Parameters Optimization of SVM using the Dual Population ACO

Huaping Zhou, Qinghong Zhu

College of computer Science and Engineering, Anhui University of Science and Technology, Huainan, Anhui, China
573023738@qq.com

Abstract. The support vector machine (SVM) parameters optimization of previous lacks of theoretical guidance. The algorithm is easy to fall into local optimum. The dual population ant colony algorithm is helpful to SVM parameters optimization in ability. The independent solution and exchange of two population information and the change of pheromone can avoid the stagnation of the algorithm or get into local optimum, and find the global optimal solution. The accuracy of SVM classification is used as the objective function. And the experimental results show that the algorithm improves the accuracy of the classification of support vector machine parameters.

1. Introduction
Support vector machine (SVM) is Vapnik et first proposed in 1995, involving statistical theory [1], in solving the small sample, non-linear and high dimensional space recognition shows unique advantages. SVM can better extended to function fitting and other machine learning [2]. In the actual problem of support vector machine (SVM) still need to improve, such as its parameters. There are a lot of research method in parameters optimization [3] of it, such as grid search method [4], experimental method, based on the bionics intelligent algorithm and so on. The intelligent algorithm based on bionics includes genetic algorithm [5,6,7], particle swarm algorithm [8,9], fish swarm algorithm [10], etc. Which can be optimized for SVM parameters [11] according to their respective algorithm characteristics, but it is easy to get into local optimal.

Ant colony algorithm is a simulated evolutionary method based on population, In recent years, many researchers on the algorithm proposed improvements. Changes in pheromones and volatilization [12,13] factors, avoiding the precocious algorithm, and overcoming the shortcomings of algorithm searching time too long. Besides, it has been used in the travel business [14] and other problems, and has a good global search capability, which is easy to combine with other algorithms. Ant colony algorithm [15,16] has been applied to parameter optimization of support vector machine (SVM), but the study of algorithm is slow to convergence speed and classification accuracy is not high. In order to improve the classification accuracy of SVM, the ant colony algorithm using the support vector machine (SVM) parameters optimization. Through the simulation results show that the improved algorithm has better classification accuracy compared with other algorithms.

2. Support Vector Machine Introduction
A set of training samples is set in support vector machines for example:

\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}
$y_i \in \{+1, -1\}, i=1, 2, 3, \ldots, n$, represent sample categories. A linear classifier in the N dimensional space to find a hyperplane. This hyperplane formula for:

$$g(x) = w^T x + b = 0$$

In the formula, $w$ stands for weight vectors. $T$ stands for transpose, and $b$ stands for threshold. In order to minimize structural risk, constraint conditions are introduced:

$$y_i (w^T x + b) \geq 1$$

Considering the existence of categorical errors in a certain range, a nonnegative relaxation variable $\xi_i$ is introduced. Then optimal hyperplane quadratic programming problem is:

$$\begin{aligned}
\min_{w, b, \xi} & \frac{1}{2} w^T w + c \sum_{i=1}^{n} \xi_i \\
y_i (w^T x_i + b) & \geq 1 - \xi_i
\end{aligned}$$

By introducing the Lagrangian multiplier, the formula (3) is transformed into a dual form:

$$\begin{aligned}
\min_{\alpha} & \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j (\varphi(x_i)\varphi(x_j)) + \sum_{i=1}^{n} a_i \\
\sum_{i=1}^{n} \alpha_i y_i & = 0 (c \geq \alpha_i \geq 0)
\end{aligned}$$

After the introduction of the kernel function, the final decision function is:

$$f(x) = \text{Sgn}(\sum_{i=1}^{n} \alpha_i y_i k(x_i, x) + b)$$

Different kernel functions will generate different support vector machine classifiers. In the radial basis function, only one parameter $\delta$ is needed, which saves a lot of trouble. This kernel function is used in most studies, so the radial basis function is also chosen in this paper. The formula is defined as:

$$k(x, x_i) = \exp\left(-\frac{|x-x_i|^2}{2\delta^2}\right)$$

The penalty factor $c$ and the radial basis function parameter $\delta$ have great influence on the classification results. Therefore, it is important to select the appropriate parameters for $c$ and $\delta$.

3. Improvement of Ant Colony Algorithm

3.1. Local Search Rule

The initialization of the $c$ and $\delta$ allows the training set to derive an error model from SVM learning:

$$\Delta t(i) = \alpha^{\text{error}(i)}$$

A pheromone based on the model to determine the location of the ants:

$$T_0(i) = \alpha^{\text{error}(i)} (\alpha = 3, \text{from the formula, the smaller the error, the greater the pheromone})$$

The next transfer probability $p(i)$ of the ant is:

$$p(i) = \frac{e^{T_0(Best)-T_0(i)}}{e^{T_0(Best)}} (\text{Best is the most pheromone ant})$$

In order to find the optimal solution in the dual population, the pheromone volatilization factor was defined as:

$$\rho = k \times \frac{\log_{2} \times NC}{e^{\text{NC}_{\text{max}}}}$$

NC represents the current number of iterations. And NC_max represents the maximum number of
iterations. k=0.1.

