

# Parameters Optimization and Application of SVM Based on PCA-Particle Swarm Algorithm

Qingmi Yang

School of Communication Engineering, Chengdu University of Information Technology, Chengdu, Sichuan, China

## ABSTRACT

The parameter optimization of Support Vector Machine (SVM) has been a hot research direction. To improve the optimization rate and classification performance of SVM, the Principal Component Analysis (PCA) - Particle Swarm Optimization (PSO) algorithm was used to optimize the penalty parameters and kernel parameters of SVM. PSO which is to find the optimal solution through continuous iteration combined with PCA that eliminates linear redundancy between data, effectively enhance the generalization ability of the model, reduce the optimization time of parameters, and improve the recognition accuracy. The simulation comparison experiments on 6 UCI datasets illustrate that the excellent performance of the PCA-PSO-SVM model. The results show that the proposed algorithm has higher recognition accuracy and better recognition rate than simple PSO algorithm in the parameter optimization of SVM. It is an effective parameter optimization method.

**Keywords** : Support Vector Machine; Principal Component Analysis; Particle Swarm Optimization; parameter optimization

## I. INTRODUCTION

In 1995, Vapnik proposed a SVM based on the theory of statistical VC dimension and the principle of structural risk minimization [1]. It has excellent generalization ability and robustness in solving small sample, nonlinear and high-dimensional mode classification and recognition problems. Currently, it is widely used in the fields of pattern recognition, regression estimation, etc [2, 3]. As a kind of attention-grabbing classification and recognition technology, SVM has many advantages, but it has strict selection of kernel function parameters and penalty factors. It will have a greater impact on the performance of SVM that Choose different kernel function parameters and penalty factors. At present, there is no unified method for selecting SVM

parameters in the world. Commonly used algorithms are: genetic algorithm, grid optimization method, etc. Huang used the genetic algorithm to optimize the SVM parameters and proposed a new adaptive value evaluation function based on the accuracy rate and the weight penalty factor [4]. Friedrichs proposed using a fixed step size for grid search to find suitable parameters, but this only applies to the case of fewer parameters, and there are problems such as complicated calculation and time consuming [5]. Kennedy and Eberhar proposed the PSO algorithm [6], which is a stochastic optimization algorithm based on swarm intelligence. It starts from the random solution, finds the optimal solution by iteration, and uses the fitness to evaluate the result good or bad. Later, Shi introduced the inertia weight to make the detection ability of the algorithm better

[7, 8]. PSO is widely used in various fields due to its simple and easy to implement and strong versatility.

When using the SVM to deal with the classification identification problem, although the effect is good, the training speed will be slow when the input data dimension is too high. Combined with the advantages of the PCA, this paper proposes an improved PSO-PCA algorithm, and the results of simulation were compared to other algorithms. Comparative results indicate that the algorithm has better convergence rate and higher recognition accuracy than simple PSO algorithm.

The second section (**METHODS**) will introduce will introduce the theory of classification used in this paper. The third section (**EXPERIMENTAL IMPLEMENTATION**) will detail the experimental implementation. The fourth section (**CONCLUSION**) will summarize the experimental conclusions

## II. METHODS

### A. SVM method

SVM is a two-class model. Its purpose is to find a hyper plane to segment the sample. The principle of segmentation is to maximize the interval and finally transform it into a convex quadratic programming problem. Whether the data is linear or not, SVM can be divided into linear SVM and nonlinear SVM.

Linear SVM can be divided into linear separable SVM and linear non-separable SVM. The separable SVM can separate the samples with a linear function. In a multidimensional space, the linear functions are collectively called hyper planes. For example, there is a data set  $D=\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ .  $x_i$  is the input vector,  $y_i \in \{-1, 1\}$  is a category. The SVM learns to get the separated hyper plane  $W^T X + b = 0$ , where  $W$  and  $b$  are represented as weight vectors and offsets, respectively. Optimal classification hyper plane is:

$$\begin{cases} \min \frac{\|w\|^2}{2} \\ s.t. y_i(w^T x_i + b) - 1 \geq 0 \end{cases} \quad (1)$$

When the data is linearly inseparable, a slack variable ( $\xi_i \geq 0$ ) is introduced to transform the formula (1) into the following formula:

$$\begin{cases} \min(\frac{\|w\|^2}{2} + C \sum_{i=1}^n \xi_i) \\ s.t. y_i(w^T x_i + b) \geq 1 - \xi_i \end{cases} \quad (2)$$

$C$  is the penalty parameter.

