



Quick Maximum Distribution Reduction in Inconsistent Decision Tables

Baizhen Li¹, Wei Chen^{2(✉)}, Zhihua Wei¹, Hongyun Zhang¹, Nan Zhang³,
and Lijun Sun¹

¹ Tongji University, Shanghai 201804, China

² Shanghai Institute of Criminal Science and Technology, Shanghai 200003, China
weichen_82@163.com

³ Yantai University, Yantai 264005, Shandong, China

Abstract. Attribute reduction is a key issue in rough set theory, and this paper focuses on the maximum distribution reduction for complete inconsistent decision tables. It is quite inconvenient to judge the maximum distribution reduct directly according to its definition and the existing heuristic based judgment methods are inefficient due to the lack of acceleration mechanisms that mainstream heuristic judgment methods have. In this paper, we firstly point out the defect of judgment method proposed by Li et al. [15]. After analyzing the root cause of the defect, we proposed two novel heuristic attribute reduction algorithms for maximum distribution reduction. The experiments show that proposed algorithms are more efficient.

Keywords: Rough sets · Attribute reduction · Maximum distribution reduction · Heuristic algorithm

1 Introduction

Rough set theory, introduced by Z. Pawlak [1] in 1982, is an efficient tool to imprecise, incomplete and uncertain information processing [2–5]. Currently, rough set theory has been successfully applied to many practical problems, including machine learning [6, 7], pattern recognition [8, 9], data mining [10], decision support systems [11], etc.

Attribute reduction, the process of obtaining a minimal set of attributes that can preserve the same ability of classification as the entire attribute set, is one of the core concepts in rough set theory [12]. Maximum distribution reduction, proposed as a compromise between the capability of generalized decision preservation reduction and the complexity of distribution preservation reduction [13] by Zhang et al. [14] in 2003, guarantees the decision value with maximum probability of object in inconsistent decision tables unchanged. Subsequently, Pei et al. proposed a theorem for maximum distribution reduct judgment in 2005. Next, Li et al. [15] paid attention to the computational efficiency of reduction definition and designed a new definition of maximum distribution reduction to

speed up attribute reduction. Taking into consideration the general reduction on inconsistent decision tables, Ge et al. [16] proposed new definition of maximum distribution reduction.

Heuristic approaches is one of import method in attribute reduction. The heuristic approach is composed of two parts: the attribute reduction heuristic and the search strategy [17]. The attribute reduction heuristic is the fitness function of a heuristic approach. Existing definitions of heuristics are mainly based on three aspects: dependency degree [18], entropy [19–21], and consistency [22, 23]. The search strategy is the control structure of the heuristic approach. Speaking loosely, the search strategy mainly includes three kinds of methods: the deletion method, the addition method, and the addition-deletion method [24].

Existing methods for the judgment of maximum distribution were weak association with mainstream heuristics. As a result, the efficiency of heuristic maximum distribution reduction algorithm was limited due to lack of the support of acceleration policies that mainstream heuristics have. This paper focuses on the quick reduction algorithms for maximum distribution reduction. At first, we analyze the defect of the quick maximum distribution reduction algorithm (Q-MDRA) proposed in [15] and explore the root cause of its defect. Next, based on the existing mainstream heuristic function, we develop three heuristic maximum distribution reduction algorithms. Finally, we conduct some experiments to evaluate the effectiveness and efficiency of proposed algorithms.

The rest of this paper is organized as follows. In Sect. 2, we review some basic notions related to maximum distribution reduction and three classic heuristic functions. In Sect. 3, we show the defect of Q-MDRA with a calculation example of maximum distribution reduction. After exploring the root cause of its defect, we present three novel algorithms for maximum distribution reduction. In Sect. 4, we evaluate the efficiency of proposed algorithms through algorithm complexity analysis and comparison experiments.

2 Preliminary

In this section, we review some basic notions related to maximum distribution reduction and three classic heuristic functions.

The research object of the rough set theory is called the information system. The information system IS can be expressed as four tuple, *i.e.* $\langle U, A, V, f \rangle$, where U stands for the universe of discourse, a non-empty finite set of objects. A is the set of attributes, $V = \bigcup_{a \in A} V_a$ is the set of all attribute values, and $f : U \times A \rightarrow V$ is an information function that maps an object in U to exactly one value in V_a . For $\forall x \in U, \forall a \in A$, we have $f(x, a) \in V_a$. Specifically in the classification problem, the information table contains two kinds of attributes, which can be characterized by a decision table $DT = (U, C \cup D, V, f)$ with $C \cap D = \emptyset$, where an element of C is called a condition attribute, C is called a condition attribute set, an element D is called a decision attribute, and D is called a decision attribute set.

For the condition attribute set $B \subseteq C$, the indiscernibility relation and discernibility relation of B is respectively defined by $IND(B) = \{ \langle x, y \rangle \in U \times U \mid \forall a \in B, f(x, a) = f(y, a) \}$ and $DIS(B) = \{ \langle x, y \rangle \in U \times U \mid \exists a \in B, f(x, a) \neq f(y, a) \}$. For an object $x \in U$, the equivalence class of x , denoted by $[x]_B$, is defined by $[x]_B = \{ y \in U \mid \langle x, y \rangle \in IND(B) \}$. The family of all equivalence classes of $IND(B)$, *i.e.*, the partition determined by B , is denoted by $U/IND(B)$ or simply U/B . Obviously, $IND(B)$ is reflexive, symmetric and transitive. Meanwhile, $DIS(C)$ is irreflexive, symmetric, but not transitive. Something else needed to be reminded of is that $DIS(C) \cup IND(C) = U \times U, DIS(C) \cap IND(C) = \emptyset$.

