Recognition of Human Activities by Smartphone Sensors Using LSTM Neural Network

Hong Zhao^{1,*}, Chunling Hou¹ and Donglin Ma²

¹School of Computer and Communication, Lanzhou University of Technology, Lanzhou, 730050, China

²Information Center, Lanzhou University of Technology, Lanzhou, 730050, China

*Corresponding author

Abstract—Human activities have been a hot research field. Many sensors are embedded in the smartphone, which makes mobile sensor become available. Sensors of smartphone can early get the human activities information, which can analyze the human behaviors and provide the useful information to the human. In this paper, we propose a new method to recognize the human activities, which is based on the LSTM neural network to extract features and classify using accelerometer sensor data and gyroscope sensor data. Experimental results show that using LSTM neural network and TensorFlow deep learning open source architecture to extract motion state characteristics, this method achieves human activities classification with an accuracy of up to 90.4%.

Keywords—deep learning; LSTM neural network; activity recognition; sensors of smartphone; TensorFlow

I. INTRODUCTION

With the development of the mobile phone industry, a lot of sensors are embedded in the smartphone, instead of wearing a special device to human's body, a smartphone or a mobile device can be easily taken in human's dress [1]. And now many people cannot do without a smartphone, which provides the researchers a new field to analyze the human activities or behaviors using smartphone. Recognition of human activity is classification work, which likes other classification task as the same including getting data, preprocessing, extracting features and recognizing, which makes use of the sensors data from mobile devices such as accelerometer sensor and gyroscope sensor, preprocesses the sensors data including noise reduction and data regularity, extracts the activity features using LSTM neural network, and finally, recognizes activities. Due to the data of sensors has time-series attributes, they are strongly bound up with nearby data [2]. When people execute a same activity, the signals should be broadly consistent. Therefore, it is vital to extract features from sensor signals.

Deep learning is in popular now. Especially the big data industry, cloud computing and the progress of the software and hardware, which makes the deep learning rapid development. And deep learning is extraordinary in many fields such as image, video, neural language progress and speech. There are many neural network structure such as auto-Encoder, convolutional neural network and recurrent neural network including LSTM. In this paper, we use LSTM neural network to recognize human activities using the data from smartphone sensors.

II. RELATED WORKS

Many researchers have tried to recognize human activities using different methods. Jiang [3] assembled signal sequences of accelerometers and gyroscopes into a novel activity image, then they made a Fourier transform of the picture, finally they put the pictures into convolutional neural networks. Although this method perfected good, some detail information has been ignored. Kouris [4] compared naïve Bayes, support vector machine, C4.5 and functional trees for activity recognition, and it found that functional trees outperformed other classifiers. Khan [5] used the Wii Remote to identify basic physical activities using decision tree classifier, however, they used a special device and it was not convenient for collecting data. Wu [6] used k-nearest neighbors as an excellent classifier, but it still cannot effectively classify very similar activities. Zheng [7], Zeng [8], Ronao [9] and Yang [10] both adopted convolutional neural networks to recognize activities using sensor data. Zheng used a convolutional layer, which cannot get more features. Zeng only introduced computation steps and did not refer to the convolutional neural networks structure. Plotz [11] used restricted Boltzmann machines to extract features. Li [12] and Bhattacharya [13] both used sparse-coding to extract features.

With the deep learning skills steady development, more and more fields use deep learning skills to improve efficiency. In this paper, we will adopt deep learning skills and LSTM neural networks to recognize activities.

III. HUMAN ACTIVITY RECOGNITION WITH LSTM NEURAL NETWORK

In this section, this paper discusses LSTM-based feature extraction approach. Fig.1 shows the training and validation procedures, consisting of training phase and test phase. The features are extracted from raw time series sensors data set and then used to train a perfect classification model until the stopping criteria are achieved during the training phase. In the test phase, features have been extracted by training phase and produce a model which contains activity features. When the test set was input into the neural networks, it also was extracted feature. When the features of the test set were matched with the model, then the model will give the data a predicted label. If the predicted label was equal with real label, it would show the result is right.

