

SMO algorithm of SVM and Naïve Bayes for MNIST handwritten digit recognition

Module: Artificial Intelligence

Name: Qinsong Wang

Student ID: 2106475

Date: 2022.01.04

chapter1 Introduction

Nowadays, with the rapid development of information technology and multimedia technology, traditional paper text can no longer meet the increasing needs of mankind, so we urgently need to convert traditional paper text information into digital information. At present, printed paper characters are relatively mature, but handwritten characters are more complicated to recognize due to different writing forms, and further research is needed. The handwritten Arabic numerals are frequently used in postal codes, test papers, and banknotes. The particularity of this use determines that the application requires extremely high recognition accuracy. These data consume a lot of manpower and material resources and also require 100% accuracy. The manual operation has many drawbacks, not only inefficient but also difficult to guarantee accuracy and safety.

Handwritten number recognition is a traditional number recognition technology. Digital recognition technology is a type of pattern recognition. Pattern recognition is the use of computer algorithm technology to automatically process and interpret graphics. We collectively refer to the environment and objects as "models." With the development of computer technology, people can study complex information processing. An important form of information processing is the biological recognition of the environment and objects. For humans, it is particularly important to identify optical information (obtained through visual organs) and harmonic information (obtained through auditory organs) ^[1]. These are two important aspects of pattern recognition. The representative products on the market are optical character recognition and speech recognition systems. Number recognition belongs to the category of optical character recognition. There are only ten types of numbers, the strokes are simple, and it is not very difficult to recognize them. But in fact, some test results show that the correct recognition rate of numbers is not as high as that of printed Chinese characters. The main reasons are: first, the strokes of the numbers are simple and smooth, with little difference in font shapes, and it is very difficult to accurately distinguish certain numbers; second, although there are only ten categories and simple strokes, there are many differences in the writing of a number. People from all over the world are using it, and their writing has obvious regional characteristics. Therefore, researching high-performance handwritten digit recognition algorithms is a challenging task

chapter 2 Research status

2.1 Feature extraction technology

Feature extraction methods for handwritten digits began more than 40 years ago until now. Researchers have proposed many recognition methods and techniques. Generally speaking, they can basically be summarized into two categories, namely, statistical recognition methods and structure recognition methods. , These two methods have their advantages and disadvantages. The statistical recognition method ^[2] is the overall statistical analysis of character information, mainly including pixel density features, template matching, mathematical transformation, and moment features. The

statistical recognition method is relatively stable, insensitive to noise, but cannot detect characters. The strokes change, so the probability of misclassification is relatively large. The structure recognition method [3] is to extract the basic information that best represents the shape of the character from the skeleton or outline of the character, such as strokes, depressions, circles, intersections, endpoints, nodes, arcs, etc. This method can detect characters. The subtle changes in strokes are sensitive to noise. Handwritten numbers are often prone to burrs and broken pens, which seriously affect the recognition results. Therefore, this method has a low recognition rate for images with complex backgrounds. At present, the overall trend of feature extraction technology research is the fusion of structure recognition and statistical recognition methods. Many researchers have also tried various improvements and proposed many fusion methods and methods, such as local and global feature fusion [4], character outline features and statistical features [5], skeleton features and coarse grid features [6], etc.

2.2 Classifier design

In terms of classifier design, researchers in recent years have worked hard to apply new knowledge to the recognition of handwritten numbers, such as artificial neural networks, support vector machines, Bayes, decision trees, K nearest neighbor algorithms, etc., among which artificial neural network recognition methods [7] Has been widely used, and currently the most widely used in character recognition is the multi-layer perceptron network. The reason why the neural network is very suitable for dealing with character recognition and classification problems is because of its learning ability. Neural network recognition methods have become mature in solving the recognition problems of small character sets, but when the character set to be recognized is large, The recognition effect is not ideal, and the reason is still restricted by the current recognition technical conditions. At present, the research of the neural network recognition method mainly focuses on two aspects of improvement. One is to study and improve the network structure of the neural network, and the other is to study and improve the corresponding learning algorithm.