One the basis of above notions, the concept of maximum distribution reduction was proposed by Zhang et al. [14] in 2003.

Definition 1. Let $DT = (U, C \cup D, V, f)$ be a decision table, $B \subseteq C$ is a maximum distribution reduct of C if and only if B satisfies

$$\begin{aligned} \forall x \in U, \gamma_B(x) &= \gamma_C(x); \\ \forall B' \subset B, \exists x \in U, \gamma_{B'}(x) &\neq \gamma_C(x), \end{aligned}$$

where $\gamma_C(x) = \{ P_i : P_i \in U/D \wedge |P_i \cap [x]_C| = \max_{P_j \in U/D} (|P_j \cap [x]_C|) \}$.

It is said that B is a maximum distribution consistent attribute set if B satisfies condition (1) mentioned above only. There are two methods of maximum distribution reduction: the discernibility matrix based methods and the heuristic methods. For that the discernibility matrix based methods are low-efficiency, heuristic methods are the more reasonable choice for processing the larger scale data. The heuristic attribute reduction algorithms comprises two parts: the heuristic function and the control strategy. We take the addition strategy based heuristic algorithms as the research object of paper. For the heuristic functions, we take three classic heuristic functions, *i.e.*, the dependency degree, the condition entropy, and the consistency as the alternatives for the construction of improved algorithms.

Definition 2. Given a decision table $DT = (U, C \cup D, V, f)$ and $B \subseteq C, U/B = \{ X_1, X_2, \dots, X_m \}, U/D = \{ Y_1, Y_2, \dots, Y_n \}$, three classic heuristic functions (dependency degree, the consistency and conditional entropy) are defined by:

$$\begin{aligned} (1) \Gamma_B(D) &= \frac{|POS_B(D)|}{|U|}; \\ (2) \delta_B(D) &= |\{ D_j \mid \frac{|[x]_B \cap D_j|}{|[x]_B|} = \max_{k=1}^{|U/D|} \{ \frac{|[x]_B \cap D_k|}{|[x]_B|} \} \}| / |U|; \\ (3) H(D|B) &= - \sum_{i=1}^m P(X_i) \sum_{j=1}^n P(Y_j|X_i) \log P(Y_j|X_i), \text{ where } P(Y_j|X_i) = \\ &|X_i \cap Y_j| / |X_i|, \text{ where } H(B) = - \sum_{i=1}^m P(X_i) \log P(X_i), P(X_i) = |X_i| / |U|. \end{aligned}$$

3 Novel Heuristic Maximum Distribution Reduction Algorithms

In this section, we present two defects in Q-MDRA firstly. After analyzing its cause, we construct two quick heuristic maximum distribution reduction algorithms based on classic heuristic functions.

At first, we want to review the quick maximum distribution reduction algorithm (Q-MDRA) proposed by Li et al. Here. Based upon Definition 1, Li et al. [15] proposed following theorem for the judgment of the maximum distribution reduct.

Theorem 1. *Let $DT = (U, C \cup D, V, f)$ be a decision table and $B \subseteq C$, B is a maximum distribution reduct of C if and only if B satisfies*

$$\begin{aligned} \forall x \in U, \gamma_B^{Md}(D) &= \gamma_C^{Md}(D); \\ \forall B' \subset B, \gamma_{B'}^{Md}(D) &\neq \gamma_C^{Md}(D), \end{aligned}$$

where $\gamma_{B'}^{Md}(D) = \sum_{X \in U/B} \frac{|X \cap P_i: \operatorname{argmax}_{P_i \in U/D} |X \cap P_i||}{|U|}$.

This theorem is expressed by the Theorem 6.11 of Ref. [15]. $\gamma_B^{MD}(D) = \gamma_C^{MD}(D)$ maintains unchanged the scale of the maximum decision classes instead of the maximum decision classes for all of the objects in decision tables. That is to say, B may be not a maximum distribution reduct of C in some special conditions. We present the detail information in Sect. 3.1. Based on the variant of dependency degree heuristic function in Theorem 1, Algorithm 1 was constructed by the way of the addition strategy. Something needed to be reminded of in Algorithm 1 is that we denote the assignment operation as “:=” and use the “=” to represent that two items are on equal term.

Algorithm 1. Quick Maximum Distribution Reduction Algorithm (Q-MDRA)

Require: Decision table $DT = (U, C \cup D, V, f)$

Ensure: A maximum distribution reduct of DT

```

1:  $red := \phi$ 
2: while  $True$  do
3:    $T := red$ 
4:   for  $a \in C - red$  do
5:     if  $\gamma_{red \cup \{a\}}^{Md}(D) > \gamma_T^{Md}(D)$  then
6:        $T := red \cup \{a\}$ 
7:     end if
8:   end for
9:   if  $red = T$  then
10:     $break$ 
11:  else
12:     $red := T$ 
13:  end if
14: end while
15: return  $red$ 

```

3.1 The Defects of Q-MDRA

Here, in the way of calculation example, we show the detail information about that Q-MDRA may not perform well as our expectation. Assume that there is a decision table given as Table 1, we are assigned to get the maximum distribution reduct of Table 1.

