UDC 004.512-026.26

HUMAN PHYSICAL ACTIVITY RECOGNITION ALGORITHM BASED ON SMARTPHONE DATA AND LONG SHORT TIME MEMORY NEURAL NETWORK

Z.Y. Chen
Master student of the faculty of Information Security of the Belarusian State University of Informatics and Radioelectronics
chenzheying1@gmail.com

Z.X. Yang
Master student of the faculty of Information Security of the Belarusian State University of Informatics and Radioelectronics

H. Li
PhD student of the faculty of Information Security of the Belarusian State University of Informatics and Radioelectronics

Z.Y. Chen

Master student of the Department of Information Security of the Belarusian State University of Informatics and Radioelectronics, graduated from Nanchang Hangkong University. Her research interests include neural networks, object classification based on sensor data, and object detection based on images and video streams.

Z.X. Yang

Master student of the Department of Information Security of the Belarusian State University of Informatics and Radioelectronics, graduated from Nanchang Hangkong University. His research interests neural networks and signal processing and signal processing for sensors.

H. Li

Received the B.S. and M.S. degrees in computer science from the Henan University. He is currently pursuing the Ph.D degree with the Belarusian State University of Informatics and Radioelectronics. His research interests include Neural Network, mobile crowd-sensing networks and signal processing.

Abstract. The continuous advancement of smartphone sensors has brought more opportunities for the universal application of human motion recognition technology. Based on the data of the mobile phone's three-axis acceleration sensor, using combining a double-layer Long Short Time Memory (LSTM) and full connected layers allow us to improve human actions recognition accuracy, including walking, jogging, sitting, standing, and going up and down stairs. This is helpful for smart assistive technology. It is shown that physical activity classification accuracy is equal to 97 %.

Keywords: mobile acceleration sensor, long short time memory, action recognition and classification accuracy.

Introduction.

Traditional sensor devices are bulky and expensive. With the continuous development of smart phones in recent years, the acceleration sensors of mobile phones have also continued to improve. It has the obvious advantages of small size, high penetration rate and lower price, which provides a new idea for the application of intelligent assistance technology and so on. Recognition of human activities from sensor data is at the core of intelligent assistive technologies, such as smart home, rehabilitation, health support, skills assessment or industrial environments [1]. For example, the project of Inooka et al. predicts the energy consumption of users by recognizing their activities [2], and Mathie et al. judges whether users are safe or not by recognizing their actions [3]. This work is motivated by two requirements of activity recognition: improving recognition accuracy and reducing reliance on engineered features to address increasingly complex recognition problems.

Human Activity Recognition (HAR) is based on the assumption that specific body movements translate into characteristic sensor signal patterns, which can be sensed and classified using machine learning techniques. We use data collected from accelerometer sensors. Almost every modern smartphone has a three-axis accelerometer that measures acceleration in all three spatial dimensions.

We selected the data set from the Wireless Sensor Data Mining (WISDM) project, which collected 1,098,207 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 use a window of size 200 with an overlap of 90 % 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 %).

We trained a double layer LSTM neural network (implemented in TensorFlow) for HAR from accelerometer data with the purpose of providing an algorithm with higher recognition accuracy. The trained model will be exported/saved and added to the Android app. The network model consists of double layer LSTM network layers and double fully connected layer (FCL), and predicts the corresponding human actions from the x, y and z axis acceleration count values from the data set. The proposed algorithm achieved 97 % accuracy and a loss of 0,2 on the test set.

Double FCL and double layer LSTM neural network architecture based on mobile phone accelerometer.

We use combining a double LSTM layer and double FCL to establish a deep learning model, aiming to predict the user's action type at a certain moment through the three-axis acceleration data from the mobile phone sensor. The proposed neural network model is shown in Figure 1. We partition the training set consist of 3-axis acceleration data (x, y and z) with a batch size of 1024, resulting in 50 iterations, each of iteration takes a tensor of shape (1024, 200, 3) as the input. We first flatten it to 204800 tensors (1,3) as input to the FCL1. FCL1 abstracts it into 204800 tensors (1,64), then we stack it into 200 tensors (1024,64) as input to the double layer LSTM and get the features of 1024 samples over 200 of time steps with one tensor (1024,64). We flatten it again to 1024 tensors (1,64) as input to the FCL2 to divided it into 6 physical human activities. Softmax layer converts the input from the previous layer into probability set of 6 physical human activities as the output. Finally, we estimate the performance of the recognition results.

