



ELSEVIER

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

A novel method for attribute reduction of covering decision systems

Changzhong Wang^{a,*}, Qiang He^b, Degang Chen^c, Qinghua Hu^d^a Department of Mathematics, Bohai University, Jinzhou 121000, PR China^b Department of Mathematics and Computer Science, Hebei University, Baoding 071002, PR China^c Department of Mathematics and Physics, North China Electric Power University, Beijing 102206, PR China^d School of Computer Science and Technology, Tianjin University, Tianjin 300072, PR China

ARTICLE INFO

Article history:

Received 17 April 2011

Received in revised form 6 August 2013

Accepted 24 August 2013

Available online 2 September 2013

Keywords:

Covering rough set

Attribute reduction

Discernibility matrix

Consistent covering decision system

Inconsistent covering decision system

ABSTRACT

Attribute reduction has become an important step in pattern recognition and machine learning tasks. Covering rough sets, as a generalization of classical rough sets, have attracted wide attention in both theory and application. This paper provides a novel method for attribute reduction based on covering rough sets. We review the concepts of consistent and inconsistent covering decision systems and their reducts and we develop a judgment theorem and a discernibility matrix for each type of covering decision system. Furthermore, we present some basic structural properties of attribute reduction with covering rough sets. Based on a discernibility matrix, we develop a heuristic algorithm to find a subset of attributes that approximate a minimal reduct. Finally, the experimental results for UCI data sets show that the proposed reduction approach is an effective technique for addressing numerical and categorical data and is more efficient than the method presented in the paper [D.G. Chen, C.Z. Wang, Q.H. Hu, A new approach to attribute reduction of consistent and inconsistent covering decision systems with covering rough sets, *Information Sciences* 177(17) (2007) 3500–3518].

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Attribute reduction plays an important role in pattern recognition and machine learning. Classical rough set theory, as proposed by Pawlak [31], has been used as a mathematical tool to conceptualize and analyze various types of imprecise and uncertain data. It can be a useful tool for studying attribute reduction in information systems. The main goal of attribute reduction is to remove redundant information in a data set, so that correct decisions can quickly be made while preserving or even improving classification ability. Rough set theory has become a popular mathematical framework for feature selection, rule extraction, data mining and knowledge discovery [5–23,26–38,44–46,51–53].

Originally, rough set theory was based on equivalence relations, which can partition the objects of a universe into mutually exclusive equivalence classes. Objects in the same equivalence class are indiscernible. Equivalence classes are also called elementary information granules; an arbitrary subset of the universe can be approximated by elementary information granules. Rough set-based data analysis starts from a data table, called an information system, which contains objects that are described by a finite set of categorical attributes [17–22,28–38,44,51,58–60].

However, in many real-world situations, the databases are not suitable for being handled by classical rough sets [1–7,10–27,39–43,45–58,61–68]. For example, some objects in a database have multiple attribute values. If we consider all of the

* Corresponding author. Tel.: +86 15041656439.

E-mail address: changzhongwang@126.com (C. Wang).

attribute values for a multi-valued attribute, we create a covering of the universe instead of a partition. Some illustrative examples of this type of database are given by Chen and Wang [5,45]. In addition, classical rough sets work only for categorical data and cannot be directly applied to reducing numerical attributes. Numerical data need to be discretized before attribute reduction, and this causes information loss [14–16]. For this reason, similarity relation rough sets [43], dominance rough sets [12,13,41], and neighborhood rough sets [14,15,55] were developed. These models induce a covering of a universe instead of a partition. Zakowski [64] used coverings of a universe to introduce the notion of covering rough sets. Later, many authors conducted studies on the properties of covering approximation operators [1–5,25,40,45,48–50,57,63–68]. Bonikowski et al. [1] mainly studied the structures of the coverings. Most authors studied some properties of the upper and lower approximations of covering rough sets and examined some axioms that are satisfied by Pawlak rough sets [25,48,63–65,67,68]. However, few people employ covering rough sets to conduct research on attribute reduction. Zhu and Wang [66] investigated the problem of reducing covering elements from a covering. Their objective was to remove excessive covering elements in a covering, under the condition that the approximations of an arbitrary subset were kept unchanged. A pioneering study on attribute reduction with covering rough sets was conducted by Chen and Wang [5]. Formal concepts of attribute reduction based on covering rough sets were introduced, and some approaches to computing all reducts using a discernibility matrix were developed. However, these approaches are very complex and time-consuming due to their poor design.

In this paper, we present a novel method for the attribute reduction of decision systems based on covering rough sets. We review some concepts related to attribute reductions with covering rough sets, present some theorems describing attribute reductions, and then construct discernibility matrices in covering decision systems. Compared with the attribute reduction method presented in [5], the computational complexity of the proposed method is greatly reduced. Finally, based on a discernibility matrix, we develop a heuristic algorithm to find a minimal subset of attributes that approximate an optimal reduct. The experimental results show that the proposed reduction method can effectively handle data and is more efficient than the reduction method presented in [5].

This paper is organized as follows. In Section 2, we recall some basic concepts related to traditional and covering rough sets. In Section 3, we develop a new method of attribute reduction in consistent covering decision systems and examine some basic properties of attribute reduction with covering rough sets. In Section 4, we discuss a theory for reducing the conditional attributes of inconsistent covering decision systems. In Section 5, we present some experiments on some public data sets.

2. Preliminaries

First, we review some basic concepts related to classical rough sets that can be found in [30–32].

An information system is a pair (U, A) , where U is a nonempty set of samples $\{x_1, x_2, \dots, x_n\}$, called a universe or a sample space and A is a nonempty set of attributes or features. For any subset $B \subseteq A$, we can define an equivalence relation as follows:

$$\text{Ind}(B) = \{(x, y) \in U \times U : a(x) = a(y), \forall a \in B\}$$

Clearly, $\text{Ind}(B) = \cap_{a \in B} \text{Ind}(\{a\})$. We denote the equivalence class of x with respect to $\text{Ind}(B)$ as $[x]_B$. For $X \subseteq U$, the lower and upper approximations of X with respect to $\text{Ind}(B)$ are defined as follows:

$$\underline{B}(X) = \{x \in U : [x]_B \subseteq X\}, \quad \overline{B}(X) = \{x \in U : [x]_B \cap X \neq \emptyset\}$$

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

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

where $\vee(c_{ij})$ is the disjunction operation on c_{ij} .

A decision system is a pair (U, A, D) , where D is called a decision attribute and the elements in A are called conditional attributes. An attribute $a \in B (\subseteq A)$ is called relatively dispensable in B if $\text{Pos}_B(D) = \text{Pos}_{B-\{a\}}(D)$; otherwise, it is said to be relatively indispensable in B , where $\text{Pos}_B(D) = \cup_{x \in U/D} \underline{B}x$. If each attribute in B is relatively indispensable in B , B is said to be relatively independent in (U, A, D) . A subset $B \subseteq A$ is called a relative reduct in (U, A, D) if B is relatively independent in (U, A, D) and $\text{Pos}_B(D) = \text{Pos}_A(D)$. The set of all relatively indispensable attributes in A is called the relative core of (U, A, D) .

Classical rough sets can only be used to address discrete data. Rough sets have been extended to address complex data with discrete and numerical features. Covering rough set theory is a natural extension of classical rough sets and can effectively handle continuous data. Next, we recall some basic notions related to covering rough sets that can be found in [5,45].

Definition 2.1. Let U be a universe of discourse and C be a family of subsets of U . Then, C is called a covering of U if no subset in C is empty and $\cup C = U$.

Definition 2.2. Let $C = \{K_1, K_2, \dots, K_n\}$ be a covering of U . For every $x \in U$, let $C_x = \cap \{K_j: K_j \in C, x \in K_j\}$. $Cov(C) = \{C_x: x \in U\}$ is then also a covering of U , and we call it the induced covering of C .

For every $x \in U$, C_x is the minimal set including x in $Cov(C)$. Each element in $Cov(C)$ cannot be written as the union of other elements in $Cov(C)$. $Cov(C) = C$ if and only if 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 $C_x = C_y$.

This definition uses a covering to granulate a universe; each sample is associated with a neighborhood. This type of granulation mode defines a relation between samples. The relation has the following properties.

- (1) Reflexivity: $\forall x \in U, x \in C_x$.
- (2) Anti-symmetry: if $y \in C_x$ and $x \in C_y$, then $C_x = C_y$.
- (3) Transitivity: $\forall x, y, z \in U$; if $x \in C_y$ and $y \in C_z$, then $x \in C_z$.

In fact, the relation induced by Definition 2.2 is a dominance relation and the elements in $Cov(C)$ are dominance classes of the dominance relation [50].

Definition 2.3. Let $\Delta = \{C_i: i = 1, 2, \dots, m\}$ be a family of coverings of U . For every $x \in U$, let $\Delta_x = \cap \{(C_i)_x: (C_i)_x \in Cov(C_i), x \in (C_i)_x\}$. $Cov(\Delta) = \{\Delta_x: x \in U\}$ is then also a covering of U , and we call it the induced covering of Δ .

Clearly Δ_x is the intersection of all coverings including x in Δ . So for every $x \in U$, Δ_x is the minimal set including x in $Cov(\Delta)$. $Cov(\Delta)$ can be viewed as the intersection of coverings in Δ . Every element in $Cov(\Delta)$ cannot be written as the union of other elements in $Cov(\Delta)$. If every covering in Δ is a partition, then $Cov(\Delta)$ is also a partition and Δ_x is the equivalence class that includes 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 $\Delta_x = \Delta_y$.

If a covering produces a dominance relation, a family of coverings can induce a set of dominance relations. Definition 2.3 presents a method for constructing dominance classes with a family of coverings [50]. In this sense, we can say that each element in $Cov(\Delta)$ is a dominance class. Based on Definitions 2.2 and 2.3, discrete or continuous data can be granulated into elementary information granules—dominance classes.

For any $X \subseteq U$, the lower and upper approximations of X with respect to Δ are defined as follows:

$$\underline{\Delta}(X) = \{x \in U : \Delta_x \subseteq X\}, \quad \overline{\Delta}(X) = \{x \in U : \Delta_x \cap X \neq \emptyset\}.$$

The pair approximation operators are dual to each other. The readers can refer to [55] for a discussion of the operators' important properties. The positive, negative and boundary domains of X with respect to Δ are computed using the following formulas, respectively:

$$POS_{\Delta}(X) = \underline{\Delta}(X), \quad NEG_{\Delta}(X) = U - \overline{\Delta}(X), \quad \text{and } BN_{\Delta}(X) = \overline{\Delta}(X) - \underline{\Delta}(X).$$

3. Attribute reduction in consistent covering decision systems

Attribute reduction is an important application of rough set theory. However, in practical applications, a large number of databases cannot be directly handled by classical rough sets. For this reason, similarity relation rough sets, dominance rough sets, and neighborhood rough sets are widely used. These models all induce coverings of a universe instead of partitions and can thus be categorized into covering rough sets. In the following, we present an illustrative example to further show which types of data sets are suitable for being handled by covering rough sets.

Example 3.1. Table 1 shows a petroleum exploration data set, where 'No' represents the number of wells, 'c₁' represents the average amplitude, 'c₂' represents the semi-attenuation time, 'c₃' represents the instantaneous frequency, 'c₄' represents the Root Mean Square (RMS) amplitude, 'D' represents the well type, '1' stands for a petroleum well, '2' stands for a dry well.

