UDC 334.029.3

PHOTOPLETHYSMOGRAPHY AND ACCELEROMETER SENSORS SIGNALS FOR RECOGNIZING PHYSICAL ACTIVITY

S.S. WEI

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received March 17, 2023

Abstract. The utilization of wearable devices to monitor human physiological parameters has been popularized, and due to their low cost, the most common method of monitoring human information in such devices is the use of photoplethysmography (PPG) signals. However, accurate estimation of the PPG signal recorded from the subject's wrist during various physical exercises is often a challenging problem, as the original PPG signal is heavily corrupted by motion artefacts. The article starts with an introduction to how PPG and Accelerometer (ACC) work, and then moves on to the programming, which is then used to provide data processing support for subsequent deep learning by importing data and calculating operations. Long short time memory (LSTM) is built for the paper to recognize activities. The experimental results showed that over 95 % accuracy was achieved in the classification of the test data.

Keywords: Photoplethysmography, Accelerometer, LSTM.

Introduction

The domain of Human Activity Recognition (HAR) has emerged as one of the most popular research topics since the availability, low cost and low energy consumption of sensors and accelerometers, real-time streaming of data, and advances in computer vision, machine learning, artificial intelligence, and the Internet of Things. In HAR, a variety of human activities, such as sitting, standing, walking, running, squatting, and resting, etc., are recognized. Data can be collected from wearable sensors, accelerometers, or images.

PPG is an electro-optical technique, in which the sensor is positioned above the skin and illuminates the skin surface by emitting green light, the sensor receives intensity changes of the reflected light, and the body state is gained through the periodic detection and analysis of the PPG signal. Such non-invasive method of real-time detection on human parameters assumes great practical importance. Numerous studies on the clinical application of photoelectric volumetric pulse waves have shown that the PPG signal contains many human physiological parameters and is an important tool for real-time monitoring of heart rate, blood oxygen saturation, blood pressure, vascular elasticity, etc. The acquisition of PPG signals requires only a special light source and a corresponding sensor, which can be easily integrated into everyday wearable devices to enable continuous monitoring of normal activities without causing discomfort, making PPG signals the preferred choice for health monitoring in everyday life.

However, accurate estimation of the PPG signal recorded on the wrist is often a challenging problem when people wear the wearable device for physical exercise, as the original PPG signal is heavily corrupted by motion artefacts (MAs), mainly due to the relative motion between the PPG source and the wrist skin [1–5]. In order to reduce MAs, a number of signal processing techniques based on data from different sensor types, particularly ACC data, have proven to be very useful [6].

ACC delivers information on the acceleration of the human body during movement. In smartphones and smartwatches, the built-in tri-axial ACC is probably the most common sensor for activity monitoring [7–8]. A combined approach for obtaining PPG and acceleration data is directly available on smartphones and smartwatches devices.

HAR can be regarded as a pattern recognition problem in which machine learning techniques have been proved particularly successful. Various machine learning methods models have been developed for HAR. The primary goal of this paper is to maintain good performance of RNN framework in terms of recognition accuracy, and a RNN was designed for detecting human activities using ACC and PPG four-dimensional data.

Dataset pre-processing

A recent publicly available dataset [9] was used which was from seven different subjects consisting of 105 PPG signals (15 per subject) and a corresponding 105 tri-axial ACC signals sampled at 400 Hz. The seven adult subjects included three males and four females, aged between 20 and 52 years, performing five series of resting, squatting, and stepping activities. The signals were acquired simultaneously and the dataset contained 210 audio clips with a total duration of 17,201 seconds. We use python language for our work.

PPG signals are continually captured during activities from the wrist using Maxim Integrated MAXREFDES100 device. To guarantee a perfect fit of the sensor unit to the skin surface, a specific weightlifting cuff, adjustable by tear-open closure, is used to hold the sensor in place by fully tightening the strap with a cable protruding from the back end of the strap. The PPG signal value is equivalent to the output of an ADC (Analog to Digital Converters) photodetector with a pulse width of 118 μs, a resolution of 16 bits and a full scale of 8192 nA, illuminated by a green LED (Light-Emitting Diode). The ACC signal values on the three axes correspond to MEMS (Micro-ElectroMechanical System) outputs with 10-bit resolution, left-aligned, and a scale of ±2 g.

