

Research on Data Classification Method of Optimized Support Vector Machine Based on Gray Wolf Algorithm

Jinqiang Ma, China People's Police University, China

Linchang Fan, China People's Police University, China*

Weijia Tian, Chengdu Public Security Bureau, China

Zhihong Miao, China People's Police University, China

ABSTRACT

The data classification method based on support vector machine (SVM) has been widely used in various studies as a non-linear, high precision, and good generalization ability machine learning method. Among them, the kernel function and its parameters have a great impact on the classification accuracy. In order to find the optimal parameters to improve the classification accuracy of SVM, this paper proposes a data multi-classification method based on gray wolf algorithm optimized SVM(GWO-SVM). In this paper, the iris data set is used to test the performance of GWO-SVM, and the classification result is compared with those based on genetic algorithm (GA), particle swarm optimization (PSO) and the original SVM model. The test results show that the GWO-SVM model has a higher recognition and classification accuracy than the other three models, and has the shortest running time, which has obvious advantages and can effectively improve the classification accuracy of SVM. This method has practical significance in image classification, text classification, and fault detection.

KEYWORDS

data classification, Gray Wolf algorithm, Kernel function, Support vector machine

1. INTRODUCTION

Classification is a relatively old problem that has been widely studied in areas such as machine learning, pattern recognition, data mining, and artificial intelligence. Classification problems can be defined as follows: given a dataset, a given dataset is called a training dataset (Jiawei & Kamber 2001; Zhongzhi, 2002). The training dataset consists of a set of database tuples (often referred to as training samples, instances, or objects), each training sample is a feature vector consisting of attribute values or eigenvalues, and each training sample also has a class label attribute. A specific sample

DOI: 10.4018/IJGHP.318408

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

form can be expressed as: $(V_1, V_2, \dots, V_n; c)$; where V_i represents the attribute value and c represents the class label. A given training dataset is used to build a classification function (often referred to as a classification model or classifier (Huajun & Yinkui, 2003), and the established classifier is used to predict the classes of tuples of data with unknown class numbers in the database.

As an age-old problem, taxonomy has been extensively studied in many fields. So far, the classical classification methods that have been studied mainly include: decision tree method, which is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome (classical decision tree algorithm mainly includes: ID3 algorithm, C4.5 algorithm and CART algorithm, etc.), neural network method (BP algorithm), genetic algorithm (GABIL system), Bayesian classification, K-nearest neighbor algorithm, case-based inference and support vector machine.

Support vector machines (SVM) are a new learning machine proposed by Vapnik et al. in the 1990s. It is based on the study of statistical theory and the principle of structural risk minimization, which fully guarantees its good generalization ability in theory, and has a solid theoretical foundation and good promotion ability. Compared with the traditional learning method, the support vector machine can overcome the problems of small samples, dimensionality disasters, local minimum points and overfitting, and by constructing the optimal classification surface, the classification error of unknown samples is minimized, and the generalization ability is shown. Because of its good performance, the support vector machine algorithm has attracted more and more researchers' attention as soon as it is proposed, and in recent years, the support vector machine has been widely used in pattern recognition, prediction field fault classification and other fields. However, as an emerging learning machine, the support vector machine also has some places to be improved, such as the selection of its parameters still does not have a unified standard, and the traditional parameter selection mostly relies on experience to take the trial method, which is not only time-consuming but also difficult to obtain satisfactory results, and studies have shown that the type and parameters of the kernel function are greatly influence on the classification accuracy of SVM, and the relationship between its parameters and the classification accuracy of SVM is an irregular multimodal function. If the parameter values are not suitable, the trained model will generalize. Poor ability, resulting in reduced identification accuracy of faulty samples (Ma, 2019; Morais & Rolim, 2006; Dash & Samantaray et al., 2007). Obviously, the traditional parameter selection method cannot adapt to the development of support vector machines. In engineering applications, the optimal parameters are determined empirically in some cases, but this's not reliable. At present, many scholars use computers to obtain the optimal parameters, which reduces the workload, and makes the selection of the optimal solution more dependable.

Grid search (GS) (Anhua, 2010) and cross validation (CV) (Dengda, 2011) are uncomplicated techniques. While moving ahead with the artificial intelligence technology, more and more intelligent algorithms have application in parameter optimization, such as genetic algorithm (GA) (Guiming, 2012) and particle swarm optimization (PSO) (Hang, Chaoyong, & Yonggang, 2013), Jinying, Hongxia, Xiuye, & Fujie, 2009) and so on. But then, the performance of these ways is not very desirable, because grid step size will affect the performance of GS, the smaller the grid step size, the higher the accuracy and the longer the time; CV needs that the amount of training samples must be plentiful; genetic algorithm contains a round of operations such as code, selection, crossover and mutation, and its speed and precision are limited; the representation of particle swarm algorithm is affected by multiple parameters of particles, and it is difficult to obtain optimal parameters in practical applications. In this regard, Mirjalili (2014) proposed a new heuristic algorithm, namely the GWO algorithm. GWO is a swarm intelligence optimization algorithm that simulates the social hierarchy and hunting behavior of the gray wolf family. In recent years, the gray wolf algorithm has also been continuously improved. In 2017, Heidari (2014) improved the GWO algorithm based on the Ryan flight method to solve the problem of prematurely falling into local extreme values when solving complex optimization problems; In 2019, in order to better reflect the actual search process of GWO,

Qiuping et al. (2017) proposed an improved method of convergence factor based on the variation of cosine law. Compare with grid search algorithm, cross-validation algorithm, genetic algorithm and particle swarm optimization algorithm, The outstanding advantage of the GWO algorithm is that each iteration takes into account both the global search (global search methods can be an effective tool for investigating the predictor space and identifying subsets of predictors that are optimally related to the response) and the local search, which not only greatly increases the probability of finding the optimal solution, but also largely avoids the premature phenomenon.

