Ministry of Education of the Republic of Belarus
Educational institution
Belarusian State University of
Informatics and Radioelectronics

UDC 004.896

WAN
Ziwei

HUMAN ACTIVITY RECOGNITION SYSTEM BASED ON ACCELEROMETER SMARTPHONE DATA AND CONVOLUTIONAL NEURAL NETWORK

Abstract
for a Master's Degree
in the Specialty 1-45 80 01 Infocommunication Systems and Networks

Supervisor
Doctor of technical sciences,
Professor
Boriskevich Anatoliy Antonovich

Minsk 2023

INTRODUCTION

Human physical activity recognition (HAR) is the problem of predicting what a person is doing based on a trace of their movement using sensors. In the view of the researches of the recent years, there has been a burgeoning development in the area of sensors for smartphones, smartwatches etc. having myriad of applications. Their key features are high computational power, inexpensive and portable size; making them feasible enough for the humans to interact with them and include them in daily routine. These devices have a wide range of applications, ranging from medical to security to military purposes. They are very useful in detecting the type of human activities. Some instances are the times when one needs to monitor workout routine, or maybe when a patient needs to record his different daily activities suggested by physician. Moreover, by keeping track of these activities one can apply different statistical or Machine Learning (ML) algorithms to personalize diets and exercises which cater best to their lifestyle.

Additionally, a lot of challenges are associated with this. There have been issues in improving the accuracy in real-world, obtaining clean data during data collection, deciding which attributes to consider and also how much importance to be given to each, providing the flexibility where new users are added without the need to re-train the whole system, building the design model for feature engineering and analysis; lastly, we also face problems in implementing these in mobile phone devices where energy and processing requirements are the key constraints.

Mainly, two kinds of sensors are available for human activity recognition purposes. First, external sensors, these are fixed in predetermined points of interest, so the inference of activities entirely depends on the voluntary interaction of users with the sensors. Intelligent homes and external cameras are the classic examples of external sensing. Second, wearable sensors, which are worn by the user attached to the user, for instance, smartphones and smartwatches. These wearable sensors record many signals such as accelerometer, gyroscope, magnetometer and even light intensity. Combinations of these sensor data allow us to determine which activity is being performed at any given time. The sensors our research used are wearable sensors.

We selected the data set from the Wireless Sensor Data Mining (WISDM) project, which collected 1098207 experimental data generated from 29 volunteers carrying smartphones to perform specified actions every 50 ms, and each piece of data consists of 6 parts: username, specified action, timestamp and accelerometer values for x, y and z axis. We compared the advantages and disadvantages of different low-pass filters, and finally adopted a third-order Butterworth filter with a cutoff frequency of 4 to filter the noise in the original data. We use a window of size 200 with an overlap of 50 % to divide the x, y and z axis accelerometer and label part in the original data, store them as acceleration data and label data respectively for preprocessing. We get 54901 windows and split both data into a training set (80 %) and a test set (20 %).

Our research aims to find out a precise human physical activity recognition algorithm based on accelerometer smartphone data and Convolutional Neural Network (CNN). to recognize the type of movements (Walking, Running, Jogging, etc.) based on a given set of accelerometer data from a mobile device carried around a person's waist.

The aim of the work is to increase recognition accuracy of different human activities using the acceleration sensor data from the smartphone and deep learning.

To achieve this aim, the following tasks were solved in the master thesis:

– Data collection on different human activities by using smartphones;
– Pre-processing algorithm of the sensor data;
– The algorithm of physical activity recognition based on CNN;
– Evaluation of algorithm performance using confusion matrix and accuracy.

GENERAL DESCRIPTION OF WORK

Relevance of the subject

The work corresponds to paragraph 1 «Digital information and communication and interdisciplinary technologies, production based on them» of the State Program of innovative development of the Republic of Belarus for 2021–2025.

The work was carried out in the educational institution Belarusian State University of Informatics and Radioelectronics within the framework of research work 21-2033 "Processing, coding and transmission of information in network-centric systems".

The aim and tasks of the work

The aim of the work is to increase recognition accuracy of different human activities using the acceleration sensor data from the smartphone and deep learning.

To achieve this aim, the following tasks were solved in the master thesis:

1 Data collection on different human activities by using smartphones.

2 Pre-processing algorithm of the sensor data.

3 The algorithm of physical activity recognition based on CNN.

4 Evaluation of algorithm performance using confusion matrix and accuracy.

Personal contribution of the author

The content of the dissertation reflects the personal contribution of the author.

