UDC 004.62

HUMAN PHYSICAL ACTIVITY RECOGNITION ALGORITHM BASED ON SMARTPHONE SENSOR DATA AND CONVOLUTIONAL NEURAL NETWORK

Wan Z.W., gr.167011, master

Belarusian State University of Informatics and Radioelectronics[1], Minsk, Republic of Belarus

Baryskievic A.A. – PhD in Technical Sciences

Annotation. Human activity recognition (HAR) is a prominent application of advanced Machine Learning (ML) and Artificial Intelligence (AI) techniques that utilizes computer vision to understand the semantic meanings of heterogeneous human actions. This paper describes a supervised learning method that can distinguish human actions based on data collected from practical human movements. This study proposes a HAR classification model based on a Convolutional Neural Network (CNN) and uses the collected human action signals. The model was tested on the WISDM dataset, which resulted in a 92 % classification accuracy. This approach will help to conduct further researches on the recognition of human activities based on their biomedical signals.

Keywords. Human activity recognition, machine learning, convolutional neural network

There are several approaches to collect HAR data from the participating subjects; broadly, they fall into one of the two categories – namely camera-based recording or sensor-based recording [1]. In the former approach, one or more video cameras are set up to record the activities of a subject for a certain amount of time, and then the recognition is performed using video analysis and processing techniques. The later one utilizes various types of sensors to track the movements of the subject. This approach can be further classified based on the type of sensors used, whether they involve wearable body sensors or the external ones [2]. External sensors are placed in predetermined points of interest on the subjects' body, whereas wearable sensors require to be attached to the subject while collecting data. Each of these techniques has its advantages, shortcomings, and apposite applications. Some recognition techniques even combine multiple recording techniques to collect more relevant data and make the corresponding actions more interpretable to the machines. The applications of HAR include intelligent surveillance, haptics, human-computer interaction, motion or gesture-controlled devices, automatic health-care monitoring systems, prosthetics, and robotics. Despite many advancements, HAR is still a challenging task because of the articulated nature of human activities, the involvement of external objects in human interactions, and complicated spatiotemporal structures of the action signals [3]. Success in recognizing these activities requires advanced signal and image processing techniques, as well as sophisticated ML algorithms. Since the absolute performance is yet to be achieved, HAR remains a tending field to the researchers.

The WISDM dataset contains mobility information that was collected from 30 people of different ages (ranging from 19 to 48 years), genders, heights and weights using a wrist-mounted smartphone. The smartphone has integrated accelerometer and gyroscope. Action data was recorded using these sensors while each of the subjects was performing six predefined tasks, which according to the jargon of ML, represent six different classes. Three-axial linear acceleration and three-axial angular velocity data were acquired at a steady rate of 20 Hz. The collected samples were labeled manually afterward. Before putting in the dataset, the samples were pre-processed using a median filter for noise cancellation and a thirdorder low-pass Butterworth filter having a 20 Hz cutoff frequency.

This study aims to classify the HAR signals of the WISDM dataset employing a CNN model, as shown in Figure 1. The training stage requires a set of data samples containing various attributes measured from subjects while performing various predefined activities. The supervised learning technique then try to make some "sense" out of the data, find out how the samples that belong to the same class are similar to each other while samples from different classes are diverse, then builds one or more internal models focusing on the crucial attributes that can highlight those contrasting properties to carry out the classification [1]. In the training stage, a preordained portion of the dataset is used to train the machine and build a feasible model, which is then evaluated over the remaining samples.

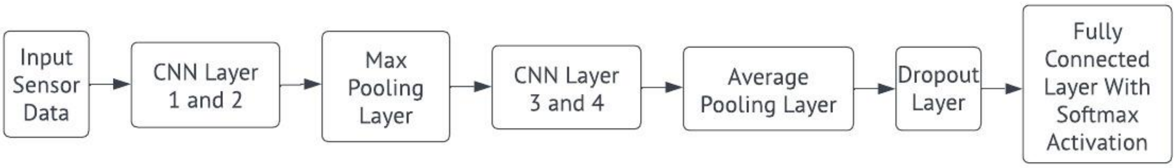

Figure 1. Block diagram of the proposed CNN-based HAR algorithm

The data has been preprocessed in such a way that each data record contains 80 time slices (data was recorded at 20 Hz sampling rate, therefore each time interval covers four seconds of accelerometer reading). Within each time interval, the three accelerometer values for the x axis, y axis and z axis are stored. This results in an 80×3 matrix. The data must be passed into the neural network as a flat vector of length 240. The first layer in the network must reshape it to the original shape which was 80×3. The model is tested on the human behavior pose dataset WISDM. We selected 80 % of the data in the WISDM dataset for model training and 20 % for model testing. Using Adam as optimizer, batch size equals to 400, epochs equal to 50 (training for 50 rounds).

Accuracy is a measure for how many correct predictions your model made for the complete test dataset. It is measured by the following formula:

$$\text{Accuracy} = (TP + TN)/(TP + TN + FP + FN), \tag{1}$$

where TP – True Positives, FP – False Positives, FN – False Negatives, and TN – True Negatives.

The model learns well with accuracy reaching above 92 % and loss hovering at around 0,39.

List of used sources:

1. Human activity recognition: using wearable sensors and smartphones. / M. Labrador, Y.O. Lara // CRC Press, 2013.
2. Human activity recognition and prediction. / Fu Y // Springer, 2016.
3. Human Action Recognition with Depth Cameras. / J. Wang [et al.] // Springer International Publishing, 2014.