Activity Recognition using 1D convolution from Accelerometers Data

Jian Wang¹, Xue Hua liu²*

¹Network Technology Department, South China Institute of Software Engineering, Guangzhou, GuangDong, 510900, China
²Computer Department Department, South China Institute of Software Engineering, Guangzhou, GuangDong, 510900, China
*Corresponding author’s e-mail: wjn@sise.com.cn

Abstract. Aiming at the problem of activity a recognition method based on a convolutional neural network was proposed in this paper, which can effectively classify 6 types of human movements: Downstairs, Jogging, Sitting, Standing, Upstairs and Working. The network consists of an input layer, two convolutional layers, two pooling layers, a fully connected layer, and an output layer. The sliding window is used to transform the sensor data into a three-channel RGB image format, and the features of the three-axis speed sensor data are automatically extracted to classify each action. The model was reproduced using Tensorflow, and the recognition rate of 89.35% was achieved on the open source database WISDM. The experimental results show that the amplifier has a better effect on human movement recognition.

1. Introduction
With the development of the Internet of Things and the application of various embedded devices, we are able to collect a large amount of sensing data. More and more edge computing devices, such as mobile phones, integrate a variety of powerful sensors, including accelerometers, gyroscopes, light sensors, GPS, etc. These sensors have become data sources for measurement applications. Especially the acceleration sensor in the mobile phone, the real-time acceleration data returned by it along the x-axis, y-axis or z-axis can be used for activity recognition. Compared to video recognition methods, this method is not susceptible to interference from natural environments, such as light and darkness, angles, and obstructions.

2. Related Work
A large number of successful cases have appeared in this field et [1]. The method was first proposed in the early 2000s et [2] [3]. Numerous research results have proved that the method is feasible, for example: Tapia et al. et [4] It is proposed that daily activities can be identified through sensors. Hollosi et [5] applied sensor data to early childhood education. Although traditional machine learning methods such as support vector machines and decision trees have made some progress et [6], the classification effect of activity is still not ideal. Because neural network technology has better feature extraction capabilities, it is currently widely used in the field of image processing. Inspired by the
above studies, this paper applies convolutional neural networks to the sequence data processing of sensors to solve the problem of activity recognition.

3. Our Proposal
In this paper, we use a convolutional neural network (CNN) as a classifier of activity, and propose a method for activity recognition using acceleration sensor data. The proposed method acquires the acceleration data of its activities by wearing sensors on the wrist and thigh of the user, and then uses 1-dimensional convolution to perform feature extraction on the input sequence data, and finally realizes human body recognition. We used the data set provided by Wireless Sensor Data Mining Lab Fordham University[7], This data set was collected in a controlled experimental environment. This dataset contains a total of 1,098,203 records, six activity categories, including Downstairs, Jogging, Sitting, Standing, Upstairs and Working, The distribution of this dataset with activities (labels) is shown in Fig. 1.

![Figure 1. The distribution of this dataset with activities](image)

In this project, researchers recorded acceleration data for three axes of daily movements. The data characterizes the forward motion of the leg through the z-axis; the y-axis characterizes the user's upward and downward motion; and the x-axis characterizes the user's horizontal leg motion.

![Figure 2. Axes of Motion Relative to User](image)

4. Data preprocessing
The main work of this part is data normalization, this can improve the convergence rate of the model and model accuracy. More important is the need for learning algorithms, the reason, in some models, the optimal solution is not equal to the original one after the inhomogeneous expansion of each dimension. For example, SVM For such a model, unless the distribution range of the data in each dimension is relatively close, it must be standardized to avoid the model parameters being dominated
by the data with a larger or smaller range. Some models, such as logistic regression, are equivalent to the original optimal solution after non-uniform scaling in various dimensions.

CNN was first applied to the classification task of 2D images et [8]. The principle of CNN work is to first use the convolution kernel to extract features from the pixels of the original picture, and then abstract the convolution data into smaller dimensions through the pooling layer. The data. Determine the spatial correlation between them by subsampling the maximums. Finally, the softmax layer is used for classification.

