Parameters optimization of SVM based on the swarm intelligence

Yu-xin WANG, Xue-zhen Li, Zheng-yi Wang
(31001troops, Beijing, 10000, China)

Abstract: In view of the existing optimization method in the optimization of support vector machine (SVM) parameters of low searching efficiency and easy falling into the master problem, put forward a kind of modified pollinate flowers algorithm based on mutation strategy—MFPA algorithm, and applied to SVM parameter optimization problem, through the relevant test data sets from UCI, test results show that MFPA-SVM parameter optimization model of its classification performance is superior to the existing PSO-SVM model and the BA-SVM model.

1. Introduction
In recent years, more and more modeling methods have been put forward by relevant scholars to solve problems such as classification identification, risk prediction, and effectiveness evaluation. These modeling methods include time series analysis, grey theory, neural network, etc. [1]. Support Vector Machine (SVM) is a modeling method based on structural risk minimization with powerful generalization ability, which can perfectly solve problems such as small samples, nonlinearity, and local extremum. However, the learning ability and generalization ability of SVM depend on the selection of suitable parameters, which directly affect the performance of the model. Thus, in recent years, more and more parameter optimization methods are used to SVM parameter selection, such as grid search method, particle swarm optimization algorithm, bat algorithm, etc. [5-8]. However, grid search method has a large amount of computation with low search efficiency. Particle swarm optimization and bat algorithm have the problems of low rate of convergence and easy local extremum in the process of parameter optimization. The flower pollinate algorithm (FPA) is a new type of meta-heuristic group intelligence algorithm put forward by the British scholar Xin-Shen Yang enlightened by pollination process of flowering plants in 2012 [9], which is simple and easy to implement with a novel optimization structure, and can achieve a good balance between global search and local search. With great research potential, it has been successfully applied to graphics coloring, feature selection, power system optimization and other issues [10-13].

Considering the deficiencies of the existing parameter optimization method in the optimization of SVM parameters, the author improved the FPA algorithm and proposed a modified flower pollinate algorithm based on mutation strategy - MFPA algorithm. After MFPA algorithm was applied to SVM parameter optimization, further MFPA - SVM optimization model is put forward. According to the test results of six standard of UCI (university of California, Irvine) test data sets, it can be shown that the classification accuracy and generalization performance of the MFPA—SVM model proposed in this paper are better than that of particle swarm optimization SVM parameter model and bat algorithm optimization SVM parameter model (BA-SVM).
2. Theoretical Introduction

2.1 Support Vector Machine

Support Vector Machine (SVM) is a classification method based on statistical learning theory proposed by Vapnik et al. in 1995, whose core idea of SVM is to construct an optimal classification hyper plane, which can accurately distinguish the two types of samples to be divided. Let the input sample set be $(x_i, y_i), \ x_i \in \mathcal{R}^d, \ y_i \in \{+1, -1\}$. For the linear separable problem, the quadratic programming problem corresponding to solving the optimal classification hyper plane is:

$$\begin{align*}
\min & \quad \frac{1}{2}\|w\|^2 + c\sum_{i=1}^{n} \xi_i \\
\text{s.t.} & \quad y_i(w^T x_i + b) + \xi_i - 1 \geq 0 \quad (1)
\end{align*}$$

In the equation, $w$ is the coefficient vector and $b$ is the offset; $\xi_i$ is a slack variable introduced when linear is not separable; $C$ is a penalty factor used to represent the penalty index for misclassification. It can be seen that $C$ is a key factor which determines the SVM learning ability and the experience risk coordination degree. If $C$ is too large, it will lead to over-learning, resulting in the decrease of the generalization ability of the classifier. Otherwise, it will lead to the decrease of correct classification of the classifier and even invalidation of the entire classifier model. Using the Lagrange method to solve equation (1), the resulting optimal classification hyper plane function is:

$$f(x) = \text{sgn}\{(w^T x_i) + b\} = \text{sgn}\{\sum_{i=1}^{n} \alpha_i^* y_i (x_i^T x) + b^*\} \quad (2)$$

In the equation, $\text{sgn}(x)$ is a sign function, $\alpha_i^*$ is the Lagrange coefficient corresponding to the support vector; $b^*$ is the classification threshold. By introducing the kernel function $K(x_i^T x)$ into equation (2), the linear inseparable samples in the low-dimensional space are mapped to the high-dimensional space, making the linear separable. By introducing the kernel function, the optimal classification hyper plane function in the nonlinear classification problem can be obtained:

$$f(x) = \text{sgn}\{\sum_{i=1}^{n} \alpha_i^* y_i K(x_i, x) + b^*\} \quad (3) \quad K(x_i, x) = f(x)^T \times f(x)$$

In equation (3), the radial basis function (RBF) is used as the kernel function. With fewer parameters and good performance for nonlinear prediction problems, the RBF kernel function is the most commonly used kernel function. The corresponding mathematical expression is:

$$K(x_i, x) = e^{-\|x_i - x\|^2/2\delta^2} \quad (4)$$

In equation (4), $\delta$ is the kernel function parameter, which affects the distribution of complexity of sample data in the characteristic space.