All of the conditional attributes in Table 1 are numerical. According to classical rough sets, these conditional attributes need to be discretized before attribute reduction. However, discretization can cause information loss. Here, we directly address the data set with covering rough sets. First, we employ two granulation methods to generate elementary covering information granules. Let $U = \{x_i: i = 1, 2, \dots, 7\}$ be the set of well samples.

Table 1
Petroleum exploration data.

No	c ₁	c ₂	c ₃	c ₄	D
1	85.4051	54.9076	27.0976	61.2443	1
2	66.4600	39.0294	25.2136	53.5892	1
3	109.7228	34.7891	39.0904	69.2056	1
4	66.8497	65.5205	24.2548	48.1985	2
5	55.3513	66.4574	32.5181	39.7762	2
6	40.1256	77.9718	33.1104	24.6535	2
7	48.6229	35.0866	30.1238	35.4501	2

- (1) For each conditional attribute $c_i (i = 1, 2, 3, 4)$, we associate a neighborhood with each sample $x_k (k = 1, 2, \dots, 7)$, called a covering element of x_k . For example, a covering element of x_k with respect to c_i can be defined as $(C_i)_{x_k} = \{x : d(x_k, x) \leq \varepsilon\}$, where $d(\cdot, \cdot)$ is a distance function and $\varepsilon \geq 0$ is a specified threshold. Obviously, $C_i = \{(R_i)_s(x_k) : k = 1, 2, \dots, 7\}$ is a covering of U , and $\Delta = \{C_1, C_2, C_3, C_4\}$ is a family of coverings of U . By Definition 2.2, we can obtain a set of dominance classes for each conditional attribute. Then, by Definition 2.3, we can obtain a set of elementary information granules—dominance classes generated by all conditional attributes. In some sense, dominance classes reflect dominance relationships among the well samples in the classification.
- (2) By sorting the well samples according to their values, we define an order relation R_i for each conditional attribute c_i . For example, we can define a relation R_1 for c_1 such that the relationship between well samples is that $x_3 > x_1 > x_4 > x_2 > x_5 > x_7 > x_6$. It is easily observed that R_1 is a special dominance relation. We can thus compute the dominance class of each well sample related to R_1 as follows: $(R_1)_s(x_3) = \{x_3, x_1, x_4, x_2, x_5, x_7, x_6\}$, $(R_1)_s(x_1) = \{x_1, x_4, x_2, x_5, x_7, x_6\}$, $(R_1)_s(x_4) = \{x_4, x_2, x_5, x_7, x_6\}$, $(R_1)_s(x_2) = \{x_2, x_5, x_7, x_6\}$, $(R_1)_s(x_5) = \{x_5, x_7, x_6\}$, $(R_1)_s(x_7) = \{x_7, x_6\}$, and $(R_1)_s(x_6) = \{x_6\}$. Thus, $C_1 = \{(R_1)_s(x_k) : k = 1, 2, \dots, 7\}$ is a covering of U , and $\Delta = \{C_1, C_2, C_3, C_4\}$ is a family of coverings of U . In this case, we can directly use Definition 2.3 to compute dominance classes for all of the conditional attributes.

With the above-mentioned granulation mechanisms of data sets, an attribute reduction approach was proposed to compute all of the reducts in [5]. However, this approach is very complex and time-consuming due to its poor design. Covering rough sets can also be used to address data sets with numerical and symbolic attributes. We discuss covering rough sets using real data sets in Section 5. In this section, we propose a new model of attribute reduction with covering rough sets in consistent decision systems.

Based on the above observations, if we apply covering rough sets to numerical data sets, each conditional attribute can induce a covering rather than a partition. Suppose U is a nonempty set of samples $\{x_1, x_2, \dots, x_n\}$, $\Delta = \{C_i : i = 1, 2, \dots, m\}$ is a family of coverings generated by condition attributes, D is a decision attribute, and U/D is the decision partition of U . We call the triple (U, Δ, D) a covering decision system. If for any object x , there exists $D_j \in U/D$ such that $\Delta_x \subseteq D_j$, then the triple (U, Δ, D) is called a consistent covering decision system, denoted as $\text{Cov}(\Delta) \leq U/D$. Otherwise, (U, Δ, D) is called an inconsistent covering decision system. Let $D = \{d\}$. The decision function is defined as $d : U \rightarrow V_d$, $d(x) \in V_d$, where $V_d = \{1, 2, \dots, r\}$ is the decision domain of d . For any $x \in U$, the image of set Δ_x under function d is denoted as $d(\Delta_x)$. If (U, Δ, D) is a consistent covering decision system, then $d(\Delta_x)$ is a singleton and $d(\Delta_x) = \{d(x)\}$ holds.

Theorem 3.1. Let (U, Δ, D) be a consistent covering decision system, $P \subseteq \Delta$. Then,

- (1) $\text{Cov}(P) = \text{Cov}(\Delta) \Leftrightarrow \Delta_x = P_x, \forall x \in U$.
- (2) For $\forall x, y \in U, y \notin \Delta_x$ if and only if there is at least a covering $C_i \in \Delta$ such that $y \notin (C_i)_x$.
- (3) $\text{Cov}(P) = \text{Cov}(\Delta)$ if and only if for $\forall x, y \in U$, if $y \notin \Delta_x$, then $y \notin P_x$.

Proof. Straightforward.

Theorem 3.1(1) and (3) present two necessary and sufficient conditions that enable one to judge whether two induced coverings are equal and inform us that two induced coverings are equal if and only if their basic information granularities are equal. As follows from Theorem 3.1(2), if two samples can be distinguished under the granularity level of Δ_x , then there is at least one covering $C_i \in \Delta$ such that the two samples are also distinguished under the granularity level of $(C_i)_x$ and vice versa. \square

Definition 3.2 5. Let (U, Δ, D) be a consistent covering decision system, $C_i \in \Delta$. If $\text{Cov}(\Delta - \{C_i\}) \leq U/D$, then C_i is called superfluous relative to D in Δ ; otherwise, C_i is called indispensable relative to D in Δ . Let $P \subseteq \Delta$; if $\text{Cov}(P) \leq U/D$ and $\text{Cov}(P - \{C_i\}) \not\leq U/D$ for any $C_i \in P$, then P is called a reduct of Δ relative to D . The collection of all indispensable elements in Δ is called the core of Δ relative to D and is denoted as $\text{Core}_D(\Delta)$.

Theorem 3.3. Suppose $\text{Cov}(\Delta) \leq U/D, C \in \Delta$. Then, C is indispensable relative to D if and only if there is at least one pair of $x_i, x_j \in U$ satisfying $d(x_i) \neq d(x_j)$ such that $x_j \notin \Delta_{x_i} \Rightarrow x_j \in (\Delta - \{C\})_{x_i}$.

Proof. \Rightarrow If $\text{Cov}(\Delta - \{C\}) \leq U/D$ is not true, then by the definition of (U, Δ, D) there is an x_i such that $(\Delta - \{C\})_{x_i} \subseteq [x_i]_D$ is not true. Thus, $\exists x_j \in U$, such that $x_j \in (\Delta - \{C\})_{x_i}$ and $x_j \notin [x_i]_D$, which implies that $d(x_i) \neq d(x_j)$ and $x_j \notin \Delta_{x_i}$. Hence, there is at least a pair of $x_i, x_j \in U$ satisfying $d(x_i) \neq d(x_j)$ such that $x_j \notin \Delta_{x_i} \Rightarrow x_j \in (\Delta - \{C\})_{x_i}$.

\Leftarrow Because $d(x_i) \neq d(x_j)$ and $\text{Cov}(\Delta) \leq U/D$, it follows that $\Delta_{x_i} \subseteq [x_i]_D \neq [x_j]_D \supseteq \Delta_{x_j}$. By the reflexivity of Δ_{x_j} we have $x_j \notin \Delta_{x_i}$. Because $x_j \notin \Delta_{x_i} \Rightarrow x_j \in (\Delta - \{C\})_{x_i}$, it follows that $(\Delta - \{C\})_{x_i} \subseteq [x_i]_D$ is not true. By the definition of (U, Δ, D) , we know that C is indispensable relative to D . \square

Theorem 3.3 shows that a covering C in a consistent covering decision system is indispensable if and only if there is at least one pair of objects x_i and x_j with different decisions such that x_j does not belong to Δ_{x_i} but instead belongs to $(\Delta - \{C\})_{x_i}$. This finding implies that C is a sole covering that meets $x_j \notin C_{x_i}$.

Theorem 3.4 (Judgment theorem of attribute reduction). *Suppose that $\text{Cov}(\Delta) \leq U/D$, $\mathbf{P} \subseteq \Delta$. Then, $\text{Cov}(\mathbf{P}) \leq U/D$ if and only if $x_j \notin P_{x_i}$ for any $x_i, x_j \in U$, satisfying $d(x_i) \neq d(x_j)$.*

Proof. \Rightarrow Because $\text{Cov}(\mathbf{P}) \leq U/D$, it follows that $P_{x_i} \subseteq [x_i]_D$ for any $x_i \in U$. We also have $P_{x_j} \subseteq [x_j]_D$ for any $x_j \in U$ satisfying $[x_i]_D \cap [x_j]_D = \emptyset$. This means that $x_j \notin P_{x_i}$. Because $\Delta_{x_i} \subseteq P_{x_i}$ and $\Delta_{x_j} \subseteq P_{x_j}$, we have $d(\Delta_{x_i}) \neq d(\Delta_{x_j})$, which implies $d(x_i) \neq d(x_j)$ by the reflexivity of Δ_{x_i} and Δ_{x_j} . That is, we have $x_j \notin P_{x_i}$ for any $x_i, x_j \in U$ satisfying $d(x_i) \neq d(x_j)$.

\Leftarrow Suppose $\text{Cov}(\mathbf{P}) \leq U/D$ is not true. It follows from the definition of a consistent covering decision system that there must exist $x_{i_0} \in U$ such that $P_{x_{i_0}} \subseteq [x_{i_0}]_D$ is not true. This means that there is $x_{j_0} \in U$ such that $x_{j_0} \in P_{x_{i_0}}$ and $[x_{i_0}]_D \cap [x_{j_0}]_D = \emptyset$, which implies $d(x_{i_0}) \neq d(x_{j_0})$. Thus, $x_{j_0} \in P_{x_{i_0}}$ but x_{i_0} and x_{j_0} satisfy $d(x_{i_0}) \neq d(x_{j_0})$. This contradicts the fact that $x_j \notin P_{x_i}$ for any $x_i, x_j \in U$ satisfying $d(x_i) \neq d(x_j)$. Hence, $\text{Cov}(\mathbf{P}) \leq U/D$. \square

The objective of relative reduction is to find a minimal subset of Δ that ensures that the original relationship between every two objects with different decisions is invariant. For any two objects with different decisions, we can thus find at least one condition attribute (covering) to distinguish them. To search for all reducts in a consistent covering decision system, we define its discernibility matrix according to **Theorems 3.1(2)** and **3.4** as follows:

Definition 3.5. Let $(U, \Delta, D = \{d\})$ be a consistent covering decision system, where $U = \{x_1, x_2, \dots, x_n\}$ and $\Delta = \{C_i : i = 1, 2, \dots, m\}$. By $M(U, \Delta, D)$, we denote a $n \times n$ matrix (c_{ij}) , called the discernibility matrix of (U, Δ, D) , which is defined as

$$c_{ij} = \begin{cases} \{C \in \Delta : x_j \notin C_{x_i}\}, & d(x_i) \neq d(x_j) \\ \Delta, & d(x_i) = d(x_j) \end{cases}$$

for $x_i, x_j \in U$.