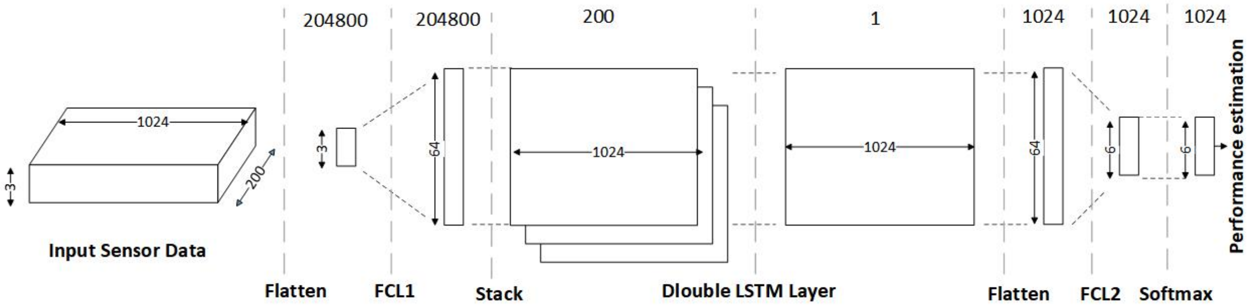

Figure 1. Block diagram of human activity recognition algorithm based on smartphone data and deep learning

Suppose a volunteer in the data set performs *N* predicted physical human activities within a certain period of time *t*, see (1):

$$A = \{a_1, ..., a_n, \dots, a_N\} \tag{1}$$

where *a* is the specified action type, a_n is the *n*th physical human activity.

The acceleration sensor from the mobile phone collects the three-axis acceleration data D corresponding to N predicted physical human activities that the user wishes to complete within the time period t, see (2):

$$D=\{d_1,...,d_n,\ldots,d_N\} \tag{2}$$

where $d_n=(x_n,y_n,z_n), x_n=(x_1,...,x_k), y_n=(y_1,...,y_k), z_n=(z_1,...,z_k)$ is the k three-axis acceleration data of nth physical human activity.

This algorithm aims to identify n actions within the time period t based on the three-axis acceleration data D by constructing a neural network model F, see (3):

$$F(D)=\{a'_1,...,a'_n,...,a'_N\}, a'_n\in A \tag{3}$$

where a' is the predicted action type, a'_n is the nth predicted physical activity.

The learning parameters and hyperparameter of the double layer LSTM neural network based on mobile phone accelerometer are shown in Table 1 and Table 2.

Table 1. The learning parameters of double FCL and double layer LSTM neural network based on mobile phone accelerometer

Modules	Learning parameters	matrix size
Fully Connected Layer 1	Weight configuration	[3,64]
	bias	[64]
Double LSTM Layer	Number of params	66048
Fully Connected Layer 2	weight	[64,6]
	bias	[6]

Table 2. The hyperparameter of double FCL and double layer LSTM neural network based on mobile phone accelerometer

Modules	Hyperparameter	Value
Fully Connected Layer 1	hidden unit size	64
	activation function	ReLU
Stacked LSTM Layer	hidden unit size	64
Fully Connected Layer 2	hidden unit size	6
Training	optimizer	Adam
	batch size	1024
	learning rate	0.0025
	number of epochs	50

Fully connected layer 1.

For fully connected layer 1, we have input 204800 tensors of shape (1,3), all of data type float32. For input 204800 samples, each sample has a length of 3, representing the acceleration of the x, y and z axis respectively. FCL1 contains 64 neurons and abstracts the 3 features of the input to 64, so as to better divide different types of data. The structure and parameters are shown in Table 3.

For the input, we multiply a weight matrix of size [3,64] and add a bias matrix of size [64]. The output Y_j of the jth neuron of FCL1 is shown in the formula (4), we use ReLU (Rectified Linear Activation Function) as the activation function f to get an output with shape [1,64].

$$Y_\mathrm{j}=f\left(\sum_{i=1}^{n}x_iw_{ij}+w_{j0}\right) \tag{4}$$

where x_i is the *i*th input variable, w_{ij} is the weight between the *i*th input variable and the *j*th FCL1 neuron, and w_{j0} is the bias of the *j*th FCL1 neuron, *n* is the number of input dimensions 3.

