

Comparison and Analysis of Chinese and English Handwriting Recognition Algorithms in Deep Learning

Xinyue Qu*

School of Maritime Economics and Management, Dalian Maritime University Dalian, Dalian, Liaoning, 116026, China

*Corresponding author: qxinyue@ldy.edu.rs

Abstract. Handwriting recognition is an important topic of current research on image recognition. This technology provides convenience for production and life, as well as guarantees for us in some aspects of safety. The author hopes to help people get a preliminary understanding of English handwriting recognition by summarizing and comparing two commonly used models in deep learning. This paper focuses on two models of widely used English handwriting recognition, which are the simple Recurrent Neural Network (RNN) model and the Long Short-Term Memory (LSTM) model. Firstly, the principles of the two models are introduced, and then gradient problems, fitting problems, training difficulty, and accuracy, as well as the reasons for these problems are all discussed and compared within the two models. Nowadays, the accuracy rate of English handwriting recognition is relatively high using deep learning methods and models. With the development of deep learning models, if more accurate string segmentation can be achieved, the accuracy rate of handwriting recognition can also be further improved..

Keywords: Handwriting recognition, deep learning, neural network, Simple RNN, LSTM.

1. Introduction

The innovation of handwriting recognition is a technology that people employ regularly in many aspects of their daily lives. Software for taking notes, such as Goodnotes and Notability, has a handwriting recognition feature that greatly boosts convenience. Additionally, handwriting recognition can be utilized to increase security, accuracy, and efficiency in problems like validating identities in a variety of industries. The field of machine learning is developing at a rapid pace, which is helping to advance handwriting identification. As for today, deep learning, a subfield of machine learning, provides several techniques for recognizing English handwriting, such as Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM). Other techniques in the field of Machine Learning, including K-NearestNeighbor (KNN), have also gone through a protracted development phase in this area. With the use of machine learning techniques, handwriting recognition accuracy may now exceed 96% [1].

This article will compare and analyze various deep learning techniques in terms of their underlying concepts, benefits, and drawbacks when it comes to the recognition of English characters and entire sentences.

By summarizing the two basic methods of English handwriting recognition, this paper hopes to help people understand the basic process and simple method principle of English handwriting recognition, to promote the development of English handwriting recognition.

2. Convolutional Recurrent Neural Network (CRNN)

2.1. Principles of CRNN in Handwriting Recognition

First, text image characteristics are extracted using Convolutional Neural Network (CNN)'s powerful nonlinear mapping and feature-expressing abilities. Next, handwritten image recognition is achieved through the application of recurrent neural networks (RNN) and transcription techniques [1].



Fig. 1 The basic structure of the CRNN model [1]

2.2. CNN

The input layer of CNN de-averages the input content.

The convolutional layer uses a filter, also known as the feature identifiers, with the same depth as the input content to perform convolution in the unit of receptive field. After being processed by the convolutional layer, the size of the characteristic pattern would be calculated as shown in the expression.

$$Y = \left(\frac{h-r}{s} + 1 \right) \bullet \left(\frac{w-r}{s} + 1 \right) \tag{1}$$

Assume that the size of the filter is r times r, plus the color dimension. In this expression, the terms ‘h’ and ‘w’ in this formula denote the image's original length and width, respectively.

The relu function acts as the layer's activation function, and the relu layer is typically placed after the convolutional layer. Because the activation map from the preceding convolution layer typically serves as the input for the subsequent convolution layer, it can identify more intricate characteristics.

The pooling function of the pooling layer is usually to take the maximum or average, reduce the amount of data, extract important features, and reduce the computational complexity. The frequency of the pooling layer is determined due to the actual situation.

The FC layer observes the activation mapping output of the high-level features of the previous layer and assigns different weights to different features to get the correct probabilities for different classifications.