In addition, after Vapnik proposed the statistical learning theory, the SVM theory quickly gained widespread attention. Since statistical learning theory can better solve the problem of linear inseparability, SVM has a wide range of applications in regression estimation, pattern recognition, probability density function estimation, etc. [8-10]. At present, the performance of SVM in the application has been better than some known neural network methods, and SVM has also shown good promotion ability in many fields. Many researchers have carried out in-depth research on it, but SVM is used in handwritten digit recognition. There are also problems such as long training time and slow recognition speed in the application

chapter 3 The theoretical basis of support vector machines

3.1 Overview OF SVM

Support Vector Machine (SVM) is abbreviated as SVM, which is the latest content in statistical learning theory and the most practical part. Support vector machine is a new machine learning technology, proposed by Vapnik and his colleagues in 1995. The support vector machine method is based on the VC dimension theory of statistical learning theory and the principle of structural risk minimization. According to the limited sample information, the complexity of the model (that is, the learning accuracy of a specific training sample) and the learning ability (that is, no Seek the best compromise between the ability to incorrectly identify any sample) to obtain the best promotion ability.

3.2 Introduction of SVM

3.2.1 SVM in the case of linearly separable

If a linear function (such as a straight line in a two-dimensional space, a plane in a three-dimensional space, and a hyperplane in a higher-dimensional space) can completely separate the two types of samples, then these samples are linearly separable. ^[11], on the contrary, if we cannot find a linear function to completely separate the two types of samples, then we call these samples non-linearly separable.

Given a linearly separable data set $\{(\vec{x}_1, y_1), (\vec{x}_2, y_2), \dots, (\vec{x}_N, y_N)\}$ Sample feature vector

$\vec{x} \in R^D$, $\vec{x}$ is a Vector in D-dimensional space, class label $y \in \{-1, +1\}$ That is, there are only two types of samples. At this time, the samples with the class label +1 are usually called positive examples, and the samples with the class label -1 are called negative examples. There are only two types of samples. At this time, the samples with the class label +1 are usually called positive examples, and the samples with the class label -1 are called negative examples. The problem is to correctly classify the two types of samples, then the goal is to find the optimal classification hyperplane. The hyperplane with the maximum classification interval can be determined according

to the training samples. We set the optimal hyperplane equation as $\vec{w}^T \vec{x} + b = 0$ According to the distance formula from the point to the plane, the distance between the sample $\vec{x}$ and the best

hyperplane $(\vec{w}, b)$ can be calculated as $\frac{|\vec{w}^T \vec{x} + b|}{\|\vec{w}\|}$ Then normalizing the hyperplane and selecting

the closest sample to the hyperplane $\vec{x}_k$. Satisfy the $\vec{w}$ and b of the formula $\left| \vec{w}^T \vec{x}_k + b \right| = 1$, thus getting the normalized plane.

At this time, the distance from the nearest sample to the edge is expressed as:

$$\frac{\left| \vec{w}^T \vec{x} + b \right|}{\left\| \vec{w} \right\|} = \frac{1}{\left\| \vec{w} \right\|}$$

And the classification interval becomes:

$$m = \frac{2}{\left\| \vec{w} \right\|} \quad (3-8)$$

The goal is to find the normal vector w that maximizes formula (3-8), and then substitute it into formula a to obtain the value of b . Maximizing formula (3-8) is equivalent to minimizing:

$$J(\vec{w}) = \frac{1}{2} \left\| \vec{w} \right\|^2 \quad (3-9)$$

In addition, there are constraints as follows

$$y_i \left(\vec{w}^T \vec{x}_i + b \right) \geq 1, \forall i \in (1, 2, \dots, N) \quad (3-10)$$

