

Research on Handwriting Recognition Algorithm Based on Convolution Neural Network

Pengfei Hu ^{1, a)}, Hao Tang ^{2, b)}

¹Faculty of Computer, Guangdong University of Technology, Guangzhou, 510006, China.

²Synergy Innovation Institute of GDUT, Heyuan, 517000, China.

^{a)} Corresponding author: 1114306517@qq.com

^{b)} 122260851@qq.com

Abstract. Since Geoffrey Hinton first proposed the concept of deep learning in 2006, which in Reducing the dimensionality of data with neural networks, deep learning has been the continuous concern of the researchers. Deep learning uses a multi-layer neural network to simulate the multi-layer abstract learning process of the human brain. The network structure design and feature extraction are the key problems in the application of data mining and pattern recognition. The accuracy of deep learning for handwritten digital recognition has always been an important criterion for measuring a deep learning algorithm or network structure. In this paper, a handwritten character recognition algorithm based on convolution neural network is studied and compared with other recognition algorithms, and the error rate in their recognition is compared.

Key words: deep learning, convolutional neural network, Handwritten character

INTRODUCTION

Machine learning is a complex discipline that studies how to use machine to conduct human activities. It is a new knowledge acquired by machine, and it is used to identify existing knowledge and optimize the performance standard of computer program. As a branch of machine learning, deep learning is a more complex machine learning algorithm. It has achieved certain results in image recognition and speech recognition. Deep learning is regarded as an advanced and novel technology. Deep learning has made great progress in personalized technology, voice, Natural Language Processing, Machine Translation, multimedia learning, machine learning, data mining, search technology and many other aspects. Through deep learning, machines can imitate human activities, such as thinking activities and audiovisual activities, essentially solves the complex pattern recognition problem, the artificial intelligence technology to a certain extent, the progress of the research and practical application of deep learning organically is an integral part of the work.

Besides, character recognition is the most important part in pattern recognition. Character recognition involves not only pattern recognition, but also logic, image processing, cognition and so on. With the development of technology, character recognition technology is changing rapidly. Recognition rate is also improving. And it has been widely used in writing characters. However, simple character recognition can have a very high recognition rate, complex character recognition has great difficulty, this is because if the character in complex environment, such as the natural environment, so its background is vague, the resolution is relatively low, in the dim scene illumination intensity not enough etc.

DEPTH LEARNING ALGORITHM

Since professor Jeffrey Hilton first proposed the concept of deep learning, the application of deep learning [1] technology in the image is becoming more and more extensive. Deep learning can achieve the approximation of

complex functions by learning a deep nonlinear network structure, representing the distributed expression of input data, and showing the ability to learn the essential characteristics of data from a few sample data sets [2].

Deep Learning Training Process

The BP algorithm is a typical algorithm of traditional training multilayer network, on the effect of model training only containing several layers of the network is not ideal, and the depth of the structure, which involves multiple nonlinear processing unit, exists non-convex objective cost function in the local minimum is a major source of hard training. Therefore, in the depth of the network model, if for all at the same time training, the time complexity is too high; if only the training layer deviation will transfer layer, it will face and the traditional algorithm on the contrary, serious over fitting, because too many parameters and network layers. In 2006, Professor Hinton put forward the effective method to establish a multilayer neural network in unsupervised data, divided into two steps: one is a layer of each training network, the two is the original tuning, said to the senior generation and said the senior said the downward generated as consistent as possible. The specific process is as follows:

- 1) First of all, monolayer neurons are constructed by layer by layer, so each training is a single layer network.
- 2) When all the training has been completed, the algorithm is tuned;

The weight of the other layers on the top layer becomes two-way, so the top layer network is still a single layer neural network, and the other layer becomes a "graph model". The upward weight is used for "cognition", the downward weight is used for "generation", and finally the wake-sleep algorithm is used to adjust all the weights. It is consistent with "cognition" and "generation", that is, to ensure that the most top-level feature of "generation" represents the maximum possible reduction of the underlying input nodes. The Wake-sleep algorithm is divided into two parts: Wake and sleep;

wake: The "cognition" process generates the abstract representation of each layer, namely the node state through the learning characteristics and the upward weight ("cognition" weight), and uses the gradient descent to update the "down" weight between the layers

sleep: The "generation" process, through the top-level representation (time learned features expression) and down weight, generates the underlying node state and updates the upwards weight between the layers.

