An improved support vector machine classifier based on artificial bee colony algorithm

Yongcun Cao\textsuperscript{1}, Saisai Ji\textsuperscript{*} and Yong Lu\textsuperscript{1}

\textsuperscript{1}School of Information Engineering, Minzu University of China, Beijing, 100081, China

\textsuperscript{*}Corresponding author’s e-mail: saisai_ji@163.com

Abstract. Support vector machine (SVM) has unique advantages in the classification of small sample data. The selection of parameters has an important impact on the classification accuracy and generalization ability of SVM. Since the selection of parameters of SVM is usually based on experience, an improved SVM classification model based on artificial bee colony algorithm (GOABC-SVM) is proposed in this paper. In this model, firstly, the traditional artificial bee colony algorithm is optimized using the ideas of global optimal solution guidance and opposite learning. Secondly, we set the reciprocal of classification error rate as the fitness function, and use the improved artificial bee colony algorithm to obtain the optimal parameter combination of SVM. Experiments on a set of datasets of UCI show that the proposed model has higher classification accuracy and better generalization ability.

1. Introduction

Support vector machine (SVM) is a machine learning method proposed by Vapnik et al. based on statistical learning theory [1]. SVM can effectively solve problems such as small samples, non-linearity, and high dimensions [2], and has good generalization ability and classification accuracy. As a pattern classification method, the classification results of SVM are mainly affected by factors such as kernel function parameters, penalty factor, and the type of kernel function selected. The generalization ability of the model also depends on the model parameters we choose to some extent. Therefore, how to choose the optimal combination of parameters to achieve the best classification results is a hot issue in the current research of SVM.

Because the penalty factor and kernel function parameters of SVM are usually set based on experience, the parameters obtained based on prior knowledge are not necessarily the actual optimal parameters. In addition, the generalization performance of SVM depends on multiple parameters, and these parameters also interact with each other. The optimal of a parameter cannot build an optimal SVM classification model, which further increases the complexity of parameter optimization. In practical applications, experimental method, grid search method, and gradient descent method are often used to optimize the parameters of SVM. The main idea of the experimental method is to try different parameters to train the classification model, and finally choose a suitable combination of parameters according to the experimental results. This method lacks theoretical guidance. It depends on the practical experience of the operator entirely. The grid search method [3] successively traverses the points in the grid range and selects the point with the minimum error as the optimal parameter of the SVM. This method generally improves the search accuracy by reducing the search step size or increasing the parameter range. As a result, the calculation complexity is high. The gradient descent method [4] is very sensitive to the initial value chosen, and easily falls into local optimum. As a kind
of parameter optimization method, swarm intelligence algorithm has the characteristics of fast optimization and global optimization. It has been widely used in the parameter optimization of SVM. At present, the commonly used optimization methods include particle swarm optimization algorithm (PSO) [5], ant colony optimization algorithm (ACO) [6] and genetic algorithm (GA) [7]. However, due to the multi-peak function relationship between the classification accuracy of SVM and the parameters to be optimized (kernel function parameters and penalty factor), the above-mentioned optimization algorithm will fall into the local optimal solution during the process of optimization, and resulting in failure to achieve optimal classification result.

