

Sleepiness Prediction Using Empatica E4

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

April 6, 2020

Abstract

In this study, we built a statistical model and applied machine learning methods to predict whether a person is sleeping based on the data from the wearable device Empatica E4. A group of five subjects took participation in the data collection, including 3-axis Acceleration (ACC), Electrodermal Activity (EDA), Temperature (TEMP) and Heart Rate (HR)₁, all measured by Empatica E4. Then, we constructed an automated and reproducible data preprocessing pipeline, dealing with data import, missingness, batch effect, and feature extraction. And we fitted a logistic regression model, as well as several machine learning models, including SVM, KNN, and Random Forest. At last, we chose Random Forest as the best model according to its prediction accuracy. In conclusion, Empatica E4 is capable of detecting people's sleeping status with the average heart rate as the most important feature.

Keywords: sleepiness prediction, wearable device, logistic regression, machine learning, sleep quality

Introduction

Sleep allows our brain to rest and recharge. The quality of sleep determines the quality of work during the daytime and consequently, the quality of life. Therefore, the importance of obtaining adequate sleep among college students cannot be overemphasized. Measuring the quality of sleep using electronic devices is increasingly popular recently and the accuracy of such measurements remains skeptical. In this study, the data is collected by the Empatica E4 wristband. It is a wearable wireless 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). Our study aims to integrate five subjects' data and then to detect the sleeping status with the best model and selected features.

Data Description

The dataset contains data from five participants in our group. Most of us wore the E4 device for 40 hours or more. We selected the following features of interest: heart rate, temperature, electrodermal activity, and 3-axis acceleration. There is no outlier detected. Also, missing values only appeared at the beginning and the end of the data series. The range and mean of those features were summarized in the following **Table 1**.

	Duration Total (sleep) Hours	HR Range (mean) bpm	TEMP Range (mean) °C	EDA Range (mean) μ S	ACC Range (mean) 1/64 g
Qingzhi Liu	34.30 (6.00)	[45.12, 201.65] (74.29)	[34.95, 39.16] (35.85)	[0.00267, 9.61] (0.27)	[1.41, 221.13] (62.74)
Ningyuan Wang	45.24 (13.33)	[50.56, 159.28] (71.73)	[26.41, 36.45] (33.99)	[0.001281, 7.39] (0.40)	[2.45, 220.55] (65.10)
Chenyi Yu	43.27 (15.05)	[52.00, 166.3] (73.61)	[27.57, 38.29] (35.66)	[0.01565, 11.46] (0.81)	[3.74, 193.38] (63.29)
Bangyao Zhao	46.50 (17.83)	[44.42, 186.48] (78.24)	[31.37, 36.16] (32.66)	[0.000267, 8.05] (0.78)	[2.24, 221.7] (65.10)
Litian Zhou	39.60 (8.26)	[41.53, 196.83] (76.89)	[34.55, 38.41] (35.23)	[0.00027, 9.61] (0.68)	[2.83, 221.70] (65.65)

Table 1. Distributive Statistics

Data Preprocessing

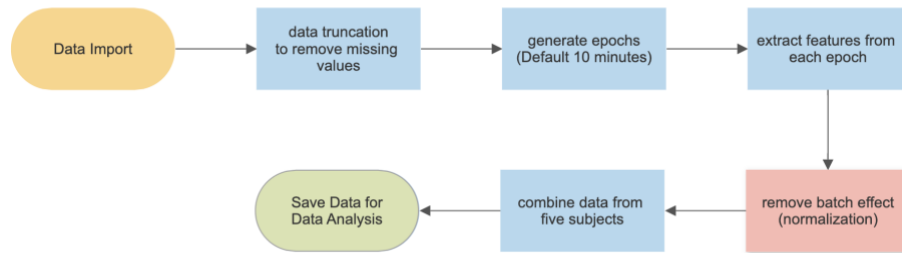


Figure 1. Project Workflow

As a group of five, we decided to merge the data before any data analysis to avoid “re-building the wheels.” However, each person has different starting times, durations and numbers of interruptions. Moreover, the raw values of biometric data due to individual health conditions may introduce batch effects to the model fitting. Lastly, we need to code a binary outcome: Sleep_Status, according to personal diaries. Thus, prior to merging, a procedure of data preprocessing was performed.

First, the R function `data2df()` reads a person’s data and transforms all data into a data frame, whose rows represent one observation in one second. By truncating the heads and tails of data series, we sacrificed a small amount of data to avoid missingness problems.

The R function `epoch_generation()` reads the data frames output by the `data2df()` and the user-specified length of the epoch to calculate the epoch statistics. For each type of data series, the function calculates its standard error and mean. The function returns a new data frame with 9 variables, which are epoch standard error and mean for all four features plus the subject ID (name). We assumed each epoch is independent of each other.