Therefore, the training process of deep learning network can be simply understood as two parts of training and tuning. Based on the parameters obtained from the first step, the parameters of the fine-tune model are further constructed, which is a supervised training process. This part is like random initialization of the weights of the neural network, and the first step of Deep Learning is through the study of structural characteristics of the input data to initialize the network rather than random initialization, so the initial weight closer to the global optimum, which can get better feature representation, achieve better classification or recognition effect. Therefore, the effect of deep learning is largely attributable to the first step of the characteristic learning process.

Handwritten Character Recognition Based on Convolution Neural Network

As a typical convolutional neural network model, the LeNet-5 model has a very good effect in handwritten character recognition. Next, we introduce the LeNet-5 model in detail.

The transverse pattern has seven layers in all, of which the image of INPUT is 32*32 pixels. The second layer C1, a total of six maps, because from the INPUT layer from the 5*5 size field as input, convolution, specific features can be extracted by the INPUT convolution, the layer feature map size is 28*28, and the parameters of the layer 156 clever 6 of them. The feature map, each including 25 parameters, a parameters bias.

The third layer is the feature extraction layer S2. The role of this layer is to retain some useful information. On the other hand, it can reduce the amount of data, and the selection of feature extraction should be moderate, not large or smaller. If it is too large, you will lose the information you need, and if it is too small, it will not achieve the effect. The layer contains 6 feature maps, because the size of the feature map is reduced by 3/4 compared with the previous one, and only 1/4 (14*14) is left. In addition, there are 12 parameters in the layer, and there are 5880 connections.

The feature extraction layer then goes to the convolution layer C3, which is the same as the previous convolution layer. From the top level, the field of 5*5 is taken as input, and the size of all the features to this layer is 10*10. Due to the increase of convolution kernel, the feature map of this layer has changed from the original 6 to 16. It is worth noting that the features extracted from layer to layer roll process, and not corresponding to each other, so there is no need to introduce a feature extraction layer in a feature graph with a layer of connection. By this method, it can make the feature graph learn different features, and make the network unsymmetrical.

Next, we extract the layer S4 for the next feature. From the graph, we can see that there are 16 feature maps in the layer, but the size of the feature map is changed from the 10*10 on the upper level to 5*5, which is because the cell of the layer relates to the field of the 2*2.

The sixth level is the volume base C5. Through the above figure, there are 120 feature maps in the layer, and the feature map size of this layer is 1*1 the feature extraction layer on the top level is fully connected, and the number of connections in the layer is 48120.

The reciprocal second layer is a fully connected layer F6, which has 86 units based on the output needs, and it is also fully connected with the C5. In addition, the layer needs the point product input vector and the weight vector, then the result is added to the bias bias, and the resulting structure is the input of the excitation function sigmoid.

The last layer is the output layer OUTPUT, and each unit in the output layer represents a class, and the final output is obtained by calculation.

Experimental Results and Analysis

Through in-depth understanding of convolutional neural network and the algorithm of self-programming, completed the experiment, and referring to the related literature, the individual results obtained with other methods of the experiments were carried out to verify the full comparison, obtained the experimental effect.

This experiment is carried out based on 64 computers and 16G memory. Due to the influence of experimental environment, the effects of different configurations of computers may be different.

```
epoch 100/100
Elapsed time is 87.877039 seconds.
1.08% error
```

FIGURE 1. The error rate of the experiment

It can be seen from the above diagram that through multiple experiments, we can use convolution neural network to recognize MINST handwritten character dataset. After 100 iterations, the final experimental error rate is 1.08%.