The artificial bee colony algorithm (ABC) was proposed by Turkish scholar Karaboga in 2005 [8]. It is a population-based evolutionary algorithm and is inspired by the foraging behaviour of bees. As a swarm intelligence algorithm, the ABC algorithm finds the optimal solution of the optimization problem through the cooperation between different types of bees and the collection and sharing of food source information. Therefore, using ABC algorithm to optimize the parameters of SVM has certain advantages. Since the ABC algorithm was proposed, it has received the attention of many researchers and has been successfully applied to many fields. Yang et al. [9] proposed a method for optimizing SVM parameters based on ABC algorithm, which was applied to gear fault diagnosis. Compared with other swarm intelligence algorithms, the experimental results show that this method can obtain higher accuracy in the least time. Mustaffa et al. [10] proposed an ABC algorithm to optimize the parameters of the least squares support vector machine (LS-SVM) and used it in the prediction of commodity prices. The experimental results show the ability of the proposed method to obtain higher prediction accuracy. Liu et al. [11] used an improved ABC algorithm based on fitness prediction strategies to optimize the parameters of LS-SVM and applied it to the prediction of telecommunication traffic. Xiao et al. [12] used empirical mode decomposition (EMD) and ABC algorithm to optimize the model parameters of SVM and used them in short-term prediction and fault early warning to improve the prediction accuracy of traditional SVM. In order to improve the accuracy of cancer classification, Gao et al. [13] used SVM based on PSO and ABC algorithm to classify the cancer data set. Experiments showed that the proposed model dealt with various types of cancer data well and the classification model was effective and robust.

In this paper, the ABC algorithm is applied to select the parameters of SVM. The optimization of parameters is considered as an optimization problem. The classification accuracy of the training set is used as the evaluation criteria, and use the ABC algorithm to select the optimal parameter combination. In this way, the classification accuracy and generalization ability of SVM are improved, and a classifier with better classification performance is constructed.

2. Proposed Method

2.1. The improved artificial bee colony algorithm
The ABC algorithm finds the optimal solution of the problem to be optimized by simulating the honey collecting behavior of the bees. The algorithm divides the bee colony into three types: employed bees, onlooker bees, and scout bees. The employed bees are responsible for looking for food resource and sharing information with other bees in the hive. The onlooker bees are responsible for collecting honey based on the information provided by the employed bees. After the nectar source is abandoned, the scout bees is responsible for randomly finding a new nectar source to replace the original nectar source. Like other swarm intelligence algorithms, the ABC algorithm is iterative. After initializing the number of bees and the location of the honey source, three processes are repeatedly executed to find the optimal solution of the problem. They are the employed bee stage, the onlooker bee stage, and the scout bee stage. The process of each phase is as follows:

In the stage of employed bees, employed bees use the following Eq. (1) to find new food sources:

\[ v_{ij} = x_{ij} + \phi_{ij} (x_{ij} - x_{k,j}) \]  

(1)
Where $x_{kj}$ represents the neighbourhood food source, $k \in (1, 2, \ldots, SN)$ and $i \neq k$. $\Phi_{ij}$ is a random number in [-1, 1].

Onlooker bees select food sources for exploiting according to the probability $P$ Eq. (2) based on information shared by employed bees.

$$P_i = \frac{\text{fit}_i}{\sum_{n=1}^{SN} \text{fit}_n}$$  \hspace{1cm} (2)

Onlooker bees use Eq. (1) to find new food sources to exploit, and use greedy selection strategies to select food sources with higher fitness function values.

If the food source is not updated after many times, the scout bee phase is started. In this phase, the scout bee uses Eq. (3) to randomly find a new food source to replace the abandoned food source.

$$x_{ij} = x_{ij}^{\min} + \text{rand}(0, 1)(x_{ij}^{\max} - x_{ij}^{\min})$$  \hspace{1cm} (3)

Aiming at that the ABC algorithm is easy to fall into local optimum and has slow convergence speed, inspired by PSO algorithm, GABC algorithm [14] introduces Gbest term to enhance the exploitation ability of ABC algorithm. The GABC algorithm finds new food sources by using Eq. (4).

$$v_{ij} = x_{ij} + \Phi_{ij}(x_{ij} - x_{kj}) + \psi_{ij}(y_j - x_{ij})$$  \hspace{1cm} (4)

Where $y_j$ is the global optimal solution, and $\psi_{ij}$ is a random number uniformly distributed between [0, C], and C is a nonnegative constant.

