Paper Currency Denomination Recognition Based on GA and SVM

Jian-Biao He, Hua-Min Zhang(✉), Jun Liang, Ou Jin, and Xi Li

Computer Science and Technology Department, College of Information Science and Engineering, Central South University, Changsha 410083, China
huaminzhang@csu.edu.cn

Abstract. SVM is a new general learning method based on the statistic learning system which can be used as an effective means to process small sample, non-linear and high dimensional pattern recognition. This paper did research on the learning algorithm of support vector machine, extracted characteristic data of banknote which is on account of PCA according to the characteristics of the support vector machine (SVM), and proposed to put support vector machine (SVM) into banknotes denomination recognition by combining SMO training algorithm with one-to-many multi-value classification algorithm. Besides, this article used genetic algorithm in parameters optimization such as the punishment coefficient C of soft margin SVM and the width parameter of Gaussian kernel function. The ultimate purpose is to recognize the denomination of banknote efficiently and accurately. The experimental results verified that this kind of recognition method increases the recognition accuracy up to 90% or more.

Keywords: SVM banknote denomination recognition · GA Gaussian kernel function · Punishment coefficientc

1 Introduction

The content of banknote recognition research includes banknote denomination recognition, orientation recognition, authenticity recognition, serial number recognition, etc. The main purpose of this article is to seek out a solution to the banknote value recognition. Being a typical pattern recognition, the process of denomination recognition is made up with four steps, they are information acquisition, preprocessing, feature extraction and selection, and classification decision [1]. At present, there are three kinds of dominating ways to recognize the paper currency denomination, which are listed as following, (1) Notes denomination recognition method based on template matching. (2) Banknotes denomination recognition method based on neural network. (3) Notes denomination recognition method based on support vector machine (SVM). The shortcoming of template matching method is that the amount of internal storage and calculation is gigantic because of a large number of data processing. The anti-interference ability of template matching method is poor. As for the second method, a

Fund project: The National Natural Science Foundation of China under Grant No.61272147.

T. Tan et al. (Eds.): IGTA 2015, CCIS 525, pp. 366–374, 2015.
DOI: 10.1007/978-3-662-47791-5_41

lot of training sample is needed to neural network method in the way finding classification function, and the necessary characteristic quantity is excessive, which means complicated calculation is hard to avoid, and over-fitting is very easy to come about.

On the strength of structure risk minimization principle, support vector machine (SVM) that has stronger theoretical basis and better generalization was been brought up. In addition to using the theory of support vector machine (SVM), this article used the method based on PCA to get the banknote's features and did parameter optimization by adopting genetic algorithm to improve recognition accuracy. The experimental results confirm that this method is effective to avoid the "over-fitting" problem in neural network way, and the recognition accuracy is higher than BP, LVQ and PCA models apparently. In the case of searching optimization using the punishment coefficient C in GA algorithm and the width parameter of Gaussian kernel function, the recognition accuracy of banknote recognition is up above 90%.

2 Image Preprocessing

2.1 Edge Detection

The first step of preprocessing image is make image grizzled processing, and then do edge detection. Edges are important foundations of images include graphical segmentation, texture feature extraction and shape feature extraction, mainly exist between in target and target, target and background, region and region [2]. The edge detection of paper currency image is to separate the four edges of the banknote from background image accurately in the original figure collected, measure the four straight line of edges for paper currency tilt correction by Hough transformation, detect the edges using Sobel operator. The final image after edge detection is shown in Fig.2.

Fig. 1. The original image

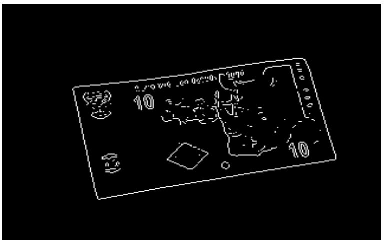

Fig. 2. The image after edge detection

2.2 Lean Correction

Because of the influences such as the angle between camera lens and paper currency while taking photos, as well as the swing of the camera lens, the tilt of paper currency image relative to the horizontal plane is likely to happen. It is necessary to do some adjustment and correction to the slant image of the banknote.Through edge detection, the paper outline is apparent whose four sides are all lines. Hough transformation is considered in detecting the straight lines. After that, do rotate correction of the angle of lines to horizontal direction. The final image profile is shown in Fig. 4.