Table 1. A decision table

U	a_1	a_2	a_3	d
x_1	0	0	0	0
x_2	1	0	0	1
x_3	1	1	0	0
x_4	1	1	0	1
x_5	1	1	1	0
x_6	1	1	1	1
x_7	1	1	1	1

For Table 1, we know that $U = \{x_1, x_2, \dots, x_7\}$, $C = \{a_1, a_2, a_3\}$, $D = \{d\}$, and obviously we have $U/C = \{X_1, X_2, X_3, X_4\} = \{\{x_1\}, \{x_2\}, \{x_3, x_4\}, \{x_5, x_6, x_7\}\}$ and $U/D = \{P_1, P_2\} = \{\{x_1, x_3, x_5\}, \{x_2, x_4, x_6, x_7\}\}$. According to Definition 1, we know that $\gamma_C(x_1) = \{P_1\}$, $\gamma_C(x_2) = \{P_2\}$; for $x \in X_3$, $\gamma_C(x) = \{P_1, P_2\}$; for $x \in X_4$, we have $\gamma_C(x) = \{P_2\}$.

The process of Q-MDRA for obtaining maximum distribution reduct of Table 1 is shown as follows.

Step 1. $red := \emptyset$.

Step 2. $T := red$, $\gamma_T^{Md}(D) = |P_2|/|U| = 4/7$; $\gamma_{T \cup \{a_1\}}^{Md}(D) = (|P_1 \cap X_1| + |\{x_2, x_3, \dots, x_7\} \cap P_2|)/|U| = 5/7$; $T := T \cup \{a_1\}$; $\gamma_{T \cup \{a_2\}}^{Md}(D) = 4/7$; $\gamma_{T \cup \{a_3\}}^{Md}(D) = 4/7$. Because of $T \neq red$, we operate the assignment of $red := T = \{a_1\}$.

Step 3. $T := red$, $\gamma_T^{Md}(D) = |P_2|/|U| = 5/7$; $\gamma_{T \cup \{a_2\}}^{Md}(D) = 5/7$; $\gamma_{T \cup \{a_3\}}^{Md}(D) = 5/7$. Because T is equal to red , program is over.

Using Q-MDRA we get a collection of attributes $\{a_1\}$. According to Theorem 1, $\{a_1\}$ is a maximum distribution reduct of Table 1 for that $\{a_1\}$ satisfies $\gamma_{\{a_1\}}^{Md}(D) = \gamma_{\{C\}}^{Md}(D) = 5/7$ and $\gamma_{\phi}^{Md}(D) \neq 5/7$. But checking it with original Definition 1, we know that $\{a_1\}$ is not a maximum distribution reduct for Table 1 because $\gamma_{\{a_1\}}(x_3) = \{P_2\} \neq \gamma_C(x_3) = \{P_1, P_2\}$. Consequently, Theorem 1 is incorrect.

Here we analyze the root of the defect of Theorem 1. Given a decision table $DT = (U, C \cup D, V, f)$, $U/C = \{X_1, X_2, \dots, X_n\}$, $U/D = \{P_1, P_2, \dots, P_m\}$. Let $maxcf(X_i) = \max_{P_j \in U/D} (|P_j \cap X_i|)$, we have $\gamma_C^{Md}(D) = \sum_{X_i \in U/C} \frac{maxcf(X_i)}{|U|}$.

Assume that $x_1 \in X_1, x_2 \in X_2, \gamma_C(x_1) \neq \gamma_C(x_2), |\gamma_C(x_1)| > 1, |\gamma_C(x_2)| > 1, |\gamma_C(x_1) \cap \gamma_C(x_2)| \geq 1$ and $B \subseteq C, U/B = \{X_1 \cup X_2, X_3, \dots, X_n\}$, it is obvious that $maxcf(X_1) + maxcf(X_2) = maxcf(X_1 \cup X_2)$ and $\gamma_C^{Md}D = \gamma_B^{Md}(D)$. But for $x \in X_1 \cup X_2, \gamma_B(x) = \gamma_C(x_1) \cap \gamma_C(x_2)$, it is not equal to $\gamma_C(x_1)$ or $\gamma_C(x_2)$. The measure $\gamma_C^{Md}(D)$, used in Theorem 1, is not sensitive to the change of the maximum decision classes of objects that have two or more than two maximum decision classes.

On the other side, an attribute set *red* outputted by Q-MDRA does not always satisfy $\gamma_{red}^{Md}(D) = \gamma_C^{Md}(D)$. The reason is that $\forall a \in C - red, \gamma_{red \cup \{a\}}^{Md}(D) = \gamma_{red}^{Md}(D)$ does not guarantee $\gamma_{red}^{Md}(D) = \gamma_C^{Md}(D)$. That is to say, $\forall a \in C - red, \gamma_{red \cup \{a\}}^{Md}(D) = \gamma_{red}^{Md}(D)$ is not conflicted with $\exists B \subseteq C - red, \gamma_{red \cup B}^{Md}(D) > \gamma_{red}^{Md}(D)$.

3.2 Novel Maximum Distribution Reduction Algorithms

To solve the problems identified in Q-MDRA, the concept of indiscernibility relation and discernibility relation of maximum distribution with respect to the specific attribute set are defined. Firstly, Next, the maximum distribution reduct is defined using the indiscernibility relation of maximum distribution. Finally, we construct heuristic maximum distribution reduction algorithms with classic heuristic functions.

