Handwritten Digits Recognition Base on Improved LeNet5

Naigong Yu [1] ,Panna Jiao[2], Yuling Zheng[3]

1.Beijing University of Technology, Beijing, 100124, China

E-mail: yunaigong@bjut.edu.cn,

2.Beijing University of Technology, Beijing, 100124, China

E-mail: pannajiao@126.com

1.Beijing University of Technology, Beijing, 100124, China

E-mail: 965290419@qq.com,

Abstract: LeNet5 is a kind of Convolutional Neural Network (CNN) and has been used in handwritten digits recognition. In order to improve the recognition rate of LeNet5 in handwritten digits recognition, this article presents an improved LeNet5 by replacing the last two layers of the LeNet5 structure with Support Vector Machines (SVM) classifier. And LeNet5 performs as a trainable feature extractor and SVM works as a recognizer. To accelerate the network's convergence speed, the stochastic diagonal Levenberg-Marquardt algorithm is introduced to train the network. A series of studies has been conducted on the MINST digit database to test and evaluate the proposed method performance. The results show that this method can outperform both SVMs and LeNet5. Moreover, the improved method gets a faster convergence speed in training process.

Keywords: Handwritten digit recognition, convolutional neural networks, Support vectors machines, Stochastic diagonal Levenberg-Marquardt

1 Introduction

Handwritten digit recognition is a research hotspot due to its vast practical applications and financial implications. It demands a decent recognition rate with the highest reliability[1]. Numerous experts and scholars were engaged in improving the recognition accuracy for handwritten data and also made great progress with many different methodology, techniques ranging from statistical methods such as Principal Component Analysis (PCA) [2] to machine learning like Neural Networks (NNs) [3] or Support Vector Machine (SVM) [4], etc.

The important areas in handwritten digit recognition include the feature extraction and the classification. The selected feature should contain the most distinguishable characteristics among different classes while retaining invariant characteristics within the same class as much as possible. Concolutional Neural Networks (CNNs) was brought about by LeCun and caused huge attention immediately in 1995 [5]. The advantage of CNN is that it automatically extracts the salient features which are invariant and a certain degree to the shift and shape distortions of the input characters. The advantage of Support Vector Machines (SVM) is that it takes into account both experimental data and structural behavior for better generalization capability based on the principle of Structural Risk Minimization (SRM) [6]. Fabien Lauer et al. [7] proposed a system such that the LeNet5 Convloutional Neural Network trained a trainable feature extraction and SVM performed the classification task. Mori et al. [8] trained the convolutional spiking neural network with time domain encoding schemes module by module using different fragment images, and then the outputs of each layer in the model were fed as features to the SVM. Niu in article [1] proposed a novel hybrid CNN-SVM classifier for recognition handwritten digits and the system obtained the low error rate.

CNN is trained by a popular neural network -back propagation algorithm. But it requires making many seemingly arbitrary choices and there is no foolproof recipe for deciding them, because they have large samples problem. In addition, the convergence speed of BP algorithms tediously slow and it always requires hundreds even thousand epochs for convergence. A CNN and SVM model are designed for handwritten digit recognition in this paper. The model automatically retrieves features based on the LeNet5 architecture, and recognizes the unknown pattern using the SVM classifier. In this model, the stochastic diagonal Levenberg-Marquardt algorithm is introduced to accelerate the convergence speed of the neural network.

The rest of this paper is organized as follows: Section 2 gives a brief description of LeNet5, SVM and presents a improved method; Section 3 shows the experiments result base on MNIST database and then certain analysis were also given in this part; Section 4 draw the conclusions.

978-1-4799-7016-2/15/$31.00 ©2015 IEEE

2 The Improved LeNet5

2.1 The improved LeNet5

Convolutional Neural Network [9] is a multi-layer feed-forward neural network that can extract topological properties from the raw image. It has a deep supervised learning architecture that can be viewed as the composition of two parts: an automatic feature extractor and a trainable classifier. The feature extractor is usually an alternation of convolutional layers and subsampling layers. The convolutional layer is organized in planes, or called feature maps, of simple units called neurons. Each unit receives 25 inputs from a 5 by 5 area in its previous layer, which is the local receptive field. Then, by multiplying 25 coefficients plus a trainable bias to calculate the value of this unit, all units of one feature map share the same set of coefficients and the bias. Moreover, the coefficients sharing technique allows to reduce the number of trainable parameters. The trainable classifier is formed by one fully connected layer and output layer.

LeNet5 is a common model of CNNs [10], as shown in Fig.1.

