ATLANTIS PRESS
Advances in Computer Science Research, volume 87
3rd International Conference on Mechatronics Engineering and Information Technology (ICMEIT 2019)

Research on Human Behavior Recognition based on Deep Neural Network

Shanshan Guan[1, a], Yinong Zhang[2, b, *] and Zhuojing Tian[1, c]

[1] Beijing Key Laboratory of Information Service Engineering, Beijing Union University, Beijing 100101, China;

[2] College of Urbun Rall Translt and Logistics, Beijing Union University, Beijing 100101, China.

[a]mountain_guan@126.com, [b, *]zdhyinong@126.com, [c]tianzhuojing@foxmail.com

Abstract. In order to improve the recognition rate of human behavior by intelligent terminals, a network model for deep learning of human behavior recognition is proposed. Time series data is transformed into a deep network model by performing motion segmentation using a sliding window algorithm. Feature vectors are imported into the SoftMax classifier through end-to-end research, which identifies six daily behaviors such as walking, sitting, going upstairs, going downstairs, standing and lying down. By comparing the recognition effects of different models, it was found that the convolutional neural network introduced into Dropout achieved better recognition results in UCI HAR dataset.

Keywords: Behavior recognition; Deep learning; Filter; Behavior segmentation; SoftMax classifier.

1. Introduction

With the development of science and technology as well as the improvement of people's living standards, intelligent products have penetrated into every aspect of people's lives. In the aspect of human behavior recognition, acceleration sensitivity with small size, low power consumption and high sensitivity have been widely used. Li Dong et al. [1] designed a set of elderly fall detection system based on the acceleration sensor via detecting the posture of fall. At present, many smart devices are equipped with various sensors such as accelerometers, gyroscopes, and direction sensors, like smartphone and smartwatch. Therefore, behavior recognition based on intelligent terminal is possible.

Behavior recognition based on intelligent terminals is an emerging research branch of pattern recognition. Acceleration sensor is used to obtain acceleration data information when the user is active, and the data is analyzed to determine the user's behavior category. Because the acceleration sensor can only collect acceleration signals from specific parts of the human body. Therefore, the difference in the number of acceleration sensors and the placement position means that the representation of the motion to be recognized is different. Li Shuang et al. [2] placed the thigh and calve using two accelerometers to obtain the movement information of human lower limbs; Fan Lin et al. [3] identified the 20 common behaviors of human body in daily life by accelerometers carried in five positions of the human body. Wang Xichang et al. [4] introduced the accelerometer sensor to the three-axis acceleration information acquired by the front and rear arms of the right hand to realize the recognition of the upper limb movement. Su Benyue et al. [5] used a single lumbar sensor to obtain gait information, using functional data analysis and hidden Markov model (HMM) to combine human gait recognition.

The key to behavior recognition is the extraction of feature vectors that characterize behavioral features. Early feature extractions are usually designed for specific purposes and may not be extended to other scenarios. Learning and acquiring features directly from data is more general than manual annotation, so learning features from data becomes a viable solution. In 2006, Deep Learning [6] proposed that it does not require any manual setting of any features and performance can outweigh the manual extraction of features, so it has been widely used in many fields. Thus, this paper will use the deep learning network model to recognize the posture of human body.

2. Deep Neural Network Model of Human Behavior

Deep learning refers to a learning function model composed of multiple network layers, which is used to extract the characteristics of input data and the abstract features of high-latitude for data classification and combination to obtain more structured results. As a result, in order to better obtain the characterization of different behaviors, this paper will use long-term short-term memory model [7,8] (Long Short-Term Memory, LSTM) and deep convolution network model to extract features.

2.1 Long Short-Term Memory Neural Networks

Recurrent Neural Networks (RNNs) have achieved great success and wide application in Natural Language Processing (NTL) due to their superiority in time modeling [9,10]. Considering the superiority of cyclic neural networks for time modeling, a circular neural network will be used to extract time domain information better [11]. Given a continuous input sequence $x=(x_1,\cdots,x_t)$, the traditional RNN network contains hidden vector sequences $h=(h_1,\cdots,h_t)$ and output vector sequences $y=(y_1,\cdots,y_t)$. From the time interval $t=(1,T)$, we know

$$h_t = H(W_{ih}x_t + W_{hh}h_{t-1} + b_h) \tag{1}$$

$$y_t = W_{ho}h_t + b_o \tag{2}$$

where W is the weight matrix, b is the base vector, and H is the hidden layer unit activation function.