Fig. 3. Image after lean correction

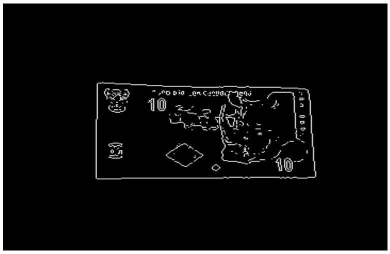

Fig. 4. The profile of image after lean correction

3 Paper Currency Recognition Based on SVM and GA

3.1 The Classification Model of SVM

The principle of SVM is separate the two types of sample points in plane correctly by classification hyper-plane method, and get maximum edge. That problem can be ascribed to the solution of a quadratic equation problem whose mathematical form is as Eq.1.

$$\text{Minimize} \quad \Phi(\omega, b) = \frac{1}{2}\|\omega\|^2 \tag{1}$$

The constraint is as Eq.2.

$$y_i\left(\omega^T * x_i + b\right) - 1 \geq 0 \quad i = 1,2,3,\ldots\ldots.n \tag{2}$$

The ω that meets the constraints condition is the normal vector of the optimal separating hyper-plane. The target function is strictly concave-up quadratic form while the constraint function is concave-down. It is a strictly convex programming. In accordance with the solution to a convex quadratic programming in the theory of optimum. That problem can be translated into Wolfe dual problem as Eq.3.

$$\text{Maximize} \sum_{i,j}^{n} \alpha_i \alpha_j y_i y_j x_i x_j \tag{3}$$

The constraint is as Eq.4

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \quad \alpha_i \geq 0 \quad i = 1,2,3,\ldots\ldots.n \tag{4}$$

α_i is Lagrange multiplier of sample points x_i. According to the Kuhn-Tunker condition, the Invalid constraints corresponding Lagrange multiplier is 0, only samples that

$\alpha_i \geq 0$ work for classification. Those are called support vector. Classification rules only decided by the minority of support vector which are just at the edge of hyper-plane, and has nothing to do with the other samples. However, There are many linearly inseparable situations under the actual situation. One solution proposed by Cortes&Vapnik1 who introduce non-negative slack law in the conditional expression, $\xi_i \geq 0$, $i = 1,2,3,.......n$, at this time, constraint becomes

$$y_i(\omega^T * x_i + b) \geq 1 - \xi_i \quad i = 1,2,3,......n \tag{5}$$

Classification hyperplane that allows wrong separating is called linearly soft interval classification hyperplane. Because it allows the existence of fault samples, the soft interval classification hyperplane at this moment is the largest classification interval of hyperplane after weeding out the wrong separated samples. Meanwhile, the target function turned into

$$\text{Minimize} \quad \Phi(\omega,\xi) = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}\xi_i \tag{6}$$

The penalty parameter C controls the penalty coefficient for wrong separated samples. The dual problem of linear soft interval classification hyperplane is the same as linear separable target function. The only difference is the constraints need to be translated into $0 \leq \alpha_i \leq C$.

In SVM, The parameter C controls the degree of punishment for wrong separated samples. The selection of C impact Generalization ability of SVM directly [3].

According to the pattern recognition theory, low dimensional space linearly inseparable mode might be linearly separable by nonlinear map to high-dimensional feature space, however, if use the technology to classify directly or regress in the high-dimensional space, there exist some problems such as the form, parameter and feature space dimension of certain nonlinear mapping function. The biggest obstacle is *the dimension disaster* when operating in high dimensional feature space. Using Kernel function technology can solve that problem effectively, avoid the problem of changing calculation in higher dimensional and greatly simplify the problem[4]. Assuming that x ,z∈X,X Belongs to R $^{(n)}$ space. The Nonlinear function Φ make the input X mapping in the feature space F. F belongs to R $^{(n)}$, n<<m. According to Kernel function technology as

$$K(x,z) = <\Phi(x),\Phi(z)> \tag{7}$$

$<\Phi(x),\Phi(z)>$ is inner product, $K(x,z)$ is Kernel function. As can be seen from Eq.7, Kernel function turns the inner product of high m dimensional space into Kernel function calculation of low n dimension input space, and solves *the dimension disaster* problem in high-dimensional feature space skillfully, lays the theoretical foundation solving complicated classification or regression problem in high-dimensional feature space. Accord*ing to the Hilbert-Schmidt* theory, a Kernel function K(x, z)which meets the Mercer conditions can be used as the inner product operation here.

