DTEK0086 Biosignal Analytics

Blood pressure estimation using photoplethysmogram

Background

Blood pressure (BP) is an important cardiovascular parameter indicating the pressure of blood on blood vessel wall per unit area. BP is the main driving force that promotes the circulation of blood in the body. The most common non-invasive method to measure BP is to use a cuff-based BP monitor to detect Korotkoff sounds caused by turbulent blood flow. Pulse transit time – extracted from electrocardiogram (ECG) and photoplethysmogram (PPG) – can be also used to estimate systolic and diastolic BP values. However, these methods cannot be used to monitor BP continuously in everyday settings due to the limitations in their setups and data collection. Recent studies proposed to use machine-learning techniques to extract BP using only PPG signals [1].

Objective

The objective of this project is to extract systolic and diastolic BP from PPG signals. Using machine-learning algorithms, you need to train a regressor to estimate the BP parameters. 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 biosignal.
    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 mean absolute error of the machine learning methods.
    d. Your evaluation and conclusion on the findings and methods.

Data collection setup

Multi-parameter Intelligent Monitoring in Intensive Care (MIMIC) II online waveform database [2] provided by PhysioNet organization is used in this project. The data were collected from different hospitals between 2001 to 2008. The waveform signals include PPG

signals from fingertip and arterial BP detected by invasive sensors with unit of mmHg. All signals are sampled at 125Hz.

Structure of the data

The project includes the PPG of 29599 records (i.e., 23675 records for training and 5924 records for test). Each record (i.e., file) consists of 10 seconds of PPG signals. The dataset includes separate "Train" and "Test" folders. The folders contain 2 subfolders as "PPG" and "BP." The "PPG" subfolder includes the PPG records. Each record is saved as a CSV file, including one column corresponding to the amplitude value of PPG signal. The filename includes the index number of PPG segment. For example, the first 10-second PPG segment is named as "PPG_0.csv". The "BP" subfolder includes the BP records of each 10-second PPG segment. The CSV file contains three columns, including the systolic blood pressure (SBP), diastolic blood pressure (DBP), and the corresponding PPG segment name (e.g., PPG_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 PPG signals: e.g., interbeat interval (IBI), time interval between systolic and diastolic peaks, amplitude ratio between systolic and diastolic peaks, skewness and kurtosis, entropy etc. See [3] for more information.
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 regressors using the training set. Each regressor should estimate the SBP and DBP values using a PPG record.
5. Compare the two regressors by evaluating the results using the test set.
    a. Obtain the mean absolute error and concordance index (c-index). They can be calculated using 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] Zhang, Y. and Feng, Z., 2017, February. A SVM method for continuous blood pressure estimation from a PPG signal. In Proceedings of the 9th international conference on machine learning and computing (pp. 128-132).

[2] Goldberger, Ary L., et al. "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals." *circulation* 101.23 (2000): e215-e220.

[3] Elgendi, Mohamed. "On the analysis of fingertip photoplethysmogram signals." *Current cardiology reviews* vol. 8,1 (2012): 14-25. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3394104/)