The primary signal data collected is mixed, ideally regular and stable, and undesirable data is messy and unstable, but has some sort of regular trend, and can be made slightly more regular by filtering the undesirable data to suppress noise. The signal data are sampled according to time and traced to form the time domain data, which is a time referenced regional data.

In a preliminary cleaning step, the following cleaning steps were done on the raw data. If there are occasional spikes or NaN (Not a Number) points in the integrated four-dimensional data, the data from the former five different subjects are used for model training, and the data from the latter two groups of subjects for validation of the model, so only the data from the former five groups are processed.

The PPG signal values for the same subject are highly variable across series and vary considerably over short periods of time within the same series. Normalization allows for better separation of the PPG signal from the motion artefacts, with the following equations:

$$\mathrm{PPG}_{\mathrm{cal}} = \frac{\mathrm{PPG} - \mu_{\mathrm{PPG}}}{\sigma}, \tag{8}$$

$$\mu_{\mathrm{PPG}} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{PPG}_i, \tag{9}$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\mathrm{PPG}_i - \mu_{\mathrm{PPG}}\right)^2}, \tag{10}$$

where $\mathrm{PPG}_{\mathrm{cal}}$ is the calculated PPG data, PPG_i means PPG value of the *i*-th data, μ_{PPG} and σ are the average value and standard deviations of the original PPG data, respectively.

The accelerometer is affected by low level noise, the gravitational acceleration is in the three spatial axes of projection, so the data usually has some offset, we remove it by subtracting the average value from the data, the method gives a fine filtering of the signal with the following equation:

$$\mathrm{ACC}_{\mathrm{cal}} = \mathrm{ACC} - \mu_{\mathrm{ACC}}, \tag{11}$$

$$\mu_{\mathrm{ACC}} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{ACC}_i, \tag{12}$$

where $\mathrm{ACC}_{\mathrm{cal}}$ is the calculated ACC data, ACC_i means ACC value of the i-th data, μ_{ACC} is the average value of the original ACC data.

The sampling rate of 400 Hz for data can place a significant burden on hardware devices. For the purpose of efficiently downsampling the data, a resampling algorithm that requires a digital filter was not chosen because it would add significant computational cost to the final embedded system implementation. Accordingly, a simple extraction process is implemented in which 1 of the R samples is retained and the rest discarded.

A limited combination of parameters was examined in the vicinity of those already tested, and with a downsampling factor of 10, the best accuracy was achieved when the sample window (before downsampling) was 1200, corresponding 3 seconds with 50 % overlap, and a total of 100 training epochs.

The number of inputs for resting, squatting, and walking activities for the five subjects used for training varies greatly, and the network may end up being biased toward a particular class due to the large difference in numbers. A simple technique to solve this problem is oversampling, a form of data augmentation in which data from less frequent classes are repeated as needed so that the data used for training is more evenly distributed across classes. Oversample only for the first five objects, then the oversampled data are used to train the final network.

Long Short Time Memory Framework

The network model employed in the paper is depicted in Figure 1. It is based on a commonly used architecture for time-based sensor data and is composed of a combination of fully connected layers and LSTM units. The input data consisted of three acceleration axes and PPG, forming a four-dimensional time sequence. The data are then fed into the network in a window of $w \times 4$, with the parameter w being the size of the time point of a single data window.