At present, the more commonly used kernel functions are mainly three categories: (1) The linear kernel function. (2) Polynomial kernel function. (3) Gauss kernel function. The expression of Gauss kernel function is as Eq.8

$$K\left(\|x - x_c\|\right) = \exp\left\{ -\|x - x_c\|^2 \Big/ 2 * \sigma^2 \right\} \tag{8}$$

x_c is the center of Kernel function, σ is the Width parameter of function, controlling the radial range of action.

The training algorithm use SMO (Sequential minimal optimization) algorithm which can be said to be a special case of the decomposition algorithm whose work aggregation only have one sample. The advantage of training algorithm is that it have analytical solution form for quadratic programming problem of multiple samples, and avoid numerical instability and time-consuming problem in the case of problem having diversity samples, at the same time, the training algorithm doesn't need large matrix storage space, suits for sparse sample especially[5]. The selection of work aggregation is not traditional steepest descent method, but in heuristic way, looking for samples to be optimized by two nested loops, selecting another sample in the inner loop, completing the first optimization. Then recycle, and carry on the next optimization until all the samples are satisfied to the optimal condition. The algorithm mainly consumes time in optimal condition judgment. Therefore, looking for the most reasonable, i.e. minimal computing cost optimal condition is necessary.

Adopting one-versus-rest method (OVR SVMs). Turn a type of sample as a class successively during training, the other remaining samples belong to another class. In this way, *k* category samples constructed *k* SVM. Classify the unknown samples with maximum classification function of the class in the process of classification.

3.2 Feature Extraction and Selection Based on PCA

In banknote recognition, it would be best to find the key original feature subset, reduce unnecessary feature calculation and resources cost, rather than get original all feature mapping. Principal Components Analysis (PCA) is one of the more commonly used Dual Sensor algorithm recently. PCA map paper currency image to the feature space which set a good characterization of training images. The defect of PCA banknote recognition is that all the characteristics of the original space map to lower dimensional feature space, it is the feature subset based on best descriptive. This paper applied a new feature selection method based on PCA, which combine feature selection and feature extraction, carry on feature selection in feature space. The original features of selected paper money focus the most critical features[6].

PCA transform input data vector x_i into a new vector s_i by

$$s_i = U^T x_i \tag{9}$$

U is an orthogonal matrix . U_i, the i columns, is the i feature vectors of the sample Covariance
matrix $\sum$, writen as Eq.10 and Eq.11

$$\sum = \frac{1}{M}\sum_{i=1}^{M}(x_i - u)(x_i - u)^T \tag{10}$$

$$u = \frac{1}{M}\sum_{i=1}^{M} x_i \tag{11}$$

In the equation (10), x_i is the i th training sample images, u is the training sample mean vector, M is the total number of training samples. Making $A = [\Phi_1, \Phi_2, \Phi_3, \ldots, \Phi_M]$, $\Phi_i = x_i - u$. Covariance matrix can be expressed as Eq.12

$$\sum = A^T * A \tag{12}$$

In order to get the eigenvalue and orthogonal normalizing feature vector of $N * N$ dimension matrix (N is the characteristics dimensions of the sample) ∑, too large amount of calculation is needed if doing calculate directly. Nonzero eigenvalues and nonzero eigenvalues corresponding eigenvectors of A ^{T} A could be gotten by solving nonzero eigenvalues and nonzero eigenvalues corresponding eigenvectors of AA ^{T} .

