Research on the Captcha Recognition Method Based On Neural Network

Zhifeng Liu\textsuperscript{1,2, a}, Xiaohu Tang\textsuperscript{1,2, b} and Zhenhua Wei\textsuperscript{1,2, c}

\textsuperscript{1}Jiangxi Engineering Technology Research Center of Nuclear Geoscience Data Science and System, East China University of Technology, Nanchang 330013, China.
\textsuperscript{2}East China University of Technology, Nanchang 330013, China.
\textsuperscript{a}Corresponding author: zfliu@ecit.cn, \textsuperscript{b}408254510@qq.com, \textsuperscript{c}zhhwei@ecit.cn

Abstract. In this paper, the CAPTCHA recognition is studied by using neural network algorithm. Because the traditional BP neural network itself has problems such as uncertain structural parameters, low convergence rate and easy to fall into local minimum, this paper identifies the words in the image according to the pixel value of the verification code. In this paper, the image is divided into individual letters, and then the identified letters are spelled into words. Experimental results show that the accuracy of individual letters was higher, but the accuracy of letters identified from each small image was much lower. The accuracy can be improved by adding dictionaries to the prediction function to find the most matching words.

1. Introduction

In the process of image recognition, there will be a large number of information operations, which has high requirements on processing speed and recognition accuracy, and neural network has real-time features and high precision, which can guarantee the smooth completion of image recognition. BP neural network algorithm can be introduced to improve the algorithm in terms of positioning and recognition of rotating distorted images, which can add additional momentum terms, guarantee the improvement of learning rate and make the network sink into local minimum points [1].

CAPTCHA stands for Completely Automated Public Turing test to tell Computers and Humans Apart. Captcha is a public, fully automatic program that differentiates whether a user is a computer or a person. This paper identifies the words in the image according to the pixel value of the verification code, transforms the difficult problem of identifying words into the small problem of identifying letters, creates a neural network to identify the letters in the image, spells the identified letters into words, and then uses the dictionary to find the most matched words and improve the recognition accuracy.

2. The principle of neural network

Neural Network is a research hotspot in the field of artificial intelligence since 1980s. It abstracts the neural network of human brain from the perspective of information processing, establishes some simple model, and forms different networks according to different connection modes. A neural network is an operational model consisting of a large number of nodes, or neurons, connected to each other. Each node represents a particular output function, called the activation function. The connection...
between each node represents a weighted value for the signal passing through the connection, called
the weight, which is equivalent to the memory of the artificial neural network. The output of the
network is different according to the connection of the network, the weight value and the excitation
function. The network itself is usually an approximation of some algorithm or function in nature, or an
expression of a logical strategy.

In recent ten years, the research work of artificial neural network, has made great progress, the
pattern recognition, intelligent robot, automatic control, forecast estimates, biology, medicine,
economy and other fields has been successfully solved many modern computer is difficult to solve the
practical problems [2, 3, 4], showed good intelligence features.

Neural network is a kind of machine learning algorithm, which can realize infinite approximation
through structural network to solve nonlinear problems. It is widely used in image processing, face
recognition and other fields [5, 6]. The structure of BP neural network is shown in figure 2. It can be
seen that it mainly includes input layer, output layer and hidden layer [7].

![Figure 1. Structure of Neural Network](image)

Define the input of BP neural network as \( (x_1, x_2, \cdots, x_m) \), output as \( (o_1, o_2, \cdots, o_n) \). The number of
hidden layer neuron nodes is determined by the number of inputs \( m \) and the number of outputs \( n \). In
general, Sigmoid function can be used as the hidden layer excitation function, and the specific form is
as follows:

\[
f(x) = \frac{1}{1 + \exp(-x)}
\]

(1)

For the feedback function of BP neural network, the feedback error can be expressed as the sum of
squares of the difference between the expected output value of the output layer and the actual output
value, and the specific form is:

\[
E = \sum_{i=1}^{n} (y_i - o_i)^2
\]

(2)

3. Solution

3.1. Create Dataset

This article only uses 4-letter English words as CAPTCHA.

The goal of this article is to write a program to restore the words in the image
(1) Divide the large image into four smaller ones containing only one letter.
(2) Sort each letter.
(3) Rearrange the letters into words.
(4) Correct word recognition errors with a dictionary.
This paper makes the following assumptions for the verification code recognition algorithm. First, the word in the captcha is a complete, valid English word with a length of four letters. Second, all letters of a word are capitalized, without symbols, Numbers, or Spaces. To increase the difficulty, use different shear effects on words when generating images.

3.1.1. Draw CAPTCHA. Next, write the function to create the verification code. The goal is to draw an image that contains the word. Generate a captcha image and display it.

Generated image:

![Image of HELP in a captcha style](image)

Figure 2. Captcha

3.1.2. Divide The Image Into Individual Letters. Although the captchas are words, this article is not intended to construct a classifier capable of identifying thousands of words, but to turn the big problem into a smaller one: identifying letters.