FIGURE I. FIGURE TRAINING AND VALIDATION PROCEDURES

LSTM is a kind of recurrent neural network which is proposed by Schmidhuber in 1997 [14]. The LSTM neural network is designed to address long-term dependencies and can remember long-term information without the need for lots of hyper-parameters and adjusting hyper-parameters process. Figure 2 shows the LSTM unit structure. LSTM unit contains three gates which were used to maintain and control status information. An LSTM memory cell contained four parts such as an input gate, a forget gate, an output gate and a neuron with a self-recurrent connection. The input gate can allow incoming signal to alter the state of the memory cell or block it. The forget gate can adjust the self-healing connection of memory cells so that the cell remembers or forgets its previous state as needed. The self-recurrent connection has a weight of 1.0 and ensures that the state of the memory cell can remain unchanged from one time step to another except with external disturbances. The gates are used to regulate the interaction between the memory cell itself and its environment. The Figure 2 can be expanded like Figure 3.

FIGURE II. AN LSTM MEMORY CELL

FIGURE III. AN EXPANDED LSTM STRUCTURE

FIGURE IV. THE DETAIL OF A LSTM MEMORY CELL

Figure 4 shows the detail of a LSTM memory cell. LSTM neural networks perform the computation processes which are defined as (1), (2), (3), (4), (5) and (6).

Equation (1) computes the value for i_t , the input gate and Equation (2) computes the value for the states of the memory cells at time t .

$$i_t = \sigma_i(w_i \cdot [h_{t-1}, x_t] + b_i). \quad (1)$$

$$D_t = \sigma_c(w_c \cdot [h_{t-1}, x_t] + b_c). \quad (2)$$

Equation (3) computes the value for f_t , the forget gate at time t .

$$f_t = \sigma_f(w_f \cdot [h_{t-1}, x_t] + b_f). \quad (3)$$

Equation (4) computes C_t the memory cells' new state at time t .

$$C_t = f_t * C_{t-1} + i_t * D_t. \quad (4)$$

Equation (5) computes the values for o_t and h_t , the output gates at time t .

$$o_t = \sigma_o(w_o \cdot [h_{t-1}, x_t] + b_o), \quad (5)$$

$$h_t = o_t * \sigma_h(C_t), \quad (6)$$

where i, f, o and C are input gate, forget gate, output gate and cell state, h is hidden value, σ is the activation function, w is weight matrix, and b is the bias term.

Then, a simple softmax classifier is used to recognize activities, which is placed at the last layer. The final result is a probability value, which can tell people that the probability that the data will be considered an activity. The probability is defined as (7).

$$P = \arg \max_c p(y = c | x) = \arg \max_c \frac{\exp(o_t)}{\sum_{k=1}^K o_t}, \quad (7)$$

where c is a class label, x is a sample feature, y is label variable, K is the number of class, o_t is same value of (5).

In order to get a better neural networks model, regularization and hyper-parameters are considered. When large weights happens, regularization can add an extra term into the cost function which punishes large weights. In this paper, L2 regularization was adopted, and it is defined as (8). Hyper-parameters have a great influence on the model all the time, and we have to adjust them to get a better result. Unfortunately, there are not any standards that can tell us how to adjust hyper-parameters, so we have to try adjusting them according to our experience.

$$E = E_0 + \lambda \sum_w w^2, \quad (8)$$

where E_0 is the original cost function, λ is the weight decay coefficient, and $\lambda \sum_w w^2$ is the penalization term.

IV. EXPERIMENT

In this paper, the public data set of UCI (University of CaliforniaIrvine) was used [15]. The dataset contains accelerometer and gyroscope data which produces from 30 persons performing 6 different activities, including walking (WK), walking upstairs (WU), walking downstairs (WD), sitting (SI), standing (ST) and laying (LY). There are 21 subjects which are used for training, and the others are used for testing [16]. These sensor data were sampled at a rate of 50 Hz, and the length of window was 128 values which stood for one example for one activity. In this paper, we put 6-axes (3-axes of accelerometer and 3-axes of gyroscope) in a line and three layers of LSTM neural network were used. Fig. 5 shows the structure of experimental neural network. Table 1 shows the experimental setup.

TABLE I. EXPERIMENT CONFIGURATION

Parameter	Value
The size of input vector	128*6
The size of batch	128
Learning rate	0.5-0.005
Activation function	ReLU
Regularization rate	0.5-0.005
Hidden points	128-300
Epochs	1000-1400
The probability of dropout(DR)	0.5-1

In order to get a better result, three layers LSTM neural networks was designed for this paper. Then, the learning steps were carried out according to Figure 1. Figure 6 and Table 2 show the LSTM-based model achieves classification accuracy of 90.4%.