Due to the gradient extinction problem of RNN [12], an improved model was proposed, which is a long-term and short-term memory model. Compared to traditional RNNs, LSTM uses a structure called memory cells to preserve the current state x_t and the saved state h_{t-1} of the previous frame, allowing for better modeling of long-term sequence dependency problems.

The LSTM update status is as follows:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}C_{t-1} + b_i) \tag{3}$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}C_{t-1} + b_f) \tag{4}$$

$$O_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}C_{t-1} + b_o) \tag{5}$$

$$C_t = f_tC_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{6}$$

$$h_t = O_t \tanh(C_t) \tag{7}$$

where σ is the activation function, i is the input gate, f is the forgotten gate, O is the output gate, and C is the memory cell activation function.

2.2 Convolutional Neural Network Model

Deep learning has a multi-level structure [13] that can handle complex feature extraction problems. As a typical model of deep learning, convolutional neural networks have been widely used in many fields such as speech recognition, natural language processing, and pattern recognition. The convolutional neural network is a multi-layered deep network, generally consisting of a convolutional layer and a pooled layer, and finally connected to the fully connected layer. Convolutional neural network is a deep learning model in which convolutional layer and pooled layer form a multi-layer network alternately.

In the convolution process, after input data, input data are processed by multiple trainable convolution kernels for convolution operation, then obtain a convolution layer. The function of the convolutional layer is feature extraction. The neurons of each convolutional layer are connected with the data in the local receptive field of the previous layer to extract local features in the local receptive field. Each convolution kernel can extract a certain feature. After the convolutional layer, usually the pooling layer, the down-sampling operation is performed on the obtained feature map. The purpose of down-sampling is to perform further feature extraction on the resulting features. The pooling layer reduces the amount of data that needs to be processed while extracting useful feature information. This special network structure allows the convolutional neural network to obtain a higher recognition rate when it is identified.

2.3 Overfitting Problem

Overfitting is a common problem in deep neural networks (DNN). The neural networks reduce parameters by weight sharing, so that the training time is much less, but there is still an over-fitting problem. In the past, it was common to solve fitting problems by combination method, that is, combining a plurality of models, but the training time is long and the test is troublesome.

Srivastava et al. [14] proposed to solve the over-fitting problem by adding a Dropout layer, which has been proved to effectively improve the over-fitting problem that occurs in neural networks. In order to avoid the fitting problem, this paper adds dropout layer to the network model.

3. Experiment and Result Analysis

3.1 Dataset

The data used in this paper were collected from the UCI HAR dataset [15] by 30 volunteers aged between 19 and 48. The data set was randomly divided into two groups, 70% of which were training sets and 30% of which were test sets.

3.2 Classifier

The SoftMax classifier maps the signals to be separated to the corresponding labels. During training, the data signal passes through the data processing process of the deep network model to obtain a classification result, which is compared with the corresponding label data to calculate the corresponding relative error, and the weights in the network are continuously adjusted by iterating a certain number of times so that relative errors continue to decrease and eventually converge, the test set is then input into the network for test classification.

3.3 Analysis Experimental Results

As the number of iterations increases, the accuracy of each model continues to increase. Figure 1 shows the classification error curves obtained by the two models and their corresponding two improved models on the test set.

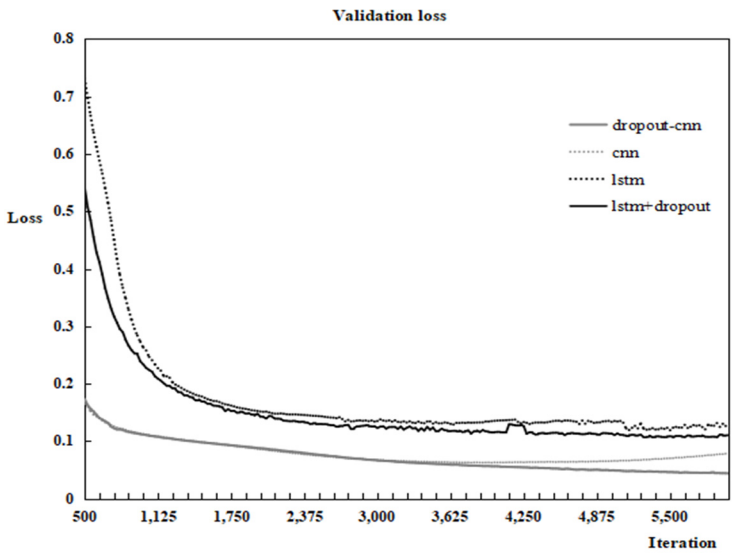

Fig. 1 Test error curves for four models

It can be seen from Figure 1 that both the traditional CNN model and the improved LSTM model have slightly lower error slowing speed and convergence speed than the traditional model. From the iteration of 4000, it can be found that the test error of the traditional CNN model and the LSTM network model begins to increase, which is the result of network over-fitting.

