An algorithm of Handwritten Digital Recognition Based on BP-Bagging

Zuojun Liu[1,a*], Lihong Li[2,b]Mi Yu[3,c]

[1]Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, Huai'an, Jiangsu, China

[2]Faculty of Foreign Language, Huaiyin Institute of Technology, Huai'an, Jiangsu, China

[3]Faculty of Foreign Language, Huaiyin Institute of Technology, Huai'an, Jiangsu, China

[a]hyliuzj@126.com, [b]hylilihong@126.com, [c]hyyumin@126.com

Keywords: BP-Bagging; Classifier; Image; Feature

Abstract. The handwritten number recognition algorithm based on BP-bagging generates the basic classification by BP network and generates multiple classifiers by bagging. The algorithm treats a handwritten character as an image. By scanning the images, 25 dimension features are extracted, and then compresses the 25 dimensions features into 5. The handwrite digital can be recognized by input it into BP-bagging classifier and by multiple voting integration. Proved by repeating experiments, the algorithm bears a high recognition rate, which functions better than single classifiers and other basic classifier combination algorithm.

Introduction.

The computer is widely used in all kinds of fields. Automation and intellectualization is becoming universal. The recognition technology of handwritten digital has been widely used in zip code recognition and banking business etc. The recognition technology has become an important subject in many fields. The handwritten digital is differing in thousands of way, such as deformation, translation and scale change. The number of the digital is very few only 0 to 9 and the stroke of the digital is very simple. The writing of each digital is differing in various ways. The people living in different area all are using digitals. The writing takes on obvious regional characteristic. So, it's hard to develop a universal digital recognition system. In many fields, people require high recognition rate. How to improve the recognition rate has become a problem to be solved urgently.

Massive scale parallel processing, fault tolerance and learnability are characteristics of BP neural network. After training, the neural network can be used in handwriting digital recognition. The invariant features of handwriting digital can be extracted by the neural network.

Due to the weakness of itself, the neural network classifier belongs to weak classifier. It is difficult to improve the recognition rate of pure neural network. So, bagging is introduced in this paper. The weak classifier can be transformed into stronger one by introducing bagging.

Classifier Design of Bagging.

Bagging is a method to produce lots of classifiers. The training set of bagging is composed of some automatically-selected samples. The scale of the training set is almost the same as the original data set. The samples in the training set are allowed to be selected repeatedly. So, some samples may appear more than once, and other samples may never appear.

Theory and Thoughts of Bagging. An algorithm of polynomial level can be used to recognize a series of concepts. If the recognition rat is pretty high, the group of concepts can be called strong learnable. Otherwise, the group of concepts can be called weak learnable. In general, a strong learning algorithm is hard to get. If a weak learning algorithm can be transformed into a strong one, it is unnecessary to find a stronger one.

A weak classifier and a group of training set are selected. Selecting n samples each time from the initial training set randomly. A predict function can be got after one time training. T predict functions

$h_1, h_2, h_3, \mathbf{L}\, h_T$ can be acquired after T times training. After predicting the samples with these predict functions, we can get the predict result $h*$ according to the majority voting rules.

Implementation of Bagging. The thoughts of bagging include two aspects: one is to generate predict function, the other is to combine the result of predict function to generate conclusions.

Bagging provides a method for generating prediction function. Breiman gives a theoretical analysis about classification problem. He pointed out that the highest accuracy of classification problem can be expressed as formula(1), and the highest accuracy of classification problem using bagging can be expressed as formula(2),

$$r = \int \max P(j \mid x) P_x(x) \tag{1}$$

$$r_B = \int_{x \in S} \max P(j \mid x) + \int_{x \in S'} [\sum_j I(f_A(x) = j) P(j \mid x)] P_x(x) \tag{2}$$

S is the correct input set. S' is the complement set of S. $I(*)$ is the indicator function. It can be seen that the accuracy rate using bagging can reach to the highest. From the angle of deviation and variance, Breiman pointed out that the variance will be larger if the deviation of the unstable prediction function is smaller.

The predict function can be generated by using bagging technology. The samples of the training set of each weak classifier is selected randomly form the initial training set. The sample can be selected repeatedly. So, some samples in the initial training set can be selected many times, and other samples may never be selected. The repeated selection can increase the difference of individual function, which can reduce the generalization error.

There are two methods of generating bagging conclusion. One is absolute majority vote; the other is relative majority vote. The method of absolute majority vote refers that the category can be determined only when the poll reaches to the most and exceeds half of the total poll. The method of relative majority vote refers that the category can be determined when the poll reaches to the most.