In this paper, In order to solve the problems of low resolution caused by improper selection modal component C and penalty factor σ and low accuracy of fault diagnosis caused by poor setting of SVM super parameters, the gray wolf algorithm is applied to the optimization of the kernel function and its parameters, A parameter optimization method of support vector machine on account of gray wolf algorithm is proposed, and the effectiveness of the method is verified by comparing the optimization results of genetic algorithm, particle swarm algorithm and original support vector machine. Compared with the same type of optimization algorithm, the iterative effect and optimization ability of GWO-SVM are greatly improved.

2. SVM ALGORITHM INTRODUCTION

SVM maps the historical series $x_1, x_2, \dots, x_n$ as input samples to the high-dimensional feature space H through the nonlinear mapping function $\varphi(x)$, and performs linear regression in H , and the regression of SVM in the high-dimensional feature space The function is:

$$f(x) = w \cdot \varphi(x) + b \quad (1)$$

In the expression, w represents the weight vector; b represents the bias vector.

According to the principle of institutional risk minimization, it can be transformed into the following optimization problem:

$$mJ = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i^* + \xi_i) \quad (2)$$

$$\text{s.t.} \begin{cases} y_i - w \cdot \varphi(x) - b \leq \varepsilon + \xi_i \\ w \cdot \varphi(x) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, n \end{cases} \quad (3)$$

In the formula: $\|w\|$ is the term related to the complexity of the function f , ε is the insensitive loss coefficient; ξ_i, ξ_i^* represent the relaxation factor; C represents the penalty factor. With the introduction of Lagrange multipliers, the optimization problem becomes a convex quadratic optimization problem:

$$\begin{aligned} L(w, b, \xi, \xi^*, \alpha, \alpha^*, \gamma, \gamma^*) = & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) - \sum_{i=1}^n \alpha_i (\xi_i + \varepsilon - y_i + f(x_i)) - \\ & \sum_{i=1}^n \alpha_i^* (\xi_i^* + \varepsilon - y_i + f(x_i)) - \sum_{i=1}^n (\xi_i \gamma_i - \xi_i^* \gamma_i^*) \end{aligned} \quad (4)$$

In the formula, α_i and α_i^* represent the Lagrange multipliers; γ represents the loss factor. In order to speed up the solution speed, the equation (4) is converted into a dual situation, there are:

$$w(\alpha, \alpha^*) = -\frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*) \times (\alpha_j - \alpha_j^*) (\varphi(x_i), \varphi(x_j)) + \sum_{i=1}^n (\alpha_i - \alpha_i^*) y_i - \sum_{i=1}^n (\alpha_i - \alpha_i^*) \quad (5)$$

$$\text{s.t.} \begin{cases} w = \sum_{i,j=1}^n (\alpha_i - \alpha_i^*) x_i \\ \sum_{i=1}^n (\alpha_i - \alpha_i^*) \alpha_i = 0 \\ 0 \leq \alpha_i, \alpha_i^* \leq C \end{cases}$$

For the linear regression problem, the SVM regression function is:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) (\varphi(x_i), \varphi(x)) + b \quad (6)$$

The establishment of a classification prediction model based on SVM is to find the optimal support vector parameters C and σ . So that the expression (6) is established, where $f(x)$ is the predicted value. The research shows that when the prior knowledge of the process is lacking, the radial basis kernel function has fewer parameters and better performance than other kernel functions. Therefore, this paper chooses the radial basis kernel function as the kernel function of SVM, which is defined as follows:

$$K(x_i, x_j) = \exp \left(-\frac{x_i - x_j^2}{2\sigma^2} \right) \quad (7)$$

where σ is the width parameter of the radial basis kernel function.

It can be seen from the SVM modeling process that the SVM learning performance is closely related to the penalty coefficient C and the selection of the kernel function parameter σ (Xiaodong, Guangliang, & Zixiang, 2012). Different kernel functions have little effect on SVM performance, but the parameters of the kernel function and error penalty factor C are the key factors affecting SVM performance. Therefore, choosing the appropriate kernel function parameters and error penalty factor C is critical to the performance of the learning machine. This paper optimizes the kernel parameters and error penalty factor C of the most widely used radial basis function SVM.

The performance of the support vector machine depends on multiple parameters, the kernel function parameters σ mainly affect the complexity of the distribution of sample data in the high-dimensional feature space, and the change of the kernel parameters actually implies the change of the VC dimension of the feature space, which affects the confidence range and ultimately the structural risk range. The compromise between the maximization of the control interval of the penalty factor C and the classification error, the larger the C , the greater the penalty for the misdivided sample; A small value of C indicates a small penalty, a small complexity of the learning machine and a large empirical risk value, the former being called “over-learning”, while the latter being “under-learning”. In addition to optimizing C in the same feature space to obtain the optimal SVM corresponding to the corresponding space, the kernel function parameters are also optimized to obtain the global optimal SVM.