From the above analysis, it can be seen that the setting of penalty parameter $C$ and RBF kernel function parameter $\delta$ are the main factors that affect the performance of SVM classifier.

2.2 Flower Pollination Algorithm

Flower pollination algorithm (FPA) simulates the natural pollination process of flowers in nature. The natural flower pollination process mainly consists of cross pollination and self-pollination. Cross-pollination is a process in which some insects or birds conduct a wide range of cross-pollinations as pollinators, which can be regarded as a global search in the optimization theory. Self-pollination mainly involves local pollination in a small range between the same species of flowers under natural action, which can be regarded as local search in optimization theory. The pollination method is selected between cross pollination and self-pollination with a certain conversion probability, which can well balance the relationship between local search and global search.

In the cross-pollination process, that is, the global search process, pollen can be transmitted to different locations by pollinators in a flight mechanism called Levy flight, which can be described by the following equation:

$$x_i^{t+1} = x_i^t + L(x_i^t - g_i); \quad (5)$$

In equation (5), $x_i^t$ and $x_i^{t+1}$ are the solutions of the $t$ generation and $t + 1$ generation of the
algorithm, respectively. \( g \) represents the optimal solution in the current population. \( L \) can be considered as a parameter related to pollination intensity, which is essentially equal to the step size parameter based on Levy flight, that is to say, in the global search process, pollinators explore by Levy flight mechanism, and the equation for calculating \( L \) is as follows.

\[
L \sim \frac{\lambda \Gamma(\lambda) \sin\left(\frac{\pi \lambda}{s^{1+\lambda}}\right)}{\pi s^{1+\lambda}}
\]

Among them, \( \lambda=3/2, \) and \( \Gamma(\lambda) \) is the standard gamma function.

The self-pollination process, that is, local search process, is described by the following equation:

\[
x_{i}^{t+1} = x_{i}^{t} + \varepsilon(x_{i}^{t} - x_{k}^{t});
\]

Equation (7) is similar to the mutation strategy in differential evolution algorithm, among which \( x_{i}^{t} \) is the solution of generation \( t \) and can be considered as the base vector. \( x_{i}^{t+1} \) is the solution of generation \( t+1 \). \( x_{i}^{t} \) and \( x_{k}^{t} \) are the solutions chosen randomly within the population, which are different from \( x_{i}^{t} \). Differential vectors \( \varepsilon(x_{i}^{t} - x_{k}^{t}) \) are obtained by making a difference between the two, and \( \varepsilon \) is a random number which obeys \((0,1)\) uniform distribution. By adding the differential vector to the base vector, a new solution \( x_{i}^{t+1} \) is obtained.

Cross-pollination and self-pollination are converted by a conversion probability \( p_e[0,1] \) in order to control the balance between global search and local search. A large number of experiments have been carried out in the literature [9] to prove that it is the most suitable for \( p \) to choose 0.8.

The flow chart of FPA algorithm is as follows:

The objective function: \( f(x) = (x_{1}, x_{2}, \ldots, x_{d}) \);

Initialize the population and find the optimal solution \( g \) in the initial population

Set the conversion probability \( p_e[0,1] \), and set the maximum iterations (or fixed constraints)

While (\( t < \) maximum iterations)

for \( i = 1:n \) (all n pollen in the population)

If \( \text{rand}<p \) _global search_

\[
x_{i}^{t+1} = x_{i}^{t} + L(x_{i}^{t} - g);
\]

Else _local search_

\[
x_{i}^{t+1} = x_{i}^{t} + \varepsilon(x_{i}^{t} - x_{k}^{t});
\]

End if evaluates new solutions, if new solutions are good, update them in the population