To keep the classification of objects unchanged, any two objects must be distinguished by some attributes if they have different decisions. This idea is consistent with the viewpoint of classical rough sets. In addition, covering and classical rough sets bear some formal resemblance to discernibility matrices. Thus, the proposed reduction method is a generalization of classical rough sets.

In [5], a method based on a discernibility matrix was proposed for computing all relative reducts. To compare it with the proposed method in this paper, we present the reduction method in [5] as follows.

Definition 3.6. Let $(U, \Delta, D = \{d\})$ be a consistent covering decision system, where $U = \{x_1, x_2, \dots, x_n\}$ and $\Delta = \{C_i : i = 1, 2, \dots, m\}$. By $M(U, \Delta, D)$, we denote a $n \times n$ matrix (c'_{ij}) , called the discernibility matrix of (U, Δ, D) , which is defined as

$$c'_{ij} = \begin{cases} \{C \in \Delta : (C_{x_i} \not\subseteq C_{x_j}) \wedge (C_{x_j} \not\subseteq C_{x_i})\} \cup \{C_s \wedge C_t : (C_{s_{x_i}} \subseteq C_{s_{x_j}}) \wedge (C_{t_{x_j}} \subseteq C_{t_{x_i}})\}, & d(\Delta_{x_i}) \neq d(\Delta_{x_j}) \\ \Delta, & d(\Delta_{x_i}) = d(\Delta_{x_j}) \end{cases}$$

for $x_i, x_j \in U$.

By observing **Definitions 3.5** and **3.6**, we find the time complexities in computing neighborhoods of samples are the same and that the main difference lies in the following two points:

- (1) In **Definition 3.5**, we only need to compute C_{x_i} for any covering $C \in \Delta$ and an arbitrary object $x_i \in U$, while both C_{x_i} and Δ_{x_i} need to be computed in **Definition 3.6**.
- (2) The computational complexities of the two discernibility matrices are different. In **Definition 3.6**, to distinguish any two objects with different decisions, we need m^2 comparisons of their neighborhoods. The time complexity is thus $O(m^2 \cdot n^2)$. However, we need m comparisons for the two objects in **Definition 3.5**. Hence, the time complexity in **Definition 3.5** is $O(m \cdot n^2)$. Therefore, the proposed method of reduction is simpler than the method presented in [5].

In fact, the reducts obtained by these two discernibility matrices are the same. Next, we provide a theoretical explanation of this point.

It is easily observed that the discernibility matrix in **Definition 3.6** is symmetric and the discernibility matrix in **Definition 3.5** is not. For any $x_i, x_j \in U$, because $(U, \Delta, D = \{d\})$ is a consistent covering decision system, we know that $d(x_i) \neq d(x_j)$ is equivalent to $d(\Delta_{x_i}) \neq d(\Delta_{x_j})$. Let $c'_{ij} = c_{ij} \cup c_{ji}$; we need to show that $c'_{ij} = c'_{ji}$.

For any $C \in c'_{ij}$, it follows from **Definition 3.5** that $x_j \notin C_{x_i}$ or $x_i \notin C_{x_j}$. This means that $C_{x_j} \not\subseteq C_{x_i}$ or $C_{x_i} \not\subseteq C_{x_j}$ by **Definition 2.2**, which implies $C \in c'_{ji}$ by **Definition 3.6**. Hence, we have that $c'_{ij} \subseteq c'_{ji}$.

Conversely, if $C \in c'_{ji}$, by **Definition 3.6**, we have that $C_{x_i} \not\subseteq C_{x_j}$ and $C_{x_j} \not\subseteq C_{x_i}$. It follows from **Definition 2.2** that $x_i \notin C_{x_j}$ and $x_j \notin C_{x_i}$, which implies that $C \in c'_{ij}$ by **Definition 3.5**. If $C_s \wedge C_t \in c'_{ji}$, it follows from **Definition 3.6** that $(C_s)_{x_i} \subseteq (C_s)_{x_j}$ and $(C_t)_{x_j} \subseteq (C_t)_{x_i}$. By **Definition 2.2**, we have that $x_j \notin (C_s)_{x_i}$ and $x_i \notin (C_t)_{x_j}$. From **Definition 3.5**, this means that $C_s \in c_{ij}$ and $C_t \in$

$\in c_{ji}$. Hence, $C_s \wedge C_t \in c_{ij}^*$, and it follows that $c_{ij}^* = c_{ij}'$. We can thus say that these two discernibility matrices are equivalent and that the reducts obtained by them are the same.

The following theorems examine some structural properties of attribute reduction in a consistent covering decision system. The complexities of the following theorems are lower than the corresponding theorems in [5].

Theorem 3.7. *Let (U, Δ, D) be a consistent covering decision system.*

- (1) Let $\mathbf{P} \subseteq \Delta$; then $\text{Cov}(\mathbf{P}) \leq U/D$ if and only if $\mathbf{P} \cap c_{ij} \neq \emptyset$ for every $i, j \leq n$.
- (2) $\text{Core}_D(\Delta) = \{C \in \Delta : c_{ij} = \{C\}, i, j \leq n\}$.

Proof.

(1) \Rightarrow For any $x_i, x_j \in U$:

If $d(x_i) = d(x_j)$, then by Definition 3.5, we have $c_{ij} = \Delta$. Hence, $\mathbf{P} \cap c_{ij} \neq \emptyset$.

If $d(x_i) \neq d(x_j)$, then by Theorem 3.4, we have that $x_j \notin P_{x_i}$. Thus, by (2) of Theorem 3.1(2), there must exist $C \in \mathbf{P}$ such that $y \notin C_x$. By Definition 3.5, it follows that $C \in c_{ij}$, which implies $\mathbf{P} \cap c_{ij} \neq \emptyset$.

\Leftarrow If $\mathbf{P} \cap c_{ij} \neq \emptyset$ for any $c_{ij} \neq \emptyset$, we suppose that $C \in \mathbf{P} \cap c_{ij}$. If $d(x_i) \neq d(x_j)$, then by Definition 3.5, we have $x_j \notin C_{x_i} \Rightarrow x_j \notin \Delta_{x_i} \Rightarrow x_j \notin P_{x_i}$. By Theorem 3.4, we have $\text{Cov}(\mathbf{P}) \leq U/D$.

(2) If $C \in \text{Core}_D(\Delta)$, then $\text{Cov}(\Delta - \{C\}) \leq U/D$ is not true. By Theorem 3.3, there is at least one pair of x_i, x_j satisfying $d(x_i) \neq d(x_j)$ such that $x_j \in (\Delta - \{C\})_{x_i}$. Combined with $(\Delta - \{C\})_{x_i} = \bigcap_{C' \in \Delta - \{C\}} C'_{x_i}$, it follows that $x_j \in C'_{x_i}$ for any covering C' in $\Delta - \{C\}$. Because $\text{Cov}(\Delta) \leq U/D$, we have $x_j \notin \Delta_{x_i}$. Assume that $x_j \in C_{x_i}$. Because $\Delta_{x_i} = (\Delta - \{C\})_{x_i} \cap C_{x_i}$, it follows that $x_j \in \Delta_{x_i}$, which is a contradiction. Thus, $x_j \notin C_{x_i}$. This means that there is only $C \in \Delta$ satisfying $x_j \notin C_{x_i}$. Thus, $c_{ij} = \{C\}$, which implies that $C \in \{C \in \Delta : c_{ij} = \{C\}\}$. Hence, $\text{Core}_D(\Delta) \subseteq \{C \in \Delta : c_{ij} = \{C\}\}$.

Conversely, if $c_{ij} = \{C\}$ for $x_i, x_j \in U$, by Definition 3.5, we have that $d(x_i) \neq d(x_j)$ and $x_j \notin C_{x_i}$, which implies that $x_j \in (\Delta - \{C\})_{x_i}$. It follows from Theorem 3.3 that C is indispensable relative to D , i.e., $C \in \text{Core}_D(\Delta)$. Thus, $\text{Core}_D(\Delta) \supseteq \{C \in \Delta : c_{ij} = \{C\}\}$. Therefore, $\text{Core}_D(\Delta) = \{C \in \Delta : c_{ij} = \{C\}\}$. \square

Corollary 3.8. *Let $\mathbf{P} \subseteq \Delta$. \mathbf{P} is a relative reduct of Δ if and only if it is a minimal set satisfying $\mathbf{P} \cap c_{ij} \neq \emptyset, i, j = 1, 2, \dots, n$.*

Let $\Delta = \{C_1, C_2, \dots, C_n\}$ be a family of coverings on U . The corresponding Boolean variable $\bar{C}_i (i \leq n)$ is defined for each covering $C_i \in \Delta$, and $f(U, \Delta, D)$ is a function on (U, Δ, D) , defined as $f(U, \Delta, D)(\bar{C}_1, \bar{C}_2, \dots, \bar{C}_n) = \bigwedge \{\vee(c_{ij})\}, (i, j \leq n, c_{ij} \neq q\emptyset)$. Then, $f(U, \Delta, D)$ is a Boolean function of (U, Δ, D) and is called a discernibility function or discernibility formula of (U, Δ, D) , where $\vee(c_{ij})$ represents a disjunction operation among the elements in c_{ij} . By applying disjunctive and conjunctive operations on the discernibility function, we can obtain the minimal disjunctive normal form of $f(U, \Delta, D)$ to compute all reducts.

The following example is employed to illustrate our idea in this section:

Example 3.2 [5]. Suppose $U = \{x_1, \dots, x_9\}, \Delta = \{C_i : i = 1, \dots, 4\}$, and

$$\begin{aligned} C_1 &= \{\{x_1, x_2, x_4, x_5, x_7, x_8\}, \{x_2, x_3, x_5, x_6, x_8, x_9\}\}, \\ C_2 &= \{\{x_1, x_2, x_3, x_4, x_5, x_6\}, \{x_4, x_5, x_6, x_7, x_8, x_9\}\}, \\ C_3 &= \{\{x_1, x_2, x_3\}, \{x_4, x_5, x_6, x_7, x_8, x_9\}, \{x_7, x_8, x_9\}\}, \\ C_4 &= \{\{x_1, x_2, x_4, x_5\}, \{x_2, x_3, x_5, x_6\}, \{x_4, x_5, x_7, x_8\}, \{x_5, x_6, x_8, x_9\}\}, \\ U/D &= \{\{x_1, x_2, x_3\}, \{x_4, x_5, x_6\}, \{x_7, x_8, x_9\}\} \end{aligned}$$

By Definition 3.5 we can compute the discernibility matrix of (U, Δ, D) as follows:

$$\begin{pmatrix} \Delta & \Delta & \Delta & \{C_3\} & \{C_3\} & \{C_1, C_3, C_4\} & \{C_2, C_3\} & \{C_2, C_3, C_4\} & \Delta \\ \Delta & \Delta & \Delta & \{C_1, C_3, C_4\} & \{C_3\} & \{C_1, C_3, C_4\} & \{C_1, C_3, C_4\} & \{C_2, C_3\} & \Delta \\ \Delta & \Delta & \Delta & \{C_1, C_3, C_4\} & \{C_3\} & \{C_3\} & \Delta & \{C_2, C_3\} & \{C_2, C_3\} \\ \{C_2, C_3, C_4\} & \{C_2, C_3, C_4\} & \Delta & \Delta & \Delta & \Delta & \{C_2, C_4\} & \{C_2, C_4\} & \{C_1, C_2, C_4\} \\ \Delta & \{C_2, C_3, C_4\} & \Delta & \Delta & \Delta & \Delta & \{C_1, C_2, C_4\} & \{C_2, C_4\} & \{C_1, C_2, C_4\} \\ \Delta & \{C_2, C_3, C_4\} & \{C_2, C_3, C_4\} & \Delta & \Delta & \Delta & \{C_2, C_4\} & \{C_2, C_4\} & \{C_2, C_4\} \\ \{C_2, C_3, C_4\} & \{C_2, C_3, C_4\} & \Delta & \{C_3\} & \{C_3\} & \{C_1, C_3, C_4\} & \Delta & \Delta & \Delta \\ \Delta & \{C_2, C_3, C_4\} & \Delta & \{C_1, C_3, C_4\} & \{C_3\} & \{C_1, C_3, C_4\} & \Delta & \Delta & \Delta \\ \Delta & \{C_2, C_3, C_4\} & \{C_2, C_3, C_4\} & \{C_1, C_3, C_4\} & \{C_3\} & \{C_3\} & \Delta & \Delta & \Delta \end{pmatrix}$$

and

$$\begin{aligned} f(U, \Delta)(\overline{C_1}, \dots, \overline{C_4}) &= \wedge\{\vee(C_{ij}) : 1 \leq j < i \leq 9\} \\ &= C_3 \wedge (C_1 \vee C_3 \vee C_4) \wedge (C_2 \vee C_3 \vee C_4) \wedge (C_2 \vee C_3) \wedge (C_2 \vee C_4) \wedge (C_1 \vee C_2 \vee C_4) = C_3 \wedge (C_2 \vee C_4) \\ &= (C_2 \wedge C_3) \vee (C_3 \wedge C_4) \end{aligned}$$

Thus, $Red(\Delta) = \{\{C_3, C_4\}, \{C_2, C_3\}\}$, $Core_D(\Delta) = \{C_3\}$.