1 Preprocessing and temporal feature extraction of acceleration sensor data in public datasets.

2 Construction and implementation of algorithm structure.

3 Evaluate the results of the proposed CNN algorithm.

Task setting and discussion of the results were carried out together with the supervisor *Doctor of Technical Sciences, Professor, Boriskevich Anatoly Antonovich.*

Testing and implementation of results

The main provisions and results of the dissertation work were reported and discussed at: 59th scientific conference of postgraduates, undergraduates and students, (Minsk, April 18–22, 2023) and International scientific and technical seminar "Technologies of information transmission and processing" (Minsk, March – April 2023).

Author’s publications

According to the results of the research presented in the dissertation, 2 author’s works was published, including: 2 articles in scientific journals recommended by the Higher Attestation Commission, with a total amount of 9 author’s pages.

Structure and size of the work

The dissertation work consists of introduction, general description of the work, three chapters with conclusions for each chapter, conclusion, bibliography, eight appendixes.

The total amount of the thesis is 65 pages, of which 42 pages of text, 14 figures on 12 pages, 12 tables on 9 pages, a list of used bibliographic sources (22 titles on 2 pages), a list of the author's publications on the subject of the thesis (2 titles on 1 pages), 1 appendix on 11 pages, graphic material on 8 pages.

Plagiarism

An examination of the dissertation « Human activity recognition system based on accelerometer smartphone data and Convolutional Neural Network » by Wan Ziwei was carried out for the correctness of the use of borrowed materials using the network resource «Antiplagiat» (access address: https://antiplagiat.ru) in the on-line mode 14.06.2023. As a result of the verification, the correctness of the use of borrowed materials was established (the originality of the thesis is 90 %)

SUMMARY OF WORK

The **introduction** addresses the problems of physical human activities recognition.

The **general description of work** shows the connection between the work and the priority areas of scientific research, the aim and tasks of the research, the personal contribution of the applicant for a scientific degree, the approbation of the dissertation results.

In the first chapter, we studied the research status of HAR for physical human activities recognition in different domains.

We selected the dataset from the WISDM lab, which data is obtained by volunteers putting the Android phone in the front trouser pocket to complete the 6 specified activities including walking, jogging, upstairs, downstairs, sitting, standing. Each sample from the WISDM dataset includes user number, the type of activities completed by the user, and the time when the sample was collected (in nanoseconds), and acceleration values on the x, y, and z-axis (see Figure 1). Accelerometer data is collected every 50 ms, and is stored as raw data by specially developed software. It means that when a volunteer completes a given action, we will get 20 samples per second.

user-id	activity	timestamp	x-axis	y-axis	z-axis
33	Jogging	49105962326000	-0.7	12.7	0.5
33	Jogging	49106062271000	5.0	11.3	1.0
33	Jogging	49106112167000	4.9	10.9	-0.1

Figure 1 – Description of WISDM Raw Data

This is an unbalanced dataset where walking contains 38.6 % (424400 samples) of the dataset, Jogging contains 31.2 % (342177 samples) of the dataset, Upstairs contains 11.2 % (122869 samples) of the dataset and Downstairs contains 9.1 % (100427 samples) of the dataset, Sitting contains 5.5 % (59939 samples) of the dataset, Standing contains 4.4 % (48395 samples) of the dataset.

We analyzed several low-pass filters, and finally chose the Butterworth low-pass filter for preprocessing to reduce the impact of noise. After testing different orders and cut-off frequencies, the preprocessing effect of the third-order Butterworth low-pass filter with a cut-off frequency of 4 is the best.

We also used sliding window algorithm to improve feature extraction from filtered data, and one-hot encoding to digitize features for action labels. We choose an overlap rate of 90 %, and get 54901 windows with a data type of 32 bits. Taking the acceleration data with the label "standing" as an example, its sequence segmentation diagram is shown in Figure 3. In addition, we performed the hyperparameter settings of the neural network model, including learning rate, batch size and Adam optimizer. Finally, we split the data into training (80 %) and test (20 %) set.

We analyzed the WISDM dataset as well as data preprocessing methods. We used a Butterworth low-pass filter for data preprocessing, a sliding window technique to improve feature extraction from filtered data, and one-hot encoding to digitize features for action labels.

We performed the hyperparameter settings of the neural network model, including learning rate, batch size and Adam optimizer.

In the second chapter, we analyzed the structures of CNN, LSTM, and CNN-LSTM respectively, compared their advantages and disadvantages, and introduced the proposed CNN structure.