In this paper, inspired by the idea of opposite learning [15], the opposite solution is introduced in the employed bee stage to enhance the exploration ability of the algorithm. When the fitness value of the newly generated solution is smaller than the fitness value of the current solution, its opposite solution is generated according to Eq. (5). At the same time, combined with the idea of GABC algorithm, the Eq. (4) is used to search for new food sources during the stage of onlooker bees and employed bees.

$$\alpha x_{ij} = x_{ij}^{\max} + x_{ij}^{\min} - x_{ij}$$  \hspace{1cm} (5)

2.2. The SVM optimized by ABC algorithm

The radial basis function has a wide adaptable range and fewer parameters to be optimized, it can handle nonlinear separable problems. Therefore, this paper uses the radial basis function Eq. (6) as the kernel function.

$$k(x_i, x_j) = \exp(-\frac{||x_i - x_j||^2}{2\sigma^2}), \sigma > 0$$  \hspace{1cm} (6)

Therefore, when the ABC algorithm is used to optimize the parameters of the SVM, the parameters that need to be optimized include the penalty factor $C$ and the kernel function parameter $\sigma$. The setting of the penalty factor $C$ and the kernel function parameter $\sigma$ plays an important role in improving the classification accuracy and generalization ability of the SVM classifier.

The penalty factor $C$ represents the interval size and the weight of classification accuracy. When $C$ approaches 0, the classification accuracy is low and prone to underfitting. As the value of $C$ increases, the classification accuracy gradually increases. At the same time, the phenomenon of overfitting is easy to occur. The parameter $\sigma$ is the width parameter of the kernel function. When $\sigma$ is small, overfitting is easy to occur; when $\sigma$ is large, underfitting is easy to occur. The penalty factor $C$ and the kernel function parameter $\sigma$ have a mutual influence relationship in SVM classification. In order to train a classifier with good performance, we first need to select the appropriate kernel function parameter $\sigma$, map the classification data samples to the high-dimensional feature space, and then find
the appropriate penalty factor $C_i$ in the feature space to balance the learning and generalization ability of SVM.

In the process of using the ABC algorithm for parameter optimization, it is necessary to select an appropriate objective function as the optimization target. Since the SVM is used in the classification task in this paper, Eq. (7) is selected as the fitness function, where $V_{acc}$ is classification accuracy.

$$f_{obj} = \frac{1}{1-V_{acc}}$$ (7)

The classification accuracy can be expressed as a binary function about $(C, \sigma)$, as shown in Eq. (8). Through the optimization of the ABC algorithm, the optimal parameter combination $(C, \sigma)$ is selected to train the SVM classifier, so that the trained model can obtain better scores on the test datasets.

$$V_{acc} = F(C, \sigma)$$ (8)

2.3. The proposed model GOABC-SVM

The process of SVM classification model based on ABC algorithm (GOABC-SVM) is as shown in Figure 1.

![Figure 1. Flow chart of GOABC-SVM](image-url)
3. Experiments and discussion

3.1. Experimental data and parameter setting
In order to test the performance of the optimized ABC algorithm, some classic test functions are selected for numerical experiments [16]. The selected test functions are shown in Table 1. At the same time, the population size is set to 40. The limit is 300. The maximum number of iterations maxCycle is 3000. And the range of the penalty factor C and kernel function parameter σ is [0.001,100].

| Name      | Function                                                                 | Search Space | Optimum |
|-----------|--------------------------------------------------------------------------|--------------|---------|
| Sphere    | \( f_1(x) = \sum_{i=1}^D x_i^2 \)                                       | [-100,100]D | 0       |
| Rosenbrock| \( f_2(x) = \sum_{i=1}^D \left[ 100(x_{i+1} - x_i^2)^2 + (x_i - 1)^3 \right] \) | [-50,50]D   | 0       |
| Griewank  | \( f_3(x) = \frac{1}{4000} \sum_{i=1}^D x_i^2 - \prod_{i=1}^D \cos \left( \frac{x_i}{\sqrt{i}} \right) + 1 \) | [-600,600]D | 0       |
| Rastrigin | \( f_4(x) = \sum_{i=1}^D [x_i^2 - 10 \cos(2\pi x_i) + 10] \)              | [-5.12,5.12] | 0       |
| SumSquare | \( f_5(x) = \sum_{i=1}^D x_i^2 \)                                       | [-10,10]D   | 0       |
| SumPower  | \( f_6(x) = \sum_{i=1}^D |x_i|^{(r+1)} \)                         | [-1,1]D     | 0       |
| Schwefel  | \( f_7(x) = \sum_{i=1}^D |x_i| + \prod_{i=1}^D |x_i| \)                           | [-10,10]D   | 0       |
| Alpine    | \( f_8(x) = \sum_{i=1}^D [x_i \sin(x_i) + 0.5] \)                         | [-10,10]D   | 0       |
| Quartic   | \( f_9(x) = \sum_{i=1}^D x_i^4 + \text{random}[0,1] \)                   | [-1.28,1.28]D | 0       |
| Ackley    | \( f_{10}(x) = -20 \cdot \exp \left( -0.2 \cdot \left[ \frac{1}{D} \sum_{i=1}^D x_i^2 \right] - \exp \left( \frac{1}{D} \sum_{i=1}^D \cos \left( \frac{2\pi x_i}{\sqrt{i}} \right) \right) + 20 + e \) \) | [-32,32]D  | 0       |

In order to test the classification performance of the GOABC-SVM, this paper uses the classic UCI dataset to perform experiments. The description of the UCI dataset is shown in Table 2.

| Dataset  | Instances | Number of classes(k) | Number of features(d) |
|----------|-----------|----------------------|-----------------------|
| Diabetes | 768       | 2                    | 8                     |
| Heartstatlog | 270     | 2                    | 13                    |
| Sonar    | 208       | 2                    | 60                    |
| Vehicle  | 846       | 3                    | 18                    |
| CMC      | 1473      | 3                    | 9                     |

3.2. Comparison with other ABCs
In this set of experiments, the GOABC algorithm, GABC algorithm and the ABC algorithm were tested on the test functions. The performance was compared by calculating the optimal value, worst value, average value, and standard deviation. Tables 3 shows the comparative experimental results when the dimension is 30.

| Function | ABC Optim Mean | ABC Max | ABC Std | GABC Optim Mean | GABC Max | GABC Std | GOABC Optim Mean | GOABC Max | GOABC Std |
|----------|----------------|---------|---------|-----------------|----------|---------|-----------------|----------|----------|
| f(x)     | 4.27E-16       | 5.19E-16| 7.13E-16| 5.92E-17        | 2.69E-16| 4.31E-16| 5.29E-16        | 7.46E-17| 3.62E-18|

[Note: The table values are provided in scientific notation.]

5
It can be seen from Table 3 that the GOABC algorithm proposed in this paper has achieved higher optimal values than the ABC and GABC algorithm on the test functions f1, f3, f5, f6, f7, f8, f9, and f10. On the test function f4, ABC, GABC, and GOABC can all reach the theoretical optimal values. In terms of stability, for the test functions f1, f3, f5, f6, f7, f9, f10, the standard deviations of GOABC are smaller than the ABC and GABC algorithm, indicating that GABC has strong stability. In summary, it can be seen from Tables 3 that GOABC has better optimization accuracy and stability than traditional ABC and GABC algorithm.

3.3. **Comparison of GOABC-SVM classification performance**

In order to test the classification performance of GOABC-SVM, this paper tests the proposed classification model on the UCI dataset, and compares the GOABC-SVM model with the SVM, SVM optimized by ABC and SVM optimized by GABC. Table 5 shows the experimental comparison results.

| Dataset     | SVM   | ABC-SVM | GABC-SVM | GOABC-SVM |
|-------------|-------|---------|----------|-----------|
| Diabetes    | 75.32 | 78.65   | 79.17    | 79.95     |
| Heartstatlog| 66.67 | 85.19   | 85.93    | 86.67     |
| Sonar       | 57.14 | 60.58   | 61.54    | 62.5      |
| Vehicle     | 75.29 | 82.23   | 83.18    | 83.65     |
| CMC         | 48.23 | 51.49   | 51.77    | 54.21     |

It can be seen from Table 4 that for the datasets Diabetes, Heartstatlog, Sonar, Vehicle, and CMC, the accuracy of classification using the model proposed in this paper is higher than using traditional SVM, SVM optimized by traditional ABC and SVM optimized by GABC. It indicates that the parameter optimization of SVM using GOABC is effective and can improve the accuracy of classification.

4. **Conclusion**

This paper combines the ideas of opposite learning and global optimal guidance, and proposes an improved ABC algorithm, which further balances the algorithm's exploration and exploitation capabilities, thereby improving the algorithm's optimization precision and convergence speed. And then, the SVM parameters are optimized using the improved ABC algorithm to make the SVM have better classification performance. Experimental results show that the proposed classification model has better performance than other models. In future work, we will consider applying the ABC algorithm to feature selection to further improve the classification performance of the SVM.
References

[1] VAPNIK V N. Statistical learning theory[M]. New York: Wiley Press, 1999.
[2] Cristianini N. An introduction to support vector machines and other kernel-based learning methods[M]. Beijing: Mechanical Industry Press, 2005.
[3] Ito K, Nakano R. Optimizing Support Vector regression hyperparameters based on cross-validation[C]// International Joint Conference on Neural Networks. IEEE, 2003.
[4] Chapelle O, Vapnik V, Bousquet O, et al. Choosing Multiple Parameters for Support Vector Machines[J]. Machine Learning, 2002, 46(1-3):131-159.
[5] Souravlias D, Parsopoulos K E. Particle swarm optimization with neighborhood-based budget allocation[J]. International Journal of Machine Learning and Cybernetics, 2016, 7(3):451-477.
[6] Merkle D, Middendorf M, Schmeck H. Ant colony optimization for resource-constrained project scheduling[J]. IEEE Transactions on Evolutionary Computation, 2002, 6(4):304-346.
[7] Gong D W, Sun J, Miao Z. A Set-Based Genetic Algorithm for Interval Many-Objective Optimization Problems[J]. IEEE Transactions on Evolutionary Computation, 2016, PP(99):1-1.
[8] Karaboga D, Basturk B. A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm[J]. Journal of Global Optimization, 2007, 39(3):459-471.
[9] Yang D, Liu Y, Li S, et al. Gear fault diagnosis based on support vector machine optimized by artificial bee colony algorithm [J]. Mechanism & Machine Theory, 2015, 90:219-229.
[10] Mustaffa Z, Yusof Y, Kamaruddin S S. Enhanced artificial bee colony for training least squares support vector machines in commodity price forecasting[J]. Journal of Computational Science, 2014, 5(2):196-205.
[11] Liu J X, Jia Z H. Telecommunication Traffic Prediction Based on Improved LSSVM[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2017:1850007.
[12] Xiao H, Sun Y, Jin Y, et al. Optimization about Fault Prediction and Diagnosis of Wind Turbine Based on Support Vector Machine[C]// International Conference on Measuring Technology & Mechatronics Automation. 2018.
[13] Gao L, Ye M, Wu C. Cancer Classification Based on Support Vector Machine Optimized by Particle Swarm Optimization and Artificial Bee Colony[J]. Molecules, 2017, 22(12):2086.
[14] Cao Y, Lu Y, Pan X, et al. An improved global best guided artificial bee colony algorithm for continuous optimization problems[J]. Cluster computing, 2018: 1-9.
[15] Tizhoosh H R, Ventresca M. Oppositional Concepts in Computational Intelligence 2008[M]. Springer Berlin Heidelberg, 2008.
[16] Bare bones artificial bee colony algorithm with parameter adaptation and fitness-based neighborhood[J]. Information Sciences, 2015, 316:180-200.