

2012 AASRI Conference on Modelling, Identification and Control

## Character Feature Extraction Method based on Integrated Neural Network

Junchuan Yang<sup>a,\*</sup>, Xiao Yan<sup>b</sup>, Bo Yao<sup>c</sup>

<sup>a</sup>*School of Natural & Mathematical Sciences, King's College London, Strand, London WC2R 2LS*

<sup>b</sup>*School of Information Science and Engineering, Yunnan University, No.2, Chestwood North Road, Kunming, Yunnan, China, 650091*

<sup>c</sup>*School of Computer Science and Electronic Engineering, University of Essex, Wivenhoe Park, Colchester, United Kingdom, CO4 3SQ*

---

### Abstract

The research of pattern recognition shows: For a complex identification and classification system, only one method is always difficult to identify and separate objects well, while among the different classification methods, the high complementary exist, the integration of multiple classifiers can clearly improve the recognition rate. Therefore, integration methods of multiple classifiers have become a hot spot which the researchers are focusing on. This paper sets up an integrated neural network handwriting recognition system, comprehensively use a variety of pattern recognition, fully reflect various aspects features of handwritten digital characters, this method has improved the accuracy of handwritten numeral recognition system.

© 2012 The Authors. Published by Elsevier B.V. Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).  
Selection and/or peer review under responsibility of American Applied Science Research Institute

*Keywords:* BP neural network, characters, feature extraction

---

---

\* Corresponding author. Tel.: +086-156-76190042 ;  
E-mail address: [junchuan.yang@kcl.ac.uk](mailto:junchuan.yang@kcl.ac.uk) .

## 1. Introduction

With the progress of the science and technology, computers are widely used in every walk of life. Character recognition is an area of computer technology applications in many environments and have various kinds of applications. Character recognition is an important part of new generation intelligent. It involves digital image processing, pattern recognition, artificial intelligence, fuzzy mathematics, combinatorial mathematics, information theory, natural language understanding and many other subjects. In nearly two decades, the wide and in-depth research was held for various characters, come up with a lot of effective recognition algorithms; character recognition software has become increasingly rich. Character recognition can be divided into two categories, namely, printed and handwritten character recognition; handwritten character recognition can be divided into a number of handwritten text and handwriting recognition. Handwritten digit recognition is one of the more in-depth areas, because the recognition of handwritten digits can be used in many aspects. Such as zip code recognition, automated processing of bank statements and so on. Although only ten handwritten digits, and simple strokes, like the digital character recognition is not very difficult, but its difficulty is the same number written vastly different, different people have written with a number of differences, even if the same person, in the different times will have different ways of writing. Between numbers without context relevance, identification of each word is very essential. And the shape of the figure is simple, available information is relatively little, in addition, the differences of the people's habit of writing, all the pens, paper also brought handwritten digital recognition digits considerable difficulty. Some test results show that the number of recognition accuracy is not high. This paper studies the recognition problem of the handwritten digit, have an experiment and demonstration for integrated BP neural network classifiers recognition and have achieved good results.

## 2. Bp neural network

Multilayer feed forward networks are an important class of neural networks. This network typically consists of three parts: a set of input nodes, one or more layers of computing nodes in the hidden layer, a layer of compute nodes in the output layer. The input signal on the basis of the progressive layers through the network propagates to the front. Under the way of supervised learning use an error back propagation algorithm to train the network.

Error back propagation algorithm consists of two different layers through the network: A time forward pass and a reverse pass. In the forward pass, an input vector play role in the network input nodes, its influence through the network layer by layer to spread. Finally, produces an output as the actual response of the network. In the forward pass, the network synaptic weights are all fixed. On the other hand, in reverse pass, the synaptic weights all adjust according to the error correction rule. In particular, the response from the target minus the actual response of the network generated error signal. The error signal is back-propagation through network, is the opposite direction with the synaptic connections- so called "error back propagation." Synaptic weight is adjusted that make the actual response of the network from the statistical significance more close to the response target. Error back propagation algorithm in the literature was called bach-propagation algorithm, BP algorithm, or simply a back-prop.

