

UDC 004.62

HEART RATE ESTIMATION FROM PHOTOPLETHYSMOGRAM AND ACCELERATION SMARTPHONE DATA BASED ON CONVOLUTIONAL NEURAL NETWORK AND LONG SHORT TIME MEMORY NETWORK

Qiu G.W., gr.167011, master

Belarusian State University of Informatics and Radioelectronics¹, Minsk, Republic of Belarus

Baryskievic A.A. – PhD in Technical Sciences

Annotation. The wearable reflective photoplethysmography (PPG) sensor can be integrated into the watch or strap to provide instantaneous heart rate (HRs), causing minimal inconvenience to users. However, the existence of motion artifacts (MAs) leads to inaccurate heart rate estimation. In order to solve this problem, I propose a new deep learning neural network to ensure accurate estimation of HR in high-intensity exercise. The average absolute error of the algorithm for all training data sets and test data sets is less than 1.5 bpm, including 1.09 bpm for training data sets and 1.46 bpm for test data sets.

Keywords. PPG, LSTM, Convolution, Concatenation, Heart rate and motion artifacts

Reflective photoplethysmography (PPG) sensor measures the intensity change of skin reflected light and provides PPG signal representing the change of arterial blood volume between systolic and diastolic phases of the heart cycle. The reason why the sensor is concerned is that it can be installed in a watch or wrist strap to measure and monitor instantaneous heart rate (HRs), minimizing the inconvenience to users. However, the sensor is sensitive to motion artifacts (MAs), which come from the pressure and motion exerted on the wrist of the PPG sensor. Motion artifacts will eventually lead to inaccurate heart rate estimation. A few years ago, Zhang and others shared the data set of acceleration and PPG signals measured simultaneously in [1] motion, which prompted people to study the use of acceleration signals to eliminate motion artifacts in PPG sensors. However, despite the efforts and progress of algorithms, these methods do not always provide accurate results.

In this paper, the purpose is to measure heart rate more accurately, so I propose a new deep neural network based on multi-class and non-uniform multi-label classification for human heart rate estimation. In the proposed algorithm, I consider two power spectra from the input layer of PPG and acceleration signals. In addition, I use the acceleration signal strength in the input layer. I assume that the acceleration signal strength can provide information about the recent HR changes: high intensity represents intense exercise, which may change HR. The proposed algorithm includes two convolution layers, two LSTM layers, one connection layer and three full connection layers (including one softmax layer). In the proposed algorithm, the power spectrum from PPG and acceleration signals is fed to two convolutions to provide MA cancellation in the PPG power spectrum. The output is flattened and connected to a fully connected layer, which is then connected to the acceleration signal strength. Then, the output is fed to two LSTM layers, and then to other full-connection layers including softmax.

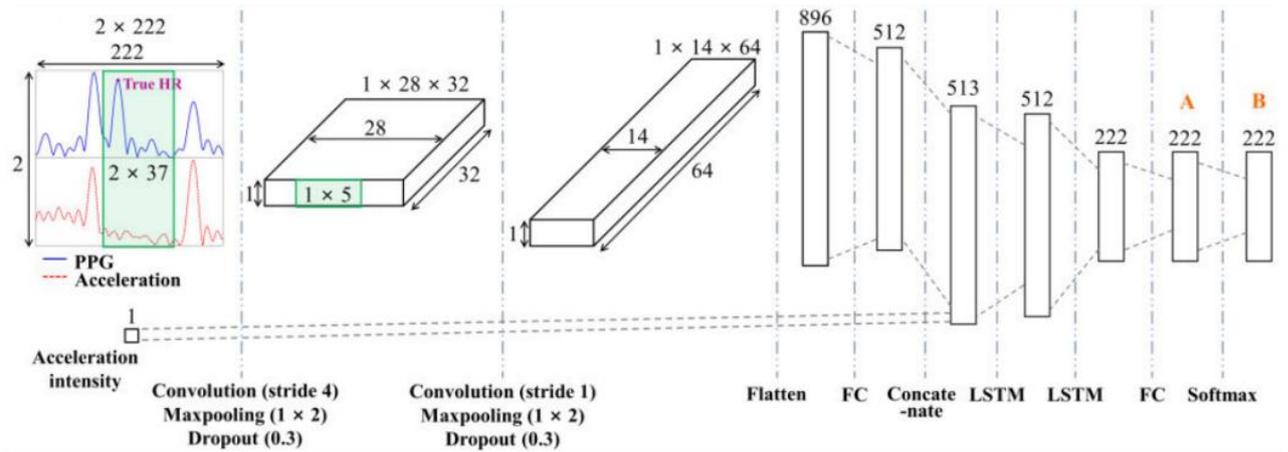


Figure 1. The architecture of the proposed algorithm; 2D convolution layer, 1D convolution layer, a full connection layer, a concatenation layer, two LSTM layers and a full connection layer are structured in sequence

The LSTM layer uses minimum outliers to track HR tracking. In this algorithm, I also propose a new scheme to modify the real HR value to Gaussian distribution to evaluate the loss value. The performance of the proposed algorithm is evaluated by comparing with the results of [1] - [18] previously reported.