3. GWO OPTIMIZATION ALGORITHM

The GWO optimization algorithm is a new meta-heuristic optimization algorithm proposed by Mirjalili et al. in 2014 (Mirjalili et al., 2014). It is a new type of swarm intelligence optimization algorithm. Relevant research shows that the algorithm has excellent performance in finding the optimal solution, and has the characteristics of simplicity and efficiency.

Gray wolf belongs to the canine family, which is at the top of the natural food chain and is regarded as the top predator. Most wolves like to live in groups. The average number of gray wolves in each group is 5 to 12. In normal life, especially in the hunting process, they follow a very strict social hierarchy and task division system. In the GWO algorithm, the highest rank is the head wolf, which has two heads, one male wolf and one female wolf, which is marked as α , which is responsible for decision-making in the process of hunting (optimization) and leading the wolf pack. The remaining wolf packs are labeled β , δ , and ω by social rank. The behavior of the next level needs to obey the leadership of the upper level, and carry out the corresponding group hunting action, as shown in Figure 1.

The GWO algorithm imitates the hunting behavior process of wolves, which is mainly divided into three steps, namely encircling, hunting, and attacking. The modeling process of each step is as follows.

3.1 Surround

During the hunting process, the wolves first surround the target. The mathematical model of this process is:

$$D = |C \cdot X_p(t) - X(t)|$$

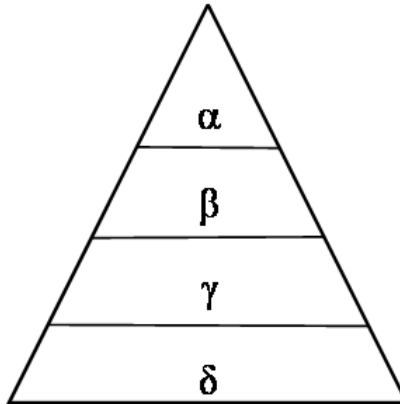
$$X(t+1) = X_p(t) - A \cdot D \quad (8)$$

Where t is the current iteration, A and C are coefficient vectors, and X_p is the global optimal solution vector (prey location), and X is the potential solution vector (wolf location). The values of A and C are calculated by the following formula:

$$A = 2a \cdot r_1 - a$$

$$C = 2 \cdot r_2 \quad (9)$$

Figure 1.
GWO algorithm for wolves rank classification



The value of a decreases linearly from 2 to 0 as the number of iterations increases, and r_1, r_2 are defined as random vectors with values ranging from [0,1].

3.2 Hunt

After the prey is surrounded, the wolves hunt in real time. In order to better search for the location of the prey, the process is usually guided by α, β and δ . As the location (optimal solution) of the prey in the algorithm is unknown, and α wolf, β wolves and δ wolves represent the three positions closest to the prey, the gray wolf algorithm considers that the location of the prey is that of α wolf, β wolves and δ wolves. The ω wolves constantly update their own positions according to the location of the three wolves, and finally complete the capture. Other search units (ω) should be based on the current optimal positions of the search units (α, β and δ) update their respective positions. Therefore, the update expression of the wolf pack position in this stage is:

$$\begin{aligned} D_\alpha &= |C_1 \cdot X_\alpha|, D_\beta = |C_2 \cdot X_\beta|, D_\delta = |C_3 \cdot X_\delta| \\ X_1 &= X_\alpha - A_1(D_\alpha), X_2 = X_\beta - A_2(D_\beta) \\ X_3 &= X_\delta - A_3(D_\delta) \\ X(t+1) &= \frac{X_1 + X_2 + X_3}{3} \end{aligned} \quad (10)$$

$X(t+1)$ is the updated potential optimal solution vector. The location update process is shown in Figure 2.

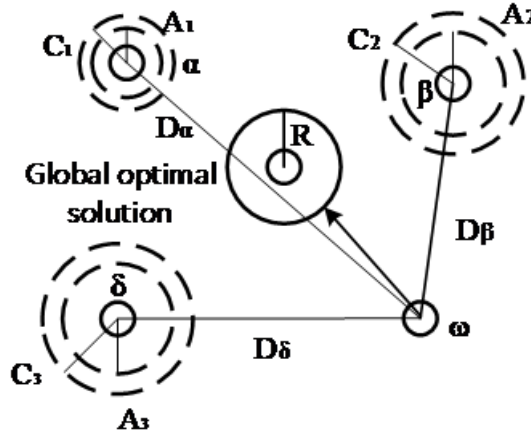
3.3 Attack

The model construction of this process is mainly realized by the decrement of the value of a in formula (3). The value of a decreases linearly from 2 to 0, and the value of A will correspondingly obtain any value within the interval $[-2a, 2a]$. When $|A| \leq 1$, the wolf pack will be located between its current position and the prey position and can focus on attacking the prey (X^*, Y^*); When $|A| > 1$, the position between the wolves and the prey deviates, and a global search is performed to find a more suitable prey.

