sleep quality are more stable than females.

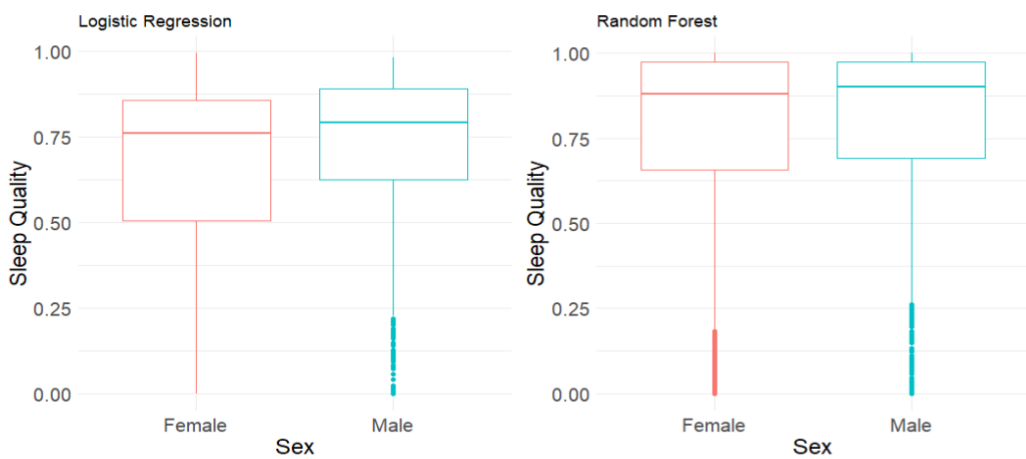

Figure 4. Boxplot of Sleep Quality vs. Sex

Conclusions & Discussion

Both logistic regression and random forest models predict sleep status with a high accuracy, which validifies the two novel types of model-based sleep quality score we defined. Our logistic model shows that almost all E4-related variables we used are significantly associated with sleep status. Both models illustrate that people who have daily vigorous exercise tend to have a higher sleep quality score compared to people who do not have exercise, and men have a slightly higher sleep quality score than women. The latter conclusion is contradicted with the previous finding that women tend to have a better sleep quality (Krishnan & Collop, 2006), which may be caused by potential confounders such as daily vigorous exercise and the small sample size. We also found roughly bell-shaped relationships (e.g. decreasing first and then increasing) of sleep quality score with daily exercise time and working time. To further validate the novel sleep quality score, in our future work we plan to find more interesting and meaningful relationships between daily diary (if available) and sleep quality score.

The limitation of study design is that the number of participants and the number of observed E4 data points of each participant are small. Also, since the original timeline of epoch is not included in this dataset, our models didn't consider the within-subject correlation, which might jeopardize the model robustness. These limitations do not affect the approach of calculating our novel sleep scores, but it could have an impact on the validity of some conclusions.

Reference

https://journals.lww.com/co-pulmonarymedicine/Fulltext/2006/11000/Gender_differences_in_sleep_disorders.3.aspx

Krishnan, V., & Collop, N. A. (2006). Gender differences in sleep disorders. *Current Opinion in Pulmonary Medicine*, *12*(6). https://journals.lww.com/co-pulmonarymedicine/Fulltext/2006/11000/Gender_differences_in_sleep_disorders.3.aspx

Zhou, L. (2020, March 5). *Data source*. https://github.com/LitianZhou/620_group_project/blob/master/data/total-data_final_normalized.rds

Appendix

Variables	Discription	Type
E4_ID	Unique Participant Identifier	Number
SLEEP	Sleep (1), Awake (0)	Factor
SEX	Female (1), Male (0)	Factor
AGE	Age in years	Number
VE	Vigorous exercises with Yes(1), No (0)	Factor
D.VE	Duration of vigorous exercises in minutes	Number
Day	Mon - Sun (1-7)	Factor
W.HOURS	Number of work hours per day	Number
ACC	ACC_MEAN, ACC_SD, ACC_Q1, ACC_Q2, ACC_Q3	Number
EDA	EDA_MEAN, EDA_SD, EDA_Q1, EDA_Q2, EDA_Q3	Number
TEMP	TEMP_MEAN, TEMP_SD, TEMP_Q1, TEMP_Q2, TEMP_Q3	Number
HR	HR_MEAN, HR_SD, HR_Q1, HR_Q2, HR_Q3	Number

Note: For last four variables, MEAN, SD, Q1,Q2, Q3 means mean, standard deviation, first Quantile, median, third quantile, repectively

Table 1. Descriptive statistics of collected data

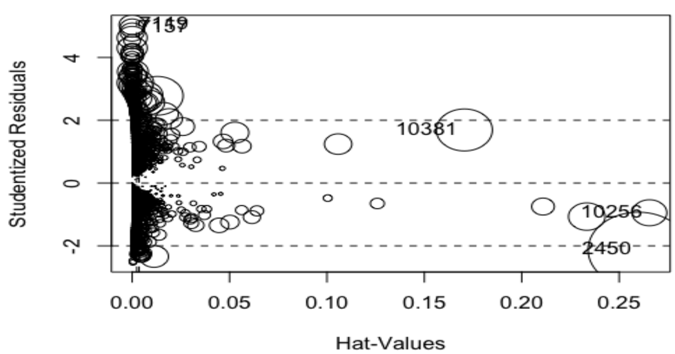

Figure 1. Logistic diagnostics figures. Studentized residuals vs fitted value, whereas the size of circle is the Cook's distance per observation.