Computers, Materials & Continua CMC, vol.62, no.1, pp.337-354, 2020

An Improved Whale Optimization Algorithm for Feature Selection

Wenyan Guo[1, *], Ting Liu[1], Fang Dai[1] and Peng Xu[1]

Abstract: Whale optimization algorithm (WOA) is a new population-based meta-heuristic algorithm. WOA uses shrinking encircling mechanism, spiral rise, and random learning strategies to update whale's positions. WOA has merit in terms of simple calculation and high computational accuracy, but its convergence speed is slow and it is easy to fall into the local optimal solution. In order to overcome the shortcomings, this paper integrates adaptive neighborhood and hybrid mutation strategies into whale optimization algorithms, designs the average distance from itself to other whales as an adaptive neighborhood radius, and chooses to learn from the optimal solution in the neighborhood instead of random learning strategies. The hybrid mutation strategy is used to enhance the ability of algorithm to jump out of the local optimal solution. A new whale optimization algorithm (HMNWOA) is proposed. The proposed algorithm inherits the global search capability of the original algorithm, enhances the exploitation ability, improves the quality of the population, and thus improves the convergence speed of the algorithm. A feature selection algorithm based on binary HMNWOA is proposed. Twelve standard datasets from UCI repository test the validity of the proposed algorithm for feature selection. The experimental results show that HMNWOA is very competitive compared to the other six popular feature selection methods in improving the classification accuracy and reducing the number of features, and ensures that HMNWOA has strong search ability in the search feature space.

Keywords: Whale optimization algorithm, Filter and Wrapper model, K-nearest neighbor method, Adaptive neighborhood, hybrid mutation.

1. Introduction

In today' s era of big data, various data can be obtained from a research object, and data can be collected from multiple research objects. The characteristics of large amounts of data and many features make research objects contain many redundant features and even noise features. It greatly increases the time and space complexity of learning and training algorithm for classification problems, and also reduces the accuracy of classification. Therefore, the feature selection algorithm [Chandrasheka and Sahin (2014)] is used to remove the irrelevant and redundant attributes of the research object, to reduce the attributes of the feature attributes, to find a feature subspace with good separability, to achieve the reduction of the dimension of high-dimensional data, and to reduce the time

[1] School of Science, Xi'an University of Technology, Xi'an, 710054, China.

[*] Corresponding Author: Wenyan Guo. Email: wyguo@xaut.edu.cn.

and space complexity of machine learning. Simplifying the classification model structure and improving the classification accuracy are the hot topics among many scholars.

In essence, the process of finding the best subset of features is a combinatorial optimization problem. The evaluation criteria and search strategy of feature subset are the keys of feature selection algorithm. The feature selection method based on the evaluation criteria is divided into Filter and Wrapper based on whether it is independent of the follow-up learning algorithm [Crone and Kourentzes (2010)]. The Filter model [Hancer, Xue and Zhang (2017)] has nothing to do with the follow-up learning algorithm, and the statistical performance of the metric features is depended on distance between features, information gain, and dependence ect. It has the advantages of simple calculation and fast speed, but low classification accuracy. The Wrapper model [Mafarja and Mirjalili (2018)] uses classification accuracy as an evaluation criterion for the merits of a subset of features, and integrates it with classification algorithms (such as K-nearest neighbor method [Wang, An, Chen et al. (2015)], support vector machine algorithm [Paul, Magdon-Ismail and Drineas (2016)], neural network algorithm [Vasilic and Kezunovic (2005)], etc.). The selected subset of features is small in scale but takes long time. The Filter and Wrapper methods have their own advantages and disadvantages. Combining the two, the Filter method is used to remove redundancy and noise characteristics, and then the Wrapper method is used to further optimize key features to achieve the mixed feature selection process. The feature selection method based on search strategy can be divided into three categories according to the formation process of characteristics: global optimal, random search and heuristic search. The global optimal search strategy uses the branch bound method to find the optimal feature subspace for data sets with fewer feature numbers; random search strategy obtains high performance feature subset by probability inference or random sampling mathematical model; the heuristic search strategy achieves a satisfactory set of features with fast and stable efficiency. Feature selection is a typical optimization problem. The methods of solving the optimization problem are divided into two types: gradient based methods [Vu-Bac, Duong, Lahmer et al. (2017)] and gradient-free algorithms. When the objective function is differentiable, the gradient-based optimization method starts from a single point and seeks optimization along the gradient direction. It has the advantage of fast convergence speed, but it is easy to fall into local optimization. The gradient-free optimization algorithm is favored by the researchers because of its weak performance requirements and diversity of search directions.

In the past two decades, intelligent algorithms based on heuristic and random search based on biological intelligence or natural phenomena (They are typical gradient-free algorithms.) have been widely used in machine learning, data mining, engineering design and other optimization fields. Intelligent algorithms have taken into account the advantages of heuristic and random search strategies. It effectively balances the global and local optimal search processes, combines them with Wrapper evaluation criteria to design feature selection methods, and improves the classification accuracy while increasing the speed of feature selection. Genetic algorithm [Spolaôr, Lorena and Lee (2018)] (GA), differential evolution algorithm [Xue, Fu and Zhang (2014)] (DE), particle swarm optimization [Fong, Wong and Vasilakos (2016)] (PSO), artificial bee colony [Muthuramalingam, Kumar and Anusheela (2011)] (ABC), ant lion optimization algorithm [Zawbaa, Emary and Parv (2015)] (ALO), grey wolf optimization [Emary,

Zawbaa and Grosan (2014)] (GWO) and hybrid algorithms [Zorarpaci and ÖzeA (2016)] are used in feature selection to find the optimal feature subset.