4. GWO-SVM MODEL CONSTRUCTION

Since the penalty factor c and the kernel parameter σ in the SVM have a great influence on the classification accuracy of the sample, this paper uses GWO to optimize the parameters in the SVM, and (c, σ) constitutes the position vector of the gray wolf. The gray wolf group mainly searches according to the positions of α, β, δ , and separates from each other to find prey. In GWO, the value of parameter A is set to a random value greater than 1 or less than 1, which can make individual gray wolves deviate from the target prey. This behavioral mechanism allows GWO to search on a global scale and also enhances the exploration performance of the algorithm. Another component of the GWO exploration is c . c is a random value in the interval $[0, 2]$. This part provides random weights for the prey to randomly strengthen ($c > 1$) or weaken ($c < 1$) the weights to determine distance. The parameter c helps GWO to show more random behavior in the optimization process, which is beneficial to improve the global exploration ability of the algorithm. Especially in the later iteration, c can effectively make the algorithm jump out of the local optimum, and then find the global optimum solution, so as to obtain the best c and σ that can make the sample fault

Figure 2.
GWO algorithm optimal solution vector position update process diagram



diagnosis with the highest accuracy, and optimize the fault diagnosis result. The implementation process of GWO-SVM is shown in Figure 3 (Cai Shaohong, Jiao Jianjun, & Long Wen, 2019; Guo Huiying, & Wang Yi, 2019; Ali Asghar Heidari, & Parham Pahlavani, 2017; Wang Qiuping, Wang Mengna, & Wang Xiaofeng, 2019).

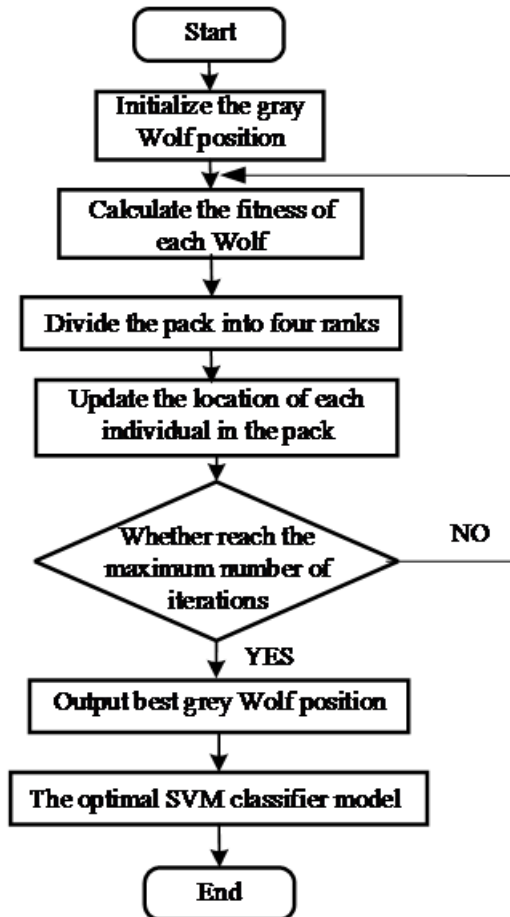
- Step 1: Input the sample data and set the training set and test set of SVM.
- Step 2: Initialize the value range of c and σ in SVM.
- Step 3: Randomly generate gray wolf groups. The individual position vector of each gray wolf group is composed of penalty factor c and kernel function σ composition and set relevant parameters of GWO.
- Step 4: According to the initial parameters c and σ , SVM trains the training set samples, and the individual fitness function is expressed by the classification error rate of the SVM algorithm to calculate the fitness of each gray wolf.
- Step 5: According to the fitness value, the gray wolf group is divided into 4 different levels: $\alpha, \beta, \delta, \omega$.
- Step 6: Update the position of each individual in the wolf pack according to equations 5, 6, 7.
- Step 7: Calculate the fitness value of the current gray wolf individual at the new position, and re-select α, β, δ from the current wolf group.
- Step 8: If the number of iterations exceeds the maximum allowable number of iterations, the training ends, and the output global optimal position is the optimal value of c and σ in the SVM; otherwise, skip to step 5 to continue parameter optimization.
- Step 9: Use the optimal parameters c and σ to establish a model, test the test set samples, and analyze and verify the test results.

5. THE EXAMPLE ANALYSIS OF IRIS CLASSIFICATION BASED ON GWO-SVM ALGORITHM

5.1 Instance Data

For the sake of proving the advantages of the gray wolf algorithm optimization support vector machine in data classification, this paper selects the data set of iris (KDD997), which is a data set of multivariate analysis, and often used as an example in both statistical learning and machine learning. Iris Dataset is considered as the Hello World for data science. It contains five columns namely – Petal Length,

Figure 3.
The implementation process of GWO-SVM



Petal Width, Sepal Length, Sepal Width, and Species Type. Iris is a flowering plant; the researchers have measured various features of the different iris flowers and recorded them digitally. There are 300 groups of data in the data set, which are divided into 3 groups according to their species and marked. Among them, there are 180 training samples and 120 test samples.

5.2 Data Preprocessing

The GWO-SVM algorithm is used to identify and classify iris varieties. In this paper, four characteristic variables are selected as input variables, namely: sepal length, sepal width, petal length, and petal width. In the process of recognition and classification, set the target output as 1, 2 and 3, where 1 represents variety 1, 2 represents variety 2 and 3 represents variety 3.