TABLE II. RESULTS OF CONFUSION MATRIX

		Predicted Class						
		WK	WU	WD	ST	SD	LY	Recall
True Class	WK	541	9	11	0	3	0	0.959
	WU	4	522	11	0	2	0	0.968
	WD	0	1	418	0	3	7	0.974
	ST	0	1	0	402	29	24	0.881
	SD	2	1	2	28	429	22	0.886
	LY	0	0	0	29	16	424	0.904
	F1	0.97	0.97	0.96	0.87	0.89	0.93	0.904
	Precision	0.98	0.97	0.94	0.87	0.90	0.89	
		9	8	6	6	9	6

FIGURE V. PERFORMANCE OF LSTM NEURAL NETWORK

To analyze the results in more detail, we find that this LSTM neural network structure is good at discriminating the activities such as walking, walking upstairs, walking downstairs, however, but not for others. Owing to the others has features that are more similar in dataset, which leads to depress the overall accuracy.

V. CONCLUSION

In this paper, we proposed LSTM neural networks to recognize human activities using accelerometer sensor and gyroscope sensor on a smartphone. The results show that this method achieves human activities classification with an accuracy of up to 90%.

In the future study, we will bring in large data sets, which can improve the recognition accuracy and enhance the robustness of the neural networks model. We will modify the LSTM neural networks structure to make the neural networks model recognize similar activities.

ACKNOWLEDGMENT

This work was supported by the Natural Science Foundation of China under Grant Nos. 51668043 and 61262016 and the CERNET Innovation Project under Grant Nos.NGII20160311 and NGII20160112.

REFERENCES

- [1] Y. Kwon, K. Kang, and C. Bae, "Unsupervised learning for human activity recognition using smartphone sensors," *Expert Systems with Applications*, vol. 41, no. 14, pp. 6067-6074, 2014
- [2] O. Dobrucali, and B. Barshan, "Sensor-Activity Relevance in Human Activity Recognition with Wearable Motion Sensors and Mutual Information Criterion," *Informationences & Systems*, vol. 264, pp. 285-294, 2013.

- [3] W. Jiang, and Z. Yin, "Human Activity Recognition Using Wearable Sensors by Deep Convolutional Neural Networks," pp. 1307-1310, 2015.
- [4] I. Kouris, and D. Koutsouris, "A comparative study of pattern recognition classifiers to predict physical activities using smartphones and wearable body sensors," *Technology & Health Care Official Journal of the European Society for Engineering & Medicine*, vol. 20, no. 4, pp. 263-275, 2012.
- [5] A. M. Khan, "Recognizing Physical Activities Using Wii Remote," *International Journal of Information & Education Technology*, vol. 3, no. 1, 2013.
- [6] W. Wu *et al.*, "Classification accuracies of physical activities using smartphone motion sensors," *Journal of Medical Internet Research*, vol. 14, no. 5, pp. e130, 2012.
- [7] Y. Zheng *et al.*, *Time Series Classification Using Multi-Channels Deep Convolutional Neural Networks*: Springer International Publishing, pp. 197-205, 2014
- [8] M. Zeng *et al.*, "Convolutional Neural Networks for Human Activity Recognition using Mobile Sensors." pp. 197-205, 2014
- [9] C. A. Ronao, and S. B. Cho, "Deep Convolutional Neural Networks for Human Activity Recognition with Smartphone Sensors." pp. 46-53, 2015.
- [10] J. B. Yang *et al.*, "Deep convolutional neural networks on multichannel time series for human activity recognition." pp. 3995-4001, 2015.
- [11] T. Plötz, N. Y. Hammerla, and P. Olivier, "Feature Learning for Activity Recognition in Ubiquitous Computing." pp. 1729-1734, 2011.
- [12] Y. Li *et al.*, *Unsupervised Feature Learning for Human Activity Recognition Using Smartphone Sensors*: DBLP, pp. 99-107, 2014
- [13] S. Bhattacharya, P. Nurmi, and N. Hammerla, "Using unlabeled data in a sparse-coding framework for human activity recognition," *Pervasive & Mobile Computing*, vol. 15, no. C, pp. 242-262, 2014.
- [14] S. Hochreiter, and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.
- [15] UCI Machine Learning Repository. [EB/OL]. <http://archive.ics.uci.edu/ml/index.php>.
- [16] D. Anguita *et al.*, "A Public Domain Dataset for Human Activity Recognition using Smartphones," pp. 437-442, 2013.