Handwritten Digit Recognition Based on Principal Component Analysis and Support Vector Machines

Rui Li and Shiqing Zhang

School of Physics and Electronic Engineering, Taizhou University
318000 Taizhou, China
{lirui,zhangshiqing}@tzc.edu.cn

Abstract. Handwritten digit recognition has always been a challenging task in pattern recognition area. In this paper we explore the performance of support vector machines (SVM) and principal component analysis (PCA) on handwritten digits recognition. The performance of SVM on handwritten digits recognition task is compared with three typical classification methods, i.e., linear discriminant classifiers (LDC), the nearest neighbor (1-NN), and the back-propagation neural network (BPNN). The experimental results on the popular MNIST database indicate that SVM gets the best performance with an accuracy of 89.7% with 10-dimensional embedded features, outperforming the other used methods.

Keywords: Handwritten digits recognition, Principal component analysis, Support vector machines.

1 Introduction

Handwritten digit recognition is an active topic in pattern recognition area due to its important applications to optical character recognition, postal mail sorting, bank check processing, form data entry, and so on.

The performance of character recognition largely depends on the feature extraction approach and the classifier learning scheme. For feature extraction of character recognition, various approaches, such as stroke direction feature, the statistical features and the local structural features, have been presented [1, 2]. Following feature extraction, it's usually needed to reduce the dimensionality of features since the original features are high-dimensional. Principal component analysis [3] is a fundamental multivariate data analysis method and widely used for reducing the dimensionality of the existing data set and extracting important information. The task of classification is to partition the feature space into regions corresponding to source classes or assign class confidences to each location in the feature space. At present, the representative statistical learning techniques [4] including linear discriminant classifiers (LDC) and the nearest neighbor (1-NN), and neural network [5], have been widely used for handwritten digit recognition. Support vector machines (SVM) [6] became a popular classification tool due to its strong generalization capability, which was successfully employed in various real-world applications. In the present study we employ PCA to extract the low-dimensional embedded data representations and explore the performance of SVM for handwritten digit recognition.

S. Lin and X. Huang (Eds.): CSEE 2011, Part I, CCIS 214, pp. 595–599, 2011.
© Springer-Verlag Berlin Heidelberg 2011

2 Principal Component Analysis

Principal Component Analysis (PCA) [3] is a basis transformation to diagonalize an estimate of the covariance matrix of the data set. PCA can be applied to represent the input digit images by projecting them onto a low-dimensional space constituted by a small number of basis images derived by finding the most significant eigenvectors of the covariance matrix.

In order to find a linear mapping M which maximizes the objective function ($trace(M^T \operatorname{cov}(X) M)$), PCA solves the following eigenproblem:

$$\operatorname{cov}(X) M = \lambda M \tag{1}$$

where $\operatorname{cov}(X)$ is the sample covariance matrix of the data X. The d principal eigenvectors of the covariance matrix form the linear mapping M. And then the low-dimensional data representations are computed by $Y = XM$.

3 Support Vector Machines

Support vector machines (SVM) [6] is based on the statistical learning theory of structural risk management and quadratic programming optimization. And its main idea is to transform the input vectors to a higher dimensional space by a nonlinear transform, and then an optimal hyperplane which separates the data can be found.

Given training data set $(x_1, y_1), \ldots, (x_l, y_l), y_i \in \{-1, 1\}$, to find the optimal hyperplane, a nonlinear transform, $Z = \Phi(x)$, is used to make training data become linearly dividable. A weight w and offset b satisfying the following criteria will be found:

$$\begin{cases} w^T z_i + b \geq 1, & y_i = 1 \\ w^T z_i + b \leq -1, & y_i = -1 \end{cases} \tag{2}$$

We can summarize the above procedure to the following:

$$\min_{w,b} \Phi(w) = \frac{1}{2}(w^T w) \tag{3}$$

Subject to $y_i(w^T z_i + b) \geq 1, \quad i = 1, 2, \ldots, n$

If the sample data is not linearly dividable, the following function should be minimized.

$$\Phi(w) = \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i \tag{4}$$

whereas ξ can be understood as the error of the classification and C is the penalty parameter for this term.

By using Lagrange method, the decision function of $w_0 = \sum_{i=1}^{l} \lambda_i y_i z_i$ will be

$$f = \operatorname{sgn}[\sum_{i=0}^{l} \lambda_i y_i (z^T z_i) + b] \tag{5}$$

From the functional theory, a non-negative symmetrical function $K(u,v)$ uniquely defines a Hilbert space H , where K is the rebuild kernel in the space H :

$$K(u,v) = \sum_i \alpha \varphi_i(u) \varphi_i(v) \tag{6}$$

This stands for an internal product of a characteristic space:

$$z_i^T z = \Phi(x_i)^T \Phi(x) = K(x_i, x) \tag{7}$$

Then the decision function can be written as:

$$f = \mathrm{sgn}[\sum_{i=1}^{l} \lambda_i y_i K(x_i, x) + b] \tag{8}$$

The development of a SVM emotion classification model depends on the selection of kernel function. There are several kernel functions, such as linear, polynomial, radial basis function (RBF) and sigmoid, that can be used in SVM models.

