Optimization of CNN and Its Application in Handwritten Numeral Recognition

Zhi-jia ZHANG, Yuan-yuan LI*, Meng-si JIA, Ling ZHONG and Ya-jie YU
Shenyang University of Technology, Shenyang, Liaoning, China
*Corresponding author

Keyword: Convolutional neural network, Handwritten numeral recognition, Network structure, Parameter optimization.

Abstract. Convolutional neural network (CNN) has been widely applied in the field of handwritten numeral recognition, but the discuss about the parameter of CNN network structure is pretty few. This paper made in-depth analysis of CNN network systematically and make in-depth research on network structure, the number of hidden feature map, the size of convolutional kernel, weight initialization, the number of bulk samples, iteration times and other parameters in recognition process and provide their value choice method. This paper also proposed a method based on CNN analysis and experiment and optimized the parameters of CNN aimed for MNIST handwritten numeral database. In addition, this paper proposed the process and method applied in handwritten numeral parameter optimization based on CNN image recognition, which has a good reference value for the further application of CNN network in the field of image recognition.

Introduction

HNR is applied to mail sorting, taxation, finance, large-scale data statistics and other fields. The significant difference between each handwritten digit makes the recognition accuracy of handwritten digit much lower than the recognition accuracy of printed digit.

Many domestic and foreign scholars made deep research [1-4], and main research methods include: Template matching method, statistical decision method, syntactic analysis method, optical character recognition method and other methods. Compared to traditional methods, neural network has better fault tolerance in pattern recognition and can recognize the input pattern with noise or distortion. In addition, it has better adaptive learning capability and high efficiency. Therefore, neural network is widely applied in pattern recognition.

CNN not only has the advantages of traditional neural network, but also has the characteristics, which are automatic feature extraction and weight sharing. Currently, some researchers have already applied CNN to the study of handwritten digit[5-6].

The network structure design and parameter settings of CNN has important effect on its recognition accuracy, but above mentioned papers have few discuss about the network structure parameter of CNN. This paper made in-depth analysis of CNN systematically and summer up the network structure, the number of hidden feature map, the size of convolutional kernel, weight initialization, the number of bulk samples, iteration times and other parameters in recognition process and analyzed these parameters based on CNN analysis and experiment. Finally, this paper provided CNN parameter optimization process and value choice method applied for HNR.

Convolutional Neural Network Structure and Its Key Parameters

This chapter will introduce network structure of convolutional neural network, network structure parameters and the settings of other related parameters during recognition process.

Network Structure

CNN is a multiple layer supervised learning network based on artificial intelligence network and is widely applied in the field of voice analysis and image recognition. It reduces the complexity of
network model, because its internal structure is similar to the internal structure of biological neural network.

When applied in image processing, each layer of CNN is composed of multiple two-dimensional planes. Each plane is composed of multiple independent neurons. As figure 1 shows, the entire CNN structure includes a input layer, two convolution layer, two sub-sampling layer and a output layer.

![Figure 1. CNN network basic model.](image)

Key Parameters

It can be found that determining the CNN network structure and related parameters is the key problem of recognition by analyzing the recognition process. The problem of determining structure is mainly to determine the hidden layer structure, including the number of network structure layer (Nlayer) and the number of hidden layer characteristics figure (Nfigure). Parameters are the size of convolution kernel (Sck), initialization of weight (Iweight), size of sample batch (Ssb) and the number of iterations (Niter). Nlayer is the sum of the number of convolution layers and sub-sampling except input layer and output layer. Usually convolutional neural network includes input layer, hidden layer (including convolution layer and sub-sampling layer) and output layer.

Nlayer is number of hidden layers.

Nfigure is the number of needed convolution kernel in convolution layer. And the number of feature map in convolution layer, sub-sampling layer and previous convolution layer are the same.

Sck is the size of convolution kernel weight matrix connected to the image of previous layer.

Iweight is the initialization of convolution kernel pixels and the weight matrix between the last two layers. Iweight has influence on training speed and output accuracy. Too wide range of initial value will contribute to the failure of training, therefore the range of initial weight value is between -1 and 1 and weight value will be optimized in the later back-propagation process until it reaches the optimal solution. For reducing training times and improving experiment efficiency, we can divide the data into groups to train.

Ssb is the number of images in each group.

Niter is the training times of all the training image samples and the end of training depends on iteration times.

Parameter Optimization

This paper studied the CNN parameter optimization facing HNR. Experiment samples come from ten types of handwritten digits 0-9 in MNIST database and the number of training images is 5000, the number of testing images is 1000.

The number of input nodes of network depends on the number of pixels of training samples. The training samples used for network training in this paper are 28 × 28, so the number of nodes of input layer is 784. Recognition results are classified into 10 categories, therefore the number of nodes of input layer is 10. This paper made research and experiment form the following aspects and optimized the network structure and parameters.
Sck and Nlayer Optimization

Convolution kernel operator is the weight used in convolution computation and can be represented by a M×N matrix. The size of this matrix is the same as the size of computed image area and it is symmetrical in the most usage. If M = N, the number of rows and columns are both odd and the value of N is commonly 3, 5, 7.