Result:

![Images of individual letters in a captcha style](image)

Figure 3. Captcha

3.2. Train Dataset

3.2.1. Single Letter Recognition, Next, construct the neural network classifier, receive the image, and predict the (single) letter in the image.

This article uses the single letter training set created earlier. The data set itself is simple. Each image is 20 pixels in size, and each pixel is represented by 1 (black) or 0 (white). Each image has 400 features and will be used as input to the neural network. The output is 26 values between 0 and 1. The larger the value, the more likely it is that the letter in the image corresponds to the letter (the first value of the output corresponds to the letter A, the second to the letter B, and so on).

Next, create a neural network. In this paper, a basic three-layer neural network is created, which consists of input layer, output layer and hidden layer. The number of neurons in the input layer and output layer is fixed. The dataset has 400 features, so the first layer needs to have 400 neurons, and 26 possible categories indicate that we need 26 neurons for output.

Determining the number of neurons in the hidden layer can be difficult. If the number of neurons is too large and the neural network appears sparse, it is difficult to train enough neurons to properly represent the data, which often leads to the problem of overfitting the training data. On the contrary, if the number of neurons is too small, each contributes too much to the classification results, and the
training is inadequate, the phenomenon of low fitting is likely to occur. The hidden layer in this paper uses 100 neurons.

The back propagation algorithm is used to train the neural network. Starting from the output layer, layer by layer goes back to the input layer, looking for the neurons that predicted the error, fine-tuning the weight of the input value of these neurons to achieve the purpose of repairing the output error [8].

Result: The accuracy rate for identifying individual letters was 96%.

3.2.2. Identify The Word. Identify the letters in each small image, and then put them into words to complete the verification code recognition.

Import words from corpus, create word data set, increase the filter condition with length 4, identify all words with length 4, and count the number of correct and wrong words.

Table 1. Result

|                    | Amount | Proportion |
|--------------------|--------|------------|
| Correct recognition| 3650   | 66.2%      |
| Wrong recognition  | 1863   | 33.8%      |
| Total              | 5513   | 100%       |

The accuracy is 66.2%.

3.3. Use a Dictionary to Improve Accuracy

We can actually check to see if the dictionary contains the entry before returning the prediction. If the word is in the dictionary, we return the predicted result, if not, we find a word similar to the predicted result and return it as an updated predicted result.

Write a distance measurement function to count the number of letters in the same position. The word length (4) minus the number of identical letters in the same position, the smaller the value, the higher the similarity between the two words.

Write a new prediction function and pass in the word list. Use the distance measurement function to calculate the distance between the predicted results and each word in the dictionary. Find the most matched word -- that is, the word with the smallest distance -- and return that word as a prediction.

Table 2. Result

|                    | Amount | Proportion |
|--------------------|--------|------------|
| Correct recognition| 3784   | 68.6%      |
| Wrong recognition  | 1729   | 31.4%      |
| Total              | 5513   | 100%       |

The accuracy is 68.6%.

4. Conclusion

In this paper, words in the image are identified according to the pixel value of the verification code image, the problem of word recognition is transformed into a small problem of letter recognition, and a neural network is created to identify letters in the image, with an accuracy rate of 96%. Despite the high accuracy rate, the accuracy rate dropped to 66.2 percent. Using dictionaries to find the best words can improve accuracy. In this paper, the commonly used edit distance is considered to represent the similarity between words. However, this article only focuses on letter errors and does not delve into which editing steps (insert, delete) are relevant, thus simplifying the distance calculation method. The final accuracy improved to 68.6%.
Acknowledgments

This work is supported by National Natural Science Fund (No. 11765001), Open Fund of Jiangxi Engineering Technology Research Center of Nuclear Geoscience Data Science and System (East China University of Technology) (No.JETRCNGDSS201804) and China Scholarship Council funding (201600880001).

References

[1] Xu Qintao, Yang Ziheng. Discussion on image recognition technology and method based on neural network. Communication design and application. 2018.
[2] Wan Xiaoqi, Song Hui. Application of Convolutional Neural Networks in Pattern Recognition of Partial Discharge Image. Power System Technology. 2018.
[3] Zhang Fan, Wang Guoqing. Research on Intelligent Robot Based on Neural Network Technology in the Background of WSN. Journal of Anhui Vocational College of Electronics & Information Technology, 2018, 17: 5-10.
[4] Yuan Zhu, Shen Yige, "Automatic Control of Trajectory Optimization for Agricultural Robot," Agricultural mechanization research, 2017 (6).
[5] Xu Shaokai, Chen Yin, "A License Plate Recognition Method Based on Improved Convolutional Neural Network," Software Engineering, 2018 (21).
[6] Lu Hong, "Face Recognition Method Base on Convolution Neural Network," Modern Information Technology, 2018 (2).
[7] Jin Hongjiao, "Image Recognition Algorithm Based on Improved BP Neural Network," Bulletin of Science and Technology, 2018 (34).
[8] Zheng Chun, Zhang Jishan. Application of Genetic Algorithm in Computer Network Optimization Design. Natural Sciences Journal of Harbin Normal University, 2018, 34: 41-45.