For nonlinear classification problems, map data in low-dimensional space to high-dimensional space to achieve linear separability. Due to the increase in the dimension, the cost of the inner product calculation increases. The kernel function  $K(x, x') = \phi(x) \phi(x')$  can be used to map to the high-dimensional space, which avoids complicated calculations and effectively overcomes the problem of dimensionality disaster. The final decision function of the classifier is:

$$f(x) = \text{sign}(\sum_{i=1}^N \alpha_i y_i K(x_i, x) + b) \quad (3)$$

Currently used kernel functions are as follows:

1) linear kernel function:

$$K(x_i, x_j) = x_i^T x_j \quad (4)$$

2) Polynomial Kernel function:

$$K(x_i, x_j) = (\gamma x_i^T x_j + c)^d \quad d > 1 \quad (5)$$

3) Radial Basis Function:

$$K(x_i, x_j) = \exp(\gamma \|x_i - x_j\|^2) \quad (6)$$

4) Sigmoid Kernel function:

$$K(x_i, x_j) = \tanh(\alpha x_i^T x_j + c) \quad (7)$$

Related research shows that the Gaussian radial basis kernel function has strong learning ability and is a kind of kernel function that performs well on classification problems. Therefore, we use the radial basis kernel function to construct the SVM model in this paper. The classification performance of support vector machine is mainly affected by two parameter: penalty factor  $C$  and kernel function parameter  $\gamma$ .

- 1) Penalty factor: it is the tolerance of the error. The higher the C, the over-fitting is easy. The smaller the C, the easier the under-fitting. C is too big or too small, the generalization ability is worse.
- 2) Kernel function parameter: it determines the distribution of features when low-dimensional data is mapped to high-dimensional space. The larger the  $\gamma$ , the higher the training accuracy and the lower the test accuracy. The smaller the  $\gamma$ , the higher the accuracy of the training set and the accuracy of the test set.

### B. Parameter optimization

We use PCA to reduce the data dimension and PSO to find the optimal parameters in this paper. The PCA method can linearly combine data with high correlation with higher dimensional into less-correlated data with less- dimension through covariance matrix [9]. It can reduce the dimension of the feature vector without changing the distribution characteristics of the data. The dimension reduction step is to use the KL transform to linearly decompose the data, obtain a series of weights and feature vectors, and form a new coordinate system by orthogonally traversing the feature vectors, and then project the original input feature vector into the PC dimension subspace to get the main information.

Process is as follows:

First, in order to solve the error between different dimensions between data, the original data is standardized. The standardization formula is as follows.

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{\sqrt{\text{Var}(x_j)}} \quad (i = 1, 2 \dots, m, j = 1, 2 \dots n) \quad (8)$$

$x_j = \frac{1}{n} \sum_{i=1}^m x_{ij}$ ,  $\text{Var}(x_j) = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$ , m is the number of samples, n is the number of variables, After data normalization, calculate the correlation coefficient matrix,

$$R = X^T X \quad (9)$$

X is a standardized matrix, R is the correlation coefficient matrix.

According to the formula  $|\lambda - R|$ , the eigenvalue is obtained, and the eigenvalue is used to obtain the eigenvector, and then a new matrix is formed based on the eigenvector, so that all the columns of the new matrix are arranged according to the descending order of the contribution rate. The contribution rate of the component is calculated by the following formula.

$$\alpha_p = \frac{\sum_{k=1}^p \lambda_k}{\sum_{k=1}^n \lambda_k} \quad (10)$$

$\lambda$  is the eigenvalue of each dimension, p represents the first p principal components, and  $\alpha_p$  represents the cumulative contribution.

In general, the principal component is obtained by setting the cumulative contribution rate threshold, that is, when the threshold is reached, the first n components are replaced by the first p principal components. The threshold is 90% in this paper.