The sample point $\vec{x}_k$ closest to the hyperplane satisfies the relationship $\vec{w}^T \vec{x} + b = 1$, but the distance $d(\vec{x}_i)$ of the other remaining sample points $\vec{x}_i$ from the hyperplane must be greater than or equal to the distance $d(\vec{x}_k)$ from $\vec{x}_k$ to the hyperplane, so:

$$\left| \vec{w}^T \vec{x}_i + b \right| \geq 1$$

By multiplying each constraint condition in formula (3-10) by a Lagrangian multiplier α_i , and then substituting it into formula (3-9), then the extreme value problem of this condition can be transformed into the following Constrained optimization problem, that is, about $\vec{w}$, b , and $\alpha_i, (i=1, 2, \dots, N)$ to minimize L .

$$L(\vec{w}, b, a) = \frac{1}{2} \left\| \vec{w} \right\|^2 - \sum_{i=1}^N \alpha_i [y_i (\vec{w}^T \vec{x}_i + b) - 1], \alpha_i \geq 0 \quad (3-12)$$

Find the partial derivatives of functions $\vec{w}$, b , and $\alpha_i, (i=1, 2, \dots, N)$ respectively, and make the partial derivatives equal to 0, you can calculate:

$$\frac{dL(\vec{w}, b, \alpha)}{d\vec{w}} = 0 \Rightarrow \vec{w} = \sum_{i=1}^N \alpha_i y_i \vec{x}_i \quad (3-13)$$

$$\frac{dL(\vec{w}, b, \alpha)}{db} = 0 \Rightarrow \sum_{i=1}^N \alpha_i y_i = 0 \quad (3-14)$$

Expand the formula (3-12), we get:

$$L(\vec{w}, b, a) = \frac{1}{2} \vec{w}^T \vec{w} - \sum_{i=1}^N \alpha_i y_i \vec{w}^T \vec{x}_i - b \sum_{i=1}^N \alpha_i y_i + \sum_{i=1}^N \alpha_i \quad (3-15)$$

Then substituting formula (3-13) and formula (3-14) into formula (3-15), we get:

$$\begin{aligned} L(\vec{w}, b, a) &= \frac{1}{2} \vec{w}^T \left(\sum_{i=1}^N \alpha_i y_i \vec{x}_i \right) - \sum_{i=1}^N \alpha_i y_i \vec{w}^T \vec{x}_i - 0 + \sum_{i=1}^N \alpha_i \\ &= -\frac{1}{2} \left(\sum_{i=1}^N \alpha_i y_i \vec{w}^T \vec{x}_i \right) + \sum_{i=1}^N \alpha_i \\ &= -\frac{1}{2} \left(\sum_{i=1}^N \alpha_i y_i \left(\sum_{j=1}^N \alpha_j y_j \vec{x}_j \right)^T \vec{x}_i \right) + \sum_{i=1}^N \alpha_i \\ &= -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \vec{x}_i^T \vec{x}_j + \sum_{i=1}^N \alpha_i \end{aligned}$$

It can be seen that the above formula is not related to $\vec{w}$ and b , it is only a function of α_i , so it can be written as formula (3-16), and its constraint is formula (3-17):

$$L(\vec{w}, b, a) = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \vec{x}_i^T \vec{x}_j + \sum_{i=1}^N \alpha_i \quad (3-16)$$

$$\alpha_i \geq 0, \sum_{i=1}^N \alpha_i y_i = 0 \quad (3-17)$$

This is a Lagrangian dual problem, and the dual problem is a convex quadratic programming problem, which can be solved by some standard optimization techniques. After solving the parameters $\vec{w}$ and b of the maximum margin splitting the hyperplane can be determined by the solution a of the dual problem:

$$\vec{w} = \sum_{i=1}^N \alpha_i y_i \vec{x}_i$$

Under the condition that the linear sample is separable, since the relation $\left| \vec{w}^T \vec{x}_k + b \right| = 1$, where