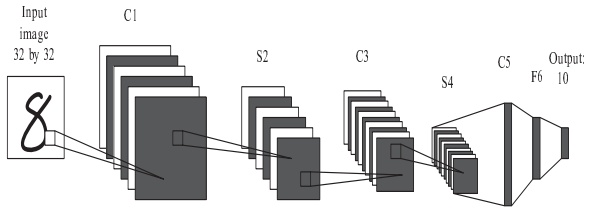

Fig.1:Structure of the LeNet5.

From the fig.1, The input is the 32×32 pixels handwritten image. The LeNet5 is consist of 7 layers: three convolutional layers, two subsampling layers, one fully connected layer and the output layer. The first layer is a convolutional layer (C1), which has 6 feature planes of 28×28 pixels. In the subsampling layer (S2), these planes are reduced into 14×14 pixels for one feature plane. The next convolutional layer (C3) extends the number of feature maps to sixteen. The subsampling layer S4 acts as S2, 16 feature planes are reduced to half their sizes. The last convolutional layer C5 has 120 feature planes, C5 is equivalent to a fully connected layer. The fully connected layer F6 contains 84 units connected to the 120 units of C5. Finally, the output layer is a Euclidean RBF layer of 10 units.

2.2 SVM Classifier

Support Vector Machines [11] with different kernel functions map the input vector into a high dimensional feature space and then finding the Optimal Separating Hyper-plane (OSH). OSH maximizes the distance between the hyper plane and the nearest data points of each class in the space. SVM was initially proposed to solve two-class problems.

At present, there are two main approaches to solve the multiclass problems. One is the single-machine approach in which a single large optimization problem is solved in [12].Three typical strategies applied are: one-against-one, one- against -all, and Directed Acyclic Graph (DAG) SVM. The other is the multiple-machine approach, which transforms a multi-class classification problem into several binary classification problems [13,14]. Hsu et al. [15] compared these two approaches and concluded that the second approach, specifically, one-against-one and DAG methods are more suitable in practice than other methods. In our experiments, the one-against-one type method is implemented for multi-class SVMs.

Consider a k-class problem, there are l training samples: $\{x_1, y_1\}\{x_2, y_2\}.....\{x_i, y_i\}$, where $x_i \in R^m$, $i = 1, 2, ..., l$ are feature vectors and $y_i \in \{1, 2, ..., k\}$ are the class corresponding labels. The one-against-one method constructs k(k-1)/2 classifiers, where each classifier uses the training data from two classes chosen out of k class. For training data from the ith and the jth classes, we need to solve the following optimization problem:

Min:

$$\frac{1}{2}(w_{ij})^T w_{ij} + C\sum_n \xi_{ij}^{\ n} \tag{1}$$

Subject to:

$$\begin{cases} (w_{ij})^T \phi(x^n) + b_{ij} \geq 1 - \xi_{ij}^{\ n}, y^n = i \\ (w_{ij})^T \phi(x^n) + b_{ij} \leq -1 + \xi_{ij}^{\ n}, y^n \neq i \\ \xi_{ij}^{\ n} \geq 0, n = 1, ...k(k-1)/2 \end{cases} \tag{2}$$

As for classification by SVM, We use the Radial Basis Function (RBF) kernel function as the kernel of SVM:

$$K(x, x_i) = \exp(-\frac{\|x - x_i\|}{\delta^2}) \tag{3}$$

2.3 The Improved LeNet5 CNN

The architecture of the improved LeNet5 was designed by replacing the last two layers of the LeNet5 structure with SVM classifier. Last convolutional layer C5 has 120 feature values, these values can be treated as features for any other classifiers. Fig.2. shows the structure of improved LeNet5 and SVM model.

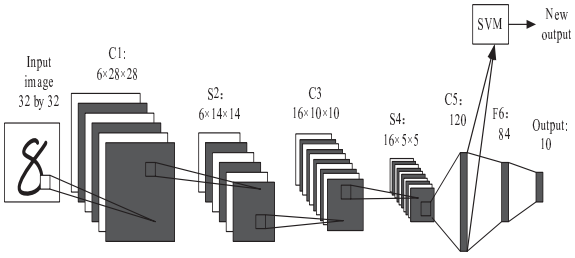

Fig.2:The structure of the improved LeNet5

First, the centered and normalized images are sent to the input layer, and the original CNN is trained with several epochs until the weights of these layers are trained, so that its output layers can get the minimum classification error.

Then, the SVM with a Radias Basis Function (RBF) kernel replaces the last two layers. The SVM takes the 120 outputs of the last convolutional layer for training. When the SVM classifier is well trained, it can perform the recognition task and make new decisions on testing images.

