Available online at www.sciencedirect.com

ScienceDirect

Information Sciences 177 (2007) 3500–3518

INFORMATION SCIENCES
AN INTERNATIONAL JOURNAL

www.elsevier.com/locate/ins

A new approach to attribute reduction of consistent and inconsistent covering decision systems with covering rough sets

Chen Degang [a,*], Wang Changzhong [b], Hu Qinghua [c]

[a] *Department of Mathematics and Physics, North China Electric Power University, 102206 Beijing, PR China*
[b] *Department of Mathematics, Harbin Institute of Technology, 150001 Harbin, PR China*
[c] *Power Engineering, Harbin Institute of Technology, 150001 Harbin, PR China*

Received 29 March 2006; received in revised form 23 February 2007; accepted 28 February 2007

Abstract

Traditional rough set theory is mainly used to extract rules from and reduce attributes in databases in which attributes are characterized by partitions, while the covering rough set theory, a generalization of traditional rough set theory, does the same yet characterizes attributes by covers. In this paper, we propose a way to reduce the attributes of covering decision systems, which are databases characterized by covers. First, we define consistent and inconsistent covering decision systems and their attribute reductions. Then, we state the sufficient and the necessary conditions for reduction. Finally, we use a discernibility matrix to design algorithms that compute all the reducts of consistent and inconsistent covering decision systems. Numerical tests on four public data sets show that the proposed attribute reductions of covering decision systems accomplish better classification performance than those of traditional rough sets.

Keywords: Rough sets; Covering rough sets; Attribute reduction; Discernibility matrix; Consistent covering decision systems; Inconsistent covering decision systems

1. Introduction

The theory of traditional rough sets, proposed by Pawlak [18], is an extension of the set theory. Rough sets are used to study intelligent systems that are characterized by insufficient and incomplete information. Rough set theory has been successfully applied to the fields of artificial intelligence including machine learning, pattern recognition, decision analysis, process control, knowledge discovery in databases, and expert systems [5,17,19–22,31,32]. A primary use of rough set theory is to reduce the number of attributes in databases thereby improving the performance of applications in a number of aspects including speed, storage, and

* Corresponding author.
E-mail address: chengdegang@263.net (C. Degang).

0020-0255/$ - see front matter
doi:10.1016/j.ins.2007.02.041

accuracy. For a data set with discrete attribute values, this can be done by reducing the number of redundant attributes and find a subset of the original attributes that are the most informative.

One limitation of attribute reduction in traditional rough sets is that, it is applicable only to databases whose attributes can induce equivalence relations or partitions because traditional rough sets are developed on equivalent relations or partitions. One response to this has been to relax the equivalence relation to a similarity relation and obtain the similarity relation-based rough sets [24,26–29]. Another response has been to relax the partition to a cover and obtain the covering rough sets [2–4,6,16,23,30,33–35]. Much attention has been paid to set approximation by covers, not much work has been done on attribute reduction in covering rough sets up to now, to the best of our knowledge. For example, Bonikowski et al. [2] studies the structures of covers, Mordeson [16] examines the relationship between the approximations of sets defined with respective to covers and some axioms satisfied by traditional rough sets. Chen Degang et al. [6] discusses the covering rough set within the framework of a complete completely distributive lattice. Zhu and Wang [34] proposes the reduction of covering rough sets to reduce the "redundant" members in a cover to find the "smallest" cover which can be used to induces the same covering lower and upper approximations. But this reduction is not in line with the original purpose of attribute reduction in traditional rough sets. Therefore, we study the reduction of covering decision systems in this paper. We define consistent and inconsistent covering decision systems and their attribute reduction, take the approach by discernibility matrix to investigate the structures of these reducts. And furthermore, we give an algorithm to compute reducts from decision systems.

The remainder of this paper is structured as follows. In Section 2, we recall set approximations and attribute reduction with traditional rough sets. In Section 3 we review existing researches on covering rough sets, and introduce the concept of induced cover and reveal the relations between two arbitrary objects related to the induced cover. In Section 4 we discuss the method for reducing a consistent covering decision system. In Section 5 we present a theory for reducing an inconsistent covering decision system. In Section 6, we show some experiments on four public data sets.

2. Set approximations and attribute reduction in information systems

In this section, we review basic concepts about traditional rough sets which can be found in [18,25].

An information system is a pair $\mathbf{A} = (U, A)$, where $U = \{x_1, \ldots, x_n\}$ is a nonempty finite set of objects and $A = \{a_1, a_2, \ldots, a_m\}$ is a nonempty finite set of attributes. With every subset of attributes $B \subseteq A$ we associate binary relation $\mathrm{IND}(B)$, called B-indiscernibility relation, and defined as $\mathrm{IND}(B) = \{(x, y) \in U \times U : a(x) = a(y), \forall a \in B\}$, then $\mathrm{IND}(B)$ is an equivalence relation and $\mathrm{IND}(B) = \cap_{a \in B}\mathrm{IND}(\{a\})$. By $[x]_B$ we denote the equivalence class of $\mathrm{IND}(B)$ including x. For $X \subseteq U$, sets $\{x \in U : [x]_B \subseteq X\}$ and $\{x \in U : [x]_B \cap X \neq \phi\}$ are called B-lower and B-upper approximations of X in $\mathbf{A}$, respectively, and they are denoted as $\underline{B}X$ and $\overline{B}X$ respectively.

By $M(\mathbf{A})$ we denote an $n \times n$ matrix (c_{ij}), called the discernibility matrix of $\mathbf{A}$, such that $c_{ij} = \{a \in A : a(x_i) \neq a(x_j)\}$for $i, j = 1, 2, \ldots, n$. A discernibility function $f(\mathbf{A})$for an information system $\mathbf{A} = (U, A)$ is a Boolean function of m Boolean variables $\overline{a_1}, \ldots, \overline{a_m}$ corresponding to the attributes $a_1, \ldots, a_m$, respectively, and defined as follows:

$$f(\mathbf{A})(\overline{a_1}, \ldots, \overline{a_m}) = \wedge\{\vee(c_{ij}) : 1 \leqslant j < i \leqslant n\},$$

where $\vee(c_{ij})$ is the disjunction of all variables $\overline{a}$ such that $a \in c_{ij}$.

Attribute $a \in B \subseteq A$ is dispensable in B if $\mathrm{IND}(B) = \mathrm{IND}(B - \{a\})$; otherwise a is indispensable in B. The collection of all indispensable attributes in A is called the core of $\mathbf{A}$. We say that $B \subseteq A$ is independent in $\mathbf{A}$ if every attribute in B is indispensable in B. Subset $B \subseteq A$ is called a reduct in $\mathbf{A}$ if B is independent and $\mathrm{IND}(B) = \mathrm{IND}(A)$. The set of all the reducts in $\mathbf{A}$ is denoted as $\mathrm{RED}(\mathbf{A})$.

Let $g(\mathbf{A})$ be the reduced disjunctive form of $f(\mathbf{A})$obtained from $f(\mathbf{A})$ by applying the multiplication and absorption laws as many times as possible. Then there exist l and $X_k \subseteq A$ for $k = 1, \ldots, l$ such that $g(\mathbf{A}) = (\wedge X_1) \vee \cdots \vee (\wedge X_l)$ where each element in X_k appears only one time. We have $\mathrm{RED}(\mathbf{A}) = \{X_1, \ldots, X_l\}$.

A decision system is a pair $\mathbf{A}^* = (U, A \cup \{a^*\})$, $a^*$is called the decision attribute, and attributes in A are called condition attributes. We say $a \in B \subseteq A$ is relatively dispensable in B if $\mathrm{POS}_B(a^*) = \mathrm{POS}_{B-\{a\}}(a^*)$.

Otherwise a is said to be relatively indispensable in B. Here $\mathrm{POS}_B(a^*)$ is the union of $B-$lower approximation of all the equivalence classes of a^*, i.e., $\mathrm{POS}_B(a^*) = \cup_{X\in U/a^*}\underline{B}X$. If every attribute in B is relatively indispensable in B, we say that $B \subseteq A$ is relatively independent in $\mathbf{A}^*$. Subset $B \subseteq A$ is called a relative reduct in $\mathbf{A}^*$ if B is relatively independent in $\mathbf{A}^*$ and $\mathrm{POS}_B(a^*) = \mathrm{POS}_A(a^*)$. The collection of all relatively indispensable attributes in A is called the relative core of $\mathbf{A}$.

Suppose $M(\mathbf{A}^*) = (c_{ij})$, we denote a matrix $\mathbf{M}(\mathbf{A}^*) = (\mathbf{c}_{ij})$ in the following way.

(1) $\mathbf{c}_{ij} = c_{ij} - \{a^*\}$, if ($a^* \in c_{ij}$ and $x_i, x_j \in \mathrm{POS}_A(a^*)$) or $\mathrm{pos}(x_i) \neq \mathrm{pos}(x_j)$;
(2) $\mathbf{c}_{ij} = \phi$, otherwise.

Here pos : $U \to \{0,1\}$ is defined as $\mathrm{pos}(x) = 1$ if and only if $x \in \mathrm{POS}_A(a^*)$. All the relative reducts can be computed in an analogous way as reductions of $M(\mathbf{A})$.

3. Current research on covering rough sets and induced covers

In this section, we first recall the concept of a cover and present three examples suitable for covering rough sets, and then review the existing research on covering rough sets. Finally, propose the definition of induced cover and reveal three basic relations between two objects with respect to the induced cover.

Definition 3.1 [2]. Let U be a universe of discourse, $\mathbf{C}$ a family of subsets of U. $\mathbf{C}$ is called a cover of U if no subset in $\mathbf{C}$ is empty and $\cup\mathbf{C} = U$.

It is clear that a partition of U is certainly a cover of U, so the concept of a cover is an extension of the concept of a partition. In [12–15], a general framework is proposed for the study of approximation operators by using the so-called neighborhood system. In a neighborhood system, each element of the universe is associated with a family of subsets of the universe. This family is called the neighborhood of the element, and each set in the family is called the neighborhood of the element. Thus if we associate every element with a family of subsets in a cover which the element belongs to, clearly we can get a neighborhood system. On the contrary, for a neighborhood system, if every element belongs to its neighborhood, then all the neighborhoods of the elements in the universe form a cover.

One kind of suitable data set for covering rough sets is the information systems that some objects have multiple attribute values for a given attribute. This kind of data set is available when some objects have multi-selections of attribute values for a given attribute. So we have to list all the possible attribute values. One example of this kind of data set is the combination of several information systems. This is illustrated with the following example.

Example 3.1. Let us consider the problem of evaluating credit card applicants. Suppose $U = \{x_1, \ldots, x_9\}$ is a set of nine applicants, E={education; salary} is a set of two attributes, the values of "education" are {best; better; good}, and the values of "salary" are {high; middle; low}. We have three specialists $\{A, B, C\}$ to evaluate the attribute values for these applicants. It is possible that their evaluation results to the same attribute values may not be the same, listed below.

For attribute "education"
A: best $= \{x_1, x_4, x_5, x_7\}$, better $= \{x_2, x_8\}$, good $= \{x_3, x_6, x_9\}$;
B: best $= \{x_1, x_2, x_4, x_7, x_8\}$, better $= \{x_5\}$, good $= \{x_3, x_6, x_9\}$;
C: best $= \{x_1, x_4, x_7\}$, better $= \{x_2, x_8\}$, good $= \{x_3, x_5, x_6, x_9\}$.

For attribute "salary"
A: high $= \{x_1, x_2, x_3\}$, middle $= \{x_4, x_5, x_6, x_7, x_8\}$, low $= \{x_9\}$;
B: high $= \{x_1, x_2, x_3\}$, middle $= \{x_4, x_5, x_6, x_7\}$, low $= \{x_8, x_9\}$;
C: high $= \{x_1, x_2, x_3\}$, middle $= \{x_4, x_5, x_6, x_8\}$, low $= \{x_7, x_9\}$.

Suppose the evaluations given by these specialists are of the same importance. If we want to combine these evaluations without losing information, we should union the evaluations given by each specialist for every attribute value as shown in Table 1. This classification is not a partition, but a cover, which reflects a kind of uncertainty caused by the differences in interpretation of data.

A fuzzy data set is another kind of data suitable for covering generalized rough sets [7]. Suppose U is a universe, C is a set of N fuzzy attributes. Every sample x_i in U is described by the fuzzy attributes in C, i.e., they are measured by partial membership degrees, μ_i^j $(j = 1,\ldots,N)$, which are graded at interval $[0,1]$: $x_i = [\mu_i^1, \mu_i^2, \ldots, \mu_i^N]$. Similarity relation Ron the universe$U$ can be defined using these fuzzy attributes [7]. Similarity relation R is a fuzzy relation satisfying reflexivity, symmetry and T-transitivity, which is a triangular norm in this paper. The fuzzy set $[x]_R$, defined as $[x]_R(y) = R(x,y), \forall y \in U$, is called the fuzzy similarity class of $x \in U$. For $\alpha \in (0,1]$, the collection of all the α-cut of $[x]_R$, i.e., $\{([x]_R)_\alpha : x \in U\}$, is a cover of U. Here $([x]_R)_\alpha = \{y \in U : [x]_R(y) = R(x,y) \geqslant \alpha\}$. And α is a criterion of similarity that can be set beforehand. $R(x,y) \geqslant \alpha$ means y is similar to x related to the criterion α. Otherwise y is not similar to x. We have the following example to illustrate this.

Example 3.2. In Example 3.1 if every applicant is described by six fuzzy attributes $C_1 =$ best education, $C_2 =$ better education, $C_3 =$ good education, $C_4 =$ high salary, $C_5 =$ middle salary and $C_6 =$ low salary, we can then compute the membership degrees of every applicant from the evaluation results given by three specialists $\{A,B,C\}$. For instance, two specialists believe x_2 has better education background, then $C_2(x_2) = 2/3$, as shown in Table 2.

Fuzzy similarity relation R can be computed using $R(x_i,x_j) = \inf_{1\leqslant k\leqslant 6}(1- \mid \mu_i^k - \mu_j^k \mid)$ as shown below:

$$(R(x_i,x_j)) = \begin{pmatrix} 1 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & 1 & 1/3 & 0 & 2/3 & 1/3 & 0 \\ & & & & 1 & 1/3 & 1/3 & 2/3 & 0 \\ & & & & & 1 & 0 & 0 & 0 \\ & & & & & & 1 & 1/3 & 0 \\ & & & & & & & 1 & 0 \\ & & & & & & & & 1 \end{pmatrix}.$$

Table 1
Classification by evaluations of all three specialists

Salary	Education		
	Best	Better	Good
High	$\{x_1,x_2\}$	$\{x_2\}$	$\{x_3\}$
Middle	$\{x_4,x_5,x_7,x_8\}$	$\{x_5,x_8\}$	$\{x_5,x_6\}$
Low	$\{x_7,x_8\}$	$\{x_8\}$	$\{x_9\}$

Table 2
Membership degrees of every applicant

	C_1	C_2	C_3	C_4	C_5	C_6
x_1	1	0	0	1	0	0
x_2	1/3	2/3	0	1	0	0
x_3	0	0	1	1	0	0
x_4	1	0	0	0	1	0
x_5	1/3	1/3	1/3	0	1	0
x_6	0	0	1	0	1	0
x_7	1	0	0	0	2/3	1/3
x_8	1/3	2/3	0	0	2/3	1/3
x_9	0	0	1	0	0	1

R is then a fuzzy T_L-similarity relation, T_L is the Lukasiewicz t-norm defined as $T_L(x, y) = \max\{0,\ x + y - 1\}$. By taking $\alpha = 1/3$, we can get a cover as

$\{\{x_1,x_2\}, \{x_3\}, \{x_4,x_5,x_7, x_8\}, \{x_4,x_5,x_6,x_7,x_8\}, \{x_5,x_6\}, \{x_9\}\}$.

Covers are also available in similarity relation-based rough sets [26]. Suppose R is a reflexivity relation on U, denote $R_s(x) = \{y : (x, y) \in R\}$. Then $\{R_s(x) : x \in U\}$ is also a cover of U. So similarity relation-based rough sets in [26] can be brought into the framework of covering rough set theory from the theoretical viewpoint.

There are just some applications of covering rough sets, and obviously covering rough sets are not just suitable for these applications. We omit further discussion on this topic.

In the following we begin to study the intersection of several covers.

Definition 3.2. Let $\mathbf{C} = \{C_1, C_2, \ldots, C_n\}$ is a cover of U. For every $x \in U$, let $C_x = \cap\{C_j : C_j \in \mathbf{C}, x \in C_j\}$. $\mathrm{Cov}(\mathbf{C}) = \{C_x : x \in U\}$ is then also a cover of U. We call it the induced cover of $\mathbf{C}$.

