International Journal of Computational Intelligence Systems
Vol. **14(1)**, 2021, *pp.* 1419–1425
DOI: https://doi.org/10.2991/ijcis.d.210419.002; ISSN: 1875-6891; eISSN: 1875-6883
https://www.atlantis-press.com/journals/ijcis/

Research Article

Attribute Reduction Method of Covering Rough Set Based on Dependence Degree

Li Fachao[1,2], Ren Yexing[2,*], Jin Chenxia[1]

[1]*School of Economics and Management, Hebei University of Science and Technology, Shijiazhuang, 050018, China*
[2]*College of Science, Hebei University of Science and Technology, Shijiazhuang, 050018, China*

ARTICLE INFO

Article History
Received 10 Jan 2021
Accepted 13 Apr 2021

Keywords
Covering rough sets
Attribute reduction
Dependence degree
Local dependence degree
ε-Boolean identification matrix

ABSTRACT

Attribute reduction is a hot topic in the field of data mining. Compared with the traditional methods, the attribute reduction algorithm based on covering rough set is more suitable for dealing with numerical data. However, this kind of algorithm is still not efficient enough to deal with large-scale data. In this paper, we firstly propose ε-Boolean identification matrix of covering rough sets, give the calculation methods of dependence degree and local dependence degree, and further discuss their properties. Secondly, we give two attribute reduction algorithms based on dependence degree and local dependence degree, respectively. Finally, we test the performance of the algorithm through several UCI data sets. Experimental results show that the efficiency of our algorithm has been greatly improved. So it is more suitable for handling large-scale data processing problems, and can have wide application value.

1. INTRODUCTION

Rough set theory was proposed by Polish mathematician Pawlak Z in 1982. It is a new mathematical tool to deal with fuzzy and uncertain knowledge, and has been successfully applied in machine learning and knowledge discovery, data mining, pattern recognition, and other fields [1,2]. And attribute reduction (removal of redundant attributes while keeping certain performance of the information system unchanged) is one of the important contents in rough set theory.

Classical rough set theory uses equivalence relations to obtain the division of objects under attribute sets. It is an effective tool for processing complete discrete data. However, in reality, due to the complexity of the problem, the data in the information system is often numerical data, which means that the numerical data needs to be discretized before attribute reduction, which will cause information loss [3]. Since the equivalence relation is no longer suitable for numerical information systems, many scholars extend the equivalence relation to the dominant relation, or change the division of the universe into the coverage of the universe, so as to deal with the problem of attribute reduction and rule acquisition of information system. Since most of the attribute reductions of information systems are not unique, and seeking the reduction with the least number of attributes is an NP-hard problem [4]. Therefore, how to realize attribute reduction through a certain heuristic algorithm is a hot research content, and many scholars have conducted many useful discussions.

Guan and Wang [5] proposed the concept of relative reduction of the maximum tolerance class, and defined a discriminant function to calculate the relative reduction using Boolean reasoning technology. Ma and Zhang [6] gave the form of generalized binary discernibility matrix for processing incomplete information systems, introduces some useful properties of attribute kernel and attribute relative reduction, and proposes an attribute kernel based on generalized binary discernibility matrix Algorithm for relative reduction of attributes. Meng and Shi [7] derived some properties of incomplete information systems, and proposed an attribute reduction algorithm based on positive regions. Chen *et al.* [8] proposed a method to reduce the attributes of the coverage decision system. The consistent and inconsistent coverage decision-making systems and their attribute reductions are defined, and the necessary and sufficient conditions for reduction are given. Wang *et al.* [9] gave some basic structural properties of attribute reductions in covering rough sets, and proposes a heuristic algorithm based on discernibility matrix to find the approximate minimum reduction attribute subsets. Gao *et al.* [10] rewrote the matrix form of neighborhood approximation set. Based on the neighborhood approximation set of matrix, the relative dependence of attributes and its algorithm are proposed. Chen *et al.* [11] proposed an acceleration strategy based on attribute groups. First, divide all candidate attributes into different groups. In the process of searching for reduction, it is only necessary to calculate the attributes in the group that contains at least one attribute in the potential reduction. Tsang *et al.* [12,13] improved the upper approximation definition of covering generalized rough sets, then proposed a new method

*Corresponding author. Email: x960312@yeah.net

of constructing a simpler discernibility matrix based on covering rough sets. Liu *et al.* [14] proposed two new matrix-based minimum and maximum description calculation methods for covering rough sets, which reduces the computational complexity of traditional methods. Li *et al.* [15] proposed a single-attribute identification matrix, gave the concept of dual-attribute comprehensive dependence, and designed an attribute reduction algorithm based on dual-attribute comprehensive dependence. Shi *et al.* [16] proposed a new attribute reduction model based on Boolean operations based on the concept of neighborhood rough set. Chen *et al.* [17] studied the attribute reduction problem of the coverage decision system based on graph theory, and proposed an algorithm for the coverage decision system. Chen and Chen [18] proposed a variable precision neighborhood rough set model, and a feature subset selection algorithm to the variable precision neighborhood rough sets is designed.

The above mentioned researches have promoted the development of rough set theory to a certain extent, but it is worth noting that the resolution matrix has the problems of large storage space and low generation efficiency. Therefore, in the context of covering rough set, this paper first proposes the ε-Boolean identification matrix, according to which the attribute reduction algorithm based on dependence degree is designed, which solves the above problems to a certain extent and greatly improves the operation efficiency of the algorithm. Secondly, the sub matrix of ε-Boolean identification matrix is used to replace ε-Boolean identification matrix, and an attribute reduction algorithm based on local dependence degree is designed, which further improves the performance of the algorithm.

The main work of this paper is stated as follows: (1) The basic concepts of rough set correlation are reviewed; (2) The calculation methods of dependence degree and local dependence degree under covering rough set are given, and the concept of ε-Boolean identification matrix is proposed; (3) The algorithm of attribute reduction based on dependency and local dependency is designed respectively; (4) The effectiveness of this algorithm is verified by several commonly used UCI datasets. Theoretical analysis and simulation results show that our methods can effectively improve the efficiency of the algorithm, and the reduction results are similar to the existing algorithms.

2. PRELIMINARIES

This section mainly reviews some basic concepts related to rough set. For convenience, we assume that (1) (U, F, A) is an information system, (where $U = \{x_1, x_2, \cdots, x_n\}$ is object set, $A = \{a_1, a_2, \cdots, a_m\}$ is a set of attributes, $F = \{f_a : U \to V_a | a \in A\}$ is the set of relationships between U and A, V_a is the range of a); (2) U is a nonempty finite universe, C is a family of nonempty subsets of U. C is called a covering of U if no subset in C is empty and $\cup C = U$.

Definition 1. [12] Suppose U is a finite universe and $C = \{k_1, k_2, \cdots k_n\}$ is a covering of U. For any $x \in U$, let $C_x = \cap\{k_j : k_j \in C, x \in k_j\}$, then $Cov(C) = \{C_x : x \in U\}$ is also a covering of U, we call it the induced covering of C.

Definition 2. [12] Suppose U is a finite universe and $\Delta = \{C_i : i = 1, 2, \cdots, m\}$ is a family of covering of U. For any $x \in U$, let $\Delta_x = \cap\{(C_i)_x : (C_i)_x \in Cov(C_i), i = 1, 2, \cdots, m\}$, then $Cov(\Delta) = \{\Delta_x : x \in U\}$ is also a covering of U, we call it the induced covering of Δ.

Generally speaking, for each numerical attribute $a \in A$, we can define a neighborhood of each sample $x \in U$. $N_a(x, \varepsilon) = \{y \in U : d(x, y) \le \varepsilon\}$, $N_A(x, \varepsilon) = \cap\{N_a(x, \varepsilon) : a \in A\}$, where $d(x, y) = |a(x) - a(y)|$ is a distance function and ε is a specified threshold. Obviously, $N_a = \{N_a(x, \varepsilon) : x \in U\}$ is the coverage of U, $\Delta = \{N_a : a \in A\}$ is the coverage family of U, and $\Delta_x = N_A(x, \varepsilon)$.

Theorem 1. *[12] Let $\Delta = \{C_i : i = 1, 2, \cdots, m\}$ be a family of coverings on U, $P \subseteq \Delta$. Then $Cov(P) = Cov(\Delta)$ if and only if $\Delta_x = P_x$ for all $x \in U$.*

Definition 3. Let (U, F, A) be an information system, $a_i \in A$, $x, y \in U$, B is a nonempty subset of A and ε is a specified threshold,

$$d(s, t) = \begin{cases} 1, & \text{if } |s - t| > \varepsilon, \\ 0, & \text{if } |s - t| \le \varepsilon. \end{cases} \tag{1}$$

1) If $d\left(f_{a_i}(x), f_{a_i}(y)\right) = 1$, then a_i can distinguish x from y; otherwise, a_i cannot distinguish x from y.

2) Let $a_t(x, y) = d\left(f^*_{a_t}(x), f^*_{a_t}(y)\right)$

$$G(a_i) = \sum_{x,y \in U} a_i(x, y) \tag{2}$$

$$G(B) = \sum_{x,y \in U} \max_{a_i \in B} a_i(x, y) \tag{3}$$

$$D\left(a_i|B\right)_x = \sum_{y \in U} a_i(x, y) \cdot d\left(\max_{a_k \in B} a_k(x, y), a_i(x, y)\right) \tag{4}$$

$$D\left(a_i|B\right) = \sum_{x,y \in U} a_i(x, y) \cdot d\left(\max_{a_k \in B} a_k(x, y), a_i(x, y)\right) \tag{5}$$

Then $G(a_i)$ is called the distinguishing ability of a_i, $G(B)$ is called the comprehensive distinguishing ability of attribute set B, $D\left(a_i|B\right)_x$ is called the local dependence degree of B on a_i and $D\left(a_i|B\right)$ is called the dependence degree of B on a_i.

Obviously, 1) $G(a_i)$ is the total number of order pairs that can be distinguished by a_i, and $G(B)$ is the total number of ordered pairs that can be distinguished by attribute set B; 2) $G(B_1) \le G(B_2)$ and $D\left(a_i|B_1\right) \ge D\left(a_i|B_2\right)$ are consistent with any $B_1 \subset B_2 \subset A$; 3) $D\left(a_i|B\right)_x$ represents the total number of objects that cannot be distinguished by B but can be distinguished by a_i when distinguishing object x from other objects, $D\left(a_i|B\right)$ is the total number of ordered pairs that cannot be distinguished by B but can be distinguished by a_i. This shows that $D\left(a_i|B\right)_x$ and $D\left(a_i|B\right)$ are important (complementary) measures to B in the sense of improving distinguishing ability, and are a selection basis for adding attributes when designing additive attribute reduction methods.

3. DEPENDENCE DEGREE ON COVERING ROUGH SETS

For convenience, in the following, we assume that,

1) $\delta(x) = 1$ for any $x > 0$, and $\delta(x) = 0$ for any $x \leq 0$.

2) For $X = (x_1, x_2, \cdots, x_n)$, $Y = (y_1, y_2, \cdots, y_n)$, $X_i = (x_{i1}, x_{i2}, \cdots, x_{in})$, $i = 1, 2, \cdots, s$, $S(X) = \sum_{i=1}^{n} x_i$ $S(X) = \sum_{i=1}^{n} x_i$, $\delta(X) = (\delta(x_1), \delta(x_2), \cdots, \delta(x_n))$, $X - Y = (x_1 - y_1, x_2 - y_2, \cdots, x_n - y_n)$, $\vee_{i\in I} X_i = (\vee_{i\in I} x_{i1}, \vee_{i\in I} x_{i2}, \cdots, \vee_{i\in I} x_{in})$.

3) If the first row of matrix Q is $(a_1, a_2, \cdots, a_m)$, and the elements of other rows are 0 or 1, then Q is called the identity Boolean matrix of $a_1, a_2, \cdots, a_m$, denoted as

$$Q = \begin{pmatrix} a_1 & a_2 & \cdots & a_m \\ P(Q, a_1) & P(Q, a_2) & \cdots & P(Q, a_m) \end{pmatrix}.$$

3.1. ε-Boolean Identification Matrix and Its Submatrix

Definition 4. Let (U, F, A) be an information system, $A = \{a_1, a_2, \cdots a_m\}$, $U = \{x_1, x_2, \cdots, x_n\}$,

$$r_{ij}^{(k)} = \begin{cases} 1, & \text{if } |f_{a_j}(x_k) - f_{a_j}(x_i)| > \varepsilon, \\ 0, & \text{if } |f_{a_j}(x_k) - f_{a_j}(x_i)| \leq \varepsilon, \end{cases} \tag{6}$$

$M(x_k) = \left(r_{ij}^{(k)}\right)_{(N-k)\times m}$ is a matrix with $r_{ij}^{(k)}$ as its element, where $i = k+1, k+2, \cdots, N, j = 1, 2, \cdots, m$. $\overrightarrow{A} = (a_1, a_2, \cdots, a_m)$,

1) The following $[n(n-1)/2 + 1] \times m$ order matrix:

$$\begin{aligned} M &= \left[\overrightarrow{A}\ M(x_1)\ M(x_2)\ \cdots\ M(x_{n-1})\right]^T \\ &\triangleq \begin{bmatrix} a_1 & a_2 & \cdots & a_m \\ P(M, a_1) & P(M, a_2) & \cdots & P(M, a_m) \end{bmatrix} \end{aligned} \tag{7}$$

is the ε-Boolean identification matrix of information system (U, F, A).

2) The following $[(n-k)+1] \times m$ order matrix:

$$\begin{aligned} M_{x_k} &= \left[\overrightarrow{A}\ M(x_k)\right]^T \\ &\triangleq \begin{bmatrix} a_1 & a_2 & \cdots & a_m \\ P(M_{x_k}, a_1) & P(M_{x_k}, a_2) & \cdots & P(M_{x_k}, a_m) \end{bmatrix} \end{aligned} \tag{8}$$

is the submatrix of ε-Boolean identification matrix of information system (U, F, A).

It is easy to see that (1) $r_{ij}^{(k)}$ is a 0-1 description about whether attribute $a_j (j = 1, 2, \cdots, m)$ can distinguish objects x_k and $x_i (i = k+1, k+2, \cdots, N)$; (2) M is a formal comprehensive description of the discriminative performance of attribute set A with respect to $U = \{x_1, x_2, \cdots x_n\}$, and M_{x_k} is a 0-1 comprehensive description about whether attribute set A can distinguish objects x_k and $x_i (i = k+1, k+2, \cdots, N)$; (3) In row $t (t = 1, 2, \cdots, N-k)$ of $M(x_k)$, the attribute corresponding to element 1 can distinguish x_k and x_{k+t}, while the attribute corresponding to element 0 cannot distinguish x_k and x_{k+t}. This indicates that the number of element 1 in this row is the number of attributes in A that can distinguish x_k and x_{k+t} (in particular, if there is only one 1 in the row, then the attribute in the column where this 1 is located must be the core attribute); (4) The *j-th* column of M is a comprehensive description of the distinguishing performance of attribute a_j with respect to $U = \{x_1, x_2, \cdots x_n\}$, and its basic characteristics are as Theorem 2.

Theorem 2. *Let (U, F, A) be an information system, $A = \{a_1, a_2, \cdots a_m\}$, $U = \{x_1, x_2, \cdots, x_n\}$, M is the ε-Boolean identification matrix, M_{x_k} is the submatrix of the ε-Boolean identification matrix. For $a_i \in A$ and $B \subseteq A$, the following statements hold.*

1) $G(a_i) = 2 \cdot S\left(P(M, a_i)\right) = 2 \cdot \sum_{k=1}^{N-1} S\left(P(M_{x_k}, a_i)\right)$;

2) $G(B) = 2 \cdot S\left(\vee_{a_t\in B} P(M, a_t)\right) = 2 \cdot \sum_{k=1}^{N-1} S\left(\vee_{a_t\in B} P\left(M_{x_k}, a_t\right)\right)$;

3) $D\left(a_j|B\right)_{x_k} = 2 \cdot S\left(\delta\left(P\left(M_{x_k}, a_j\right) - \vee_{a_t\in B} P\left(M_{x_k}, a_t\right)\right)\right)$;

4) $D\left(a_i|B\right) = 2 \cdot S\left(\delta\left(P(M, a_i) - \vee_{a_t\in B} P(M, a_t)\right)\right)$.

Proof. It is obvious by Definition 3.

Theorem 3. *Let (U, F, A) be an information system. For $x_k, x_{k+t} \in U$ and $B \subseteq A$, the following statements hold.*

1) *$x_{k+t} \in N_A(x_k, \varepsilon)$ if and only if the t-th row of matrix $M(x_k)$ is row 0;*

2) *$Cov(P) = Cov(\Delta)$ if and only if the $N_A(x, \varepsilon) = N_B(x, \varepsilon)$ is always true for any $x \in U$.*

Proof.

1) Necessity. If $x_{k+1} \in N_A(x_k, \varepsilon)$, then $x_{k+t} \in N_a\left(x_k, \varepsilon\right)$ holds for any $a \in A$. Next, we will prove that the row t of matrix M is row 0 by contradiction. In fact, if the element of a column in row t of the matrix is not 0 (it may be assumed that the element of the first column in row t of the matrix is not 0), then from Definitions 3 and 4, we can see that $|f_{a_1}(x_k) - f_{a_1}(x_{k+t})| > \varepsilon$, that is, $x_{k+t} \notin N_1(x_k, \varepsilon)$. This is in contradiction with the above condition $x_{k+t} \in N_a(x_k, \varepsilon)$.
Sufficiency. Similarly.

2) It can be proved by Theorem 1.

3.2. Properties and Calculation Methods of Dependence Degree and Local Dependence Degree

Suppose that (1) $M \ominus \{a_k\}$ is the new matrix formed by deleting the row in which the column element of attribute a_k is 1 and deleting the column in which a_k is located in matrix M; (2) Let $B \subset A$, $M \ominus B$ be the matrix formed by deleting the row corresponding to element 1 in $\vee_{a_t \in B} P(M, a_t)$ and the column corresponding to attribute in B in matrix M; (3) $M \ominus \phi = M$. According to the above convention, it is not difficult to see the following rules of operation: (1) Commutative law ($\left(\text{i.e.}, M \ominus \{a_i, a_j\} = M \ominus \{a_j, a_i\}\right)$; (2) Recurrence law $\left(\text{i.e.}, M \ominus \{a_i, a_j\} = M \ominus \{a_i\}\right) \ominus \{a_j\}$, $M \ominus \{a_i, a_j, a_k\} = \left(M \ominus \{a_i, a_j\}\right) \ominus \{a_k\}\right)$.

Using the above convention and Definition 3, we can see that for the nonempty subset of A, there are

$$D(a_i|B)_{x_k} = 2 \cdot S(P(M_{x_k} \ominus B, a_i)), \tag{9}$$

$$D(a_i|B) = 2 \cdot S(P(M \ominus B, a_i)). \tag{10}$$

It can be seen that the degree of dependence $D\left(a_i|B\right)$ is an extension of the degree of local dependence $D\left(a_i|B\right)_x$. In addition, the dependence degree can also be extended to the multi-attribute comprehensive dependence degree. For the nonempty subsets B and C of A, there is

$$D\left(C|B\right) = 2 \cdot S\left(\vee_{a_u \in C} P(M \ominus B, a_u)\right). \tag{11}$$

In the additive attribute reduction algorithm, adding the attributes with the largest local dependence degree or dependence degree of attribute set B in sequence can improve the comprehensive distinguishing ability of attribute set B faster. (9) and (10) are progressive attribute importance calculation methods for attribute addition process. The properties of dependence degree are given below.

Theorem 4. *Let (U, F, A) be an information system, $B, C \subseteq A$. Then the following statements hold:*

1) $D\left(a_j|B\right) = 2 \cdot \sum_{k=1}^{N-1} D\left(a_j|B\right)_{x_k}$;

2) $G\left(B \cup C\right) = G(B) + D(C|B)$.

Proof. It can be proved by Definition 3 and Theorem 2 .

4. ATTRIBUTE REDUCTION ALGORITHM BASED ON DEPENDENCE AND LOCAL DEPENDENCE

Since the above analysis shows that dependence degree and local dependence degree can be used as a selection basis for attribute addition, this section will design a corresponding heuristic algorithm to obtain the reduction of the information system. The detailed process is stated as follows:

Algorithm 1: Attribute reduction algorithm based on dependence degree

Input: Information system $S = (U, F, A)$, where U is the object set and A is the attribute set;

Output: A reduction B of the information system;

Step 1: Construct the corresponding ε-Boolean identification matrix M according to the information system S, and let $B = \phi$;

Step 2: First, delete the 0 and 1 rows in the ε-Boolean identification matrix M;

Step 3: Calculate the sum of the elements of each row in $M \ominus B$. If there is a row with the sum of 1, add the attribute a corresponding to 1 in the row to B, update B to $B \cup \{a\}$, and update $M \ominus B$ at the same time; otherwise, go to step 5;

Step 4: Calculate the sum of elements in each column in $M \ominus B$, select the corresponding attribute a with the largest result to add to B, update B to $B \cup \{a\}$, and update $M \ominus B$ at the same time. Turn to step 5;

Step 5: If $M \ominus B$ is not a row vector, turn to step 4; otherwise, output B.

The basic idea of this algorithm is to first add the core attributes to the reduction set B and update the matrix $M \ominus B$. Then add the attributes with the greatest dependence degree on the reduction set B in turn until the comprehensive distinguishing ability of the reduction set B is equal to that of the original attribute set to get the reduction. It is worth noting that although ε-Boolean identification matrix solves the shortcomings of large storage space and low generation efficiency of discernible matrix to a certain extent, but ε-Boolean identification matrix also has the problem of high computational complexity, so this paper designs Algorithm 2, which can further improve the operation efficiency of the algorithm. The implementation process is as follows:

Algorithm 2: Attribute reduction algorithm based on local dependence degree

Input: Information system $S = (U, F, A)$, where U is the object set and A is the attribute set;

Output: A reduction B of the information system;

Step 1: Let $k = 1$, $B = \phi$;

Step 2: Calculate the submatrix M_{x_k} of ε-Boolean identification matrix and $M_{x_k} \ominus B$;

Step 3: Delete the 0 and 1 rows in the submatrix of the ε-Boolean identification matrix M_{x_k};

Step 4: Calculate the sum of the elements of each row in $M_{x_k} \ominus B$. If there is a row with the sum of 1, add the attribute a_i corresponding to 1 in the row to B, update B to $B \cup \{a_i\}$, and update $M_{x_k} \ominus B$ at the same time; otherwise, go to step 6;

Step 5: Calculate the sum of elements in each column in $M_{x_k} \ominus B$, select the corresponding attribute a_i with the largest result to add to B, update B to $B \cup \{a_i\}$, and update $M_{x_k} \ominus B$ at the same time.

Step 6: If the matrix $M_{x_k} \ominus B$ is not a row vector, go to step 5; otherwise go to step 7;

Step 7: Let $k = k + 1$, if $k = n$, output B; otherwise go to step 2.

The basic idea of the algorithm is as follows: firstly, according to the submatrix M_{x_1} of the ε-Boolean identification matrix, the attribute with the largest local dependence degree of the reduction set is added in turn to distinguish the first object from other objects. Similarly, according to the submatrix M_{x_2}, the second object is separated from other objects (at this time, the first object does not need

to be considered) and the above process is repeated until the last object is separated from other objects. Because the calculation of local dependence degree is simpler, the attribute reduction algorithm based on local dependence degree is more efficient.

The computational complexity of Algorithms 1 and 2 is analyzed as follows: the time complexity of calculating ε-Boolean identification matrix is $O\left(n^2 \times |A|\right)$, so the time complexity and space complexity of Algorithm 1 are $O\left(n^2 \times |A|\right)$; The time complexity of calculating the submatrix of ε-Boolean identification matrix is $O(n \times |A|)$. Because the algorithm needs to cycle $n-1$, the time complexity of Algorithm 3 is $O\left(n^2 \times |A|\right)$ and the space complexity is $O(n \times |A|)$. Therefore, Algorithm 2 reduces the storage space, which will greatly improve the efficiency of the algorithm. But in the reduction results, the number of attributes in Algorithm 2 may be more than that in Algorithm 1.

5. SIMULATION AND COMPARATIVE ANALYSIS

In this section, 11 commonly used datasets in UCI database (see Table 1) are used to compare the reduction performance of algorithms 1 and 2 with CVR algorithm proposed in reference [12] and CDG algorithm proposed in reference [15] (since the algorithm in this paper is designed for information system, only conditional attribute set is selected for processing when the data set is a decision table), the results are shown in Tables 2 and 3 (where, |U| represents the number of examples, |A| represents the number of attributes in the reduction set, T represents the average running time (seconds), and * indicates that it exceeds the computer memory).

The information of the data set is given in Table 1. Before reduction, each numeric or integer attribute is normalized into interval [0, 1]. The reduction results are as follows:

CVR and CDG algorithms have important applications in covering rough sets. According to the attribute reduction results of 11 examples in Table 2, the reduction results of Algorithms 1 and 2 are similar to those of the above two algorithms. It is not difficult to see that the reduction result is related to the value of ε. With the increase of ε, the number of attributes may not decrease after reduction, because the order pairs that can be distinguished may become indistinguishable with the increase of ε, which may lead to the increase of core attributes and the number of attributes after reduction. But Algorithms 1 and 2 greatly reduce the running time of the algorithm (see Table 3). This is because compared with the identification matrix in CVR algorithm, the elements of ε-Boolean identification matrix in Algorithms 1 and 2 are 0 or 1, which reduces the storage space of the algorithm to a certain extent. CDG algorithm needs to compare $n-1$ times when distinguishing objects, while Algorithms 1 and 2 only need to compare $n-k$ times when distinguishing the k-th object from other objects. The running time of the four algorithms is given as follows:

It can be seen from the results in Table 3 that Algorithms 1 and 2 greatly reduce the running time of the algorithm compared with algorithm CVR and algorithm CDG. But Algorithm 1 is still not efficient in dealing with some large data sets, so this paper further improves Algorithm 1. The submatrix of ε-Boolean identification matrix is used to replace ε-Boolean identification matrix, and the local dependence degree is used as a selection basis for attribute addition. It can be seen from the Table that Algorithm 2 further reduces the running time of the algorithm and improves the efficiency of the algorithm. In Example 1, Algorithm 1 exceeds memory because Algorithm 1 needs to create an array of 25880415×22 (4.2GB), which exceeds the preset maximum array size.

In a word, from the data in Tables 2 and 3, it can be seen that the reduction results of Algorithms 1 and 2 proposed in this paper are similar to those of the other two algorithms (CVR, CDG), and the algorithm proposed in this paper has higher efficiency. Compared with Algorithm 1, Algorithm 2 is more efficient and more suitable for processing large-scale data sets. The experimental results are consistent with the theoretical analysis, which verifies the effectiveness of the proposed algorithm.

6. CONCLUSION

Covering rough set is a generalization of classic rough set. In the context of covering rough sets, this paper designs attribute reduction algorithms based on dependency degree and local dependency degree respectively. The algorithm proposed in this paper has the following advantages:

First, the reduction results of Algorithms 1 and 2 are comparable to those of existing algorithms (CVR algorithm, CDG algorithm). But Algorithms 1 and 2 greatly improve the operation efficiency of the algorithm, and are more suitable for processing large-scale data sets.

Table 1 Data information.

No	Data Sets	|U|	|A|
1	Anuran calls (MFCC)	7195	22
2	HCV for Egyptian patients	1385	29
3	Zoo	101	17
4	MEU-Mobile KSD	2856	71
5	Abalone	4177	8
6	Yeast	1484	8
7	Image segmentation	2310	19
8	Wine	178	13
9	Ionosphere	351	34
10	Glass	214	10
11	Audit data	777	18

Table 2 Reduction results.

No	|A|	Algorithm 1 (ε)			Algorithm 2 (ε)			CVR (ε)			CDG (ε)		
		0	0.1	0.2	0	0.1	0.2	0	0.1	0.2	0	0.1	0.2
1	22				2	22	22						
2	29	2	6	4	2	6	4	2	6	4	2	6	4
3	17	11	11	1	11	11	1	11	11	1	11	11	1
4	71	2	17	8	2	20	10				2	17	8
5	8	3	8	5	4	8	5				3	8	5
6	8	5	8	8	4	8	8	4	5	6	5	8	8
7	19	3	5	2	3	5	2	3	5	2	5	5	2
8	13	2	1	1	2	2	1	2	1	1	2	1	1
9	34	7	31	32	7	31	32	7	6	7	7	31	32
10	10	2	1	ϕ	2	1	ϕ	2	ϕ	1	2	ϕ	1
11	18	5	3	2	5	3	3	5	4	3	5	3	2

Table 3 Running time of the reduction with the four algorithms (in seconds).

No	|A|	T			
		Algorithm 1	Algorithm 2	CVR	CDG
1	22	*	25.93	*	*
2	29	1.64	0.74	496.115	3.29
3	17	0.03	0.06	0.03	1.74
4	71	167.4	22.15	*	479.39
5	8	13.03	4.96	*	42.42
6	8	0.52	0.74	334.97	2.38
7	19	3.21	1.91	1575.52	1.46
8	13	0.01	0.19	9.04	0.04
9	34	0.20	0.36	93.22	0.28
10	10	0.04	0.10	8.56	0.04
11	18	0.39	0.31	190.88	0.74

Second, Algorithm 2 further reduces the computational complexity of Algorithm 1 and is more efficient in dealing with large-scale data sets, but the number of attributes in the reduction may be more than other algorithms.

Third, the algorithm proposed in this paper is also suitable for attribute reduction of decision table, and has strong application value in data processing, data mining and other fields.

It is worth noting that this paper does not consider the application of the algorithm in the field of big data, so the next step is to reduce the sample before the attribute reduction of information system, or use the sampling method for multiple reduction to improve the efficiency of the algorithm.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare.

AUTHORS' CONTRIBUTIONS

Li Fachao: Visualization; Ren Yexing: Writing - original draft; Jin Chenxia: review & editing.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (71771078, 71371064).

REFERENCES

[1] W.X. Zhang, G.F. Qiu, Uncertain Decision Making Based on Rough Sets, Tsinghua University Press, Beijing, China, 2005.
[2] Z. Pawlak, Rough sets, Int. J. Comput. Inf. Sci. 11 (1982), 341–356.
[3] F. Jang, Y.F. Sui, A novel approach for discretization of continuous attributes in rough set theory, Knowl.-Based Syst. 73 (2015), 324–334.
[4] S.K.M. Wang, W. Ziarko, On optimal decision rules in decision tables, Bull. Pol. Acad. Sci. 33 (1985), 693–696.
[5] Y.Y. Guan, H.K. Wang, Set-valued information systems, Inf. Sci. 176 (2006), 2507–2525.
[6] F.M. Ma, T.F. Zhang, Generalized binary discernibility matrix for attribute reduction in incomplete information systems, J. China Univ. Posts Telecommun. 24 (2017), 57–75.
[7] Z.Q. Meng, Z.Z. Shi, A fast approach to attribute reduction in incomplete decision systems with tolerance relation-based rough sets, Inf. Sci. 179 (2009), 2774–2793.
[8] D.G. Chen, C.Z. Wang, Q.H. Hu, A new approach to attribute reduction of consistent and inconsistent covering decision systems with covering rough sets, Inf. Sci. 177 (2007), 3500–3518.
[9] C.Z. Wang, Q. He, D.G. Chen, *et al.*, A novel method for attribute reduction of covering decision systems, Inf. Sci. 254 (2014), 181–196.
[10] Y. Gao, C.W. Lv, Z.J. Wu, Attribute reduction of Boolean matrix in neighborhood rough set model, Int. J. Comput. Intell. Syst. 13 (2020), 1473–1482.

[11] Y. Chen, K.Y. Liu, J.J. Song, *et al.*, Attribute group for attribute reduction, Inf. Sci. 535 (2020), 64–80.
[12] E.C.C. Tsang, D.G. Chen, D.S. Yeung, Approximations and reducts with covering generalized rough sets, Comput. Math. Appl. 56 (2008), 279–289.
[13] C.Z. Wang, M.W. Shao, B.Q. Sun, *et al.*, An improved attribute reduction scheme with covering based rough sets, Appl. Soft Comput. 26 (2015), 235–243.
[14] C.H. Liu, K.C. Cai, D.Q. Miao, *et al.*, Novel matrix-based approaches to computing minimal and maximal descriptions in covering-based rough sets, Inf. Sci. 539 (2020), 312–326.
[15] F.C. Li, Y.X. Ren, C.X. Jin, Research on attribute reduction method based on dual attribute comprehensive dependency, J. Shanxi Univ. (Nat. Sci. Ed). 43 (2020), 805–812.
[16] Y.P. Shi, Y. Huang, C.Z. Wang, *et al.*, Attribute reduction based on the Boolean matrix, Granul. Comput. 4 (2019), 313–322.
[17] J.K. Chen, Y.J. Lin, G.P. Lin, *et al.*, Attribute reduction of covering decision systems by hypergraph model, Knowl.-Based Syst. 118 (2017), 93–104.
[18] Y.Y. Chen, Y.M. Chen, Feature subset selection based on variable precision neighborhood rough sets, Int. J. Comput. Intell. Syst. 14 (2021), 572–581.