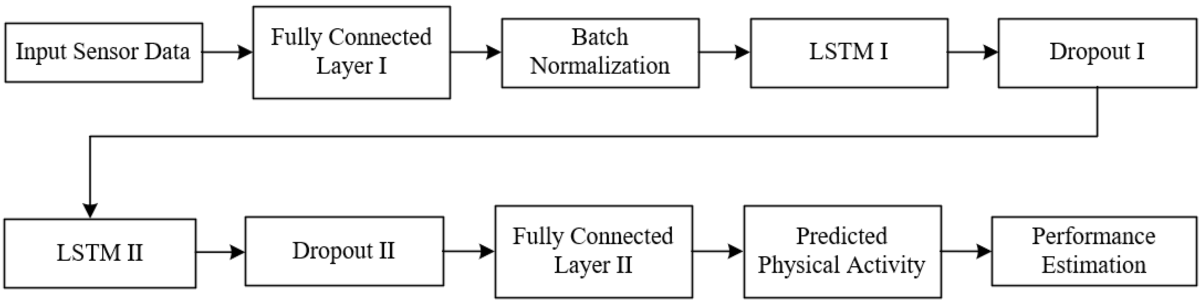

Figure 1. The network model for training dataset

The first layer is fully connected layer and aims to identify the relevant features in the input data. In this layer, the general neuron produces an output value y:

$$y = f\left(\left[w_1, w_2, \ldots, w_n\right]\left[x_1, x_2, \ldots, x_n\right]^T + b\right), \tag{13}$$

where the x_n inputs to the layer and the w_n neuron weights in association with each input, f is the activation function and b is the bias value.

The batch normalization layer, which normalizes the mean and standard deviation of the global data, operates on individual batches of data with training. Then, the recurrent neural network is represented at its core by two cascaded LSTM layers, with the LSTM followed by a dropout layer that randomly discards some of the inputs to reduce overfitting.

In the end, there is a fully connected layer of size 3 which, together with the sparse class cross-entropy loss function assigned to the network, classifies one of these three classes of layers. The loss function represents the error that must be minimized by the training process. The representation of the error varies upon the given function of the network allotted to it. For a categorical cross-entropy function $J(w)$, the error function is as follows:

$$J\left(w\right)=-\frac{1}{N}\sum_{i=1}^{N}[y_i\log y_i+(1-y_i)\log(1-y_i)]\,, \tag{14}$$

where w is the set of model parameters, N is the number of input test features, y_i and y_i are the true and predicted classes respectively, expressed numerically.

Table 1 shows the details of the individual layers. The RNN, as built in this configuration, has 25,283 trainable parameters.

Table 1. **The parameters of LSTM neural network based on ACC and PGG**

Layer	Input Size	Output Size	Parameters
Fully connected layer I	[w, 4]	[w, 32]	128
Batch Normalization	[w, 32]	[w, 32]	128
LSTM I	[w, 32]	[w, 32]	8320
Dropout I	[w, 32]	[w, 32]	0
LSTM II	[w, 32]	[1, 32]	8320
Dropout II	[1, 32]	[1, 32]	0
Fully connected layer II	[1, 32]	[1, 3]	99

Result and Analysis

It is demonstrated the matrix of confusion that arises when classifying the test data in the same setup in the Figure 2. It is evident that the squat and step activities are the activities with greater error rates, while the rest activity is correctly identified in 99 % of the cases. This may be partly due to the much smaller amount of raw input data for the squat and step activities. Accuracy is the ratio of the sum of true positives (TP) and true negatives (TN) to the total number of records (Num). Figure 3 shows the progress of accuracy (estimated on the training material itself) and loss with respect to the training epochs for the network with no downsampling (original data at 400 Hz). About 100 epochs, the values reach convergence. The accuracy is the evaluation ratio metric to all true assessment results of summarize the total grouping achievement for resting, squatting and stepping activities:

$$\text{Accuracy}=\frac{\text{TP}_{\text{resting}}+\text{TP}_{\text{squatting}}+\text{TP}_{\text{stepping}}+\text{TN}_{\text{resting}}+\text{TN}_{\text{squatting}}+\text{TN}_{\text{stepping}}}{\text{Num}}\,. \tag{15}$$

In the current setup, a maximum accuracy of 95,36 % was achieved in the test phase for the decimation factor 40. Although dividing the dataset into five training subjects and two test subjects is a natural choice, the limited size of the dataset can lead to biased results depending on the partition chosen.