For every $x \in U$, C_x is the minimal set in $\mathrm{Cov}(\mathbf{C})$ including x. If $\mathbf{C}$ is an attribute and every C_i denotes an attribute value of $\mathbf{C}$, i.e., the collection of objects in U takes a certain attribute value, and $C_x = C_1 \cap C_2$, this implies the possible values of x with respect to $\mathbf{C}$ can be C_1 or C_2, i.e., the relation between C_1 and C_2 is a disjunction. $\mathrm{Cov}(\mathbf{C}) = \mathbf{C}$ if and only if $\mathbf{C}$ is a partition. For every $x, y \in U$, if $y \in C_x$ then $C_x \supseteq C_y$, so if $y \in C_x$ and $x \in C_y$, then $C_x = C_y$. Every element in $\mathrm{Cov}(\mathbf{C})$ cannot be written as the union of other elements in $\mathrm{Cov}(\mathbf{C})$.

However, Definition 3.2 is not applicable to the incomplete information system shown in [10,11]. It is believed that every value of this attribute is available for this object if an object in an incomplete information system losses its attribute value [10,11]. All the objects with an attribute value missing will be classified as a separate class by Definition 3.2, and this class has no overlap with other classes, so the description of this class is also meaningless since elements in this class have already lost their attribute values as shown in [10].

Definition 3.3. Let $\boldsymbol{\Delta} = \{\mathbf{C}_i : i = 1, \ldots, m\}$ be a family of covers of U. For every $x \in U$, let $\Delta_x = \cap\{C_{ix} : C_{ix} \in \mathrm{Cov}(\mathbf{C}_i), x \in C_{ix}\}$, then $\mathrm{Cov}(\boldsymbol{\Delta}) = \{\Delta_x : x \in U\}$ is also a cover of U. We call it the induced cover of Δ.

Clearly Δ_x is the intersection of all the elements in every $\mathbf{C}_i$ including x, so for every $x \in U$, Δ_x is the minimal set in $\mathrm{Cov}(\boldsymbol{\Delta})$ including x. If every cover in $\boldsymbol{\Delta}$ is an attribute, then $\Delta_x = \cap\{C_{ix} : C_{ix} \in \mathrm{Cov}(\mathbf{C}_i), x \in C_{ix}\}$means the relation among C_{ix} is a conjunction. $\mathrm{Cov}(\boldsymbol{\Delta})$ can be viewed as the intersection of covers in $\boldsymbol{\Delta}$. If every cover in $\boldsymbol{\Delta}$ is a partition, then $\mathrm{Cov}(\boldsymbol{\Delta})$ is also a partition and Δ_x is the equivalence class including x. For every $x, y \in U$, if $y \in \Delta_x$, then $\Delta_x \supseteq \Delta_y$, so if $y \in \Delta_x$ and $x \in \Delta_y$, then $\Delta_x = \Delta_y$. Every element in $\mathrm{Cov}(\boldsymbol{\Delta})$ cannot be written as the union of other elements in $\mathrm{Cov}(\boldsymbol{\Delta})$. We employ an example to illustrate the practical meaning of C_x and Δ_x.

Example 3.3. In Example 3.1 if let $\boldsymbol{\Delta} = \{\mathbf{C}_1, \mathbf{C}_2\}$, where $\mathbf{C}_1$ denotes the attribute "education" and $\mathbf{C}_2$ denotes the attribute "salary", then

$$\mathbf{C}_1 = \{C_{11} = \{x_1, x_2, x_4, x_5, x_7, x_8\}\ (\text{best}), C_{12} = \{x_2, x_5, x_8\}\ (\text{better}), C_{13} = \{x_3, x_5, x_6, x_9\}\ (\text{good})\},$$
$$\mathbf{C}_2 = \{C_{21} = \{x_1, x_2, x_3\}\ (\text{high}), C_{22} = \{x_4, x_5, x_6, x_7, x_8\}\ (\text{middle}), C_{23} = \{x_7, x_8, x_9\}\ (\text{lower})\}.$$

We have $C_{1x_5} = \{x_5\} = C_{11} \cap C_{12} \cap C_{13}$, which implies the possible description of x_5 is $\{\text{best} \vee \text{better} \vee \text{good}\}$ according to attribute "education". $\Delta_{x_8} = (C_{11} \cap C_{12}) \cap (C_{22} \cap C_{23})$, which implies the possible description of x_8 is $\{(\text{best} \vee \text{better}) \wedge (\text{middle} \vee \text{lower})\}$.

For every $X \subseteq U$, the lower and upper approximation of X with respect to $\mathrm{Cov}(\boldsymbol{\Delta})$ are defined as follows:

$$\underline{\boldsymbol{\Delta}}(X) = \cup\{\Delta_x : \Delta_x \subseteq X\}, \qquad \overline{\boldsymbol{\Delta}}(X) = \cup\{\Delta_x : \Delta_x \cap X \neq \phi\}.$$

The positive, negative and boundary regions of Xrelative to $\boldsymbol{\Delta}$ are computed using the following formulas respectively:

$$\mathrm{POS}_{\Delta}(X) = \underline{\boldsymbol{\Delta}}(X), \quad NEG_{\Delta}(X) = U - \overline{\boldsymbol{\Delta}}(X), \quad BN_{\Delta}(X) = \overline{\boldsymbol{\Delta}}(X) - \underline{\boldsymbol{\Delta}}(X).$$

If every cover in $\mathbf{\Delta}$ is an attribute, then both $\underline{\mathbf{\Delta}}(X)$ and $\overline{\mathbf{\Delta}}(X)$ have certain meaning, i.e., they are the combination of disjunctions and conjunctions of some attribute values. A rule $A \Rightarrow B$ is called a possible rule if $\frac{|A \cap B|}{|A|} < 1$, and is called a certain rule if $\frac{|A \cap B|}{|A|} = 1$ [6]. So if we denote $\overline{\mathbf{\Delta}}(X) \Rightarrow X$ as the possible rule and every $\underline{\mathbf{\Delta}}(X) \Rightarrow X$ as the certain rule, these rules are meaningful.

The current researches on covering rough sets focus on approximation operators, where construction of a lower approximation operator are the same as that of traditional rough sets, but the definition of an upper approximation operator is different [2,6,16,34]. These important observations have been studied in detail in [6] under a wider framework of complete completely distributive lattice. We list below the current definitions of lower and upper approximations in covering rough sets.

Definition 3.4 [2]. The ordered pair $(U, \mathbf{C})$ is called a covering approximation space, where U is a nonempty set called a universe, and $\mathbf{C}$ is its finite cover.

Definition 3.5 [2]. Let $(U, \mathbf{C})$ be the approximation space, $x \in U$. The following family $Md(x) = \{K \in \mathbf{C} : x \in K \wedge \forall S \in \mathbf{C}(x \in S \wedge S \subseteq K) \Rightarrow K = S\}$ is called the minimal description of object x. In another words, every element in $Md(x)$ is a minimal one including x.

Definition 3.6 [2]. For any $X \subseteq U$, the set $X_* = \cup\{K \in \mathbf{C} : K \subseteq X\}$ is called the lower approximation of the set X.

Definition 3.7 [2]. For any $X \subseteq U$, set $X^* = X_* \cup \{K \in Md(x) : x \in X - X_*\}$ is called the upper approximation of set X.

Clearly in $\mathrm{Cov}(\mathbf{C})$, Δ_x is the minimal description of object x. In [16] the upper approximation of X is defined as set $\{y \in U : \exists C \in \mathbf{C}, y \in C \text{ and } C \cap X \neq \phi\}$. Our definition of upper approximation $\overline{\Delta}(X)$ is a special case of the one in [6] and it is different from the one in [2,16]. It was argued in [6] that it is possible to lose some useful information by the definition of upper approximation in [2] and it is also possible to include some unnecessary information by the definition of upper approximation in [16]. Further detail information please see [6].

According to traditional rough set theory, for every $x, y \in U$, the equivalence classes including these two objects are either equal to each other or have an empty overlap. If their equivalence classes are equal, these two objects are then called indiscernible. But for the covering rough set, it is quite different and seems more complex. For every $x, y \in U$, there are three possible relations with respect to $\mathbf{\Delta}$ between x and y: $R(1)$ $\Delta_x = \Delta_y$; $R(2)$ $\Delta_x \subset \Delta_y$ or $\Delta_x \supset \Delta_y$; $R(3)$ $\Delta_x \neq \Delta_y$ and $\Delta_x \not\subset \Delta_y$, $\Delta_y \not\subset \Delta_x$i.e., Δ_x and Δ_y cannot be included by each other. In the following we denote relation $R(3)$ by $\Delta_x \not\subset \Delta_y$, $\Delta_y \not\subset \Delta_x$ for short. If $\mathrm{Cov}(\mathbf{\Delta})$ is a partition, these relations are then just the cases in traditional rough sets mentioned above. We use the following theorem to characterize these three relations.

Theorem 3.8. *Supposing U is a finite universe and $\mathbf{\Delta} = \{\mathbf{C}_i : i = 1, \ldots, m\}$ is a family of covers of U, the following statements hold*:

(1) $\Delta_x = \Delta_y$ *if and only if for every* $\mathbf{C}_i \in \mathbf{\Delta}$ *we have* $C_{ix} = C_{iy}$.
(2) $\Delta_x \supset \Delta_y$ *if and only if for every* $\mathbf{C}_i \in \mathbf{\Delta}$ *we have* $C_{ix} \supseteq C_{iy}$ *and there is a* $\mathbf{C}_{i_0} \in \mathbf{\Delta}$ *such that* $C_{i_{0x}} \supset C_{i_{0y}}$.
(3) $\Delta_x \not\subset \Delta_y$ *and* $\Delta_y \not\subset \Delta_x$ *hold if and only if there are* $\mathbf{C}_i, \mathbf{C}_j \in \mathbf{\Delta}$ *such that* $C_{ix} \subset C_{iy}$ *and* $C_{jx} \supset C_{jy}$ *or there is a* $\mathbf{C}_{i_0} \in \mathbf{\Delta}$ *such that* $C_{i_{0x}} \not\subset C_{i_{0y}}$ *and* $C_{i_{0y}} \not\subset C_{i_{0x}}$.

Proof

(1) If for every $\mathbf{C}_i \in \mathbf{\Delta}$ we have $C_{ix} = C_{iy}$, clearly $\Delta_x = \Delta_y$ holds.
If there exists a $\mathbf{C}_{i_0} \in \mathbf{\Delta}$ such that $C_{i_{0x}} \neq C_{i_{0y}}$, then at least one of $x \notin C_{i_{0y}}$ and holds, which implies $\Delta_x \neq \Delta_y$.
(2) If for every $\mathbf{C}_i \in \mathbf{\Delta}$ we have $C_{ix} \supseteq C_{iy}$, then $\Delta_x \supseteq \Delta_y$ holds.
If there is a $\mathbf{C}_{i_0} \in \mathbf{\Delta}$ such that $C_{i_{0x}} \supset C_{i_{0y}}$, then $x \notin C_{i_{0y}}$ holds which implies $\Delta_x \supset \Delta_y$.

If $\Delta_x \supset \Delta_y$, then we have $y \in \Delta_x$. So for every $\mathbf{C}_i \in \mathbf{\Delta}$ we have $y \in C_{ix}$ which implies $C_{ix} \supseteq C_{iy}$. From $\Delta_x \supset \Delta_y$ we have $x \notin \Delta_y$, so there is $\mathbf{C}_{i_0} \in \mathbf{\Delta}$ such that $x \notin C_{i_{0y}}$, so $C_{i_{0x}} \supset C_{i_{0y}}$.

(3) If there are $\mathbf{C}_i, \mathbf{C}_j \in \mathbf{\Delta}$ such that $C_{ix} \subset C_{iy}$ and $C_{jx} \supset C_{jy}$, then $y \notin C_{i_x}$ which implies $y \notin \Delta_x$, and $x \notin C_{jy}$ which implies $x \notin \Delta_y$. We have thus $\Delta_x \not\subset \Delta_y$ and $\Delta_y \not\subset \Delta_x$.
If there is a $\mathbf{C}_{i_0} \in \mathbf{\Delta}$ such that $C_{i_{0x}} \not\subset C_{i_{0y}}$ and $C_{i_{0y}} \not\subset C_{i_{0x}}$, then we have $x \notin C_{i_{0y}}$ and $y \notin C_{i_{0x}}$, so $x \notin \Delta_y$ and $y \notin \Delta_x$ which implies $\Delta_x \not\subset \Delta_y$ and $\Delta_y \not\subset \Delta_x$.
If $\Delta_x \not\subset \Delta_y$ and $\Delta_y \not\subset \Delta_x$ hold and there is not $\mathbf{C}_{i_0} \in \mathbf{\Delta}$ satisfying $C_{i_{0x}} \not\subset C_{i_{0y}}$ and $C_{i_{0y}} \not\subset C_{i_{0x}}$, then from the proof of (2) there must be $\mathbf{C}_i, \mathbf{C}_j \in \mathbf{\Delta}$ such that $C_{ix} \subset C_{iy}$ and $C_{jx} \supset C_{jy}$. Otherwise it leads to $\Delta_x \supset \Delta_y$ or $\Delta_x \subset \Delta_y$ which is a contradiction. □

These three relations will be are referred to hereafter as the original relations with respect to $\mathbf{\Delta}$ between $x, y \in U$ in the following sections. For $\mathbf{P} \subset \mathbf{\Delta}$, suppose $\mathrm{Cov}(\mathbf{P}) = \{\mathbf{P}_x : x \in U\}$, we say the original relations with respect to $\mathbf{\Delta}$ between $x, y \in U$ is changed with respect to $\mathbf{P} \subset \mathbf{\Delta}$ if the relations between Δ_x, Δ_y and $\mathbf{P}_x, \mathbf{P}_y$ are different. For example, if $\Delta_x = \Delta_y$, and $\mathbf{P}_x \subset \mathbf{P}_y$, then the original relation between $x, y \in U$ with respect to $\mathbf{\Delta}$ is changed with respect to $\mathbf{P} \subset \mathbf{\Delta}$. We just say in short the original relation between $x, y \in U$ is changed without any confusion.

4. Attribute reduction of consistent covering decision systems

As mentioned in Section 2, the attribute set of a decision system consists of two parts; a conditional attribute set and a decision attribute. In a classical decision system, collections of objects with the same conditional attribute values have no overlapping, so every conditional attribute can induce a partition. But as indicated by the examples in Section 3 there are more complex cases where attributes can induce covers rather than partitions. We call decision systems with conditional attributes which induce covering the covering decision systems. The covering decision systems can be divided into consistent covering decision systems and inconsistent decision systems. We start next our study on attribute reduction of consistent covering systems by defining consistent covering systems first.

Definition 4.1. Let $\mathbf{\Delta} = \{\mathbf{C}_i : i = 1, \ldots, m\}$ be a family of covers of U, D is a decision attribute, U/D is a decision partition on U. If for $\forall x \in U$, $\exists D_j \in U/D$ such that $\Delta_x \subseteq D_j$, then decision system $(U, \mathbf{\Delta}, D)$ is called a consistent covering decision system, and denoted as $\mathrm{Cov}(\mathbf{\Delta}) \leqslant U/D$. Otherwise, $(U, \mathbf{\Delta}, D)$ is called an inconsistent covering decision system. The positive region of D relative to $\mathbf{\Delta}$ is defined as $\mathrm{POS}_{\mathbf{\Delta}}(D) = \bigcup_{X \in U/D} \underline{\mathbf{\Delta}}(X)$. In Example 3.1, if the final evaluation decision is $D = \{\{x_1, x_2, x_3\}$ (eligible), $\{x_4, x_5, x_6, x_7, x_8\}$ (need further evaluation), $\{x_9\}$ (incompetent)$\}$, then $(U, \mathbf{\Delta} = \{\mathbf{C}_1, \mathbf{C}_2\}, D\}$ is a consistent covering decision system. If the final evaluation decision is $D = \{\{x_1, x_2, x_3\}$ (eligible), $\{x_4, x_5, x_6\}$ (need further evaluation), $\{x_7, x_8, x_9\}$ (incompetent)$\}$, then $(U, \mathbf{\Delta} = \{\mathbf{C}_1, \mathbf{C}_2\}, D\}$ is an inconsistent covering decision system.

