A novel SVM parameter tuning method based on advanced whale optimization algorithm

Xuehao Yin¹, YanDong Hou², Jiabao Yin¹, and Chao Li¹

¹Department of Computing, Jiangsu Automation Research Institute, Lian Yungang, Jiangsu, 222000 China
²School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin, Heilongjiang, 150000, China

Abstract. The classification performance of support vector machine (SVM) algorithm is highly dependent on the careful tuning of hyper-parameters and penalty coefficient. This paper introduces a novel SVM parameter optimization method by using the advanced whale optimization algorithm (AWOA) that is an improved whale of algorithm (WOA) with external archiving strategy. A new framework for SVM parameter optimization based on AWOA is built. To demonstrate the performance of our proposed method, six typical data sets are chosen to evaluate the effect of SVM classification problem. Experimental results show that the higher accuracy and better convergence can be achieved by AWOA compared with other three usual parameter optimization methods (WOA, PSO, and DE).

1. Introduction
Support Vector Machine (SVM) is a machine learning method based on statistical learning theory with the principle of structural risk minimization. It has been widely applied in various fields such as text categorization, sample classification, image recognition, modeling and control, state estimation and prediction and so on.

However, the performance of SVM will depend seriously on parameter selection and setting. At the same time, many scholars have carried out extensive research on the parameter optimization of SVM algorithm. Currently, there are mainly three kinds of methods used to SVM parameter optimization including empirical selection, grid search, and swarm intelligence optimization. The empirical selection method requires researchers to repeatedly choose based on experience and experiments, which has a large randomness and is difficult to obtain a global optimal solution. The grid search algorithm is an alternative to improve the classification accuracy of SVM, but it is time-consuming and depends on the grid size. In comparison, the SVM parameter optimization based on swarm intelligence algorithm will be a promising way. Typically, the literature [1] used genetic algorithm (GA) to encode parameter combination, then find the optimal solution and decodes it. Although the SVM parameters can be optimized by GA algorithm to some extent, it still has the disadvantages of slow convergence and large computation burden. [2] adopted particle swarm optimization (PSO) to improve the classification accuracy of SVM, but it is easy to get stuck in local optimal without finding global optimal solution. Differential evolution (DE) is also applied to tune the parameters of SVM [3], but the performance may highly depend on the chosen trial vector generation strategy and associated parameter values. Combining the artificial bee colony algorithm (ABC) with SVM was also proposed
in [4], but it is poor at local search ability, and its efficiency decreases obviously when it is near optimal solution.

Whale optimization algorithm (WOA) based on bubble-net hunting behavior of humpback whales was proposed in 2016 [5]. Compared with other swarm intelligence algorithms, WOA has been proved to be more competitive and has been successfully applied in various engineering optimization problems. However, through exploring optimization mechanism, it’s not difficult to find that search agents get only close to the best solution over the latter half process of iterations, which makes WOA easily trap into local optima [6], resulting in losing population diversity and poor at balancing exploitation and exploration. In order to solve these defects of WOA, some strategies should be developed, meanwhile the improved WOA will be a promising way for SVM parameter optimization.

As discussed above, this paper proposed a novel method called advanced whale optimization algorithm (AWOA) to address the issue of SVM parameter optimization. The AWOA is an improved WOA with the features of external archive strategy. When performing a random search operation, AWOA algorithm selects whale individuals from the current generation and archive, and the optimal solution is selected from the optimal solution set that ensure AWOA algorithm has the advantages of rapid convergence on the early stage, high-precision optimization and jumping out of the local optimal solution during the later stage.

2. An improved SVM parameter optimization method

2.1. Support vector machine

The SVM algorithm is based on the principle of minimizing the structural risk, relying on a small number of samples and considering both training error and test error, so as to obtain the optimal promotion effect. For the binary classification problem, given a sample set \( T = \{(x_1, y_1), \ldots, (x_N, y_N)\}, y \in \{-1, 1\} \), SVM algorithm constructs a function \( f(x) \) describing the relationship between input and output through the training model:

\[
 f(x) = \text{sgn}(\omega^T \cdot \phi(x) + b), \quad i = 1, 2, \ldots, N
\]

(1)

Where \( N \) is the number of samples, \( \omega \) is the normal vector of the hyperplane, \( b \) is the deviation term, \( x_i = (x_{i1}, \ldots, x_{iD}) \) represents the input data of each sample, \( D \) is the number of features, \( \phi(x) \) is the mapping function, and \( f(x) \) represents the output.

