Information Sciences 326 (2016) 297–314

Contents lists available at ScienceDirect

Information Sciences

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

Efficient attribute reduction from the viewpoint of discernibility

Shu-Hua Teng [a], Min Lu [a], A-Feng Yang [a], Jun Zhang [a], Yongjian Nian [b], Mi He [b,*]

[a] Science and Technology on Automatic Target Recognition Laboratory, National University of Defense Technology, Changsha 410073, China
[b] School of Biomedical Engineering, Third Military Medical University, Chongqing 400038, China

ARTICLE INFO

Article history:
Received 23 May 2013
Revised 30 June 2015
Accepted 29 July 2015
Available online 8 August 2015

Keywords:
Rough set
Discernibility viewpoint
Attribute reduction
Attribute significance

ABSTRACT

Attribute reduction is an important preprocessing step in pattern recognition, machine learning and data mining. As an effective method for attribute reduction, rough set theory offers a useful and formal methodology. It retains the discernibility power of the original datasets; thus, attribute reduction has been extensively studied in rough set theory. However, the inefficiency of the existing attribute reduction algorithms limits the application of rough sets. In this paper, we first analyse the limitations of existing attribute reduction algorithms. Then, a novel measure of attribute quality, called the relative discernibility degree, is proposed based on the discernibility. Theoretical analysis shows that this measure can find relative dispensable attributes and remain unchanged after removing the relative dispensable attributes and redundant objects in the process of selecting attributes. This property can be used to reduce the search space and accelerate the heuristic process of attribute reduction. Consequently, a new attribute reduction algorithm is proposed from the viewpoint of discernibility. Furthermore, the relationships among the reduction definitions of the algebra view, information view and discernibility view are derived. Some non-equivalent relationships among these views of rough set theory in inconsistent decision tables are discovered. A set of numerical experiments was conducted on UCI datasets. Experimental results show that the proposed algorithm is effective and efficient and is applicable to the case of large-scale datasets.

1. Introduction

In recent years, the size of datasets has grown dramatically in terms of both the number of instances and the number of features, such as medical data, text classification, etc. Extracting useful information from the rapidly expanding amount of data is an important and challenging issue. As a preprocessing step in data mining, attribute reduction is very effective in reducing the dimensionality of a feature space, increasing predictive accuracy and improving comprehensibility of the induced models. Therefore, it has been extensively investigated in past decades [1–3].

Rough set theory (RST) [4] is a valid mathematic tool to handle inexact and uncertain information. It provides a formal framework for data mining. One of the most focused on applications of RST is attribute reduction, which is considered an important branch of feature selection methods [1,5–10]. In the framework of rough sets, a reduct is defined as a minimal attribute subset that has the same discernibility as the entire attribute set. As an effective method of feature selection, attribute reduction can select the features that encode the most significant information in a dataset without transforming the data. Meanwhile, it minimizes the loss of information during the selection process. Specifically, it retains the semantics of the original datasets [11].

* Corresponding author. Tel.: +86 15808032997.
E-mail address: hmcherry@126.com (M. He).

http://dx.doi.org/10.1016/j.ins.2015.07.052
0020-0255/

Several methods have been proposed for finding the set of all reducts or a single reduct [12–18]. Skowron and Rauszer [19] proposed an attribute reduction algorithm to find all reducts using a discernibility matrix. This method provides a mathematical foundation to investigate attribute reduction [20,21]. To compress the discernibility function of a decision table, Chen et al. [22] proposed a sample pair selection algorithm (SPSA) to find all reducts or one reduct. The algorithm can find a proper reduct and is effective as a preprocessing step to find reducts. It is known that finding all the reducts from a decision table is computationally expensive. In contrast, a single reduct is sufficient in most applications [23]. Thus, many heuristic methods for finding one reduct have also been proposed [22,24–27]. Shen et al. [28,29] developed a quick reduct algorithm (QRA) to compute a reduct by keeping the dependency function invariant. It is an efficient algorithm and widely used in real applications. However, the dependency function cannot well reflect the attribute importance [1,30,31], which leads to more randomness in the process of selecting attributes. Therefore, in most cases, QRA may not yield a reduct but instead a super-reduct, which contains a reduct as a subset of it. Based on this observation, an improved quick reduct algorithm (IQRA) [30] was proposed by introducing variable precision rough sets (VPRS) in the process of attribute selection. However, the search process based on VPRS in IQRA increases the time of iteration, and the problem of QRA still exists in IQRA. Both QRA and IQRA are reduct methods from the viewpoint of algebra; some researchers have also investigated attribute reduction from the viewpoint of information theory [24,27,32]. The relationship between the definitions of attribute reduction in the algebra view and information view was presented in [33]. The information view-based method is suitable for small-scale data sets, but it is very time-consuming when tested on high-dimensional data sets.