Definition 3. Given a decision table $DT = (U, C \cup D, V, f)$, the indiscernibility relation of maximum distribution of U with respect to $B \subseteq C$ is defined as $IND_{md}(B) = \{< x, y > | x, y \in U, \gamma_B(x) = \gamma_B(y)\}$, and the discernibility relation of maximum distribution of U with respect to B stands for $DIS_{md}(B) = \{< x, y > | x, y \in U, \gamma_B(x) \neq \gamma_B(y)\}$.

Obviously, $IND_{md}(C)$ is reflexive, symmetric and transitive; $DIS_{md}(C)$ is irreflexive, symmetric, but not transitive. It is worth noting that $IND_{md}(C) \cup DIS_{md}(C) = U \times U, IND_{md}(C) \cap DIS_{md}(C) = \emptyset$.

Theorem 2. Given $DT = (U, C, D, V, f)$, B is a maximum distribution consistent attribute set of C if and only if B satisfies $IND(C) \subseteq IND(B) \subseteq IND_{md}(C), DIS_{md}(C) \subseteq DIS(B) \subseteq DIS(C)$.

Proof. It is apparent that $DIS(B) \subseteq DIS(C), IND(C) \subseteq IND(B)$, and based on $IND(B) \cap DIS(B) = \emptyset, IND(B) \cup DIS(B) = U \times U, IND_{md}(C) \cap DIS_{md}(C) = \emptyset, IND_{md}(C) \cup DIS_{md}(C) = U \times U$, we know that $DIS(C) \subseteq DIS(B) \subseteq DIS_{md}(C)$ is equal to $IND(C) \subseteq IND(B) \subseteq IND_{md}(C)$. Thus what all we need is to prove that $DIS_{md}(C) \subseteq DIS(B)$ is true.

- Sufficiency(\Rightarrow): Assume that if B is a maximum distribution consistent attribute set then $DIS_{md}(C) \not\subseteq DIS(B)$. $DIS_{md}(C) \not\subseteq DIS(B)$ means $\exists < x, y > \in DIS_{md}(C), < x, y > \notin DIS(B)$. Then we know $\gamma_C(x) \neq \gamma_C(y)$ and $\gamma_B(x) = \gamma_B(y)$. It is conflicted with our assumption. So if B is a maximum distribution consistent attribute set, then $DIS_{md}(C) \subseteq DIS(B)$.

- Necessity(\Leftarrow): Assume that if B satisfies $DIS_{md}(C) \subseteq DIS(B)$ then $\exists x \in U, \gamma_B(x) \neq \gamma_C(x)$. According to the assumption, we know $\exists y \in [x]_B - [x]_C, \gamma_C(y) \neq \gamma_C(x)$. That is to say, $\langle x, y \rangle \in DIS_{md}(C), \langle x, y \rangle \notin DIS(B)$. It is conflicted with $DIS_{md}(C) \subseteq DIS(B)$. Consequently we know if B satisfies $DIS_{md}(C) \subseteq DIS(B)$ then $\forall x \in U, \gamma_B(x) = \gamma_C(x)$.

As mentioned above, Theorem 2 is true. □

Above theorem is good for understanding but it is not friendly in computing. So we represent maximum distribution reduction in the way of classic heuristic functions. According to Definition 2, we can present the definition of the maximum distribution reduct by conditional entropy.

Theorem 3. *Given a decision table $DT = (U, C, D, V, f)$, Let $TGran$ stands for $U/IND_{md}(C)$, $B \subseteq C$ is a maximum distribution reduct if and only if B satisfies*

- (1) $H(TGran|B) = 0$;
- (2) $\forall B' \subset B, B'$ doesn't satisfy condition (1).

Proof. On the basis of Theorem 2, we can prove this theorem by explaining the equivalence relation between $H(TGran|B) = 0$ and $DIS_{md}(C) \subseteq DIS(B)$.

- Sufficiency(\Rightarrow): According to the definition of $H(Q|P)$, it is easy to know that $H(TGran|B) = 0 \Leftrightarrow \forall Y \in TGran, \exists\{X : X \in U/B \wedge X \cap Y \neq \phi\}, \bigcup_{X_i \in X} X_i = Y$. Therefore, we conclude that $DIS_{md}(C) \subseteq DIS(B), IND_{md}(C) \supseteq IND(B)$. As a result, $H(TGran|B) = 0 \Rightarrow B$ is a maximum distribution consistent attribute set.
- Necessity(\Leftarrow): Assume that B is a maximum distribution consistent attribute set, and B satisfies $H(TGran|B) \neq 0$. According to the definition of conditional entropy, we know $H(TGran|B) \neq 0$ means $\exists Y \in TGran, X \in U/B$ satisfies $X \cap Y \neq \phi \wedge X \not\subseteq Y$. That is to say, $\exists p \in X - X \cap Y, q \in X \cap Y, \gamma_C(p) \neq \gamma_C(q), \gamma_B(p) = \gamma_B(q)$. This concludes a conflict with B is a maximum distribution reduct. That is to say, if B is a maximum distribution consistent attribute set, then $H(TGran|B) = 0$.

As a result, Theorem 3 is true. □

According to Definition 2, we can use dependency degree for the presentation of the maximum distribution reduct.