We proposed a human activity recognition neural network model with CNN(see Figure 2), and set the learning parameters and hyperparameters of each layer in the model separately, and set the learning parameters and hyperparameters of each layer in the model separately (see Table 1 and Table 2). We partition the training set consist of 3-axis acceleration data (x, y and z) with a batch size of 400, resulting in 50 iterations as the input. Softmax layer converts the input from the previous layer into probability set of 6 physical human activities as the output.

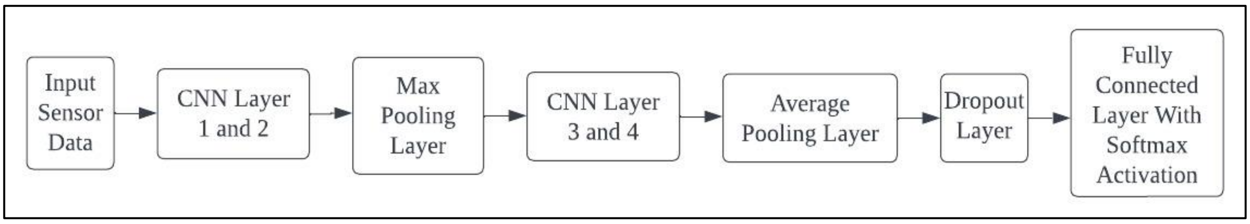

Figure 2 – Block diagram of the proposed CNN-based HAR algorithm

Table 1 – The learning parameters of CNN based on mobile phone accelerometer

Moduels	Learning parameters	matrix size
CNN Layer 1 and 2	weight	[3,64]
	bias	[64]
Max Pooling Layer	/	/
CNN Layer 3 and 4	weight	[64,6]
	bias	[6]

Table 2 – The hyperparameter of CNN based on mobile phone accelerometer

Moduels	Hyperparameter	Value
CNN Layer 1 and 2	hidden layer	1
Max Polling Layer	hidden unit size	64
CNN Layer 3 and 4	/	/
Training	optimizer	Adam
	batch size	400
	learning rate	0.0025
	number of epochs	50

The algorithm transforms the input 3 features into 64, so as to better divide different types of data. The softmax layer converts the model prediction results into the probability set of each action category;

In the third chapter, we performed the performance estimation indicators of several recognition algorithms used in this network model using confusion matrix, loss and accuracy. The confusion matrix of this model is shown in Figure 3. From the confusion matrix, it can be seen that the model predicts well on most actions, and it is worth noting that there are some misclassifications on up and down stairs.