In comparison with the discernibility matrix of (U, Δ, D) presented in [5], the discernibility matrix constructed in this paper is simpler and the computational complexity is greatly reduced. However, the reducts that it obtained are the same as those obtained in the method used in [5], so we can say that the proposed method is a more efficient way to compute all of the reducts than the method presented in [5].

4. Attribute reduction in inconsistent covering decision systems

In consistent covering decision systems, each object has a neighborhood in which all objects have the same decision as itself. In many practical problems, this situation does not always exist. There are also some objects in a neighborhood whose decisions are not consistent. For this reason, consistent covering decision systems need to be generalized to more general decision systems: inconsistent covering decision systems. In this section, we study attribute reduction for this type of decision system.

Suppose that $\Delta = \{C_i : i = 1, 2, \dots, m\}$ is a family of coverings of U , D is a decision attribute, $d: U \rightarrow V_d$ is the decision function, and $(U, \Delta, D = \{d\})$ is an inconsistent covering decision system. The positive domain of D relative to Δ is defined as $Pos_{\Delta}(D) = \cup_{X \in U/D} \underline{\Delta}(X)$. For an attribute $C_i \in \Delta$, if $Pos_{\Delta}(D) = Pos_{\Delta - \{C_i\}}(D)$, then C_i is called superfluous relative to D in Δ ; otherwise, C_i is called indispensable in Δ . Let \mathbf{P} be a subset of Δ . If each element in \mathbf{P} is indispensable in \mathbf{P} and $Pos_{\mathbf{P}}(D) = Pos_{\Delta}(D)$, then \mathbf{P} is called a reduct of Δ relative to D . The set of all indispensable elements relative to D in Δ is called the relative core of Δ and is denoted as $Core_D(\Delta)$.

From the definition of positive domain, we can easily obtain the following properties.

Theorem 4.1. *Let $(U, \Delta, D = \{d\})$ be an inconsistent covering decision system. Then,*

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

Theorem 4.1(1) shows that if a sample belongs to the positive domain, then all samples in its neighborhood belong to the positive domain. On the contrary, if a sample does not belong to the positive region, the samples in its neighborhood are not pure. Theorem 4.1(2) and (3) show that a subset of attributes can keep the positive domain of a decision invariant if and only if it needs to ensure the positive domain of each decision equivalence class is invariant. The following theorem presents a sufficient and necessary condition that enables one to judge whether a conditional attribute (covering) is superfluous.

Theorem 4.2. *Let $(U, \Delta, D = \{d\})$ be an inconsistent covering decision system, $C \in \Delta$. Then, $Pos_{\Delta}(D) \neq Pos_{\Delta - \{C\}}(D)$ if and only if there is at least one pair of $x_i, x_j \in U$ such that if they satisfy one of the following conditions, then $x_j \notin \Delta_{x_i} \Rightarrow x_j \in \{\Delta - \{C\}\}_{x_i}$.*

- (1) $x_i \in Pos_{\Delta}(D)$ and $x_j \notin Pos_{\Delta}(D)$;
- (2) $x_i, x_j \in Pos_{\Delta}(D)$ and $[x_i]_D \cap [x_j]_D = \emptyset$.

Proof. We denote $\mathbf{P} = \Delta - \{C\}$.

\Leftarrow Assume that there is a pair of $x_i, x_j \in U$ satisfying $x_i \in Pos_{\Delta}(D)$ and $x_j \notin Pos_{\Delta}(D)$. Then by Theorem 4.1(1), we have that $\Delta_{x_i} \subseteq [x_i]_D$ holds and that $\Delta_{x_j} \subseteq [x_i]_D$ is not true, which implies $x_j \notin \Delta_{x_i}$ by the reflexivity of Δ_{x_j} . Because $x_j \notin \Delta_{x_i} \Rightarrow x_j \in \mathbf{P}_{x_i}$, it follows from the transitivity of \mathbf{P}_{x_i} that $\mathbf{P}_{x_j} \subseteq \mathbf{P}_{x_i}$. Assume that $\mathbf{P}_{x_i} \subseteq [x_i]_D$. By $\Delta_{x_j} \subseteq \mathbf{P}_{x_i}$, we have $\Delta_{x_j} \subseteq [x_i]_D$, which implies $x_j \in Pos_{\Delta}(D)$. This is a contradiction. It follows that $\mathbf{P}_{x_i} \subseteq [x_i]_D$ is not true. Thus $x_i \in \underline{\Delta}([x_i]_D)$ and $x_i \notin \underline{\mathbf{P}}([x_i]_D)$, which implies $Pos_{\Delta}(D) \neq Pos_{\Delta - \{C\}}(D)$.

Assume that there is a pair of $x_i, x_j \in U$ satisfying $x_i, x_j \in Pos_{\Delta}(D)$ and $[x_i]_D \cap [x_j]_D = \emptyset$. By Theorem 4.1(1), we then have that $\Delta_{x_i} \subseteq [x_i]_D$ and $\Delta_{x_j} \subseteq [x_j]_D$, which implies $x_j \notin \Delta_{x_i}$ by the reflexivity of Δ_{x_j} . Because $x_j \notin \Delta_{x_i} \Rightarrow x_j \in \mathbf{P}_{x_i}$, it follows that $\mathbf{P}_{x_i} \subseteq [x_i]_D$ is not true. Thus, $x_i \in \underline{\Delta}([x_i]_D)$ and $x_i \notin \underline{\mathbf{P}}([x_i]_D)$, which implies $Pos_{\Delta}(D) \neq Pos_{\Delta - \{C\}}(D)$.

\Rightarrow Because $Pos_{\Delta}(D) \neq Pos_{\Delta - \{C\}}(D)$, by Theorem 4.1(2), there exists $X_0 \in U/D$ such that $\underline{\mathbf{P}}(X_0) \subset \underline{\Delta}(X_0)$. It follows that $\exists x_i \in U$ such that $\Delta_{x_i} \subseteq [x_i]_D = X_0$ and that $\mathbf{P}_{x_i} \subseteq X_0$ is not true. Hence, there must exist $x_j \in U$ such that $x_j \in \mathbf{P}_{x_i}$ and $x_j \notin [x_i]_D$, which implies $x_j \notin \Delta_{x_i}$.

If $\Delta_{x_j} \subseteq [x_j]_D$, then $x_i, x_j \in Pos_{\Delta}(D)$ and $[x_i]_D \cap [x_j]_D = \emptyset$, which implies that x_i and x_j satisfy condition (2). If $\Delta_{x_j} \subseteq [x_j]_D$ is not true, then $x_i \in Pos_{\Delta}(D)$ and $x_j \notin Pos_{\Delta}(D)$, which implies that x_i and x_j satisfy the condition (1). No matter the type of situation in which x_i and x_j meet, we always have that $x_j \notin \Delta_{x_i}$ and $x_j \in \mathbf{P}_{x_i}$, i.e., $x_j \notin \Delta_{x_i} \Rightarrow x_j \in \{\Delta - \{C\}\}_{x_i}$. \square

The first condition means that one sample belongs to the positive domain of the decision and that the other does not. The second item means that both samples belong to the positive domain but have different decisions. We call the two conditions covering reduction conditions. An indispensable attribute must be retained in reducts; otherwise, the positive domain of decision will change. An attribute is indispensable if and only if there is at least one pair of samples meeting the covering reduction conditions and this pair can only be distinguished by this attribute. Generally speaking, we always find all indispensable attributes first when searching for a reduct; doing so can guarantee the quality of the selected reducts and the efficiency of the searches for reducts.

Theorem 4.3 (Judgment theorem of attribute reduction). *Suppose that $(U, \Delta, D = \{d\})$ is an inconsistent covering decision system, $P \subseteq \Delta$. Then, $Pos_{\Delta}(D) = Pos_P(D)$ if and only if for any $x_i, x_j \in U$, no matter what case, they meet in the following conditions. Then, $x_j \notin \Delta_{x_i} \Rightarrow x_j \notin P_{x_i}$.*

- (1) $x_i \in Pos_{\Delta}(D)$ and $x_j \notin Pos_{\Delta}(D)$;
- (2) $x_i, x_j \in Pos_{\Delta}(D)$ and $[x_i]_D \cap [x_j]_D = \emptyset$.

Proof. \Leftarrow For any $x_i, x_j \in U$, if $x_i \in Pos_{\Delta}(D)$ and $x_j \notin Pos_{\Delta}(D)$, then by Theorem 4.1(1), we have that $\Delta_{x_i} \subseteq [x_i]_D$ holds and that $\Delta_{x_j} \subseteq [x_k]_D$ is not true for any $x_k \in U$. Assume that $x_j \in \Delta_{x_i}$; then, $\Delta_{x_j} \subseteq \Delta_{x_i}$ by the transitivity of Δ_{x_i} . Thus, $\Delta_{x_j} \subseteq [x_i]_D$, which implies that $x_j \in Pos_{\Delta}(D)$. This is a contradiction. Hence, $x_j \notin \Delta_{x_i}$. If $x_i, x_j \in Pos_{\Delta}(D)$ and $[x_i]_D \cap [x_j]_D = \emptyset$, then by Theorem 4.1(1), we have that $\Delta_{x_i} \subseteq [x_i]_D$ and $\Delta_{x_j} \subseteq [x_j]_D$, which implies $x_j \notin \Delta_{x_i}$. Because $x_j \notin \Delta_{x_i} \Rightarrow x_j \notin P_{x_i}$, it follows that $P_{x_i} \subseteq [x_i]_D$. This means that $\Delta_{x_i} \subseteq [x_i]_D \Rightarrow P_{x_i} \subseteq [x_i]_D$, which implies $\Delta_{x_i} \subseteq [x_i]_D \iff P_{x_i} \subseteq [x_i]_D$ by $\Delta_{x_i} \subseteq P_{x_i}$. By Theorem 4.1(3) we have $Pos_{\Delta}(D) = Pos_P(D)$.