Table 3. The input, condition and output of fully connected layer 1

FCL1	Value
Input	tensor (1,3)
	dtype: float32
Condition	Weight matrix size [3,64]
	Bias matrix [64]
	activation function: ReLU
Output	tensor (1,64)
	dtype: float32

The ReLU is a simple calculation that directly returns the value provided as input if the input is greater than 0, or returns a value of 0 if the input is 0 or less. Due to the sparsity of ReLU, the sparse model can better mine relevant features and fit the training data.

Next, we split the 204800 outputs according to the size of the time steps 200 into a list of two order tensors of the shape (1024,64).

Double layer LSTM.

We use double layer LSTM to form the LSTM layer, the number of hidden units is 64, and the data type is float32. For an input list of 200 tensors of shape (1024,64), we get the output tensor of shape (1024,64), that is, the features of 1024 samples over 200 of time steps are extracted to 64 dimensions. The parameters are shown in Table 4.

Table 4. The input, condition and output of double layer LSTM

Double LSTMs	Value
Input	List of 200 two order tensors (1024,64)
	dtype: float32
Condition	LSTM units: 64
Output	two order tensor (1024,64)
	dtype: float32

Double layer of LSTMs makes the model deeper and the extracted features deeper, resulting in more accurate predictions. It has achieved good results on a wide range of prediction problems.

Double layer LSTM is to take the output of the previous layer of LSTM as the input of the next layer of LSTM and send it into the network. The network structure is shown in Figure 2. The triaxial accelerometer value x_t at time *t* is used as the input of the first layer of LSTM cells, and the output h_t^1 of its hidden layer is simultaneously used as the input of the second layer of LSTM cells and the update of the hidden layer of the first layer of LSTM cells, and update the cell state c_t^1 of the first layer of LSTM neurons. The second layer of LSTM neurons outputs the hidden layer h_t^2 as a result, and updates the hidden layer h_t^2 and cell state c_t^2 at time *t*.

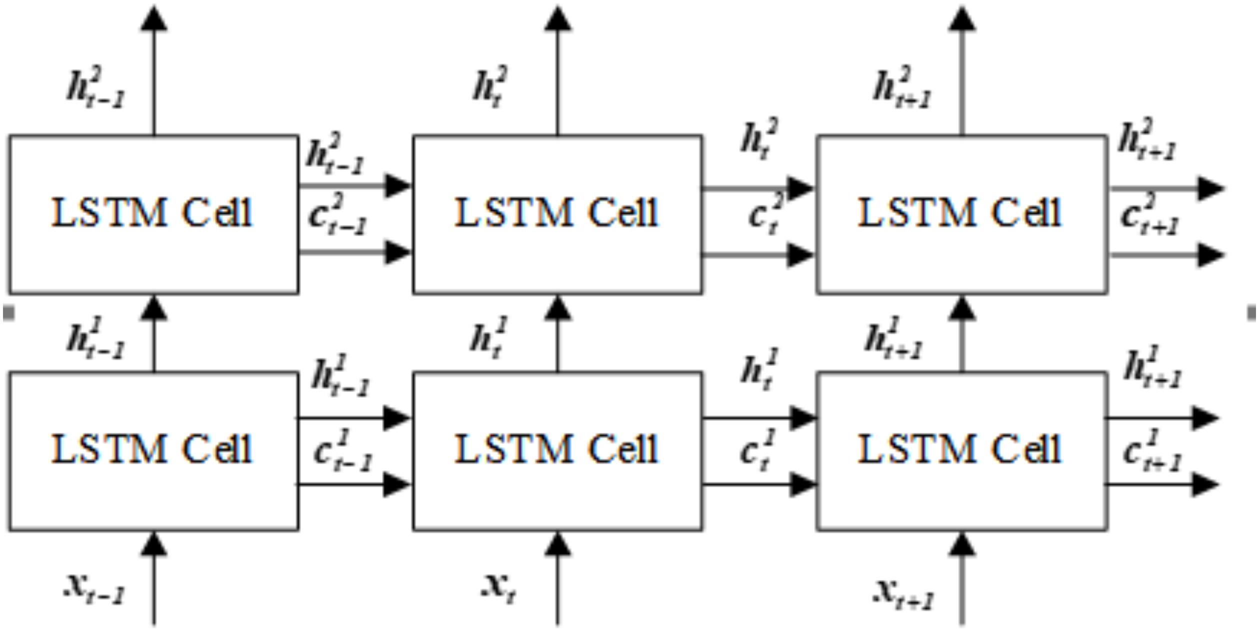

Figure 2. The structure of double LSTM layer

Fully connected layer 2.