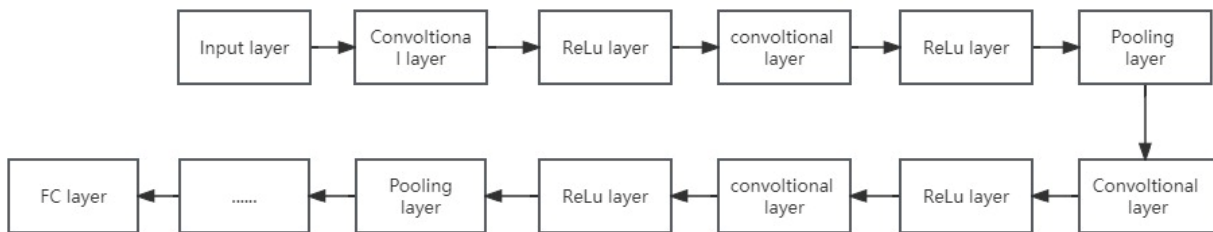


Fig. 2 The structure of CNN [2]

2.3. Simple RNN

CNN does not show good performance when dealing with sequence problems, so RNN needs to be imported to deal with strong correlation features like time series and text processing [1].

The results of the FC layer of the CNN model are passed into the RNN model as input [2].

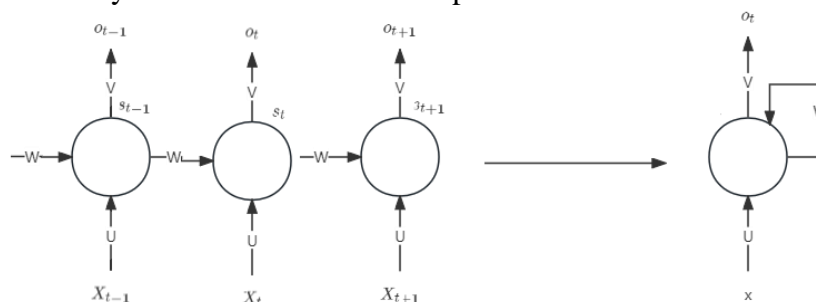


Fig. 3 The working mechanism of the RNN model [3]

The weight vectors in this image are denoted by U for the input layer to the hidden layer, W for the output layer to hidden layer, and V for the hidden layer to hidden layer. It's also important to note that to provide memory capacity, the input layer's variables must be sent to the neurons of the hidden layers along with state variables from the neuron preceding it.

In theory, RNN models have infinite memory, which means they can look back indefinitely. But in practice, it can only look back at the last few steps due to gradient disappearance.

3. Principles of LSTM in Handwriting Recognition

It's still necessary to first use CNN for feature extraction and expression, after that LSTM will be used to record the information from the text sequence and to recognize the handwriting [4].



Fig. 4 The structure of the handwriting recognition model using LSTM (Photo/Picture credit: Original)

Compared with Simple RNN, the LSTM network model introduces three gate control units, which are the input gate, forget gate, and output gate, to realize the selective memory of information.

The filtering proportion of information is obtained by the three gates using the sigmoid activation function, and the tanh activation function is related to the LSTM model's cell state and hidden state.

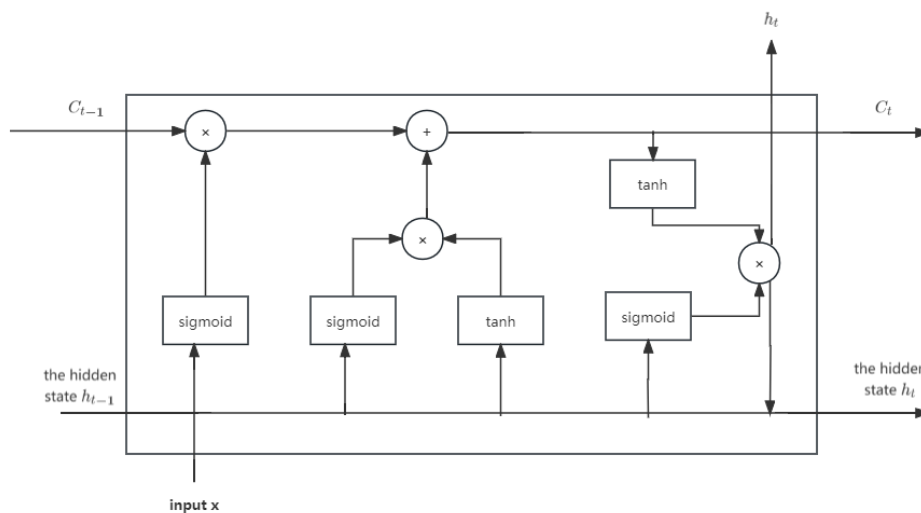


Fig. 5 The structure of the LSTM model [5]

4. Comparison of Simple RNN and LSTM

4.1. Gradient Problems

In Simple RNN models, gradient problems (Gradient disappearance) often appear. In RNN, the so-called "gradient disappearance" actually refers to the near-distance gradient taking over the gradient, making it challenging for the model to acquire a long-range dependence. Due to the setting of the three gate control units, the gradient disappearance problem rarely occurs in the LSTM model.