\Rightarrow For any $x, y \in U$, if $x_i \in Pos_{\Delta}(D)$ and $x_j \notin Pos_{\Delta}(D)$, then by Theorem 4.1(1), we have that $\Delta_{x_i} \subseteq [x_i]_D$ holds and that $\Delta_{x_j} \subseteq [x_k]_D$ is not true for any $x_k \in U$. Assume that $x_j \in \Delta_{x_i}$. Then, $\Delta_{x_j} \subseteq \Delta_{x_i}$ by the transitivity of Δ_{x_i} . Thus, $\Delta_{x_j} \subseteq [x_i]_D$; it is a contradiction. Hence, $x_j \notin \Delta_{x_i}$. Because $Pos_{\Delta}(D) = Pos_P(D)$, it follows from Theorem 4.1(3) that $\Delta_{x_i} \subseteq [x_i]_D \iff P_{x_i} \subseteq [x_i]_D$. Combined with the fact that $\Delta_{x_j} \subseteq [x_k]_D$ is not true for any $x_k \in U$, it follows that $P_{x_j} \subseteq [x_k]_D$ is not true for any $x_k \in U$. Assume that $x_j \in P_{x_i}$. Then, $P_{x_j} \subseteq P_{x_i}$ by the transitivity of P_{x_i} . Thus, $P_{x_j} \subseteq [x_i]_D$, which is a contradiction. Hence, $x_j \notin P_{x_i}$, i.e., $x_j \notin \Delta_{x_i} \Rightarrow x_j \notin P_{x_i}$.

If $x_i, x_j \in Pos_{\Delta}(D)$ and $[x_i]_D \cap [x_j]_D = \emptyset$, then by Theorem 4.1(1), we have that $\Delta_{x_i} \subseteq [x_i]_D$ and $\Delta_{x_j} \subseteq [x_j]_D$, which implies $x_j \notin \Delta_{x_i}$. Because $Pos_{\Delta}(D) = Pos_P(D)$, it follows from Theorem 4.1(3) that $\Delta_{x_i} \subseteq [x_i]_D \iff P_{x_i} \subseteq [x_i]_D$ and $\Delta_{x_j} \subseteq [x_j]_D \iff P_{x_j} \subseteq [x_j]_D$, which implies that $x_j \notin P_{x_i}$, i.e., $x_j \notin \Delta_{x_i} \Rightarrow x_j \notin P_{x_i}$. \square

This theorem is the basis for constructing a discernibility matrix. From Theorems 3.1(2) and 4.3, we have the following definition.

Definition 4.4. Let $(U, \Delta, D = \{d\})$ be an inconsistent covering decision system, where $U = \{x_1, x_2, \dots, x_n\}$ and $\Delta = \{C_i : i = 1, 2, \dots, m\}$. By $M(U, \Delta, D)$, we denote an $n \times n$ matrix (c_{ij}) , which is called the discernibility matrix of (U, Δ, D) , such that if any $x_i, x_j \in U$ satisfy one of the following conditions, then $c_{ij} = \{C \in \Delta : x_j \notin C_{x_i}\}$; otherwise, $c_{ij} = \Delta$.

- (1) $x_i \in Pos_{\Delta}(D)$ and $x_j \notin Pos_{\Delta}(D)$;
- (2) $x_i, x_j \in Pos_{\Delta}(D)$ and $[x_i]_D \cap [x_j]_D = \emptyset$.

Definition 4.4 shows that we have to distinguish each pair of samples that meet the above two conditions by some attributes. This discrimination ensures that the positive domain of the decision is invariant. The concept of constructing a discernibility matrix is similar to that used in classical rough sets.

In [5], a reduction method based on discernibility matrix was also proposed to compute all reducts. To compare these methods, the reduction method proposed in [5] is presented as follows.

Definition 4.5. Let $(U, \Delta, D = \{d\})$ be an inconsistent covering decision system, where $U = \{x_1, x_2, \dots, x_n\}$ and $\Delta = \{C_i : i = 1, 2, \dots, m\}$. By $M(U, \Delta, D)$, we denote an $n \times n$ matrix (c_{ij}) , which is called the discernibility matrix of (U, Δ, D) , such that if $x_i, x_j \in U$ satisfy

- (1) $\Delta_{x_i} \subset POS_{\Delta}(D)$ and $\Delta_{x_j} \not\subset POS_{\Delta}(D)$, then

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

If $x_i, x_j \in U$ satisfy

- (2) $\Delta_{x_i}, \Delta_{x_j} \subset POS_{\Delta}(D)$ and $[x_i]_D \cap [x_j]_D = \emptyset$, then

$$c_{ij} = \{C \in \Delta : (C_{x_i} \not\subseteq C_{x_j}) \wedge (C_{x_j} \not\subseteq C_{x_i})\} \cup \{C_s \wedge C_t : (C_{s_{x_i}} \subseteq C_{s_{x_j}}) \wedge (C_{t_{x_j}} \subseteq C_{t_{x_i}})\}$$

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

By comparing Definitions 4.4 and 4.5, we know that one needs only to compute C_{x_i} for any $x_i \in U$ in Definition 4.4 but must compute both C_{x_i} and Δ_{x_i} in Definition 4.5. Furthermore, the computational complexities of these two discernibility matrices are also different. In Definition 4.5, we need m^2 comparisons of their neighborhoods to distinguish two objects with different decisions, but only m comparisons are needed in Definition 4.4. Hence, the time complexities of the discernibility matrices in Definitions 4.4 and 4.5 are $O(m^2 \cdot n^2)$ and $O(m \cdot n^2)$, respectively. This shows that our proposed method of reduction is simpler than the method presented in [5]. If each covering is a partition, then the proposed method coincides with the classical rough sets in [44]. Thus, the proposed method is a direct generalization of the classical reduction method, and the classical reduction method is just a special case of our proposed method.

In the following, we examine some structural properties of attribute reduction in an inconsistent covering decision system. The complexities of the following theorems are lower than the corresponding theorems presented in [5].

Theorem 4.6. *Let $(U, \Delta, D = \{d\})$ be an inconsistent covering decision system. Then,*

- (1) For any $\mathbf{P} \subseteq \Delta$, $\mathbf{P} \cap c_{ij} \neq \emptyset$ for any $i, j \leq n$ if and only if $\text{Pos}_{\mathbf{P}}(D) = \text{Pos}_{\Delta}(D)$.
- (2) $\text{Core}_D(\Delta) = \{C \in \Delta : c_{ij} = \{C\}\}$.

Proof. (1) \Leftarrow If x_i and x_j do not meet conditions (1) and (2) in Definition 4.4, then $c_{ij} = \Delta$. Hence, $\mathbf{P} \cap c_{ij} \neq \emptyset$. In the following, we prove the other cases:

Assume that there exist $i_0, j_0 \leq n$ such that $\mathbf{P} \cap c_{i_0 j_0} = \emptyset$ ($c_{i_0 j_0} \neq \emptyset$).

By Definition 4.4 we have $x_{j_0} \in C_{x_{i_0}}$ for $\forall C \in \mathbf{P}$, which implies $x_{j_0} \in \mathbf{P}_{x_{i_0}}$. Because $\text{Pos}_{\mathbf{P}}(D) = \text{Pos}_{\Delta}(D)$, it follows from Theorem 4.3 that $x_{j_0} \in \mathbf{P}_{x_{i_0}} \Rightarrow x_{j_0} \in \Delta_{x_{i_0}}$. Thus, $x_{j_0} \in C_{x_{i_0}}$ for any $C \in \Delta$. Hence, $c_{i_0 j_0} = \emptyset$, which is a contradiction to the assumption.

\Rightarrow Because $\mathbf{P} \cap c_{ij} \neq \emptyset$ for any $c_{ij} \neq \emptyset$, we assume that $C \in \mathbf{P} \cap c_{ij}$. No matter the conditions of Definition 4.4 in which x_i and x_j meet, we have that $x_j \notin C_{x_i} \Rightarrow x_j \notin \Delta_{x_i} \Rightarrow x_j \notin \mathbf{P}_{x_i}$. It follows from Theorem 4.3 that $\text{Pos}_{\mathbf{P}}(D) = \text{Pos}_{\Delta}(D)$.

(2) If $C \in \text{Core}_D(\Delta)$, then $\text{Pos}_{\Delta}(D) \neq \text{Pos}_{\Delta - \{C\}}(D)$. By Theorem 4.2, there is at least one pair of $x_i, x_j \in U$ meeting one of the covering reduction conditions such that $x_j \notin \Delta_{x_i} \Rightarrow x_j \in \{\Delta - \{C\}\}_{x_i}$. Combined with $(\Delta - \{C\})_{x_i} = \bigcap_{C' \in \Delta - \{C\}} C_{x_i}$, it follows that $x_j \in C_{x_i}$ for any covering C' in $\Delta - \{C\}$. Assume that $x_j \in C_{x_i}$. By the fact that $\Delta_{x_i} = (\Delta - \{C\})_{x_i} \cap C_{x_i}$, it follows that $x_j \in \Delta_{x_i}$, which is a contradiction. Thus, $x_j \notin C_{x_i}$. This means that only $C \in \Delta$ satisfies $x_j \notin C_{x_i}$. Thus, $c_{ij} = \{C\}$, which implies that $C \in \{C \in \Delta : c_{ij} = \{C\}\}$. Hence, $\text{Core}_D(\Delta) \subseteq \{C \in \Delta : c_{ij} = \{C\}\}$.

Conversely, if $c_{ij} = \{C\}$ for $x_i, x_j \in U$, it follows from Definition 4.4 that x_i and x_j meet one of the covering reduction conditions and $x_j \notin C_{x_i}$, which implies that $x_j \in (\Delta - \{C\})_{x_i}$. It follows from Theorem 4.2 that C is indispensable relative to D , i.e., $C \in \text{Core}_D(\Delta)$. Thus $\text{Core}_D(\Delta) \supseteq \{C \in \Delta : c_{ij} = \{C\}\}$. Therefore, $\text{Core}_D(\Delta) = \{C \in \Delta : c_{ij} = \{C\}\}$. \square

Corollary 4.7. *Suppose that $\mathbf{P} \subseteq \Delta$. \mathbf{P} is thus a reduct of Δ if and only if it is a minimal subset of Δ satisfying $\mathbf{P} \cap c_{ij} \neq \emptyset$, $i, j \leq n$.*

Suppose that $\Delta = \{C_1, C_2, \dots, C_n\}$ is a family of coverings on U and that the corresponding Boolean variable \overline{C}_i ($i \leq n$) is defined for each covering C_i ($i \leq n$). $f(U, \Delta, D)$ is a function on (U, Δ, D) and is defined as

$$f(U, \Delta, D)(\overline{C}_1, \overline{C}_2, \dots, \overline{C}_n) = \bigwedge \{\vee(c_{ij})\}, \quad (i, j \leq n, c_{ij} \neq \emptyset).$$

Then, $f(U, \Delta, D)$ is a Boolean function of (U, Δ, D) and is called a discernibility function or discernibility formula of (U, Δ, D) , where $\vee(c_{ij})$ represents a disjunction operation among the elements in c_{ij} . By the discernibility function, we have the following theorem to compute all reducts.

The following example is employed to compare the reduction method proposed in this section with the method presented in [5]

Example 4.8 5. We now consider a house evaluation problem. Suppose that $U = \{x_1, x_2, \dots, x_{10}\}$ is a set of nine houses, $E = \{\text{price; color; structure; surroundings}\}$ is 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 “surroundings” are {quiet; a little noisy; noisy; quite noisy}.