$\vec{x}_k$ is any vector that is closest to the optimal classification hyperplane, one of the $\vec{x}_k$ that can

make $\left| \vec{w}^T \vec{x}_k + b \right|$ obtain the minimum value, So we can substitute $\vec{w}$ and $\vec{x}_k$ into the above formula to find b :

$$b = 1 - \min_{y_i=+1}(\vec{w} \cdot \vec{x}_i), \text{ or } b = -1 - \max_{y_i=-1}(\vec{w} \cdot \vec{x}_i)$$

Under more general conditions, since the closest distance to the segmentation hyperplane (w, b), in the two types of samples is no longer necessarily 1 and may be different, b is:

$$b = -\frac{1}{2}(\min_{y_i=+1}(\vec{w} \cdot \vec{x}_i) + \max_{y_i=-1}(\vec{w} \cdot \vec{x}_i))$$

Equation (3-20) includes the linearly time-sharing case. According to the properties of the optimized solution (Karush-Kuhn-Tucker conditions), the solution α must satisfy:

$$\alpha_i \left[y_i \left(\vec{w}^T x_i + b \right) - 1 \right] = 0, \forall i = 1, 2, \dots, N$$

3.2.2 SVM in the case of non-linearity requiring kernel function mapping

The classification ability of linear classifiers is generally very limited, and the continuous relaxation of constraints when dealing with nonlinear problems will lead to an increase in the sample error rate. At this time, we can transform the problem into a high-dimensional space through nonlinear transformation. For linear problems, find the best classification hyperplane in the transformed space.

The kernel function is a symmetric function: $K : R^n \times R^n \rightarrow R$ It maps the n -dimensional vector in two R spaces to a real number, and the Mercer kernel function calculates the dot product in the high-dimensional space:

$$K(x_i, x_j) = \psi(x_i) \cdot \psi(x_j), \psi : R^n \rightarrow R^D$$

In this way, if the Mercer kernel function for calculating the dot product can be found in the feature space, the kernel function can be used instead of the dot product operation in the support vector machine, and there is no need to care about the specific form of the nonlinear mapping ψ at all, because in the SVM training and all relevant formulas in the classification, ψ has not appeared alone, and always appears in the form of $\psi(x_i) \cdot \psi(x_j)$.

Therefore, a reasonable selection of the inner product kernel function K is essential to realize the mapping from low-dimensional space to high-dimensional space. Through this method, we can easily solve the nonlinear classification problem and transform it into solving the linear classification problem, and its calculation is complicated. The degree does not increase accordingly. At this time, the optimization objective function is the formula (3-32), and the decision (classification) function also becomes the formula (3-33):

$$L(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (3-32)$$

$$h(x) = \text{sgn} \left(\sum_{i \in SV} \alpha_i^* y_i K(x_i, x_j) + b^* \right) \quad (3-33)$$

The commonly used kernel functions and their respective expressions are shown in the following formulas:

Linear kernel function $K(x, y) = x^T \cdot y + c$ c : constant

Polynomial kernel function $K(x, y) = (\alpha x^T \cdot y + c)^d$ α :slope, c :constant

Radial basis kernel function $K(x, y) = \exp(-\gamma \|x - y\|^2)$ γ : $1/n_features$

Sigmoid kernel function $K(x, y) = \tanh(\alpha x^T \cdot y + c)$ $\alpha = 1/N$, N is data dimension c : constant

3.3 SMO algorithm

Sequential minimal optimization, referred to as the SMO algorithm, is currently a very effective method for solving the problem of support vector machine training under a large amount of data. SMO is proposed by Platt to solve the problem of large training samples. There are only two samples in the working set of the algorithm. Its advantage is that the quadratic programming problem for two samples can have an analytical solution form, thereby avoiding the instability and time-consuming problem of numerical solutions in the case of multi-samples, and it does not require large The matrix storage space is especially suitable for sparse samples, and its working set selection is also unique, not the traditional steepest descent method, but a heuristic. Two nested loops are used to find the sample variables to be optimized. Look for samples that violate KKT's optimal conditions in the outer loop, and then select another sample in the inner loop to complete optimization. Recirculate and perform the next optimization until all samples meet the optimal conditions. The SMO algorithm is mainly time-consuming to judge the optimal condition, so the optimal condition discriminant that is the most reasonable, that is, the lowest computational cost, should be found, and the commonly used parameters should be cached.

