

Research of Comparison on Convolution Network and BP Network Based on Human Body Attitude Recognition

Yang Du¹, Jiaxin Tian², Binghong Zhan³, Fei Guo^{3,*}

1. Information Engineering Institute Beijing Institute of Fashion Technology

2. Information Engineering Institute Beijing Institute of Fashion Technology

3. Academy of Art & Design Beijing Institute of Fashion Technology

Beijing, China

18310236998@163.com, 617724382@qq.com, ysylab@bift.edu.cn, lunwen95@163.com

Abstract—In this thesis, after collecting data for human wearable acceleration sensor, human posture is recognized by using traditional BP network and convolution network based on Tensorflow in Python. Through respectively introducing and using BP network and convolution network, we then made a comparison experiment. It had become clear that the effect of using convolution network is much better than using BP network for the posture mentioned in the article. Its accuracy rate is 75%.

Keywords—Acceleration Sensor; Human motion recognition; BP Network; Convolution Network; Python; TensorFlow; Deep Learning

I. INTRODUCTION

There are many experts in the field of human motion recognition. For this field, there are two aspects, the first is based on image, and another is for sensors. However, for sensors, two skills are the most significant. One is optical sensors, other skill is acceleration sensor[1]. The most main image processing is used in image. The information is extracted from image to re recognize. For image, there are many factors that can effect image, such as light, the height of the camera resolution and skill of photography etc. For optical sensors, because the data is collected by locating the point of light, the ambient light can effect information in these data. Thus, acceleration sensor is used in this thesis use for this field. In this thesis, through using platform of acceleration sensor to collect data[2], BP network and Convolution Network is used to recognize and classify. Finally, a comparison experiment is achieved, and it is proved that the effect of Convolution Network for human motion recognition is much better than traditional BP network.

II. BP NETWORK

A. The Theory of BP Network

1) The concept of BP Network

BP neural network is called error feedback network[3], it is a network of have supervision and having two procession:

①. The procession of signal forward propagation: As forward propagation, after incoming from input layer input sample is dealt by each hidden layer, then the result of hidden layer can be transfer to output layer. If the reality output of output layer is not fit expected output, the propagation transfer to the reverse propagation stage of error.

②. The procession of the reverse propagation stage of error: As reverse propagation, the error of between reality value and expected value is propagated from output layer to input layer. The error would be allocated to each layer, through this procession, this error signal is used to correction weight value.

Its schematic diagram just like figure 1:

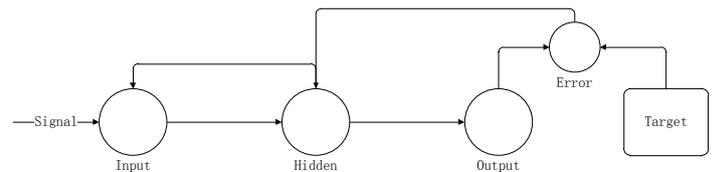


fig.1 BP Network Schematic Diagram

2) BP Network Structure

The structure of BP network has three parts. For one is input layer that is used to input signal, the second part is called hidden layer which is used to process the signal for input layer, the final part is output layer that is used to output the final result. The fig. of its structure is like figure 2:

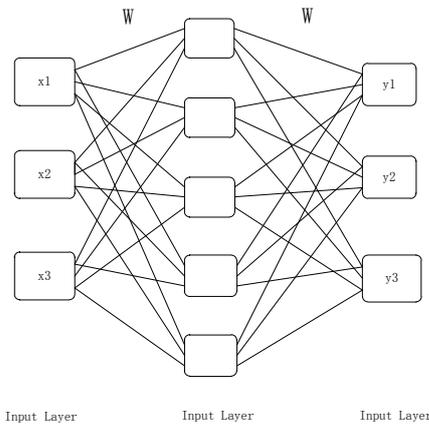


fig.2 BP Network Structure

In this figure, it just built a simple network, which input layer is 3dimension the same as output. Its transforming matrixes expressed as: the matrix of input layer is $X[3*5]$, the matrix of hidden layer is dependent on input layer and output layer. For this network, output is 3dimension so the matrix of hidden layer is $W[5*3]$, and W is called join weights matrix. Using the formula 1 to express :

$$X[3*5]*W[5*3]=Y[3*3] \quad (1)$$

B. Build BP Network

For the building of BP network, three parts must be confirmed:

