Sleepiness Classification/Quantification for College Students

BIOS620 Project II Proposal
2020.03.22

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

Summary:

How was your sleeping last night? This is the common question people always ask when they greet each other in the morning. It is a daily question many of us "brain workers" feel important. In our project, we aim to quantify the sleepiness by the biometric data including heart rate (HR), electrodermal activity (EDA) plus physical activity data from the 3-accelerometer sensor (AAC). First, we will label our data with "sleep" as 0 or "non-sleep" as 1 with reference to our daily journal. Then, we will integrate the standardized data from five students. train a potential logistic model to predict whether a person is sleeping or not. Then the fitted value will be transformed into a continuous score (sleepiness score) like the fitted value in logistics regression. It can be interpreted as either sleep quality or the degree of sleepiness. We will perform the cross-validation and if the model has a high prediction accuracy, we can help professor Song identify students with a high sleepness score in the class and give students suggestions to avoid sleepiness during the lecture.

Feasibility:

We propose to divide the data into epochs. We will treat each chunk as an observation and each summary statistic as a variable, which enables us to do logistic regression. We will choose the larger size of the chunk so that they are assumed to be independent. We can find the best length of epoch (default as 30 minutes) to optimize the prediction accuracy. After combining the data, we have a dataset with around 400 observations which is large enough compared to the number of features. Also, to eliminate the batch effect (individual differences), potentially we will standardize the data by transforming the raw data into the ratio to the sleeping average or the percentile. The final standardized method is TBD.

In summary, We make an assumption that each time epoch is independent of each other, so the validity of linear regression and logistic regression should not be affected. Also, we will test different standardization methods to reduce batch effects. In addition, our observation/feature ratio is high so the variance of estimates should not be a problem. Thus, we think the project is feasible and valid for statistical inference.