4.2. Principles of Gradient Problems

Deep networks are stacked with many nonlinear layers. In the formula shown below, f stands for the activation function, f_i represents the value of the neurons in layer i , w_i and b represents the weight vector and bias of the neurons in different layers, respectively.

$$f_{i+1} = f(f_i * w_i + b) \tag{2}$$

The Back Propagation method utilizes a gradient descent strategy and adjusts parameters in the adverse gradient direction as depicted in the formula below.

$$w = w + \Delta w$$

$$\Delta w = -\alpha * \frac{\partial Loss}{\partial w}$$

$$\Delta w_i = \frac{\partial Loss}{\partial w_i} = \frac{\partial Loss}{\partial f_n} \frac{\partial f_n}{\partial f_{n-1}} \frac{\partial f_{n-1}}{\partial f_{n-2}} \dots \frac{\partial f_3}{\partial f_2} \frac{\partial f_2}{\partial w_i} \tag{3}$$

In formula 3, $\frac{\partial f_n}{\partial f_{n-1}}$ equals the derivative of the activation function, so if the derivative is greater than one, there will be a gradient explosion as the number of layers increases, and if the derivative is less than one, there will be a gradient disappearance.

4.3. Solutions

Bidirectional RNN can be used to solve the gradient disappearance problem of simple RNN models to a certain extent. Bidirectional RNNs can read text sequences in both directions, and it's also worth mentioning that the gradient is also determined by both the front and back in this way.

To some extent, the problem can be resolved by using Skip Connection, which is also referred to as Residual Connection [6]. The fundamental concept of Residual Connection is to overcome the gradient disappearance problem and enhance model performance in some network layers by directly connecting the input of certain layers with the output of the subsequent layers[7].

The problem of gradient disappearance can be partially resolved by utilizing the ReLU activation function instead of the tanh activation function. However, relu activation function has obvious drawbacks, when the input value of the activation function is negative, there will be neuron death. In addition, the asymmetry of the output value may affect the well operation of the neural network.

The GRU model is a simple version of the LSTM model and an improved version of the RNN model. It uses the reset gate to replace the forget gate and input gate in the LSTM model. It can also be used to solve gradient problems.

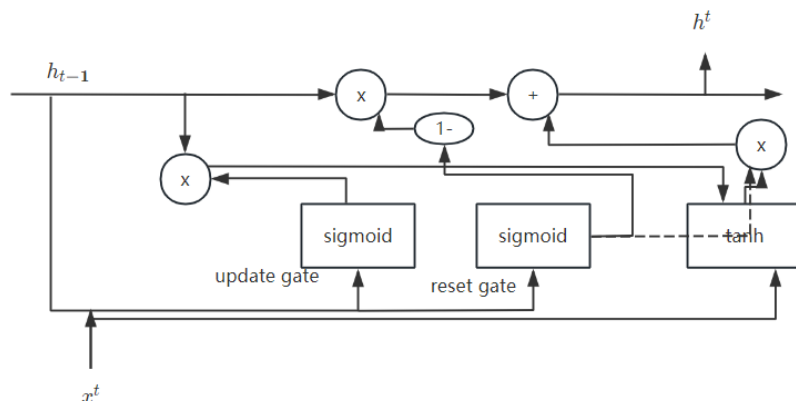


Fig. 6 The structure of the GRU model [8]

4.4. Fitting Problems

Due to the small number of parameters in the simple RNN model, an over-fitting problem is more likely to occur in the case of insufficient data compared with other models like LSTM.