Price: $C_1 = \{\{x_1, x_2, x_3, x_4, x_6, x_7, x_8, x_9, x_{10}\}, \{x_3, x_4, x_6, x_7\}, \{x_3, x_4, x_5, x_6, x_7\}\}$.

Structure: $C_2 = \{\{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}, \{x_6, x_7, x_8, x_9\}, \{x_{10}\}\}$.

Color: $C_3 = \{\{x_1, x_2, x_3, x_6, x_8, x_9, x_{10}\}, \{x_2, x_3, x_4, x_5, x_6, x_7, x_9\}\}$.

Surroundings: $C_4 = \{\{x_1, x_2, x_3, x_6\}, \{x_2, x_3, x_4, x_5, x_6, x_7\}, \{x_6, x_8, x_9, x_{10}\}, \{x_6, x_7, x_9\}\}$.

The final decision D is

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

The positive domain of D relative to Δ is

$$POS_{\Delta}(D) = \bigcup_{X \in U/D} \underline{\Delta}(X) = \{x_1, x_2, x_3, x_6, x_{10}\}.$$

By Definition 4.5, we construct the discernibility matrix as follows:

$$\begin{bmatrix} \Delta & \Delta & \Delta & \Delta & \Delta & \Delta & \Delta & \Delta & \Delta & \Delta \\ \Delta & \Delta & \Delta & \Delta & \Delta & \Delta & \Delta & \Delta & \Delta & \{C_2, C_4\} \\ \Delta & \Delta & \Delta & \Delta & \Delta & \Delta & \Delta & \Delta & \Delta & \{C_2, C_4\} \\ \{C_3, C_4\} & \{C_3, C_4\} & \{C_3, C_4\} & \Delta & \Delta & \{C_2, C_3, C_4\} & \Delta & \Delta & \Delta & \{C_2, C_3, C_4\} \\ \{C_1, C_3, C_4\} & \{C_1, C_3, C_4\} & \{C_1, C_3, C_4\} & \Delta & \Delta & \Delta & \Delta & \Delta & \Delta & \Delta \\ \Delta & \Delta & \Delta & \Delta & \Delta & \Delta & \Delta & \Delta & \Delta & \{C_2\} \\ \{C_3, C_4\} & \{C_3, C_4\} & \{C_3, C_4\} & \Delta & \Delta & \{C_3, C_4\} & \Delta & \Delta & \Delta & \{C_2, C_3, C_4\} \\ \{C_2, C_4\} & \{C_2, C_3, C_4\} & \Delta & \Delta & \Delta & \Delta & \Delta & \Delta & \Delta & \{C_2\} \\ \{C_2, C_4\} & \{C_2, C_4\} & \{C_1, C_2, C_4\} & \Delta & \Delta & \{C_1, C_2, C_4\} & \Delta & \Delta & \Delta & \{C_2\} \\ \{C_2, C_4\} & \{C_2, C_3, C_4\} & \Delta & \Delta & \Delta & \Delta & \Delta & \Delta & \Delta & \Delta \end{bmatrix}$$

and

$$\begin{aligned} f(U, \Delta)(\overline{C_1}, \dots, \overline{C_4}) &= \wedge \{\vee(c_{ij}) : 1 \leq j \leq i \leq 10\} \\ &= C_2 \wedge (C_2 \vee C_4) \wedge (C_3 \vee C_4) \wedge (C_1 \vee C_3 \vee C_4) \wedge (C_2 \vee C_3 \vee C_4) \wedge (C_1 \vee C_2 \vee C_4) = C_2 \wedge (C_3 \vee C_4) \\ &= (C_2 \wedge C_3) \vee (C_2 \wedge C_4), \end{aligned}$$

Thus, $RED(\Delta, D) = \{\{C_2, C_4\}, \{C_2, C_3\}\}$, $Core_D(\Delta) = \{C_2\}$.

We can see that the discernibility matrix computed by the proposed method is simpler than the method presented in [5]. The results are also the same as those obtained in [5]. We can say that the proposed method is more efficient at computing all of the reducts than the method presented in [5].

5. Experimental analysis

Attribute reduction and feature selection have become the focus of pattern recognition and machine learning tasks. Selecting an optimal feature subset for classification learning is required for many practical problems. Classical rough sets are only applicable to selecting nominal features. Covering rough sets can be used to address more complex data, such as numerical and categorical data. In Sections 3 and 4, we present approaches for finding all reducts of conditional attributes based on covering rough sets, but the computational complexities increase exponentially with the number of samples. In real-world applications, it is not necessary to find all reducts. It is adequate to solve a practical problem by using a minimal reduct. For this reason, we develop a heuristic algorithm to find a minimal subset of features to approximate an optimal reduct. We call the heuristic algorithm the covering discernibility algorithm (CDA) and describe it as follows:

Heuristic algorithm: (Covering discernibility algorithm).

Input: A decision table $DT = (U, A, D)$ and Epsilon $\epsilon // \epsilon$ is a threshold to control the size of the neighborhood of an object.

Output: Reduct Red

- Step1: $Red \leftarrow \{\}$, $Core \leftarrow \{\} // Red$ and $Core$ are the pools that contain the selected features and the features that belong to the core of a reduct, respectively.
- Step2: For any attribute a_k in an attribute set A , compute its discernibility matrix M_k at a threshold ϵ whose elements are 0 and 1. Here, 1 indicates that the corresponding objects are discernible, and 0 indicates that they are not.
- Step3: Compute the discernibility matrices $Total_Discern(A)$ and $Additive_Discern(A)$ of the set $\{M_k : a_k \in A\}$ by disjunction and addition operation, respectively.
- Step4: Find the core of the given decision table.
For each $a_k \in A$, if the intersection of $Additive_Discern(A)$ and M_k is nonempty and some elements in $M_k \cap Additive_Discern(A)$ are equal to 1, then put attribute a_k into $Core$.
- Step5: $Red \leftarrow Core$.
- Step6: For each $b \in A - Red$, compute the discernibility matrix $Total_Discern(Red \cup \{b\})$ of the set $\{M_k : a_k \in Red \cup \{b\}\}$. Select the attribute with maximum discernible power and put it into Red .
- Step7: For each $c \in A - Red$, if the discernible power of $Total_Discern(Red \cup \{c\})$ is equal to the discernible power of $Total_Discern(A)$, then stop. Otherwise, go to step6.

This algorithm can obtain a suboptimal reduct from a given decision table. To show that the proposed algorithm is an effective technique to handle complex data sets and is more efficient than the methods presented in [5], we employ eight datasets from the UCI Machine Learning Repository to verify the performance of the proposed algorithm. The information

Table 2
Description of data sets.

No	Data sets	Samples	Numerical	Categorical	Classes
1	Credit	690	8	7	2
2	Dermatology	366	1	33	6
3	Glass	214	9	0	6
4	Heart	270	5	8	2
5	Ionosphere	351	34	0	2
6	Sonar	208	60	0	2
7	Wdbc	569	30	0	2
8	Wine	178	13	0	3

Table 3
Number of selected features with different reduction algorithms.

Data sets	Raw data	CDA	Covering	FCM + PKRS	ϵ
Credit	15	11	11	11	0.050
Dermatology	34	21	18	8	0.375
Glass	9	8	9	8	0.275
Heart	13	9	8	7	0.125
Ionosphere	34	11	9	7	0.150
Sonar	60	21	19	6	0.300
Wdbc	30	22	23	7	0.250
Wine	13	10	8	5	0.225

Table 4
Running time of the reduction with different algorithms.

Data sets	CDA	Covering	FCM + PKRS	ϵ
Credit	29.35 ± 0.13	120.14 ± 3.81	19.20 ± 2.65	0.050
Dermatology	11.60 ± 0.12	142.52 ± 10.59	24.36 ± 0.43	0.375
Glass	0.73 ± 0.02	1.64 ± 0.01	1.38 ± 0.29	0.275
Heart	1.53 ± 0.09	11.7 ± 1.54	0.43 ± 0.07	0.125
Ionosphere	8.50 ± 0.15	79.59 ± 2.65	2.27 ± 0.08	0.150
Sonar	4.15 ± 0.13	51.65 ± 2.19	2.05 ± 0.06	0.300
Wdbc	41.73 ± 0.10	175.38 ± 11.98	12.24 ± 2.61	0.250
Wine	1.00 ± 0.07	7.65 ± 0.40	0.49 ± 0.07	0.225

from the eight datasets is described in Table 2. Of the eight data sets, five have numerical features: Glass, Ionosphere, Sonar, Wdbc and Wine; the remaining data sets have mixed features.

In the sequence of experiments, we compare the covering discernibility algorithm with both the reduction algorithm in [5] and the classical rough set method. Three indices are used in the comparison: (1) the number of selected attributes with different methods, (2) the running time of selecting a close to minimal reduct, and (3) the classification performances of attributes selected using different methods.

In the following analysis, we apply each of the three algorithms to search for attribute reducts on the eight datasets. To evaluate the quality of selected attributes, the SVM and 3NN learning algorithms are used to train classifiers and test classification accuracies based on 10-fold cross validation. As we know, the classical rough set method considers only categorical data. Hence, a fuzzy c -means clustering (FCM) technique is employed to discretize numerical data before reduction. The numeric attributes are discretized into four intervals. The parameter ϵ is used to control the size of the neighborhood of an object; it has a great impact on different data sets. Because classification performance varies with ϵ , we select different values of ϵ for these datasets. The experimental results of these datasets are shown in Tables 3–6, where the covering discernibility algorithm is abbreviated by CDA, the method proposed in [5] by Covering, and the classical rough set method by FCM + PKRS.

From Tables 3, 5 and 6, we find that all three methods can effectively reduce the datasets' attributes. The FCM + PKRS method selected the smallest number of attributes, and the classification accuracies of the reduced data are also lowest, most of which decrease a little compared with the original data. This result shows that discretization causes information loss. The reductions with the CDA and Covering methods produce higher classification accuracies and keep more attributes in reducts, which shows that there are still some useful attributes in the subsets deleted by FCM + PKRS.

It is notable that the CDA method has a slightly better classification performance than the Covering method in most cases, although the number of attributes selected by each method is comparable. Furthermore, in the majority of cases, the standard deviations of the classification accuracies produced by CDA are lower than those produced by the Covering method because CDA optimizes the Covering method, which is poorly designed. The superiority of the CDA method is also shown in

Table 5

Comparison of the classification accuracies of reduced data with SVM.

Data sets	Raw data	CDA	Covering	FCM + PKRS	ε
Credit	84.78 \pm 4.44	85.51 \pm 6.11	84.78 \pm 5.65	82.90 \pm 4.77	0.050
Dermatology	96.72 \pm 2.13	97.57 \pm 2.37	94.76 \pm 2.98	76.05 \pm 6.70	0.375
Glass	69.56 \pm 9.21	71.39 \pm 9.07	69.56 \pm 9.21	67.83 \pm 7.57	0.275
Heart	76.67 \pm 8.38	81.26 \pm 7.24	81.37 \pm 9.73	76.30 \pm 7.24	0.125
Ionosphere	90.59 \pm 4.69	92.87 \pm 4.52	92.30 \pm 4.49	90.32 \pm 8.43	0.150
Sonar	83.93 \pm 11.45	84.18 \pm 3.81	83.22 \pm 9.87	79.80 \pm 8.34	0.300
Wdbc	96.31 \pm 2.80	97.57 \pm 2.03	97.19 \pm 2.22	96.14 \pm 2.59	0.250
Wine	97.15 \pm 3.99	97.15 \pm 3.99	96.04 \pm 3.79	94.93 \pm 4.88	0.225

Table 6

Comparison of the classification accuracies of reduced data with 3NN.