The SMO algorithm consists of two main parts: one is the analytical method to optimize the two Lagrangian coefficients, and the other is the heuristic algorithm to find these two Lagrangian coefficients. The basic idea is: if the solutions of all variables satisfy the KKT condition of this optimization problem, then the solution of this optimization problem is obtained.

The SMO algorithm is summarized as follows:

(1) Calculation error

$$E_i = \left(\sum_{j=1}^N \alpha_j y_j K(x_j, x_i) + b \right) - y_i, i = 1, 2, \dots, N$$

(2) Calculate the upper and lower boundaries L, H

$$\begin{cases} L = \max(0, \alpha_2^{old} - \alpha_1^{old}), H = \min(C, C + \alpha_2^{old} - \alpha_1^{old}), y_1 \neq y_2 \\ L = \max(0, \alpha_2^{old} + \alpha_1^{old} - C), H = \min(C, \alpha_2^{old} + \alpha_1^{old}), y_1 = y_2 \end{cases}$$

(3) Calculate η

$$\eta = K_{11} + K_{22} - 2K_{12} = \|\phi(x_1) - \phi(x_2)\|^2$$

(4) update α_2

$$\alpha_2^{new,unc} = \alpha_2^{old} + \frac{y_2(E_1 - E_2)}{\eta}$$

(5) Modify according to the value range α_2

$$\alpha_2^{new} = \begin{cases} H, \alpha_2^{new,unc} > H \\ \alpha_2^{new,unc}, L \leq \alpha_2^{new,unc} \leq H \\ L, \alpha_2^{new,unc} < L \end{cases}$$

(6) update α_1

$$\alpha_1^{new} = \alpha_1^{old} + y_1 y_2 (\alpha_2^{old} - \alpha_2^{new})$$

(7) update $b_1^{new}, b_2^{new}, b^{new}$

$$\begin{aligned} b_1^{new} &= -E_1 - y_1 K_{11}(\alpha_1^{new} - \alpha_1^{old}) - y_2 K_{21}(\alpha_2^{new} - \alpha_2^{old}) + b^{old} \\ b_2^{new} &= -E_2 - y_1 K_{12}(\alpha_1^{new} - \alpha_1^{old}) - y_2 K_{22}(\alpha_2^{new} - \alpha_2^{old}) + b^{old} \\ b^{new} &= \frac{b_1^{new} + b_2^{new}}{2} \end{aligned}$$

chapter 4 Overview of Naive Bayes

Naive Bayes is a classification algorithm. The algorithm is a classification method based on Bayes' theorem and the assumption of the independence of characteristic conditions. Among them, the Bayesian algorithm is based on the Bayesian principle and uses the knowledge of probability statistics to classify the sample data set. Due to its solid mathematical foundation, the misjudgment rate of the Bayesian classification algorithm is very low. The characteristic of the Bayesian algorithm is the combination of prior probability, and posterior probability, which avoids the supervisor's bias of using only the prior probability, and also avoids the overfitting phenomenon of using sample information alone. The Bayesian classification algorithm shows a higher accuracy rate when the data set is large, and the algorithm itself is relatively simple. The Naive Bayes algorithm is based on the Bayes algorithm and is simplified, that is, it is assumed that the attributes are conditionally independent of each other when the target value is given. That is to say, no attribute variable has a larger proportion to the decision result, and no attribute variable has a smaller proportion to the decision result. Although this simplified method reduces the classification effect of the Bayesian classification algorithm to a certain extent, in actual application scenarios, it greatly simplifies the complexity of the Bayesian method.