Considering the criterion of structural risk minimization, the SVM classification problem is expressed as the following optimization problem with constraints:

\[
 \min_{w, b, \xi} J(w, b, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} \xi_i
\]

\[
 \text{s.t.} \quad \left\{ y_i (\omega^T \cdot \phi(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \right\}
\]

(2)

(3)

Where \( \xi_i \) is a relaxation variable, and \( C \) is the penalty coefficient. The larger parameter \( C \) is, the greater penalty will be for the samples of wrong classification. And the classification interval is small, making the generalization ability of classifier insufficient to adapt to the unknown samples. Due to this problem belongs to convex quadratic optimization, we can obtain the Lagrange function as follows when the Lagrange multiplier is introduced.

\[
 L := J(w, b, \xi) - \sum_{i=1}^{N} \alpha_i \left[ y_i f(x_i) - (1 + \xi_i) \right] - \sum_{i=1}^{N} \mu_i \xi_i
\]

(4)

Calculating the partial derivatives of coefficient \( \alpha \), deviation \( b \), relaxation variable and Lagrange multiplier respectively and set to 0, the following equations can be obtained:

\[
 w = \sum_{i=1}^{N} \alpha_i y_i \phi(x_i)
\]

(5)
Substituting the above equations into equation (4), the dual form of the optimization problem can be obtained, and the final prediction model can be written as:

$$f(\mathbf{x}) = \sum_{i=1}^{N} a_i y_i k(\mathbf{x}, \mathbf{x}_i) + b$$

Where $k(\mathbf{x}, \mathbf{x}_i) = \phi(\mathbf{x})^T \phi(\mathbf{x}_i)$ is the kernel function, and mainly used to map the sample space to high-dimensional feature space. The common kernel function includes polynomial kernel function, radial basis kernel function and sigmoid kernel function. At present, radial basis kernel function is the most widely used kernel function. In this paper, radial basis kernel function is adopted as follows:

$$k(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\frac{|\mathbf{x} - \mathbf{x}_i|^2}{\sigma^2}\right)$$

Where $\sigma$ is the parameter of kernel function. The smaller parameter $\sigma$ is, the faster decay of the kernel function on the sample eigenvalue will be, and any small change in the eigenvalue will cause a large change for the value of kernel function, which is easy to cause overfitting.

### 2.2. Whale optimization algorithm

The WOA optimization algorithm is inspired by the hunting behavior of humpback whales that first walk randomly to look for prey, and then rise in a spiral trajectory way after finding the targets and gradually surround the prey. WOA models the relevant behavior and can be given as follows.

1) Searching for prey: At this stage, the whales do not know the specific location of the prey, and they walk randomly according to the location of each other to forage. In this way, the whale algorithm has the ability of global search, so as to obtain the global optimal solution. The formula for location update can be described as follows:

$$\mathbf{D} = \left| \mathbf{C} \cdot \mathbf{X}_{\text{rand}} - \mathbf{X}_i \right|$$

$$\mathbf{X}_{i+1} = \mathbf{X}_{\text{rand}} - \mathbf{A} \cdot \mathbf{D}$$

Where $\mathbf{X}$ is the position of the whale individual, and $t$ represents the number of iteration, and $\mathbf{X}_{\text{rand}}$ is the position vector of whale individual randomly selected from the current generation. $\mathbf{A}$ and $\mathbf{C}$ are both coefficient vectors, and can be expressed as follow:

$$\mathbf{A} = 2a \cdot \mathbf{r} - a$$

$$\mathbf{C} = 2 \cdot \mathbf{r}$$

Where $\mathbf{r}$ is a random vector ranging from 0 to 1. The whale individuals are forced to approach random individual $\mathbf{X}_{\text{rand}}$ for foraging when $|\mathbf{A}| > 1$. $a$ is a variable linearly decreasing from 2 to 0 and can be calculated as in equation (14).

$$a = 2 - 2t / t_{\text{max}}$$

Where $t_{\text{max}}$ is the max number of iterations.

2) Encircling prey: When $|\mathbf{A}| \leq 1$, the whale population $\mathbf{X}_{\text{best}}$ will be regarded as the global optimal individual of current generation, and the mathematic model of this stage can be summarized below.

$$\mathbf{X}_{i+1} = \mathbf{X}_{\text{best}} - \mathbf{A} \left| \mathbf{C} \cdot \mathbf{X}_{\text{best}} - \mathbf{X}_i \right|$$

3) Bubble-net attacking: When surrounding the target prey, the whale populations will calculate the distance from the global optimal individual, and then hunt for prey in a spiral trajectory way. The updating position is as follows:

$$\mathbf{X}_{i+1} = \mathbf{D} \cdot e^{at} \cdot \cos(2\pi t) + \mathbf{X}_{\text{best}}$$
Where $D' = |X_{\text{best}} - X|$ shows the distance between the whale individuals and the current optimal position vector, and $b$ is a constant defining the spiral shape.