Theorem 4. *Given a decision table $DT = (U, C, D, V, f)$, Let $TGran$ stands for $U/IND_{md}(C)$, $B \subseteq C$ is a maximum distribution reduct if and only if B satisfies (1) $\Gamma_B(TGran) = 1$; (2) $\forall B' \subset B, B'$ doesn't satisfy condition (1).*

Proof. According to Theorem 2 and Theorem 3, the conclusion is clearly established.

For that $\Gamma_C(D) = 1 \Leftrightarrow \delta_C(D) = 1$, we have $\Gamma_B(TGran) = 1 \Leftrightarrow \delta_B(TGran) = 1$. As a result, there is no need to construct a theorem for maximum distribution reduction with $\delta_B(TGran)$. Based on upon theorems, the significance functions for maximum distribution reduction can be defined as follows.

- (1) $Sig_1^{outer}(a, B, TGran) = H(TGran|B) - H(TGran|B \cup \{a\}), a \notin B;$
 $Sig_2^{outer}(a, B, TGran) = \Gamma_B(TGran) - \Gamma_{B \cup \{a\}}(TGran), a \notin B.$
- (2) $Sig_1^{inner}(a, B, TGran) = H(TGran|B - \{a\}) - H(TGran|B), a \in B;$
 $Sig_2^{inner}(a, B, TGran) = \Gamma_B(TGran) - \Gamma_{B - \{a\}}(TGran), a \in B.$

For convenience of algorithm description, we denote $Sig_i^j(a, B, TGran, U')$, $i \in \{1, 2\}, j \in \{inner, outer\}$ as the significance value computed in U' . Using Theorem 3 and Theorem 4, we can construct Algorithms 2 and 3 for maximum distribution reduction. Algorithms 2 and 3, indeed, are the variant of the discernibility matrix based reduction algorithms. The difference of two algorithms to the discernibility matrix based algorithms is the focus paid toward the indiscernibility relation instead of the discernibility relation. It can be proved by extending the relation of $IND(B) \cup DIS(B) = U \times U$ to the reduction algorithms. As a result, in intuition, the correctness of two algorithms can be transmitted from the discernibility matrix based algorithm for obtaining maximum distribution reducts.

Algorithm 2. Maximum Distribution Reduction Algorithm Using Condition Entropy (MDRAUCE)

Require: Decision table $DT = (U, C \cup D, V, f)$

Ensure: A maximum distribution reduct of DT

- 1: $red := \emptyset;$
 - 2: $TGran := U/IND_{md}(C);$
 - 3: $U' := U;$
 - 4: **while** $U_r \neq \emptyset$ **do**
 - 5: Calculate $a_{max} : a_{max} = argmax_{a \in C - red} Sig_1^{outer}(a, red, TGran, U_r);$
 - 6: $red := red \cup \{a_{max}\};$
 - 7: $U' := U' - POS_{red}(TGran);$
 - 8: $TGran := TGran - POS_{red}(TGran);$
 - 9: **end while**
 - 10: **return** red
-

4 Correctness Analysis and Experiments Results

The objective of this section is to present the correctness and the efficiency of the attribute reduction algorithms proposed in this paper, *i.e.* MDRAUCE and MDRAUDD. To show the correctness of two algorithms, we calculate the maximum distribution reduct of Table 1 using MDRAUCE and MDRAUDD, and check outputs of two algorithms with the definition of maximum distribution

Algorithm 3. Maximum Distribution Reduction Algorithm Using Dependency Degree (MDRAUDD)

Require: Decision table $DT = (U, C \cup D, V, f)$
Ensure: A maximum distribution reduct of DT

- 1: $red := \emptyset$;
- 2: $TGran := U/IND_{md}(C)$;
- 3: $U' := U$;
- 4: **while** $U' \neq \emptyset$ **do**
- 5: Calculate $a_{max} : a_{max} = \operatorname{argmax}_{a \in C-red} Sig_2^{outer}(a, red, TGran, U')$;
- 6: $red := red \cup \{a_{max}\}$;
- 7: $U' := U' - POS_{red}(TGran)$;
- 8: $TGran := TGran - POS_{red}(TGran)$;
- 9: **end while**
- 10: **return** red

reduction for validation. On the other side, we employed 12 UCI data sets to verify the performance of time consumption of MDRAUCE, MDRAUDD, and existing maximum distribution reduction algorithms.

4.1 The Validation of Correctness

In this part, we show the correctness of two algorithms proposed in Sect. 3 through presenting the process of calculating the maximum distribution reduct for Table 1 using Algorithm 2 and Algorithm 3. After that, we check the outputs of two algorithms according to the maximum distribution definition.

The process of MDRAUCE for finding the maximum distribution reduct of Table 1 is presented here. In the following description of calculation process, “item1=item2” denotes that the relationship of two are on equal item, and “:=” stands for the assignment operation.

Step 1. $red := \emptyset, TGran = U/IND_{md}(C) = \{\{x_1\}, \{x_2, x_5, x_6, x_7\}, \{x_3, x_4\}\}, U' = \{x_1, x_2, \dots, x_7\}$.