Simply speaking, BP algorithm belongs to a supervised learning algorithm. Its main idea is: for the  $n$  input leaning sample:  $X_1, X_2, \dots, X_n$ , Known to the corresponding output sample  $T_1, T_2, \dots, T_m$ : the object of study is to use the network actual output  $Y_1, Y_2, \dots, Y_m$  and objective vector  $T_1, T_2, \dots, T_m$  error to modify the its weight, make the actual output and objective vector closely, that is: Make the error square sum of the output layer to the minimum. It is through the continuous error function decreasing direction to calculate the network weights and bias changes gradually to approach the goal.

BP algorithm consists of two parts: information forward transmission and error back propagation. In the process of forward transmission, input information through the hidden layer's layer by layer calculation to transfer to the output layer, the state of neurons in each layer affects only the state of the next layer of neurons. If the output layer didn't get the desired output, calculate the value of the output layer error change, and then back propagation, through the network put the error signal along the original connection to rollback, and modify the weights of neurons in each layer until up to the desired goals.

### 3. Improvement of bp algorithm

Although the BP algorithm has high recognition accuracy, the algorithm is relatively simple, the program is very well realized, but also some limitations and drawbacks, some methods can have a significant increase for BP algorithms, can be improved from the following aspects.

#### 3.1. Activation function

Generally speaking, when the activation function network neuron model is a non-symmetric and not symmetric, neural network used for back-propagation algorithm training. In order to minimize learning time, should avoid the use of non-zero mean input. Consider a multilayer neural network used in the first hidden layer neurons signal vector  $X$  (i.e. The signal vector which applied to the input layer) terms, it is easy before  $X$  applied to the network; reduce an average value of each element. However, the signal applied to the rest of the hidden layer and output layer neurons, the result is the activation function used in the network type. If the activation function is asymmetric, the output of each neuron in the world  $[0, 1]$ . Such a choice for those in the network after the first hidden layer of neurons brought about a systematic bias source. To overcome this problem, need to use an anti-symmetric function like hyperbolic tangent function. For the latter option, the output of each neuron can be the interval  $[-1, 1]$  any of the positive and negative, in which case it might mean to 0. If the number of network connections is large, using the back propagation learning of symmetric activation function learning may converge faster than using asymmetric than a similar process of activation function. A very popular example about opposed function is a hyperbolic tangent sigmoid nonlinear, and  $a$ ,  $b$  is constant. Suitable  $a$ ,  $b$  value is  $a=1.7159$   $b=2/3$  hyperbolic tangent function which defined like this has the following useful properties:  $\varphi(-1) = -1$  Activate the gradient of function at the origin (i.e. the effective gain) is close to 1, as followed:  $\varphi(0) = ab = 1.7159 \times 2/3 = 1.1424$   $\varphi(v)$  second derivative when  $v = 1$  up to maximum.

#### 3.2. Initialization

A good choice for synaptic weights and initial closure of network had a huge help for a successful network design. When the synaptic weight was given to a large initial value, then the network of neurons is likely to become saturated. If this happens, the local gradient of back-propagation algorithm presented a very small value, resulting in that back-propagation learning process is very slow. However, if the synaptic weight is given a small initial value, back-propagation algorithm may be done within a flat area of the error surface original point; especially for symmetric functions (such as the hyperbolic tangent function) Conditions, this is more likely. The origin is a saddle point, the saddle point is a stable point and the saddle point in the orthogonal error surface curvature is positive, and negative along the saddle direction. For these reasons, should avoid using the over big or too small the initialized synaptic weights. Suitable initialization choice is between these two extreme. Specifically, consider to put a hyperbolic tangent function as activation function of multilayer neural networks. Let each neuron network bias to 0, and the value of a neuron is expressed as

$$v_j = \sum_{i=1}^m w_{ji} y_i \tag{1}$$

Assuming the network output of each neuron the mean 0 and variance 1, is expressed as  $\mu_y = E[y_i] = 0$