Here obtaining the matrix composed by nonzero eigenvalues of ∑, $(\lambda_1, \lambda_2, \ldots\ldots \lambda_r)$, λ_i ($i = 1,2\ldots\ldots r$).. it is arranged by descending order. r is the order of A ^{T} A or AA ^{T} . p_i ,the corresponding eigenvectors of λ_i , makes up unit of orthogonal matrix $P = [p_1, p_2, \ldots\ldots p_r] \subset R^{N*r}$. Pack up the front k vectors of $P = [p_1, p_2, \ldots\ldots p_r] \subset R^{N*r}$ and get the feature space $U = [u_1, u_2, \ldots\ldots u_k] \subset R^{N*k}$, calculate projection in the feature space of a picture x_i consequently: $s_i = U^T x_i$.

3.3 Optimization of Parameter C and σ

Relative to the choice of kernel function, to determine the parameters of the kernel function is more important. After choosing Gaussian kernel function, it is need to make sure σ(the width of the Gaussian radial basis function).σ is very sensitive to the performance of classifier [7]. Making Analysis on the performance of the Gaussian kernel support vector machine (SVM) while taking different σ. It is found that all of the training sample points are support vectors if 0→σ, and all of them can be classified correctly. However, the phenomenon of "overlearning" appears very easily, its generalization ability is poor, and the rate of false recognition of test sample is relativly high. If ∞→σ, Gaussian kernel support vector machine (SVM) consider all samples as the same, both generalization ability and the correct recognition rate of the test samples are zero. In fact, when σ get the value smaller than the average distance between the training sample points very much, the result of 0→σ can be achieved. On the contrary, when σget the value larger than the average distance between the training sample points very much, the result of ∞→σ can be achieved [8].

In this article, the genetic algorithm is used to search out the most appropriate function parameters. In the view of biological genetics, genetic algorithm (GA) integrates the concept of the survival of the fittest and the idea of random information exchange. The evolution of the population is realized by mechanisms such as natural selection, chiasma and variation. In the process of optimization, the genetic algorithm produce multiple starting points in the solution space, do searching at the same time. It is a search algorithm whose search dierction is guided by the fitness function that is able to seek for the most optimized global solution in a complex search space.

In determining the width σ adopting genetic algorithm (GA) and chiasma verification, initialize the population firstly, which means give a set of initial values for σ. Then, do SVM training for each σ, caculate the false rate of classification respectively, and choose highest rate σ as the ultimate width of Gaussian kernel function[9]. In caculating the accuracy rate of classification, K-fold Cross Validation is adopted. The original datas are divided into *k* groups(dicided equally generally). Put each subset data as validation set, the rest of the *k*-1 subsets as the training sets. It will get *k* models whose accuracy rate of classification in their final validation sets is the accurate performance indicators of K-fold Cross Validation. The most optimal punish coefficient c can be determated at the same time. Figure 5 is the fitness curve of evolution algebra in parameter optimization with the remote transmission algorithm.