Step 2. $Sig_1^{outer}(a_1, red, TGran, U') = H^{U'}(TGran|red) - H^{U'}(TGran|red \cup \{a_1\}) = 1.38 - 0.79 = 0.59$, $Sig_{md}^{outer}(a_2, red, TGran, U') = H^{U'}(TGran|red) - H^{U'}(TGran|red \cup \{a_2\}) = 1.38 - 0.98 = 0.40$, $Sig_{md}^{outer}(a_3, red, TGran, U') = H^{U'}(TGran|red) - H^{U'}(TGran|red \cup \{a_3\}) = 1.38 - 0.86 = 0.52$. So $a_{max} = a_1$, $red := red \cup \{a_1\} = \{a_1\}$. We have $POS_{red}(TGran) = \{x_1\}$. U' and $TGran$ are updated as follows, $U' := U' - POS_{red}(TGran) = \{x_2, x_3, \dots, x_7\}$, $TGran = TGran - POS_{red}(TGran) = \{\{x_2, x_5, x_6, x_7\}, \{x_3, x_4\}\}$.

Step 3. $Sig_1^{red}(a_2, red, TGran, U') = H^{U'}(TGran|red) - H^{U'}(TGran|red \cup \{a_2\}) = 0.92 - 0.81 = 0.11$, $Sig_{md}^{outer}(a_3, red, TGran, U') = H^{U'}(TGran|red) - H^{U'}(TGran|red \cup \{a_3\}) = 0.92 - 0.46 = 0.46$. So $a_{max} = a_3$, $red := red \cup \{a_3\} = \{a_1, a_3\}$. We have $POS_{red}(TGran) = \{x_5, x_6, x_7\}$. U' and $TGran$ are updated

as follows, $U' := U' - POS_{red}(TGran) = \{x_2, x_3, x_4\}$, $TGran = TGran - POS_{red}(TGran) = \{\{x_2\}, \{x_3, x_4\}\}$.

Step 4. $Sig_{md}^{outer}(a_2, red, TGran, U') = H^{U'}(TGran|red) - H^{U'}(TGran|red \cup \{a_2\}) = 0.92 - 0 = 0.92$. So $a_{max} = a_2$, $red := red \cup \{a_2\} = \{a_1, a_3, a_2\}$. We have $POS_{red}(TGran) = \{x_2, x_3, x_4\}$. U' and $TGran$ are updated as follows, $U' := U' - POS_{red}(TGran) = \emptyset$, $TGran = TGran - POS_{red}(TGran) = \emptyset$.

Step 5. Because $U' = \emptyset$, program is over. Algorithm outputs $red = \{a_1, a_3, a_2\}$ as the result.

The process of MDRAUDD for obtaining the maximum distribution reduct of Table 1 is presented as follows.

Step 1. $red := \emptyset$, $TGran = U/IND_{md}(C) = \{\{x_1\}, \{x_2, x_5, x_6, x_7\}, \{x_3, x_4\}\}$, $U' = \{x_1, x_2, \dots, x_7\}$.

Step 2. $Sig_1^{outer}(a_1, red, TGran, U') = \Gamma_{red \cup \{a_1\}}^{U'}(TGran) - \Gamma_{red}^{U'}(TGran) = \frac{1}{7} - 0 = \frac{1}{7}$, $Sig_1^{outer}(a_2, red, TGran, U') = \Gamma_{red \cup \{a_2\}}^{U'}(TGran) - \Gamma_{red}^{U'}(TGran) = 0 - 0 = 0$, $Sig_1^{outer}(a_3, red, TGran, U') = \Gamma_{red \cup \{a_3\}}^{U'}(TGran) - \Gamma_{red}^{U'}(TGran) = \frac{3}{7} - 0 = \frac{3}{7}$. So $a_{max} = a_3$, $red := red \cup \{a_3\} = \{a_3\}$. We have $POS_{red}(TGran) = \{x_5, x_6, x_7\}$. U' and $TGran$ are updated as follows, $U' := U' - POS_{red}(TGran) = \{x_1, x_2, x_3, x_4\}$, $TGran = TGran - POS_{red}(TGran) = \{\{x_1\}, \{x_2\}, \{x_3, x_4\}\}$.

Step 3. $Sig_1^{outer}(a_1, red, TGran, U') = \Gamma_{red \cup \{a_1\}}^{U'}(TGran) - \Gamma_{red}^{U'}(TGran) = \frac{1}{4} - 0 = \frac{1}{4}$, $Sig_1^{outer}(a_2, red, TGran, U') = \Gamma_{red \cup \{a_2\}}^{U'}(TGran) - \Gamma_{red}^{U'}(TGran) = \frac{2}{4} - 0 = 0.5$. So $a_{max} = a_2$, $red := red \cup \{a_2\} = \{a_3, a_2\}$. We have $POS_{red}(TGran) = \{x_3, x_4\}$. U' and $TGran$ are updated as follows, $U' := U' - POS_{red}(TGran) = \{x_1, x_2\}$, $TGran = TGran - POS_{red}(TGran) = \{\{x_1\}, \{x_2\}\}$.

Step 4. $Sig_1^{outer}(a_1, red, TGran, U') = \Gamma_{red \cup \{a_1\}}^{U'}(TGran) - \Gamma_{red}^{U'}(TGran) = 1 - 0 = 1$. So $a_{max} = a_1$, $red := red \cup \{a_1\} = \{a_3, a_2, a_1\}$. We have $POS_{red}(TGran) = \{x_1, x_2\}$. U' and $TGran$ are updated as follows, $U' := U' - POS_{red}(TGran) = \emptyset$, $TGran = TGran - POS_{red}(TGran) = \emptyset$.