It needs to be noted that encircling prey and bubble-net attacking are executed synchronously, and its updating mode is decided by random $P$. When $P < 0.5$, the encircling prey mechanism is adopted, while the spiral predation model is selected instead to update the whale position.

2.3. Advanced whale optimization algorithm

As a new meta-heuristic optimization algorithm, WOA algorithm has the advantages of simple structure, easy implementation and few parameters to be adjusted. The main parameters of the WOA algorithm is vector $A$ that enables WOA to balance exploitation and exploration capabilities. When $|A| > 1$, equation (11) makes the updated position of whale individual be any position between the current position and the random individual, which is similar to the global search. When $|A|$ is in [-1,1], equation (15) and equation (16) make the whale individual gradually surround and attack the prey in a spiral trajectory way, which can be considered as local exploitation. And the value of $A$ depends on the variant $a$ linearly decreasing from 2 to 0, and $A$ value range is $A \in [-a, a]$. Therefore, if $a \leq 1$, then $|A| \leq 1$. According to equation (14), when the number of iterations exceeds half of $t_{\text{max}}$, then the whale populations will no longer forage randomly, but only approach to the current optimal individual. Therefore, at the early stage, WOA has better global search capability, but with slower convergence rate. At the late phase, the local development capability is enhanced and the search speed is accelerated, but the population diversity may lose and easily fall into local optimization, resulting in low convergence accuracy. It is clear that WOA will achieve better search accuracy when it balances its exploration and development capabilities.

In order to improve convergence accuracy and robustness of WOA, a novel WOA algorithm with external archiving strategy is proposed and the specific improvements are described as follows:

1) Establishing external archiving: The basic WOA algorithm simply discards the original prey in the process of location updating. That is, the quality of prey source itself is not evaluated when the position vector is updated, thus only the population $X_{\text{best}}$ is kept. This approach cannot effectively and quickly extract useful information from the search space, which will affect WOA’s search efficiency to some extent. AWOA uses the greedy selection mechanism to update the position vector. If the fitness of the updated solution is better than that before the update, the inferior solution before the update will be archived; otherwise, the position vector before the update will be retained. In the next random walk foraging process, the following formula is used to update the position vector:

$$X_{+1} = \tilde{X}_{\text{rand}} - A \cdot [C \cdot \tilde{X}_{\text{rand}} - X]$$

Where $\tilde{X}_{\text{rand}}$ represents the individuals randomly selected from the set $G$ and $I$. $G$ is the current population, and $I$ is the archived populations. The strategy not only accelerates the convergence but also increases the diversity of the population and finally improves the ability of global exploration.

2) Building optimal solution set: At the end of iterations, all populations of the basic WOA algorithm approach the optimal solution in the current generation. AWOA takes the former $k$% optimal individual in current generation as the optimal solution set, and when the whale swarm shrinks to surround prey with spiral trajectory, one prey is randomly selected from the optimal solution set. The location is updated as follows:

$$X_{+1} = \begin{cases} X_{\text{best}} - A \cdot [C \cdot X_{\text{best}} - X] & \text{if } p < 0.5 \\ D' \cdot e^{b'} \cdot \cos(2\pi t) + X_{\text{best}} & \text{if } p \geq 0.5 \end{cases}$$

Where $D' = |X_{\text{best}} - X|$ shows the distance between the whale individuals and the current optimal position vector, and $b'$ is a constant defining the spiral shape.
Where $X^k_{best}$ is a whale individual randomly selected from the pre-optimal individuals of the current generation. This strategy can effectively alleviate the precocity and help the whale population to jump out of local optimum. The specific pseudo code of AWOA algorithm is shown as follows:

**AWOA algorithm**

Initialize the whale population, and the initial external archive is empty  
Calculate the fitness of each whale individual, and record the previous k% optimal whale individual and position  
$X^* =$the best whale individual  
while (t <maximum number of iterations)  
    for each whale individual  
        Update $a$, $A$, $C$, $p$, $b$, and $l$  
        if1 ($p < 0.5$)  
            if2 ($|A| > 1$)  
                Randomly select an individual $X_{ext}$ from the set $G \cup I$  
                Update the position of current search agent by equation (17)  
            else  
                Randomly select a search agent from the optimal solution set $X^k_{best}$  
                Update the position of current search agent by equation (18)  
        end if2  
    else  
        Randomly select individuals from the optimal solution set $X^k_{best}$  
        Update the position of current search agent by equation (18)  
    end if1  
end for  
Check if any search agent goes beyond the search space and amend it  
Calculate the fitness of each search agent  
Update $X^*$ by greedy mechanism  
Store the inferior solution in the external archive. If the size of archive exceeds $N$, delete the particles in the archive randomly until the size of archive meets the condition.  
$t = t+1$  
end while  
return $X^*$

2.4. AWOA-SVM classification model

Many researchers have revealed that the parameters $\sigma$ and $C$ largely determine the classification accuracy and robustness of SVM algorithm [7]. Therefore, optimizing the parameters combination ($\sigma$, $C$) is an important means to improve the performance of SVM algorithm. In this paper, the proposed AWOA algorithm is used to adjust parameters combination ($\sigma$, $C$), so that the SVM model with the optimal parameters has the highest classification accuracy. The extraction process of optimal parameter based on AWOA algorithm is shown in figure 1.
Figure 1. SVM parameter optimization framework based on AWOA

During the process of parameter optimization with AWOA, the position of each individual, namely the feasible solution, is the parameter combination $(\sigma, C)$. The fitness function determines the search direction, and three-fold cross validation precision [8] is chosen as the objective function, meaning the sample set is divided into three parts. Take one subset as test samples respectively, and the remaining two subsets are training set. Then it will receive three classification model, and we will take the average classification accuracy of these model as the objective function. This method can effectively reduce the risk of overfitting, and the classification results obtained are more reliable.

3. Experiments and result discussion
To verify the feasibility and effectiveness of the proposed method in SVM parameters optimization problem, six typical classification data sets of UCI standard database (http://archive.ics.uci.edu/ml/) are selected as the experimental data. As shown in table 1, the data sets selected in this paper consider both high and low dimension attribute, so the evaluation of algorithm classification performance is more objective.

| Label                | Number | Dimension | Categories |
|----------------------|--------|-----------|------------|
| Ionosphere           | 351    | 34        | 2          |
| Wine                 | 178    | 13        | 3          |
| Balance Scale        | 625    | 4         | 3          |
| Seeds                | 210    | 7         | 3          |
| Yeast                | 1484   | 8         | 10         |
| Forest type mapping  | 367    | 27        | 4          |

3.1. Experimental data and settings
The experimental environment is a 3.80GHz AMD Athlon(tm) X4 760K CPU with 4G RAM memory operating on Windows7 operating system. The optimal parameter mining of SVM for our training sample data is realized in Matlab2014a, and the LibSVM model toolbox was called. Additionally, the
The proposed AWOA algorithm is compared with other typical SVM parameters optimization algorithm, including basic WOA, PSO, and DE.

The same parameters of the four algorithms include the number of population set 10 and the number of iterations is 50. Inertia weight $\omega_{\text{min}} = 0.9$, $\omega_{\text{max}} = 0.2$ and velocity parameters $c_1, c_2 = 2$ in PSO. For DE algorithm, scaling factor $F = 0.5$, crossover probability $C_r = 0.9$. The WOA algorithm parameter settings are the same as [6]. The length of the newly added parameter external archive in AWOA is the same as the number of population, and the optimal solution takes the first 30% of the optimal individuals. The range of optimized parameters combination $(\sigma, C)$ are respectively $(0, 10]$ and $(0, 1000]$.

### 3.2. Experiments on SVM parameter optimization

The settings of the six training samples are shown in table 1. Mean value, standard deviation and maximum value of classification results were selected to evaluate the performance of the four optimization method. The optimization results of the four parameter optimization algorithms are shown in table 2, where the optimization results of each parameter optimization method are denoted as PSOSVM, DE SVM, WOASVM and AWOASVM. It should be noted that table 2 shows the results of thirty experiments with three-fold cross validation.

| Data sets   | Assessment criteria | PSOSVM | DE SVM | WOASVM | AWOASVM |
|-------------|---------------------|--------|--------|--------|---------|
| Ionosphere  | max                 | 95.7265| 96.0056| 96.0114| 96.2963 |
|             | mean                | 94.9287| 95.4701| 95.6410| 95.9544 |
|             | std                 | 0.2942 | 0.6162 | 0.4414 | 0.2703  |
| Wine        | max                 | 98.3146| 99.4256| 99.3918| 99.4382 |
|             | mean                | 98.3146| 99.3258| 99.3258| 99.4382 |
|             | std                 | 1.16e-14| 0.2834| 0.3553 | 1.49e-14|
| Balance Scale| max               | 98.24  | 98.24  | 98.24  | 98.272  |
|             | mean                | 98.208 | 98.24  | 98.24  | 98.272  |
|             | std                 | 0.1126 | 0.1394 | 0.1052 | 0.0675  |
| Seeds       | max                 | 95.2381| 95.2381| 95.2943| 95.7143 |
|             | mean                | 95.0476| 95.0952| 95.2381| 95.2857 |
|             | std                 | 0.2300 | 0.2459 | 0.0049 | 0.1506  |
| Yeast       | max                 | 60.3241| 60.5489| 61.0860| 61.2534 |
|             | mean                | 60.1859| 60.3445| 60.4447| 61.0377 |
|             | std                 | 0.4132 | 0.3483 | 0.7269 | 0.3448  |
| Forest      | max                 | 87.2128| 87.3846| 87.5857| 87.6923 |
|             | mean                | 86.8462| 86.7692| 87.4869| 87.6000 |
|             | std                 | 0.2314 | 0.2064 | 0.1752 | 0.1589  |

As can be seen from table 2, compared with PSO and DE, WOA has better precision. At the same time, AWOA has higher average data classification accuracy and minimum prediction standard deviation, indicating that AWOA has higher accuracy and better robustness for SVM parameter optimization. In order to further analyze the convergence of AWOA in SVM parameter optimization, the visualization effect of the four SVM optimization models on six data sets is shown in figure 2.
According to figure 2(a) ~ figure 2(f), AWOA algorithm has higher accuracy and better convergence than other algorithms for optimizing SVM parameters. It can be seen from figure 2(b), figure 2(d) and figure 2(f), when other algorithms were stagnant, AWOA can search further and find more accurate solutions, which reflects AWOA algorithm with a stronger ability to jump out of local optimum. Additionally, from figure 2(a), figure 2(c) and figure 2(e), we can see that AWOA algorithm has fast convergence when compared with other three methods. Thus, we can concluded that the AWOA algorithm we proposed has good effectiveness and feasibility for SVM parameter optimization.

4. Conclusions
In order to improve the classification performance of SVM algorithm, this paper proposes a parameter optimization method based on AWOA algorithm. The proposed AWOA is an improved WOA algorithm with external archiving strategy, which effectively overcomes the problem that traditional
WOA algorithm has a fast convergence speed on the early stage and is prone to fall into local optimization during the later iterations. AWOA assesses the quality of the prey as the whale population updates its location, and archives the inferiority. When performing a random search operation, AWOA algorithm selects whale individuals from the current generation and archives. The optimal solution is selected from the optimal solution set during the encircling and preying operations. These two strategies make AWOA not only with the ability of rapid convergence, but also have the ability of high-precision optimization and jumping out of the local optimal solution, which provides a feasible method for SVM parameter optimization.

References
[1] Phan A V, Nguyen M L, Bui L T. Feature weighting and SVM parameters optimization based on genetic algorithms for classification problems[J]. Applied Intelligence, 2016, 46(2):1-15.
[2] Huang C L, Dun J F. A distributed PSO–SVM hybrid system with feature selection and parameter optimization[J]. Applied Soft Computing, 2008, 8(4):1381-1391.
[3] Bhadra T, Bandyopadhyay S, Maulik U. Differential Evolution Based Optimization of SVM Parameters for Meta Classifier Design[J]. Procedia Technology, 2012, 4(11):50-57.
[4] Yu M, Ai Y Q. SVM parameter optimization and application based on artificial bee colony algorithm[J]. Guangdianzi Jiguang/Journal of Optoelectronics Laser, 2012, 23(2):374-378.
[5] Mirjalili S, Lewis A. The Whale Optimization Algorithm[J]. Advances in Engineering Software, 2016, 95:51-67.
[6] Lun J, Shi BY. A hybrid whale optimization algorithm based on modified differential evolution for global optimization problems[J]. Applied Intelligence, https://doi.org/10.1007/s10489-018-1362-4.
[7] Ji Y, Chen Y, Fu H, et al. An EnKF-based scheme to optimize hyper-parameters and features for SVM classifier[J]. Pattern Recognition, 2017, 62(Complete):202-213.
[8] Stone M. 1974. Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society. Series B (Methodological), 36(2): 111–147 [DOI: 10.2307/2984809].