The global transfer factor is determined by the size of the pheromone per iteration. In a double population, the number of ants is m. According to $e^{-\tau^{(i)}}$, it gets ascending order $T_i(j)$, j=1, 2, 3, ......., m.

$$NC < \frac{\text{max} \alpha \beta \text{NC}}{2}, \quad p_0 = t_i(k), k = \frac{2}{3} m.$$ Not then $p_0 = t_i(k), k = \frac{1}{3} m$. If $p(i) < p_0$, perform local search. Otherwise global search is performed.

3.2. Information Element Update Formula

$$\tau(i, j) = (1 - \rho) \cdot \tau(i, j) + \sum_{k=1}^{m} \Delta \tau^k_{ij}$$

$$j \in \text{allowed}_k, \Delta \tau^k_{ij} = \frac{H(x_i) - H(x_j)}{\lambda}$$

(11)

$x_i, x_j$ respectively represents the position of ant (i, j) before the optimal solution. \(\lambda\) Represents the possible impact of the dual population, but the effect is small and negligible. This shows that the closer the ants are, the higher the concentration of pheromones.

Exchange rule: At the same time, in the dual population, in order to avoid falling into local optimum or stagnation, 5 times a iteration. The exchange rules can be used when the pheromone difference between the two populations is large.

4. Specific Algorithm Steps

Step 1: Initialize ants with the same number of m between two populations. Initializes a given set of values, c and $\delta$, And given the range of values, the number of iterations NC, related parameters $\alpha, \beta, NC \text{ max}$. The ants position corresponds to a set of parameters $(c, \delta)$ of SVM.

Step 2: Set SVM corresponding data sets, trains data sets, train data labels, target data sets, target data labels.

Step 3: Two population are searched globally according to the formula of transition probability. Ants that satisfy local search are searched according to local search rules. At the same time, the target function value is calculated. According to formula(11), the information element is updated. The exchange of information that satisfies the rules of exchange. Finally, the optimal ant in two species population is obtained.

Step 4: Determine whether the number of iterations NC reaches the maximum number of iterations NC_max. If not, turn second steps, and if so, proceed to the next step. Each iteration compares the optimal solution between the population and takes the optimal solution between them.

Step 5: At the end of the algorithm, the parameters $(c, \delta)$ corresponding to the optimal ant are output.

5. Parameter Optimization of Support Vector Machine based on Improved Double Population Ant Colony Algorithm.

The classification accuracy of SVM is taken as the objective function and denoted as $H(x)$. And then we set upper and lower bounds respectively. That is $\max H(c, \delta), c \in (c_{\min}, c_{\max}), \delta \in (\delta_{\min}, \delta_{\max})$.

The dual population ant colony algorithm is used to select the most suitable parameter pairs. The algorithm flow diagram is shown in Figure 1:
6. Experimental Result

6.1. Experimental Data
The experimental data are derived from 4 data sets on the UCI web site. Data sets are Wine, Iris, Liver disorders, Statlog. We use particle swarm optimization to optimize support vector machine (PSO-SVM) for comparison with genetic algorithm optimization support vector machine (GA-SVM). This approach proves the effectiveness of the improved algorithm for optimizing parameters.

| Data sets    | Data length | Dimension | Class number |
|--------------|-------------|-----------|--------------|
| Wine         | 178         | 13        | 3            |
| Iris         | 150         | 4         | 3            |
| Liver disorders | 345       | 6         | 2            |
| Statlog      | 270         | 13        | 2            |

6.2. Experimental Environment
Cpu 2.3GHz, Memory 4G, We use Libsvm-3.22 developed by professor Lin chi-jin of Taiwan. The algorithm parameters of this paper are designed as follows:

- The number of the two population is m=10 respectively. Pheromone volatilization coefficient $\rho$ =0.8, $\alpha$=2, $\beta$=4, Maximum iterations NC_max=10. Parameters $c \in (0,10)$, $\delta \in (0,1)$. When using a dataset, it run 10 times for each data set. After that, we take the average and get the result. GA-SVM of the population size is 20. The crossover probability is 0.6. The mutation rate was 0.06. PSO-SVM of the population size is also 20. Learning factor $c_1$=2.5, $c_2$=0.5.

6.3. Experimental Analysis

| Data sets    | Different algorithms | $c$       | $\delta$    | Accuracy rate |
|--------------|----------------------|-----------|-------------|---------------|
| GA-SVM       |                      | 0.7842    | 0.3971      | 96.2137       |
| Wine         | PSO-SVM              | 3.1635    | 0.5738      | 97.4660       |
|               | ACO-SVM | GA-SVM | Iris  |
|---------------|---------|--------|-------|
|               | 5.0349  | 6.1463 | 2.4316|
| PSO-SVM       | 5.9728  | 2.1076 | 4.3719|
|               | 2.1193  | 7.1193 | 6.8856|
| Liver disorders| 7.4123  | 7.5388 | 7.5388|
|               | 4.3719  | 6.8856 | 7.1193|
| Statlog       | 5.0349  | 6.1463 | 2.4316|
|               | 5.9728  | 2.1076 | 4.3719|
|               | 2.1193  | 7.1193 | 6.8856|
|               | 7.4123  | 7.5388 | 7.5388|

The parameter pairs and the optimal parameters selected by the training set are tested on the given data set. The classification results are shown in the table 2. As can be seen from the results, the accuracy of the three algorithms is higher than that of the other two algorithms in ACO-SVM. Compared with GA-SVM and PSO-SVM, this paper presents the convergence of the algorithm in SVM model. The results of classifying different data sets are shown in visual, as shown in Figure 2 ~Figure 5.