Since the SVM model is very sensitive to the data in the range of $[0, 1]$, it is necessary to normalize the feature vector to improve the training efficiency before inputting the training samples (Figure 4 shows the original data of features, and Figure 5 shows the data after feature normalization), namely:

Table 1.
Training set and test set of KDD997

iris varieties	number of training samples	number of test samples
Variety 1	60	40
Variety 2	60	40
Variety 3	60	40

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (11)$$

Where: x is the original sample data.

Input the normalized training samples into the vector machine model and use GWO to optimize the parameters c and σ in the SVM, which improves the traditional way of setting the value randomly. The parameters c and σ constitute the position vector of the individual gray wolf pack. The error rates representation of the individual fitness function SVM for the classification of iris samples. In this paper, the number of wolves is set to 10, the maximum number of iterations is set to 10, and the search interval of parameters c and σ is [0.01, 100]. The optimal parameters c and σ obtained by GWO optimization were used to construct the SVM model, and finally the constructed GWO-SVM model was tested on the test samples of iris.

Figure 4.
Data for the four characteristics

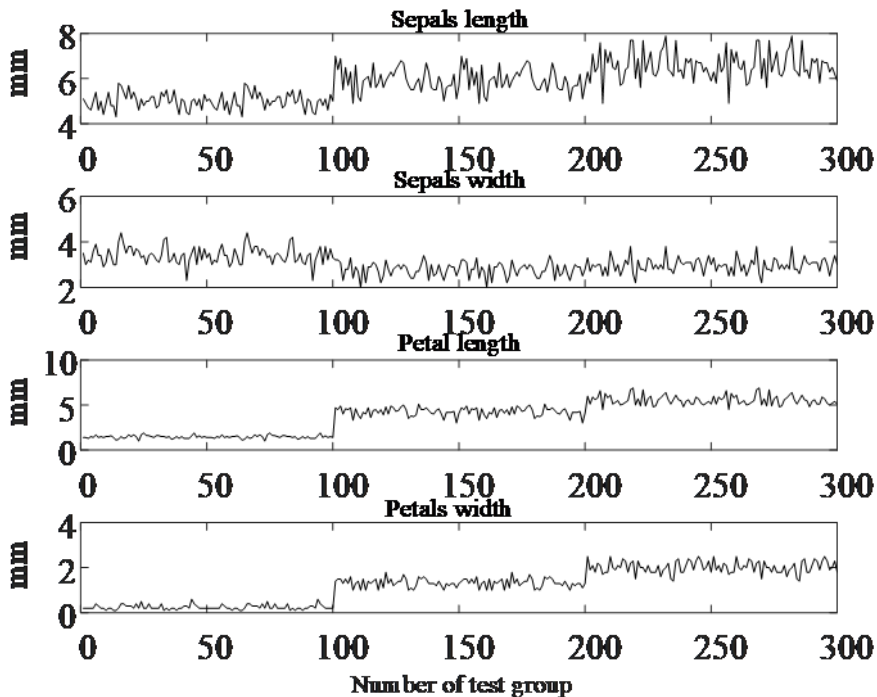
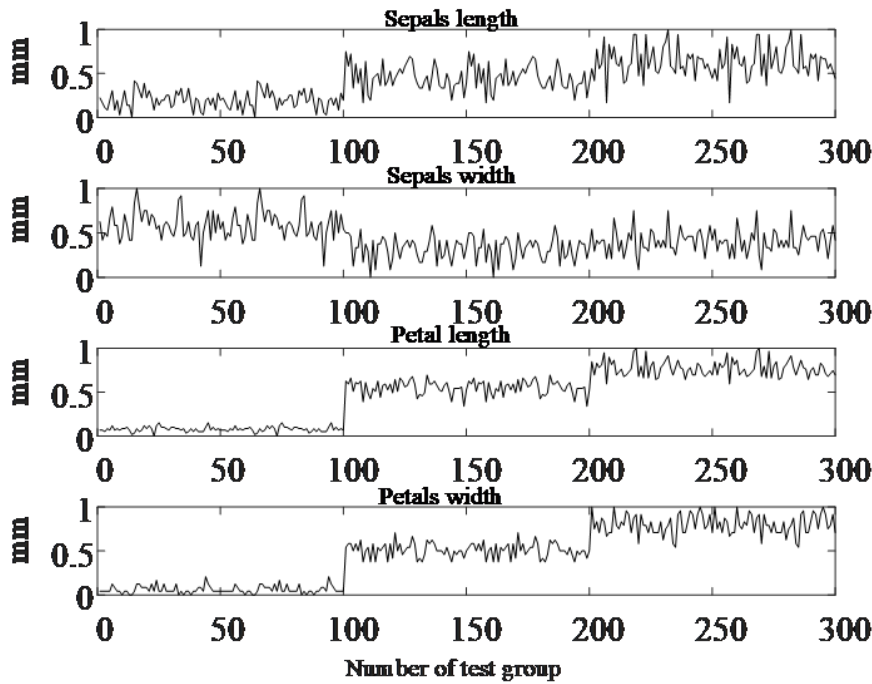


Figure 5.
Data normalization processing of four characteristics



5.3 Example Results and Analysis

For the sake of proving the effectiveness and advantages of GWO-SVM, genetic algorithm (GA) and standard particle swarm optimization (PSO) are used to optimize the kernel function and parameters of the established SVM, and used the optimized SVM model to identify and classify the varieties of iris, and compare it with the test results of the support vector machine based on GWO optimization. Among them, GA selects crossover probability and mutation probability according to the fitness values of the object function, therefore reduces the convergence time and improves the precision of GA, ensuring the accuracy of parameter selection. And PSO uses its fast global optimization feature to search the parameters of SVM, which can reduce the blindness of trial and improve the accuracy of model prediction. The SVM classification and recognition results based on the three optimization algorithms are shown in Figures 6-9.