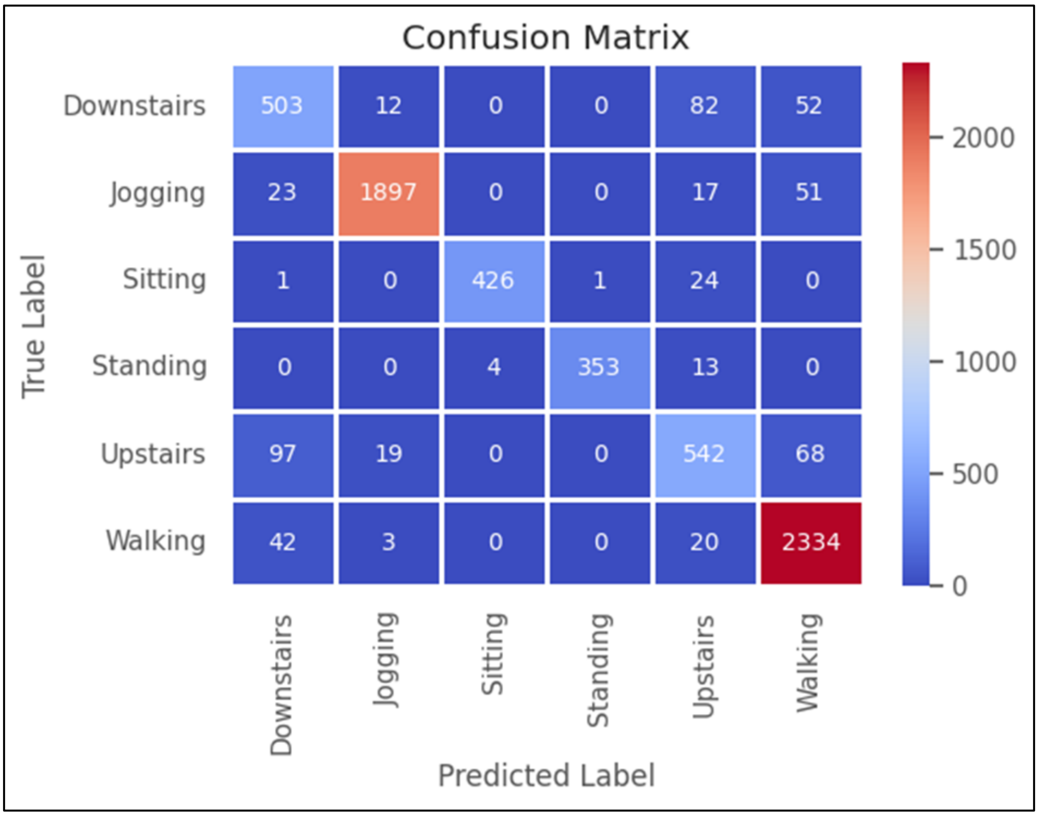

Figure 3 – The confusion matrix of the CNN model

We separately evaluated and tested the dataset in CNN model, LSTM model, and CNN-LSTM model. The accuracy and loss of the model are compared, and the results are shown in the Table 3.

Table 3 – Accuracy and loss of different models

Methods	Accuracy	Loss
CNN	0.92	0.39
LSTM	0.88	0.44
CNN-LSTM	0.88	0.35

It was set that the network model composed of CNN, which achieved an accuracy rate of 92 % and a loss of 0.39 had the best HAR effect.

CONCLUSION

The modeling result of this paper is based on the WISDM dataset, which data is obtained by volunteers putting the Android phone in the front trouser pocket to complete the 6 specified actions (Walking, Jogging, Upstairs, Downstairs, Sitting, Standing). In this paper, we mainly completed the following work:

1 In chapter 1 we studied the research status in the field of human action recognition HAR using computer vision, wearable sensor devices, smartphone-based methods. We analyzed the WISDM dataset, as well as data preprocessing methods. In terms of data preprocessing, we used a 3rd-order 4 Hz Butterworth low-pass filter for filtering to remove high-frequency noise. In terms of feature generation, the sliding window technique is used to improve extraction of features from the filtered data. We digitize the features of action labels using one-hot encoding.

2 In chapter 2 we analyzed the structures of CNN, compared their advantages and disadvantages, and proposed the CNN structure. We proposed a human physical activity recognition convolutional neural network model. The human physical activity recognition algorithm includes CNN layer 1 and 2, Max Pooling layer, CNN layers 3 and 4, Average Pooling layer, Dropout layer, and Full Connected layer with Softmax function, the Softmax layer computes the probability set of action category. We set the learning parameters and hyperparameters of the proposed CNN model, including the following parameters:

– Learning parameters: CNN Layer 1 and 2 weight matrix [3,64], bias [64]; CNN Layer 3 and 4 weight matrix [64,6], bias [6];

– Hyperparameters: The number of hidden units in CNN 1 and 2 layer is 64, the activation function is ReLU; the number of hidden units in Max Pooling Layer is 64; the number of hidden units in CNN layer 3 and 4 is 64.

We performed the hyperparameter settings of the neural network model, including learning rate 0.0001, batch size 400, and Adam optimizer;

3 In chapter 3 we proposed a human activity recognition neural network model with CNN layers. The algorithm transforms the input 3 features into 64, so as to better divide different types of data, the Softmax layer converts the model prediction results into the probability set of each action category. We set the learning parameters and hyperparameters of CNN. We separately evaluated and tested the WISDM dataset in CNN, LSTM, CNN-LSTM model. It was set that the proposed CNN model which achieved an accuracy rate of 92 % and a loss of 0.39 had the best HAR effect.

LIST OF AUTHOR'S PUBLICATIONS

1–Wan, Z. W. Human physical activity recognition algorithm based on smartphone sensor data and convolutional neural network / Z. W. Wan // Инфокоммуникации : сборник тезисов докладов 59-ой научной конференции аспирантов, магистрантов и студентов, Минск, 18–22 апреля 2023 г. – Minsk : BSUIR, 2023. – С. 182–187.

2–Wan, Z. W. Human physical activity recognition algorithm based on smartphone sensor data and convolutional neural network / Z. W. Wan // Технологии передачи и обработки информации (Technologies of information transmission and processing): материалы Международного научно-технического семинара, Минск, март – апрель 2023 г. – Minsk : BSUIR, 2023. – С. 72–77.