Causes and solutions to fitting problems

The inability of the model to accurately fit the training set and identify the key characteristics and correlations within it is referred to as under-fitting [9]. Three main factors can cause under-fitting. Firstly, the data volume is too small, which makes it nearly impossible for the model to obtain the information it needs; secondly, the model's complexity is insufficient; therefore, either the number of layers or neurons in the neural network must be increased, or a more complex model (such as a Polynomial Regression model in comparison to a simple Linear Regression model) must be used—last but not least, feature misselection. When feature misselection happens, it would be necessary to introduce more features or summarize new features from existing features.

When a model performs poorly on unseen data but too well on training data, it's called overfitting [10]. K-folds Cross-validation is an effective way to solve over-fitting problems. In addition, regularization and pruning of tree models are also effective methods to solve over-fitting problems. If the problem is caused by insufficient data, then it would be necessary to collect more data to increase the amount of data.

4.5. Training Difficulty Level and Accuracy

The RNN model has a low training difficulty and time because of its limited number of parameters that need to be taught. However, when compared with more intricate models like LSTM, the related prediction accuracy is likewise low.

5. Conclusion

Based on several research papers, two basic methods of English handwriting recognition are compared and analyzed in different ways in this paper, which are the simple RNN model and the LSTM model. In general, most English handwriting recognition models in these papers first use CNN model to capture the main features of handwritten pictures, and then use different RNN models to predict the sequence information considering the influence of time series information before and after. The accuracy of these models is relatively high. However, most papers do not mention the problem of string partitioning, especially in the case of continuous writing, which is likely to become an important factor affecting the final result under the premise that the accuracy of the model itself is high under the application of the attention mechanism. If the problems of string partitioning can be solved effectively, handwriting recognition will surely enter a new stage.

References

- [1] He Juan. Application Research of Handwritten English Automation Recognition Model Based on Deep Learning Network in Machine Translation between English and Chinese. Automation and Instrumentation, 2023.
- [2] Zhang Zhenwei, Wang Hongcheng. Handwritten Digit Recognition Method Based on Improved Convolutional Neural Network and SVM. Journal of Dongguan University of Technology, 2021, (05), 57-61+86.
- [3] Xiao Bin. Research on Commercial Bank Cash Reserve Demand Prediction Model with Fusion of Spatiotemporal Features. Shanghai University of Finance and Economics, 2023.
- [4] Zhang Xinfeng, Yan Kunpeng, Zhao Xun. Research on Handwritten Text Recognition Technology Based on Bidirectional LSTM. Journal of Nanjing Normal University (Natural Science Edition), 2019, 42(03): 64-70.
- [5] Xie Zilin. Oil Well Production Prediction Model Based on Improved Particle Swarm Algorithm and Long Short-Term Memory Neural Network. China National Petroleum Corporation Xinjiang Oilfield Branch

- (Xinjiang Lixian Oilfield Laboratory), Xi'an Petroleum University, Shaanxi Petroleum Society. Proceedings of the 2022 International Conference on Oil and Gas Field Exploration and Development III, 2022, 11.
- [6] Li Guangyan, Wang Xiuhui. Handwritten Character Recognition Method of Lightweight Residual Network with Multiple Branches. *Computer Engineering and Applications*, 2023.
- [7] Gao Tao, Shao Qian, Zhang Yanan, Zhang Sai. Research on Facial Expression Recognition Based on Deep Residual Network. *Electronic Design Engineering*, 2020, 28(23): 4.
- [8] Heng Hongjun, Hu Gang. Aircraft Airworthiness Named Entity Recognition Method Based on BiGRU-LAN. China Highway Society, World Traffic Transportation Conference Executive Committee, Xi'an Municipal People's Government, Shaanxi Association for Science and Technology. Proceedings of the World Traffic Transportation Engineering Technology Forum (WTC2021) (Volume I). People's Communications Press Co., Ltd., 2021, 7.
- [9] Yang Yixin. Research on Remote Sensing Image Fault Detection Method Based on Deep Learning. China University of Petroleum (Beijing), 2023.
- [10] Zhang Yudong, Wu Lenan, Wei Geng. Research on Neural Network Generalization Enhancement Technology. *Science Technology and Engineering*, 2009, 9(17): 4997-5002.