The results of the average recognition rate of the method used in the paper are shown in Table 1 shows the recognition rate curve of each model during the training process. From Figure 1 and Table 1, it can be found that after adding random dropout, the model has more generalization ability and better recognition effect, and effectively prevents overfitting.

Table 1. Experimental results of the methods

Behavior recognition method	Average recognition accuracy(%)
CNN	90.87
LSTM	87.58
CNN with dropout	93.00
LSTM with dropout	91.73

In addition to comparing the results before and after the improved model, the results are compared with the results of the existing method in the same public database UCI HAR dataset. As can be seen from Table 2, two improved recognition methods designed by us have achieved better results in data sets.

Table 2. Experimental results of different methods

Behavior recognition method	Average recognition accuracy(%)
CNN with dropout	93.00
LSTM with dropout	91.73
CNN [16]	90.98
LSTM [16]	87.38

4. Conclusion

Experimental results show that the improved convolutional neural network has a good effect on human behavior recognition of intelligent terminals, and it deserves further research. This article uses data from the common data set to examine the model, rather than using actual data to verify its effectiveness. To further apply this in real life, the next step is to evaluate the model using real-world scenario data.

References

[1]. Dong Li, Shan Liang. Design of fall detection device for elderly people based on accelerometer[J]. Transducer and Microsystem Technologies,2008,27(9):85-88.

[2]. Shuang Li, Zhi-zeng Luo, Ming Meng. Acquisition method for the lower limb motion information based on accelerometers[J]. Mechanical & Electrical Engineering Magazine, 2009, 26(1):5-7.

[3]. Lin Fan, Zhong-min Wang. Human activity recognition model based on location-independent accelerometer[J]. Application Research of Computers,2015,32(1):63-66.

[4]. Changxi Wang, Xianjun Yang, Qiang Xu et al. Motion Recognition System for Upper Limbs Based on 3D Acceleration Sensors[J]. Chinese Journal of Sensors and Actuators,2010,23(6):816-819.

[5]. Benyue Su, Dandan Zheng, Min Sheng. Daily Behavior Recognition with Single Sensor Based on Functional Time Series Data Modeling[J]. Pattern Recognition and Artificial Intelligence, 2018,31(7):653-661.

[6]. Hinton G E, Salakhutdinov R R. Reducing the Dimensionality of Data with Neural networks[J]. Science, 2006, 313(5786): 504-507.

[7]. Graves A. Long Short-Term Memory[M]// Supervised Sequence Labelling with Recurrent Neural Networks. Springer Berlin Heidelberg, 2012:1735-1780.

[8]. Jürgen Schmidhuber. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8):1735-1780.

[9]. Xiong W, Droppo J, Huang X, et al. The Microsoft 2016 conversational speech recognition system[C]. IEEE International Conference on Acoustics. IEEE, 2017:5255-5259.

[10]. Cho K, Van Merrienboer B, Bahdanau D, et al. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches[J]. Computer Science, 2014.

[11]. Graves A, Mohamed A.R., Hinton G. Speech Recognition with Deep Recurrent Neural Networks[J]. 2013,38(2003):6645-6649.

[12]. Pascanu R, Mikolov T, Bengio Y. On the difficulty of training Recurrent Neural Networks[J]. 2013, 52 (3): III-1310.

[13]. Weiguo Sheng, Pengxiao Shan. Negative correlation neural network ensemble based on a niching algorithm[J]. Journal of Zhejiang University of Technology,2016,44(5):482-486.

[14]. Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014, 15(1):1929-1958.

[15]. Information on: https:// archive. ics. uci. Edu /ml / datasets /Human + Activity + Recognition +Using+Smartphones.

[16]. Xiaohua Kuang, Jun He, Shaohua Hu, et al. Comparison of deep feature learning methods for human activity recognition[J]. Application Research of Computers,2018.35(9):2815-2822.