Next, we use the time of the convolution neural network to make a full understanding of it. As the following table:

TABLE1. The time-consuming time consumed by the convolution neural network in the experiment

Experimental Stage	Forward Propagation	Backpropagation	Gradient Descent	Total Consumption Time
Duration(s)	34.12	72.31	4.33	110.73

TABLE2. A variety of algorithms in MNIST Handwritten Digits character library recognition

Method	Error Rate(%)
Recognition of Convolution Neural Network	1.08
Deep Neural Network under the GPU Platform [3]	0.31
Recognition of KL+BP Algorithm [4]	1.3
Recognition of Improved SVM Algorithm Based on KNN [5]	2.55
Recognition Based on Principal Content Analysis [6]	4.84
Recognition Based on Popular Learning [7]	6.73

As we can see clearly from the previous table, the experiment is divided into three stages, namely, the forward propagation stage, the reverse propagation stage and the gradient descent stage, respectively. In the three stage, the

forward propagation time consuming 34.12s, the gradient descent stage takes 4.33s time, the backpropagation phase takes the longest time, reaches 72.31s, and it is two times longer than the forward propagation stage. The analysis of the reason is mainly because the process of reverse convolution is more complex than the process of convolution, so it takes more time.

The results of this experiment and other algorithms are compared with the results of the MINST Handwritten Digits character library.

CONCLUSION

Through the above comparison, we can see that the convolution neural network has a good effect on the recognition of the MINST Handwritten Digits character library, and the error rate is 1.08%. There are two recognition methods for the error rate lower than the convolution neural network, which are the recognition of the deep neural network and the LeNet-5 model. It is worth mentioning that the above deep neural network algorithm can achieve better recognition effect, but it can be realized on the GPU platform, and the input image needs to be processed ahead of schedule. Therefore, compared to the convolution neural network, the hardware conditions are higher and more complex. The second is the KNN based support vector machine algorithm. The idea of this kind of algorithm is that first we need to observe the KNN value of the sample. By judging it, it determines whether it can be a support vector, but the disadvantage of this kind of algorithm is that it relies more on training samples. In addition, the recognition method based on PCA first needs to standardize the characters, and then uses PCA to extract feature and reduce dimension to make it stable. The closest thing to this experiment is the recognition of KL+BP smart method.

ACKNOWLEDGMENTS

The successful completion of this paper is inseparable from my advisor, Professor Zhang Mei, who has given me many useful guidance and help in his usual scientific research and life. In the course of writing and implementation of the paper, many valuable suggestions were put forward, and her rigorous academic attitude made me benefit all the life.

REFERENCES

1. Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks. *Science*, 2006, 313: 504.
2. Dan C, Meier U, Schmidhuber J. Multi-column deep neural networks for image classification. 2012, 157(10): 3642-3649.
3. Ciresan DC, Meier U, Gambardella LM, Schmidhuber J. Handwritten digit recognition with a committee of deep neural nets on gpus. arXiv preprint arXiv:1103.4487, 2011
4. Bradley DM, Bagnell JA. Differentiable sparse coding. *Advances in Neural Information Processing System*, 2008, 21: 113-120
5. Madzarov G, Gjorgjevikj D. Evaluation of distance measures for multi-class classification in binary SVM decision tree. In: *Artificial Intelligence and Soft Computing: Springer*, 2010. 437-444
6. Almazan J, Valveny E, Fornes A. Deforming the blurred shape model for shape description and recognition. *Pattern Recognition and Image Analysis: Springer*, 2011: 1-8.
7. Sakano H, Ohashi T, Kimura A, Sawada H, Ishiguro K. Extended fisher criterion based on auto-correlation matrix information. *Structural, Syntactic, and Statistical Pattern Recognition: Springer*, 2012: 409-416