chapter 5 The realization process of handwritten digit recognition algorithm

5.1 MNIST Data set extraction

MNIST database There are four files in the data set of (Handwritten Character Recognition), which are as follows:

train-images-idx3-ubyte.gz, Training set, a total of 60,000 (28*28) image data;

train-labels-idx1-ubyte.gz, The label information of the training set (values are 0-9), 60,000*1;

t10k-images-idx3-ubyte.gz, Test set, a total of 10,000 (28*28) image data;

t10k-labels-idx1-ubyte.gz, The label information of the test set (values are 0-9), 10,000*1。

The ubyte in the file name represents the data type. The unsigned single-byte type corresponds to the uchar data type in Matlab. For the label file, there are two 32-bit integers, and for the image file, there are four 32-bit integers. Therefore, you need to move the file pointers to the two files to point to the correct position.

As you can see, the core of the above two scripts is to get the file handle fid through fopen first, then offset to get the magic number (magic) and some other information, and finally get all the data, and reshape the data to the corresponding dimensions (The pictures are all 28x28 in size). The file is a binary file, and the first offset is a magic number for verification. The image is 2051 and the label is 2049. The offset after that is the number of pictures, the number of rows, and the number of columns. After that is the value of the pixel. The value of the pixel is 0 from 0 to 255. 0 (white) as the background, 255 (black) as the foreground。

5.2 Implementation of SVM handwriting recognition algorithm

The first step is to prepare the extracted MNIST data set, determine the training data set and the test data set;

The second step, determine the amount of information in the training data set and the test data set and the size of the picture information;

The third step, create the covariance Matrix and perform covariance decomposition to obtain the projection matrix;

The fourth step is to use the projection matrix to reduce the dimensionality of the image information, The fifth step is to select the parameters of the SVM training model,

The sixth step is to use the trained SVM model to The test data is recognized, the recognition result is compared with the real result of the mark, and the recognition accuracy is counted.

5.3 Implementation of Bayes classify algorithm

(1) Enter the feature vector dataset $X = (x_1, \dots, x_m)$ and category vector $Y = (y_1, \dots, y_k)$ of the

unknown category;

(2) Calculate $P(y_j)$ and $P(x_i | y_j)$, a total of k groups;

(3) Use the Bayesian formula to calculate $P_j = P(x_i | y_j) \cdot P(y_j)$. This is because $P(x_i)$ is a constant, so $P(y_j | x_i) \propto P(x_i | y_j) \cdot P(y_j)$;

(4) Take the largest one among $P_s = \max\{P_j\}, j = 1, 2, \dots, k$, and the corresponding classification of x_i is y_s ;

(5) Loop the above steps until all the classifications of x_i are output, and the end.

5.4 Handwritten digit recognition process and results

5.4.1 Naive Bayes

It adds the Laplace correction which can greatly smooth the probabilistic function. It tests the Naive Bayes model using the cross-validation method and the final accuracy is 84.92%.

Bayes													
真实类	0	887		3	7	2	40	17	1	23		90.5%	9.5%
	1		1085	10	5		9	6		19	1	95.6%	4.4%
	2	19	8	855	27	15	4	33	14	55	2	82.8%	17.2%
	3	5	15	34	844		13	9	18	46	26	83.6%	16.4%
	4	2	6	4		793	6	20	1	24	126	80.8%	19.2%
	5	25	12	7	127	29	630	17	7	19	19	70.6%	29.4%
	6	18	18	15	2	12	35	852		6		88.9%	11.1%
	7	1	24	14	4	15			870	27	73	84.6%	15.4%
	8	17	23	12	74	17	23	7	6	759	36	77.9%	22.1%
	9	9	13	5	9	72	7		23	23	848	84.0%	16.0%