- 1) The number of neurons must be confirmed for input layer, hidden layer and output layer. N is identified as the number of neurons for input layer. P is identified as number of neurons for hidden layer. And Q is defined to output layer.
- 2) Generally according to experience, selection of initial weight value is between -1 and 1[4].
- 3) The selection of learning rate, it is selected as 0.001. A large learning rate can lead to system instability. But the small learning rate leads to a slow convergence and a longer training time.
- 4) Activation function selection, general use of the Sigmoid function.

After confirmed the information of network, its running step as follows: initialization, signal forward propagation, network output and target do poor, error signal back propagation, repeat iterative correction network. The calculation of error is used in Type 2:

$$E = \frac{1}{2m} \sum_{k=1}^m \sum_{o=1}^q (d_o(k) - y_o(k))^2 \quad (2)$$

III. CONVOLUTION NETWORK

A. The Theory of Convolution Network

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the

abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

Convolution Network is the core of the operation of convolutional operations[5], Convolution Expressions in Calculus 3:

$$S(t) = \int x(t - a)\omega(a)da \quad (3)$$

The two-dimensional convolution is defined as:

$$s(i, j) = (X * W)(i, j) = \sum_m \sum_n x(i + m, j + m)\omega(m, n) \quad (4)$$

For this formula, $*$ represents convolution operations, where w is the convolution kernel, and x is input. If x is a two-dimensional input matrix, and W is also a two-dimensional matrix. But if X is a multidimensional tensor, then W is also a multidimensional tensor.

B. The Structure of Convolution Network

In the convolution network structure, including the core of the convolution layer, the next is the activation function, then the pool layer, and finally the fully connected layer, as well as the end of the output layer. This is the approximate structure of the convolution network, as shown in figure 3:



fig.3 Convolution Network Structure Diagram

The convolution network is used in the field of image recognition, as shown in figure 4:

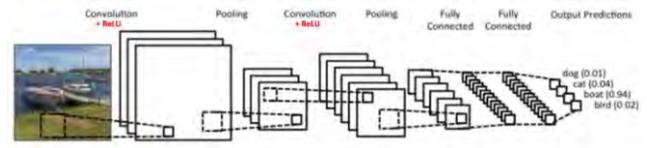


fig.4 Convolution Network in Image

The convolution layer in convolution network is computed by convolution operation, that is, the different local matrix of the input image and the elements of the convolution kernel matrix are multiplied and then added[6]. The procedure is shown in figure 5:

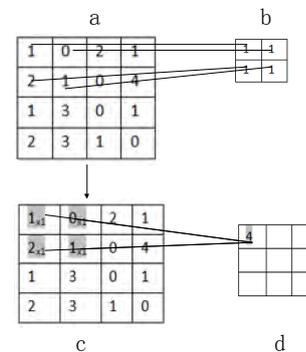


fig.5 Convolution Operations

In the figure, a represents the original data, b represents the convolution kernel, c is the convolution operation, and d is the characteristic quantity obtained after the convolution operation.

The data in the characteristic quantity is the data which is multiplied and followed by the original data and the convolution kernel, here the simple convolution dimension is introduced, which is the $a[4 \times 4] * b[2 \times 2] = d[3 \times 3]$, which is the convolution operation. After convolution, the feature information in the original data can be extracted. After the convolution layer, the output of the convolution layer is using ReLu as activating function, which accelerates the convergence of the network, and the element value corresponding to the position of less than 0 in the tensor of the output becomes 0. The Relu function expression is as follows:

$$y = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (5)$$

The pool layer in the convolution network is also an important part[7]. The so-called pooling is the compression of each sub matrix of the input tensor of the layer. If it is a 2x2 pool, then the sub-matrix of each 2x2 element into an element, if the 3x3 is the pool, then the sub-matrix of each of the 3x3 elements into an element, so that can play a dimension-reducing role. There are two ways of pooling: Maximum pool layer and Mean-value pool layer. Here, the maximum pool is introduced, and its operational ideas are shown in figure 6:

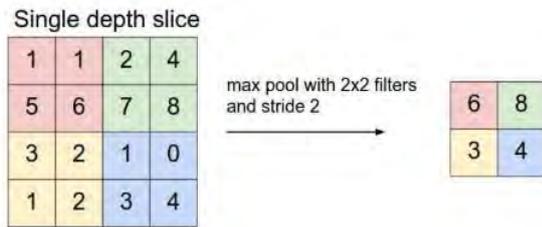


fig.6 Max Pool Process

Finally, the fully connected layer, the two layers of all neurons have a weight connection, usually the full connection layer in the convolution neural network tail. Which is the same way as the traditional neural network neurons: input, middle layer, output layer.

IV. COMPARISON BETWEEN BP NETWORK AND CONVOLUTION NETWORK

A. Comparison Between BP Network and Convolution Network

In this paper, the inertial sensor used for data collection, the collection of nine points of human body data, respectively, X, Y, Z, three axes of the acceleration value of a_x , a_y , a_z . As

shown in figure 7:

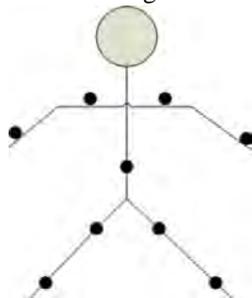


fig.7 Sensor Node Distribution map

The raw data that is obtained is normalized.

$$A = \sqrt{a_x^2 + a_y^2 + a_z^2} \quad (6)$$

The mean filter of a is obtained to the time domain information.

$$T = g(x, y) = \frac{1}{M} \sum_{f \in S} f(x, y) \quad (7)$$

After median filtering, fast Fourier transform (FFT) is carried out to obtain frequency domain information(8).

The data of Matrix $[A, T, F]_{n \times 3}$ is formed by the above data processing.

B. Classification Comparison

Through the above introduction, the BP network and the convolution network are compared to the human attitude recognition, table I is the parameter and the structure setting in two kinds of networks.

TABLE I. COMPARISON TABLE FOR NETWORK ESTABLISHMENT

	BP Network	Convolution network
Network structure and parameters	Input layer Neuron 3	Number of convolution layers 7
	Hidden layer Neuron 7	Convolution nucleus 3x3 Characteristic quantity 32x32
	Input layer Neuron 1	full connection layer size 1024
	Learning Rate 0.0001	Learning Rate 0.0001
	Maximum number of iterations 500	Maximum number of iterations 500

This article uses the Python language to carry on the programming experiment, the Python language has the very good experience to the developer with its simple and clear grammatical structure. In this paper, two kinds of networks are experimented in Python environment, and the figure 8 is obtained.

As those figure can be seen from the graph, the error classification number of BP network increases obviously, and the total error in the convolution network is decreasing, and the regularity fluctuation occurs after 100 times, which belongs to normal phenomenon. The correct rate of the convolution network identification is shown in tableII:

TABLE II. EXPERIMENTAL RESULTS

Number of Tests	Correct quantity	Correct rate
20	16	75%

V. SUMMARY

In this paper, two kinds of networks are introduced, and comparative experiments are carried out. The experimental results show that under the same conditions, the BP network based on inertial sensor can not recognize the attitude of human body, and the network cannot recognize the data in this paper. But in the convolution network under the depth study, the effect is better than the traditional BP network, and the recognition rate can reach 75% under the convolution network, so the convolution network is more suitable for this kind of problem.

