NAME: Meith Navlakha
SAP: 60004190068
AI Practical

EXPERIMENT 6

AIM: Case study on Artificial Intelligence applications.

TOPIC: Application of Convolutional Neural Network in Handwritten Chinese Character Recognition

ANALYSIS:

The paper was based on the recognition and classification of the HWDB dataset which was established by the National Laboratory for Pattern Recognition (NLPR) of the Institute of Automation, Chinese Academy of Sciences. Results obtained by the earlier models on HWDB handwritten Chinese character recognition, based on convolutional neural networks differ noticeably at different learning rates; if the learning rate was too low, the CNN's convergence speed slowed, resulting in low training efficiency. This research mainly focused on improving the accuracy by pre-processing of the dataset, selecting an apt activation function, construction and parameterizing the CNN and finally optimizing the model. The research successfully achieved an accuracy of 93% in the recognition and classification of handwritten Chinese characters.

In the HWDB dataset, the handwriting samples were made on paper by 1,020 writers using Anoto pens. The size and the label of the pictures were not uniformly processed which were needed to be pre-processed, including unify the size of the pictures and label the corresponding pictures. The font shape of Chinese characters were square, and the relative size of the original image were distributed between 40–100 pixels, thus unified to a pixel size of 50∗50. Four activation functions: sigmoid, Tanh, ReLU and Mish were selected for data analysis using the CNN model to compare the influence of each activation functions on accuracy. For the Sigmoid activation function, the accuracy rate was very low, around 10%, clearly indicating that it wasn't suitable in that scenario. When compared to Mish and ReLU activation functions, Tanh had a greater overall learning efficiency, could converge to a higher accuracy rate faster, and the final recognition accuracy could be stabilised at around 85%. Although, in the final recognition rate, Mish can be stabilized at about 90% with the fluctuation range is about 2%. Tanh, ReLU, and Mish had produced convincing results but among these three activation functions the recognition rate and stability of the Mish activation function produced the best results.

After selecting an apt activating function, an optimising algorithm was formulated in order compensate with the large fluctuations in the value of loss function around the minimum value, which hinder to reach to the optimal value. Three Loss functions: SmoothL1Loss(), BCEWithLogitsLoss() and MSELoss() were selected to measure the inconsistency between the predicted value and the real value of the model. No significant influence was observed on the final recognition accuracy, but a certain influence was seen on the initial convergence speed.

Finally, through the learning rate tuning, the weights of the CNN model were adjusted. As the learning rate reduced, the accuracy rate had a significant decrease after training 30,000 times. The appropriate range of learning rate determined was between 0.01 ~0.08. With the learning rate being 0.08, the highest accuracy is obtained, which could be stabilized to about 93%.

CONCLUSION:

In this experiment, I performed case study on ‘Application of Convolutional Neural Network in Handwritten Chinese Character Recognition’ research paper published on IEEEE. The paper was based on the recognition and classification of the HWDB dataset which was established by the National Laboratory for Pattern Recognition (NLPR) of the Institute of Automation, Chinese Academy of Sciences. The research mainly focused on improving the accuracy by pre-processing of the dataset by unifying the images to 50*50 pixels, selecting an apt activation function, construction and parameterizing CNN layers. and finally optimizing the model. The research successfully achieved an accuracy of 93% in the recognition and classification of handwritten Chinese characters using learning rate 0.08 along with Mish activation function.

2020 IEEE International Conference on Information Technology,Big Data and Artificial Intelligence (ICIBA 2020)

Application of Convolutional Neural Network in Handwritten Chinese Character Recognition

Hongliang Guo[1],Likun Ai[2],Shuxin Chen[1,2*]
1. Department of Computer Science and Technology, Renai College of Tianjin University, Tianjin, China
2. College of Mechanical and Electrical Engineering, Qiqihar University, Qiqihar, China
304473675@qq.com, ailikun034@163.com, shuxinfriend@126.com
Corresponding Author: Shuxin Chen   Email:shuxinfriend@126.com

***Abstract*—The handwritten Chinese character data set HWDB, designs and trains a convolutional neural network to recognize handwritten Chinese characters. On this basis, this article tested and studied the effects of activation function, loss function, and learning rate on the performance of convolutional neural networks, and further analyzed the reasons for their effects, and accumulated for the subsequent application of convolutional neural networks in other aspects. The valuable experimental experience and theoretical basis were laid. This paper is based on the relevant technical process, analyzed the experimental effect, summarized the influence of important parameters, and summarized the design ideas of neural network for different image types, laying a foundation for further research in the future.**