Finally, we checked the batch effect of our data. Luckily, there is not a big difference between individuals (See **Appendix Figure 1**). Nevertheless, we standardized the data to eliminate potential batch effects by dividing its mean values during sleeping time, i.e. we transferred the epoch values into its scale to resting status.

Methods

The integrated data after batch effect removal were used for analysis. Four classification algorithms were applied for predicting Sleep_Status, which are logistic regression, support vector machine (SVM), k-nearest neighbor algorithm (KNN), and Random Forest. The covariates are mean_acc, sd_acc, mean_eda, sd_eda, mean_temp, sd_temp, mean_hr and sd_hr. For each algorithm, the performance was measured based on 5-fold cross-validation. Each fold corresponds to the data of one subject since we have already removed the batch effect and also five subjects have a similar length of data. Specifically, we used the data of four students for model training and predict Sleep_Status of the subject left out using the fitted model. After five iterations of model training and testing, we get the predicted results of Sleep_Status for each subject at each epoch. The accuracy was calculated and compared in Table 2 in the Results section. We fitted the final model using the best algorithm based on five subjects' data. In the end, according to the criteria of variable importance (i.e., “mean decrease Gin Index for the random forest, and p-value of Wald test for the logistic regression), we select the variable that is highly associated with Sleep_Status.

Results

As shown by **Table 2**, the random forest model outperformed logistic regression, SVM and KNN in terms of “total accuracy”, and its “total accuracy” is equal to 90%. All four models achieve high

accuracy (>94%) when predicting Litian Zhou's Sleep_Status, while the accuracies of SVM and logistic regression are lower than 80% when predicting Chenyi Yu's Sleep_Status.

	Random Forest	Logistic Regression	KNN	SVM
Litian Zhou	0.95	0.95	0.94	0.94
Bangyao Zhao	0.91	0.87	0.89	0.87
Qingzhi Liu	0.85	0.87	0.87	0.87
Ningyuan Wang	0.96	0.96	0.86	0.88
Chenyi Yu	0.84	0.78	0.82	0.7
Total	0.9	0.89	0.87	0.85

Table 2. Prediction accuracy of four algorithms. The 2nd– 6th rows correspond to the prediction performance on different testing data based on 5-fold-cross validation. The last row corresponds to “total accuracy”.

The above optimal random forest model has two tuning parameters with $mtry = 2$ (number of variables randomly sampled as candidates at each split) and $ntree = 480$ (number of trees to grow). Based on the same parameters, the random forest model trained by all subjects' data is used to check the importance of each variable in the model. **Figure 2** shows that `mean_hr` has the highest variable importance score (Mean Decrease Gini), which is much larger than the importance scores of other variables. It illustrates that among the eight variables shown in the figure, `mean_hr` has the strongest association with Sleep_Status. In conclusion, based on the random forest algorithm, the integrated E4 data (after our preprocessing procedure) can predict sleep status with high accuracy (~90%), and HR data are more important than other types of data (e.g. TEMP, ACC, and EDA) when predicting sleep status.

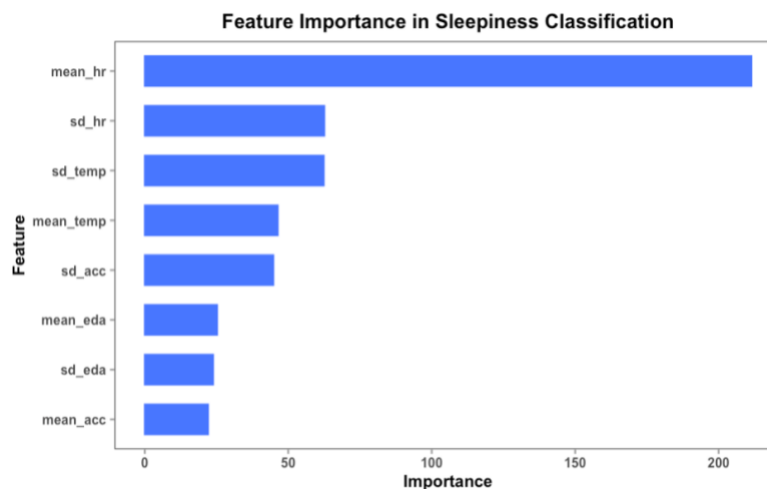


Figure 2. Feature Importance (Mean Decrease Gini) in the optimal random forest model.

Discussion

logistic probability as sleep quality score

One promising part of logistic regression is being able to compute a continuous score for each epoch, which is interpreted as the probability of sleep in classic statistical view. Is it possible to give it another more meaningful interpretation? Here is a plot of one group member's sleep in two nights.