The classifier and the weights learned in the feature extractor are always trained by a back-propagation, but the convergence of the neural network was tediously slow and also required hundreds even a few thousand of epochs iteration. In improved LeNet5 model, the classifier is trained by a second order technique that called stochastic diagonal Levenberg-Marquardt to update weights. This paper updates the weights after each presentation of a single pattern in accordance with stochastic update methods. The patterns are presented in a constant random order, and the training set is typically repeated 20 times. At each learning iteration a particular parameter w_k is updated according to the following stochastic update rule:

$$w_k \leftarrow w_k - \xi_k \frac{\partial E^p}{\partial w_k} \tag{4}$$

Where ξ_k is an individual learning rate, it is computed for each parameter before each pass through the training set [16].

3 Experiments

In order to demonstrate the capabilities of the LeNet5 and SVM model used in the recognition of the handwritten digits, the experiment performed on the well-known MNIST digit dataset. The MNIST database contains 60000 training samples and 10000 test samples, and can be downloaded from internet[17]. Some samples on this database are shown in Fig.3.

The images in the MNIST dataset are converted into binary numeral bitmaps ,then they have been centered and size normalized to 28×28 pixels at the stage of pre-processing.

In the next section, the convergence speed of the model is estimated and experimentally evaluate the recognition performance is evaluated on the MNIST dataset.

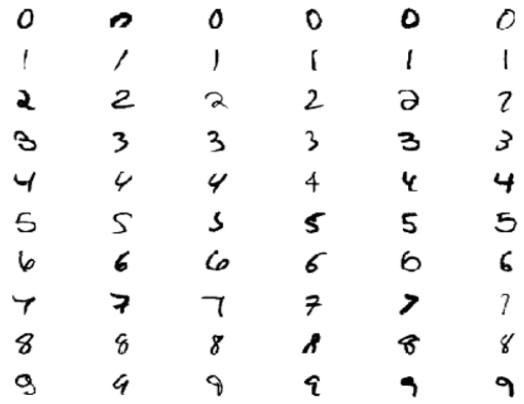

Fig.3:Samples images in MNIST database.

3.1 Parameter Estimation

The improved LeNet5 model is built and trained. The last two layers of LeNet5 CNN is replaced by SVM classifier to predict label of the input pattern. 120 values from the layer C5 of the trained CNN network are used as a new feature vector to represent each input pattern, and fed to the SVM for learning and testing. The RBF kernel is used to build the SVM in the model and determine the optimal kernel parameter δ and penalty parameter C by applying the 5-fold cross validation method on the training dataset. The grid searching range of each parameter is: $\delta = [2^3, 2^1, 2^0, 2^{-5}]$ and $C = [2^{10}, 2^9, 2^8, 2^{-5}]$. The experiment tried 8×15=120 different combinations. The best parameters are achieved when C=128 and $\delta=2^{-11}$.

The idea of implement the stochastic diagonal Levenberg -Marquardt algorithm is to compute the diagonal Hessian through a running estimation of the second derivative concerning each parameter. The learning rate ξ_k are constant but the second derivative of the loss function is changing along with the axis w_k :

$$\xi_k = \frac{\eta}{h_{kk} + \mu} \tag{5}$$

$$h_{kk} = \left\langle \frac{\partial^2 E}{\partial w_k^{\,2}} \right\rangle \tag{6}$$

In the experiment, the initial value of η is 0.01, if η is too large the network can't convergence. If η is too small, the speed of convergence will tediously slow and the network is much easier to fall into local minimum value. In the process of training, the heuristic rules are used to change the value the η, the minimum value is 5e-005 in the experiment. h_{kk} is a running estimation of the diagonal second derivative concerning w_k, it can adjust the sample quantity according to the size of the sample set. 100 samples are selected randomly to estimate it. μ is a parameter to prevent ξ_k from blowing up in case the second derivative is small.

3.2 Results Analysis

These parameters are used to train the improved LeNet5 model. The training procedure is stopped after 40 epochs

and converged to a fixed value. the result is showed in fig.4.

On the MNIST database, each epoch demands 60000 back propagations and takes around 50 minutes. In article [1], Niu trained the network by a back-propagation algorithm and the training procedure requires 500 epochs and converged to a fixed value. By contrast, the stochastic diagonal Levenberg-Marquardt algorithm performs good convergence in around 25-30 epochs in the experiment. This is amazing for too reasons: First, the confidence of the network performs correctly is increased. Second, the network converged in around 20-24 hours at 50 minutes each epoch, which is a appropriate time for running.

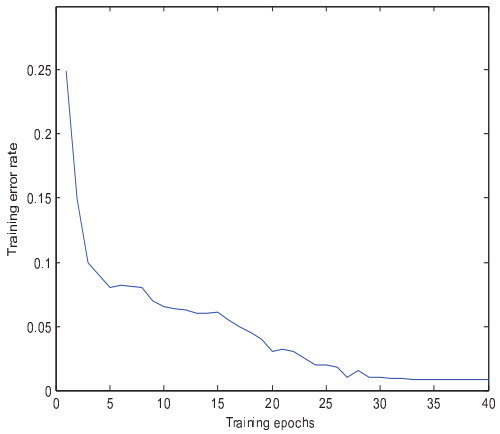

Fig.4： Training error rates of CNN on the MNIST dataset.