For the fully connected layer 2, we flatten the two-dimensional tensor (1024, 64) from the double layer of LSTMs to one-dimensional to get 1024 tensors of shape (1, 64) as input.

For the input, we multiply a weight matrix of size [64,6] and add a bias of size [6], FCL2 has 6 neurons, which fuse the features from the previous layer into 6 dimensions, the output O_k to the *k*th neuron in FCL2 see formula (5). The parameters are shown in Table 5.

Table 5. The input, condition and output of fully connected layer 2

FCL2	Value
Input	tensor (1,64)
	dtype: float32
Condition	Weight matrix size [64,6]
	Bias [6]
Output	tensor (1,6)
	dtype: float32

$$O_k = \sum_{r=1}^{m} x_r w_{rk} + w_{k0} \qquad (5)$$

where x_r is the *r*th input variable, w_{rk} is the weight between the *r*th input variable and the *k*th FCL2 neuron, $w_{\text{k}0}$ is the bias of the *k*th FCL2 neuron, *m* is the input dimension 64.

Softmax layer.

Through the output O_k of the FCL2, we get 1024 tensor samples of shape (1,6), which respectively represent the prediction results of each 6 kinds of physical human activities (Walking, Jogging, Upstairs, Downstairs, Sitting, Standing). Next, we use the Softmax function (see formula 6) to convert each prediction result into a number between 0 and 1, indicating the probability that the predicted result of the sample is the *k*th class action.

$$P(k \mid X) = \frac{e^{Z_k}}{\sum_{c=1}^{6} e^{Z_c}} \qquad (6)$$

where *X* is the input of double FCL and double layer LSTM neural network, Z_k is the *k*th value of input tensor of shape (1,6), Z_c is the *c*th value of input tensor of shape (1,6).

Through the Softmax function, the output value of the multi-category can be converted into a probability distribution ranging from 0 to 1. The Softmax function can convert the predicted results into non-negative numbers, and make the sum of the probabilities of various predicted results equal to 1. The input and output are shown in the Table 6.

Table 6. The input and output of Softmax

input	output
Predicted value under each physical human activity: $Z = \{Z_1, Z_2, .., Z_6\}$	The predicted probability of each physical human activity: $P = \{p_1, p_2, .., p_6\}$

The recognition algorithm performance estimation.

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, MSE as model measure, learning rate 0,0025, batch size 1024, training for 50 rounds. We evaluate the model by confusion matrix, training loss and accuracy.

Confusion matrix.

The confusion matrix our work is shown in Table 7. where *A* means activity, *W* means walking, *J* means jogging, *U* means upstairs, *D* means downstairs, *SIT* means sitting, *ST* means standing, *N* is the number of various activities.

Table 7. Confusion Matrix for Multiple Classification Tasks

Activities		Predicted label						total
		Walking	Jogging	Upstairs	Downstairs	Sitting	Standing	
True label	Walking	A_{WW}	A_{JW}	A_{UW}	A_{DW}	A_{SITW}	A_{STW}	N_W
	Jogging	A_{WJ}	A_{JJ}	A_{UJ}	A_{DJ}	A_{SITJ}	A_{STJ}	N_J
	Upstairs	A_{WU}	A_{JU}	A_{UU}	A_{DU}	A_{SITU}	A_{STU}	N_U
	Downstairs	A_{WD}	A_{JD}	A_{UD}	A_{DD}	A_{SITD}	A_{STD}	N_D
	Sitting	A_{WSIT}	A_{JSIT}	A_{USIT}	A_{DSIT}	A_{SITSIT}	A_{STSIT}	N_{SIT}
	Standing	A_{WST}	A_{JST}	A_{UST}	A_{DST}	A_{SITST}	A_{STST}	N_{ST}

In the field of machine learning, confusion matrix, also known as likelihood table or error matrix. It is a specific matrix used to present a visualization of algorithm performance. Each column represents the predicted value, and each row represents the actual category. The name comes from the fact that it makes it easy to indicate if multiple categories are confused.

