DTEK0086 Biosignal Analytics

Sleep stage classification using electroencephalography

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 is further divided into four stages (i.e., N1, N2, N3, and N4 stages). N1 and N2 stages consist of light sleep, while N3 and N4 include deep sleep. Both light and deep sleep are essential for various processes in the body. Different brain activities can characterize REM, N1, N2, N3, and N4. The sleep stages can be measured using the polysomnography method, containing electroencephalography (EEG), electrooculography, and electromyography.

Objective
The objective of this project is to perform sleep stages classification leveraging features derived from EEG signals. Using machine learning, you need to differentiate the EEG segments into six classes, i.e., Awake, REM, N1, N2, N3, and N4. 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 EEG signals have been recorded with scalp EEG electrodes. The records contain two EEG channels: i.e., Fpz-Cz and Pz-Oz. The EEG records were sampled at 100 Hz. The data annotations were performed by well-trained technicians. The data is extracted from the Physionets Sleep-EDF database (Sleep EDF Expanded)

Structure of the data
The project includes the EEG of 12 different subjects (i.e., 9 for training and 3 for test). Each record (i.e., file) consists of 30 seconds of EEG signals, and it corresponds to the Awake, REM, N1, N2, N3, or N4 stage. The dataset includes separate “Train” and “Test” folders. The folders contain six subfolders as “Awake,” “REM,” “N1,” “N2,” “N3,” and “N4.” Each subfolder includes the EEG records. Each record is saved as a CSV file, including two columns corresponding to the Fpz-Cz

and Pz-Oz channels. The filename includes the event number and the subject ID. For example, 0 is the event number, and SC4001E is the subject ID in “sample_SC4001E_0.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 EEG signals (the EEG rhythms): e.g., summary statistics, power of the EEG rhythms, 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 (multiclass) supervised machine learning algorithms and train two classifiers using the training set.
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] National Institutes of Health, 2014. Brain basics: understanding sleep. NIH Publication, pp.06-3440.