Step 5. Because $U' = \emptyset$, program is over. Algorithm outputs $red = \{a_3, a_2, a_1\}$ as the result.

According to Definition 1, we know $\gamma_{red}(x_1) = \{P_1\}$ and $\gamma_{red}(x_2) = \{P_2\}$; for $x \in \{x_3, x_4\}$, we have $\gamma_{red}(x) = \{P_1, P_2\}$; for $x \in \{x_5, x_6, x_7\}$, we know $\gamma_{red}(x) = \{P_2\}$. Meanwhile, we know $\gamma_C(x_1) = \{P_1\}$ and $\gamma_C(x_2) = \{P_2\}$; for $x \in \{x_3, x_4\}$ $\gamma_C(x) = \{P_1, P_2\}$; for $x \in \{x_5, x_6, x_7\}$, we have $\gamma_C(x) = \{P_2\}$. It is obvious that for $\forall x \in U$, $\gamma_{red}(x) = \gamma_C(x)$. Finally we know that MDRAUCE and MDRAUDD are correct.

4.2 The Efficiency of Proposed Algorithms

In this part, we employed 12 data sets to verify the performance of time consumption of MDRAUDD, MDRAUDD, Q-MDRA [15] and QGARA-FS [16]. We carried out all the attribute reduction algorithms in experiments on a personal computer with Windows 10, Intel(R) Core(TM) CPU i5-8265U 1.60GHZ and 8GB RAM memory. The software used was Visual Studio Code 1.3.8, and the programming language was python 3.7.

The data sets used in experiments are all downloaded from UCI repository of machine learning data sets [25] whose basic information is outlined in Table 2. For the sake that reduction algorithms can address only symbolic data, data sets containing continuous attributes were preprocessed by CAIM [26] discretization algorithm. For each data sets, the positive region dependency degree, *i.e.* $\gamma_C(D)$, is listed in the last column of Table 2. As we know, the data set is consistent if $\gamma_C(D) = 1$; otherwise, it is inconsistent. As shown in Table 2, Wpbc, Wine, and Sonar are consistent. Taking into consideration the value of $\gamma_C(D)$, we take Sat, Segment, Wdbc, and Wave as consistent data sets whose value of $\gamma_C(D)$ satisfies $0.981 \leq \gamma_C(D) \leq 1$. The other 5 data sets (Vehicle, Ion, Glass, Heart, and Pid) are inconsistent.

Table 2. Description of data sets

ID	Data sets	Cases	Attributes	Classes	$\gamma_C(D)$
1	Wpdc	198	34	2	1
2	Wine	178	13	3	1
3	Sat	6435	86	6	0.993
4	Segment	2310	19	7	0.991
5	Wdbc	569	30	2	0.989
6	Waveform	5000	21	3	0.981
7	Vehicle	846	18	4	0.946
8	Ions	351	34	2	0.940
9	Glass	214	9	7	0.937
10	Heart	270	6	2	0.935
11	Sonar	208	60	2	1
12	Pid	768	8	2	0.519

Table 3 indicate the computational time of MDRAUCE, MDRADD, Q-MDRA, and QGARA-FS for obtaining maximum distribution reduct on 12 data sets. We can see that MDRADD was the fastest of four attribute reduction algorithms for that it was the best on 11 data sets, and MDRAUCE was faster than QGARA-FS. MDRAUCE performed better than Q-MDRA in obtaining the

Table 3. Time consumption of maximum distribution reduction algorithms

Data sets	Time consumption(s)			
	MDRAUCE	MDRADD	Q-MDRA	QGARA-FS
Wpdc	0.253	0.208	0.447	1.110
Wine	0.051	0.051	0.045	0.166
Sat	5.395	3.220	59.683	16.604
Segment	1.027	0.546	1.743	1.213
Wdbc	0.591	0.452	0.915	2.411
Waveform	3.607	1.634	11.820	2.551
Vehicle	0.463	0.304	1.587	0.633
Ions	0.612	0.300	0.572	1.582
Glass	0.034	0.034	0.064	0.092
Heart	0.087	0.040	0.121	0.105
Sonar	0.649	0.411	0.621	1.832
Pid	0.101	0.081	0.100	0.137
Average	1.073	0.607	6.477	2.370

reduct of 9 data sets. Q-MDRA performed better than MDRAUCE, MDRAUCE on small data sets, *i.e.* Wine data set. However, in processing the large scale data, Q-MDRA consumed more time than MDRAUCE, MDRADD. From results of experiments on both consistent and inconsistent decision tables, the computational times of four algorithms in obtaining the maximum distribution reduct followed this order: MDRADD \geq MDRAUCE, Q-MDRA $>$ QGARA-FS. For most of the cases in experiments, the computational time of MDRAUDE can reduce half of the computation time of QGARA-FS and Q-MDRA, such as data sets Wpdc, Glass, Heart, etc. In the same condition. from the row of average time consumption in obtaining reduct of 12 data sets, we know that MDRAUCE and MDRADD are more efficient and steady in time consumption of maximum distribution reduction than existing maximum distribution reduction algorithms.

5 Conclusion

In this paper, we focus on the maximum distribution reduction for complete inconsistent decision tables. The problems in Li's algorithm for obtaining the maximum distribution reduct were pointed out, and based on classic heuristic functions, we designed two novel heuristic algorithms, *i.e.* MDRAUCE and MDRADD, to efficiently finding a maximum distribution reduct. Because the scale of data processed becomes larger and larger, the efficiency of attribute reduction algorithms is still our focus of future researches.