Loss and accuracy of the proposed neural network model.

The loss function will determine the performance of the model by comparing the predicted output of the model with the expected output, and then find the optimization direction. If the deviation between the two is very large, the loss value will be large; if the deviation is small or nearly the same, the loss value will be very low.

In this research, we use the cross-entropy loss function to calculate the loss. For each input sensor (1,6), we compare it with the label array, calculate the cross-entropy loss for the training data and evaluate the training effect in test set with formula (7). We hot-encode the labels of the training set and the test set. For each label, the encoded data is an array of length 6. In the order of the physical human activities Walking, Jogging, Upstairs, Downstairs, Sitting, Standing, if the label is the *i*th activity, the *i*th value of the array is 1, and the others are 0.

$$L = -\sum_{i=1}^{M} q_i \log(p_i) \tag{7}$$

where M is the number of physical human activities 6, q_i is the *i*th value of the label array, p_i is the *i*th value of the input tensor, which is the predicted probability of *i*th physical human activity from Softmax layer.

Accuracy is one of the most popular metrics in multi-class classification and it is directly computed from the confusion matrix. The formula of the Accuracy considers the sum of True Positive and True Negative elements at the numerator and the sum of all the entries of the confusion matrix at the denominator.

$$A\text{ccuracy} = \frac{A_{WW} + A_{JJ} + \ldots + A_{STST}}{N_W + N_J + \ldots + N_{ST}} \tag{8}$$

The proposed algorithm learns well with recognition accuracy reaching above 97 % and loss hovering at around 0,2 in test set.

Conclusion.

The modeling result 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). The proposed physical activity recognition algorithm is based on combining FCL1, double layer LSTM neural network, FCL2 and Softmax layer to extract spatial and temporal features for improving classification accuracy. The proposed algorithm using 64 relevant features works well on most actions types and allows us to achieve 97 % classification accuracy, but there were 119 activities of downstairs that were wrongly predicted to go upstairs. The algorithm can be used in human activity recognition system using smartphone sensor data for monitoring physical behavior.

References

[1] Anderson, I., Maitland. Shakra. Tracking and sharing daily activity levels with unaugmented mobile phones. / Anderson, I., Maitland.: Mobile Networks and Applications. 2007. –12 p.

[2] Inooka, H., Ohtaki, Y. Development of advanced portable device for daily physical assessment. / Inooka, H., Ohtaki, Y.: SICE-ICASE International Joint Conference. 2006. –5878-5881 p.

[3] Mathie, M., Celler B. Classification of basic daily movements using a triaxial accelerometer / Mathie, M., Celler B.: Medical & Biological Engineering and Computing, 2004. – 42 p.

АЛГОРИТМ РАСПОЗНАВАНИЯ ФИЗИЧЕСКОЙ АКТИВНОСТИ ЧЕЛОВЕКА НА ОСНОВЕ ДАННЫХ СМАРТФОНА И ДЛИТЕЛЬНОЙ КОРОТКОЙ ПАМЯТИ НЕЙРОННОЙ СЕТИ

Ч.Э.Чэнь
магистрант факультета
информационной безопасности
БГУИР

Ц.С. Ян
магистрант факультета
информационной
безопасности БГУИР

Х. Ли
аспирант факультета
информационной безопасности
БГУИР

Кафедра инфокоммуникационных технологий
Факультет информационной безопасности
Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь
Электронная почта: chenzheying1@gmail.com

Аннотация. Постоянное совершенствование датчиков смартфонов открыло больше возможностей для универсального применения технологии распознавания движений человека. Основываясь на данных трехосевого датчика ускорения мобильного телефона, использование сочетания двухслойной долговременной памяти (LSTM) и полносвязных слоев позволяет нам повысить точность распознавания действий человека, включая ходьбу, бег трусцой, сидение, стояние и подниматься и спускаться по лестнице. Это полезно для интеллектуальных вспомогательных технологий. Показано, что точность классификации физической активности составляет 97 %.

Ключевые слова: мобильный датчик ускорения, долгая короткая память, распознавание действий и точность классификации.