Data sets	Raw data	CDA	Covering	FCM + PKRS	ε
Credit	84.49 \pm 4.74	84.77 \pm 5.95	85.22 \pm 4.96	84.49 \pm 5.64	0.050
Dermatology	96.49 \pm 2.56	96.18 \pm 3.42	95.56 \pm 4.01	73.15 \pm 6.64	0.375
Glass	67.03 \pm 6.73	67.03 \pm 5.93	67.03 \pm 6.73	66.73 \pm 0.67	0.275
Heart	77.41 \pm 9.48	81.48 \pm 8.46	81.83 \pm 7.25	75.93 \pm 9.44	0.125
Ionosphere	85.74 \pm 5.74	86.87 \pm 7.78	91.16 \pm 5.64	84.93 \pm 7.40	0.150
Sonar	84.94 \pm 7.93	86.04 \pm 7.28	84.17 \pm 11.06	77.32 \pm 8.11	0.300
Wdbc	97.01 \pm 2.19	97.34 \pm 1.38	97.08 \pm 2.34	95.96 \pm 2.87	0.250
Wine	96.11 \pm 4.57	96.04 \pm 3.79	94.38 \pm 3.71	93.75 \pm 5.68	0.225

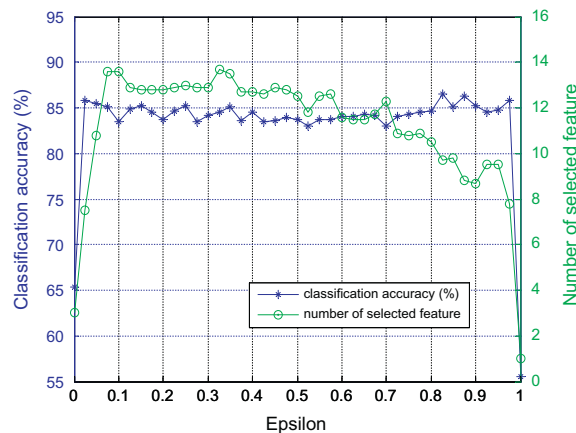
**Fig. 1.** Number of selected features and accuracy varying with the threshold ε (credit).

Table 4, from which we can see that in most datasets, the CDA method greatly reduces the time consumption of the Covering method and that the standard deviations of running time have the similar trends. However, both the CDA and Covering methods spend more time than FCM + PKRS. Let us analyze the reasons in the following.

The procedures that the three methods use to find reducts are basically the same: (1) compute the discernibility matrix, (2) compute the core, and (3) compute the reducts. The difference lies in that the CDA and Covering methods need to use some time to compute the neighborhoods of samples. However, because the CDA method overcomes the design flaws of the Covering method, the efficiency of searching for reducts is greatly improved as showed in Table 4, which corresponds to the theoretical analysis in Sections 3 and 4.

Now, we conduct a series of experiments to find the optimal parameter ε . For perfectly observing the performance variation with ε , we set the value of ε to vary from 0 to 1 with a step of 0.025. In the following, we display only the experimental results of these datasets with SVM in Figs. 1–8. The experimental results obtained using 3NN are roughly consistent those obtained using SVM.

Figs. 1–8 show the numbers of selected features and classification accuracies varying with ε . We find that there are similar trends in these curves. Generally speaking, with an increase of the value of ε , the numbers of selected features rise and, correspondingly, the classification accuracies increase, arrive at a peak, and remain unchanged. Finally, the numbers of se-

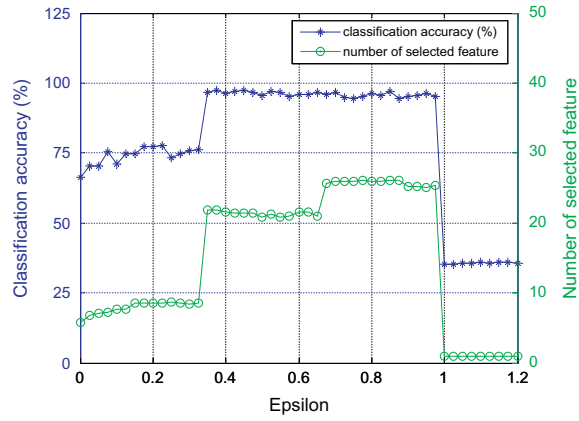


Fig. 2. Number of selected features and accuracy varying with the threshold ϵ (dermatology).

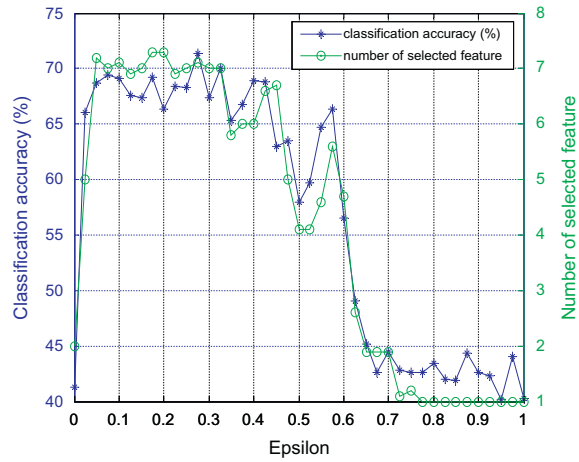


Fig. 3. Number of selected features and accuracy varying with the threshold ϵ (glass).

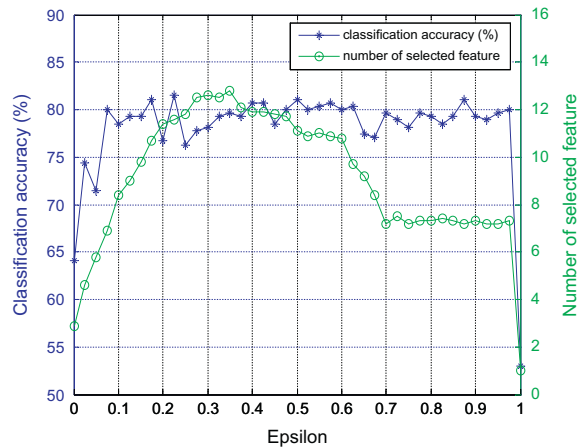


Fig. 4. Number of selected features and accuracy varying with the threshold ϵ (heart).

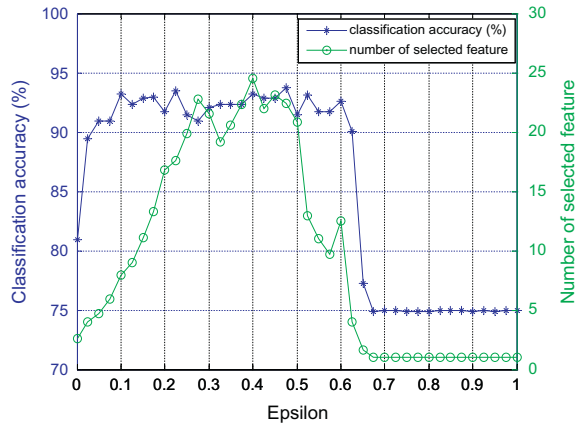


Fig. 5. Number of selected features and accuracy varying with the threshold ϵ (ionosphere).

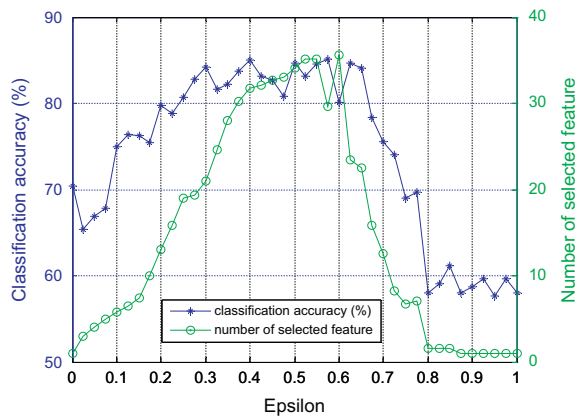


Fig. 6. Number of selected features and accuracy varying with the threshold ϵ (sonar).

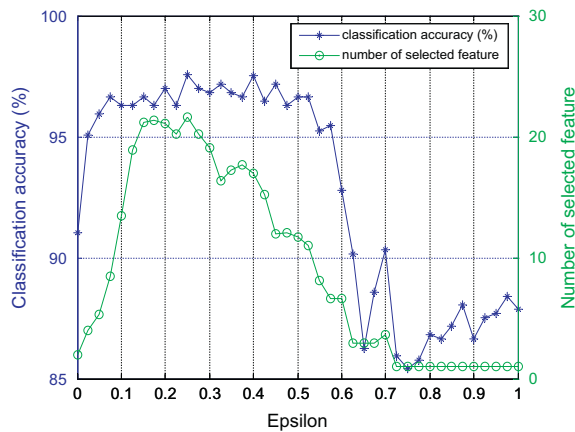


Fig. 7. Number of selected features and accuracy varying with the threshold ϵ (wdbc).

lected features and the accuracies drastically decrease after a threshold. These curves show that the classification performance is stable and can provide a selection of an appropriate subset of features. The optimal positions of classification performance are different between these curves. Here, we recommend that ϵ should take values in the interval [0.05,0.6].

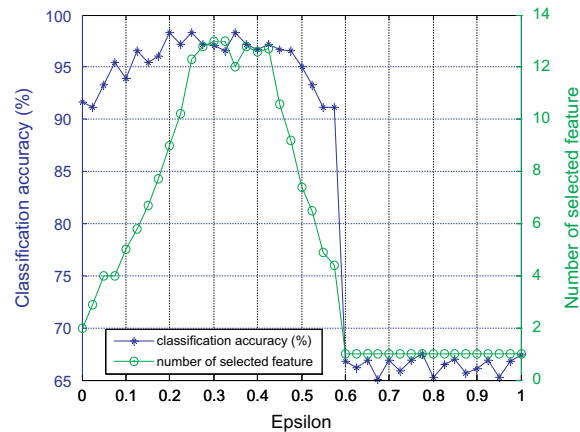


Fig. 8. Number of selected features and accuracy varying with the threshold ε (wine).

6. Conclusion and future work

Attribute reduction is an important step in classification learning. In this paper, we propose a novel method of attribute reduction using covering rough sets to handle datasets with discrete and numerical attributes. We construct the discernibility matrix to compute the reducts in decision systems and analyze the structural properties of attribute reduction. In comparison with the approaches for reduction presented in [5], the time complexity of the proposed reduction method is relatively low. Based on a discernibility matrix, we develop a heuristic algorithm to find a minimal subset of attributes that approximate an optimal reduct. Experiments using UCI datasets show that the proposed reduction approaches represent an effective technique to address complex datasets.

If each covering is a partition, then the proposed method for computing reducts coincides with the traditional rough sets presented in [44]. Thus, the proposed method is a direct generalization of the traditional reduction method. From a theoretical viewpoint, we can compute all of the reducts of a covering decision system. Our study on this topic plays the same important theoretical role as previous studies [44,46] in traditional and generalized rough sets.

As shown in the experimental analysis, the proposed heuristic algorithm can effectively address medium-sized datasets. However, the computational complexity of discernibility matrix increases exponentially with the numbers of samples and attributes. Therefore, it is necessary to develop methods that avoid the construction of a discernibility matrix to conduct attribute reduction and feature selection for large datasets in the future.