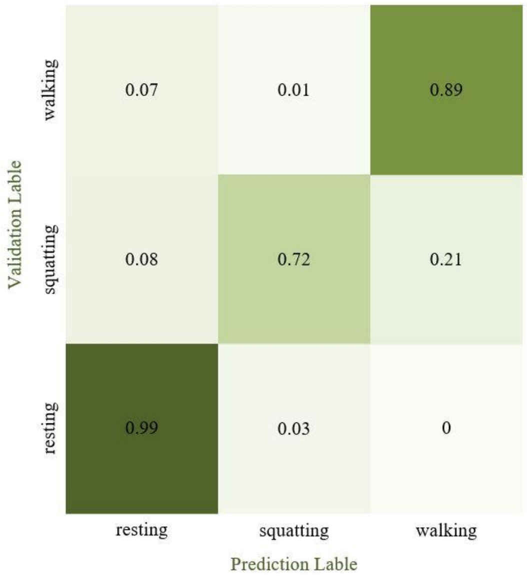

Figure 2. The matrix of confusion for classification

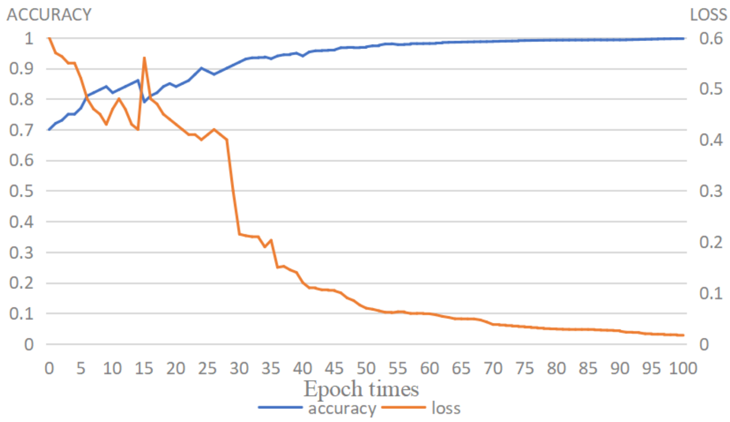

Figure 3. Accuracy and loss progress with respect to training epochs

Conclusion

For recognition, human activity a network model based on LSTM is proposed in the paper. The virgin PPG signal has been severely corrupted by MAs, mainly due to the relative motion between the PPG source and the wrist skin. In order to reduce MAs, the ACC data and PPG were integrated into four-dimensional data, which were processed and analyzed. In an investigation of Python based data analysis of PPG and ACC signals, the LSTM was used for recognizing physical activity. The results revealed that 95,36 % accuracy in classification of the test data was achieved.

References

1. Allen J. // Physiological measurement. 2007. P. 28.
2. Hwang S., Seo J., Jebelli H., [et. al] // Automation in construction. 2016. P. 372–381.
3. Karimian N., Guo Z., Tehranipoor M., [et. al] // Speech and Signal Processing (ICASSP). 2017. P. 4636–4640.
4. Rundo F., Conoci S., Ortis A., [et. al] // Sensors. 2018. P. 405.
5. Abdel-Nasser M., Mahmoud K. // Neural Computing and Applications. 2019. P. 2727–2740.
6. Biagetti G., Crippa P., Falaschetti L., [et. al] // Biomedical Signal Processing and Control. 2019. P. 293–301.
7. Troiano R P., McClain J J., Brychta R J., [et. al] // British journal of sports medicine. 2014. P. 1019–1023.
8. Figo D., Diniz P C., Ferreira D R., [et. al] // Personal and Ubiquitous Computing. 2010. P. 645–662.
9. Biagetti G., Crippa P., Falaschetti L., [et. al] // Data in brief. 2020. P. 105044.