# Efficient Parameter Tuning of Support Vector Machine Based on Nesting Particle Swarm Optimization

Pin Liao[1, 2, a], Sensen Wang[2, b], Xin Zhang[1, c], Kunlun Li[1, d], Mingyan Wang[2, e]

[1]College of Science and Technology, Nanchang University, Nanchang, China

[2]Information Engineering School, Nanchang University, Nanchang, China

[a]pinliao@ncu.edu.cn, [b]wsensen@hotmail.com, [c]zhangxin77@gmail.com, [d]likunlunnihao@126.com
, [e]mingyanw@ncu.edu.cn

**Abstract.** Parameter optimization of Support Vector Machine (SVM) is an important research problem, because it is critical for establishing an effective SVM with excellent generalization. This study develops a novel and efficient method to solve this problem by nesting two particle swarm optimization (PSO) algorithms to optimize SVM kernel parameter(s) and penalty parameter separately, in comparison with traditional methods that try to optimize all the parameters concurrently. The experimental results on gender classification of facial images validate the feasibility and quickness of the new method.

## Introduction

Support Vector Machine (SVM) [1] [2] is a powerful and popular machine learning method for classification and regression problems. One of the important research problems in SVM is parameter optimization, which has significant influence on achieving outstanding generalization performance of SVM. And the SVM parameters to be optimized consist of the penalty parameter and the kernel parameter(s). Many techniques have been developed to solve the parameter optimization of SVM, such as: priori knowledge method [2], grid search [3], gradient descent methods [4], simulated annealing [5], genetic algorithm (GA) [6], artificial immune algorithm [7], ant colony optimization (ACO) [8], particle swarm optimization (PSO) [9], and so on.

In this paper, a novel method is presented to solve parameters optimization based on nesting two particle swarm optimization algorithms (NPSO). NPSO is far more efficient than the traditional methods which optimize all the parameters simultaneously. In the NPSO method, the inner loop PSO is responsible for the optimization of the penalty factor, while the outer loop PSO deals with the optimization of the kernel function parameters. Thereby the computation complexity can be reduced by orders of magnitude compared to the traditional methods.

## Support Vector Machine for Classification

Given a training set S = { $(x_i, y_i) \mid x_i \in \mathbb{R}^D$, $y_i \in \{-1, 1\}$, $i = 1,\ldots, N$ }, where $x_i$ is a input vector, $y_i$ the labels of $x_i$. D dimension of feature vectors, and N number of training samples, determination of optimum hyper plane of SVM for classification is required to solve the constrained quadratic programming problem as follows.

$$\min_{w,b,\xi} \quad \frac{1}{2}w^T \cdot w + C\sum_{i=1}^{N}\xi_i$$

$$s.t. \quad y_i((w^T \cdot \varphi(x_i)) + b) \geq 1 - \xi_i, \tag{1}$$

$$\xi_i \geq 0, i = 1, \cdots N$$

where $\xi$ are non-negative slack variables, $C$ is the penalty parameter, and function $\varphi()$ (if non-linear) maps training samples $x_i$ into a high-dimensional feature space.

The usually adopted kernel functions in SVM comprise at least four types: linear, polynomial, sigmoid and radial basis kernel function (RBF). Particularly, RBF (defined as $K(x, y) = \exp(-\gamma \| x - y \|^2)$, $\gamma > 0$) has been used widely, because of its amazing generalization performance for non-linear problems, and its simplicity for optimization with only one parameter, $\gamma$. Therefore, this study applies RBF as the default kernel in the SVM without loss of generality, as the other kernels can also be employed in the same way.

## Particle Swarm Optimization (PSO)

PSO is a stochastic populated search method proposed by Kennedy and Eberhart [10], which is motivated by social behavior such as bird flocking or fish schooling to search optimal solutions in a multidimensional space. Like other evolutionary algorithms, PSO evolves a population (called swarm) of potential solutions (called particles) iteratively to seek the optimal solution. In contrast with other optimization methods, PSO is easy to exploit with few parameters to tune. Each particle in a swarm moves during every evolving iteration based on two factors: (1) its personal previous best position (pbest), called the cognition part, (2) and its best global position (gbest), called the social part. The updating of the velocity and position of particles can be defined as follows for each particle $i$ and dimension $d$.