Suppose that T weak classifiers have been generated, that the predict functions of each weak classifier give the correct classification results with a probability of 1-p, that the errors of each weak classifier are not related, and adopting the method of majority vote, the error probability of final predict function can be computed with formula (3),

$$p_e = \sum_{k>T/2}^{T} C_T^k P^k (1-p)^{T-k} \tag{3}$$

When p<1/2, P_e will decrease when T increases. Therefore, the more is the number of the weak classifier to be integrated, the higher is the integration precision. So, the relative majority vote can get better results.The equation of relative voting is shown as formula (4),

$$h*(x) = \begin{cases} j, m\ (\ x \in C_j\ ) = \max m(x \in C_i) > l \bullet K \\ 10, otherwise \end{cases} \tag{4}$$

In the formula, x refers to the training sample. $l \bullet K$ is the vote threshold. K is the number of network vote, $i = 0,1,2, \mathbf{L}\, 9$.

Choosing Weak Classifier.

Bagging weak classifier can be the nearest neighbor classifier, decision tree, neural network, etc. The BP neural network is not stable. For a unstable algorithm, the effect of bagging is very obvious. Therefore, the BP network can be used as the weak classifier, whose topology is shown as figure 1.

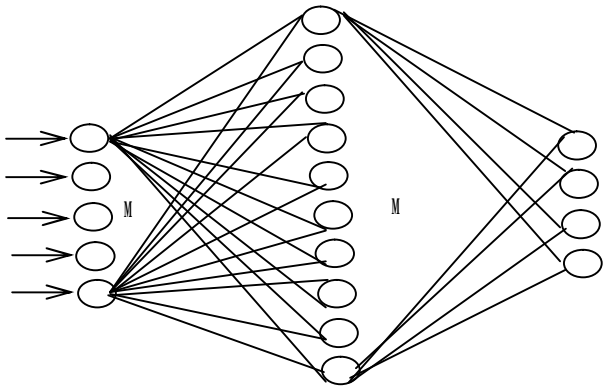

Fig1 topological structure

The topological structure of the network is forward feed network of three-layer including input layer, hidden layer and output layer. The layers are fully connected. The neurons in each layer are not connected. The number of neuron in input layer is decided by dimension of eigenvector. Each neuron is corresponded to the one dimension component of eigenvector. The number of neuron is decided by the number of models in the model set. Each neuron is corresponding to one model in the model set.

In hidden layer and output layer, the output O_j of some neuron j can be calculated by formula (5),

$$O_j = f_j(Net_j) = f_j(\sum w_{ij} x_i + q_j) \tag{5}$$

In this formula, f_j is an activation function corresponding with neuron j .At present, the activation function is generally the *sigmoid* function, $f(x) = \frac{1}{1+e^{-x}}$. q_j is the threshold of neuron j . x_i is the input of neuron j . w_{ij} is the connection weights between input and neuron j .

Digital Feature Extraction.

Each digital is processed into an eight bit bitmap file, which is composed of the file header, bitmap information header, color information and image data. Each pixel in eight bit bitmap file is expressed with eight bits. So, the color of each pixel can be expressed as one of 256 colors. The eight bit bitmap file contains a color table, which is used to show the color of each pixel. The bit number of each pixel can be 1, 4, 8. The top three have corresponding color tables. The color table of 24 bit bitmap file is empty. Its pixel value is just the color value.

The handwritten digital that the user inputted in is eight bit bitmap file. In order to realize digital recognition and improve the recognition rate, a work of pretreatment must be done to the bitmap file so as to filter the noise and interference. Then, we can scan the bitmap file to divide it into five rows and five columns (total 25 small squares). The proportion of black pixel in the small square can be used as the feature of the image. The 25 dimension eigenvector will lead to too much learning time. Therefore, the eigenvector must be compressed. The 25 dimension eigenvector projected onto the horizontal direction. The first eigenvalue to the fifth eigenvalue are added as the new first eigenvalue, and so on. The eigenvector of 25 dimensions is transformed into a new eigenvector of 5 dimensions.

Digital Recognition.

The digital recognition is to input the eigenvector to the classifier to get a recognition result. Only one digital can be recognized each time. There are three recognition results: correct recognition, false recognition and unable to recognize.

Network Topology. The eigenvector of handwriting is five dimensions. Therefore, the input layer of BP network has five nodes. According to the twice relation between input layer and hidden layer, the

hidden layer should be ten nodes. The result of output has ten probabilities from 0 to 9. So, the output layer should have four nodes. The output of the network is four binary numbers, which represents the number to recognition. For example, if the output of the network is '1000', the recognition result is number 8. The number cannot be recognized, if the recognition result is out of 0 to 9. We can get the recognition result, if we input the eigenvalue of samples into the network.