It can be seen from Figure 6 to Figure 9 that GWO-SVM has the best recognition and classification effect, and only 2 test samples are incorrectly. PSO-SVM algorithm model has 4 test sample recognition errors; GA-SVM algorithm model has 8 test sample recognition errors; The SVM model without algorithm optimization has 10 test sample recognition errors and the recognition effect is the worst. Therefore, choosing the appropriate kernel function parameters and error penalty factor C is critical to the performance of the learning machine. The optimization results of the penalty factor and kernel parameters in the SVM model by the three optimization algorithms are shown in Table 2. Table 3 shows the accuracy of GWO-SVM, PSO-SVM, GA-SVM and SVM algorithm models in identifying and classifying test sets and the running time of each program. As can be seen from Table 3, the accuracy rate of the original SVM model without optimization for iris variety identification and classification is 91.67%, and the running time is 1.78s, which is the lowest accuracy rate and longest running time among the four models. The accuracy rate of

Figure 6.
Predicted classification graph of GWO-SVM

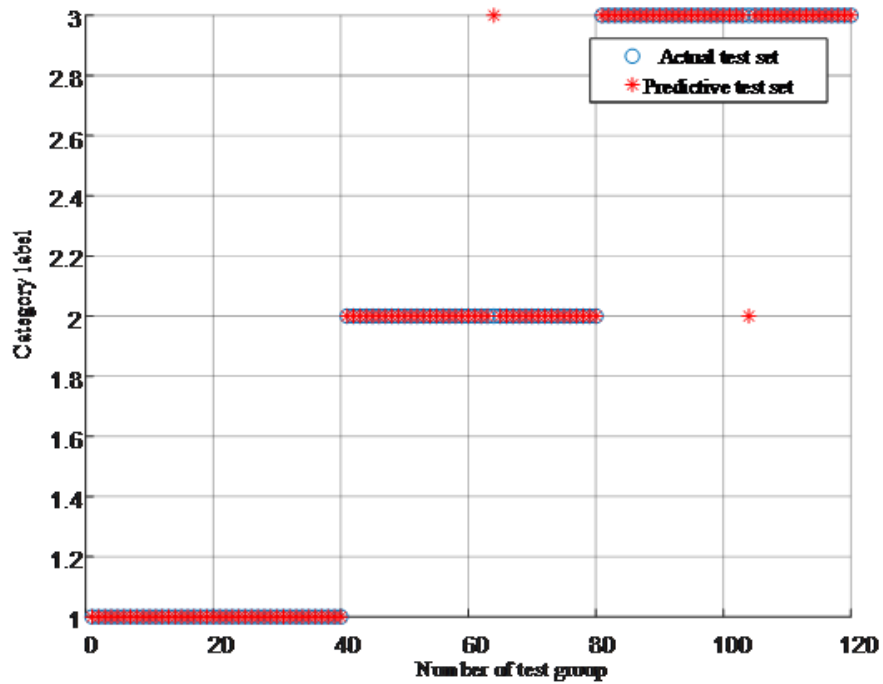


Figure 7.
Predicted classification graph of PSO-SVM

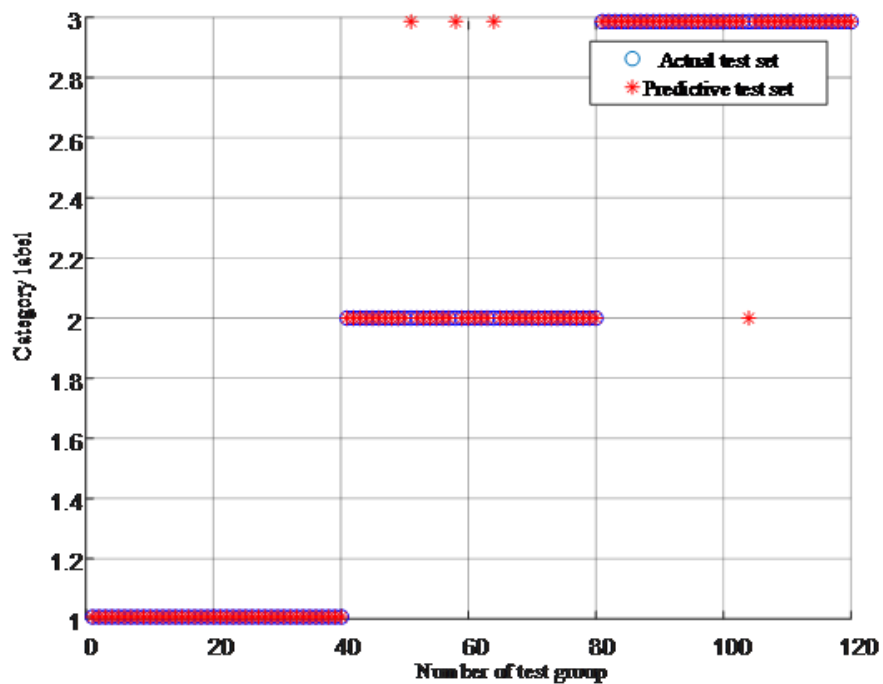


Figure 8.
Predicted classification graph of GA-SVM