$$v_{id}^{t+1} = w^t \cdot v_{id}^t + c_1 \cdot rand_1 \cdot \left( pbest_{id}^t - p_{id}^t \right) + c_2 \cdot rand_2 \cdot \left( gbest_d^t - p_{id}^t \right) \tag{2}$$

$$p_{id}^{t+1} = p_{id}^t + v_{id}^{t+1} \tag{3}$$

where $t$ is the evolving generation; $c_1$ and $c_2$ denote personal and social learning factors , whose values are positive and generally assigned to 2; $rand_1$ and $rand_2$ are random variables generated from a uniform distribution in the range [0,1]; Termed the "inertia weight", $w$ is introduced to compromise and balance the global exploration and local exploitation; $p_{id}$ represents the position of particle $i$ on dimension $d$, in a value range of $[p_{min}, p_{max}]$; and $v_{id}$ means the velocity of particle $i$ on dimension $d$, in a value range of $[-v_{max}, v_{max}]$. Here $p_{min}$ and $p_{max}$ are set empirically, $v_{max}$ set as ($p_{max}$ - $p_{min}$) * 0.1, and the inertia weight $w$ defined as:

$$w^t = \frac{t_{max} - t}{t_{max}} \cdot \left( w_{max} - w_{min} \right) + w_{min} \tag{4}$$

## Parameter Optimization Based on NPSO

In order to decrease the expensive cost of SVM training, the sequential minimal optimization (SMO) [11] algorithm was proposed, which can decompose a large convex quadratic programming (QP) problem into a number of sub-problems. SMO adopts a minimal working set with only two data points, and resolves the sub-problems in a direct analytic way. Moreover, LIBSVM [12], a modified SMO algorithm, was presented to accelerate convergence by exploiting second order information. Thereby, this work employs the LIBSVM algorithm to train SVM.

Different from the traditional methods which optimize the penalty factor and the kernel parameter simultaneously, NPSO optimizes the two kinds of SVM parameters separately by nesting two loops of PSO, where the inner loop PSO optimizes C and the outer loop PSO optimizes γ. Thereby, as the kernel matrix remains unchanging with one constant γ in each inner loop, it can be reused at all the iterations in the inner loop with a swarm of different C. Because the updating of the kernel matrix is the most computation expensive part of SMO, NPSO can decrease the computation complexity significantly by orders of magnitude dependent on the size of the C swarm.

The algorithm of NPSO is given as follows:

1   (Initialization) Generate a swarm of particles for  γ  with random positions and velocities in the search space.

2   (Fitness) Measure the fitness of each  γ  in the current swarm.

**Loop for each $\gamma$ in the current swarm:**

    2.1    (Initialization) Generate a swarm of particles for *C* with random positions and velocities in the search space.

    2.2    (Fitness) Measure the fitness of each *C* in the swarm with the fixed $\gamma$ with the average testing rates of k-fold cross validation.

    2.3    Assign the top fitness of the fixed $\gamma$ with each *C* to the fitness of the $\gamma$.

    2.4    (Termination) Go to Step 2 if termination criterion is satisfied, or go to Step 2.5 otherwise.

    2.5    (Update) The *C* swarm involves in accordance with Eq. (2) and Eq. (3). Go to Step 2.2.

3    (Termination) Go to Step 5 if termination criterion is satisfied, or go to Step 4 otherwise.

4    (Update) The $\gamma$ swarm involves in accordance with Eq. (2) and Eq. (3). Go to Step 2.

5    (Optimal result) Achieve the ultimate $\{C, \gamma\}$ according to the top fitness (i.e. the average testing accuracy rate of a k-fold cross-validation).

## Experimental Results

Extensive experiments for gender classification of facial images are conducted in order to evaluate the performance of NPSO, the proposed novel parameter optimization approach of SVM. A large-scale database is constructed by us with over 20,000 facial images gathered online. To expand the size of the training set to over 60,000, some synthetic samples are generated from the original images with minor geometric transforms such as translation, scaling, rotation and mirror-reflection. The testing set, independent of the training set, is constructed with 4404 original facial images. Gabor filters are exploited to extract feature with the dimension of 5856.

The k-fold cross validation is used for choosing optimal parameters in this work, as described in the algorithm of NPSO. In k-fold cross validation, the training data set is randomly partitioned into k roughly equal-sized subsets. Among the k subsets, a single subset is preserved as the validation data for testing, and the remaining k-1 subsets are used as training data. Then the cross-validation procedure is repeated k times, with each subset used exactly once for validation. The fitness is defined as the average testing correct rate of the k-fold cross validation on the training data set. In this study, k is set as 10. At last, the chosen optimal parameters by the k-fold cross validation are used to train an SVM classifier on the whole training set, and then the final SVM classifier is evaluated on the testing set.