End for find the current optimal solution \( g^* \)

end while output found optimal solution

3. Parameter optimization of SVM based on improved flower pollination algorithm

3.1 Improved flower pollination algorithm

The key of optimal performance of the flower pollination algorithm is the global search mechanism and the local search mechanism. It can be seen from equation (7) that the local search mechanism of the algorithm draws on the differential mutation idea in the differential evolution algorithm, which obtains the difference vector through two random variables. And in the process of population evolution, the differential variation information of the difference vector is added to the new population so as to generate a new solution. However, the large randomness of the difference vector also results in lower search efficiency and slower convergence speed. The author draws on the targeted strategy proposed in literature [14] and applies it to the local search mechanism of the algorithm to improve the local search ability of the algorithm. Targeted mutation (TM) strategy is a new differential evolution mutation strategy proposed by Wei-Jie Zheng and other researchers in 2015, whose overall performance is better than the traditional differential evolution mutation strategy, and the mutation strategy has simple structure, strong portability and great research potential. Targeted mutation strategy is described as equation (8)

\[
\nu_{i}^{t} = x_{\text{best}}^{g} + F(x_{r}^{g} - x_{i}^{g})
\]
In equation (8), \( x_{\text{best}}^g \) is the optimal individual in the current population, \( x_{r}^g \) is a randomly generated vector, and \( x_{\text{g}}^g \) is the target vector, that is, the current solution vector. In the targeted mutation strategy, the global optimal solution \( x_{\text{best}}^g \) is used as the base vector of the mutation strategy, which can make the new solution search in the optimal direction and accelerate the convergence speed. And the difference vector \( x_{r}^g - x_{\text{best}}^g \) is obtained by using the random vector \( x_{r}^g \) with random information and the target vector \( x_{\text{g}}^g \) with certainty information. In this way, the relationship between randomness and certainty of difference vector can be well balanced. As a result, it can avoid the problems of inefficient search efficiency and low convergence speed caused by large randomness.

In order to further improve the performance of the mutation strategy, the author improves the original targeted mutation strategy and proposes an improved targeted mutation strategy based on the random mutation mechanism. Compared with the original targeted mutation strategy, this mutation strategy has two major improvement measures:

(a) Improve the original targeted mutation and propose a new targeted mutation varietas as the following equation:

\[
 x_{i}^{t+1} = x_{i}^{t} + F(x_{i}^{t} - x_{r}^{t}); \tag{9}
\]

In equation (9), \( F \) is a zoom factor, and equation (9) is a modified variant of equation (8). The base vector \( x_{\text{best}}^g \) in equation (8) is replaced by the target base vector \( x_{i}^{t} \), which can broaden the search domain of mutation operator.

(b) Using the idea of random variation proposed in literature [15], a random variation mechanism is introduced into the original targeted mutation strategy.

We apply the improved targeted mutation strategy based on random mutation mechanism to the local search process of FPA algorithm, and the improved local search process is given as follows:

\[
\text{if } \text{rand}<w, \quad x_{i}^{t+1} = x_{\text{best}}^g + \varepsilon(x_{i}^{t} - x_{r}^{t});
\]

\[
\text{else } \quad x_{i}^{t+1} = x_{i}^{t} + F(x_{i}^{t} - x_{r}^{t});
\]

Here \( \text{rand} \) is a random number and \( w \) is the golden ratio coefficient. It has been proved in reference [16] that a better new solution can be obtained according to the direction of the golden ratio coefficient toward the optimal solution. By introducing random mutation mechanism, random selection of different targeted mutation strategies can improve the ability of the algorithm to jump out of local extremum.

The above improvements only improve the local search mechanism of the algorithm. As for the global search mechanism for algorithms, the author introduces uniform mutation operators in the global search process to improve the diversity of algorithm populations. By introducing uniform mutation operators, difference mutation information can be introduced into new individuals to greatly increase the diversity of populations. The improved global search mechanism is as follows:

\[
 x_{i}^{t+1} = x_{i}^{t} + L(x_{i}^{t} - g_{\ast}) + \varepsilon(x_{i}^{t} - x_{k}^{g}) \quad \tag{10}
\]

Compared with the original global pollination mechanism, the improved global pollination mechanism adds a uniform variation operator \( \varepsilon(x_{i}^{t} - x_{k}^{g}) \), among which \( \varepsilon \) is a random number with uniform distribution between (0,1), and \( x_{i}^{t}, x_{k}^{g} \) are randomly generated random vectors.

By improving the local search mechanism and global search mechanism of the original flower pollination algorithm, the final improved algorithm can be obtained, which is called the modified flower pollination algorithm based on mutation strategy -- MFPA algorithm

### 3.2 MFPA-SVM Implementation Steps

The combined variable (\( C, \delta \)) of the SVM penalty factor \( C \) and the RBF kernel parameter \( \delta \) is used as the search target of the MFPA algorithm so as to find the combinatorial variable value which has the highest classification accuracy of SVM. It is the ultimate objective of the MFPA algorithm to optimize the SVM parameter model. In this paper, Accuracy rate of classification of SVM is taken as an
evaluation criterion. Accuracy of classification is obtained by the average cross validation recognition rate obtained by K fold cross validation method.

Fitness value: \( \text{Accuracy} = f(x) \), \( x = (C, \delta) \);

Initialize the population and find the optimal \((C, \delta)\) combined variable \(g\) in the initial population

Set the conversion probability \(p \in [0,1]\), set the search range of \(C\)、\(\delta\), and set the maximum iterations (or fixed constraints)

While (\(t < \text{maximum iterations}\))
For \(i=1:n\) (all \(n\) individuals in the population)
If \(\text{rand} < p\)  
   \(x_{i}^{t+1} = x_{i}^{t} + L(x_{i}^{t} - g) + \varepsilon(x_{i}^{t} - x_{k}^{t})\) 
else  
   local search process 
   if  \(\text{rand} < w\),  
     \(x_{i}^{t+1} = x_{\text{best}}^{t} + \varepsilon(x_{i}^{t} - x_{i}^{t})\); 
   else  
     \(x_{i}^{t+1} = x_{i}^{t} + F(x_{j}^{t} - x_{i}^{t})\); 
   end if 
End if evaluates the new solution by fitness value, and if the new solution is good, update them in the population
End for find the current optimal solution \(g\)
end while outputs the fould optimal solution, that is, the current optimal parameter combined variable \((C, \delta)\).

4. Simulation Verification

In order to verify the effectiveness and evaluate the performance of the MFPA-SVM model proposed in this paper, PSO-SVM model in literature [7] and BA-SVM model in literature [8] were selected for comparison validation. Six standard data sets from UCI were used to complete the experimental test. The relative information of data sets is shown in table 1. Parameters of each experimental model are set as follows:

   MFPA-SVM parameters: \(p = 0.8\); PSO-SVM parameters: \(c1 = c2 = 1.4559, w = 0.9, V_{\text{max}}\) (maximum speed) = 1. BA-SVM parameters: \(A = 0.25, r = 0.5, \alpha_f = 0.95\). The population number is uniformly set to 30, the maximum iterations is 100, the value range of penalty factor \(C\) is [0.01, 100], and the value range of RBF kernel function parameter \(\delta\) is [0.0001, 100].

| Database   | Train sample | Test sample | Attribute dimension |
|------------|--------------|-------------|---------------------|
| diabetes   | 420          | 348         | 8                   |
| WBDC       | 380          | 303         | 9                   |
| heart      | 150          | 120         | 13                  |
| ionosphere | 190          | 161         | 34                  |
| sonar      | 120          | 88          | 60                  |
| iris       | 90           | 60          | 4                   |

The experimental results are as shown in Table 2. In table 2, it shows the classification accuracy of the three models in train set and test set. In order to demonstrate more vividly that MFPA-SVM model has better classification ability than PSO-SVM model and BA-SVM model, the relationship between the number of loop iterations and the optimal classification accuracy rate of each iteration is drawn in the form of graph, as shown in figure 1 to figure 6.
Table 2 Test result

| Database      | PSO-SVM (%) | BA-SVM (%) | MFPA-SVM (%) |
|---------------|-------------|------------|--------------|
| Iris(train)   | 93.3333     | 94.4444    | 98.8889      |
| test          | 93.3333     | 83.3333    | 96.6667      |
| sonar(train)  | 53.3333     | 73.3333    | 85.8333      |
| test          | 53.4091     | 53.4091    | 89.7727      |
| ionosphere(train) | 53.1579     | 90.5263    | 93.6842      |
| test          | 80.7453     | 70.8075    | 97.5155      |
| heart(train)  | 60          | 85.3333    | 88           |
| test          | 55          | 53.3333    | 75           |
| WBDC(train)   | 87.3684     | 95         | 96.5789      |
| test          | 89.1089     | 76.8977    | 98.0198      |
| diabetes(train)| 70.2381     | 75.9524    | 76.9048      |
| test          | 69.8276     | 69.5402    | 77.0115      |

Figure 1 sonar dataset

Figure 2 WBDC Dataset
It can be seen from table 2 that whether in the classification accuracy obtained in the train set (train
line), or the test accuracy obtained by the optimal parameter combination of \((C, \delta)\) on the test samples (test line), or use training to get the optimal parameter combination \((C, \text{the delta})\) to test the test sample classification accuracy (test line), MFPA - the SVM model is better than the other two models, indicating that compared with PSO-SVM model and BA-SVM model, the MFPA-SVM model proposed in this paper also strengthens the learning performance and generalization ability of SVM in addition to improving the classification accuracy of SVM. From Figure. 1 to Figure 6, it can be seen that MFPA algorithm can jump out of local optimum compared with PSO algorithm and BA algorithm in the process of SVM parameter optimization with more efficient search speed, and it can find better parameter combination with better classification performance.

5. Conclusion
In this paper, a modified flower pollination algorithm based on mutation strategy (MFPA) is successfully applied to the parameter optimization of SVM. The MFPA algorithm is used to search and optimize the penalty coefficient and kernel parameters of SVM, so as to find the optimal parameters combination of SVM. Relevant test data sets in UCI are tested in this paper, and the experimental results show that compared with the existing PSO—SVM model and BA—SVM model, the MFPA—SVM model put forward in this paper has better learning ability and generalization ability.

References:
[1] Kun-Chieh W. The grey-based artificial intelligence modeling of thermal error for machine tools[J]. JOURNAL OF GREY SYSTEM, 2010, 22(4): 353-366.
[2] Guo Q, Xu R, Mao C, et al. Application of information fusion to volumetric error modeling of CNC machine tools[J]. The International Journal of Advanced Manufacturing Technology, 2015, 78(1-4): 439-447.
[3] Chen Z, Liu X, Peng D, et al. Dynamic model of NC rotary table in angle measurements with time series[J]. Transactions of the Institute of Measurement and Control, 2012: 0142331212437992.
[4] Huang Y, Zhang J, Li X, et al. Thermal error modeling by integrating GA and BP algorithms for the high-speed spindle[J]. The International Journal of Advanced Manufacturing Technology, 2014, 71(9-12): 1669-1675.
[5] Suganyadevi M V, Babulal C K. Support vector regression model for the prediction of loadability margin of a power system[J]. Applied Soft Computing, 2014, 24: 304-315.
[6] Jiang M, Luo J, Jiang D, et al. A Cuckoo Search-Support Vector Machine Model for Predicting Dynamic Measurement Errors of Sensors[J]. IEEE Access.
[7] Xue Z, Du P, Su H. Harmonic analysis for hyperspectral image classification integrated with PSO optimized SVM[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2014, 7(6): 2131-2146.
[8] Enache A C, Sgârciu V. Anomaly intrusions detection based on support vector machines with bat algorithm[C]//System Theory, Control and Computing (ICSTCC), 2014 18th International Conference. IEEE, 2014: 856-861.
[9] Yang X S. Flower pollination algorithm for global optimization[C]//International Conference on Unconventional Computing and Natural Computation. Springer Berlin Heidelberg, 2012: 240-249.
[10] Bensouyad M, Saidouni D E. A discrete flower pollination algorithm for graph coloring problem[C]//Cybernetics (CYBCONF), 2015 IEEE 2nd International Conference on. IEEE, 2015: 151-155.
[11] Prathiba R, Moses M B, Sakthivel S. Flower pollination algorithm applied for different economic load dispatch problems[J]. International Journal of Engineering and Technology, 2014, 6(2): 1009-1016.
[12] Rodrigues D, Yang X S, De Souza A N, et al. Binary flower pollination algorithm and its application to feature selection[M]//Recent Advances in Swarm Intelligence and
Evolutionary Computation. Springer International Publishing, 2015: 85-100.
[13] Ochoa A, González S, Margain L, et al. Implementing flower multi-objective algorithm for selection of university academic credits[C]/Nature and Biologically Inspired Computing (NaBIC), 2014 Sixth World Congress on. IEEE, 2014: 7-11.
[14] Zheng W, Fu H, Yang G. Targeted Mutation: A Novel Mutation Strategy for Differential Evolution[C]/Tools with Artificial Intelligence (ICTAI), 2015 IEEE 27th International Conference on. IEEE, 2015: 286-293.
[15] Ou-yang Gaibin, Gao Liqun, Kong Xiangyong.A Differential Evolution Algorithm of Random Evolution [J]. Journal of Northeastern University (Natural Science), 2013, 34(3):330-334.
[16] Walton S, Hassan O, Morgan K, et al. Modified cuckoo search: a new gradient free optimisation algorithm[J]. Chaos, Solitons & Fractals, 2011, 44(9): 710-718.