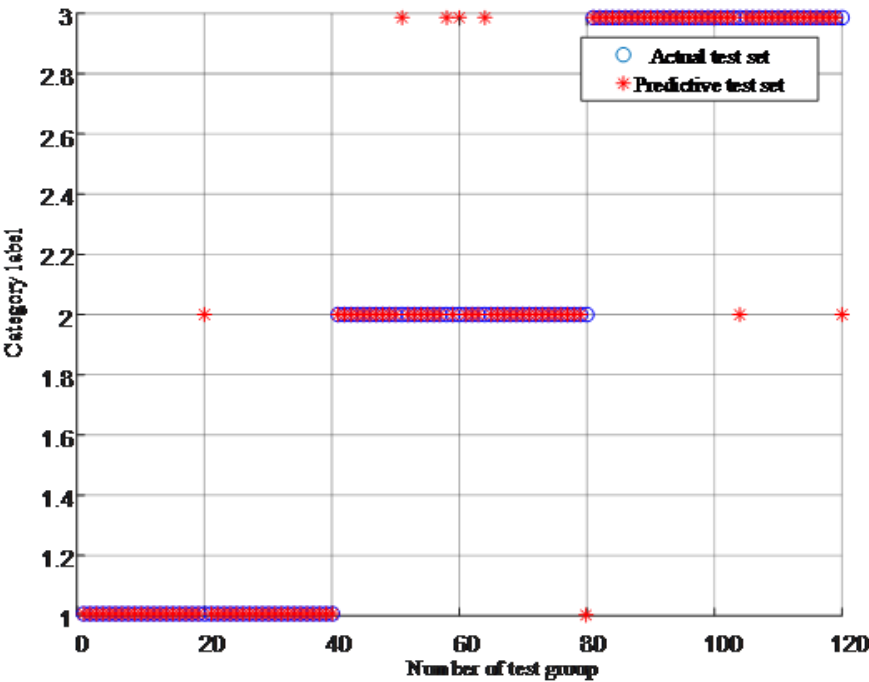


Figure 9.
Predicted classification graph of original SVM

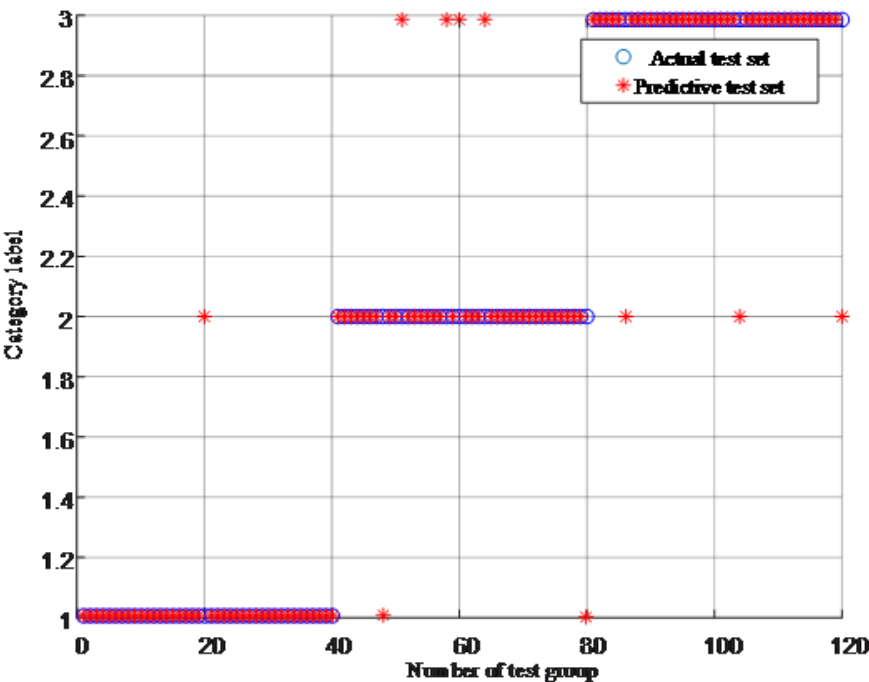


Table 2.
Optimized SVM model parameters

Algorithm model	c	σ
GWO-SVM	13.9545	0.7414
PSO-SVM	19.6289	0.0263
GA-SVM	2.1561	0.3887
SVM	13.3436	0.0100

Table 3.
The recognition accuracy and running time of the four models

Algorithm model	Accuracy /%	Running time /s
GWO-SVM	98.33	0.63
PSO-SVM	96.67	0.96
GA-SVM	93.34	1.32
SVM	91.67	1.78

the original GA-SVM model for iris variety identification and classification is 93.34%, and the running time is 1.32s. This is because GA algorithm cannot make timely use of the feedback information of the network, the search speed of the algorithm is relatively slow, and more training time is required to obtain a more accurate solution. The accuracy rate of the original PSO-SVM model for iris variety identification and classification is 96.67%, and the running time is 0.96s. But PSO algorithm is easy to fall into local optimal solution, and it is difficult to obtain accurate optimal solution. The accuracy of GWO-SVM model for iris variety identification and classification was 98.33%, which was the highest among the four models. In addition, the running time of GWO-SVM model is 0.63s, which is also very advantageous in running time. Therefore, GWO-SVM model has obvious advantages and better application prospects.