***Keywords*—deep learning; convolutional neural network; handwritten Chinese character recognition; learning rate**

I. Introduction

Deep learning is an important research direction in the field of computer science. By designing artificial neural networks about the internal logic of sample data, and then achieve the effect of predicting the same type of data, such as text recognition, image classification, and speech recognition. The data field is in the stage of rapid development of the Internet and computer performance, and various types of data are also growing rapidly[1]. The mass data of text, pictures, and sound have also been sorted out more reasonably, which also provided a necessary for the further development of deep learning. These data can mine effective information.

Artificial intelligence emerged in the 1950s, has experienced three major developments, which is now in the third stage of development[2]. In the past five years, related applications of deep learning have become more and more closely related to our lives. For example, specific applications for image feature learning include OCR recognition, face recognition, and gesture recognition. Especially the auto-driving technology that is being vigorously developed in the automotive-related field mainly relies on the analysis of the characteristics of road conditions. HWDB handwritten Chinese character recognition based on convolutional neural network has obvious differences in the results produced at different learning rates, if the learning rate is too small, the convergence speed of the convolutional neural network will slow down, resulting in low training efficiency. It can be expected that the development of this technology will radiate to all aspects of society in the future.

II. The Application of CNN in Handwritten Chinese Character Recognition

The HWDB dataset was established by the National Laboratory for Pattern Recognition (NLPR) of the Institute of Automation, Chinese Academy of Sciences. The handwriting samples were made on paper by 1,020 writers using Anoto pens. The dataset used in this experiment with 10 sets of data in the HWDB dataset. The PNG image of the dataset are obtained through the open source dataset, as shown in Fig.1.

Fig. 1. HWDB raw data picture

A. HWDB image Dataset Preprocessing

By observing the above Fig.1, we can find the original data set, the size and label of the pictures are not uniformly processed. Therefore, if we want to use the dataset, we need to perform certain preprocessing. Including unify the size of the pictures and label the corresponding pictures.

Fig. 2. Pre-processed HWDB image data

978-1-7281-5224-0/20/$31.00 ©2020 IEEE

Taking into account that the font shape of Chinese characters is square, and the relative size of the original image is distributed between 40-100 pixels, here it is unified to a pixel size of 50*50. The labeling method is shown in Fig.2.

B. Importance of experimental results

(1) Sigmoid

The function image of sigmoid is shown in Fig.3. Its function is expressed as formula (1):

$$f(x) = \frac{1}{1+e^{-x}}) \tag{1}$$

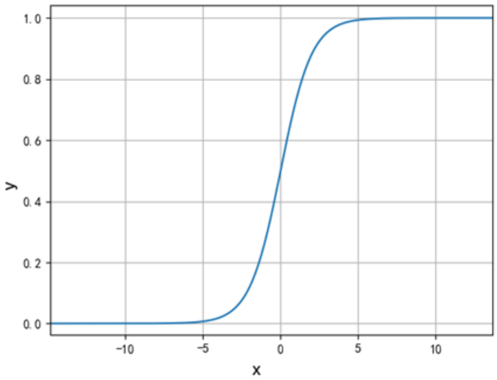

Fig. 3. Sigmoid function image

(2) Tanh

The Tanh function is expressed as formula (2):

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{2}$$

(3) ReLU

The ReLU function is expressed as formula (3):

$$f(x) = max(0, x) \tag{3}$$

(4) Mish

The Mish function is expressed as formula (4):

$$f(x) = x * tanh(ln(1 + e^x)) \tag{4}$$

The boundlessness of the Mish activation function avoids the saturation due to capping. In theory, negative values are slightly allowed instead of hard zero boundaries like in ReLU. At the same time, the smooth activation function allows more information to penetrate the neural network, resulting in better accuracy and generalization ability. The function image of Mish is shown in Fig. 4.