Table 1 :shows published results of other paper on the MNIST database together with the best results of this paper.

Classifier	Reference	Test error rate(%)
SVM	[18]	1.4
CNN	[9]	0.95
CNN-SVM	This paper	0.85

From Table1, the proposed method outperforms both CNNs and SVMs without using new samples to expand the training set.

Base on the above experiment, the CNN and SVM improve the performance of the recognition: lower test error rate and fast convergence speed.

4 Conclusion

This paper presented two of the best classifiers used in handwriting recognition: the LeNet5 Convolutional Neural Network and the Support Vector Machine. The LeNet5 CNN and SVM method has been proposed to solve the handwritten digits recognition problem. The method took the LeNet5 as an automatic feature extractor and SVM acted as the output predictor. This paper uses a second order method the stochastic diagonal Levenberg-Marquardt algorithm to accelerate learning in certain cases. The proposed method was evaluated in two aspects: the convergence speed and the recognition rate. Experimental result base on the MNIST digit database shows the improved LeNet5 method outperformed both LeNet5 and SVM and it is able to achieve good convergence in around 25-30 epochs.

The experimental results indicates that the proposed method is quite a promising classification method in the handwritten recognition field due to three advantages: First, it is combination of the advantage of the CNN and SVM which are the best classifier in handwritten character recognition field. Second, the salient features can be automatically extracted by the improved LeNet5. Third, the improved LeNet5 required only 25-30 epochs for convergence, which is much less time than other method.

Acknowledgment

The authors thank the anonymous reviewers for their constructive comments and suggestions.

REFERENCES

[1] X. X. Niu and C.Y.Suen, A novel hybrid CNN-SVM classifier for recognizing handwritten digits, *Pattern Recognition*, 45 (2012) 1318-1325.

[2] V. Deepu, M. Sriganesh, A. G. Ramakrishman, Principal component analysis for online handwritten character recognition, Proc. *IEEE 17th International Conference on Pattern Recognition (ICPR04)* , IEEE press, Aug. 2004, pp.327-330.

[3] M. Kang , D. Palmer-Brown, A modal learning adaptive function neural network applied to handwritten digit recognition, *Information Sciences,* (178) 2008, pp: 3802-3812.

[4] D. Gorgevik, D. Cakmakov, Handwritten digit recognition by combining SVM classifiers Proc. *IEEE International Conference on Computer as a Tool, IEEE press*, Nov.2005, pp.1393-1396.

[5] Y. LeCun and Y. Bengio. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks,* 1995.

[6] V. N. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, New York, 1995.

[7] F. Lauer, C. Y. Suen, and G. Bloch, A trainable feature extractor for handwritten digit recognition, *Pattern Recognition,* 40(6): June 2007, 1816-1824.

[8] K. Mori, M. Matsugu, T. Suzuki, Face recognition using SVM fed with intermediate output of CNN for face detection, in: *Proceedings of the IAPR Conference on Machine Vision Applications*, Tsukuba Science City, Japan, May 2005, 224-229.

[9] Y.LeCun, L.Bottou, Y.Bengio, and P. Haffner, Gradient-based learning applied to document recognition, *Proceedings of the IEEE,* 86(11):2278-2324, November 1998.

[10] D. Bouchain. Character recognition using convolutional neural networks.*Institute for Neural Information Processing*, 2006.

[11] Vladimir N.Vapnik, The Nature of Statistical Learning Theory, *Spinger*, 38 (1996), pp.123-167.

[12] Y.Guermeur. Cmbining discriminant models with new multi-class SVMs, *Patt. Anal*. Appl. 5 (2002), 168-179.

[13] R Debnath, N Takahide, H.Takahashi. A decision based one-against-one method for multi-class support vector machine, *Patt. Anal*. Appl. 7(2), 2004, 164-175.

[14] B. Liu, Z. F. Hao, X. W. Yang. Nesting algorithm for multi-classification problems, *Soft Comput.* 11(4), 2007 383-389.

[15] C.W.Hsu, C.J. Lin, A comparision of methods for multi-class support vector machines, *IEEE Transactions on Neural Networks* , 13 (2002) , 415-425.

[16] S. Becker and Y.LeCun, Improving the convergence of back propagation learning with second-order methods, *Tech. Rep.* CRG-TR-88-5, University of Toronto Connectionist Research Group, September 1988.

[17] The MNIST Database of Handwritten Digits, http://yann.lecun.com/ exdb/ mnist/.

[18] D.Decoste, B.Scholkopf, Training invariant support vector machines, *Machine Learning* ,46 (2002) 161-190.