Suppose $\mathrm{Cov}(\mathbf{\Delta}) \leqslant U/D$, clearly for every $x \in U$, $\Delta_x \subseteq [x]_D$ is always true, thus we have $[x]_D = \cup\{\Delta_x : x \in [x]_D\}$. So for a consistent covering decision system, the lower approximation of every element in U/D is just equal to both itself and its upper approximation, which implies $\mathrm{POS}_{\mathbf{\Delta}}(D) = U$. For every $X \in U/D$, if we believe every $\overline{\mathbf{\Delta}}(X) \Rightarrow X$ as the possible rule and every $\underline{\mathbf{\Delta}}(X) \Rightarrow X$ as a certain rule, then all the decision rules extracted from a consistent covering decision system are consistent.

If every cover in Δ is a partition, then $\mathrm{COV}(\Delta)$ is also a partition, and $\mathrm{Cov}(\mathbf{\Delta}) \leqslant U/D$ is just the case of a consistent decision system in traditional rough set theory. Let $D = \{d\}$, then $d(x)$ is a decision function $d : U \to V_d$ of the universe U into value set V_d. For every$x_i, x_j \in U$, if $\Delta_{x_i} \subseteq \Delta_{x_j}$, then $d(x_i) = d([x_i]_D) = d(\Delta_{x_i}) = d(\Delta_{x_j}) = d(x_j) = d([x_j]_D)$. If $d(\Delta_{x_i}) \neq d(\Delta_{x_j})$, then $\Delta_{x_i} \cap \Delta_{x_j} = \phi$, i.e. $\Delta_{x_i} \not\subset \Delta_{x_j}$ and $\Delta_{x_j} \not\subset \Delta_{x_i}$. But if $\Delta_{x_i} \not\subset \Delta_{x_j}$ and $\Delta_{x_j} \not\subset \Delta_{x_i}$, then either $d(\Delta_{x_i}) = d(\Delta_{x_j})$or $d(\Delta_{x_i}) \neq d(\Delta_{x_j})$ are possible. For this case, if $\Delta_{x_i} \cap \Delta_{x_j} \neq \phi$, we have $d(\Delta_{x_i}) = d(\Delta_{x_j})$. If $d(\Delta_{x_i}) = d(\Delta_{x_j})$, then both $\Delta_{x_i} \not\subset \Delta_{x_j}$ and $\Delta_{x_j} \not\subset \Delta_{x_i}$, or $\Delta_{x_i} \subseteq \Delta_{x_j}$, or $\Delta_{x_i} \supseteq \Delta_{x_j}$ are possible. We define next the relative reduction of a consistent covering decision system.

Definition 4.2. Let $(U, \mathbf{\Delta}, D = \{d\})$ be a consistent covering decision system. For $\mathbf{C}_i \in \mathbf{\Delta}$, if $\mathrm{Cov}(\mathbf{\Delta} - \{\mathbf{C}_i\}) \leqslant U/D$, then $\mathbf{C}_i$ is called superfluous relative to D in $\mathbf{\Delta}$, otherwise $\mathbf{C}_i$ is called indispensable relative to D in $\mathbf{\Delta}$. For every $\mathbf{P} \subseteq \mathbf{\Delta}$ satisfying $\mathrm{Cov}(\mathbf{P}) \leqslant U/D$, if every element in $\mathbf{P}$ is indispensable, i.e., for every $\mathbf{C}_i \in \mathbf{P}$, $\mathrm{Cov}(\mathbf{P} - \{\mathbf{C}_i\}) \leqslant U/D$ is not true, then $\mathbf{P}$ is called a reduct of $\mathbf{\Delta}$ relative to D, relative reduct in short. The collection of all the indispensable elements in $\mathbf{\Delta}$ is called the core of $\mathbf{\Delta}$ relative to D, denoted as $\mathrm{Core}_D(\mathbf{\Delta})$. The relative reduct of a consistent covering decision system is the minimal set of conditional covers(attributes) to ensure every decision rule still consistent. For a single cover $\mathbf{C}_i$, we present some equivalence conditions to judge whether it is indispensable.

Theorem 4.3. *Suppose* $\mathrm{Cov}(\mathbf{\Delta}) \leqslant U/D$, $\mathbf{C}_i \in \Delta$, *and* $\mathrm{Cov}(\Delta - \{\mathbf{C}_i\}) = \{\mathbf{P}_x : x \in U\}$, *then* $\mathbf{C}_i$ *is indispensable, i.e.*, $\mathrm{Cov}(\mathbf{\Delta} - \{\mathbf{C}_i\}) \leqslant U/D$ *is not true if and only if* $\exists x \in U$ *such that* $\mathbf{P}_x \subseteq [x]_D$ *is not true.*

Proof. If $\mathrm{Cov}(\mathbf{\Delta} - \{\mathbf{C}_i\}) \leqslant U/D$ is not true, then there exists $x \in U$, for every $y \in U$ such that $\mathbf{P}_x \subseteq [y]_D$ is not true. Especially, $\mathbf{P}_x \subseteq [x]_D$ is also not true.

If there exists $x \in U$ such that $\mathbf{P}_x \subseteq [x]_D$ is not true, by $x \in \mathbf{P}_x$, for every $y \in U$, $\mathbf{P}_x \subseteq [y]_D$ is not true. So $\mathrm{Cov}(\mathbf{\Delta} - \{\mathbf{C}_i\}) \leqslant U/D$ is not true. □

It should be noted $\mathbf{P}_x \subseteq [x]_D$ is not true means $(U, \mathbf{\Delta} - \{\mathbf{C}_i\}, D)$ is an inconsistent decision system, $\mathbf{C}_i$ is thus indispensable implies it is a key cover to ensure $(U, \mathbf{\Delta}, D)$ is a consistent decision system.

Theorem 4.4. *Suppose* $\mathrm{Cov}(\mathbf{\Delta}) \leqslant U/D$, $\mathbf{C}_i \in \Delta$, $\mathbf{C}_i$ *is then indispensable, i.e.*, $\mathrm{Cov}(\mathbf{\Delta} - \{\mathbf{C}_i\}) \leqslant U/D$ *is not true if and only if there is at least a pair of* $x_i, x_j \in U$ *satisfying* $d(\Delta_{x_i}) \neq d(\Delta_{x_j})$, *of which the original relation with respect to* $\mathbf{\Delta}$ *changes after* $\mathbf{C}_i$ *is deleted from* $\mathbf{\Delta}$.

Proof. We denote $\mathrm{Cov}(\Delta - \{\mathbf{C}_i\}) = \{\mathbf{P}_x : x \in U\}$, if $\mathrm{Cov}(\mathbf{\Delta} - \{\mathbf{C}_i\}) \leqslant U/D$ is not true, then there are $x_0, y_0 \in U$ such that $y_0 \in \mathbf{P}_{x_0}$ and $y_0 \notin [x_0]_D$, which implies $\mathbf{P}_{y_0} \subseteq \mathrm{P}_{x_0}$ and $\Delta_{x_0} \not\subset \Delta_{y_0}$, $\Delta_{y_0} \not\subset \Delta_{x_0}$ respectively, and x_0, y_0 satisfy $d(\Delta_{x_0}) \neq d(\Delta_{y_0})$, so the original relation of x_0, y_0 with respect to $\mathbf{\Delta}$ changes after C_i is deleted from $\mathbf{\Delta}$.

Suppose $x_0, y_0 \in U$ satisfy $d(\Delta_{x_0}) \neq d(\Delta_{y_0})$, which implies $[x_0]_D \cap [y_0]_D = \phi$, $\Delta_{x_0} \not\subset \Delta_{y_0}$ and $\Delta_{y_0} \not\subset \Delta_{x_0}$, if their original relation with respect to $\mathbf{\Delta}$ changes after deleting $\mathbf{C}_i$ from $\mathbf{\Delta}$, which means $\mathbf{P}_{x_0} \not\subset \mathbf{P}_{y_0}$ or $\mathbf{P}_{y_0} \not\subset \mathbf{P}_{x_0}$ does not hold, so if $\mathbf{P}_{x_0} \subseteq \mathbf{P}_{y_0}$, then $\mathbf{P}_{y_0} \subseteq [y_0]_D$ does not hold, otherwise it implies $x_0 \in [y_0]_D$ which is a contradiction; if $\mathbf{P}_{x_0} \supseteq \mathbf{P}_{y_0}$, then $\mathbf{P}_{x_0} \subseteq [x_0]_D$ does not hold; if $\mathbf{P}_{x_0} = \mathbf{P}_{y_0}$ then both $\mathbf{P}_{x_0} \subseteq [x_0]_D$ and $\mathbf{P}_{y_0} \subseteq [y_0]_D$ do not hold. So $\mathrm{Cov}(\mathbf{\Delta} - \{\mathbf{C}_i\}) \leqslant U/D$ is not true. □

Theorem 4.4 implies that an indispensable cover can be characterized by the original relation between two elements in the universe according to Theorem 3.8. Thus we have the following theorem to characterize a consistent decision system.

Theorem 4.5. *Suppose* $\mathrm{Cov}(\mathbf{\Delta}) \leqslant U/D$, $\mathbf{P} \subseteq \Delta$, *then* $\mathrm{Cov}(\mathbf{P}) \leqslant U/D$ *if and only if for* $\forall x_i, x_j \in U$ *satisfying* $d(\Delta_{x_i}) \neq d(\Delta_{x_j})$, *the relation between* x_i *and* x_j *with respect to* $\mathbf{\Delta}$ *is equivalent to their relation with respect to* $\mathbf{P}$, *i.e.*, $\Delta_{x_i} \not\subset \Delta_{x_j}, \Delta_{x_j} \not\subset \Delta_{x_i} \Longleftrightarrow \mathbf{P}_{x_i} \not\subset \mathbf{P}_{x_j}, \mathbf{P}_{x_j} \not\subset \mathbf{P}_{x_i}$.

Proof. Necessary condition is obvious. We have to prove the sufficient condition.

Suppose $\mathrm{Cov}(\mathbf{P}) \leqslant U/D$ is not true, then there are $x_0, y_0 \in U$ such that $y_0 \in \mathbf{P}_{x_0}$ and $y_0 \notin [x_0]_D$, which implies $\mathbf{P}_{y_0} \subseteq \mathbf{P}_{x_0}$ and $\Delta_{x_0} \not\subset \Delta_{y_0}$, $\Delta_{y_0} \not\subset \Delta_{x_0}$, and x_0, y_0 satisfying $d(\Delta_{x_0}) \neq d(\Delta_{y_0})$, which is a contradiction. □

Theorem 4.6. *For a consistent covering decision system* $(U, \mathbf{\Delta}, D)$, *there is always a reduct.*

Proof. If for every $\mathbf{C}_i \in \mathbf{\Delta}$, $\mathrm{Cov}(\mathbf{\Delta}) \leqslant U/D$ is not true, then $\mathbf{\Delta}$ is just the reduction of itself. If $\mathbf{C}_i \in \mathbf{\Delta}$ and $\mathrm{Cov}(\mathbf{\Delta}) \leqslant U/D$ holds, then we deal with $\mathbf{\Delta} - \{\mathbf{C}_i\}$. Since $\mathbf{\Delta}$ is a finite set, a subset $\mathbf{P}$ of $\mathbf{\Delta}$ can be obtained as a relative reduction of $\mathbf{\Delta}$. □

The purpose of relative reduction of $\mathbf{\Delta}$ is to find the minimal subset of $\mathbf{\Delta}$ to keep every decision rule invariant. By Theorems 4.3–4.5 we know that it is equivalent to keeping the original relation between every two elements which different decision values invariant, and this relation is just relation $R(3)$ in Section 3. By (3) of Theorem 3.8 we can design an algorithm to compute all the relative reducts.

Definition 4.7. Let $(U, \mathbf{\Delta}, D = \{d\})$ be a consistent covering decision system. Suppose $U = \{x_1, \ldots, x_n\}$, by $M(U, \mathbf{\Delta}, D)$ we denote a $n \times n$ matrix (c_{ij}), called the discernibility matrix of $(U, \mathbf{\Delta}, D)$, defined as

$$c_{ij} = \begin{cases} \{\mathbf{C} \in \mathbf{\Delta} : (C_{x_i} \not\subset C_{x_j}) \wedge (C_{x_j} \not\subset C_{x_i})\} \cup \{\mathbf{C}_s \wedge \mathbf{C}_t : (C_{s_{x_i}} \subset C_{s_{x_j}}) \wedge (C_{t_{x_j}} \subset C_{t_{x_i}})\} & d(\Delta_{x_i}) \neq d(\Delta_{x_j}) \\ \mathbf{\Delta} & d(\Delta_{x_i}) = d(\Delta_{x_j}) \end{cases}$$

for $x_i, x_j \in U$.

If $\mathbf{C} \in c_{ij}$ for $d(\Delta_{x_i}) \neq d(\Delta_{x_j})$, $\mathbf{C}$ is one of the covers to maintain the relation between x_i and x_j with respect to $\mathbf{\Delta}$. Here we should point out that if $c_{ij} = \{\mathbf{C}_s : s = 1, \ldots, l\}$, the relations between elements in $c_{ij}$is a disjunction; if $c_{ij} = \{\mathbf{C}, \mathbf{C}_s \wedge \mathbf{C}_t : s, t \leqslant n\}$, we mean it is conjunction between $\mathbf{C}_s$ and $\mathbf{C}_t$, i.e. if $\mathbf{C}_s \in c_{ij}$, then $\mathbf{C}_t \in c_{ij}$; while it is a disjunction between $\mathbf{C}_s \wedge \mathbf{C}_t$ and $\mathbf{C}$, $\mathbf{C}_{s_0} \wedge \mathbf{C}_{t_0}$ and $\mathbf{C}_{s_1} \wedge \mathbf{C}_{t_1}$, $s_0, t_0, s_1, t_1 \leqslant n$. Since $M(U, \mathbf{\Delta}, D)$ is symmetric and $c_{ii} = \mathbf{\Delta}$, for $i = 1, \ldots, n$, we represent $M(U, \mathbf{\Delta}, D)$ only by elements in the lower triangle of $M(U, \mathbf{\Delta}, D)$, i.e., the c_{ij}'s with $1 \leqslant j \leqslant i \leqslant n$.

Theorem 4.8. *Let $(U, \mathbf{\Delta}, D = \{d\})$ be a consistent covering decision system, then we have*:

(1) *For $\mathbf{P} \subseteq \mathbf{\Delta}$, $\mathbf{P} \cap c_{ij} \neq \phi$ holds for every $i, j \leqslant n$ if and only if $\mathrm{COV}(\mathbf{P}) \leqslant U/D$. Here $\mathbf{P} \cap c_{ij} \neq \phi$ means if $\mathbf{C}_s \wedge \mathbf{C}_t \in c_{ij}$ and $\mathbf{C}_s \in \mathbf{P} \cap c_{ij}$, then $\mathbf{C}_t \in \mathbf{P} \cap c_{ij}$ holds.*
(2) $\mathrm{Core}_D(\mathbf{\Delta}) = \{\mathbf{C} \in \mathbf{\Delta} : c_{ij} = \{\mathbf{C}\} \vee c_{ij} = \{\mathbf{C} \wedge \mathbf{C}_t : t = 1, \ldots, k\}\}$ *for some i, j.*

Proof

(1) Suppose $\mathrm{COV}(\mathbf{P}) \leqslant U/D$.If $d(\Delta_{x_i}) = d(\Delta_{x_j})$, then $c_{ij} = \Delta$, hence $\mathbf{P} \cap c_{ij} \neq \phi$ is true.If $d(\Delta_{x_i}) \neq d(\Delta_{x_j})$, since $\Delta_{x_i} \subseteq \mathbf{P}_{x_i} \subseteq [x_i]_D$, $\Delta_{x_j} \subseteq \mathbf{P}_{x_j} \subseteq [x_j]_D$, then we have $d(\mathbf{P}_{x_i}) \neq d(\mathbf{P}_{x_j}) \Longleftrightarrow d(\Delta_{x_i}) \neq d(\Delta_{x_j})$, i.e., $\Delta_{x_i} \not\subset \Delta_{x_j}$, $\Delta_{x_j} \not\subset \Delta_{x_i} \Longleftrightarrow \mathbf{P}_{x_i} \not\subset \mathbf{P}_{x_j}, \mathbf{P}_{x_j} \not\subset \mathbf{P}_{x_i}$, hence by (3) of Theorem 3.8 there are $\mathbf{C}_i, \mathbf{C}_j \in \mathbf{P}$ such that $C_{ix} \subset C_{iy}$ and $C_{jx} \supset C_{jy}$ or there is a $\mathbf{C}_{i_0} \in \mathbf{P}$ such that $C_{i_{0x}} \not\subset C_{i_{0y}}$ and $C_{i_{0y}} \not\subset C_{i_{0x}}$, thus we have $\mathbf{P} \cap c_{ij} \neq \phi$. On the contrary, if $\mathbf{P} \cap c_{ij} \neq \phi$, for $x_i, x_j \in U$ satisfying $d(\Delta_{x_i}) \neq d(\Delta_{x_j})$, it means that there are enough covers in $\mathbf{P}$to keep the relation of x_i, x_j with respect to $\mathbf{\Delta}$ equivalent to the relation of x_i, x_j with respect to $\mathbf{P}$. Thus we know $\mathrm{COV}(\mathbf{P}) \leqslant U/D$.
(2) If $\mathbf{C} \in \mathrm{Core}_D(\Delta)$, then $\mathrm{Cov}(\Delta - \{\mathbf{C}\}) \leqslant U/D$ is not true, which implies there is at least a pair of $x_i, x_j \in U$ satisfying $d(\Delta_{x_i}) \neq d(\Delta_{x_j})$ whose original relation with respect to Δ changes after $\mathbf{C}$ is deleted from $\mathbf{\Delta}$. $\mathbf{C}$ is thus the only cover in $\mathbf{\Delta}$ satisfying $C_{x_i} \not\subset C_{x_j}$ and $C_{x_j} \not\subset C_{x_i}$ or $\mathbf{C}$ is the only cover in $\mathbf{\Delta}$ satisfying there is $\{\mathbf{C}_t : t = 1, \ldots, k\} \subseteq \Delta$ such that $C_{x_i} \subset C_{x_j}$ and $(C_t)_{x_i} \supset (C_t)_{x_j}$. By Definition 4.7, $c_{ij} = \{\mathbf{C}\}$ or $c_{ij} = \{\mathbf{C} \wedge \mathbf{C}_t : t = 1, \ldots, k\}$. Hence we have $\mathrm{Core}_D(\mathbf{\Delta}) \subseteq \{\mathbf{C} \in \mathbf{\Delta} : c_{ij} = \{\mathbf{C}\} \vee c_{ij} = \{\mathbf{C} \wedge \mathbf{C}_t : t = 1, \ldots, k\}\}$.