4 Experiment Study

4.1 MNIST Database

The popular MNIST database of handwritten digits, which has been widely used for evaluation of classification and machine learning algorithms, is used for our experiments. The MNIST database of handwritten digits, available from the web site: http://yann.lecun.com/exdb/mnist, has a training set of 60000 examples, and a test set of 10000 examples. It is a subset of a larger set available from NIST. The original black and white images from NIST were size normalized to fit in a 20×20 pixel box while preserving their aspect ratio. The images were centered in a 28×28 image by computing the center of mass of the pixels, and translating the image so as to position this point at the center of the 28×28 field. In our experiments, for computation simplicity we randomly selected 3000 training samples and 1000 testing samples for handwritten digits recognition. Some samples from the MNIST database are shown in Fig.1.

4.2 Experimental Results and Analysis

To verify the performance of SVM on handwritten digits recognition task, three typical methods, i.e., linear discriminant classifiers (LDC), the nearest neighbor (1-NN) and the back-propagation neural network (BPNN) as a representative neural network were used to compare with SVM. For BPNN method, the number of the hidden layer nodes is 30. We employed the LIBSVM package, available at http://www.csie.ntu.edu.tw/cjlin/libsvm, to implement SVM algorithm with RBF kernel, kernel parameter optimization, one-versus-one strategy for multi-class classification problem. The RBF kernel was used for its better performance compared with other kernels. For simplicity, the feature dimension of the original grey image features (28×28=784) is reduced to 10 as an illustration of evaluating the performance of SVM.

Fig. 1. Some samples from the MNIST database

Table 1. Handwritten Digits Recognition Results with 10-Dimensional Embedded Features

Methods	LDC	1-NN	BPNN	SVM
Accuracy (%)	77.6	82.7	84.8	89.7

Table 2. Confusion matrix of Handwritten Digits Recognition Results with SVM

Digits	0	1	2	3	4	5	6	7	8	9
0	**90**	0	0	0	0	5	0	0	1	0
1	0	**110**	1	0	0	0	0	0	4	0
2	0	0	**84**	1	2	0	1	1	0	0
3	0	0	5	**108**	0	3	0	0	7	0
4	0	0	3	0	**69**	1	1	0	1	12
5	2	0	3	2	2	**88**	0	0	1	1
6	2	0	0	0	1	0	**84**	0	1	0
7	0	2	1	0	1	0	0	**101**	0	6
8	2	0	2	2	0	4	1	0	**75**	3
9	0	2	0	0	6	1	0	2	4	**88**

Table 1 presents the different recognition results of four classification methods including LDC, 1-NN, BPNN as well as SVM. From the results in Table 1, we can observe that SVM performs best, and achieves the highest accuracy of 89.7% with 10-dimensional embedded features, followed by BPNN, 1-NN and LDC. This demonstrates that SVM has the best generalization ability among all used four classification methods. In addition, the recognition accuracies for BPNN, 1-NN and LDC, are 84.8%, 82.7% and 77.6%, respectively.

To further explore the recognition results of different handwritten digits with SVM, the confusion matrix of recognition results with SVM is presented in Table 2. As shown in Table 2, we can see that three digits, i.e., "1", "3" and "7", could be discriminated well, while other digits could be classified poor.

5 Conclusions

In this paper, we performed reduction dimension with PCA for the grey digits image features and explored the performance of four different used classification methods, i.e., LDC, 1-NN, BPNN and SVM, for handwritten digits recognition from the popular MNIST database. The experimental results on the MNIST database demonstrate that SVM can achieve the best performance with an accuracy of 89.7% with 10-dimensional reduced features, due to its good generalization ability. In our future work, it's an interesting task to study the performance of other more advanced dimensionality reduction techniques than PCA on handwritten digits recognition.

Acknowledgments. This work is supported by Zhejiang Provincial Natural Science Foundation of China (Grant No. Y1111058).

References

1. Trier, O.D., Jain, A.K., Taxt, T.: Feature extraction methods for character recognition—a survey. Pattern Recognition 29(4), 64–662 (1996)
2. Lauer, F., Suen, C.Y., Bloch, G.: A trainable feature extractor for handwritten digit recognition. Pattern Recognition 40(6), 1816–1824 (2007)
3. Partridge, M., Calvo, R.: Fast dimensionality reduction and simple PCA. Intelligent Data Analysis 2(3), 292–298 (1998)
4. Jain, A.K., Duin, R.P.W., Mao, J.: Statistical pattern recognition: a review. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(1), 4–37 (2000)
5. Kang, M., Palmer-Brown, D.: A modal learning adaptive function neural network applied to handwritten digit recognition. Information Sciences 178(20), 3802–3812 (2008)
6. Vapnik, V.: The nature of statistical learning theory. Springer, New York (2000)