The CNN is essentially an iterative process, for models with scalable invariants, it is best to standardize the data. The case where the maximum and minimum values are unknown, or the case where there is outlier data, so in this paper, we followed zero-mean normalization. Zero-mean normalization is also called Z-score normalization. The processed data conform to the standard normal distribution, that is, the mean value is 0 and the standard deviation is 1. The values of attribute $A$ are normalized using the mean and standard deviation of $A$. A new value $v'_i$ is obtained using the equation (1), where $\bar{A}$ and $\sigma_A$ are the mean and standard deviation of attribute $A$, and using the equation (2)(3)

$$v'_i = \frac{v_i - \bar{A}}{\sigma_A} \quad (1)$$

$$\bar{A} = \frac{1}{N} \sum_{i=1}^{N} x_i \quad (2)$$

$$\sigma_A = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2} \quad (3)$$

We divide the sensor data into data segments of the same length and change the dynamic data sequence into static data in order to extract the spatiotemporal characteristics of the data. This method can better process sequence data, and analyze and recognize activity from it. The down stairs are taken as an example. Figure 3 shows the processed sequence data in a visual way.

Figure 3. Time series plots of Acceleration data
5. Model Structure & Training Process

5.1. Model Structure

In this section, we discuss our CNN architecture. Fig. 4, shows the structure of the proposed approach. It has three kinds of layers: 1) input layer whose values are fixed by the input data; 2) hidden layers whose values are derived from previous layers; 3) output layer whose values are derived from the last hidden layer.

Consider the actual situation of human movement, we propose setting the window size to 4.5s. The window size is set to 90. The overlap between each time window is half the size of the window. Therefore, the sample size of input layer in CNN is 1*90*3, the input data is accelerated by a preprocessing depth of 3, then the behavior features are extracted by convolution and pooling, after the data is preprocessed, it enters the first volume of the convolutional neural network. The window size of convolution $L_1$ is 1*10, there are 60 convolution windows with different values. The data is processed by the convolution $L_1$ and then enters the pool layer. Sampling the Max-pooling with the maximum value, and then enters the convolution $L_2$ to process. The window size of convolution $L_2$ is 1*6, there are 10 convolution windows with different values in convolution $L_2$. Finally, the recognition of actions is carried out through the identification of full link layer and Softmax level.

Formally, extracting a feature map using a convolution operation is given equation (4) et [9]

$$a_j^{(l+1)}(\tau) = \sigma \left( b^l + \sum_{f=1}^{F^l} K_{jf}^l(\tau) \ast a_j^l(\tau) \right) = \sigma \left( b^l + \sum_{f=1}^{F^l} \left[ \sum_{p=1}^{P^l} K_{jf}^l(p)a_j^l(\tau-p) \right] \right) \quad (4)$$

Where $a_j^l(\tau)$ denotes the feature map $j$ in layer $l$, $\sigma$ is a Rule non-linear function, $F^l$ is the number of feature $l$, $K_{jf}^l$ is the kernel convolved over feature map $f$ in layer $l$ to create the feature map $j$ in layer $(l+1)$, $P^l$ is the length of kernels in layer $l$ and $b^l$ is a bias vector.

5.2. Training Process

As you know, when the learning rate is set too low, the model's convergence speed will be very slow, and when the learning rate is set too high, a gradient explosion will occur, making the model unable to converge. Therefore, in this paper we use Adam as the model (set the initial learning rate to 0.001) to avoid the problem of improper learning rate value. Which flow as equation (5), $\alpha_{\text{start}}$ is the initial learning rate, $\alpha_{\text{decayed}}$ As the decay rate of each round of learning, $\lambda_{\text{global}}$ is the current learning steps, $\lambda_{\text{decay}}$ is the steps per round of learning, it flow as equation (6), $S$ is the total number of samples, $B$ is the size of each batch.