The concept of the PSO algorithm stems from the study of the foraging behavior of birds. It finds the optimal solution through collaboration and information sharing among individuals in the group. The algorithm uses particles to simulate birds. Particles have two properties of velocity and direction. The velocity is the speed at which the particles move, and the position is the direction in which the particles move. Each particle finds the optimal solution separately in space, records it as the current individual extremum, and then shares the individual extremum with other particles in the particle swarm. The current overall optimal solution is the optimal individual extremum in the group. And then update the speed and position of the particles by iterating over and over again. Finally, the optimal solution satisfying the termination condition is obtained. In space, an individual's position is  $x_i$ , the speed is  $v_i$ , the individual's extreme value is  $P_{best}$ , and

the global extremum is  $g_{best}$ . Speed and position are updated by the following formula.

$$v_i = wv_i + c_1r_1(p_{best} - x_i) + c_2r_2(g_{best} - x_i) \quad (11)$$

$$x_i = x_i + v_i \quad (12)$$

$w$  is the inertia factor,  $c_1$  and  $c_2$  are acceleration factors, and  $r_1$  and  $r_2$  are random numbers between 0 and 1.

### C. Process

The final implementation requires that everything is brought together and training a model using the PSO algorithm and the PCA method. The model process is formed as follows.

1. Standardize data.
2. The normalized data is analyzed by the PCA method to obtain a principal component with a threshold of 90%, and the obtained data is divided into test data and training data.
3. PSO algorithm is used to train the sample parameters to optimize, and then the parameters are used for model training.
4. The trained model is used to calculate the fitness and judge whether the fitness is optimal. If not, repeat step 3.
5. Test samples for classification and identification tests to obtain accuracy.

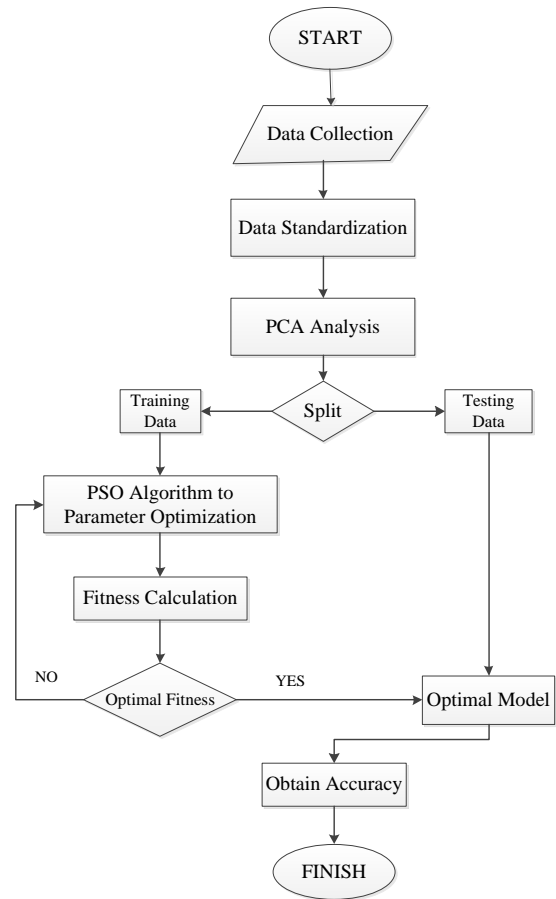


Fig.1 Model process

### III. EXPERIMENTAL IMPLEMENTATION

In order to verify the performance of the PCA-PSO algorithm, this paper selects six data sets in the UCI database to conduct experiments. Name, Number of Instances, Number of Attributes, Number of Categories, and Number of Training/Testing samples of the data are shown in the table below.

Table 1 UCI Data

Name	Number of Instances	Number of Attributes	Number of Categories	Number of Training / Testing samples
WDBC	569	32	2	300/269
Ionosphere	351	34	2	180/171
Robot navigation	5456	24	4	2660/2796
Vehicle	946	18	4	490/456
heart	270	13	2	130/140
PID	768	8	2	300/468

**A. Parameter Settings**

The method in this paper optimizes the penalty factor C and kernel function parameter  $\gamma$  of SVM. The relevant parameters are set as follows: the maximum number of population is 20, the maximum evolution algebra is 100, the learning factor  $c1=1.5$ ,  $c2=1.7$ , The range of SVM parameter C is [0.1, 1000], the range of  $\gamma$  is [0.01, 10], the number of cross-validation is 10, and the accuracy of the SVM training model is used as the fitness function.