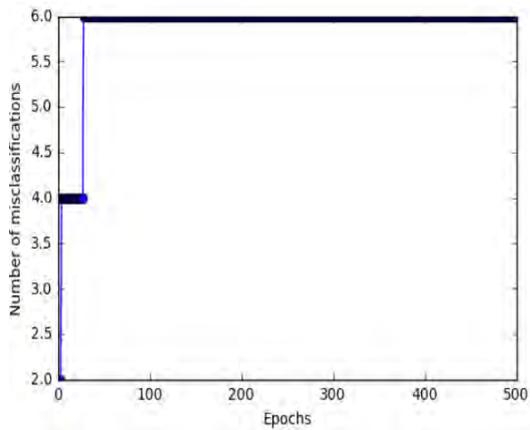
ACKNOWLEDGMENT

In this paper, the research was sponsored by 2013 National Science and Technology Support Program funded projects (Project No. 2013BAH41F02) and 2017 Beijing Municipal Education Commission Science and Technology Program on the surface of the project (Project No. SQKM201710012006) and Research on the construction of the university library resources recommending platform based on user-driven (Project No. KYJH02160201/002/022) and Innovative design of garment ergonomics and apparel functions Beijing key Laboratory (Project No. KYTG02170202) and Research on the application of mobile learning to the improvement of college students ' creativity and practical ability the key project of education and teaching reform in Beijing Garment College (Project No. ZDJG-1505) and 2018 Postgraduate Innovation project of Beijing Institute of Fashion Technology(Project No. 120301990122/009)

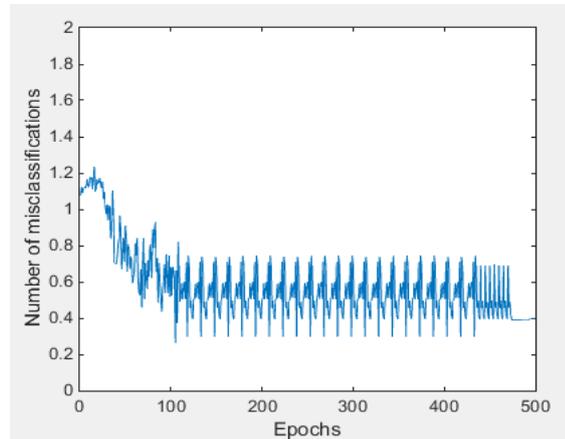
REFERENCES

[1] Yang CHENG. Research on Motion Capture System of Equestrians Based on Nine-Axis Inertial Sensor[A]. Advanced Science and Industry Research Center.Proceedings of 2017 International Conference on Mathematics, Modelling and Simulation Technologies and Applications

(MMSTA 2017)[C].Advanced Science and Industry Research Center.,2017:4.
 [2] Chen C, Jafari R, Kehtarnavaz N. A survey of depth and inertial sensor fusion for human action recognition[J]. Multimedia Tools & Applications, 2017, 76(3):4405-4425.
 [3] Dalong Zhang. A Fault-Tolerant Recognition Method for Services Packets Based on Center Distance And BP Net[A]. IEEE Beijing Section.Proceedings of the 7th IEEE/International Conference on Advanced Infocomm Technology(IEEE/ICAIT2014)[C].IEEE Beijing Section.,2014:7.
 [4] Li J, Cheng J H, Shi J Y, et al. Brief Introduction of Back Propagation (BP) Neural Network Algorithm and Its Improvement[J]. 2012, 169:553-558.
 [5] SHUMIN SHI. Multi-Features Group Emotion Analysis Based on CNN for Weibo Events[A]. Wuhan Zhicheng Times Cultural Development Co., Ltd.Proceedings of the 2nd International Conference on Computer Engineering,Information Science and Internet Technology(CII 2017)[C].Wuhan Zhicheng Times Cultural Development Co., Ltd.,2017:11.
 [6] Zhiquan Wang. Research on Web Text Classification Algorithm Based on Improved CNN and SVM[A]. IEEE Beijing Section、Sichuan Institute of Electronics.Proceedings of 2017 17th IEEE International Conference on Communication Technology (ICCT 2017)[C].IEEE Beijing Section、Sichuan Institute of Electronics.,2017:4.
 [7] Xiaohong W. Gao,Rui Hui,Zengmin Tian. Classification of CT brain images based on deep learning networks[J]. Computer Methods and Programs in Biomedicine,2017,138



a. BP



b.Conv

fig.8 Network Contrast Experiment Diagram

$$F = X_k = \sum_{n=0}^{\frac{N}{2}-1} x_{2n} * e^{-i2\pi km / (N/2)} + e^{-i2\pi k / N} \sum_{m=0}^{\frac{n}{2}-1} x_{2m+1} * e^{-i2\pi km / (N/2)} \quad (8)$$