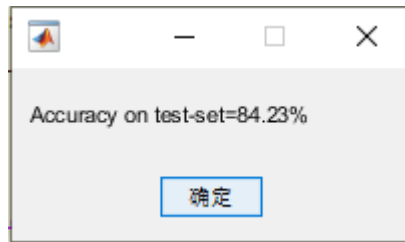


Figure 2 Accuracy on test-set

5.4.2 SMO algorithm of SVM

	1	2	3	4	5	6	7	8	9	10
1	980	979	979	976	980	982	976	979	981	974
2	1136	1135	1131	1132	1134	1131	1136	1136	1129	1135
3	1031	1029	1032	1030	1036	1032	1028	1028	1026	1025
4	1014	1013	1008	1010	1008	1009	1011	1010	1006	1003
5	982	981	986	984	982	981	979	985	983	985
6	890	896	892	891	891	892	890	893	891	881
7	954	959	962	959	959	960	958	958	954	958
8	1029	1029	1022	1028	1031	1027	1028	1028	1028	1024
9	973	979	980	971	973	972	970	973	974	985
10	1014	1009	1016	1017	1006	998	1009	1014	997	1009

test num.mat

Run the code will generate a sheet, the num1 to 10 means '0' to '9', the data, for example, '(1,2); 979' means using the SMO algorithm classifies the digits '0', and digits '1'. The final identification is the number of the latter digits '1', which means the total of '0' and '1' is 980 plus 1135 equals 2115 and find 979 digits '0'.

the total accuracy rate:

1	99.8000
2	99.7900
3	99.6900
4	99.7600
5	99.8200
6	99.7700
7	99.8300
8	99.8400
9	99.6600
10	99.4900

accuracy on test-set

chapter 6 Conclusion

This article mainly uses two methods of support vector machine and Naive Bayes to recognize and process the 0-9 handwritten digits. From the results obtained through programming and calling the toolbox, the recognition effects are different, which may be related to some parameters. The selection is related.

As a classic branch of pattern recognition, digital recognition has great application value. The research of digital recognition will help the development of pattern recognition, robotics, and how to better carry out human-computer interaction. It is of great theoretical value to make computers have the same abilities as humans.

Reference

- [1] Xu Qinzhen Yang Luxi Handwritten Arabic numeral recognition method based on improved hybrid learning model [J] . Journal of Electronics and Information 2010, 32(2): 4-10
- [2] Zhou Qi. Research on Offline Handwritten Character Recognition Based on Support Vector Machine [D]. Chongqing University, 2017.
- [3] Zhu Hui, Yang Yang, Jie Bin, etc. Research on the application of SVM in the recognition of handwritten Chinese characters in small character sets [J]. Microcomputer Letter, 2004, 20(4): 74-88.
- [4] Xia Guoen, Jin Weidong, Zhang Gexiang. Handwritten digit recognition method based on combined features[J]. Computer Application Research, 2016, (6): 170-172.
- [5] Zhang Chongyang, Lou Zhen, Yang Jingyu. Handwritten digit recognition based on contour features and statistical features[J]. Computer Engineering and Applications, 2014, (2): 82-84
- [6] Yuan H. Research on Handwritten Digit Recognition Technology Based on Feature Fusion and Neural Network [D]. Wuhan University of Science and Technology, 2017.
- [7] Zhao Xiaojuan. Research on the recognition of handwritten numbers and English characters [D]. Northeast Normal University, 2010.
- [8] Wang Hongman, Ou Zongying. Research on Face Recognition Method Based on Support Vector Machine [J]. Small Microcomputer System, 2014, 25(1): 139-142.
- [9] Cheng Xueyun, Ji Genlin, Ling Xiaohan. Multi-class cost-sensitive learning based on SVM and its application [J]. Journal of Nanjing Normal University: Engineering Technology Edition, 2016, 6(4): 79-82.
- [10] Peng Lu. Research and Application of Support Vector Machine Classification Algorithm [D]. Hunan University, 2018.
- [11] C. Cortes, V. Vapnic. Support vector network [J]. Machine Learning, 1995,20(3):273-297.