Whale optimization algorithm is a new kind of heuristic random intelligent algorithm based on population. By simulating the predation behavior of humpbacks, the local search ability of the algorithm is enhanced by the shrinking encircling mechanism and spiral ascending mechanism, while the global search ability of the algorithm is enhanced by the random learning strategy. It has the advantages of few control parameters, simple calculation, and strong ability to search optimal solution. It has successfully solved the optimization problems such as wind speed prediction [Wang, Du and Niu (2017)], classification of biotransformation liver drug toxicity [Tharwat, Moemen and Hassanien (2017)], feature selection [Sharawi, Zawbaa and Emary (2017)], and image segmentation [Aziz, Ewees and Hassanien (2017)]. However, whale optimization algorithm has the disadvantage of slow convergence speed. Many scholars have improved it. IWOA [Xiong, Zhang, Shi et al. (2018)] improves the individual's updating method through differential evolution strategy and improves the global optimization ability of the algorithm; LWOA [Ling, Zhou and Luo (2017)] integrates Levy flight to enhance the diversity of the population and prevent premature convergence; CPWOA [Huang, Li, Song et al. (2018)] integrates cosine control factors so that the algorithm slows down the convergence speed in the early iteration to carry out a full global exploration. The polynomial variation is used to enhance the ability of the algorithm to jump out of the local optimal solution. In this paper, in order to better balance the local development and global exploration ability of WOA algorithm, a new adaptive neighborhood radius based on the mean distance from itself to other whales is designed to calculate the adaptive neighborhood. An improved whale optimization algorithm (HMNWOA) based on adaptive neighborhood and hybrid mutation strategy is proposed by using the optimal solution learning strategy to its own neighborhood instead of random learning strategy to enhance the algorithm exploration, while improving the algorithm convergence speed and combining Gaussian and Cauchy mutation operation to enhance the algorithm's ability to jump out of the local optimal solution and improve the calculation accuracy of the algorithm. A feature selection algorithm based on binary HMNWOA is proposed. Twelve standard datasets from the UCI data repository test the effectiveness of HMNWOA for feature selection. The experimental results show that HMNWOA is superior to the six feature selection algorithms in improving the classification accuracy and reducing the number of features, and ensures that HMNWOA has strong search ability in the search feature space.

This article is arranged as follows. The second section briefly introduces the whale optimization algorithm. The third section gives the detailed steps of HMNWOA. The fourth section describes the feature selection algorithm based on binary HMNWOA. The fifth section is experimental results and analysis. Finally, the sixth section summarizes the full text.

2 Whale optimization algorithm

For optimization problem

$$\begin{aligned} &\min f(\boldsymbol{x}) \\ &\text{s.t.} \boldsymbol{l} \le \boldsymbol{x} \le \boldsymbol{u} \end{aligned} \tag{1}$$

where $f(\boldsymbol{x})$ is single-objective optimization function; $\boldsymbol{x} \in \mathbf{R}^n$ is n-dimensional decision variable; $\boldsymbol{u}, \boldsymbol{l} \in \mathbf{R}^n$ is the upper and lower bounds of $\boldsymbol{x}$.

The whale optimization algorithm for solving the above optimization problems includes three renewal strategies: shrinking encircling mechanism, spiral update mechanism, and random search strategy.

The mathematical model of the three strategies:

$$x_{i,j}(t+1) = x_{*,j}(t) - A \cdot d_{i,j} \tag{2}$$

$$x_{i,j}(t+1) = x_{*,j}(t) + e^{bl} \cdot cos(2\pi l) \cdot d'_{i,j} \tag{3}$$

$$x_{i,j}(t+1) = x_{rand,j}(t) - A \cdot d_{i,j} \tag{4}$$

where t is the current iteration; $\boldsymbol{x}_*$ is the global best solution; b is a constant that determines the shape of solenoid; l is a random number in [-1,1], $\boldsymbol{x}_{rand}$ represents randomly selecting a whale from the current population.

When $p < 0.5$ and $A \le 1$, whales update their position via Eq. (2), else if $A > 1$, whales update their position via Eq. (4), $d_{i,j} = \left| 2 \cdot rand \cdot x_{*,j}(t) - x_{i,j}(t) \right|$, $rand$ is a random number in [0,1]. Else, whales update their position via Eq. (3), $d'_{i,j} = \left| x_{*,j}(t) - x_{i,j}(t) \right|$.

3 The improved whale optimization algorithm

The whale optimization algorithm has high computation accuracy, which has compromised the exploration and exploitation ability of the algorithm by controlling parameter A. However, it is strongly dependent on the global optimal solution of the population in the exploitation stage. By integrating hybrid mutation strategy [Kumar (1998)] into the whale optimization algorithm, a high-quality population can be obtained, thus accelerating the convergence of the algorithm and improving the calculation precision of the algorithm. In the exploration stage, the whale optimization algorithm adopts stochastic strategy to update positions, which has strong blindness. The quality of the solution at the exploration stage can be improved by introducing the reasonable neighborhood surrounding the whale and making it learn from the optimal individuals in the neighborhood. Therefore, for the whales in the exploration stage, the search radius of the adaptive neighborhood is given to them to learn from the optimal individuals in the adaptive neighborhood, so as to improve the quality of the solution quickly. In this paper, an improved whale optimization algorithm based on the adaptive neighborhood strategy and the hybrid mutation strategy (HMNWOA) is put forward.

3.1 The adaptive neighborhood strategy

In the exploration phase, whale individual uses a whale that are randomly selected in the current population to update position, and it is easy to miss the optimal solution. The results show that the neighborhood strategy has an improved influence on the performance of the algorithm. At present, the commonly used neighborhood topologies include ring, star, and von neumann topologies. However, adaptive neighborhood

selection is more suitable for the evolution of population. So, we designed an adaptive neighborhood radius calculation method and proposed an adaptive neighborhood search strategy to improve the computation performance of the whale optimization algorithm by using the useful information of the neighborhood.

In the process of searching for prey, each whale individual searches within a range and communicates with other whales within the range to share information. In order to find the optimal position faster and improve the convergence speed of the algorithm, therefore, the design of effective search radius is the key to the planning of whale search range. For the i-th whale $\boldsymbol{x}_i(t)$ in the current population, calculates its distance from other individuals $\boldsymbol{x}_k(t), k \neq i, k = 1, 2, \cdots, N$

$$d_i^k(t) = \sqrt{\sum_{j=1}^{n} \left(x_{i,j}(t) - x_{k,j}(t) \right)^2} \tag{5}$$

Define the search radius $R_i(t)$ of the *i*-th whale as follow:

$$R_i(t) = \frac{\sum_{k=1}^{N} d_i^k(t)}{N-1} \tag{6}$$

that is, $R_i(t)$ represents the average distance from the *i*-th whale to others. With the adaptive change of the number of iterations t and individual i, the average value can remove individuals that are father away from the individual i, so that the individual i forms a reachable neighborhood with the visual range, narrowing the individual's learning range, and reducing the individual's blind search. The neighborhood Nb_i of individual i is:

$$Nb_i(t) = \left\{ \boldsymbol{x}_k(t) \middle| d_i^k(t) \leq R_i(t), k \neq i, k = 1, 2, \cdots, N \right\} \tag{7}$$

the best individual in neighborhood is $\boldsymbol{x}_i^{lbest}(t) = \underset{\boldsymbol{x}_k \in Nb_i}{\arg} f(\boldsymbol{x}_k)$.

When $A > 1$, whale performs global adaptive neighborhood search. In order to enhance the ability to jump out of local optimality and increase the speed of convergence. The Eq. (4) are improved as follows:

$$d_{i,j} = \left| 2 \cdot rand \cdot x_i^{lbest}{}_{,j}(t) - x_{i,j}(t) \right| \tag{8}$$

$$x_{i,j}(t+1) = x_i^{lbest}{}_{,j}(t) - A \cdot d_{i,j} \tag{9}$$

The WOA algorithm randomly learns from an individual in the population. Assuming that the individual fitness value of the learned individual is poor, convergence speed will be reduced. The introduced adaptive neighborhood method can make the whale learn from the optimal solution to the neighborhood compared with the WOA algorithm. The individual being learned is not the worst, at least within the neighborhood, thus enhancing the ability of the algorithm to jump out of the local optimal and improving the speed of convergence.

3.2. Hybrid mutation strategy

Based on the optimization algorithm of population, its global exploration ability and local development ability are contradictory to some extent, and it is necessary to reasonably balance the development and exploration ability. The WOA algorithm has strong development ability. At the same time, its exploration ability is weak and it is easy to fall into the local optimal solution. In order to overcome this shortcoming, a hybrid mutation strategy was designed, combining Gaussian and Cauchy mutation operation to enhance the algorithm's ability to jump out of the local optimal solution.

Gaussian mutation operation:

$$x_{i,j}^{'}(t+1) = x_{i,j}(t+1) + c_{i,j}^{1} \cdot G_{i,j}(\mu_{i,j}, \sigma_{i,j}) \tag{10}$$

Cauchy mutation operation:

$$x_{i,j}^{'}(t+1) = x_{i,j}(t+1) + c_{i,j}^{2} \cdot C_{i,j}(\mu_{i,j}^{'}, \sigma_{i,j}^{'}) \tag{11}$$

Hybrid mutation operation:

$$x_{i,j}^{'}(t+1) = x_{i,j}(t+1) + w_{i,j}^{1} \cdot c_{i,j}^{1} \cdot G_{i,j}(\mu_{i,j}, \sigma_{i,j}) + w_{i,j}^{2} \cdot c_{i,j}^{2} \cdot C_{i,j}(\mu_{i,j}^{'}, \sigma_{i,j}^{'}) \tag{12}$$

where $G_{i,j}(\mu_{i,j}, \sigma_{i,j})$ is a random number of Gaussian distribution; $C_{i,j}(\mu_{i,j}^{'}, \sigma_{i,j}^{'})$ is a random number of Cauchy distribution. $\mu_{i,j}$, $\sigma_{i,j}$ are the mean and variance of Gaussian distribution; $\mu_{i,j}^{'}$, $\sigma_{i,j}^{'}$ are the mean and variance of Cauchy distribution. $c_{i,j}^{1}$, $c_{i,j}^{2}$ are the coefficients of Gaussian and Cauchy mutation. $c_{ij}^{1} = \sigma_{ij} \cdot e^{(\alpha \cdot N(0,1) + \beta \cdot N_j(0,1))}$, $c_{ij}^{2} = \sigma_{ij}^{'} \cdot e^{(\alpha^{'} \cdot N(0,1) + \beta^{'} \cdot N_j(0,1))}$, $\alpha = \frac{1}{\sqrt{2n}}$, $\beta = \frac{1}{\sqrt{2\sqrt{n}}}$. $w_{i,j}^{1}$, $w_{i,j}^{2}$ are weights, for any i, j the equation $w_{i,j}^{1} + w_{i,j}^{2} = 1$ needs to be satisfied.

3.3 The proposed of HMNWOA plgorithm

In order to improve the convergence speed of whale optimization algorithm and increase the diversity of population, an improved whale optimization algorithm based on adaptive neighborhood and hybrid mutation strategy (HMNWOA) is proposed. The HMNWOA algorithm implementation steps are as follows:

Step 1: Initialize whale optimization algorithm parameters, such as population size N, dimension n, maximum number of iterations $t_{\max}$, the shape of logarithmic spiral b.

Step 2: Randomly initialize the whale population G^0 and record the best solution as $\boldsymbol{x}_*(0)$.

Step 3: If $p < 0.5$ and $|A| \le 1$, according to Eq. (2) to update position.

Step 4: If $p < 0.5$ and $|A| > 1$, according to Eqs. (5)- (7) to calculate the adaptive neighborhood Nb_i, then update their position by Eqs. (8)- (9).

Step 5: If $p \ge 0.5$, update position according to Eq. (3).

Step 6: Update the position according to Eq. (12).

Step 7: Fix the search individuals that go beyond the boundaries of the search space.

Step 8: Update the global best solution $\boldsymbol{x}_*(t+1)$.

Step 9: If terminal condition is met, output the best solution $\boldsymbol{x}_*(t+1)$, $f(\boldsymbol{x}_*(t+1))$. Otherwise set $t=t+1$, $\boldsymbol{x}_*(t)=\boldsymbol{x}_*(t+1)$, and go to Step 3.

4 HMNWOA algorithm for wrapper feature selection

In this section, we discuss the binary HMNWOA method for feature selection [Zarshenas and Suzuk (2016)] and the fitness function calculation. Similarly [Mafarja and Mirjalili (2017)], for a feature set consisting of *n* features, each feature subset is regarded as the position of whale that is an *n* dimensional vector that each element is 0 or 1. "1" indicates that the corresponding feature is selected, while "0" indicates that the feature is not selected.

Feature selection is a multi-objective optimization problem, requiring that the number of selected features be as few as possible, and that the accuracy of classification using this feature subset be as high as possible. These two conflicting goals can be translated into the following minimization issues:

$$\min f(\boldsymbol{x})=\left(\lambda_1\cdot error_x+\lambda_2\cdot\frac{R}{n}\right) \tag{13}$$

where $error_x$ represents the error rate of using the characteristic subset $\boldsymbol{x}$ for the KNN classifier; R represents the number of selected features; n is the number of features in the original dataset. $\lambda_1\in[0,1]$, $\lambda_2=1-\lambda_1$ indicate penalty factor [Emary, Zawbaa and Hassanier (2016)]. Using $f(\boldsymbol{x})$ as the measure of the best subset, the relationship between feature numbers and classification accuracy is effectively balanced.

For each feature $x_{i,j}$, the value can only be 0 or 1. Therefore, we need to binarize the variables $x_{i,j}$.In the initialization phase, set $\boldsymbol{l}=0$, $\boldsymbol{u}=1$. The $x_{i,j}$ is the *j*-th dimension of the *i*-th individual in the initialization population P.

$$x_{i,j}=\begin{cases}1, & x_{i,j}>0.5\\ 0, & x_{i,j}\le 0.5\end{cases}\qquad i=1,2,\cdots,N;\ j=1,2,\cdots,n \tag{14}$$

From the Eq. (8), we can know that $d_{i,j}\in[0,1]$. So, the method of binarization is as follows:

$$d_{i,j}=\begin{cases}1, & d_{i,j}>0.5\\ 0, & d_{i,j}\le 0.5\end{cases}\qquad i=1,2,\cdots,N;\ j=1,2,\cdots,n \tag{15}$$

Since the calculation of $x_{i,j}(t+1)$ in Eq. (2) and Eq. (9) is related to A, so

$$x_{i,j}^{'}(t+1)=\begin{cases}1, & x_{i,j}(t+1)>\dfrac{1}{2}-A\\ 0, & x_{i,j}(t+1)\le\dfrac{1}{2}-A\end{cases}\qquad i=1,2,\cdots,N;\ j=1,2,\cdots,n \tag{16}$$

The method of binarization of Eq. (3) is consistent with Eq. (15).

After Eq. (12) is updated, we perform a binary method as follows:

$$x_{i,j}(t+1)=\begin{cases}1, & x'_{i,j}(t+1)>\dfrac{t}{t_{\max}}\\ 0, & x'_{i,j}(t+1)\le\dfrac{t}{t_{\max}}\end{cases}\quad i=1,2,\cdots,N;\ j=1,2,\cdots,n \tag{17}$$

In summary, the optimal model of feature selection can be expressed as:

$$\min f(\boldsymbol{x})=\left(\lambda_1\cdot error_{\boldsymbol{x}}+\lambda_2\cdot\frac{R}{n}\right)$$
$$\text{s.t. } x_i\in\{0,1\} \tag{18}$$

For this feature selection optimization problem, HMNWOA algorithm is used to solve the problem. The pseudo code of HMNWOA for feature selection is shown in **Algorithm 1**.

Algorithm 1: HMNWOA algorithm for feature selection

Begin

Initialize the parameters, population size N ,dimension n ,maximum number of iterations $t_{\max}$,the shape of logarithmic spiral b

Randomly initialize the whale population G^0 and set $t=1$, than use Eq. (14) to update G^0

Assess the objective function value for each search individual, $\boldsymbol{x}_*(0)$ is the best search individual

While ($t<t_{\max}$)

For each search individual

If $p<0.5$

If ($|A|\le 1$)

Use Eq. (2), Eq. (16) to update the search individual

Else

Find neighborhoods via Eqs. (5)- (7) and update the local best in neighborhoods

Use Eq. (8), Eq. (15), Eq. (9), Eq. (16) to update the search individual

End If

Else

Use Eq. (3), Eq. (15) to update the search individual

End If

Use Eq. (12), Eq. (17) to update the search individual

Fix the search individuals that go beyond the boundaries of the search space

End For

Update the best search individual $\boldsymbol{x}_*(t+1)$

$t=t+1$

End While

Output the best search individual $\boldsymbol{x}_*$

End

5 Experimental results and analysis

In order to test the performance of the HMNWOA algorithm for feature selection, 12 sets of data were selected for experimentation from the UCI database [Blake and Merz (1998)]. The description of the data is shown in Tab. 1.

Table 1: List of datasets used in experiments

Dataset	No. of attribute	Size
Breastcancer	9	699
BreastEW	30	569
CongressEW	16	435
HeartEW	13	270
lonosphereEW	34	351
KrvskpEW	36	3196
Lymphography	18	148
SonarEW	60	208
Tic-tac-toe	9	958
WaveformEW	40	5000
WineEW	13	178
Zoo	16	101

The whale optimization algorithm is combined with the simulated annealing algorithm, and the tournament selection strategy is introduced to obtain the new algorithm WOASAT-2 in Mafarja et al. [Mafarja and Mirjalili (2017)]. We will compare HMNWOA algorithm with WOA, ALO, GA, PSO, Full, WOASAT-2 algorithms in [Mafarja and Mirjalili (2017)]. The Full indicates that all features are selected. The results of comparison are shown in Tab. 2. In this paper, set $\mu_{i,j} = \mu_{i,j}^{'} = 0$, $\sigma_{i,j} = \sigma_{i,j}^{'} = 1$, $w_{i,j}^{1} = w_{i,j}^{2} = 0.5$, population size $N = 10$, dimension n is the characteristic number of each test data set; the number of iterations is 100, and each algorithm runs 5 times independently.

Table 2: Average of classification accuracy and number of features of 7 algorithms

Dataset	Algorithm						
	WOA	ALO	GA	PSO	Full	WOASAT-2	HMNWOA
Breastcancer	0.96/6.40	0.96/6.82	0.96/5.09	0.95/5.72	0.94/9.00	0.97/4.20	**0.97/4.10**
BreastEW	0.93/23.80	0.93/16.08	0.94/16.35	0.94/16.56	0.96/30.00	**0.98**/11.60	0.95/**5.00**
CongressEW	0.93/10.00	0.93/6.98	0.94/6.62	0.94/6.83	0.92/16.00	**0.98**/6.40	0.97/**3.20**
HeartEW	0.79/9.40	0.83/10.31	0.82/9.49	0.78/7.94	0.82/13.00	**0.85**/5.40	0.84/**4.80**
lonosphereEW	0.87/22.40	0.87/9.42	0.83/17.31	0.84/19.18	0.87/34.00	0.96/12.80	**0.96/6.20**
KrvskpEW	0.93/24.20	0.96/24.70	0.92/22.43	0.94/20.81	0.92/36.00	0.98/18.40	**0.99/18.20**
Lymphography	0.78/10.80	0.79/11.05	0.71/11.05	0.69/8.98	0.68/18.00	0.89/7.20	**0.89/6.80**
SonarEW	0.86/46.40	0.74/37.92	0.73/33.30	0.74/31.20	0.62/60.00	**0.97**/26.40	0.95/**20.00**
Tic-tac-toe	0.76/8.40	0.73/6.99	0.71/6.85	0.73/6.61	0.72/9.00	0.79/6.00	**0.80/5.60**
WaveformEW	0.71/33.60	0.77/35.72	0.77/25.28	0.76/22.72	0.77/40.00	0.76/**20.60**	**0.80**/28.60
WineEW	0.95/7.40	0.91/10.70	0.93/8.63	0.95/8.36	0.93/13.00	**0.99**/6.40	0.97/**4.60**
Zoo	0.96/8.80	0.91/13.97	0.88/10.11	0.83/9.74	0.79/16.00	0.97/5.60	**0.97/4.80**
Average	0.87/17.63	0.86/15.89	0.85/14.38	0.84/13.72	0.83/24.50	0.92/10.92	**0.92/9.38**

In Tab. 2, A/B indicates the accuracy of the classification/number of features, bold data represents maximum accuracy and minimum number of features. From Tab. 2, we can see that the selection of all features is less effective, wasting neither time nor improving the accuracy of the classification. Compared with WOA, ALO, GA, and PSO algorithms, the proposed algorithm has significantly improved the accuracy of classification. 67% of datasets have obtained an accuracy rate of more than 95%, and some have even reached 99%. Compared with the WOASAT-2 algorithm, the HMNWOA algorithm has 50% better than the WOASAT-2 algorithm in terms of classification accuracy. In the remaining datasets, WOASAT-2 has better performance, but HMNWOA ranks second, and the gap with WOASAT-2 algorithm is within 0.03. Based on the average accuracy of algorithms in Tab. 2, it can be seen that the average accuracy of WOASAT-2 and HMNWOA algorithms are 92%, while the accuracy of other algorithms is about 85%. In terms of classification accuracy, the HMNWOA and WOASAT-2 algorithms performed

significantly better. In order to compare the classification accuracy of 7 algorithms more intuitively, Fig. 1 gives the accuracy of the classification of 12 sets of data by 7 algorithms. As can be seen from Fig. 1, for all data sets, the HMNWOA algorithm has a high classification accuracy, ranking first or second.

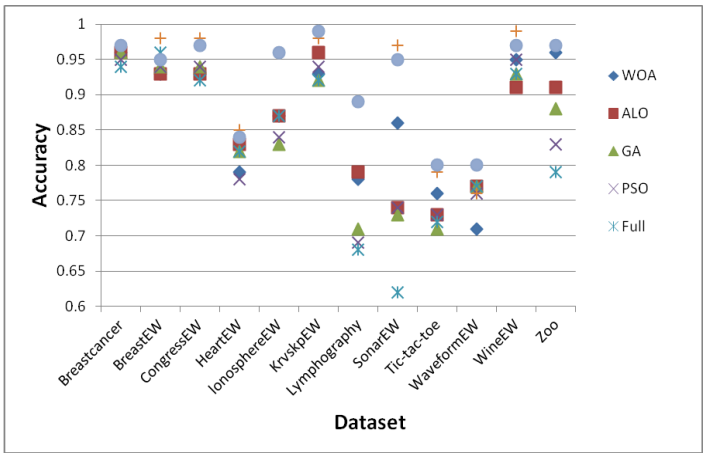


Figure 1: Average of classification accuracy of 7 algorithms

As can be seen from Tab. 2, with the exception of the data set WaveformEW, the number of features selected for the remaining data set HMNWOA is less than the number of features selected by other algorithms, so the HMNWOA algorithm performs well. Moreover, the number of features selected by the two data sets of CongrissEW and SonarEW are significantly less than that of other algorithms. From the average value, we can see that the HMNWOA algorithm selects significantly fewer features. In terms of feature number, the HMNWOA algorithm performs better than other algorithms. In order to compare the feature numbers of the seven algorithms more clearly, Fig. 2 gives the feature numbers of the seven algorithms that classify the 12 sets of data. It can be seen from Fig. 2 that, except for the data set WaveformEW, the HMNWOA algorithm has the least number of features for other data sets, indicating that the method can reduce the number of features, use fewer features for accurate classification, and reduce classification time.

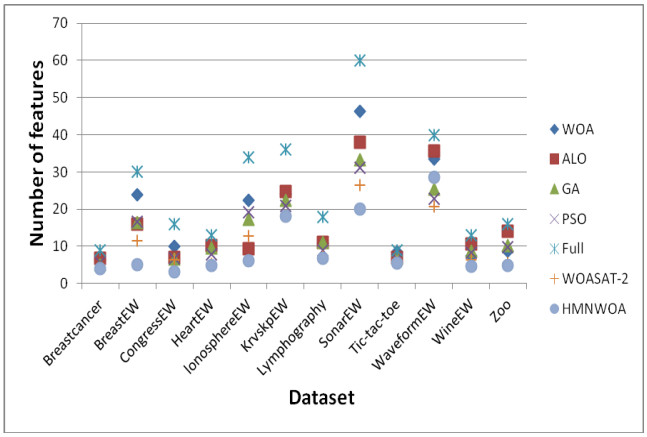


Figure 2: Average of number of features of 7 algorithms

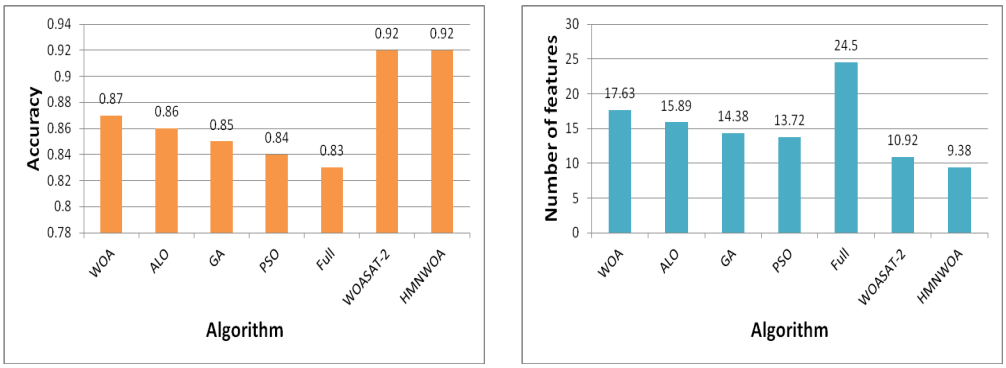
In summary, the HMNWOA algorithm is significantly superior to the WOA, ALO, GA, and PSO algorithms speaking of classification accuracy and feature numbers. Compared with the WOASAT-2 algorithm, the two algorithms perform equally in terms of classification accuracy, while the HMNWOA algorithm performs well in terms of feature number. Therefore, in order to analyze the accuracy of the classification and the number of features at the same time, the results of the two algorithms are statistically different, as shown in Tab. 3.

Table 3: Difference between the results of WOASAT-2 and HMNWOA algorithms

Dataset	Accuracy	Attributes
Breastcancer	0	0.1
BreastEW	0.03	**6.6**
CongressEW	0.01	**3.2**
HeartEW	**-0.09**	0.6
IonosphereEW	0	**6.6**
KrvskpEW	**-0.01**	0.2
Lymphography	0	0.4
SonarEW	0.02	**6.4**
Tic-tac-toe	**-0.01**	0.4
WaveformEW	**-0.04**	-8
WineEW	0.02	**1.8**
Zoo	0	0.8
Average	0	1.54

The difference between the two algorithms is shown in Tab. 3. The data in the Table are the values subtracted from the results calculated by the WOASAT-2 and HMNWOA algorithms. From Tab. 3, it can be seen that the accuracy data marked in bold is negative, indicating that the accuracy of the WOASAT-2 algorithm is less than that of the HMNWOA algorithm. The number of features corresponding to these four data is greater than 0. Therefore, for these four datasets, the HMNWOA algorithm performs well in terms of accuracy and selection characteristics, achieving the goal of reducing the number of features while improving the classification accuracy. For the datasets used for the five sets of data marked by feature numbers, the HMNWOA algorithm has fewer feature numbers, and the accuracy gap is not large, within 0.03. It shows that the HMNWOA algorithm uses a few features to classify, achieves a relatively high accuracy rate, and saves classification time. There are four sets of data sets with the same classification accuracy of the two algorithms, but the HMNWOA algorithm reduces the number of features. For the average value, the HMNWOA algorithm selects fewer features under the same accuracy. In summary, the two algorithms have their own advantages. In general, the HMNWOA algorithm can reduce the number of features, and at the same time, the accuracy is relatively high.

In order to compare the performance of all algorithms as a whole, Fig. 3 gives the average accuracy and feature numbers of the 12 data sets tested by 7 algorithms. As can be seen from Fig. 3(a), in terms of accuracy, Full<PSO<GA<ALO<WOA<WOASAT-2=HMNWOA. From Fig. 3(b), it can be seen that from the feature number, the performance of the algorithm is Full<WOA<ALO<GA<PSO<WOASAT-2<HMNWOA. Therefore, the HMNWOA algorithm has the best performance regardless of the accuracy or the number of features, indicating that the algorithm can accurately classify while reducing the number of features.



(a) Accuracy (b) Number of features

Figure 3: Average accuracy and number of features of the 12 data sets tested by 7 algorithms

Tabs. 4-6 show the average, minimum, and maximum of the fitness values calculated by the five algorithms for all datasets. Among them, bold data represent the minimum value calculated by all algorithms. From Tabs. 4-6, we can see that the HMNWOA algorithm calculates that the average and minimum values of fitness are better than other algorithms, and the maximum value is also smaller than other algorithms, indicating that the

algorithm is relatively stable and when applied to feature selection, the effect is very good, so that the objective function value can be minimized.

Table 4: Mean fitness values obtained from the different approaches

Dataset	Algorithm				
	ALO	GA	PSO	WOASAT-2	HMNWOA
Breastcancer	**0.02**	0.03	0.03	0.04	0.03
BreastEW	**0.03**	0.04	**0.03**	**0.03**	0.05
CongressEW	0.05	0.04	0.04	0.03	**0.03**
HeartEW	**0.12**	0.14	0.15	0.16	0.16
lonosphereEW	0.11	0.13	0.14	**0.04**	0.05
KrvskpEW	0.05	0.07	0.05	0.02	**0.02**
Lymphography	0.14	0.17	0.19	0.11	**0.11**
SonarEW	0.18	0.13	0.13	**0.03**	0.06
Tic-tac-toe	0.22	0.24	0.24	0.21	**0.21**
WaveformEW	0.21	**0.20**	0.22	0.25	**0.20**
WineEW	0.02	**0.01**	0.02	**0.01**	0.03
Zoo	0.07	0.08	0.10	0.04	**0.02**

Table 5: Minimum fitness values obtained from the different approaches

Dataset	Algorithm				
	ALO	GA	PSO	WOASAT-2	HMNWOA
Breastcancer	0.02	0.02	0.03	0.03	**0.02**
BreastEW	0.03	**0.02**	**0.02**	**0.02**	0.04
CongressEW	0.03	0.03	0.03	0.02	**0.02**
HeartEW	**0.11**	0.12	0.13	0.13	0.13
lonosphereEW	0.10	0.09	0.12	**0.03**	0.04
KrvskpEW	0.03	0.03	0.03	0.02	**0.01**
Lymphography	**0.08**	0.12	0.14	0.09	0.09

SonarEW	0.13	0.07	0.07	**0.01**	0.05
Tic-tac-toe	0.20	0.21	0.21	0.20	**0.20**
WaveformEW	0.19	0.19	0.21	0.23	**0.19**
WineEW	**0.00**	**0.00**	**0.00**	**0.00**	0.02
Zoo	0.04	0.00	0.03	0.00	**0.00**

Table 6: Maximum fitness values obtained from the different approaches

Dataset	Algorithm				
	ALO	GA	PSO	WOASAT-2	HMNWOA
Breastcancer	0.03	**0.04**	0.03	**0.04**	**0.04**
BreastEW	0.04	0.05	0.05	0.04	**0.06**
CongressEW	**0.06**	**0.06**	0.04	0.05	0.04
HeartEW	0.13	0.14	0.18	0.18	**0.19**
lonosphereEW	0.12	0.16	**0.17**	0.05	0.05
KrvskpEW	0.07	**0.13**	0.07	0.02	0.02
Lymphography	0.16	**0.27**	**0.27**	0.14	0.13
SonarEW	**0.26**	0.23	0.22	0.05	0.08
Tic-tac-toe	0.24	0.26	**0.27**	0.23	0.23
WaveformEW	0.22	0.21	0.23	**0.26**	0.21
WineEW	0.03	0.03	0.03	0.03	**0.46**
Zoo	0.12	0.18	**0.21**	0.10	0.09

6 Conclusions

At present, the swarm intelligence algorithm is the most influential method for solving optimization problems. It has become the focus of research on balancing the global search and local search capability, improving the convergence speed and calculation accuracy, and expending the application field of the algorithm. In this paper, a new adaptive neighborhood generation strategy is designed to reduce the blindness brought by random learning in the WOA algorithm, and the hybrid mutation strategy based on Gaussian mutation operator and Cauchy mutation operator is used to improve exploration capability. A whale optimization algorithm based on optimal neighborhood and hybrid mutation strategies (HMNWOA) is proposed to enhance the exploration and exploitation

ability of the algorithm. The standard data set verifies that the new method improves the classification accuracy while effectively reducing the number of features, which is superior to the current intelligent algorithm for feature selection. In the future, we can combine HMNWOA with other classification algorithms (other than KNN), or use it in feature selection algorithms based on a mixture of filtering and wrapper methods, and further explore the feature selection method on the basis of the WOA algorithm. It provides an effective data pretreatment method for big data research. Secondly, in this paper, we use whale optimization algorithm to solve the problem of engineering optimization, while geometric computing [Ghasemi, Park and Rabczuk (2017); Ghasemi, Park and Rabczuk (2018)] is a mathematical optimization problem. So in the next work, we try to use whale optimization algorithm to solve geometric optimization problems and further broaden the application field of whale optimization algorithm. In addition, the idea of genetic algorithm is applied to whale optimization algorithm, and it is coded to solve the integer optimization problem [Anitescu, Atroshchenko, Alajlan et al. (2019)].

Acknowledgement: This work was supported by the National Natural Science Foundation of China (Grant No. 2017YFC0403605 and No. 11601419).

References

Anitescu, C.; Atroshchenko, E.; Alajlan, N.; Rabczuk, T. (2019): Articial neural network methods for the solution of second order boundary value problems. *Computers, Materials & Continua*, vol. 59, no. 1, pp. 345-359.

Aziz, M. A. E.; Ewees, A. A.; Hassanien, A. E. (2017): Whale optimization algorithm and moth-flame optimization for multilevel thresholding image segmentation. *Expert Systems with Applications*, vol. 83, pp. 242-256.

Blake, C. L.; Merz, C. J. (1998): UCI Repository of machine learning databases. http://www.ics.uci.edu/.

Chandrashekar, G.; Sahin, F. (2014): A survey on feature selection methods. *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16-28.

Crone, S. F.; Kourentzes, N. (2010): Feature selection for time series prediction-A combined filter and wrapper approach for neural networks. *Neurocomputing*, vol. 73, no. 10, pp. 1923-1936.

Emary, E.; Zawbaa, H. M.; Hassanier, A. E. (2016): Binary ant lion approaches for feature selection. *Neurocomputing*, vol. 213, pp. 54-65.

Emary, E.; Zawbaa, H. M.; Grosan, C. (2014): Feature subset selection approach by gray-wolf optimization. *Afro-European Conference for Industrial Advancement*, pp. 1-13.

Fong, S.; Wong, R.; Vasilakos, A. V. (2016): Accelerated PSO swarm feature selection for data stream mining big data. *IEEE Transactions on Services Computing*, vol. 9, no. 1, pp. 33-45.

Ghasemi, H.; Park, H. S.; Rabczuk, T. (2017): A level-set based IGA formulation for topology optimization of flexoelectric. *Computer Methods in Applied Mechanics and Engineering*, vol. 313, pp. 239-258.

Ghasemi, H.; Park, H. S.; Rabczuk, T. (2018): A multi-material level set-based topology optimization of flexoelectric composites. *Computer Methods in Applied Mechanics and Engineering*, vol. 332, pp. 47-62.

Hancer, E.; Xue, B.; Zhang, M. (2017): Differential evolution for filter feature selection based on information theory and feature ranking. *Knowledge-Based Systems*, pp. 1-17.

Huang, Q. B.; Li, J. X.; Song, C. N.; Xu, C. H.; Lin, X. F. (2018): Whale optimization algorithm based on cosine control factor and polynomial mutation. *Control and Decision*.

Kumar, C. (1998): Combining mutation operators in evolutionary programming. *IEEE Transactions on Evolutionary Computation*, vol. 2, no. 3, pp. 91-96.

Ling, Y.; Zhou, Y. Q.; Luo, Q. F. (2017): Levy flight trajectory-based whale optimization algorithm for global optimization. *IEEE Access*, vol. 5, pp. 6168-6186.

Mafarja, M. M.; Mirjalili, S. (2017): Hybird whale optimization algorithm with simulated annealing for feature selection. *Neurocomputing*, vol. 260, pp. 302-312.

Mafarja, M.; Mirjalili, S. (2018): Whale optimization approaches for wrapper feature selection. *Applied Soft Computing*, vol. 62, pp. 441-453.

Muthuramalingam, A.; Kumar, S. S.; Anusheela, N. (2011): A novel feature subset selection algorithm using artificial bee colony in keystroke dynamics. *Proceedings of the International Conference on Soft Computing for Problem Solving*, pp. 759-766.

Paul, S.; Magdon-Ismail, M.; Drineas, P. (2016): Feature selection for liner SVM with provable guarantees. *Pattern Recognition*, vol. 60, pp. 205-214.

Sharawi, M.; Zawbaa, H. M.; Emary, E. (2017): Feature selection approach based on whale optimization algorithm. *Proceedings of the 2017 International Conference on Advanced Computational Intelligence*, pp. 4-6.

Spolaôr, N.; Lorena, A. C.; Lee, H. D. (2018): Feature selection via Pareto multi-objective genetic algorithms. *Applied Artificial Intelligence*, vol. 31, no. 32, pp. 1-28.

Tharwat, A.; Moemen, Y. S.; Hassanien, A. E. (2017): Classification of toxicity effects of biotransformed hepatic drugs using whale optimized support vector machines. *Journal of Biomedical Informatics*, vol. 68, pp. 132-149.

Vasilic, S.; Kezunovic, M. (2005): Fuzzy ART neural network algorithm for classifying the power system faults. *IEEE Transactions on Power Delivery*, vol. 20 no. 2, pp. 1306-1314.

Vu-Bac, N.; Duong, T. X.; Lahmer, T.; Zhuang, X.; Sauer, R. A. et al. (2017): A NURBS-based inverse analysis for reconstruction of nonlinear deformations of thin shell structures. *Computer Methods in Applied Mechanics and Engineering*, vol. 331, pp. 427-455.

Wang, A.; An, N.; Chen, G.; Li, L. (2015): Accelerating wrapper-based feature selection with K-nearest-neighbor. *Knowledge-Based Systems*, vol. 83, no. C, pp. 81-91.

Wang, J. Z.; Du, P.; Niu, T. (2017): A novel hybrid system based on a new proposed algorithm multi-objective whale optimization algorithm for wind speed forecasting. *Applied Energy*, vol. 208, pp. 244-360.

Xiong, G. J.; Zhang, J.; Shi, D. Y.; He, Y. (2018): Parameter extraction of solar photovoltaic models using an improved whale optimization. *Energy Conversion and Management*, vol. 174, pp. 388-405.

Xue, B.; Fu, W.; Zhang, M. (2014): Difference evolution for multi-objective feature selection in classification. *Companion Publication of the Conference on Genetic and Evolution*, pp. 83-84.

Zarshenas, A.; Suzuk, K. (2016): Binary coordinate assent: an efficient optimization technique for feature subset selection for machine learning. *Knowledge-Based Systems*, vol. 110, pp. 191-201.

Zawbaa, H. M.; Emary, E.; Parv, B. (2015): Feature selection based on antlion optimization algorithm. *Proceeding of the 2015 Third World Conference on Complex Systems*, pp. 1-7.

Zorarpaci, E.; ÖzeA, S. A. (2016): Hybrid approach of differential evolution and artificial bee colony for feature selection. *Expert Systems with Applications*, vol. 62, pp. 91-103.