6. CONCLUSION

In order to improve the classification accuracy of SVM, this paper optimized the penalty factor and kernel parameters of support vector machine on account of the gray Wolf algorithm, and obtained the GWO-SVM multi-classification model. The performance of GWO-SVM was tested by using iris data set, and compared with the classification results of GA-SVM, PSO-SVM and original SVM model. Finally, simulation experiments verify that the GWO-SVM classification model has better stability and superiority. The test results show that: The SVM model in terms of small sample data classification is of great advantage, but its accuracy and running time are under the influence of penalty factor and parameters of the SVM. After algorithm optimization, the recognition and classification accuracy of GWO-SVM model is improved to 98.3%, and the running time of program is the shortest. Compared with other models, GWO-SVM model can significantly improve the classification performance of support vector machine, which has obvious advantages and better application prospects.

REFERENCES

- Anhua, C. Lingli, J. Yilun, L. & Xuejun, L. (2010) Gear fault diagnosis based on SVM and multi-sensor information fusion. *Cent. South Univ. Sci. Technol.*, 41 (06), 2184–2188.
- Cai, S., Jiao, J., & Wen, L. (2019). An Improved Grey Wolf Optimization Algorithm. *Chinese Journal of Electronics*, 47(01), 169–175.
- Dash, P., Samantaray, K., & Panda, G. (2007). Fault Classification and Section Identification of an Advanced Series-Compensated Transmission Line Using Support Vector Machine. *IEEE Transactions on Power Delivery*, 22(1), 67–73. doi:10.1109/TPWRD.2006.876695
- Dengda, G. Lingli, J. Xuejun, L., & Dalian, Y. (2011) Fault diagnosis method based on multi-sensors installed on the base and KPCA. *Chin. J. Sci. Instrum.*, 32 (07), 1551–15 57.
- Guiming, C. Qingjie, L. Xiaofang, L. & Qing, Y. (2012) Genetic algorithm based SVM parameter composition optimization. *Comput. Appl. Softw.* 29 (04) (94-96 + 100).
- Han, J., & Kamber, M. (2001). *DATA MINING—Concepts and Techniques*. Higher Education Press.
- Hang, C., Chaoyong, H., & Yonggang, Z. (2013). Constructed of SVM decision tree based on particle swarm optimization algorithm for gear box fault diagnosis. *J. Vib. Meas.Diagn.*, 33(01), 153–156.
- Heidari, A. A., & Pahlavani, P. (2017). An Efficient Modified Grey Wolf Optimizer with Lévy Flight for Optimization Tasks. *Applied Soft Computing*, 60, 115–134. doi:10.1016/j.asoc.2017.06.044
- Huiying, G. & Yi, W. (2019). Transformer fault diagnosis based on DGA support vector machine. *Modern Electronic Technology*, 42(19), 154-158+163.
- Jinying, H., Hongxia, P., Xiuye, W., & Fujie, W. (2009). Fault condition recognition with kernel principal component analysis based on particle swarm optimization. *Machining Science and Technology*, 28(12), 1546–1550.
- Li, X., Wang, G., & Zhao, Z. (2012). Improved undersampling method for imbalanced data classification based on support vector machine. [Natural Science Edition]. *Journal of Sun Yat-Sen University*, 51(06), 10–16.
- Ma, X. (2019). Theory and Application of Support Vector Machines. *Science and Technology Innovation*, 2019(02), 13–14.
- MA Y. (2014). *Support vector machines applications*. Springer.
- Mirjalili, S., Mirjalili, S. M., & Lewis, A. LEWIS A. (2014). Grey wolf optimizer. *Advances in Engineering Software*, 69, 46–61. doi:10.1016/j.advengsoft.2013.12.007
- Morais, D. R., & Rolim, J. G. (2006). A hybrid tool for detection of incipient faults in transformers based on the dissolved gas analysis of insulating oil. *IEEE Transactions on Power Delivery*, 21(2), 673–680. doi:10.1109/TPWRD.2005.864044
- Shi, Z. (2002). *Knowledge Discovery*. Tsinghua University Press.
- Shuai, D. & Xu, D. (2017) Research on short-term prediction of cloud computing resource load based on improved GWO optimized SVM. *Computer engineering and application*, 53 (7), 68-73.
- Wang, Q., Wang, M., & Wang, X. (2019). Grey Wolf Optimization Algorithm with Improved Convergence Factor and Proportional Weight. *Computer Engineering and Applications*, 55(21), 60–65.
- Zeng, H., & Zhang, Y. (2003). *Machine Learning*. China Machine Press.