$$\alpha_{\text{decayed}} = \alpha_{\text{start}} \times (\alpha_{\text{decay}})^{\lambda_{\text{global}}/\lambda_{\text{decay}}} \quad (5)$$
\[ \lambda_{\text{decay}} = \frac{\lambda}{\theta} \quad (6) \]

The loss function is defined by Softmax loss as equation (7), because we assume that the samples are subject to Bernoulli distribution. \( L \) is the loss, \( S_j \) is the value of the output vector \( S \) from softmax, \( y_i \) is the desired output.

\[ L = \sum_{j=1}^{T} y_i log S_j \quad (7) \]

6. Experimental Analysis

All our experiments are conducted in Tensorflow on an ordinary computer with a GTX 1060 GPU and 8GB memory. We take cross validation, using random sampling method, the whole sample was divided into three parts by random sampling method. Two of them were trained and one was tested. Cross validation has lower accuracy than sample test set, but it is closer to classification accuracy of unlabeled samples, to analyze the results in more detail, we show the confusion matrix (Table 1) and Normalized Confusion matrix (Fig. 5) for the Actitracker dataset using CNN.

| Activity     | Accuracy rates |
|--------------|----------------|
| Jogging      | 70.79%         |
| LyingDown    | 97.40%         |
| Sitting      | 94.17%         |
| Stairs       | 93.78%         |
| Standing     | 65.25%         |
| Walking      | 93.82%         |

Figure 5. Confusion Matrix for result of experiment

7. Conclusion

Inspired by convolutional neural networks in machine vision, we designed and implemented a convolutional neural network model for human activity recognition based on smart phone triaxial acceleration sensor information. It has good universality and high recognition accuracy for walking, sitting and standing, the accuracy rate for upstairs and downstairs is not satisfactory although it reached an accuracy rate of 83.94%. It has advantages of higher total recognition rate, better fault tolerability and stability, stronger robustness, and is more convenient and flexible to use.

Acknowledgments

School-level quality engineering project of South China Institute of Software Engineering in 2018(No.JXTD201804),The Key Scientific Research Projects of Guangdong Universities in 2018(No.2018KTSCX341)

References

[1] The Ambient Assisted Living Joint Programme, http://www.aal-europe.eu/
[2] A. Hey Laurence, Reduced complexity algorithm for cognitive packet network [J]. Computer
Communications. 2008, 31(16): 3822-3830
[3] L. Bao and S. Intille, “Activity recognition from user annotated acceleration data”, In Proc. Pervasive, LNCS 3001, pp. 1–17, 2004
[4] D.J. Patterson, L. Liao, D. Fox and H. Kautz, “Inferring High-Level Behavior from Low-Level Sensors”, Proc. of the 5th Conference on Ubiquitous Computing (UbiComp), PP. 73-89, 2003
[5] E. Munguia-Tapia, S.S. Intille and K. Larson, “Activity recognition in the home using simple and ubiquitous sensors”, Proc. of Pervasive, pp. 158-175, 2004
[6] D. Hollosi, J. Schroder, S. Goetze, and J.E. Appell, “Voice Activity Detection Driven Acoustic Event Classification for Monitoring in Smart Homes”, The 3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies, pp. 1-5, 2010
[7] Jennifer R. Kwapisz, Gary M. Weiss and Samuel A. Moore (2010). Activity Recognition using Cell Phone Accelerometers, Proceedings of the Fourth International Workshop on Knowledge Discovery from Sensor Data (at KDD-10), Washington DC.
[8] Kwapisz J R, Weiss G M, Moore S A. Activity recognition using cell phone accelerometers[J]. Acm Sigkdd Explorations Newsletter, 2011, 12(2): 74-82.
[9] Francisco Javier Ordóñez *, Daniel Roggen. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition[J]. Sensors 2016, 16, 115; doi:10.3390/s16010115