DTEK0086 Biosignal Analytics

Sleep stage classification using electromyography and electrooculography

Background
Sleep can be divided into distinct categories, including rapid eye movement (REM) sleep and non-rapid eye movement (NREM) sleep. In REM sleep, the eyes move rapidly from side to side, breathing becomes faster and irregular, and heart rate and blood pressure increase to near waking levels [1]. REM sleep is essential for healthy emotion regulation. NREM sleep consists of light sleep and deep sleep. Both light and deep sleep are essential for various processes in the body. The sleep stages can be measured using the polysomnography method, containing electromyography (EMG), electrooculography (EOG), and electroencephalography. Facial EMG shows differences between the sleep stages. EOG is a technique to collect the electrical activity of the eye's muscles. It can also be used to track eye movements.

Objective
The objective of this project is to perform sleep stage classification leveraging features derived from EOG and EMG signals. Using machine learning, you need to differentiate the records into three classes: i.e., awake, REM, and NREM. The analysis should be done in Python (more details in the Instruction Section).

For this course project, you need to:

1. Submit your Python script and your report of the observations, graphs, and conclusions made upon analyzing the given signals. It is suggested to submit a Jupyter Notebook file, including your code and report.
2. Give a 20-minute presentation about your work. Your presentation should include a description of
    a. The problem and the biosignals
    b. The steps in your analysis: e.g., what pre-processing methods you use, which features you extract, which machine-learning algorithms you use
    c. The results that you obtain: e.g., the accuracy of two machine learning methods
    d. Your evaluation and conclusion on the findings and methods

Data collection setup
The single-channel EOG was measured from the left side of the face (referenced to the contralateral ear lobe). The single-channel EMG was collected from the chin. The signals were measured in microvolts with the sampling frequency of 200 Hz. The data annotations were performed by human experts. The data is extracted from the Physionets You Snooze You Win challenge database (challenge-2018)

Structure of the data
The project includes the records of 100 subjects (i.e., 80 for training and 20 for test). There are 4-9 records for each subject. Each record (i.e., file) consists of one minute of EOG and EMG signals, and it corresponds to the awake, REM, or NREM stage. The dataset includes separate "Train" and

“Test” folders. The folders contain three subfolders as “awake,” “rem,” and “nonrem.” Each record is saved as a CSV file, including two columns corresponding to the EOG and EMG signals. The filename includes the event number and the subject number. For example, 100 is the event number, and 61 is the subject number in “100_subj_61.csv.”

Instruction

For the analysis, you should:

1. Use pre-processing techniques (such as filtering) if necessary.
2. Extract relevant time-domain and frequency-domain features from the EOG and EMG signals (e.g., summary statistics, RMS value, and resonance frequency).
3. Standardize your data: i.e., use the mean and standard deviation of the training data to standardize the training data and the test data.
4. Select two supervised machine learning algorithms and train two classifiers using the training set. Each classifier should predict 0 for “awake”, 1 for “rem”, or 2 for “nonrem”.
5. Compare the two classifiers by evaluating the results using the test set.
   a. Obtain the confusion matrix, accuracy, precision, recall, and F1-score. These can be calculated from the predicted and true values.

Hint: You can utilize packages such as scipy, tsfresh, and tsfel for the pre-processing and feature-extraction steps, and packages such as scikit-learn for the machine-learning step.

[1] Maier, C., M. Bauch, and H. Dickhaus. "Recognition and quantification of sleep apnea by analysis of heart rate variability parameters." Computers in Cardiology 2000. Vol. 27 (Cat. 00CH37163). IEEE, 2000.