For all neurons i and

$$\sigma_x^2 = E[(y_i - u_i)^2] = E[y_i^2] = 1 \tag{2}$$

For all neurons i, further, assume that the input values are not relevant, that is

$$E[y_i y_k] = \begin{cases} 1, & k = i \\ 0, & k \neq i \end{cases} \tag{3}$$

Synaptic weight and the value set is the uniform distribution with mean 0 of a group of numbers drawn

$$u_w = E[w_{ji}] = 0 \tag{4}$$

For all (j, i) and Variance

$$\sigma_w^2 = E[(w_{ji} - u_w)^2] = E[w_{ji}^2] \tag{5}$$

For all (j, i) And the mean and variance of value was expressed as

$$u_v = E[v_j] = E\left[\sum_{i=1}^m w_{ji} y_i\right] = \sum_{i=1}^m E[w_{ji}] E[y_i] = 0 \tag{6}$$

And

$$\sigma_v^2 = E[(v_j - u_v)^2] = E[v_j^2] = E\left[\sum_{i=1}^m \sum_{k=1}^m w_{ji} w_{jk} y_i y_k\right] = \sum_{i=1}^m \sum_{k=1}^m E[w_{ji} w_{jk}] E[y_i y_k] = \sum_{i=1}^m E[w_{ji}^2] = m \sigma_w^2 \tag{7}$$

Here m is a neuron's synaptic weight connection number. Based on the above results, how to describe synaptic weights as mixed-signal strategy to make the neurons and the standard deviation of the value locate the transition region of its linear part and the saturation part of sigmoid activation function. I.e. As mentioned earlier the parameters a and b the value established by the hyperbolic tangent function, but the above equation when  $\sigma_v = 1$  can meet this objective, so can get  $\sigma_w = m^{-1/2}$ , Therefore, for a uniform distribution, it needs its mean 0 and variance will be inversely proportional to the number of neurons, thus the distribution choose the value of synaptic weight.

#### 4. Extraction method basing on BP neural network character feature

Feature extraction is the key to the character recognition system; recognition algorithm is based on the type of selection feature to be selected. If the selected features are stable, whether on behalf of the characteristics of a class of characters, is the key of high and low system recognition rate. Handwritten character feature extraction method can be divided into two kinds of structural features and statistical features. Below separately states the features of these two extraction methods and general method. From the perspective of

statistical pattern recognition, character recognition is actually a pattern classification problem, people recognize natural object recognition is based on the learning objects, features analysis, and computer and human pattern recognition process share a lot of similarities. Actually, by learning or other methods, to form a memory knowledge base, when implement pattern recognition, clearly express an image from the object to the memory knowledge base, resulting in the recognition results. When human carry out object recognition, use memory base which the brain formed through the study, and process an image for the identified objects, to find out the matching category from the memory base. The computer express this image when human recognize the objects, is generally accomplished by two steps: first, to describe the objects in appropriate characteristics, that is, from xi to f (xt) image; the second step, the computer performs some operations to complete the image form the f (xi) min c (x) = X. This process is actually a general method when the traditional statistical pattern recognition carried out object recognition, specifically, feature extraction and the design problem of classification functions. The feature extraction is a difficult and key issue. If the feature is known, we can use the existing mathematical theory to guide the design of the image function  $R(x_i)$ . However, for feature selection and extraction, but no theory guide to follow, difficult to compare the characteristics of an object in what is substantive, are representative of those features, those features may be unimportant or irrelevant and recognition. This requires a lot of experimental and theoretical guidance.

Supposed the original image is  $F(m, n)$  (the character of this article is stored by the way of lattice form, lattice formed grayscale image, Through the transformation, each point can be extracted from the gray value)  $m = 1, 2L, M, n = 1, 2LN$ , Normalized image  $G(i, j), i = 1, 2L, I, j = 1, 2LJ, H(m), V(n)$  Strokes, respectively horizontal and vertical pixels in the projection function,

$$H(m) = \sum_{n=1}^N F(m, n) \tag{8}$$

$$m = 1, 2LM \quad V(n) = \sum_{m=1}^M F(m, n) \quad n = 1, 2LN \tag{9}$$

The purpose of nonlinear regularization approach: By coordinate transformation, make the normalized image feature space as even as possible distribution.

$$i = \sum_{k=1}^m H(k) \times I \div \sum_{k=1}^m H(k) \tag{10}$$

$$j = \sum_{k=1}^n V(k) \times J \div \sum_{k=1}^n V(k) \tag{11}$$

This equation shows that the original image pixel (m, n) in the normalized image in the corresponding position (i, j). In order to reduce characteristics dimension, put the normalized image compression 16 × 16 to 8 × 8 dimensions, which constitutes the character of the 64 dimensional global features.

The edge can outline the target object, contains wealth information, and is the important feature to extract the image feature in the image segmentation, identification and analysis. Edge detection use extremism of image first-order derivative or second derivative zero-crossing edge points information to provide the fundamental basis is an important class of image segmentation. The edge of the image is the result that the image local characteristics are not continuous (or mutation), such as the gray value of the mutation, the color mutation, texture mutation. Gradient corresponds to the first derivative, for a continuous image function f (x, power, the gradient is a vector, defined as:

$$\nabla f(x, y) = [G_x G_y]^T = \left[ \frac{\partial f}{\partial x} \frac{\partial f}{\partial y} \right]^T \quad (12)$$

$G_x$  And  $G_y$  separately are the gradient along x direction and y direction. Gradient magnitude  $|\nabla f(x, y)|$  and direction angle  $\phi(x, y)$  separately is

$$|\nabla f(x, y)| = \text{mag}(\nabla f(x, y)) = (G_x^2 + G_y^2)^{1/2} \quad (13)$$

$$\phi(x, y) = \arctan(G_y / G_x) \quad (14)$$

By the above equation the value of gradient is  $f(x, y)$  unit distance increased volume in the maximum rate direction. For digital images, the gradient is a difference instead of differential to achieve, the type can be written as

$$|\nabla f(x, y)| = \left\{ [f(x, y) - f(x+1, y)]^2 + [f(x, y) - f(x, y+1)]^2 \right\}^{1/2} \quad (15)$$

This article cites the classifier identification reliability theory and expands it, and applies to handwritten digits recognition. This method is simple, less computation, and adds a new classifier to make combination without affecting the existing data. Experimental data further confirmed the feasibility and potentials of classifier combination. Classifier combination have many ways, the key is to find characteristics' value which reflect the sample recognition performance for the classifier. Now the performance of a single classifier has been greatly improved, classifier combination is certainly a meaningful way.

## 5. Conclusion

The research in the field of pattern recognition shows: For a complex problem of identification and classification, only one method always is difficult to identify the object well separated, while among the different classification methods, the high complementary exist, the integration of multiple classifiers can clearly improve the recognition rate. Therefore, integration methods of multiple classifiers have become a hot spot which the researchers are focusing on. His paper set up an integrated neural network handwriting recognition system, comprehensively use a variety of pattern recognition, fully reflect various aspects features of handwritten digital characters, this system improve the accuracy of handwritten numeral recognition.

## References:

- [1] Caichun Gong, Rongxing Liu, quick handwritten number character recognition basing on the overall features, Computer Engineering and Application 2004.19
- [2] Fengying Xie, Danpei Zhao, Visual C++ digital image processing, Electronic Industry Press, 2008
- [3] Hui Dong, Handwritten Numeral Recognition Feature Extraction and Feature Selection Research, Beijing University of Posts and Telecommunications Master ETD 2007
- [4] Feng Lu, Weixing Yi, Xiaohui Yuan, Neural Network Pattern Recognition Method Research for Digital Identification in Financial Instruments, Suzhou, Jiangsu Branch of Science), 2003.8
- [5] Xiaobin Luo, Shoubin Dong, "Integrated Neural Network Handwritten Digit Recognition, " Computer Engineering, 28, No. 8, August 2002
- [6] Li Zhang, JingYu Yang, Zhen Lou, " Confidence-based Recognition of Handwritten Digits Dynamic Combination of Multiple Classifiers " Computer Engineering, 29, No. 16, September 2003