C. Importance of Sampling Construction Convolutional Neural Network Model

Based on the experience summarized in the previous design of the MNIST neural network, the structure of the convolutional neural network is designed as:

(1) Input layer: Process the HWDB data set in advance to obtain grayscale data with the size of 50*50*1;

(2) Convolutional layer 1: Use 6 layer(5*5) convolution kernels to perform convolution operation on the image of the input layer, with the moving step of 1;

(3) Pooling layer 1: Dimensionality reduction is performed on the output features of convolutional layer 1 through the maximum pooling operation. The size of the pooling window is 2*2, and the moving step of the pooling operation is 2.

(4) Convolutional layer 2: Use 16 layer (3*3) convolution kernels to perform convolution operations on the image, with a moving step of 1.

(5) Pooling layer 2: Same pooling layer 1.

(6) Fully connected layer: There are 200 nodes in total to connect the data passing through the pooling layer 2.

(7) Output layer: 10 nodes, corresponding to the classification of the picture, in order to achieve data normalization, the Mish activation function is used the dataset and parameters of the training process:

I. Number of training sets: 1800 sheets;

II. The number of test sets: 200 sheets;

III. Training times: 30000 times;

IV. Every 100 times of training, use the test set to count the change of accuracy rate.

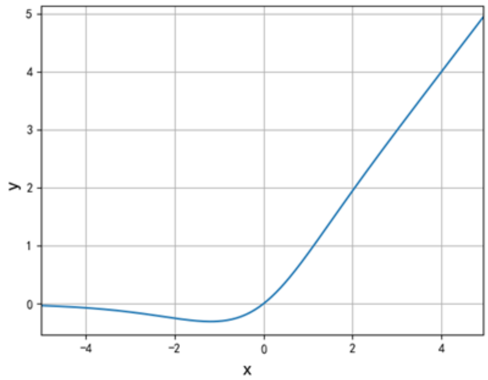

Fig. 4. Mish activation function image

D. Experimental results using different activation functions

The above four activation functions sigmoid, Tanh, ReLU and Mish were selected for data analysis. The use of convolutional neural network HWDB for handwritten Chinese character recognition is shown in Fig. 5, which compared the influence of four activation functions on accuracy. The X-axis value is the number of training times, and the Y-axis value is the accuracy. Conclusion of the experiment:

(1) For the Sigmoid activation function, the accuracy rate is basically 10%. Compared with other activation functions, it can be clearly found that the sigmoid activation function is not suitable for image classification under the convolutional neural network in this problem scenario.

(2) For the Tanh activation function, its overall learning efficiency is higher. Compared with other Mish and ReLU

activation functions, it can converge to a higher accuracy rate faster, and the final recognition accuracy can be stabilized at about 85%. Compared with the other two, the accuracy rate is slightly lower, and the fluctuation range is about 3%, which is relatively stable.

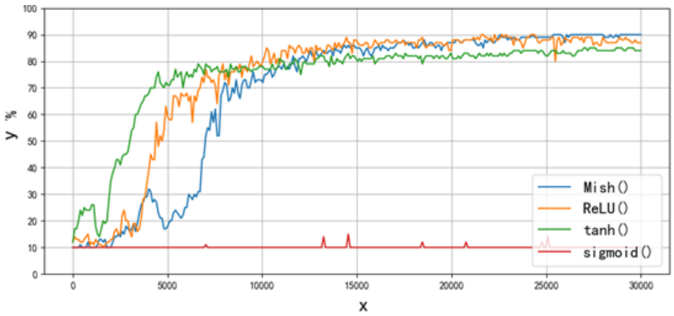

Fig. 5. Comparison of effects of different activation functions

(3) For ReLU activation function and Mish activation function: In terms of convergence speed, ReLU is faster; in the final recognition rate, Mish can be stabilized at about 90%, the fluctuation range is about 2%, and the accuracy of ReLU can be basically stabilized at About 87%, but its fluctuation range is about 5%.

(4) For the three activation functions of Tanh, ReLU, and Mish, the recognition of handwritten Chinese characters can be successfully realized. Among the three, the recognition rate and stability of the Mish activation function are the best, but for The sigmoid activation function has a final accuracy rate of only 10%, which is a training failure.

III. EXPERIMENT OPTIMIZATION

Optimization algorithm is almost the most necessary algorithm for machine learning model training. In the early stage of algorithm optimization, learning is accelerated, making the model more accessible to local or global optimal solutions. However, in the later period, there will be a large fluctuation, and even the value of loss function will fluctuate around the minimum value, which makes it difficult to reach the optimal value. In the preceding chapter, the difference between model distribution and target distribution are calculated by samples. However, in practice, the label of an event is known to be invariant, that is, the entropy of target distribution is constant. Therefore, instead of calculating KL divergence, the loss values of model distribution and target distribution can be obtained by calculating cross entropy. It is known that the difference between model distribution and target distribution can be replaced by cross entropy under the condition that the target distribution is constant. If the target distribution changes, then cross entropy cannot be used.

(1) Mean square error loss function

In PyTorch, the mean square error loss function is the MSELoss loss function, which is expressed as formula (5):

$$MSE(x,y) = \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n} \tag{5}$$

(2) Huber loss function

In PyTorch, it is the SmoothL1Loss loss function. The function is expressed as formula (6):

$$Huber(x_i, y_j) = \begin{cases} 0.5(x_i - y_i)^2, & if \quad |x_i - y_i| < 1 \\ |x_i - y_i| - 0.5, & otherwise \end{cases} \tag{6}$$

(3) Cross entropy loss function

For the cross-entropy loss function used for classification problems, the sigmoid layer is added, which can make the numerical effect more stable. The corresponding specific function is named BCEWithLogitsLoss, and the function is expressed as formula (7):

$$loss(x_i, y_j) = -w_i[y_i * log x_i + (1 - y_i) * log(1 - x_i)] \tag{7}$$

Among them, x is the actual output, y is the label of the sample, and w is the weight of the item.

A. Experimental Results Using Different Loss Functions

The above three common loss functions, SmoothL1Loss(), BCEWithLogitsLoss() and MSELoss(), are selected. The loss function is used to measure the inconsistency between the predicted value and the real value of the model, so as to measure the prediction quality of the model. The convolutional neural network is used for handwriting Chinese character recognition as shown in Fig. 6, the feedback influence of different loss functions on accuracy. The X-axis value is the number of training times, and the Y-axis value is the accuracy.

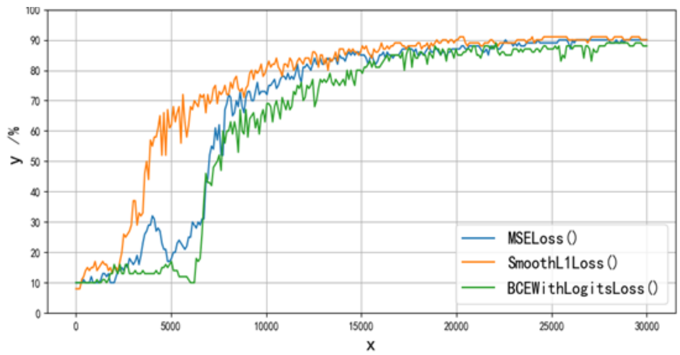

Fig. 6. Effect comparison diagram of different loss functions.The convolutional neural network structure and dataset default parameters;Activation function: Mish; Loss function: Huber loss function.

According to the analysis of the change curve of accuracy in Fig. 6, the three loss functions have no significant influence on the final recognition accuracy, but have a certain influence on the initial convergence speed. On the whole, it can be seen better than the effect of using Huber loss function.

B. Experimental Results Using Different Learning Rate

The learning rate is a super parameter, which controls the degree of adjusting the network weight. The lower the value, the slower it will descend along the gradient, so as to ensure that we will not miss any local lowest point, but it will take a long time to converge. 0.1 in the optimizer is the learning rate for the training parameter. The learning rate is big shock, which does not converge, the learning rate is small convergence speed is slow. Experimental data show the influence of learning rate on the accuracy of handwritten Chinese character recognition using convolutional neural network. The X-axis is the number of training times, and the Y-axis is the accuracy. As FIG. 7 shows the influence of learning rate on the recognition accuracy within the range of 0.1~0.01, and FIG. 8 shows the influence of learning rate within the range of 0.01~0.001. These two groups of comparative experiments comprehensively cover the effect on learning rate of the recognition accuracy about the range of 10^{-2} and 10^{-3}, and also verify the adverse effect whether high or low learning rate on the training results.

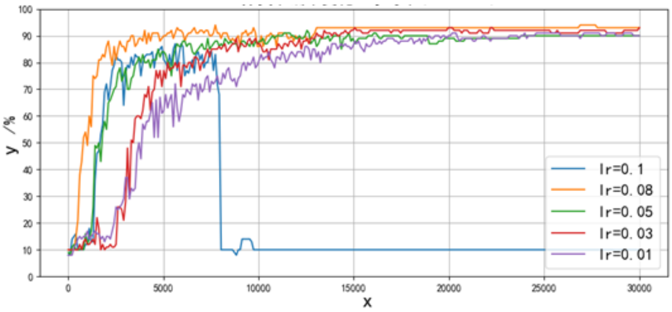

Fig. 7. Effect comparison diagram of the learning rate (0.1~0.01) on accuracy

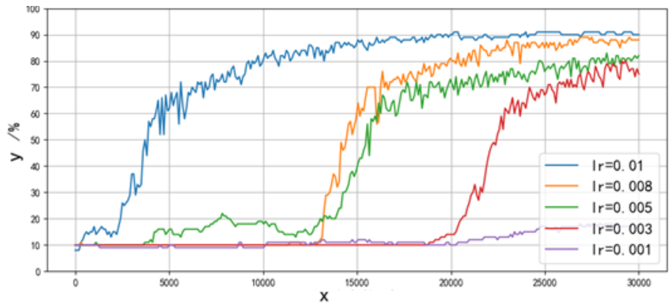

Fig. 8. Effect comparison diagram of the learning rate (0.01~0.001) on accuracy

Through the analysis of Fig. 8, it is concluded that as the learning rate shrinks, the main improvement phase of the learning rate during the training process is significantly shifted back, and the accuracy rate also has a significant decrease after training 30,000 times. According to the theory, the learning rate is too small, the convergence speed of the convolutional neural network will slower down, resulting in low training efficiency.

In summary, in this experiment, it is more appropriate to set the learning rate between 0.01~0.08. A neural network with relatively stable performance and better generalization ability can be trained within a reasonable number of training times. While the learning rate is 0.08, the accuracy is the highest, which can be stabilized at about 93%.

IV. ALGORITHM SUMMARY

Through the study of deep learning, this paper realizes the recognition and classification of the HWDB dataset summarized the experimental process of the former, and successfully achieved 93% accuracy in the recognition and classification of handwritten Chinese characters.

(1) Through exploring the development process of deep learning and major technological breakthroughs at various stages, we have a certain understanding of deep learning and the range of problems.

(2) By summarizing the experimental process of the former, we successfully realized the training and recognition and classification of handwritten Chinese characters by convolutional neural network. This process includes data preprocessing, convolutional neural network construction, parameter comparison and debugging, and finally realized 93% accuracy rate.

Through the research and practice of image feature learning, it is found that convolutional neural networks have better results in processing image classification problems. Compared with other neural networks, convolutional neural networks can more extract image features. And the feature extraction is not defined by humans, but through massive data, the neural network learns independently through the error back propagation algorithm, this learning strategy is directly determined, and the algorithm can be applied to other practical problems, such as food energy analysis, garbage classification, item identification. But it is not without its drawbacks. how to use these data to mine effective information, this has also become a research direction with great potential value. The content that still needs to be studied in the future.

ACKNOWLEDGMENT

This work is supported by the National Science Foundation for Young Scientists of China under Grant No. 11800313 and University-Industry Collaborative Education Program under Grant No. 201802252028.

REFERENCES

[1] B. Du, M. Zhang, L. Zhang, et al. " Pltd: Patch-based low-rank tensor decomposition for hyper-spectral images. " IEEE Transactions on Multimedia, vol.19, issue 1, pp. 67-79, 2016.

[2] L.Y. Zhang, and B.R. Tao, "A framework for ontology integration based on genetic algorithm." Journal of Intelligent & Fuzzy Systems 30, pp.1651, 2016.

[3] Q. Zhang, L. T. Yang, Z. Chen, et al. "High-order possibilistic c-means algorithms based on tensor decompositions for big data in iot. " Information Fusion, vol.39, pp. 72-80, 2018.

[4] L.Y. Zhang, J.D. Ren and Xianwei Li, "OIM-SM: A method for ontology integration based on semantic mapping", Journal of Intelligent & Fuzzy Systems vol.32, pp. 1983–1995, 2017.

[5] S.X. Chen, W.M. Sun, X. Kong, "Clustering analysis of line indices for LAMOST spectra with AstroStat." Research in Astronomy and Astrophysics, vol.18, issue 6, pp.78, 2018.