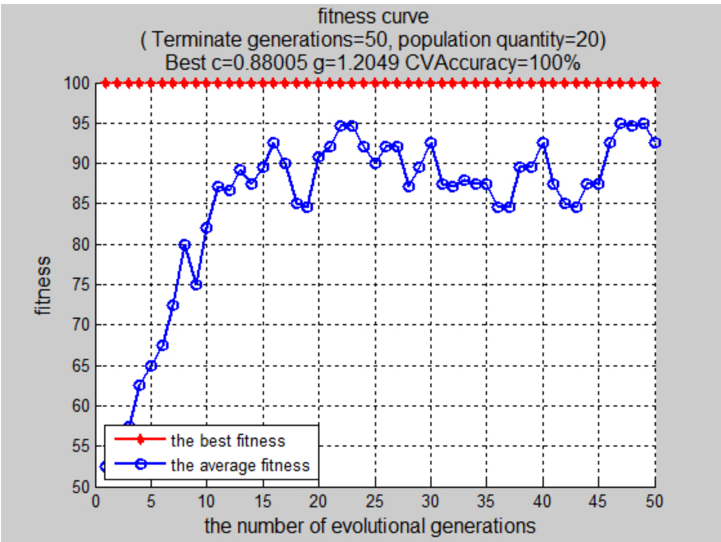

Fig. 5. The fitness graph

4 Test Results and Analysis

The following article will take training test for the punish coefficient c selected by support vector machine (SVM) and genetic algorithm,as well as width paremater σ of Gauss kernel function. Then, draw a comparision with experience choise of c and σ. The experiment al hardware conditions is Pentium(R) Dual-Core PC and limsvm based on matlab. The banknote used in this article is Rand of South Africa in 2005 whose denomination is 100,50,20 and 10 respectively. The number of each denomination is 60 pieces. In denomination recognition, each banknote has two sides pros and cons, and every side has two directions. That means there exit four kinds of situations for each banknote denomination, making a total 4 * 4 categories for each paper currency combining denomination with orientation and side.

Table 1. The result of experiment using genetic algorithm to parameter optimization

Algorithm	Train set	C	σ	Test set 数	Accuracy rate
GA	160	7.6452e+003	1.5259e-006	80	91.25%
	160	5.1147e+003	1.9418e-006	80	87.5%
	140	7.2236 e+003	1.8610e-006	60	93.3%
	140	8.5582 e+003	1.5368e-006	60	91.67%

Table 2. The result of experiment using PSO algorithm toparameter optimization

Algorithm	Train set setset	C	σ	Test set	Accuracy rate rate
PSO	160	8.5024e+003	1.8011e-006	80	60%
	160	5.5305e+003	4.8576e-006	80	62.5%
	140	10000	1.0000e-006	60	60%
	140	10000	4.3125e-006	60	63.33%

From table 1 and table 2, it can be found that the accuracy rate is higer by adopting genetic algorithms to optimization of C and σ. Because genetic algorithms operate offline, it has no influence on real-time response of the system even if the velocity of optimization is slow.

5 Conclusion and Prospects

This paper studied the banknote recognition based on SVM, especially the genetic algorithm that used to do parameter optimization of penalty coefficient C of SVM and σ(the width of the Gaussian kernel function). The adaptive genetic algorithm proposed in this paper provides a solution to SVM parameters selection. The experimental results confirm the feasibility and high efficiency of the scheme. From the simulation results, it is also apparent that the model adopting GA and optimized SVM is more accurate compared with the empirical estimates SVM, and the former has stronger generalization ability. It has a certain value to application, which can be used in parameter optimization of other types of SVM. The direction of banknote recognition research mainly focus on speeding the velocity of multiclass classification and researching the training algorithm in the future. It is necessary to make a further step in optimization algorithm, so as to speed up the training, improve the ability of real-time processing, and raise recognition accuracy [10].

References

1. Zhang, E.H., Jiang, B., Duan, J.H., et al.: Research on paper currency recognition by neural networks. In: 2003 International Conference on Machine Learning and Cybernetics, vol. 4, pp. 2193–2197. IEEE (2003)
2. Ray, K.: Unsupervised edge detection and noise detection from a single image. Pattern Recognition **46**(8), 2067–2077 (2013)
3. Onoda, R.G., Muller, T.K.: Soft Margins for AdaBoost. Machine Learning **42**(3), 287–320 (2001)
4. Lin, C.-L., Qingyun, S.: Computer engineering, 11 (2003)
5. Liu, Y.-H., Chen, Y.-T.: Face recognition using total margin-based adaptive fuzzy support vector machines. IEEE Transactions on Neural Networks **18**(1), 192 (2007)
6. Yu, C.-L.: Computer technology and development, 04 (2011)
7. Pei, Y.M.S.: The SVM parameters optimization research based on improved genetic algorithm. Computer Simulation **008**, 150–152 (2010)
8. Lin, H.-T., Lin, C.-J., Weng, R.C.: A note on Platt's probabilistic outputs for support vector machines. Machine Learning **68**(3), 267–276 (2007)
9. Frontzek, T., Navjn, T, Eekmiller, R.: Predicting the nonlinear dynamics of biological neurons using support vector machines with different kernels. In: Proc of the International Joint Conference on Neural Networks, Washington DC
10. Mirza, R., Nanda, V.: Paper Currency Verification System Based on Characteristic Extraction Using Image Processing. International Journal of Engineering, 1