**B. Experimental Analysis and Results**

Combining the parameter settings in the previous section, the parameters of the SVM are optimized by the particle swarm optimization algorithm

and the PCA combined with the particle swarm optimization algorithm. According to the results, the SVM model is used to predict and classify the test set data. The classification results are shown in Table 2.

It can be seen from Table 2 that under the same data conditions, compared with the optimization of SVM parameters by the particle swarm optimization algorithm, the classification accuracy obtained by the PCA-PSO method proposed in this paper is improved. At the same time, it is found that the time spent on the classification of the PCA-PSO method was reduced, especially when the recognition of Robot navigation and the WDBC class was improved by nearly 1/3.

Table 2 Comparison of Classification Results

Name	PSO-SVM				PCA-PSO-SVM			
	C	$\gamma$	Accuracy /%	Time /s	C	$\gamma$	Accuracy /%	Time /s
WDBC	3.7758	1.0479	97.20	302.88	2.6331	1.1559	98.14	193.29
Ionosphere	2.5274	1.2823	88.95	54.696 5	9.2678	1.3520	91.16	45.4029
Robot navigation	227.68 39	0.1000	85.84	6163.6 1	320.242 5	0.2238	86.41	4037.86
Vehicle	14.205 8	1.4407	74.13	968.78	604.732 1	0.1010	85.47	1217.10
heart	6.5938	1.3373	84.28	83.95	15.5426	1.8426	85.00	64.93
PID	6.6318	0.2017	78.19	94.23	10.3528	0.1316	81.06	88.54

**IV. CONCLUSION**

In this paper, principal component analysis and particle swarm optimization are combined into the parameter optimization of SVM classification to form a PCA-PSO algorithm. Compared with the PSO algorithm, the method of this paper first uses the PCA method to reduce the dimension, and then uses the PSO algorithm to optimize the parameters of the SVM. Not only does it speed up the recognition speed, but

also the recognition accuracy is significantly improved. It shows the relative advantages of the algorithm in recognition accuracy and optimization speed, as well as good classification performance in practical applications.

## V. REFERENCES

- [1]. Vapnik V. The nature of statistical learning theory M. Springer science & business media, 2013.
- [2]. Huanrui H. New mixed kernel functions of SVM used in pattern recognitionJ. *Cybernetics and Information Technologies*, 2016, 16(5): 5-14.
- [3]. Korpela J, Miyaji R, Maekawa T, et al. Toothbrushing performance evaluation using smartphone audio based on hybrid HMM-recognition/SVM-regression modelJ. *Journal of Information Processing*, 2016, 24(2): 302-313.
- [4]. Huang C L, Wang C J. A GA-based feature selection and parameters optimization for support vector machinesJ. *Expert Systems with applications*, 2006, 31(2): 231-240.
- [5]. Friedrichs F, Igel C. Evolutionary tuning of multiple SVM parametersJ. *Neurocomputing*, 2005, 64: 107-117.
- [6]. Eberhart R, Kennedy J. Particle swarm optimizationC//*Proceedings of the IEEE international conference on neural networks*. 1995, 4: 1942-1948.
- [7]. Shi Y, Eberhart R. A modified particle swarm optimizerC//1998 IEEE international conference on evolutionary computation proceedings. IEEE world congress on computational intelligence (Cat. No. 98TH8360). IEEE, 1998: 69-73.
- [8]. Shi Y, Eberhart R C. Empirical study of particle swarm optimizationC//*Proceedings of the 1999 Congress on Evolutionary Computation-CEC99* (Cat. No. 99TH8406). IEEE, 1999, 3: 1945-1950.
- [9]. Gaspar P, Carbonell J, Oliveira J L. On the parameter optimization of Support Vector Machines for binary classificationJ. *Journal of integrative bioinformatics*, 2012, 9(3): 33-43.

## Cite this article as :

Qingmi Yang, "Parameters Optimization and Application of SVM Based on PCA-Particle Swarm Algorithm", *International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET)*, Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 6 Issue 4, pp. 325-330, July-August 2019. Available at doi : <https://doi.org/10.32628/IJSRSET196431>  
Journal URL : <http://ijsrset.com/IJSRSET196431>