Compared with the support vector machine parameter optimization based on genetic algorithm and particle swarm algorithm, SVM classification model parameters (C, δ) select the optimal solution to achieve more accurate classification analysis. From Figure 2 ~Figure 5, it is proved that the algorithm has faster convergence speed and better learning in the four data sets in UCI. Furthermore, it is proved that the improved double population ant colony algorithm is effective and feasible for SVM parameter optimization.

Figure 2 Wine classification effect diagram
Figure 3 Iris classification effect diagram
7. Summary
In support vector machines (SVM) classification, SVM has many parameters. The kernel parameter and penalty factor are the most discussed. Parameter optimization has great significance for the whole classification effect. The algorithm of this paper has improved ant colony algorithm. In local search, pheromone changes lead to a change in the overall rule. In addition, the pheromone exchange between two populations satisfying certain conditions also avoids falling into local optima and stagnation. The whole algorithm has a better effect on the parameter optimization of SVM, and finally directly influences the accuracy of classification. The experimental show that the ACO-SVM has faster searching ability than GA-SVM and PSO-SVM, and the classification effect is more significant. Although the experiment proves that the classification effect is better than the other two intelligent algorithms, the effect of classification would be different for different kernel functions. Therefore, the later task is how to choose the kernel function to get better classification effect.

Acknowledgment
Project supported by the National Natural Science Foundation of China(5574257); project of outstanding young talents in Universities in Anhui(ZY289)

References
[1] Vapnik V N. The Nature of Statistical Learning Theory. Springer, 1995
[2] Yan You-biao, Chen Yuan-y. Summarization of main strategies for machine learning. computer application research, 2004, 21 (7): 4-10
[3] Li Jiao. Research on Parameter Optimization of support vector machines. Huazhong Normal University, 2011
[4] Wang Xing-ling, Li Zhan-bin. Determination of kernel function parameters of support vector machines based on mesh search. Journal of Ocean University of China (NATURAL SCIENCE EDITION), Natural Science Edition, 2005, 35 (5): 859-862.
[5] Dai Shang-ping, Song Yong-dong. Parameter selection of support vector machines based on genetic algorithm and particle swarm optimization. computer engineering and science, 2012, 34 (10): 113-117.
[6] Zheng C, Jiao L. Automatic parameters selection for SVM based on GA// Global Conference on intelligent control and automation wcica2004. 2004:1869-1872 Vol.2.
[7] Zhang Ying-l-u. Prediction of network traffic based on genetic algorithm optimized support vector machines. computer science, 2008, 35 (5): 177-179.
[8] Shao Xin-guang, Yang Hui-zhong, Chen Gang. Support vector machine parameter selection and its
application based on particle swarm optimization. control theory and application, 2006, 23 (5): 740-743.

[9] Wang Zhen-wu, Sun Jia-jun, Yin Gang. Improved support vector machines with particle swarm optimization and their applications. Journal of Harbin Engineering University, 2016, 37(12): 1728-1733.

[10] Li Xiao-lei. A new intelligent optimization method artificial fish swarm algorithm. Zhejiang University, 2003.

[11] Li Yan-ying. Research on swarm intelligence optimization algorithm based on Parameter Optimization of support vector machines. Tianjin University, 2007.

[12] Zhu Qing-bao, Yang Zhi-jun. Ant colony optimization algorithm based on mutation and dynamic pheromone updating. Journal of software, 2004, 15 (2): 185-192.

[13] Zhang Yong-qiang, Wang Xiao-dong. Improved ant colony algorithm based on pheromone updating and volatility factor. Journal of Xi'an Polytechnic University, 2016, 30 (3): 400-404.

[14] Xu Kai-bo, Lu Yan, Cheng Bi, et al. Improved pheromone, two update and local optimization ant colony algorithm for TSP. Journal of Computer applications, 2017, 37 (6): 1686-1691.

[15] Liu Chun-bo, Wang Xian-fang, Pan Feng. Parameter selection and Simulation of support vector machines based on ant colony optimization algorithm. Journal of Central South University (NATURAL SCIENCE EDITION), 2008, 39 (6): 1309-1313.

[16] Zhang Pei-lin, Cao Jian-jun, et al. Money Fang Lin, ant colony algorithm for parameter optimization of support vector machine based on. Journal of Nanjing University of Science and Technology (NATURAL SCIENCE EDITION), 2009, 33 (4): 464-468.