In our experiments, empirically the exploring range of *C* is set as $[2^{-3}, 2^{10}]$, and the exploring range of $\gamma$ $[2^{-13}, 2^{-1}]$. In NPSO, the size of *C* swarm in the inner PSO is set as 6, and the size of $\gamma$ swarm in the outer PSO is set as 10. And for the traditional PSO, the size of swarm set as 60. The range of the inertia weight is often within [0.0, 1.0]. In this work, $w_{max} = 0.9$, and $w_{min} = 0.4$.

As illustrated in Table 1, NPSO achieves the best accuracy rate compared to the grid search and the traditional PSO which tune the penalty parameter and the kernel function parameter(s) in SVM all together. NPSO is much more efficient than the traditional methods, as the computation complexity is dramatically reduced by orders of magnitude. Therefore it can search a more optimal result from much more candidate solutions in a same long time than the traditional methods.

**Table 1**. Optimal SVM parameters and accuracy rates corresponding to different searching methods on the testing set

| Method | *C* | γ | Accuracy (%) |
| --- | --- | --- | --- |
| Grid Search | 3.48220225 | 0.00170029 | 96.9119 |
| PSO | 8.91840492 | 0.00109746 | 97. 2298 |
| NPSO | 4.87523026 | 0.00142664 | 97. 2979 |

## Conclusion

This paper presents an original optimization method to tune the penalty factor and the kernel parameter of SVM separately by nesting two PSO algorithms (NPSO), compared to the traditional methods all of which optimize all the parameters simultaneously. NPSO can reduce the computing cost of the SVM parameter optimization by orders of magnitude by reusing the kernel matrix. The experimental results indicate that the NPSO method can achieve better classification performance than the other methods on gender classification of facial images.

## Acknowledgement

## References

[1] Corinna Cortes and Vladimir Vapnik, Support-vector networks. Machine Learning (Historical Archive) 20(3) (1995) 273-297.

[2] V. N. Vapnik, The Nature of Statistical Learning Theory, 2nd ed., New York: Springer-Verlag, 1999.

[3] C.-W. Hsu, C.-C. Chang and C.-J. Lin. A practical guide to support vector classification, Technical report available at http://www.csie.ntu.edu.tw/?cjlin/libsvm/, Department of Computer Science and Information Engineering, National Taiwan University, Taipei, 2003.

[4] S.S. Keerthi, Efficient tuning of SVM hyper parameters using radius/margin bound and iterative algorithms, IEEE Transactions on Neural Networks 13 (5) (2002), 1225–1229.

[5] H.-Q. Wang, D.-S. Huang and B. Wang, Optimisation of radial basis function classifiers using simulated annealing algorithm for cancer classification, Electronics Letters, 41 (11) (2005) 630 - 632.

[6] Mingyuan Zhao, Chong Fu, Luping Ji, Ke Tang, Mingtian Zhou. Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes. Expert Systems with Applications 38 (2011) 5197–5204.

[7] Ilhan Aydin, Mehmet Karakose, Erhan Akin. A multi-objective artificial immune algorithm for parameter optimization in support vector machine. Applied Soft Computing 11 (2011) 120–129.

[8] X. Zhang, X. Chen, and Z. He, An ACO-based algorithm for parameter optimization of support vector machines. Expert System with Applications, Vol.37, 2010, No.9, pp.6618-6628.

[9] S.W. Lin, K.C. Ying, S.C. Chen, & Z.J. Lee, Particle swarm optimization for parameter determination and feature selection of support vector machines, Expert Systems with Applications, vol. 35, pp. 1817–1824, 2008.

[10] J. Kennedy, & R. C. Eberhart, Particle swarm optimization. Proceedings of IEEE Conference on Neural Network, 4, (1995) 1942–1948.

[11] John C. Platt. Fast training of support vector machines using sequential minimal optimization, in B. Schölkopf, C. Burges, and A. Smola, editors, Advances in Kernel Methods - Support Vector Learning. MIT Press, (1999) 185–208.

[12] C. C. Chang, C. J. Lin, "LIBSVM: A Library for Support Vector Machines", Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm/