



Attribute Selection Method Based on Artificial Bee Colony Algorithm and Neighborhood Discrimination Matrix Optimization

Yuxuan Ji¹, Jun Ye^{1,2}(✉), Zhenyu Yang¹, Jiaxin Ao¹, and Lei Wang^{1,2}

¹ Nanchang Institute of Technology, Nanchang 330000, Jiangxi, China
2003992646@nit.edu.cn

² Key Laboratory of Water Information Cooperative Perception and Intelligent Processing of Jiangxi Province, Nanchang 330000, Jiangxi, China

Abstract. At present, the existing attribute reduction algorithm combining artificial bee colony and neighborhood rough set basically uses the attribute dependence and the number of attribute subsets as parameters to construct the fitness function, while ignoring the role of heuristic information. As a result, the number of bee colony iterations increases, and the convergence speed is slow. Aiming at this kind of problem, an improved method is proposed. First, a discernibility matrix under the neighborhood rough set is defined; secondly, an attribute importance measurement method of the discernibility matrix under the neighborhood decision system is proposed; The attribute importance of the domain discrimination matrix constructs a new fitness function for the heuristic factor; finally, an attribute selection algorithm for artificial bee colony algorithm and neighborhood discrimination matrix importance optimization is designed. Compared with the original algorithm, the new method reduces the number of generations, accelerates the convergence speed, and retains the minimum attribute reduction collected during each iteration, and multiple minimum attribute reductions can be obtained. The experimental results on the UCI data set prove the feasibility and effectiveness of the algorithm.

Keywords: Neighborhood decision system · Discernibility matrix · Artificial bee colony algorithm · Feature selection · Attribute importance

1 Introduction

Among the mathematical tools for dealing with fuzzy and inaccurate problems, rough set theory is widely used in many fields. However, the classic rough set

Supported by the National Natural Science Foundation of China (Nos. 61562061) and Technology Project of Ministry of Education of Jiangxi province of China (No.GJJ211920, GJJ170995).

model is not suitable for dealing with continuous data problems. For this reason, some scholars have introduced the concept of neighborhood rough set in order to solve such problems. Lin [1] proposed the concept of neighborhood model relation based on topology from the perspective of granularity. Subsequently, Hu et al. [2] proposed a fast reduction algorithm based on the neighborhood rough set model and a hybrid data reduction algorithm based on the neighborhood rough set, which can directly process unprocessed numerical data, thereby extending the classical rough set model. Application areas. In the neighborhood rough model, because the computational workload is much larger than that of discrete data, the attribute reduction and feature dimensionality reduction in the neighborhood rough model have a higher time complexity, especially with the dimensionality of the decision table. The increase in time complexity has increased geometrically. Therefore, reducing time complexity and reducing computational workload are the key research contents of many scholars.

In recent years, swarm intelligence algorithms have achieved good results in solving some optimization problems [4–8]. The researchers introduced the artificial bee colony optimization algorithm [4] into the attribute reduction of the neighborhood rough set, by constructing a moderate function to guide the colony to search for the smallest feature subset to achieve the purpose of reduction. The designed moderate function can reduce the number of iterations and reduce the reduction. Reduce the time complexity, improve the reduction rate of the algorithm, and obtain multiple minimum attribute reductions. For example, literature [5] proposed an information gain-guided bee colony optimization algorithm, which constructs the mutual information between conditional features and decision-making features based on the information entropy of the features, and finally obtains a feature subset. Literature [6] uses a combination of fuzzy rough set and bee colony algorithm for attribute reduction. Through dependence and reduction rate, an objective function that can reflect the size and importance of the attribute set is constructed, and the attribute reduction problem is transformed into an optimization problem. Finally, the goal is The function is an iterative criterion. Literature [7] proposes a new combined bee colony algorithm to solve the MAR problem, in which the lead bee, follow bee and scout bee adopt a search mode based on mutation calculation. Literature [8] proposed a method of combining rough sets in the classic field with artificial bee colonies, and its algorithm can obtain multiple minimum feature subsets.

The neighborhood rough set attribute subset selection algorithms optimized by these artificial bee colony algorithms all use attribute dependence and the number of attribute subsets as parameters to construct fitness functions. Although more satisfactory results have been obtained in the attribute search subset, there are limitations. For this reason, literature [9] introduces attribute importance based on information entropy as heuristic information, and constructs a new fitness function in combination with attribute dependence, thereby proposing an attribute selection algorithm, which improves The speed of attribute reduction reduces the number of iterations. However, taking the importance of information entropy attribute as the fitness function of the heuristic information, the feature subset searched by the bee colony in some decision tables may contain redundant attributes. This paper draws on the fitness func-

tion constructed in the literature [9], based on the attribute distinguishing different neighborhood objects in the discernibility matrix, defines a new attribute importance, and replaces the information entropy attribute in the fitness function with the new attribute importance Importance, an improved attribute selection algorithm of artificial bee colony and neighborhood rough set discernibility matrix is proposed. The main advantage of this algorithm is to provide an idea for the neighborhood rough set attribute reduction method, which reduces the number of generations and speeds up convergence speed and retention of the minimum attribute reduction collected during each iteration will make the result of generating attribute reduction more accurate. The rest of this article is organized as follows. Section 2 introduces the knowledge of neighborhood rough set and artificial bee colony algorithm, Sect. 3 gives attribute reduction algorithm, Sect. 4 gives experimental analysis, and Sect. 5 gives conclusions.

2 Related Knowledge

In order to facilitate understanding, the following briefly introduces related knowledge of neighborhood rough set, artificial bee colony optimization and its improved algorithm related to this article. For more detailed knowledge, please refer to literature [1,3].

2.1 Basic Knowledge of Neighborhood Rough Set

In the rough set of neighborhood, the equivalence relationship is extended to the neighborhood relationship, and the information of the original data is retained to the greatest extent.

Definition 1. [2] Given N in the real-dimensional space Ω , $\Delta = R^N \times R \rightarrow R$, then called Δ is a metric on R^N . A non-empty finite set $U = \{x_1, x_2, \dots, x_n\}$ on a given real number space Ω . The neighborhood δ of $\forall x_i$ is defined $\forall x_i$:

$$\delta(x_i) = \{x \mid x \in U, \Delta(x, x_i) \leq \delta\} \quad (1)$$

Δ represents the distance, and the Euclidean distance is used in this article, which is $p = 2$.

$$\Delta_p(x_1, x_2) = \left(\sum_{i=1}^N |f(x_1, a_k) - f(x_2, a_k)|^p \right)^{\frac{1}{p}} \quad (2)$$

Definition 2. [2] Neighborhood decision system $NDS = (U, C \cup D, V, f)$, to $B \subseteq C, X \subseteq U$, Then the lower approximation and upper approximation of the neighborhood are respectively defined as:

$$\underline{N}_B(X) = \{x_i \mid \delta_B(x_i) \subseteq X, x_i \in U\} \quad (3)$$

$$\bar{N}_B(X) = \{x_i \mid \delta_B(x_i) \cap X \neq \emptyset, x_i \in U\} \quad (4)$$

By Definition 2, the positive domain and negative domain of the neighborhood decision system are:

$$\text{Pos}_B(D) = \underline{N}_B D \quad (5)$$

$$\text{Neg}_B(D) = U - \bar{N}_B D \quad (6)$$

Definition 3. [2] Neighborhood decision system $NDS = (U, C \cup D, V, f)$, The dependence of decision attribute D on conditional attribute subset B is defined as:

$$\gamma_B(D) = \frac{|\text{Pos}_B(D)|}{|U|} \quad (7)$$

Definition 4. [2] Neighborhood decision system $NDS = (U, C \cup D, V, f)$, to $\forall c \in C$, If there is $\gamma_{C-\{c\}}(D) \neq \gamma_C(D)$, then c is called a necessary attribute, otherwise it is an unnecessary attribute, and all necessary attributes form the core attribute set $\text{CORE}(C)$.

Definition 5. [2] Neighborhood decision system $NDS = (U, C \cup D, V, f)$, $\forall A \subseteq C$, If the conditional attribute subset A satisfies: (1) $\gamma_A(D) = \gamma_C(D)$ (2) $\forall a \in A, \gamma_A(D) > \gamma_{A-\{a\}}(D)$.

The conditional attribute subset A is said to be a relative reduction of conditional attribute set, and all reduction sets are denoted as $\text{RED}(C)$.

2.2 Artificial Bee Colony and Its Improved Algorithm

Karaboga [4] and others proposed the artificial bee colony algorithm (ABC) to simulate the nectar-collecting behavior of bee colonies in nature. The algorithm converts the nectar source location problem into a function to generate the optimal solution. The process of bee colony searching for nectar source is the optimal solution. The process of. The algorithm idea is:

Suppose the dimension of the problem to be solved is D , the position of the honey source i in t iterations is expressed as $X_i^t = [x_{i1}^t, x_{i2}^t, \dots, x_{iD}^t]$. Represents the current number of iterations. $x_{im} \in (K_m, U_m)$, K_m and U_m respectively represent the upper and lower bounds of the cost parameter, $m = 1, 2, \dots, D$. Randomly generate the initial position of the desired nectar source i . At the beginning of the search, the hired bee searches around the nectar source i to generate a new nectar source.

When the fitness of the new honey source $W_i = [w_{i1}, w_{i2}, \dots, w_{iD}]$ is better than X_i , the greedy selection method is used to replace X_i with W_i , otherwise X_i is retained. After the hired bee completes the calculation, he flies back to the information exchange area to share the nectar source information. Follow the bee based on the nectar source information shared by the hired bee. The calculated probability is followed.

Then, the follower bee uses the roulette method to select the hired bee. During the search process, if the nectar source X_i reaches the threshold limit after trial iterations and no better nectar source is found, the nectar source

X_i will be abandoned and the corresponding employment The role of the bee is transformed into a scout bee. The scout bee will randomly generate a new nectar source in the search space to replace X_i .

Flow chart of artificial bee colony algorithm (Fig. 1.)

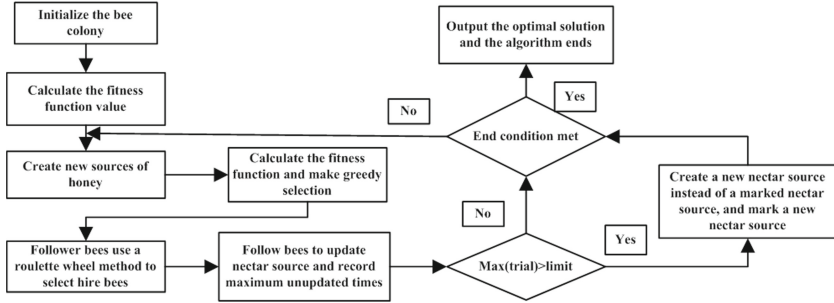


Fig. 1. Flow chart of artificial bee colony algorithm.

Artificial bee colony algorithm has the advantages of strong robustness and strong search ability. However, there are problems such as easy to fall into the local optimum and slow convergence speed in the later stage. Researchers have proposed a variety of improvement methods. Among them, the improved method of literature [10] has been widely used. Gao et al. [10] combined the differential evolution algorithm with the global optimal particle-improved bee colony algorithm, and proposed an algorithm with faster convergence speed. The information of the best solution is introduced into the generation of candidate honey sources, and the position update formula is modified to:

$$v_{ij} = x_{ij} + \varphi_{ij} (x_{ij} - x_{kj}) + \alpha_{ij} (x_{gj} - x_{kj}) \quad (8)$$

Among them, x_{ij} is the best nectar source found by the bee colony so far, $a_{ij} \in [0, 1]$. This paper draws on the Formula (8) for updating the location of the honey source constructed by Gao et al. to optimize the algorithm proposed in this paper.

3 Improved Attribute Selection Algorithm for Artificial Bee Colony and Neighborhood Discrimination Matrix

3.1 Definition of Attribute Importance of Neighborhood Discernibility Matrix

A. Skowron [11] first proposed the use of discernibility matrix to express knowledge system, which is defined as follows:

Definition 6. [11] Given a decision-making system $DS = (U, C \cup D, V, f)$, where C is the conditional attribute set, D is the decision attribute set, and the universe is a non-empty finite set of objects $U = \{x_1, x_2, \dots, x_n\}$, where is $|U| = n$. Then define the discernibility matrix of the decision-making system as:

$$(M_{ij})_{n \times n} = (c_{ij})_{n \times n} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix} \quad (9)$$

Among, $i, j = 1, 2, 3, \dots, n$

$$C_{ij} = \begin{cases} \{a \mid a \in C \wedge f(x_i) \neq f(x_j)\}, & f_D(x_i) \neq f_D(x_j) \\ \emptyset, & f_D(x_i) = f_D(x_j) \end{cases} \quad (10)$$

In this paper, the above Definition 7 is extended to the rough set of the neighborhood, and the lower discernibility matrix of the neighborhood relationship is defined. The definition method is as follows:

Definition 7. Neighborhood decision system $NDS = (U, C \cup D, V, f)$, where C is the condition attribute set, D is the decision attribute set, for any $x_i, x_j \in U$ and attributes $a \in C$, the discrimination matrix M_{ij} of the decision system NDS is defined as:

$$\{M_{ij}\}_{n \times n} = \begin{cases} \{a \mid a \in C \wedge |f_a(x_i) - f_a(x_j)| > \delta(x_i)\}, & f_D(x_i) \neq f_D(x_j) \\ 0, & f_D(x_i) \neq f_D(x_j) \wedge f_C(x_i) - f_C(x_j) < \delta(x_i) \\ \emptyset, & f_D(x_i) = f_D(x_j) \end{cases} \quad (11)$$

According to Definition 7, M_{ij} can be known that it is an upper triangle or a lower triangle matrix.

Theorem 1. In the neighborhood discernibility matrix M_{ij} , The set of all single attribute elements is the necessary (core) attribute set of C relative to D , namely $CORE_C(D) = \{a \mid a \in C \wedge (\exists c_{ij}, ((c_{ij} \in M) \wedge (c_{ij} = \{a\})))\}$

Through the analysis of Eq. (11), we know that for objects with equal decision attribute values, these objects do not need to be distinguished, so there is $c_{ij} = \varphi$. For objects whose decision attribute values are not equal and belong to the same neighborhood, if these objects are indistinguishable, there is $c_{ij} = 0$. For objects with unequal decision attribute values and belonging to different neighborhoods, these objects can be distinguished, then c_{ij} is composed of conditional attributes in C that can distinguish these objects, which may include the attributes in $CORE(C)$ or include Hor or K . All the attributes in c_{ij} play a role in distinguishing objects x_i and x_j in different neighborhoods, but the effects are different. The necessary attributes have the greatest effect, the relative necessary attributes are the second, and the unnecessary attributes contribute the least. In order to accurately measure the role of different types of conditional attributes in distinguishing objects in different neighborhoods, this paper uses the frequency of occurrence of each attribute in c_{ij} and the weight of each attribute to measure its importance, which is defined as follows:

Definition 8. Neighborhood decision-making system $NDS = (U, C \cup D, V, f)$, C is a set of conditional attributes, and D is a set of decision-making attributes. c_{ij} is a set of non-empty elements, K is the total number of sets of non-empty elements, for $\forall c \in C$, the importance of attribute c to D is defined as:

$$Nsig(c) = \begin{cases} 1, & |c_{ij}| = 1 \\ \frac{\sum_{i,j=1,2,\dots,n} r_{c \cap c_{ij}}}{K}, & |c_{ij}| > 1 \end{cases} \quad (12)$$

Among, $r_{c \cap c_{ij}} = \begin{cases} \frac{1}{|c_{ij}|}, & c \cap c_{ij} \neq \emptyset \\ 0, & c \cap c_{ij} = \emptyset \end{cases}$, $i, j = 1, 2, \dots, n, |c_{ij}| = 1$ represents the number of attributes in the collection c_{ij} .

It can be seen from the Formula (12) in Definition 8 that when $|c_{ij}| = 1$ is a single attribute, its importance is 1, which means that a single attribute contributes the most when distinguishing objects in different neighborhoods. This is the same as the single attribute in Theorem 2. The attributes match. When $|c_{ij}| > 1$, that is, the set of c_{ij} is composed of multiple attributes, it means that distinguishing objects in different neighborhoods is completed by multiple attributes, which may be necessary attributes, relatively necessary attributes or unnecessary attributes. Each attribute functions as $\frac{1}{|c_{ij}|}$. Obviously, the importance of a nuclear attribute is the sum of its importance as a single attribute 1 and the importance of multiple attributes, namely: $1 + \frac{\sum_{i,j=1,2,\dots,n} r_{c \cap c_{ij}}}{K}$, which also reflects that the nuclear attribute has the strongest ability to distinguish neighboring objects. For Definition 8, we can get two properties:

Property 1: For the set c_{ij} of non-empty elements in the discernibility matrix M , if $\forall c \in C$ and $c \in c_{ij}$, then $Nsig(c) > 0$.

It can be seen from Property 1 that Definition 9 can not only obtain the importance of necessary attributes, but also the relative importance of necessary attributes and unnecessary attributes. It shows that the attributes in the set all play a role in distinguishing objects in different neighborhoods, avoiding the situation that the importance of non-core attributes in Definition 6 are all zero.

Property 2: For the non-empty element set c_{ij} in the discernibility matrix M , there is $\forall a, b \in c_{ij}$, if there are $a \in CORE(C)$, $b \in N$, then there is $Nsig(a) > Nsig(b) > 0$. Among them, $N = C - CORE(C)$ represents the non-core attribute set.

Property 2 shows that the importance of nuclear attributes is the largest, and the importance of non-nuclear attributes is less than it. This is consistent with the properties of the rough set of neighborhoods. Therefore, Definition 9 accurately measures the role of attributes in neighborhood resolution.

3.2 Fitness Function Construction

The fitness function directly determines the evolution direction of the colony, the number of iterations, and the pros and cons of the solution. Therefore, the

appropriate fitness function plays an important role in the attribute selection method. At present, the attribute selection algorithm optimized by the artificial bee colony algorithm uses the fitness function of the following Formula (13):

$$fit = \alpha \cdot \gamma_B(D) + \beta \cdot \frac{|C| - |B|}{|C|} \quad (13)$$

Among them, $\alpha + \beta = 1$ and $\alpha, \beta \in [0, 1]$, $|C|$ is the total number of feature attributes of the data set, $|B|$ is the length of the currently selected feature subset, and $\gamma_B(D)$ is the attribute dependency. Equation (13) has been widely used in the research of various feature selection methods based on neighborhood rough set attribute dependence, and has achieved good results. However, the fitness function also has some shortcomings; one is that the attribute dependence in the fitness function is only calculated for the selected attributes, and the analysis and evaluation of other attributes in the attribute set are lacking [9]; the other is that the fitness function is only The feature is selected based on the attribute dependence and the number of attributes, ignoring the role of heuristic information that can speed up finding the reduced subset. For this reason, literature [9] introduces attribute importance based on information entropy as heuristic information, and combines attribute dependence to construct a new fitness function, as shown in the following Formula (14):

$$fit^2 = \lambda_1 \cdot \gamma_B(D) + \lambda_2 \cdot \frac{1}{1 + e^{sig(C-B)}} + \lambda_3 \cdot \frac{|C| - |B|}{|C|} \quad (14)$$

Among them, $\lambda_1, \lambda_2, \lambda_3$ are weighting factors, where $\lambda_1 + \lambda_2 + \lambda_3 = 1$ and $\lambda_1, \lambda_2, \lambda_3 \in [0, 1]$; $sig(C - B)$ is the attribute importance of the conditional feature set outside the selected feature subset B ;

However, the attribute importance of information entropy is a quantitative measure of the change in the uncertainty domain. The change of any attribute may cause the uncertainty domain to change. Therefore, the information entropy method improves the non-kernel attributes under the rough set algebraic view. When the importance is zero. However, in some decision tables, the information entropy method has the situation that the importance of redundant attributes is greater than the importance of core attributes [16]. The subset of attributes found by the bee colony may contain redundant attributes, resulting in selection accuracy. Decreased significantly.

Aiming at the shortcomings of the fitness function constructed by the conditional entropy attribute importance in the literature [9], this paper replaces the fitness function with the attribute importance of the discernibility matrix defined in Sect. 3.1. The new fitness function is as follows (15).

$$fitness' = \omega_1 \cdot \gamma_B(D) + \omega_2 \cdot \frac{1}{1 + e^{Nsig(c, C-B, D)}} + \omega_3 \cdot \frac{|C| - |B|}{|C|} \quad (15)$$

Among them, the weight factor $\omega_1, \omega_2, \omega_3$ satisfies the condition $\omega_1 + \omega_2 + \omega_3 = 1$ and $\omega_1, \omega_2, \omega_3 \in [0, 1]$; $Nsig(c, C - B, D)$ is the attribute importance

of the conditional feature set outside the selected feature subset B under the neighborhood rough set discernibility matrix. let $C = \{c_1, c_2, \dots, c_n\}$, $\{0, 1\}^n$ denote m -dimensional Boolean space, define $x = (x_1, x_2, \dots, x_n) \notin \{0, 1\}^n$.

3.3 Neighborhood Discernibility Matrix Importance and Artificial Bee Colony Feature Selection Algorithm

According to the fitness function constructed by the new neighborhood discernibility matrix attribute importance, we design an artificial bee colony feature search algorithm (ABCNRT algorithm) based on the discernibility matrix

Algorithm 1: ABCNRT algorithm

Input: Neighborhood decision system $NDS = (U, C \cup D, V, f)$, the population size of the bee colony is $nPop$, the number of iterations $iter$ and the location of the nectar source is Pop

Output: the best nectar source x_b in history

```

1 while  $U \neq \emptyset$  do
2   for each  $M_j \in U$  do
3     | Select the neighborhood for the decision table  $\delta$ ;
4     |  $POS_i \cup M_j \rightarrow POS_i$ 
5   end
6   The dependence degree  $\gamma_B(D)$  is calculated according to Equation (7);
7   The importance  $Nsig(c, C - B, D)$  is calculated according to Equation (12)
8 end
9 while  $iter < iter_{max}$  do
10  for each  $x_i \in nPop$  do
11    |  $\gamma_B(D)$  and
12    |  $Nsig(c, C - B, D) \rightarrow \lambda_1 \cdot \gamma_B(D) + \lambda_2 \cdot \frac{1}{1 + e^{Nsig(c, C - B, D)}} + \lambda_3 \cdot \frac{|C| - |B|}{|C|}$ ;
13  end
14  for each  $x_i \in C$  do
15    |  $L_i = (a * rand([-1, 1][1, x_b])) * (Pop[i] - Pop[k]) + (a * rand([1, x_b])) * (Pop[i] - x_b)$ ;
16    | Judging the value of  $L_i$  gives  $W_i$ ;
17    | Use formula (20) to calculate the value of the  $W_i$  fitness function;
18    | if  $f(W_i) \leq f(X_i)$  then
19      | Replace  $X_i$  with  $W_i$ 
20    | else
21      |  $trial = trial + 1$ 
22    | end
23  end
24  for each  $x_i \in nPop$  do
25    | Execute roulette algorithm, Repeat the above steps to hire bees;
26    | Go through the stage of detection bee;
27    | Output historical optimal solution  $x_b$ ;
28 end

```

attribute importance. The main idea is: First, obtain the discernibility matrix and calculate Resolve the importance of each attribute in the matrix; secondly, construct the fitness function; finally, iterate according to the direction of the fitness function value, until the search for the smallest feature set stops the algorithm. The specific steps of the ABCNRT algorithm are as follows. In the experiment part, the fitness function of the algorithm is used to conduct two sets of experiments with Eqs. (13) and (15).

Algorithm Complexity Analysis. The time complexity of the algorithm in this paper mainly depends on the number of bee colonies $nPop$, the variable dimension D , the number of conditional attributes M , and the number of universe objects $|U|$. Among them, the time complexity of calculating the discernibility matrix in the fitness function is $O(|M||U|^2)$, Suppose the sample is divided into k sets, so the time complexity of importance in the fitness function is $O\left((|M| + N)\frac{|U|^2}{k}\right)$. In addition to the neighborhood rough set calculation, the outer loop also has the number of bee colonies and variable dimensions. Therefore, in general, the time complexity of the ABCNRT algorithm is $O\left((|M| + N)\frac{|U|^2}{k} * k * D * nPop\right)$. Among them, it can be calculated in parallel in the process of calculating the degree of dependence, reducing the time complexity as $O\left((|M| + N)\frac{|U|^2}{k} * D * nPop\right)$.

4 Experiment Analysis

In order to verify the superiority and reliability of the proposed algorithm. We select 8 data sets from UCI machine learning library (Table 1) for experiments, and compare the algorithm results obtained from experiments with the neighborhood rough set NRS algorithm proposed in literature [3]. Feature selection FSRSWOA algorithm based on rough set and improved Whale optimization algorithm in literature [14] and feature selection PSORSFS algorithm based on rough set and particle swarm optimization algorithm in literature [15] were performed 20 times in the experiment. The software environment is Win 10 operating system, the programming software is Python3.6, the hardware is Intel I7 1040 3.6 Ghz, and the memory is 8 GB.

Table 1. Data set

Serial number	Data set	Number of samples	Number of attributes	Number of categories
1	Zoo	101	16	7
2	Wine	178	13	3
3	Sonar	208	60	2
4	Ionosphere	351	34	2
5	Heart	270	13	7
6	WDBC	569	30	2
7	German	1000	24	2
8	CMC	1473	9	2

4.1 Selection of δ

In order to remove the influence of dimensions on the data, the data set data is normalized first. Experiment on the data set again, and the experimental results are shown in Fig. 2.

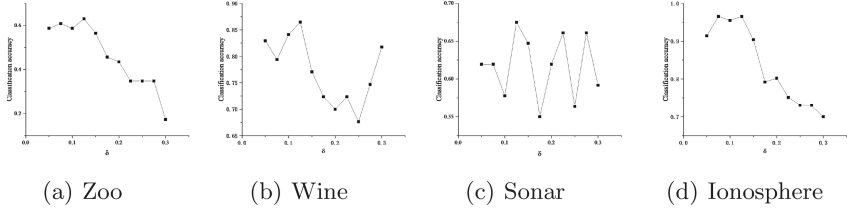


Fig. 2. Classification accuracy varies with threshold

In the experiment, due to the selected threshold δ , the classification accuracy of different data sets will be different. When calculating the classification accuracy, four data sets in the data set are selected, and the CART classifier is used to classify these four data sets to select the best δ value. The classification accuracy changes with the threshold δ were cross-validated and analyzed ten times.

It can be seen from Fig. 2 that the threshold is within the range of $[0.1, 0.15]$, and the classification accuracy of the data set is better. In this paper, $\delta = 0.125$ is selected for the experiment.

4.2 Algorithm Comparison Results

Compare this algorithm with NRS algorithm [3], FSRSWOA algorithm [14] and PSORSFS algorithm [15]. The value of ABCNRT algorithm is: initial colony size $N = 20$, the maximum number of iterations $M = 50$. When the values of the neighboring particles δ are all set to 0.125, three algorithms are used to perform attribute reduction on 8 data sets.

Original Fitness Function. The original fitness function $fit = \alpha \cdot \gamma_B(D) + \beta \cdot \frac{|C| - |B|}{|C|}$ is selected, and the minimum attribute reduction problem is transformed into the maximum fitness function value problem. Take $\alpha = 0.7$ and $\beta = 0.3$. First, use the ABCNRT algorithm to reduce the 8 UCI data. Table 2 shows the reduction of ABCNRT algorithm.

Table 2. ABCNRT algorithm reduction result

Serial number	Data set	Number of samples	Number of optimal attribute subsets	Optimal attribute subset	Reduction rate	Optimal classification accuracy
1	Zoo	16	3	c_5, c_7, c_9	81.25	0.7619
2	Wine	13	4	c_1, c_5, c_7, c_{10}	69.23	0.9444
3	Sonar	60	7	$c_{10}, c_{12}, c_{22}, c_{28}, c_{31}, c_{32}, c_{48}$	88.33	0.7380
4	Ionosphere	34	6	$c_4, c_5, c_7, c_{16}, c_{29}, c_{34}$	82.35	0.9295
5	Heart	13	9	$c_1, c_4, c_5, c_7, c_8, c_9, c_{11}, c_{12}, c_{13}$	30.77	0.6393
6	WDBC	30	4	$c_8, c_{21}, c_{22}, c_{27}$	86.67	0.9649
7	German	24	2	c_4, c_6	91.67	0.6950
8	CMC	9	2	c_4, c_7	77.78	0.4033

According to Table 2, 5 of the 8 UCI data sets have a reduction rate of more than 80%, and 2 of them have a reduction rate of about 70%. Only the Heart data set has a lower reduction rate of about 30%. There are more subsets of reductions. The first six data are compared with the initial classification accuracy of Fig. 1. According to Table 2, it can be seen that all data sets are almost better than the initial classification accuracy. It shows that the above table fully reflects the better reduction results achieved by ABCNRT algorithm reduction.

Table 3, 4 shows the comparative experimental results of PSORSFS algorithm, FSRSWOA algorithm, NRS algorithm, ABCNRT algorithm attribute reduction and the best classification accuracy.

According to Table 3 below, the attribute reduction number analysis of the four algorithms for the eight data sets shows that when comparing the NRS algorithm, it is found that except for the Heart data set, the other data ABCNRT algorithm is better than the NRS algorithm. When comparing the PSORSFS algorithm, except for Wine and Heart, the number of other attribute reductions is more than that of the ABCNRT algorithm. When comparing the FSRSWOA algorithm, except for the Heart, Wine and Ionosphere data sets, the number of reductions in some running results is less than that of the ABCNRT algorithm, and the number of other attribute reductions is more than that of the ABCNRT algorithm.

According to the following Table 4, the classification accuracy analysis of the eight data sets in the four algorithms shows that the classification accuracy of the Heart data set is better than the ABCNRT algorithm. The classification accuracy of the Zoo data set is better than the PSORSFS algorithm and the FSRSWOA algorithm. For the ABCNRT algorithm, the Sonar data set FSRSWOA algorithm is better than the ABCNRT algorithm, and the classification accuracy of the ABCNRT algorithm in other data sets is better than the other three algorithms.

Table 3. Attribute reduction

Data set	PSORSFS	FSRSWOA	NRS	ABCNRT
Zoo	8 ^[10] , 9 ^[10]	7 ^[5] , 8 ^[12] , 9 ^[3]	5	3^[15], 4^[5]
Wine	3 ^[5] , 4 ^[15] , 5 ^[5]	3^[12], 4^[8]	5	4
Sonar	11 ^[15] , 12 ^[5]	10 ^[13] , 11 ^[7]	16	7^[17], 8^[3]
Ionosphere	7 ^[8] , 8 ^[12]	5	9	5 ^[4] , 6 ^[16]
Heart	6 ^[5] , 7 ^[15]	7 ^[5] , 8 ^[15]	6	9
WDBC	10 ^[13] , 11 ^[7]	9 ^[12] , 10 ^[8]	6	4^[18], 5^[2]
German	5 ^[16] , 6 ^[4]	3 ^[13] , 4 ^[7]	9	2^[10], 3^[10]
CMC	5	6 ^[16] , 7 ^[4]	9	1^[8], 2^[12]

Table 4. Optimal classification accuracy

Data set	PSORSFS	FSRSWOA	NRS	ABCNRT
Zoo	0.8443	0.8743	0.7425	0.7619
Wine	0.8764	0.8967	0.8764	0.9444
Sonar	0.7234	0.7829	0.6782	0.7380
Ionosphere	0.9072	0.9156	0.8576	0.9295
Heart	0.6672	0.7023	0.6516	0.6393
WDBC	0.8976	0.9425	0.8796	0.9649
German	0.6950	0.6909	0.5672	0.6950
CMC	0.3782	0.3965	0.3928	0.4033

In summary, it can be proved that the ABCNRT algorithm has advantages.

Improve Fitness Function. The fitness function formula $\text{fitness}' = \omega_1 \cdot \gamma_B(D) + \omega_2 \cdot \frac{1}{1+e^{\text{Nsig}(c,C-B,D)}} + \omega_3 \cdot \frac{|C|-|B|}{|C|}$ improved in this paper is adopted, in which the problem of minimum attribute reduction is transformed into the problem of maximum fitness function value, and $\omega_1 = 0.6, \omega_2 = 0.3, \omega_3 = 0.1$ is taken. Table 4 shows the reduction situation obtained with the improved fitness function formula.

Table 5. ABCNRT algorithm reduction result

Serial number	Data set	Number of samples	Number of optimal attribute subsets	Optimal attribute subset	Reduction rate	Optimal classification accuracy
1	Zoo	16	3	c_5, c_7, c_9	81.25	0.7619
2	Wine	13	4	c_1, c_5, c_7, c_{10}	69.23	0.9444
3	Sonar	60	6	$c_{10}, c_{12}, c_{22}, c_{31}, c_{32}, c_{48}$	90.00	0.7619
4	Ionosphere	34	5	$c_4, c_7, c_{16}, c_{29}, c_{34}$	85.29	0.9437
5	Heart	13	6	$c_1, c_2, c_3, c_5, c_8, c_{10}$	53.84	0.7213
6	WDBC	30	4	$c_8, c_{21}, c_{22}, c_{27}$	86.67	0.9649
7	German	24	2	c_4, c_6	91.67	0.6950
8	CMC	9	1	c_4	88.89	0.4644

It can be seen from Table 5 that the reduction rate of the three data sets Sonar, Ionosphere, Heart, and CMC in the table has been significantly improved. Among them, the reduction rate of 6 data sets has reached more than 80%, and the reduction rate of the Heart data set The reduction rate has also increased to about 50%. In terms of classification accuracy, Sonar, Ionosphere, Heart, and CMC have significantly improved the classification accuracy with the original fitness function based on the improvement of the reduction rate.

This paper improves the predecessor’s fitness function formula, changes the predecessor’s use of information entropy to calculate the importance of the attribute, and calculates the importance of the attribute more accurately. It also provides a new way to calculate the fitness function. Experiments have also proved that the new fitness function is better than the original fitness function.

In the algorithm comparison, according to Table 6, it can be seen that in addition to the Wine data set, the number of attribute reductions in the FSRSWOA algorithm is slightly less than that of the ABCNRT algorithm. The Heart and Ionosphere data sets are partially equal to other algorithms, which can be clearly seen in other data sets. The ABCNRT algorithm is better than the compared algorithm. It can be seen from Table 7 that the optimal classification accuracy of Zoo and Sonar data sets is better than the ABCNRT algorithm in other algorithms, and the classification accuracy of the ABCNRT algorithm in the remaining data sets is better than the comparison algorithm.

Table 6. Attribute reduction

Data set	PSORSFS	FSRSWOA	NRS	ABCNRT
Zoo	8 ^[10] , 9 ^[10]	7 ^[5] , 8 ^[12] , 9 ^[3]	5	3 ^[15] , 4 ^[5]
Wine	3 ^[5] , 4 ^[15] , 5 ^[5]	3 ^[12] , 4 ^[8]	5	4
Sonar	9 ^[15] , 10 ^[5]	8 ^[13] , 9 ^[7]	16	6 ^[17] , 7 ^[3]
Ionosphere	6 ^[8] , 7 ^[12]	5 ^[13] , 6 ^[7]	9	5 ^[15] , 6 ^[5]
Heart	6 ^[15] , 7 ^[5]	6 ^[9] , 7 ^[5] , 8 ^[6]	6	6
WDBC	10 ^[13] , 11 ^[7]	9 ^[12] , 10 ^[8]	6	4 ^[18] , 5 ^[2]
German	5 ^[16] , 6 ^[4]	3 ^[13] , 4 ^[7]	9	2 ^[10] , 3 ^[10]
CMC	4 ^[14] , 5 ^[6]	5 ^[14] , 6 ^[3] , 7 ^[3]	9	1

Table 7. Optimal classification accuracy

Data set	PSORSFS	FSRSWOA	NRS	ABCNRT
Zoo	0.8447	0.8746	0.7425	0.7619
Wine	0.8764	0.8967	0.8764	0.9444
Sonar	0.7356	0.7889	0.6782	0.7619
Ionosphere	0.9192	0.9256	0.8576	0.9437
Heart	0.6972	0.7123	0.6516	0.7213
WDBC	0.8776	0.9425	0.8796	0.9649
German	0.6951	0.6920	0.5672	0.6951
CMC	0.3982	0.4173	0.3928	0.4644

In summary, it fully illustrates the advantages and reliability of the algorithm in this paper after comparing the other three algorithms.

4.3 Algorithm Performance Comparison

In order to further verify the advantages of the algorithm in this paper, the performance of the above-mentioned various algorithms is compared, and the experiment uses the methods of Friedman test and Nemenyi follow-up test to verify.

First, according to the classification accuracy in Table 4 and Table 7, sort and assign the classification accuracy of the 4 algorithms involved in the experiment on 8 data sets. Tables 8 and 9 show the classification accuracy of each algorithm in each data set. The ranking value of the test performance and its average value.

Table 8. The original function experimental algorithm performance ranking results

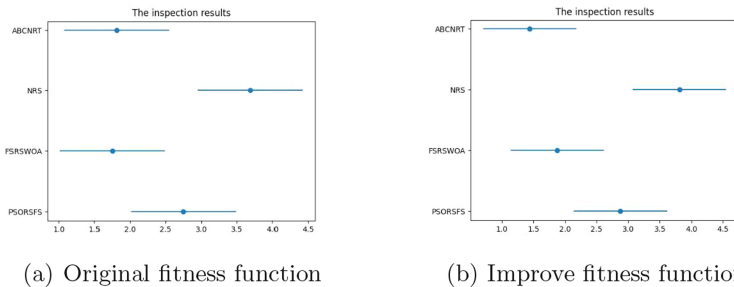
Data set	PSORSFS	FSRSWOA	NRS	ABCNRT
Zoo	2	1	4	3
Wine	3.5	2	3.5	1
Sonar	3	1	4	2
Ionosphere	3	2	4	1
Heart	2	1	3	4
WDBC	3	2	4	1
German	1.5	3	4	1
CMC	4	2	3	1
Average	2.75	1.75	3.6875	1.1825

Table 9. Improved function experiment algorithm performance ranking results

Data set	PSORSFS	FSRSWOA	NRS	ABCNRT
Zoo	2	1	4	3
Wine	3.5	2	3.5	1
Sonar	3	1	4	2
Ionosphere	3	2	4	1
Heart	3	2	4	1
WDBC	4	2	3	1
German	1.5	3	4	1.5
CMC	3	2	4	1
Average	2.875	1.875	3.8125	1.4375

Then use Friedman test to determine whether the above algorithms have the same performance. Through the above observations, there is no algorithm with the same performance. Suppose we compare n algorithms on M data sets, the commonly used variable $\tau_F = \frac{(M-1)\tau_X^2}{M(n-1)-\tau_X^2}$, where τ_F obeys the F distribution with $n-1$ and $(n-1)(M-1)$ degrees of freedom.

If the algorithm performance is not the same, use the Nemenyi follow-up test method, use the Nemenyi follow-up test to calculate the critical value range CD , $CD = q_\alpha \sqrt{\frac{n(n+1)}{6M}}$ of the average ordinal difference, where n is the number of algorithms and M is the number of data sets. When $n = 4$ and $\alpha = 0.1$ are given after looking up the table, the value of q_α is 2.291, so the critical value $CD = 1.4789$ is obtained.

**Fig. 3.** Inspection result graph

According to the result of the test (Fig. 3.), it can be seen that in the original fitness function test result graph, the performance of the FSRSWOA algorithm is slightly better than the ABCNRT algorithm, and the two are similar, but in the improved fitness function test result graph It can be clearly seen that the ABCNRT algorithm is better than the other three algorithms. It fully shows that

the performance of the attribute reduction algorithm in this paper has significant advantages compared with other algorithms.

In summary, through the above experiments, it can be seen that the ABC-NRT algorithm proposed in this paper has more advantages in selecting the optimal attribute reduction and obtaining the optimal classification accuracy. In the above experiments, the algorithm in this paper has selected a better attribute reduction and obtained a better classification accuracy, which shows the effectiveness of the algorithm in this paper.

5 Concluding Remarks

This paper proposes an attribute importance measurement method based on discrimination matrix, and introduces the attribute importance as heuristic information into the bee colony algorithm, constructs a new fitness function, and designs based on artificial bee colony and discrimination matrix. The attribute selection method of the optimization algorithm, this method avoids the redundant attributes in the searched attribute subset, reduces the number of iterations, and reduces the reduction time complexity. The experimental results of the UCI data set show that this method improves the reduction rate. After testing, it can be found that the attribute selection method in this paper has better performance and better classification accuracy. However, the algorithm in this paper also has shortcomings. First, it is unable to calculate the importance of attributes that do not appear in the neighborhood discrimination matrix; second, although the number of colony iterations is reduced, the importance of the neighborhood discrimination matrix must be calculated, which increases the amount of calculation. To further reduce the computational workload and find the optimal attribute subset is the method for future research.

References

1. Lin, T.Y.: Granular computing on binary relations I: data mining and neighborhood systems. *Rough Sets Knowl. Disc.* **18**(1), 107–121 (1998)
2. Hu, Q.H., Yu, D.R., Liu, J., et al.: Neighborhood rough set based heterogeneous feature subset selection. *Inf. Sci.* **178**(18), 3577–3594 (2008)
3. Liu, Y., Huang, W., Jiang, Y., et al.: Quick attribute reduct algorithm for neighborhood rough set model. *Inf. Sci.* **271**(7), 65–81 (2014)
4. Karaboga, D., Basturk, B.: A comparative study of artificial bee colony algorithm. *Appl. Math. Comput.* **214**(1), 108–132 (2009)
5. Tao, W., Fan, H.L.: Attribute reduction of rough set based on bee colony optimization. *Comput. Meas. Control* **20**(01), 193–195 (2012)
6. Wang, S.Q., et al.: Attribute reduction based on fuzzy rough set and bee colony algorithm. *J. Central South Univ. (Sci. Technol.)* **44**(01), 172–178 (2013)
7. Ye, D.Y., Chen, Z.J.: An efficient combined artificial bee colony algorithm for minimum attribute reduction problem. *Acta Electronica Sinica* **43**(05), 1014–1020 (2015)

8. Peng, X.R., Liu, Z.R., Gao, Y.: Artificial bee colony decision table reduction algorithm based on neighborhood rough model. *J. Qingdao Univ. (Nat. Sci.)* **30**(03), 55–59 (2017)
9. Fang, B., Chen, H.M., Wang, S.W.: Feature selection method based on rough set and fruit fly optimization algorithm. *Comput. Sci.* **46**(07), 157–164 (2019)
10. Gao, W.F., Huang, L.L., Wang, J., et al.: Enhanced artificial bee colony algorithm through differential evolution. *Appl. Soft Comput.* **48**, 137–150 (2016)
11. Wang, G.Y., Yao, Y.Y., Yu, H.: *Chin. J. Comput.* **32**(07), 1229–1246 (2009)
12. Wang, H., Wu, Z.J., Rahnamayan, S., et al.: Improving artificial Bee colony algorithm using a new neighborhood selection mechanism. *Inf. Sci.* **527**, 227–240 (2020)
13. Zhou, X.Y., Lu, J.X., Huang, J.H., Zhong, M.S., Wang, M.W.: Enhancing artificial bee colony algorithm with multi-elite guidance. *Inf. Sci.* **543**(prepublish), 242–258 (2021)
14. Wang, S.W., Chen, H.M.: Feature selection method based on rough set and improved whale optimization algorithm. *Comput. Sci.* **47**(02), 44–50 (2020)
15. Wang, X.Y., Yang, J., Teng, X.L., Xia, W.J., Jensen, R.: Feature selection based on rough sets and particle swarm optimization. *Patt. Recogn. Lett.* **28**(4), 459–471 (2006)
16. Ye, J., Wang, L.: Approach of ascertaining combinatorial attribute weight based on discernibility matrix. *Comput. Sci.* **41**(11), 273–277 (2014)
17. Cui, Z.H., Cao, Y., Cai, X.J., Cai, J.H., Chen, J.J.: Optimal LEACH protocol with modified bat algorithm for big data sensing systems in internet of things. *J. Parallel. Distrib. Comput.* **132**, 217–229 (2019)
18. Cui, Z.H., Chang, Y., Zhang, J.J., Cai, X.J., Zhang, W.S.: Improved NSGA-III with selection-and-elimination operator. *Swarm. Evol. Comput.* **49**, 23–33 (2019)