If $c_{ij} = \{\mathbf{C}\}$ or $c_{ij} = \{\mathbf{C} \wedge \mathbf{C}_t : t = 1, \ldots, k\}$ for some i, j, it is clear $\mathbf{C} \in \mathrm{Core}_D(\Delta)$. $\square$

Corollary 4.9. *Suppose $\mathbf{P} \subseteq \mathbf{\Delta}$, then $\mathbf{P}$ is a relative reduct of $\mathbf{\Delta}$ if and only if it is the minimal set satisfying $\mathbf{P} \cap c_{ij} \neq \phi$ for $i, j \leqslant n$.*

A discernibility function $f(U, \mathbf{\Delta}, D)$ for $(U, \mathbf{\Delta}, D)$ is a Boolean function of m Boolean variables $\overline{\mathbf{C}_1}, \ldots, \overline{\mathbf{C}_m}$ corresponding to covers $\mathbf{C}_1, \ldots, \mathbf{C}_m$, respectively, and defined as $f(U, \mathbf{\Delta}, D)(\overline{\mathbf{C}_1}, \ldots, \overline{\mathbf{C}_m}) = \wedge\{\vee(c_{ij}) : 1 \leqslant j < i \leqslant n,\}$ where $\vee(c_{ij})$ is the disjunction of all elements in c_{ij} as $\mathbf{C}$ or $\mathbf{C}_s \wedge \mathbf{C}_t$. By using the discernibility function, we have the following theorem to compute all the relative reducts.

Theorem 4.10. *Let $(U, \mathbf{\Delta}, D = \{d\})$ be a consistent covering decision system, $M(U, \mathbf{\Delta}, D) = (c_{ij} : i, j \leqslant n)$ is the discernibility matrix of $(U, \mathbf{\Delta}, D)$, $f(U, \mathbf{\Delta}, D)$ is the discernibility function of $(U, \mathbf{\Delta}, D)$. If $f(U, \mathbf{\Delta}, D) = \bigvee_{k=1}^{l} (\wedge \Delta_k)(\Delta_k \subseteq \mathbf{\Delta})$ is obtained from $f(U, \mathbf{\Delta}, D)$ by applying the multiplication and absorption laws as many times as possible such that every element in $\mathbf{\Delta}_i$ appears only one time, then the set $\{\mathbf{\Delta}_k : k \leqslant l\}$ is the* *collection of all the reductions of system* $(U,\mathbf{\Delta},D)$, *i.e., if* $\mathrm{Red}(\mathbf{\Delta},D)$ *is the collection of all the reducts of system* $(U,\mathbf{\Delta},D)$, *then* $\mathrm{Red}(\mathbf{\Delta},D)=\{\mathbf{\Delta}_1,\ldots,\mathbf{\Delta}_l\}$.

Proof. For every $k=1,\ldots,l$, we have $\wedge\mathbf{\Delta}_k\leqslant\vee c_{ij}$, so $\mathbf{\Delta}_k\cap c_{ij}\neq\phi$. Since $f(U,\mathbf{\Delta},D)=\overset{l}{\underset{k=1}{\vee}}(\wedge\Delta_k)$, for every Δ_k, if we reduce an element $\mathbf{C}$ in Δ_k, let $\mathbf{\Delta}'_k=\mathbf{\Delta}_k-\{\mathbf{C}\}$, then $f(U,\mathbf{\Delta},D)\neq\overset{k-1}{\underset{r=1}{\vee}}(\wedge\mathbf{\Delta}_r)\vee(\wedge\mathbf{\Delta}'_k)\vee\left(\overset{l}{\underset{r=k+1}{\vee}}\mathbf{\Delta}_r\right)$ and $f(U,\mathbf{\Delta},D)<\overset{k-1}{\underset{r=1}{\vee}}(\wedge\mathbf{\Delta}_r)\vee(\wedge\mathbf{\Delta}'_k)\vee\left(\overset{l}{\underset{r=k+1}{\vee}}\mathbf{\Delta}_r\right)$. If for every c_{ij}, we have $\mathbf{\Delta}'_k\cap c_{ij}\neq\phi$, then $\wedge\mathbf{\Delta}'_k\leqslant\vee c_{ij}$, which implies $f(U,\mathbf{\Delta},D)\geqslant\overset{k-1}{\underset{r=1}{\vee}}(\wedge\mathbf{\Delta}_r)\vee(\wedge\mathbf{\Delta}'_k)\vee\left(\overset{l}{\underset{r=k+1}{\vee}}\mathbf{\Delta}_r\right)$ and $f(U,\mathbf{\Delta},D)=\overset{k-1}{\underset{r=1}{\vee}}(\wedge\mathbf{\Delta}_r)\vee(\wedge\mathbf{\Delta}'_k)\vee\left(\overset{l}{\underset{r=k+1}{\vee}}\mathbf{\Delta}_r\right)$, which is a contradiction. Hence there exist $c_{i_0j_0}$ such that $\mathbf{\Delta}'_k\cap c_{i_0j_0}=\phi$ which implies $\mathbf{\Delta}_k$ is a reduct of $(U,\mathbf{\Delta},D)$.

For every $\mathbf{X}\in\mathrm{Red}(\mathbf{\Delta})$, we have $\mathbf{X}\cap c_{ij}\neq\phi$ for every c_{ij}, so we have $f(U,\mathbf{\Delta},D)\wedge(\wedge\mathbf{X})=\wedge(\vee c_{ij})\wedge(\wedge\mathbf{X})=\wedge\mathbf{X}$. This implies $\wedge\mathbf{X}\leqslant f(U,\mathbf{\Delta},D)$. Suppose that for every k we have $\mathbf{\Delta}_k-\mathbf{X}\neq\phi$. Then for every k one can find $\mathbf{C}_k\in\mathbf{\Delta}_k-\mathbf{X}$. By rewriting $f(U,\mathbf{\Delta},D)=(\vee_{k=1}^{l}\mathbf{C}_k)\wedge\Phi$, we have $\wedge\mathbf{X}\leqslant\vee_{k=1}^{l}\mathbf{C}_k$. So there is $\mathbf{C}_{k_0}$ such that $\wedge\mathbf{X}\leqslant\mathbf{C}_{k_0}$, i.e., $\mathbf{C}_{k_0}\in\mathbf{X}$, which is a contradiction. So $\mathbf{\Delta}_{k_0}\subseteq\mathbf{X}$ for some k_0. Since both $\mathbf{X}$ and $\mathbf{\Delta}_{k_0}$ are reducts, we have $\mathbf{X}=\mathbf{\Delta}_{k_0}$. Hence $\mathrm{RED}(\mathbf{\Delta})=\{\mathbf{\Delta}_1,\ldots,\mathbf{\Delta}_l\}$. $\square$

We have the following example to illustrate our idea in this section.

Example 4.11. Suppose $U=\{x_1,\ldots,x_9\}$, $\mathbf{\Delta}=\{\mathbf{C}_i:i=1,\ldots,4\}$, and

$$
\begin{aligned}
&\mathbf{C}_1=\{\{x_1,x_2,x_4,x_5,x_7,x_8\},\{x_2,x_3,x_5,x_6,x_8,x_9\}\},\\
&\mathbf{C}_2=\{\{x_1,x_2,x_3,x_4,x_5,x_6\},\{x_4,x_5,x_6,x_7,x_8,x_9\}\},\\
&\mathbf{C}_3=\{\{x_1,x_2,x_3\},\{x_4,x_5,x_6,x_7,x_8,x_9\},\{x_7,x_8,x_9\}\},\\
&\mathbf{C}_4=\{\{x_1,x_2,x_4,x_5\},\{x_2,x_3,x_5,x_6\},\{x_4,x_5,x_7,x_8\},\{x_5,x_6,x_8,x_9\}\},\\
&U/D=\{\{x_1,x_2,x_3\},\{x_4,x_5,x_6\},\{x_7,x_8x_9\}\},
\end{aligned}
$$

where Δ_i is Δ_{x_i} for short, $\mathbf{P}_i$ is $\mathbf{P}_{x_i}$ for short.

Then $\Delta_1=\{x_1,x_2\},\Delta_2=\{x_2\},\Delta_3=\{x_2,x_3\},\Delta_4=\{x_4,x_5\},\Delta_5=\{x_5\},\Delta_6=\{x_5,x_6\}$,

$$\Delta_7=\{x_7,x_8\},\Delta_8=\{x_8\}\Delta_9=\{x_8,x_9\}.$$

The discernibility matrix of $(U,\mathbf{\Delta},D)$ is as follows:

$$
\begin{pmatrix}
\mathbf{\Delta} & & & & & & & & \\
\mathbf{\Delta} & \mathbf{\Delta} & & & & & & & \\
\mathbf{\Delta} & \mathbf{\Delta} & \mathbf{\Delta} & & & & & & \\
\{\mathbf{C}_3\} & \{\mathbf{C}_1\wedge\mathbf{C}_2,\mathbf{C}_3,\mathbf{C}_4\} & \{\mathbf{C}_1,\mathbf{C}_3,\mathbf{C}_4\} & \mathbf{\Delta} & & & & & \\
\{\mathbf{C}_3\} & \{\mathbf{C}_3\} & \{\mathbf{C}_3\} & \mathbf{\Delta} & \mathbf{\Delta} & & & & \\
\{\mathbf{C}_1,\mathbf{C}_3,\mathbf{C}_4\} & \{\mathbf{C}_1\wedge\mathbf{C}_2,\mathbf{C}_3,\mathbf{C}_4\} & \{\mathbf{C}_3\} & \mathbf{\Delta} & \mathbf{\Delta} & \mathbf{\Delta} & & & \\
\{\mathbf{C}_2,\mathbf{C}_3,\mathbf{C}_4\} & \{\mathbf{C}_2,\mathbf{C}_3,\mathbf{C}_4\} & \{\mathbf{C}_1,\mathbf{C}_2,\mathbf{C}_3,\mathbf{C}_4\} & \{\mathbf{C}_2\wedge\mathbf{C}_3,\mathbf{C}_3\wedge\mathbf{C}_4\} & \{\mathbf{C}_1\wedge\mathbf{C}_3,\mathbf{C}_2\wedge\mathbf{C}_3,\mathbf{C}_3\wedge\mathbf{C}_4\} & \{\mathbf{C}_1,\mathbf{C}_2\wedge\mathbf{C}_3,\mathbf{C}_4\} & \mathbf{\Delta} & & \\
\{\mathbf{C}_2,\mathbf{C}_3,\mathbf{C}_4\} & \{\mathbf{C}_2,\mathbf{C}_3,\mathbf{C}_4\} & \{\mathbf{C}_2,\mathbf{C}_3,\mathbf{C}_4\} & \{\mathbf{C}_1\wedge\mathbf{C}_3,\mathbf{C}_2\wedge\mathbf{C}_3,\mathbf{C}_4\} & \{\mathbf{C}_2\wedge\mathbf{C}_3,\mathbf{C}_3\wedge\mathbf{C}_4\} & \{\mathbf{C}_1\wedge\mathbf{C}_3,\mathbf{C}_2\wedge\mathbf{C}_3,\mathbf{C}_4\} & \mathbf{\Delta} & \mathbf{\Delta} & \\
\{\mathbf{C}_1,\mathbf{C}_2,\mathbf{C}_3,\mathbf{C}_4\} & \{\mathbf{C}_2,\mathbf{C}_3,\mathbf{C}_4\} & \{\mathbf{C}_2,\mathbf{C}_3,\mathbf{C}_4\} & \{\mathbf{C}_1,\mathbf{C}_2\wedge\mathbf{C}_3,\mathbf{C}_4\} & \{\mathbf{C}_1\wedge\mathbf{C}_3,\mathbf{C}_2\wedge\mathbf{C}_3,\mathbf{C}_3\wedge\mathbf{C}_4\} & \{\mathbf{C}_2\wedge\mathbf{C}_3,\mathbf{C}_3\wedge\mathbf{C}_4\} & \mathbf{\Delta} & \mathbf{\Delta} & \mathbf{\Delta}
\end{pmatrix}
$$

and

$$
\begin{aligned}
f(U,\mathbf{\Delta})(\overline{\mathbf{C}_1},\ldots,\overline{\mathbf{C}_4})&=\wedge\{\vee(c_{ij}):1\leqslant j<i\leqslant 9\}\\
&=\mathbf{C}_3\wedge(\mathbf{C}_1\vee\mathbf{C}_3\vee\mathbf{C}_4)\wedge(\mathbf{C}_2\vee\mathbf{C}_3\vee\mathbf{C}_4)\wedge(\mathbf{C}_1\vee\mathbf{C}_2\vee\mathbf{C}_3\vee\mathbf{C}_4)\\
&\quad\wedge((\mathbf{C}_1\wedge\mathbf{C}_2)\vee\mathbf{C}_3\vee\mathbf{C}_4)\wedge((\mathbf{C}_2\wedge\mathbf{C}_3)\vee(\mathbf{C}_3\wedge\mathbf{C}_4))\wedge((\mathbf{C}_1\wedge\mathbf{C}_2)\vee(\mathbf{C}_2\wedge\mathbf{C}_3)\vee\mathbf{C}_4)\\
&\quad\wedge(\mathbf{C}_1\vee(\mathbf{C}_2\wedge\mathbf{C}_3)\vee\mathbf{C}_4)\wedge((\mathbf{C}_1\wedge\mathbf{C}_3)\vee(\mathbf{C}_2\wedge\mathbf{C}_3)\vee(\mathbf{C}_3\wedge\mathbf{C}_4))\\
&=\mathbf{C}_3\wedge((\mathbf{C}_2\wedge\mathbf{C}_3)\vee(\mathbf{C}_3\wedge\mathbf{C}_4))\wedge((\mathbf{C}_1\wedge\mathbf{C}_2)\vee(\mathbf{C}_2\wedge\mathbf{C}_3)\vee\mathbf{C}_4)\\
&=\mathbf{C}_3\wedge(\mathbf{C}_3\wedge(\mathbf{C}_2\vee\mathbf{C}_4))\wedge([\mathbf{C}_2\wedge(\mathbf{C}_1\vee\mathbf{C}_3)]\vee\mathbf{C}_4)\\
&=\mathbf{C}_3\wedge(\mathbf{C}_2\vee\mathbf{C}_4)\wedge((\mathbf{C}_2\vee\mathbf{C}_4)\wedge(\mathbf{C}_1\vee\mathbf{C}_3\vee\mathbf{C}_4))\\
&=\mathbf{C}_3\wedge(\mathbf{C}_2\vee\mathbf{C}_4)\wedge(\mathbf{C}_1\vee\mathbf{C}_3\vee\mathbf{C}_4)
\end{aligned}
$$

$$= \mathbf{C}_3 \wedge (\mathbf{C}_2 \vee \mathbf{C}_4)$$
$$= (\mathbf{C}_3 \wedge \mathbf{C}_4) \vee (\mathbf{C}_2 \wedge \mathbf{C}_3).$$

So $\mathrm{Red}(\boldsymbol{\Delta}) = \{\{\mathbf{C}_3, \mathbf{C}_4\}, \{\mathbf{C}_2, \mathbf{C}_3\}\}$, $\mathrm{Core}_D(\boldsymbol{\Delta}) = \{\mathbf{C}_3\}$.

5. Attribute reduction of inconsistent covering decision systems

We always have inconsistent covering decision systems in many of practical problems. In this section we propose attribute reductions for inconsistent covering decision systems. By the definition of inconsistent covering decision systems, we know that some rules extracted from them may not be consistent. Similar to the case of traditional rough sets [18,25], we define its attribute reduct as the minimal set of conditional attributes to keep the positive region of the decision attribute invariant. We have thus the following definition of attributes reduction. In the following we always suppose U is a finite universe and $\boldsymbol{\Delta} = \{\mathbf{C}_i : i = 1, \ldots, m\}$ is a family of covers of U. Then the induced cover of $\boldsymbol{\Delta}$ is defined as $\mathrm{Cov}(\boldsymbol{\Delta}) = \{\Delta_x : x \in U\}$, D is a decision attribute, $U/D = \{[x]_D : x \in U\}$ is the decision partition. We always suppose $\mathrm{POS}_{\boldsymbol{\Delta}}(D) \neq \phi$ in this section.

Definition 5.1. Suppose U is a finite universe and $\boldsymbol{\Delta} = \{\mathbf{C}_i : i = 1, \ldots, m\}$ is a family of covers of U, $\mathbf{C}_i \in \boldsymbol{\Delta}$, D is a decision attribute relative to $\boldsymbol{\Delta}$ on U and $d : U \to V_d$ is the decision function V_d defined as $d(x) = [x]_D$, $(U, \boldsymbol{\Delta}, D)$ is an inconsistent covering decision system, i.e., $\mathrm{POS}_\Delta(D) \neq U$. If $\mathrm{POS}_\Delta(D) = \mathrm{POS}_{\Delta-\{\mathbf{C}_i\}}(D)$, then $\mathbf{C}_i$ is superfluous relative to D in $\boldsymbol{\Delta}$. Otherwise $\mathbf{C}_i$ is indispensable relative to D in $\boldsymbol{\Delta}$. For every $\mathbf{P} \subseteq \boldsymbol{\Delta}$, if every element in $\mathbf{P}$ is indispensable relative to D, and $\mathrm{POS}_{\mathbf{P}}(D) = \mathrm{POS}_\Delta(D)$, then $\mathbf{P}$ is a reduct of $\boldsymbol{\Delta}$ relative to D, called relative reduct in short. The collection of all the indispensable elements relative to D in $\boldsymbol{\Delta}$ is the core of $\boldsymbol{\Delta}$ relative to D, denoted by $\mathrm{Core}_D(\boldsymbol{\Delta})$.

Theorem 5.2. *Inconsistent covering decision system* $(U, \boldsymbol{\Delta}, D = \{d\})$ *has the following properties*:

(1) *For* $\forall x_i \in U$, *if* $\Delta_{x_i} \subset \mathrm{POS}_\Delta(D)$, *then* $\Delta_{x_i} \subseteq [x_i]_D$; *if* $\Delta_{x_i} \not\subset \mathrm{POS}_\Delta(D)$, *then for* $\forall x_k \in U$, $\Delta_{x_i} \subseteq [x_k]_D$ *is not true.*
(2) *For any* $\mathbf{P} \subseteq \boldsymbol{\Delta}$, $\mathrm{POS}_{\mathbf{P}}(D) = \mathrm{POS}_\Delta(D)$ *if and only if* $\underline{\mathbf{P}}(X) = \underline{\boldsymbol{\Delta}}(X)$ *for* $\forall X \in U/D$.
(3) *For any* $\mathbf{P} \subseteq \boldsymbol{\Delta}$, $\mathrm{POS}_{\mathbf{P}}(D) = \mathrm{POS}_\Delta(D)$ *if and only if* $\forall x_i \in U$, $\Delta_{x_i} \subseteq [x_i]_D \iff \mathbf{P}_{x_i} \subseteq [x_i]_D$.

Proof

(1) Since $\mathrm{POS}_\Delta(D) = \underset{X \in U/D}{\cup} \underline{\boldsymbol{\Delta}}(X)$ and $\underline{\boldsymbol{\Delta}}(X) = \cup\{\Delta_x : \Delta_x \subseteq X\}$, if $\Delta_{x_i} \subset \mathrm{POS}_\Delta(D)$, then there exists $x_j \in U$, such that $\Delta_{x_i} \subseteq [x_j]_D$; since $x_i \in \Delta_{x_i}$, then $x_i \in [x_j]_D$, thus $[x_i]_D = [x_j]_D$, so we have $\Delta_{x_i} \subseteq [x_i]_D$.
If $\Delta_{x_i} \not\subset \mathrm{POS}_\Delta(D)$, suppose $\exists x_0 \in U$, such that $\Delta_{x_i} \subseteq [x_0]_D$, then $\Delta_{x_i} \subset \mathrm{POS}_\Delta(D)$, which is a contradiction. Thus for $\forall x_k \in U$, $\Delta_{x_i} \subseteq [x_k]_D$ is not true.
(2) Since $\mathbf{P} \subseteq \boldsymbol{\Delta}$, then $\forall x_i \in U$, we have $\Delta_{x_i} \subseteq \mathbf{P}_{x_i}$; then by $\underline{\boldsymbol{\Delta}}(X) = \cup\{\Delta_x : \Delta_x \subseteq X\}$ and $\underline{\mathbf{P}}(X) = \cup\{\mathbf{P}_x : \mathbf{P}_x \subseteq X\}$, we have $\underline{\mathbf{P}}(X) \subseteq \underline{\boldsymbol{\Delta}}(X)$ for $\forall X \in U/D$. If $\exists X_0 \in U/D$ such that $\underline{\mathbf{P}}(X_0) \subset \underline{\boldsymbol{\Delta}}(X_0)$, then $\mathrm{POS}_{\mathbf{P}}(D) \subset \mathrm{POS}_\Delta(D)$, which is a contradiction, hence the result is true.
The converse is obviously true.
(3) If $\mathrm{POS}_{\mathbf{P}}(D) = \mathrm{POS}_\Delta(D)$, then $\forall X \in U/D$, $\underline{\mathbf{P}}(X) = \underline{\boldsymbol{\Delta}}(X)$, hence for $\forall x_i \in U$ satisfies $\Delta_{x_i} \subseteq [x_i]_D$, since $\underline{\mathbf{P}}([x_i]_D) = \underline{\boldsymbol{\Delta}}([x_i]_D)$, we have $x_i \in \underline{\mathbf{P}}([x_i]_D)$, so $\mathbf{P}_{x_i} \subseteq [x_i]_D$; it is clear if $\mathbf{P}_{x_i} \subseteq [x_i]_D$, then $\Delta_{x_i} \subseteq [x_i]_D$.

If $\forall x_i \in U$, $\Delta_{x_i} \subseteq [x_i]_D \iff \mathbf{P}_{x_i} \subseteq [x_i]_D$, by $\underline{\boldsymbol{\Delta}}(X) = \cup\{\Delta_x : \Delta_x \subseteq X\}$ and $\underline{\mathbf{P}}(X) = \cup\{\mathbf{P}_x : \mathbf{P}_x \subseteq X\}$, we can have $\underline{\mathbf{P}}(X) = \underline{\boldsymbol{\Delta}}(X)$, hence $\mathrm{POS}_{\mathbf{P}}(D) = \mathrm{POS}_\Delta(D)$. $\square$

Theorem 5.2 implies that for every $x_i, x_j \in U$, if $\Delta_{x_i} \subset \mathrm{POS}_\Delta(D)$ and $\Delta_{x_j} \not\subset \mathrm{POS}_\Delta(D)$, then either $\Delta_{x_i} \subset \Delta_{x_j}$ or $\Delta_{x_i} \not\subset \Delta_{x_j}$, $\Delta_{x_j} \not\subset \Delta_{x_i}$ are possible. If $\Delta_{x_i}, \Delta_{x_j} \subset \mathrm{POS}_\Delta(D)$ and $[x_i]_D \neq [x_j]_D$, then $\Delta_{x_i} \not\subset \Delta_{x_j}$ and $\Delta_{x_j} \not\subset \Delta_{x_i}$.

Theorem 5.3. *Suppose* $(U, \boldsymbol{\Delta}, D = \{d\})$ *is an inconsistent covering decision system*, $\mathbf{C}_i \in \Delta$, *if* $\mathrm{POS}_\Delta(D) \neq \mathrm{POS}_{\Delta-\{\mathbf{C}_i\}}(D)$, *then there is at least a pair of* $x_i, x_j \in U$ *satisfy one case of the following conditions* (1) *and* (2), *and their original relation with respect to* $\boldsymbol{\Delta}$ *changes after* $\mathbf{C}_i$ *is deleted from* $\boldsymbol{\Delta}$:

(1) $\Delta_{x_i} \subset \mathrm{POS}_\Delta(D)$ *and* $\Delta_{x_j} \not\subset \mathrm{POS}_\Delta(D)$;
(2) $\Delta_{x_i}, \Delta_{x_j} \subset \mathrm{POS}_\Delta(D)$ *and* $[x_i]_D \cap [x_j]_D = \phi$.

Proof. We denote $\mathbf{P} = \Delta - \{\mathbf{C}_i\}$, $\mathrm{Cov}(\Delta - \{\mathbf{C}_i\}) = \{\mathbf{P}_x : x \in U\}$,

Since $\mathrm{POS}_\Delta(D) \neq \mathrm{POS}_{\Delta-\{\mathbf{C}_i\}}(D)$, then there exists $X_0 \in U/D$ such that $\underline{\mathbf{P}}(X_0) \subset \underline{\mathbf{\Delta}}(X_0)$, by $\underline{\mathbf{P}}(X_0) = \cup\{\mathbf{P}_x : \mathbf{P}_x \subseteq X_0\}$ and $\underline{\mathbf{\Delta}}(X_0) = \cup\{\Delta_x : \Delta_x \subseteq X_0\}$, it follows that $\exists x_0 \in U$ such that $\Delta_{x_0} \subseteq [x_0]_D = X_0$ holds, and $\mathbf{P}_{x_0} \subseteq [x_0]_D = X_0$ is not true. For$y_0 \in \mathbf{P}_{x_0}$ and $y_0 \notin [x_0]_D$, if $\Delta_{y_0} \subseteq [y_0]_D$, then the relationship of x_0, y_0 relative to Δ is suitable for case (2), i.e., $\Delta_{x_0} \not\subset \Delta_{y_0}$ and $\Delta_{y_0} \not\subset \Delta_{x_0}$ holds. If $\Delta_{y_0} \subseteq [y_0]_D$ is not true, the relationship of x_0, y_0 relative to Δ is suitable for case (1), i.e., either $\Delta_{x_0} \subset \Delta_{y_0}$or $\Delta_{x_0} \not\subset \Delta_{y_0}$ and $\Delta_{y_0} \not\subset \Delta_{x_0}$ holds, but $x_0, y_0 \in \mathbf{P}_{x_0}$, i.e., $\mathbf{P}_{y_0} \subseteq \mathbf{P}_{x_0}$, so no matter x_0, y_0 are suitable for any given case, the relation of x_0, y_0 with respect to $\mathbf{\Delta}$ is not equivalent to the relation of x_0, y_0 with respect to $\mathbf{P}$,i.e., their original relation with respect to $\mathbf{\Delta}$ changes after $\mathbf{C}_i$ is deleted from $\mathbf{\Delta}$. □

It should be pointed out that if there is at least a pair of $x_0, y_0 \in U$ satisfying (2) in Theorem 5.3, i.e., $\Delta_{x_0} \not\subset \Delta_{y_0}$ and $\Delta_{y_0} \not\subset \Delta_{x_0}$, which original relation with respect to $\mathbf{\Delta}$ changes after $\mathbf{C}_i$ is deleted from $\mathbf{\Delta}$, which means $\mathbf{P}_{y_0} \not\subset \mathbf{P}_{x_0}$ and $\mathbf{P}_{x_0} \not\subset \mathbf{P}_{y_0}$ does not hold. So if $\mathbf{P}_{x_0} \subseteq \mathbf{P}_{y_0}$, then $\mathbf{P}_{y_0} \subseteq [y_0]_D$ does not hold; if $\mathbf{P}_{x_0} \supseteq \mathbf{P}_{y_0}$, then $\mathbf{P}_{x_0} \subseteq [x_0]_D$ does not hold; if $\mathbf{P}_{y_0} = \mathbf{P}_{x_0}$, then both $\mathbf{P}_{y_0} \subseteq [y_0]_D$ and $\mathbf{P}_{x_0} \subseteq [x_0]_D$ do not hold, by (3) of Theorem 5.2, we have $\mathrm{POS}_\Delta(D) \neq \mathrm{POS}_{\Delta-\{\mathbf{C}_i\}}(D)$, which implies if $x_0, y_0 \in U$ satisfy (2) in Theorem 5.3, and their original relation with respect to $\mathbf{\Delta}$ changes after $\mathbf{C}_i$ is deleted from $\mathbf{\Delta}$, $\mathbf{C}_i$ must be an indispensable element of $\mathbf{\Delta}$ relative to D. If $x_0, y_0 \in U$ satisfy (1) of Theorem 5.3, we suppose $\Delta_{x_0} \subset \mathrm{POS}_\Delta(D)$ and $\Delta_{y_0} \not\subset \mathrm{POS}_\Delta(D)$, then we have $\Delta_{x_0} \subset \Delta_{y_0}$or $\Delta_{x_0} \not\subset \Delta_{y_0}$ and $\Delta_{y_0} \not\subset \Delta_{x_0}$. If their original relation with respect to $\mathbf{\Delta}$ changes after $\mathbf{C}_i$ is deleted from $\mathbf{\Delta}$, this means if $\Delta_{x_0} \subset \Delta_{y_0}$ holds, then $\mathbf{P}_{x_0} \subset \mathbf{P}_{y_0}$ does not hold, by (2) of Theorem 3.1 this means only $\mathbf{P}_{x_0} = \mathbf{P}_{y_0}$ holds. Since $\Delta_{y_0} \subseteq \mathbf{P}_{y_0}$, then $\mathbf{P}_{y_0} \not\subset \mathrm{POS}_\Delta(D)$, thus $\mathbf{P}_{x_0} \not\subset \mathrm{POS}_\Delta(D)$, so we also have $\mathrm{POS}_\Delta(D) \neq \mathrm{POS}_{\Delta-\{\mathbf{C}_i\}}(D)$. If $\Delta_{x_0} \not\subset \Delta_{y_0}$ and $\Delta_{y_0} \not\subset \Delta_{x_0}$ hold, while $\mathbf{P}_{y_0} \not\subset \mathbf{P}_{x_0}$ and $\mathbf{P}_{x_0} \not\subset \mathbf{P}_{y_0}$ do not hold, which implys only $\mathbf{P}_{x_0} \subset \mathbf{P}_{y_0}$, or $\mathbf{P}_{x_0} \supset \mathbf{P}_{y_0}$, or $\mathbf{P}_{x_0} = \mathbf{P}_{y_0}$ holds, if $\mathbf{P}_{x_0} \supset \mathbf{P}_{y_0}$, or $\mathbf{P}_{x_0} = \mathbf{P}_{y_0}$ holds, since $\Delta_{y_0} \subseteq \mathbf{P}_{y_0}$, then $\mathbf{P}_{y_0} \not\subset \mathrm{POS}_\Delta(D)$, thus $\mathbf{P}_{x_0} \not\subset \mathrm{POS}_\Delta(D)$, so we also have $\mathrm{POS}_\Delta(D) \neq \mathrm{POS}_{\Delta-\{\mathbf{C}_i\}}(D)$; but if $\mathbf{P}_{x_0} \subset \mathbf{P}_{y_0}$ holds, $\mathrm{POS}_\Delta(D) \neq \mathrm{POS}_{\Delta-\{\mathbf{C}_i\}}(D)$ may not hold.

So we can say if $x_0, y_0 \in U$ satisfy (1) in Theorem 5.3, although their original relation with respect to $\mathbf{\Delta}$ may change after $\mathbf{C}_i$ is deleted from $\mathbf{\Delta}$, $\mathbf{C}_i$ may not be an indispensable element of $\mathbf{\Delta}$ relative to D (see Example 5.9). Clearly in this case $\mathbf{C}_i$is an indispensable element of $\mathbf{\Delta}$ relative to$D$ if and only if $\mathbf{P}_{y_0} \subseteq \mathbf{P}_{x_0}$ holds. For every $x_i, x_j \in U$, if $\Delta_{x_i} \subset \mathrm{POS}_\Delta(D)$ and $\Delta_{x_j} \not\subset \mathrm{POS}_\Delta(D)$, then either $\Delta_{x_i} \subset \Delta_{x_j}$ and $\Delta_{x_i} \not\subset \Delta_{x_j}$, $\Delta_{x_j} \not\subset \Delta_{x_i}$ are possible. Suppose $\mathbf{P} \subseteq \mathbf{\Delta}$, if $\mathrm{POS}_{\mathbf{P}}(D) = \mathrm{POS}_\Delta(D)$, then $\mathbf{P}_{x_i} \subset \mathrm{POS}_\Delta(D)$, $\mathbf{P}_{x_j} \not\subset \mathrm{POS}_\Delta(D)$ hold, thus we also have both $\mathbf{P}_{x_i} \subset \mathbf{P}_{x_j}$ and $\mathbf{P}_{x_i} \not\subset \mathbf{P}_{x_j}$, $\mathbf{P}_{x_j} \not\subset \mathbf{P}_{x_i}$ are possible. If $\Delta_{x_i} \subset \Delta_{x_j}$ holds, we have $\mathbf{P}_{x_i} \subset \mathbf{P}_{x_j}$; if $\Delta_{x_i} \not\subset \Delta_{x_j}$, $\Delta_{x_j} \not\subset \Delta_{x_i}$ holds, then either $\mathbf{P}_{x_i} \subset \mathbf{P}_{x_j}$or $\mathbf{P}_{x_i} \not\subset \mathbf{P}_{x_j}$ and $\mathbf{P}_{x_j} \not\subset \mathbf{P}_{x_i}$ are possible (see Example 5.9).

From the above discussion we have Theorem 5.4.

Theorem 5.4. *Suppose* $(U, \mathbf{\Delta}, D = \{d\})$ *is an inconsistent covering decision system.* $\mathbf{P} \subseteq \mathbf{\Delta}$, *then* $\mathrm{POS}_\Delta(D) = \mathrm{POS}_{\mathbf{P}}(D)$ *if and only if*

(1) *For* $x_i, x_j \in U$ *satisfying* $\Delta_{x_i} \subset \mathrm{POS}_\Delta(D)$ *and* $\Delta_{x_j} \not\subset \mathrm{POS}_\Delta(D)$, *both* $\Delta_{x_i} \subset \Delta_{x_j} \Rightarrow \mathbf{P}_{x_i} \subset \mathbf{P}_{x_j}$ *and* $\Delta_{x_i} \not\subset \Delta_{x_j}$, $\Delta_{x_j} \not\subset \Delta_{x_i} \Rightarrow \mathbf{P}_{x_i} \subset \mathbf{P}_{x_j}$ *or* $\mathbf{P}_{x_i} \not\subset \mathbf{P}_{x_j}$, $\mathbf{P}_{x_j} \not\subset \mathbf{P}_{x_i}$ *hold.*
(2) *For* $x_i, x_j \in U$ *satisfying* $\Delta_{x_i} \subset \mathrm{POS}_\Delta(D)$, $\Delta_{x_j} \subset \mathrm{POS}_\Delta(D)$ *and* $[x_i]_D \cap [x_j]_D = \phi$, $\Delta_{x_i} \not\subset \Delta_{x_j}$, $\Delta_{x_j} \not\subset \Delta_{x_i} \Longleftrightarrow \mathbf{P}_{x_i} \not\subset \mathbf{P}_{x_j}$, $\mathbf{P}_{x_j} \not\subset \mathbf{P}_{x_i}$ *holds.*

Proof

⇒(1) If $\Delta_{x_i} \subset \Delta_{x_j}$holds, by (2) of Theorem 3.8, it follows that $\mathbf{P}_{x_i} \subseteq \mathbf{P}_{x_j}$ holds; if $\mathbf{P}_{x_i} = \mathbf{P}_{x_j}$, since $\Delta_{x_j} \not\subset \mathrm{POS}_\Delta(D)$, then $\mathbf{P}_{x_j} \not\subset \mathrm{POS}_\Delta(D)$, thus $\mathbf{P}_{x_i} \not\subset \mathrm{POS}_\Delta(D)$, by (3) of Theorem 5.2, we have $\mathrm{POS}_\Delta(D) \neq \mathrm{POS}_{\mathbf{P}}(D)$, which is a contradiction. So $\Delta_{x_i} \subset \Delta_{x_j} \Rightarrow \mathbf{P}_{x_i} \subset \mathbf{P}_{x_j}$ holds. If $\Delta_{x_i} \not\subset \Delta_{x_j}$, $\Delta_{x_j} \not\subset \Delta_{x_i}$, since $\mathrm{POS}_\Delta(D) = \mathrm{POS}_{\mathbf{P}}(D)$, then $\mathbf{P}_{x_i} \subset \mathrm{POS}_{\mathbf{P}}(D)$ and $\mathbf{P}_{x_j} \not\subset \mathrm{POS}_{\mathbf{P}}(D)$ hold, which implies $\mathbf{P}_{x_i} \subset \mathbf{P}_{x_j}$or $\mathbf{P}_{x_i} \not\subset \mathbf{P}_{x_j}$, $\mathbf{P}_{x_j} \not\subset \mathbf{P}_{x_i}$ holds, thus $\Delta_{x_i} \not\subset \Delta_{x_j}$, $\Delta_{x_j} \not\subset \Delta_{x_i} \Rightarrow \mathbf{P}_{x_i} \subset \mathbf{P}_{x_j}or\mathbf{P}_{x_i} \not\subset \mathbf{P}_{x_j}$, $\mathbf{P}_{x_j} \not\subset \mathbf{P}_{x_i}$ hold.
(2) Clearly it is true.

$\Leftarrow$ Suppose $\mathrm{POS}_\Delta(D) \neq \mathrm{POS}_{\mathbf{P}}(D)$, by the proof of Theorem 5.3, there must exist $x_i, x_j \in U$ satisfying $\Delta_{x_i} \subset \mathrm{POS}_\Delta(D)$ and $\Delta_{x_j} \not\subset \mathrm{POS}_\Delta(D)$, or $\Delta_{x_i}, \Delta_{x_j} \subset \mathrm{POS}_\Delta(D)$ and $[x_i]_D \cap [x_j]_D = \phi$ such that $\mathbf{P}_{x_i} \supseteq \mathbf{P}_{x_j}$, which is a contradiction under the given condition. $\square$

From the discussion above we know that to make $\mathbf{P} \subset \mathbf{\Delta}$ a relative reduct we should pay attention to the following statements: (1) for $\Delta_{x_i} \subset \mathrm{POS}_\Delta(D)$ and $\Delta_{x_j} \not\subset \mathrm{POS}_\Delta(D)$, if $\Delta_{x_i} \subset \Delta_{x_j}$, we should keep $\mathbf{P}_{x_i} \subset \mathbf{P}_{x_j}$; if $(\Delta_{x_i} \not\subset \Delta_{x_j}) \wedge (\Delta_{x_j} \not\subset \Delta_{x_i})$, we should ensure the original relation between $x_i, x_j \in U$ does not change to $\mathbf{P}_{x_i} \supseteq \mathbf{P}_{x_j}$. (2) for $\Delta_{x_i}, \Delta_{x_j} \subset \mathrm{POS}_\Delta(D)$ and $[x_i]_D \cap [x_j]_D = \phi$, we should keep the original relation between $x_i, x_j \in U$ invariant. So by Theorem 3.8 we have the following definition.

Definition 5.5. Let $(U, \mathbf{\Delta}, D = \{d\})$ be an inconsistent covering decision system. Suppose $U = \{x_1, \ldots, x_n\}$, by $M(U, \mathbf{\Delta}, D)$ we denote a $n \times n$ matrix (c_{ij}), called the discernibility matrix of $(U, \mathbf{\Delta}, D)$, such that if $x_i, x_j \in U$ satisfies:

(1) $\Delta_{x_i} \subset \mathrm{POS}_\Delta(D)$ and $\Delta_{x_j} \not\subset \mathrm{POS}_\Delta(D)$, then

$$c_{ij} = \begin{cases} \{\mathbf{C} \in \mathbf{\Delta} : C_{x_i} \subset C_{x_j}\}, & \Delta_{x_i} \subset \Delta_{x_j}; \\ \{\mathbf{C} \in \mathbf{\Delta} : C_{x_i} \subset C_{x_j}\} \cup \{\mathbf{C} \in \mathbf{\Delta} : (C_{x_i} \not\subset C_{x_j}) \wedge (C_{x_j} \not\subset C_{x_i})\} & \\ \quad \cup \{\mathbf{C}_s \wedge \mathbf{C}_t : (C_{s_{x_i}} \subset C_{s_{x_j}}) \wedge (C_{t_{x_j}} \subset C_{t_{x_i}})\}, & \Delta_{x_i} \not\subset \Delta_{x_j} \wedge \Delta_{x_j} \not\subset \Delta_{x_i}. \end{cases}$$

If $x_i, x_j \in U$ satisfies:

(2) $\Delta_{x_i}, \Delta_{x_j} \subset \mathrm{POS}_\Delta(D)$ and $[x_i]_D \cap [x_j]_D = \phi$, then

$$c_{ij} = \{\mathbf{C} \in \mathbf{\Delta} : (C_{x_i} \not\subset C_{x_j}) \wedge (C_{x_j} \not\subset C_{x_i})\} \cup \{\mathbf{C}_s \wedge \mathbf{C}_t : (C_{s_{x_i}} \subset C_{s_{x_j}}) \wedge (C_{t_{x_j}} \subset C_{t_{x_i}})\}.$$

(3) otherwise $c_{ij} = \mathbf{\Delta}$.

$\mathbf{C} \in c_{ij} \neq \mathbf{\Delta}$ implies $\mathbf{C}$ is one of the covers to ensure (1) for $\Delta_{x_i} \subset \mathrm{POS}_\Delta(D)$ and $\Delta_{x_j} \not\subset \mathrm{POS}_\Delta(D)$, if $\Delta_{x_i} \subset \Delta_{x_j}$, then $\mathbf{C}_{x_i} \subset \mathbf{C}_{x_j}$; if $(\Delta_{x_i} \not\subset \Delta_{x_j}) \wedge (\Delta_{x_j} \not\subset \Delta_{x_i})$, then $\mathbf{C}_{x_i} \supseteq \mathbf{C}_{x_j}$ do not hold. (2) for $\Delta_{x_i}, \Delta_{x_j} \subset \mathrm{POS}_\Delta(D)$ and $[x_i]_D \cap [x_j]_D = \phi$, then $\mathbf{C}$ keep the original relation between $x_i, x_j \in U$ invariant. Since $M(U, \mathbf{\Delta}, D)$ is symmetric and $c_{ii} = \mathbf{\Delta}$, for $i = 1, \ldots, n$, we represent $M(U, \mathbf{\Delta}, D)$ only by the elements in the lower triangle of $M(U, \mathbf{\Delta}, D)$, i.e., the c_{ij}'s with $1 \leqslant j \leqslant i \leqslant n$.

Theorem 5.6. *Let $(U, \mathbf{\Delta}, D = \{d\})$ be an inconsistent covering decision system, then we have*:

(1) *For any $\mathbf{P} \subseteq \mathbf{\Delta}$, $\mathbf{P} \cap c_{ij} \neq \phi$ holds for every $i, j \leqslant n$ if and only if $\mathrm{POS}_{\mathbf{P}}(D) = \mathrm{POS}_\Delta(D)$. Here $\mathbf{P} \cap c_{ij} \neq \phi$ means if $\Delta_{x_i} \subset \mathrm{POS}_\Delta(D)$ and $\Delta_{x_j} \not\subset \mathrm{POS}_\Delta(D)$, $\mathbf{C}_s \wedge \mathbf{C}_t \in c_{ij}$, $\mathbf{C}_t \in \mathbf{P} \cap c_{ij}$ satisfies $C_{tx_j} \subset C_{tx_i}$, then $\mathbf{C}_s \in \mathbf{P} \cap c_{ij}$ holds.*

(2) $\mathrm{Core}_D(\mathbf{\Delta}) = \{\mathbf{C} \in \mathbf{\Delta} : c_{ij} = \{\mathbf{C}\} \vee c_{ij} = \{\mathbf{C} \wedge \mathbf{C}_t : t = 1, \ldots, k\}\}$ *for some i, j.*

Proof

(1) $\Leftarrow$ Suppose $\mathrm{POS}_{\mathbf{P}}(D) = \mathrm{POS}_\Delta(D)$, $\forall i, j \leqslant n$, if x_i and x_j are suitable for (3) in Definition 5.5, then $c_{ij} = \mathbf{\Delta}$, hence $\mathbf{P} \cap c_{ij} \neq \phi$. Otherwise suppose $\exists i_0, j_0 \leqslant n$, satisfying $\mathbf{P} \cap c_{i_0 j_0} = \phi (c_{i_0 j_0} \neq \phi)$.
Suppose x_{i_0}, x_{j_0} satisfy (1) in Definition 5.5. If $\Delta_{x_{i_0}} \subset \Delta_{x_{j_0}}$, then by (2) of Theorem 3.1 we have for $\forall \mathbf{C} \in \mathbf{P}$, $C_{x_{i_0}} \not\subset C_{x_{j_0}}$, which implies $C_{x_{i_0}} \supseteq C_{x_{j_0}}$, then $\mathbf{P}_{x_{i_0}} \supseteq \mathbf{P}_{x_{j_0}}$ holds, by Theorem 5.4 $\mathrm{POS}_\Delta(D) \neq \mathrm{POS}_{\mathbf{P}}(D)$ holds, which is a contradiction. If $\Delta_{x_{i_0}} \not\subset \Delta_{x_{j_0}}$, $\Delta_{x_{j_0}} \not\subset \Delta_{x_{i_0}}$, by (3) of Theorem 3.8 we have for $\forall \mathbf{C} \in \mathbf{P}$, $C_{x_{i_0}} \not\subset C_{x_{j_0}}$ and $C_{x_{j_0}} \not\subset C_{x_{i_0}}$ do not hold; $\forall \mathbf{C}_s, \mathbf{C}_t \in \mathbf{P}$, $C_{sx_{i_0}} \subset C_{sx_{j_0}}$ and $C_{tx_{j_0}} \subset C_{tx_{i_0}}$ do not hold; and $\forall \mathbf{C} \in \mathbf{P}$, $C_{x_{i_0}} \subset C_{x_{j_0}}$ does not hold either, which implies $\forall \mathbf{C} \in \mathbf{P}$, we only have $C_{x_{i_0}} \supseteq C_{x_{j_0}}$, so $\mathbf{P}_{x_{i_0}} \supseteq \mathbf{P}_{x_{j_0}}$, by Theorem 5.4 we have $\mathrm{POS}_\Delta(D) \neq \mathrm{POS}_{\mathbf{P}}(D)$, which is a contradiction. If x_{i_0}, x_{j_0} satisfy (2) in Definition 5.5, similar to the above proof we can also get a contradiction, thus for every $i, j \leqslant n$, $\mathbf{P} \cap c_{ij} \neq \phi$ holds.
$\Rightarrow$For $\Delta_{x_i} \subset \mathrm{POS}_\Delta(D)$ and $\Delta_{x_j} \not\subset \mathrm{POS}_\Delta(D)$, if $\Delta_{x_i} \subset \Delta_{x_j}$, since $\mathbf{P} \subseteq \mathbf{\Delta}$, $c_{ij} = \{\mathbf{C} \in \Delta : C_{x_i} \subset C_{x_j}\}$, and $\mathbf{P} \cap c_{ij} \neq \phi (c_{ij} \neq \phi)$, which implies, $\mathbf{P}_{x_i} \subset \mathbf{P}_{x_j}$i.e., $\Delta_{x_i} \subset \Delta_{x_j} \Rightarrow \mathbf{P}_{x_i} \subset \mathbf{P}_{x_j}$ is true.If $\Delta_{x_i} \not\subset \Delta_{x_j}$, $\Delta_{x_j} \not\subset \Delta_{x_i}$, since

$c_{ij} = \{\mathbf{C} \in \Delta : C_{x_i} \subset C_{x_j}\} \cup \{\mathbf{C} \in \Delta : (C_{x_i} \not\subset C_{x_j}) \wedge (C_{x_j} \not\subset C_{x_i})\} \cup \{\mathbf{C}_s \wedge \mathbf{C}_t : (C_{s_{x_i}} \subset C_{s_{x_j}}) \wedge (C_{t_{x_j}} \subset C_{t_{x_i}})\}$, and $\mathbf{P} \cap c_{ij} \neq \phi (c_{ij} \neq \phi)$, then $\mathbf{P}_{x_i} \subset \mathbf{P}_{x_j}$ or $\mathbf{P}_{x_i} \not\subset \mathbf{P}_{x_j}$, $\mathbf{P}_{x_j} \not\subset \mathbf{P}_{x_i}$ holds, i.e., $\Delta_{x_i} \not\subset \Delta_{x_j}$, $\Delta_{x_j} \not\subset \Delta_{x_i} \Rightarrow \mathbf{P}_{x_i} \subset \mathbf{P}_{x_j}$ or $\mathbf{P}_{x_i} \not\subset \mathbf{P}_{x_j}$, $\mathbf{P}_{x_j} \not\subset \mathbf{P}_{x_i}$, by Theorem 5.4 we know $\mathrm{POS}_{\mathbf{P}}(D) = \mathrm{POS}_{\Delta}(D)$. For $\Delta_{x_i}, \Delta_{x_j} \subset \mathrm{POS}_{\Delta}(D)$ and $[x_i]_D \neq [x_j]_D$, since
$c_{ij} = \{\mathbf{C} \in \Delta : (C_{x_i} \not\subset C_{x_j}) \wedge (C_{x_j} \not\subset C_{x_i})\} \cup \{\mathbf{C}_s \wedge \mathbf{C}_t : (C_{s_{x_i}} \subset C_{s_{x_j}}) \wedge (C_{t_{x_j}} \subset C_{t_{x_i}})\}$, $\mathbf{P} \cap c_{ij} \neq \phi (c_{ij} \neq \phi)$, then $\mathbf{P}_{x_i} \not\subset \mathbf{P}_{x_j}$, $\mathbf{P}_{x_j} \not\subset \mathbf{P}_{x_i}$, since $\mathbf{P}_{x_i} \not\subset \mathbf{P}_{x_j}$, $\mathbf{P}_{x_j} \not\subset \mathbf{P}_{x_i} \Rightarrow \Delta_{x_i} \not\subset \Delta_{x_j}$, $\Delta_{x_j} \not\subset \Delta_{x_i}$ is also true, so we have $\Delta_{x_i} \not\subset \Delta_{x_j}$, $\Delta_{x_j} \not\subset \Delta_{x_i} \iff \mathbf{P}_{x_i} \not\subset \mathbf{P}_{x_j}$, $\mathbf{P}_{x_j} \not\subset \mathbf{P}_{x_i}$, by Theorem 5.4, we also know $\mathrm{POS}_{\mathbf{P}}(D) = \mathrm{POS}_{\Delta}(D)$.

(2) The proof is similar to that of (2) of Theorem 4.8. □

Corollary 5.7. *Suppose $\mathbf{P} \subseteq \Delta$, then $\mathbf{P}$ is a relative reduct of Δ if and only if it is the minimal set satisfying $\mathbf{P} \cap c_{ij} \neq \phi$ for $c_{ij} \neq \phi \quad i, j \leqslant n$.*

Theorem 5.8. *Let $(U, \Delta, D = \{d\})$ be an inconsistent covering decision system, $M(U, \Delta, D) = (c_{ij} : i, j \leqslant n)$ is the discernibility matrix of (U, Δ, D), $f(U, \Delta, D)$ is the discernibility function of (U, Δ, D) defined as $f(U, \Delta, D)(\overline{\mathbf{C}_1}, \ldots, \overline{\mathbf{C}_m}) = \wedge\{\vee(c_{ij}) : 1 \leqslant j < i \leqslant n, \}$. If $f(U, \Delta, D) = \bigvee_{k=1}^{l} (\wedge\Delta_k)(\Delta_k \subseteq \Delta)$ is obtained from $f(U, \Delta, D)$ by applying the multiplication and absorption laws as many times as possible which satisfies every element in Δ_i, appears only one time, then set $\{\Delta_k : k \leqslant l\}$ is the collection of all the relative reducs of Δ, i.e., if $\mathrm{Red}(\Delta, D)$ is the collection of all the reductions of system (U, Δ, D), then $\mathrm{Red}(\Delta, D) = \{\Delta_1, \ldots, \Delta_l\}$.*

Proof. The proof is similar to that of Theorem 4.10. □

It should be pointed out that if the covering decision system is consistent, then the method proposed in this section is the same as the one in Section 4. If every cover is a partition, then the method proposed in this section is just the method for computing relative reducts of traditional rough sets in [23]. We present the following example to illustrate our idea in this section.

Example 5.9. Now we consider a house evaluation problem. Suppose $U = \{x_1, \ldots, x_{10}\}$to be a set of ten houses, and E={price; structure; color; surrounding} to be a set of attributes. The values of "price" are {high; middle; low}, the values of "color" are {good; bad}, the values of "structure" are {reasonable; ordinary; poor}, and the values of "surrounding" are {quiet; a little noisy; noisy; quite noisy}. We have four specialists {A, B, C, D} to evaluate the attributes of these houses. Possibly, their evaluation results are not the same as one another. The evaluation results are listed below.

For attribute "price":

A: high $= \{x_1, x_2, x_4, x_6, x_7, x_8, x_9, x_{10}\}$, middle $= \{x_3\}$, low $= \{x_5\}$;
B: high $= \{x_1, x_2, x_3, x_5, x_6, x_8, x_9, x_{10}\}$, middle $= \{x_4\}$, low $= \{x_6, x_7\}$;
C: high $= \{x_1, x_2, x_3, x_4, x_8, x_9, x_{10}\}$, middle $= \{x_8\}$, low $= \{x_2, x_3, x_5, x_6, x_9\}$;
D: high $= \{x_1, x_2, x_3, x_4, x_8, x_9, x_{10}\}$, middle $= \{x_7\}$, low $= \{x_5, x_6\}$.

For attribute "structure":

A: reasonable $= \{x_1, x_2, x_3, x_4, x_5\}$, ordinary $= \{x_6, x_7, x_8, x_9\}$, poor $= \{x_{10}\}$;
B: reasonable $= \{x_1, x_2, x_3, x_4, x_5, x_6\}$, ordinary $= \{x_7, x_8, x_9\}$, poor $= \{x_{10}\}$;
C: reasonable $= \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$, ordinary $= \{x_8, x_9\}$, poor $= \{x_{10}\}$;
D: reasonable $= \{x_1, x_2, x_3, x_4, x_5, x_7\}$, ordinary $= \{x_6, x_8, x_9\}$, poor $= \{x_{10}\}$.

For attribute "color":

A: good $= \{x_1, x_2, x_3, x_6, x_8, x_{10}\}$, bad $= \{x_4, x_5, x_7, x_9\}$;
B: good $= \{x_1, x_2, x_3, x_8, x_{10}\}$, bad $= \{x_4, x_5, x_6, x_7, x_9\}$;
C: good $= \{x_1, x_6, x_8, x_9, x_{10}\}$, bad $= \{x_2, x_3, x_4, x_5, x_7\}$;
D: good $= \{x_1, x_2, x_6, x_8, x_{10}\}$, bad $= \{x_3, x_4, x_5, x_7, x_9\}$.

For attribute "*surrounding*":

A: quiet $= \{x_1, x_2\}$, a little noisy $= \{x_3, x_4, x_5\}$, noisy $= \{x_6, x_8, x_{10}\}$, very noisy $= \{x_7, x_9\}$;
B: quiet $= \{x_1, x_3\}$, a little noisy $= \{x_2, x_4, x_5, x_6\}$, noisy $= \{x_7, x_8, x_{10}\}$, very noisy $= \{x_9\}$;
C: quiet $= \{x_1, x_3\}$, a little noisy $= \{x_2, x_4, x_5, x_7\}$, noisy $= \{x_6, x_8, x_{10}\}$, very noisy $= \{x_9\}$;
D: quiet $= \{x_1, x_2, x_6\}$, a little noisy $= \{x_3, x_4, x_5\}$, noisy $= \{x_8, x_9, x_{10}\}$, very noisy $= \{x_7\}$.

Assuming the evaluation of every specialist is of the same importance. If we want to combine these evaluations without losing information, we should union the evaluations given by every specialist for every attribute value. Then for every attribute we get a cover instead of a partition, which embodies a kind of uncertainty caused by different interpretation of the data.

Price: $\mathbf{C}_1 = \{\{x_1, x_2, x_3, x_4, x_6, x_7, x_8, x_9, x_{10}\}, \{x_3, x_4, x_6, x_7\}, \{x_3, x_4, x_5, x_6, x_7\}\}$,
Structure: $\mathbf{C}_2 = \{\{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}, \{x_6, x_7, x_8, x_9\}, \{x_{10}\}\}$,
Color: $\mathbf{C}_3 = \{\{x_1, x_2, x_3, x_6, x_8, x_9, x_{10}\}, \{x_2, x_3 x_4, x_5, x_6, x_7, x_9\}\}$,
Surrounding: $\mathbf{C}_4 = \{\{x_1, x_2, x_3, x_6\}, \{x_2, x_3, x_4, x_5, x_6, x_7\}, \{x_6, x_8, x_9, x_{10}\}, \{x_6, x_7, x_9\}\}$.

Final decision D is given as

$$U/D = \{\{x_1, x_2, x_3, x_6\} \text{ (sale)}, \{x_4, x_5, x_7\} \text{ (further evaluation)}, \{x_8, x_9, x_{10}\} \text{ (reject)}\}.$$

Δ_i is Δ_{x_i} for short, $\mathbf{P}_i$ means $\mathbf{P}_{x_i}$ for short. Let $\boldsymbol{\Delta} = \{\mathbf{C}_i : i = 1, \ldots, 4\}$, then

$$\Delta_1 = \{x_1, x_2, x_3, x_6\}, \quad \Delta_2 = \{x_2, x_3, x_6\}, \quad \Delta_3 = \{x_3, x_6\}, \quad \Delta_4 = \{x_3, x_4, x_6, x_7\}, \quad \Delta_5 = \{x_3, x_4, x_5, x_7\},$$
$$\Delta_6 = \{x_6\}, \quad \Delta_7 = \{x_6, x_7\}, \quad \Delta_8 = \{x_6, x_8, x_9\}, \quad \Delta_9 = \{x_6, x_9\}, \quad \Delta_{10} = \{x_{10}\}.$$

The positive domain of D relative to Δ is :$\mathrm{POS}_\Delta(D) = \bigcup_{X \in U/D} \underline{\boldsymbol{\Delta}}(X) = \{x_1, x_2, x_3, x_6, x_{10}\}$.

Let $\mathbf{P} = \Delta\text{-}\mathrm{C}_1\}$, then $\mathbf{P}_1 = \{x_1, x_2, x_3, x_6\}$, $\mathbf{P}_2 = \mathbf{P}_3 = \{x_2, x_3, x_6\}$, $\mathbf{P}_4 = \mathbf{P}_5 = \{x_2, x_3, x_4, x_5, x_6, x_7\}$, $\mathbf{P}_6 = \{x_6\}$, $\mathbf{P}_7 = \{x_6, x_7\}$, $\mathbf{P}_8 = \{x_6, x_8, x_9\}, \mathbf{P}_9 = \{x_6, x_9\}, \mathbf{P}_{10} = \{x_{10}\}$, the positive domain of D relative to $\mathbf{P}$ is: $\mathrm{POS}_{\mathbf{P}}(D) = \bigcup_{X \in U/D} \underline{\mathbf{P}}(X) = \{x_1, x_2, x_3, x_6, x_{10}\}$.

Thus $\mathrm{POS}_\Delta(D) = \mathrm{POS}_{\mathbf{P}}(D)$, hence $\mathbf{C}_1$ is a superfluous element in Δ relative to D. Here we can see $\Delta_{x_1} \not\subset \Delta_{x_4}, \Delta_{x_4} \not\subset \Delta_{x_1} \Rightarrow \quad \mathbf{P}_{x_1} \not\subset \mathbf{P}_{x_4}, \mathbf{P}_{x_4} \not\subset \mathbf{P}_{x_1}, \Delta_{x_2} \not\subset \Delta_{x_4}, \Delta_{x_4} \not\subset \Delta_{x_2} \Rightarrow \mathbf{P}_{x_2} \subset \mathbf{P}_{x_4}, \Delta_{x_3} \subset \Delta_{x_4} \Rightarrow \mathbf{P}_{x_3} \subset \mathbf{P}_{x_4}$. Although the original relation of x_2, x_4 with respect to $\boldsymbol{\Delta}$ changes after $\mathbf{C}_1$ is deleted from $\boldsymbol{\Delta}$, $\mathbf{C}_1$ is a superfluous element of $\boldsymbol{\Delta}$ relative to D.

The discernibility matrix of $(U, \boldsymbol{\Delta})$ is as follows:

$$\begin{bmatrix}
\boldsymbol{\Delta} & & & & & & & & & \\
\boldsymbol{\Delta} & \boldsymbol{\Delta} & & & & & & & & \\
\boldsymbol{\Delta} & \boldsymbol{\Delta} & \boldsymbol{\Delta} & & & & & & & \\
\{\mathbf{C}_3,\mathbf{C}_4\} & \{\mathbf{C}_1\wedge\mathbf{C}_3,\mathbf{C}_1\wedge\mathbf{C}_4, \mathrm{C}_3, \mathrm{C}_4\} & \{\mathbf{C}_3,\mathbf{C}_4\} & \boldsymbol{\Delta} & & & & & & \\
\{\mathbf{C}_1,\mathrm{C}_3,\mathbf{C}_4\} & \{\mathbf{C}_1,\mathrm{C}_3,\mathbf{C}_4\} & \{\mathbf{C}_1,\mathrm{C}_2,\mathbf{C}_4\} & \boldsymbol{\Delta} & \boldsymbol{\Delta} & & & & & \\
\boldsymbol{\Delta} & \boldsymbol{\Delta} & \boldsymbol{\Delta} & \{\mathrm{C}_2,\mathbf{C}_3,\mathrm{C}_4\} & \{\mathbf{C}_1,\mathrm{C}_2,\mathbf{C}_3,\mathrm{C}_4\} & \boldsymbol{\Delta} & & & & \\
\{\mathbf{C}_3,\mathbf{C}_4\} & \{\mathbf{C}_1\wedge\mathbf{C}_3,\mathbf{C}_1\wedge\mathbf{C}_4,\mathrm{C}_3,\mathrm{C}_4\} & \{\mathbf{C}_1\wedge\mathbf{C}_3,\mathbf{C}_1\wedge\mathbf{C}_4,\mathrm{C}_3,\mathrm{C}_4\} & \boldsymbol{\Delta} & \boldsymbol{\Delta} & \{\mathbf{C}_3,\mathbf{C}_4\} & \boldsymbol{\Delta} & & & \\
\{\mathbf{C}_2,\mathbf{C}_4\} & \{\mathbf{C}_2,\mathbf{C}_3, \mathrm{C}_4\} & \{\mathbf{C}_1, \mathrm{C}_2,\mathbf{C}_3, \mathrm{C}_4\} & \boldsymbol{\Delta} & \boldsymbol{\Delta} & \{\mathbf{C}_1, \mathrm{C}_2,\mathbf{C}_3, \mathrm{C}_4\} & \boldsymbol{\Delta} & \boldsymbol{\Delta} & & \\
\{\mathbf{C}_2,\mathbf{C}_4\} & \{\mathbf{C}_2,\mathbf{C}_4\} & \{\mathbf{C}_1,\mathrm{C}_2,\mathbf{C}_4\} & \boldsymbol{\Delta} & \boldsymbol{\Delta} & \{\mathbf{C}_1,\mathrm{C}_2,\mathbf{C}_4\} & \boldsymbol{\Delta} & \boldsymbol{\Delta} & \boldsymbol{\Delta} & \\
\{\mathbf{C}_2,\mathbf{C}_4\} & \{\mathbf{C}_2,\mathbf{C}_4\} & \{\mathbf{C}_2,\mathbf{C}_4\} & \{\mathrm{C}_2,\mathbf{C}_3, \mathrm{C}_4\} & \{\mathbf{C}_1,\mathrm{C}_2,\mathbf{C}_3,\mathrm{C}_4\} & \{\mathbf{C}_2\} & \{\mathbf{C}_3,\mathbf{C}_4\} & \{\mathbf{C}_2\} & \{\mathbf{C}_2\} & \boldsymbol{\Delta}
\end{bmatrix}$$

and

$$\begin{aligned}
f(U, \boldsymbol{\Delta})(\overline{\mathbf{C}_1}, \ldots, \overline{\mathbf{C}_4}) &= \wedge\{\vee(c_{ij}) : 1 \leqslant j \leqslant i \leqslant 10, \quad c_{ij} \neq \phi\} \\
&= \mathbf{C}_2 \wedge (\mathbf{C}_2 \vee \mathbf{C}_4) \wedge (\mathbf{C}_3 \vee \mathbf{C}_4) \wedge (\mathbf{C}_1 \vee \mathbf{C}_3 \vee \mathbf{C}_4) \wedge (\mathbf{C}_2 \vee \mathbf{C}_3 \vee \mathbf{C}_4) \wedge (\mathbf{C}_1 \vee \mathbf{C}_2 \vee \mathbf{C}_4) \\
&\quad \wedge (\mathbf{C}_1 \vee \mathbf{C}_2 \vee \mathbf{C}_3 \vee \mathbf{C}_4) \wedge ((\mathbf{C}_1 \wedge \mathbf{C}_3) \vee (\mathbf{C}_1 \wedge \mathbf{C}_4) \vee \mathbf{C}_3 \vee \mathbf{C}_4) \wedge ((\mathbf{C}_2 \wedge \mathbf{C}_3) \vee (\mathbf{C}_2 \wedge \mathbf{C}_4) \vee \mathbf{C}_3 \vee \mathbf{C}_4) \\
&= \mathbf{C}_2 \wedge (\mathbf{C}_3 \vee \mathbf{C}_4) = (\mathbf{C}_2 \wedge \mathbf{C}_3) \vee (\mathbf{C}_2 \wedge \mathbf{C}_4),
\end{aligned}$$

so

$$\mathrm{RED}(\boldsymbol{\Delta}, D) = \{\{\mathbf{C}_2, \mathbf{C}_4\}, \{\mathbf{C}_2, \mathbf{C}_3\}\}, \quad \mathrm{Core}_D(\boldsymbol{\Delta}) = \{\mathbf{C}_2\}.$$

If these ten houses are the training samples, then we have two different kinds of evaluation references for other input samples: {structure; surrounding}, {color; structure}, and it is clear that the attribute "structure" is the key attribute for the evaluation of houses.

6. Experimental analysis

In order to verify the effectiveness of the proposed method for attribute reduction with covering rough sets, we download several data sets from UCI Repository of machine learning databases [1], which are described in Table 3. All conditional attributes in 4 data sets are numerical.

The objective of the experiments is to show the difference between attribute reduction with traditional rough set model and covering rough sets. As traditional model can just deal with discrete attributes, we employ three discretization techniques, equal-width, equal-frequency, and fuzzy c-means clustering (FCM) [8], to transform the numerical data into discrete one. Then we employ the dependency based algorithm [9] on the discretized data sets. The number of the selected features are presented in Table 4.

Then we apply the proposed algorithm to the numerical decision systems. We construct a covering dependency based forward greedy search algorithm. The difference between the proposed one and that in [8,9] lines only in computing dependency. The proposed algorithm uses covering rough sets to calculate the dependency of numerical attributes.

There is a parameter to be specified in using covering rough sets to compute dependency, the size of the covering of an object. We try the size of covering delta at a step of 0.05.

Before reduction, all of the numerical attributes are standardized at interval [0, 1]. During the experiment, we employ radial basis function(RBF) kernel in support vector machine(SVM) learning algorithm to validate the selected features. The number of selected features and classification accuracy with 10-fold cross validation are as shown in Figs. 1–4.

From which, we can see that the number of selected features and classification accuracies vary with the size of covering. There are some similar patterns in these figures. At the beginning, fewer features are selected; and correspondingly, the classification accuracies are lower. Then the numbers of selected features increase and the classification performance improves. However, if the size of covering continues to increase, the number of selected features and classification performance synchronously decrease, which means that there is an optimal covering size.

It can be seen from Tables 4 and 5 that the reductions with the proposed algorithm produce the best classification performances in most cases. At the same time, the selected features are more than those obtained

Table 3
Data description

Data set	Abbreviation	Samples	Features	Classes
1. Ionosphere	Iono	351	34	2
2. Sonar, Mines vs. Rocks	Sonar	208	60	2
3. Wisconsin Diagnostic Breast Cancer	WDBC	569	31	2
4. Wisconsin Prognostic Breast Cancer	WPBC	198	33	2

Table 4
Comparison of the numbers of feature with different reduction algorithms

	Raw data	Equal width	Equal frequency	FCM	Covering
Iono	34	1	1	10	18
Sonar	60	7	0	6	20
WDBC	31	12	6	7	23
WPBC	33	9	0	4	19

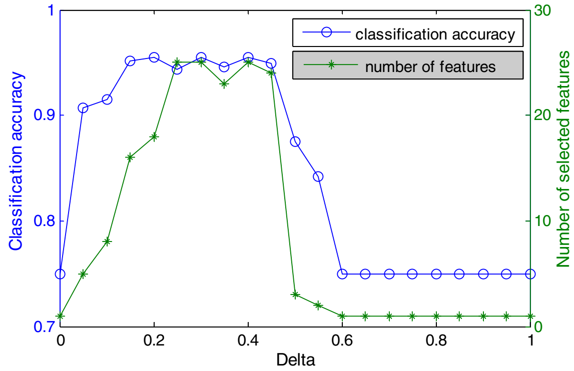

Fig. 1. Variation of number of selected features and classification accuracies with size of covering (data set iono).

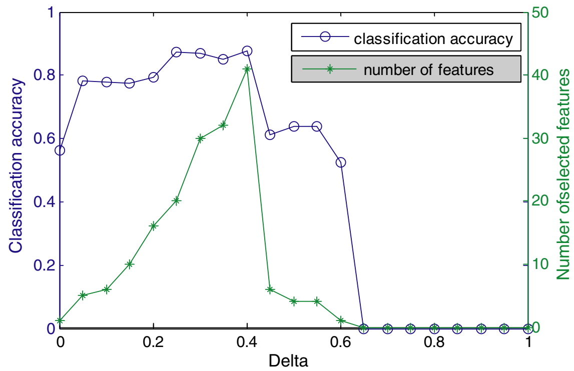

Fig. 2. Variation of number of selected features and classification accuracies with size of covering (data set Sonar).

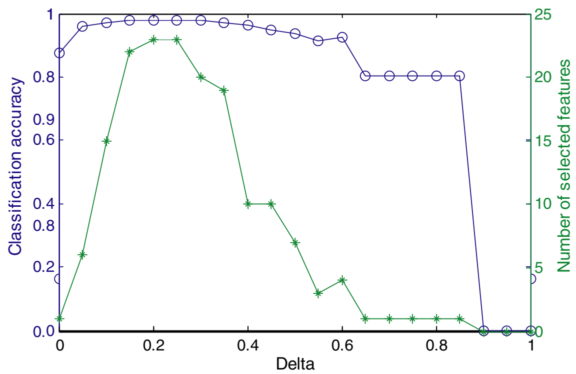

Fig. 3. Variation of number of selected features and classification accuracies with size of covering (data set WDBC).

from Pawlak's rough set model. The results show that discretization causes information loss, while the covering method can be used to keep or even improve the classification performance of the reduced data sets. Furthermore, we can also see the number of selected features can be reduced if we relax the condition of maximal accuracy. i. e., there exist some attribute subsets which can be used to get the similar classification power, but the subsets are far less than the reductions with highest accuracy. Sometimes, we can select a suboptimal reduct to get a subset of smaller features.

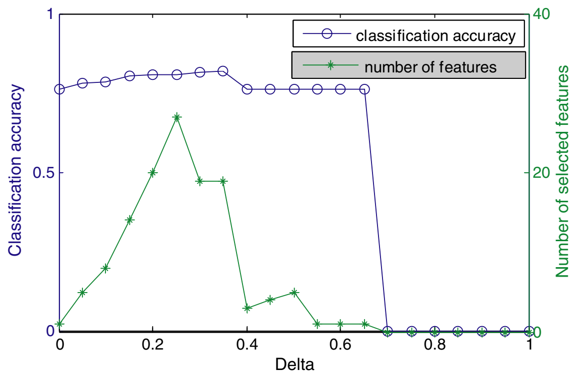

Fig. 4. Variation of number of selected features and classification accuracies with size of covering (data set WPBC).

Table 5
Comparison of classification accuracies with different reduction algorithms

	Raw data	Equal width	Equal frequency	FCM	Covering
Iono	0.9379 ± 0.0507	0.7499 ± 0.0866	0.7499 ± 0.0866	0.9348 ± 0.0479	0.9546 ± 0.0420
Sonar	0.8510 ± 0.0948	0.7398 ± 0.1106	0	0.7074 ± 0.1004	0.8745 ± 0.0833
WDBC	0.9808 ± 0.0225	0.9668 ± 0.0223	0.9597 ± 0.0231	0.9649 ± 0.0183	0.9790 ± 0.0215
WPBC	0.7779 ± 0.0420	0.7737 ± 0.0514	0	0.7837 ± 0.0506	0.819 ± 0.0703

7. Conclusion

According to traditional rough set theory, every conditional attribute in a decision system is requested to induce a partition of the universe. But this condition could not always be hold in many practical problems and it limits the applications of rough set theory. For this reason the partition is generalized to a cover. In this paper, by defining the intersection of covers we propose a new method to reduce redundant covers in a covering decision system. We also develop the algorithms to compute all the reductions of covering decision system by discernibility matrix. With the above discussion, we set up the theoretical foundation for reduction of covering rough sets for its further application. Another important thing we should point out is that our methods in this paper are the natural generalization of the corresponding algorithm in traditional rough set theory [20].

Acknowledgements

We would like to thank all the referees for their valuable comments and the first author is supported by a foundation of North China Electric Power University.

References

[1] C.L. Blake, C.J. Merz, UCI Repository of machine learning databases, Tech. Rep., Univ. California, Dept. Inform. Comp. Sci., Irvine, CA. Available from: <http://www.ics.uci.edu/mlearn/MachineLearning.html>, 1998.
[2] Z. Bonikowski, E. Bryniarski, U. Wybraniec, Extensions and intentions in the rough set theory, Information Sciences 107 (1998) 149–167.
[3] Z. Bonikowski, Algebraic structures of rough sets, in: W. Ziarko (Ed.), Rough Sets, Fuzzy Sets and Knowledge Discovery, Springer-Verlag, London, 1994, pp. 243–247.
[4] E. Bryniarski, A calculus of rough sets of the first order, Bulletin of the Polish Academy of Sciences 16 (1989) 71–77.
[5] G. Cattaneo, Abstract approximate spaces for rough theories, in: Polkowski, Skowron (Eds.), Rough Sets in Knowledge Discovery 1: Methodology and Applications, Physicaverlag, Heidelberg, 1998, pp. 59–98.
[6] Chen Degang, Zhang Wenxiu, S. Yeung, C.C. Tsang, Rough approximations on a complete completely distributive lattice with applications to generalized rough sets, Information Sciences 176 (2006) 1829–1848.

[7] J.M. Fernandez Salido, S. Murakami, Rough set analysis of a general type of fuzzy data using transitive aggregations of fuzzy similarity relations, Fuzzy Sets and Systems 139 (2003) 635–660.
[8] Q.H. Hu, D.R. Yu, Z.X. Xie, Information-preserving hybrid data reduction based on fuzzy-rough techniques, Pattern Recognition Letters 27 (5) (2006) 414–423.
[9] Q.H. Hu, X.D. Li, D.R. Yu, Analysis on Classification Performance of Rough Set Based Reducts, in: Q. Yang, G. Webb (Eds.), PRICAI 2006, LNAI 4099, Springer-Verlag Heidelberg, Berlin, 2006, pp. 423–433.
[10] M. Kryszkiewicz, Rough set approach to incomplete information systems, Information Sciences 112 (1998) 39–49.
[11] M. Kryszkiewicz, Rule in incomplete information systems, Information Sciences 113 (1999) 271–292.
[12] T.Y. Lin, Neighborhood systems and relational database, in: Proceeding of CSC'88, 1988, p. 725.
[13] T.Y. Lin, Neighborhood systems-application to qualitative fuzzy and rough sets, in: P.P. Wang (Ed.), Advances in Machine Intelligence and Soft Computing, Department of Electrical Engineering, Duke University, Durham, NC, USA, 1997, pp. 132–155.
[14] T.Y. Lin, Q. Lin, K.J. Huang, W. Chen, Rough sets, neighborhood systems and application, in: Z.w. Ras, M. Zemankova, M.L. Emrichm (Eds.), Methodologies for Intelligent Systems, Proceeding of ISMIS 1990, Knoxville, TN, October 25–27, North-Holland, New York, 1990, pp. 130–141.
[15] T.Y. Lin, Y.Y. Yao, Mining soft rules using rough sets and neighborhoods, in: Proceedings of CESA'96, IMASCS Multiconference, Lille, France, July 9–12, 1996, pp. 1095–1100.
[16] J.N. Mordeson, Rough set theory applied to (fuzzy) ideal theory, Fuzzy Sets and Systems 121 (2001) 315–324.
[17] H.S. Nguyen, D. Slezak, Approximation reducts and association rules correspondence and complexity results, in: N. Zhong, A. Skowron, S. Oshuga (Eds.), Proceedings of RSFDGrC'99, Yamaguchi, Japan. LNAI1711, 1999, pp. 137–145.
[18] Z. Pawlak, Rough sets, International Journal of Computer and Information Sciences 11 (1982) 341–356.
[19] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning About Data, Kluwer Academic Publishing, Dordrecht, 1991.
[20] Z. Pawlak, Andrzej Skowron, Rudiments of rough sets, Information Sciences 177 (2006) 3–27.
[21] Z. Pawlak, Andrzej Skowron, Rough sets: Some extensions, Information Sciences 177 (2006) 28–40.
[22] Z. Pawlak, Andrzej Skowron, Rough sets and Boolean reasoning, Information Sciences 177 (2006) 41–73.
[23] J.A. Pomykala, Approximation operations in approximation space, Bulletin of the Polish Academy of Sciences 9–10 (1987) 653–662.
[24] A. Skowron, J. Stepaniuk, Tolerance approximation spaces, Fundamenta Informaticae 27 (1996) 245–253.
[25] A. Skowron, C. Rauszer, The discernibility matrices and functions in information systems, in: R. Slowinsk (Ed.), Intelligent Decision support, Handbook of Applications and Advances of the Rough Sets Theory, Kluwer Academic Publishers, 1992.
[26] R. Slowinski, D. Vanderpooten, A generalized definition of rough approximations based on similarity, IEEE Transactions Data Knowledge Engineering 2 (2000) 331–336.
[27] A. Wasilewska, Topological rough algebras, in: Lin, Cercone (Eds.), Rough Sets & Data Mining, Kluwer Academic Publishers, Boston, 1997, pp. 411–425.
[28] Y.Y. Yao, Constructive and algebraic method of theory of rough sets, Information Sciences 109 (1998) 21–47.
[29] Y.Y. Yao, Relational interpretations of neighborhood operators and rough set approximation operators, Information Sciences 101 (1998) 239–259.
[30] W. Zakowski, Approximations in the space (U,Π), Demonstratio Mathematica 16 (1983) 761–769.
[31] W. Ziarko, Variable precision rough set model, Journal of Computer and System Sciences 46 (1993) 39–59.
[32] W.-X. Zhang, W.-Z. Wu, J.-Y. Liang, D.-Y. Li, Theory and method of rough sets, Science Press, Beijing, 2001.
[33] William Zhu, Fei-Yue Wang, Some results on the covering generalized rough sets, Pattern Recognition Artificial Intelligence 5 (2002) 6–13.
[34] William Zhu, Fei-Yue Wang, Reduction and axiomization of covering generalized rough sets, Information Sciences 152 (2003) 217–230.
[35] William Zhu, Topological approaches to covering rough sets, Information Sciences 177 (2007) 1892–1915.