Recognition Algorithm Based on BP-Bagging. For the characteristics of BP network and bagging, we unite bagging with BP-network to form a recognition algorithm based on BP-bagging. The steps of the algorithm are shown as follow:

(1)Assuming that the training sample set is $D = \{x(i)\}, i = 1,2,\mathbf{L},N_1$, that the total training number is N, that the initial value of training times is 0, that every training generate a basic classifier, then, the total number of classifier is N.

(2)In the i time, k samples are extracted randomly from the training set D to form a training subset, which is used to train the classifier to get a classifier of number i. To repeat this step if $i < N$.

(3)The verification samples can be classified by the N classifiers. Then, the results are integrated with the simple majority vote. Therefore, the result of classification can be got.

Performance Indicators of Classifier. The performance index of handwritten digital recognition can be precision, recall and $F1$, which are to evaluate the performance of the algorithm in each category. Precision is a ratio of the number of sample digital classified correctly and the total number of sample digital. The equation is shown as formula (6),

$$\text{Precision} = \frac{the\quad number\quad of\quad digital\quad classified\quad correctly}{the\quad total\quad number\quad of\quad digital} \tag{6}$$

Recall is a ratio of the number of digital classified correctly and the total number of digital that should be. Its equation is shown as formula (7),

$$\text{Recall} = \frac{the\quad number\quad of\quad digital\quad classifierd\quad correctlly}{the\quad total\quad number\quad of\quad digital\quad that\quad should\quad be} \tag{7}$$

$F1$ is the harmonic mean of precision and recall. Its equation is shown as formula (8),

$$F1 = \frac{precision \times recall \times 2}{precision + recall} \tag{8}$$

Experimental Results and Conclusions.

In order to verify the validity of this algorithm, we developed a system. The samples used in the experiment are selected randomly form seven thousands samples sets with full consideration of the representative and diversity. There are five hundreds samples in the training set and one thousand samples in the verifying set. The experiment is divided two stages: one is the pure BP-network recognition, the other is bagging recognition. The experiment times after introducing bagging increase constantly to verify the recognition effect of bagging. The verification results of each stage are shown as in table 1.

Table 1 Verification Results

		BP-network	Bagging (15times)	Bagging (30times)
Training Set	precision	0.946	0.978	0.985
	recall	0.957	0.985	0.990
	F1	0.950	0.980	0.989
Verification Set	precision	0.893	0.975	0.988
	precision	0.911	0.981	0.991
	F1	0.917	0.979	0.989

Proved by the experiment, the handwritten digital recognition rate introducing bagging is improved obviously compared with pure BP classifier. At the same time, the error rate decreased obviously. This shows that the pure BP network is hard to improve the recognition rate for itself instability and defects, which result in an increased network variance. Bagging improves the effectiveness of the learning algorithm by reducing the difference of learning system. It can be seen from the data parameters in table 2 that the error rate is reduced and the precision and recall are constantly rise with the increasing of training times.

Table 2 Experiment Results

Times	2	4	6	8	10	12	14	16	18	20
Error	0.079	0.061	0.052	0.043	0.041	0.035	0.040	0.039	0.038	0.020

Conclusions.

It's difficult to improve the recognition rate of pure BP network classifier for the flaws of itself. So, the recognition rate can't meet the needs of various fields. In this paper, we combined the bagging with BP network. An algorithm of handwritten digital cognition based on BP-Bagging is proposed. By this method, the side effects of local minimum points are eliminated. The time complexity and space complexity increase to some extent, however, the precision ratio and recall ratio of the algorithm are greatly improved, which can satisfy the requirements of various fields.

References

[1] Kearns M, Valiant L G. Learning Boolean Formulae or Factoring[R].Technical Report TR-1488, Harvard University Aiken Computation Laboratory, 2009, 43(3): 28-33.

[2] Raj Dharmarajan Jyer Jr. An Eifficient Boosting Algorithm for Combining Preferences [D]. Massachusetts Institute of Technology, 1999, 41(3): 28-33.

[3] Gareth James.Majority Vote Classifiers: Theory and Application [D] Stanford University, 1998,36(3):28-33.

[4] Alexander Goltsev , Donald C. wunsch. Generalization of Features in the Assembly Neural Networks [J]. International Journal of Neural Systems (IJNS), 2008, 24 (4): 39-56.

[5] I. Guyon. Applications of Neural Networks of Character Recognition [J]. International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI), 2005, 44 (2): 353-382.