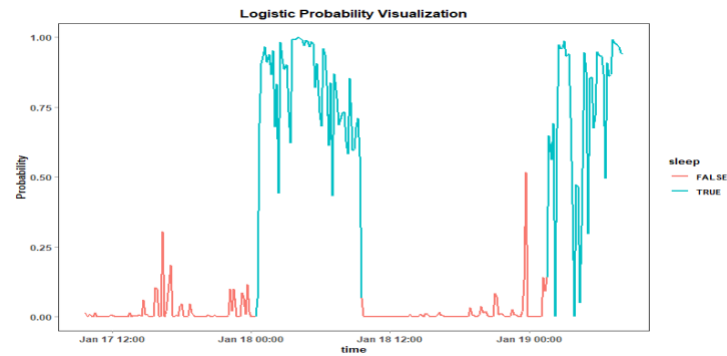


Figure 3. *Logistic Probability Visualization*

In his/her diary, the second-night sleep quality was bad due to the effect of coffee, while the first-night sleep was good. Interestingly, in the plot above, the first night's logistics probability curve is more well shaped and has higher mean than the second night, which indicates that the traditional logistic probability could have other specific interpretation as sleep quality score. Similar patterns have been found in other group members' data.

limitations

Lacking sufficient sleeping data, we cannot label deep sleep and light sleep, thus we are not able to justify the absolute rightness of this interpretation. Sleep periods are defined by user tags, and epochs in sleep periods will be labeled as sleep = T. Therefore, there doesn't exist a real sleep quality score for us to check the correctness of our algorithm. However, the sleep quality score is a very promising part of future exploration. Besides, our models do not consider the within-subject correlation, so it may affect the validity of statistical inference.

Conclusion

In conclusion, we achieved the study objective that Empatica E4 can be used for sleepiness prediction. We also achieved high prediction accuracy using the Random Forest and statistical inference from the logistic regression.

Thinking and Feeling

The data preprocessing took a long time. Not only do we need to take missingness and batch effects into consideration, but we also need to consider whether the constructed observations satisfied the model assumption. However, after the data was streamlined, we can easily try many different types of models.

The treasure of data is usually hidden behind enormous noise and unknown phenomenon, which is often difficult to explore and discover. After this project, I think this is true. After a long time and sometimes tedious data analysis, we are so glad to see the power of E4 data and how much it can benefit human health.

Reference

1. <https://support.empatica.com/hc/en-us/articles/201608896-Data-export-and-formatting-from-E4-connect->
2. All codes generating this report are curated in https://github.com/LitianZhou/620_group_project

Appendix

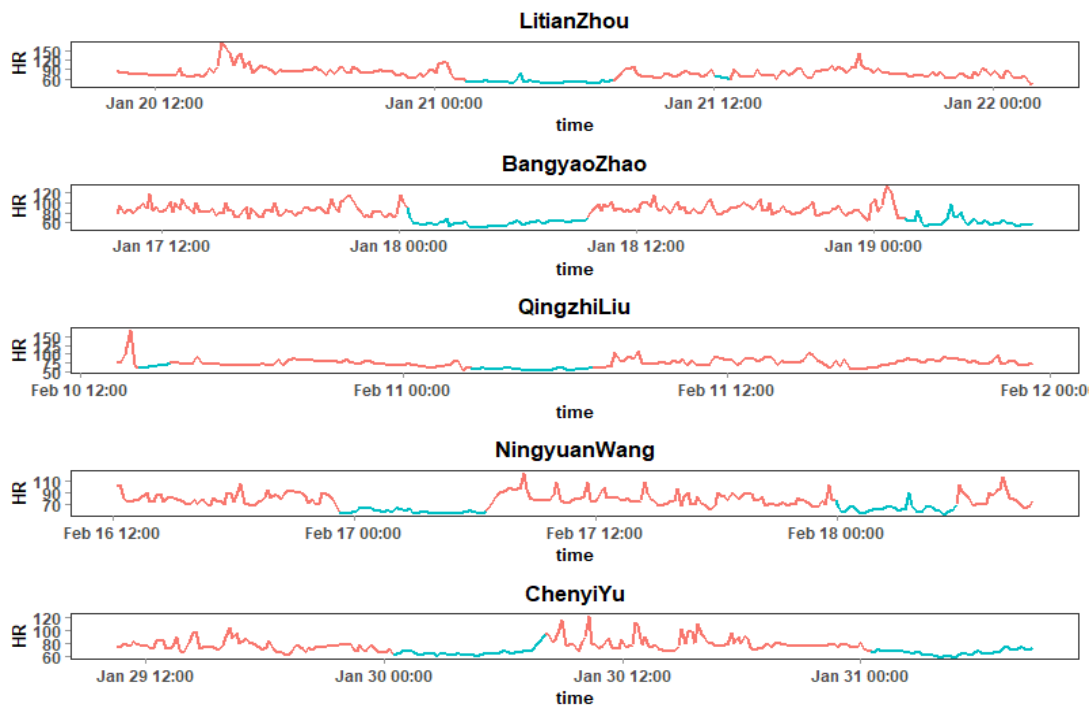


Figure 1. The heart rate of group members.

The color of the lines indicates whether the participant was sleeping (Red: not sleep; Blue: sleep). We can see that heart rate is lower during sleep times.