The increase of the number of layers of neural network in a certain range can effectively improve the recognition accuracy. However, if the number of layers is too large, not only the network structure is complicated, but also the recognition accuracy will decrease. The key problem of image recognition is the image feature extraction. A brief description is that use a set of simple data (image description variable) to describe the entire image, therefore the simpler the set of data is, the more representative the set of date is. And the number of layers of network structure is determined by the size of training sample and Sck.

| Sample size | Sck  | C1    | S2   | C3    | S4 | Nlayer |
|-------------|------|-------|------|-------|----|--------|
| 28×28       | 3×3  | 26×26 | 13×13| 11×11 | -  | 3      |
| 28×28       | 5×5  | 24×24 | 12×12| 8×8   | 4×4| 4      |
| 28×28       | 7×7  | 22×22 | 11×11| 5×5   | -  | 3      |

As table 1 shows, the output feature map of samples based on 28×28 image under these three sck are respectively: And the output feature map under 5×5 is simplest, therefore the effect of choosing 5 × 5 convolution kernel is the best. Thus two convolution layers and sub-sampling layers are needed.

Ssb Optimization

For reducing training times and improving training efficiency of samples, data is divided into groups to train. Under these conditions which iteration times is 1, 10, 20, 50, 100, each group is respectively 10 images, 20 images, 40 images and 50 images. And the error rate after testing is as figure 3 shows. Draw the experiment operation time under different Ssb is as figure 4 shows.

From figure 3,4, Ssb is selected by recognition efficiency and precision, so Ssb=40.

Niter Optimization

Due to the too many training images in training process, the experiment in this paper used custom setting Niter and needs to analyze the effect of Niter on the result in order to make sure the efficient operation in the condition of high recognition accuracy.

The experiment result of Niter is as Figure 5 shows.

From figure 5, it can be concluded that the more the Niter, the more convergent the error rate, the higher the recognition accuracy and the longer the responding time; the less the Niter, the lower the recognition accuracy and the shorter the responding time. Therefore, select the Niter, which just enter convergence zone in the condition of relative high efficiency. In this experiment, the error rate of HNR starts convergence when the iteration times is 50, so Niter = 50.

Nfigure Optimization

This paper uses two convolution layers and two sub-sampling layers. The number of feature maps of sub-sampling layer depends on the number of feature maps of upper layers and the number of feature maps of convolution layer depends on the experiment. In order to ensure the integrity and representative of the experiment results, take the number of feature map of first convolution layer as invariants and get the error rate of neural network by changing the number of feature maps of second convolution layer and get the optimal neural network structure.
In the experiment, the range of the number of feature maps in first convolution layer is [1, 10] and range of the number of feature maps in second convolution layer is [1, 40]. When the number of feature map in first convolution layer is 1 and the number of feature map in second convolution layer is from 1 to 40, the optimal combination is obtained and recorded. Successively take the value to take experiments and compare the results to select the combination with lowest error rate. The experiment result is shown as Figure 6.

Experimental results show that handwritten digit recognition is insensitive to network structure. When the number of feature map in first convolution layer is set as 9 and the number of feature map in second convolution layer is 20, the error rate of recognition is lowest, which is 1.2%.

From the experimental results, it can be analyzed that increase the number of feature maps in network can not only increase the number of required learning parameters in network, but also make the network training process cannot reach a relative stable state in the condition of limited number of samples. In addition, it cannot achieve a good recognition accuracy. Therefore, the number of feature map is not the more the better, but need to consider the experiment and select the optimal structure.

Conclusions

This paper used CNN to recognize handwritten digit and made in-depth analysis of network structure parameters of CNN systematically based on MNIST handwritten digit database. In addition, it also summarized the network structure and its main related parameters in recognition process and found the optimal value of these parameters based on analysis and experiment. Study the parameter optimization process and method based on CNN image recognition can get the following conclusions:

1. Nlayer depends on both the size of input image and the size of convolution kernel. The less the network layers, the lower the recognition accuracy; the more the network layers, the higher the recognition accuracy. The computation and output of the size of feature map determines the number of network structure layers.
2. The number of Ssb plays a vital role in the whole system. The smaller the number of Ssb, the more the training batch, the higher the experiment recognition accuracy and the longer the responding time; The larger the number of Ssb, the less the training batch, the lower the experiment recognition accuracy and the shorter the responding time. Combine the recognition accuracy and operation time to select Ssb.

3. Niter determines the time and effect of the whole system network. The less the Niter, the lower the recognition accuracy and the shorter the required responding time; the more the Niter, the higher the recognition accuracy and the longer the required responding time. When the curve enter convergence zone, select the Niter, which just entered convergence zone as the experiment standard.

4. Hidden layer is the most critical factor in determining the recognition accuracy of system and Nfigure plays a vital role in hidden layer. From above figures, it can be concluded that Nfigure is not the more the better, aimlessly increase the number of Nfigure cannot improve the recognition accuracy, but also increase the operation time. Therefore, select the appropriate Nfigure need a lot of experiments. Finally, this paper select the number of feature map of first and second convolution layer with minimum recognition accuracy as Nfigure.

Acknowledgement
This work was financially supported by National Natural Science Foundation of China (61540069).

References
[1] A. Cardoso, A. Wichert, Handwritten digit recognition using biologically inspired features, Neurocomputing. 99(1)575-580.
[2] X.C. Shuang, K. Zhang, Research on handwritten digit recognition based on statistics and structure feature, Computer Engineering and Design. 33(4)1533-1537.
[3] U. Porwal, Z.X. Shi, S. Setlur, Machine Learning in Handwritten Arabic Text Recognition, Handbook of Statistics. 31(2013)443-468.
[4] F.Z. Martinez, V. Frinken, S.E. Boquera, Neural network language models for off-line handwriting recognition, Pattern Recognition. 47(4)1642-1652.
[5] M. Kang, Handwritten digit recognition technology research, Wuhan: Wuhan University of Technology. (2010)
[6] J.C. Fei, T. Rui, Y. Zhou, Multi-input convolution neural network based on gradient, Photoelectricity Engineering. 42(3)33-38.