I used two data sets to evaluate the proposed HR estimation algorithm - the IEEE Signal Processing Cup (ISPC) 2015 data set (n=23) and the direct measurement data set obtained by the developed equipment (n=48). These two sets of data include multi-channel PPG signal and three-axis accelerometer signal, which are measured by the equipment on the wrist during high-intensity physical exercise. At the same time, ECG signals are measured in the chest as real HR reference.

The data includes 12-minute three-channel PPG signal and 50-Hz sampling three-axis acceleration signal. The dataset is divided into two groups, BAMI-I and BAMI-II.

During every 4 minutes of running and walking, the subjects walked or ran with a treadmill stick in the last two minutes. I designed this link to reflect the heart rehabilitation exercise of patients with heart disease who have poor exercise ability. They usually walk or run with a treadmill stick. There are 17 males and 6 females, with an average age of 22.0 ± 1.7 years. The whole exercise process is also carried out on the treadmill. For these two data sets, the reference real heart rate was measured by ECG data recorded simultaneously by 24-h the ambulatory ECG monitor (SEER Light, GE Medical, Milwaukee, WI, USA).

In this study, I selected the ISPC data set (n=23) as the training data, because it includes various movements, such as waking up, running, resting, jumping, push up, shaking hands, stretching, pushing and boxing. I also added the own dataset BAMI-I (n=25) to the training data to increase the size of the training data, and used another dataset named BAMI-II (n=23) to test the training algorithm. For all data sets, the ecg-based HRs are calculated using an 8-second window, with a 2-second offset (6-second overlap) to obtain the HRs every 2 seconds. In the whole study, the same window length (8 seconds) and shift (2 seconds) are used to evaluate the performance of the proposed algorithm relative to the existing algorithm.

Figure 1 summarizes the architecture of the proposed algorithm. It consists of eight layers: two convolutional layers, two LSTM layers, one tandem layer and three full connection layers, including a softmax layer. More specifically, a 2D convolution layer, a one-dimensional convolution layer, a flat layer, a splice layer, a fully connected layer, two LSTM layers, and a fully connected layer followed by a softmax.

For the input layer, the power spectrum from PPG signal and acceleration signal is divided into two (size 2×222): the top signal is from PPG ($P_s(i)$), and the bottom signal is from acceleration ($P_a(i)$). Note that the power spectrum is based on each 8-second window and then shifted for 2 seconds (6-second overlap). Input layer input to 32×37 cores, two-dimensional convolution with a step of 4, and then perform the nonlinear function leaky corrected linear unit (ReLU) and 1×2 max pooling, in steps of 2. The resulting size is $1 \times 28 \times 32$. The characteristic diagram of 32 shows the intermediate PPG power spectrum after MA cancellation.

Send characteristic map into 64×5 -core, one-dimensional convolution with a side of 1, and then carry out leaky ReLU and $1 \times 2 \times 2$ max pooling, get other sizes of $1 \times 14 \times 64$, showing the PPG power spectrum of MA elimination. The resulting feature map is spread evenly to 896 nodes, which are fully connected to 512 nodes with leaking ReLU.

Subsequently, the acceleration intensity (I) is connected with 512 activations. I assume that the acceleration intensity represents how the current HR value will change in the near future. For example, high acceleration intensity means high intensity exercise, which may increase HR. 513 connected nodes are sent to two LSTM layers. These two LSTM layers act as HR tracking algorithms by considering the local HR tracking mode. Due to the existence of LSTM layer, even if the signal to noise ratio (SNR) of PPG signal is extremely low, the dominant frequency corresponding to HR will not deviate seriously in the continuous window. The first layer LSTM contains 512 nodes, each node has 6 time steps, and all nodes are connected to the second layer LSTM. Note that six time steps provide the best accuracy. The numerical analysis of the effect of time step will be introduced. The second LSTM layer also includes six time steps, each of which provides an output with a length of 222. Then, only the output from the last time step is fed to the full connection layer, which is connected to 222 nodes with leaking ReLU. In the frequency range of 0.6-3.3 Hz, the frequency resolution is 0.012 Hz, and the result of the final PPG power spectrum after MA cancellation is 222 activations. The activation is input to the softmax layer, which provides the final probability of the real HR value.

To avoid over-fitting, I applied dropout to two convolution layers and two LSTM layers. For convolution layer, the exit rate is set to 0.3. For LSTM layer, the drop rate of input linear transformation is 0.3, and the drop rate of loop state is 0.2.

List of used sources:

1. TROIKA: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise. / Z. Zhang, Z. Pi, B. Liu // *IEEE Trans. Biomed. Eng.* Feb. 2015. Vol. 62, no. 2, pp.
2. Photoplethysmography-based heart rate monitoring in physical activities via joint sparse spectrum reconstruction. / Z. Zhang // *IEEE Trans. Biomed. Eng.* Aug. 2015. Vol. 62, no. 8, pp.
3. A novel time-varying spectral filtering algorithm for reconstruction of motion artifact corrupted heart rate signals during intense physical activities using a wearable photoplethysmogram sensor. / S. Salehizadeh, D. Dao, J. Bolkhovsky, C. Cho, Y. Mendelson, K. Chon. // *Sensors*, 2016. Vol. 16, no. 1, p. 10.