Acknowledgements

The authors thank the Language Editor and other anonymous reviewers for their valuable suggestions. This work was supported by the National Basic Research Program of China (2009CB219801-3), the National Natural Science Foundation of China under Grants 61070242, 71171080 and 61222210, the Program for Liaoning Excellent Talents in University under Grant (LR2012039), the Natural Science Foundation of Hebei Province (F2012201023), and the Scientific Research Project of Hebei University (09265631D-2).

References

- [1] Z. Bonikowski, E. Bryniarski, U. Wybraniec, Extensions and intentions in the rough set theory, *Information Sciences* 107 (1998) 149–167.
- [2] J. Błaszczyński, R. Słowiński, M. Szeląg, Sequential covering rule induction algorithm for variable consistency rough set approaches, *Information Sciences* 181 (5) (2011) 987–1002.
- [3] 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.
- [4] D.G. Chen, W.X. Zhang, D. Yeung, E.C.C. Tsang, Rough approximation on a complete completely distributive lattice with applications to generalized rough sets, *Information Sciences* 176 (2006) 1829–1848.
- [5] D.G. Chen, C.Z. Wang, Q.H. Hu, A new approach to attribute reduction of consistent and inconsistent covering decision systems with covering rough sets, *Information Sciences* 177 (17) (2007) 3500–3518.
- [6] D.G. Chen, Q.H. Hu, Y. Yang, Parameterized attribute reduction with Gaussian kernel based fuzzy rough sets, *Information Sciences* 181 (23) (2011) 5169–5179.
- [7] M. Diker, A.A. Uğur, Textures and covering based rough sets, *Information Sciences* 184 (1) (2012) 44–63.
- [8] I. Düntsch, G. Gediga, Uncertainty measures of rough set prediction, *Artificial Intelligence* 106 (1998) 109–137.
- [9] I. Düntsch, G. Gediga, E. Orłowska, Relational attribute systems, *International Journal of Human–Computer Studies* 55 (2001) 293–309.
- [10] J. Dai, Rough set approach to incomplete numerical data, *Information Sciences* 241 (2013) 43–57.
- [11] J. Dai, Q. Xu, Approximations and uncertainty measures in incomplete information systems, *Information Sciences* 198 (2012) 62–80.
- [12] S. Greco, B. Matarazzo, R. Słowiński, Rough approximation of a preference relation by dominance relations, *European Journal of Operational Research* 117 (1999) 63–83.

- [13] S. Greco, B. Matarazzo, R. Slowiński, Rough approximation by dominance relations, *International Journal of Intelligent Systems* 17 (2002) 153–171.
- [14] Q. Hu, W. Pedrycz, D. Yu, J. Lang, Selecting discrete and continuous features based on neighborhood decision error minimization, *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics* 40 (1) (2010) 137–150.
- [15] Q. Hu, D. Yu, J. Liu, C. Wu, Neighborhood rough set based heterogeneous feature subset selection, *Information Sciences* 178 (2008) 3577–3594.
- [16] R. Jensen, Q. Shen, Fuzzy-rough sets assisted attribute selection, *IEEE Transactions on Fuzzy Systems* 15 (1) (2007) 73–89.
- [17] G. Jeon, D. Kim, J. Jeong, Rough sets attributes reduction based expert system in interlaced video sequences, *IEEE Transactions on Consumer Electronics* 52 (4) (2006) 1348–1355.
- [18] M. Kryszkiewicz, Rough set approach to incomplete information systems, *Information Sciences* 112 (1998) 39–49.
- [19] M. Kryszkiewicz, Rule in incomplete information systems, *Information Sciences* 113 (1999) 271–292.
- [20] Y. Leung, D.Y. Li, Maximal consistent block technique for rule acquisition in incomplete information systems, *Information Sciences* 153 (2003) 85–106.
- [21] Y. Leung, W.Z. Wu, W.X. Zhang, Knowledge acquisition in incomplete information systems: a rough set approach, *European Journal Operational Research* 168 (2006) 164–180.
- [22] J.Y. Liang, C.Y. Dang, K.S. Chin, C.M. Yam Richard, A new method for measuring uncertainty and fuzziness in rough set theory, *International Journal of General Systems* 31 (4) (2002) 331–342.
- [23] G.L. Liu, Using one axiom to characterize rough set and fuzzy rough set approximations, *Information Sciences* 223 (2013) 285–296.
- [24] D. Liu, T. Li, D. Ruan, Probabilistic model criteria with decision-theoretic rough sets, *Information Sciences* 181 (17) (2011) 3709–3722.
- [25] J.N. Mordeson, Rough set theory applied to (fuzzy) ideal theory, *Fuzzy Sets and Systems* 121 (2001) 315–324.
- [26] Z. Meng, Z. Shi, Extended rough set-based attribute reduction in inconsistent incomplete decision systems, *Information Sciences* 204 (2012) 44–69.
- [27] J.S. Mi, Y. Leung, W.Z. Wu, An uncertainty measure in partition-based fuzzy rough sets, *International Journal General Systems* 34 (2005) 77–90.
- [28] J.S. Mi, W.Z. Wu, W.X. Zhang, Approaches to knowledge reductions based on variable precision rough sets model, *Information Sciences* 159 (2004) 255–272.
- [29] 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, 1999*, pp. 137–145.
- [30] Z. Pawlak, Rough sets, *International Journal of Computer and Information Sciences* 11 (1982) 341–356.
- [31] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning About Data*, Kluwer Academic Publishing, Dordrecht, 1991.
- [32] Z. Pawlak, A. Skowron, Rudiments of rough sets, *Information Science* 177 (2007) 3–27.
- [33] J.A. Pomykala, Approximation operations in approximation space, *Bulletin of the Polish Academy of sciences* 9–10 (1987) 653–662.
- [34] W. Pedrycz, *Granular Computing: Analysis and Design of Intelligent Systems*, CRC Press, Francis Taylor, Boca Raton, 2013.
- [35] W. Pedrycz, G. Vukovich, Granular worlds: Representation and communication problems, *International Journal of Intelligent Systems* 15 (2000) 1015–1026.
- [36] Y. Qian, J. Liang, W. Pedrycz, C. Dang, Positive approximation: an accelerator for attribute reduction in rough set theory, *Artificial Intelligence* 174 (2010) 597–618.
- [37] Y.H. Qian, J.Y. Liang, D.Y. Li, H.Y. Zhang, C.Y. Dang, Measures for evaluating the decision performance of a decision table in rough set theory, *Information Sciences* 178 (2008) 181–202.
- [38] Y.H. Qian, J.Y. Liang, C.Y. Dang, Converse approximation and rule extracting from decision tables in rough set theory, *Computer and Mathematics with Applications* 55 (2008) 1754–1765.
- [39] K.Y. Qin, Z. Pei, J. Yang, Y. Xu, Approximation operators on complete completely distributive lattices, *Information Sciences* 247 (2013) 123–130.
- [40] Z. Shi, Z. Gong, The further investigation of covering-based rough sets: uncertainty characterization, similarity measure and generalized models, *Information Sciences* 180 (19) (2010) 3745–3763.
- [41] A. Skowron, J. Stepaniuk, Tolerance approximation spaces, *Fundamenta Informaticae* 27 (1996) 245–253.
- [42] A. Skowron, J. Stepaniuk, R. Swiniarski, Modeling rough granular computing based on approximation spaces, *Information Sciences* 184 (1) (2012) 20–43.
- [43] R. Slowinski, D. vanderpooten, A generalized definition of rough approximations based on similarity, *IEEE Transactions on Data and Knowledge Engineering* 2 (2000) 331–336.
- [44] A. Skowron, C. Rauszer, The discernibility matrices and functions in information systems, in: R. Slowinski (Ed.), *Intelligent Decision support, Handbook of Applications and Advances of the Rough Sets Theory*, Kluwer Academic Publishers, 1992.
- [45] E.C.C. Tsang, D. Chen, D.S. Yeung, Approximations and reducts with covering generalized rough sets, *Computers and Mathematics with Applications* 56 (2008) 279–289.
- [46] C. Wang, C. Wu, D. Chen, A systematic study on attribute reduction with rough sets based on general binary relations, *Information Sciences* 178 (2008) 2237–2261.
- [47] C. Wang, C. Wu, C. Wu, Q. Hu, Communicating between Information Systems, *Information Sciences* 178 (2008) 3228–3239.
- [48] C. Wang, D. Chen, B. Sun, Q. Hu, Communication between information systems with covering based rough sets, *Information Sciences* 216 (2012) 17–33.
- [49] L. Wang, X. Yang, J. Yang, C. Wu, Relationships among generalized rough sets in six coverings and pure reflexive neighborhood system, *Information Sciences* 207 (2012) 66–78.
- [50] C. Wang, D. Chen, Q. Hu, A comparative study of ordered and covering information systems, *Fundamenta Informaticae* 123 (2013) 351–363.
- [51] W.Z. Wu, W.X. Zhang, H.Z. Li, Knowledge acquisition in incomplete fuzzy information systems via rough set approach, *Expert Systems* 20 (2003) 280–286.
- [52] W.Z. Wu, M. Zhang, H.Z. Li, J.S. Mi, Knowledge reduction in random information systems via Dempster–Shafer theory of evidence, *Information Sciences* 174 (2005) 143–164.
- [53] W.Z. Wu, Y. Leung, Theory and applications of granular labeled partitions in multi-scale decision tables, *Information Sciences* 181 (18) (2011) 3878–3897.
- [54] A. Wasilewska, Topological rough algebras, in: Lin, Cercone (Eds.), *Rough Sets & Data Mining*, Kluwer Academic Publishers, Boston, 1997, pp. 411–425.
- [55] Y.Y. Yao, Relational interpretations of neighborhood operators and rough set approximation operators, *Information Sciences* 101 (1998) 239–259.
- [56] Y.Y. Yao, Constructive and algebraic method of theory of rough sets, *Information Sciences* 109 (1998) 21–47.
- [57] Y.Y. Yao, B. Yao, Covering based rough set approximations, *Information Sciences* 200 (2012) 91–107.
- [58] Y.Y. Yao, The superiority of three-way decisions in probabilistic rough set models, *Information Sciences* 181 (6) (2011) 1080–1096.
- [59] D. Ye, Z. Chen, S. Ma, A novel and better fitness evaluation for rough set based minimum attribute reduction problem, *Information Sciences* 222 (2013) 413–423.
- [60] W. Ziarko, Variable precision rough set model, *Journal of Computer and System Sciences* 46 (1993) 9–59.
- [61] X. Zhang, B. Zhou, P. Li, A general frame for intuitionistic fuzzy rough sets, *Information Sciences* 216 (2012) 34–49.
- [62] X. Zhang, Y. Yao, H. Yu, Rough implication operator based on strong topological rough algebras, *Information Sciences* 180 (19) (2010) 3764–3780.
- [63] Y. Zhang, M. Luo, On minimization of axiom sets characterizing covering-based approximation operators, *Information Sciences* 181 (14) (2011) 3032–3042.
- [64] W. Zakowski, Approximations in the space (U, Π) , *Demonstratio Mathematica* 16 (1983) 761–769.
- [65] W. Zhu, Fei-Yue Wang, Some results on the covering generalized rough sets, *Pattern Recognition and Artificial Intelligence* 5 (2002) 6–13.
- [66] W. Zhu, F. Wang, Reduction and axiomization of covering generalized rough sets, *Information Sciences* 152 (2003) 217–230.
- [67] W. Zhu, Topological approaches to covering rough sets, *Information Sciences* 177 (2007) 1892–1915.
- [68] W. Zhu, F.Y. Wang, The fourth type of covering-based rough sets, *Information Sciences* 201 (2012) 80–92.