Acknowledgement. A special thank is owed to professor Li Min, the discussing with him on Q-MDRA contributes a lot to this paper. The work is partially supported by the National Key Research and Development Project (No. 213), the National Nature Science Foundation of China (No. 61976160, No. 61673301) and Key Lab of Information Network Security, Ministry of Public Security (No. C18608).

References

1. Pawlak, Z.: Rough sets. *Int. J. Comput. Inf. Sci.* **11**(5), 341–356 (1982)
2. Huang, K.Y., Li, I.-H.: A multi-attribute decision-making model for the robust classification of multiple inputs and outputs datasets with uncertainty. *Appl. Soft Comput.* **38**, 176–189 (2016)
3. Dai, J., Xu, Q.: Approximations and uncertainty measures in incomplete information systems. *Inf. Sci.* **198**, 62–80 (2012)
4. Shi, J., Lei, Y., Zhou, Y., Gong, M.: Enhanced rough-fuzzy c-means algorithm with strict rough sets properties. *Appl. Soft Comput.* **46**, 827–850 (2016)
5. Zhan, J., Ali, M.I., Mehmood, N.: On a novel uncertain soft set model: Z-soft fuzzy rough set model and corresponding decision making methods. *Appl. Soft Comput.* **56**, 446–457 (2017)
6. Das, R.T., Ang, K.K., Quek, C.: ieRSPOP: a novel incremental rough set-based pseudo outer-product with ensemble learning. *Appl. Soft Comput.* **46**, 170–186 (2016)
7. Xie, X., Qin, X., Yu, C., Xu, X.: Test-cost-sensitive rough set based approach for minimum weight vertex cover problem. *Appl. Soft Comput.* **64**, 423–435 (2018)
8. Hu, Y.C.: Flow-based tolerance rough sets for pattern classification. *Appl. Soft Comput.* **27**, 322–331 (2015)
9. Huang, K.Y.: An enhanced classification method comprising a genetic algorithm, rough set theory and a modified PBMF-index function. *Appl. Soft Comput.* **12**(1), 46–63 (2012)
10. Wang, F., Liang, J., Dang, C.: Attribute reduction for dynamic data sets. *Appl. Soft Comput.* **13**(1), 676–689 (2013)
11. Kaya, Y., Uyar, M.: A hybrid decision support system based on rough set and extreme learning machine for diagnosis of hepatitis disease. *Appl. Soft Comput.* **13**(8), 3429–3438 (2013)
12. Thangavel, K., Pethalakshmi, A.: Dimensionality reduction based on rough set theory: a review. *Appl. Soft Comput.* **9**(1), 1–12 (2009)
13. Kryszkiewicz, M.: Comparative study of alternative types of knowledge reduction in inconsistent systems. *Int. J. Intell. Syst.* **16**(1), 105–120 (2001)
14. Zhang, W.X., Mi, J.S., Wu, W.Z.: Knowledge reductions in inconsistent information systems. *Chin. J. Comput.-Chin. Ed.* **26**(1), 12–18 (2003)
15. Li, M., Shang, C.X., Feng, S.Z., Fan, J.P.: Quick attribute reduction in inconsistent decision tables. *Inf. Sci.* **254**, 155–180 (2014)
16. Ge, H., Li, L.S., Xu, Y., Yang, C.J.: Quick general reduction algorithms for inconsistent decision tables. *Int. J. Approx. Reason.* **82**, 56–80 (2017)
17. Yao, Y., Zhao, Y., Wang, J.: On reduct construction algorithms. In: Gavrilova, M.L., Tan, C.J.K., Wang, Y., Yao, Y., Wang, G. (eds.) *Transactions on Computational Science II*. LNCS, vol. 5150, pp. 100–117. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-87563-5_6
18. Hu, X.H., Cercone, N.: Learning in relational databases: a rough set approach. *Comput. Intell.* **11**(2), 323–338 (1995)

19. Tian, J., Wang, Q., Yu, B., Yu, D.: A rough set algorithm for attribute reduction via mutual information and conditional entropy. In: 2013 10th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), pp. 567–571. IEEE (2013)
20. Sun, H.R., Wang, R., Xie, B.X., Tian, Y.: Continuous attribute reduction method based on an automatic clustering algorithm and decision entropy. In: Control Conference (2017)
21. Yan, T., Han, C.Z.: Entropy based attribute reduction approach for incomplete decision table. In: International Conference on Information Fusion (2017)
22. Dash, M., Liu, H.: Consistency-based search in feature selection. *Artif. Intell.* **151**(1), 155–176 (2003)
23. Hu, Q., Zhao, H., Xie, Z., Yu, D.: Consistency based attribute reduction. In: Zhou, Z.-H., Li, H., Yang, Q. (eds.) PAKDD 2007. LNCS (LNAI), vol. 4426, pp. 96–107. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-71701-0_12
24. Wang, J., Miao, D.Q.: Analysis on attribute reduction strategies of rough set. *J. Comput. Sci. Technol.* **13**(2), 189–192 (1998)
25. Asuncion, A., Newman, D.J.: UCI machine learning repository (2007)
26. Feng, S.Z., Li, M., Deng, S.B., Fan, J.P.: An effective discretization based on class-attribute coherence maximization. *Pattern Recogn. Lett.* **32**, 1962–1973 (2011)