Finding reducts in large datasets is very challenging in RST [34,35]. The above attribute reduction methods are commonly computationally expensive, and they are not acceptable in the case of large-scale datasets. To improve the efficiency of reduct algorithms, attribute reduction algorithms under the algebra and information viewpoints in RST have been enhanced by filtering out redundant objects [36]. However, the enhanced algorithms in [36] only reduce the computation time to a certain extent. Because they choose the same attribute reduct as the original version, the aforementioned problem still exists. Moreover, the enhanced algorithms only aim to reduce redundant objects in the attribute selection process. They do not reduce the redundancy attributes. It has been observed that the number of attributes in datasets can also significantly affect the efficiency of attribute reduction. Motivated by these observations, this paper further improves the performance of heuristic attribute reduction methods by gradually reducing both the size of the universe and the number of attributes in each iteration of attribute reduction. The computational complexity of finding reducts can therefore be reduced.

This paper further improves the efficiency of heuristic reduct algorithms. This paper has three major contributions, which are summarized as follows. First, the limitations of existing attribute reduction algorithms are analysed. We observed that the attribute reduction algorithms from both algebra and information viewpoints select attributes in a random manner, which leads to high computational complexity when tested on large-scale data sets. Second, a novel measure for attribute quality is proposed. Using this measure, we can either remove the relative dispensable attributes or filter out the redundant objects in the process of reduct computation. Therefore, the search space is significantly reduced. Subsequently, a novel attribute reduction algorithm is proposed from the viewpoint of discernibility, which overcomes the limitations of existing reduction algorithms. Moreover, the relationships among the concepts of reducts from the viewpoints of algebra, information and discernibility are discussed. Third, the performance of our algorithm is compared with that of the discernibility matrix, algebra and information viewpoints on the UCI datasets. Numerical experiments show that the proposed algorithm obtained the smallest number of selected attributes in the shortest time in most cases. It achieved higher classification accuracy and can be applied to large-scale data sets with large numbers of attributes or objects.

The rest of this paper is organized as follows. Basic concepts from the viewpoints of algebra and information in RST are briefly reviewed in Section 2. The limitations of the existing attribute reduction algorithms are discussed in Section 3. The relative discernibility degree and its main properties are discussed in Section 4, together with a novel heuristic attribute reduction algorithm from the viewpoint of discernibility. In Section 5, we further study the relationships among the attribute reductions from the algebra viewpoint, information viewpoint and discernibility viewpoint of RST. Experimental analysis is presented in Section 6. Section 7 concludes this paper.

2. Preliminaries

2.1. Preliminary concepts of RST

An information system can be represented by $S = (U, A)$, where $U = \{x_1, x_2, \ldots, x_{|U|}\}$ is a non-empty finite set of objects ($|\cdot|$ denotes the cardinality of the set) and $A = \{a_1, a_2, \ldots, a_{|A|}\}$ is a non-empty finite set of attributes such that $a_j{:}a_j \to V_{a_j}$ for each $a_j \in A$. The set V_{a_j} is called the value set of a_j.

Each subset of attributes $P \subseteq A$ determines a binary indiscernibility relation IND(P), as follows:

$$\mathrm{IND}(P) = \{(x, y) \in U \times U | \forall a \in P, f(x, a) = f(y, a)\}.$$

Obviously, IND(P) is an equivalence relation. If $(x, y) \in \mathrm{IND}(P)$, then x and y are indiscernible by P. The partition generated by IND(P) is denoted by $U/\mathrm{IND}(P)$, which is further abbreviated as U/P. $U/P = \{P_1, P_2, \ldots, P_m\}$ denotes knowledge associated with the equivalence relation IND(P), where P_i is the equivalence class, $1 \le i \le m$, $1 \le m \le |U|$. Each equivalence class is called an information granule. The attribute set P is therefore, called knowledge P. The equivalence class determined by x with respect to (wrt) attribute set P is denoted by $[x]_P = \{y \in U | (x, y) \in \mathrm{IND}(P)\}$. If $x \in P_i$, then $[x]_P = P_i$.

Let $\mathrm{DIS}(P) \subseteq U \times U$ denote a discernibility relation [37], defined for a non-empty set of attributes $\emptyset \neq P \subseteq A$ as:

$$\mathrm{DIS}(P) = \{(x, y) \in U \times U | \exists a \in P, f(x, a) \neq f(y, a)\}.$$

If a pair of objects belongs to $\mathrm{DIS}(P)$, then the two objects differ in at least one attribute from the set P. Although the discernibility relation is symmetric, it is not reflexive and not transitive. The indiscernibility and discernibility relations are related by a complementary relationship [38], which means that any two objects are either indiscernible or discernible, i.e. $\mathrm{IND}(P) \cup \mathrm{DIS}(P) = U \times U$.

Let $P, Q \subseteq A$, $U/P = \{P_1, P_2, \ldots, P_m\}$ and $U/Q = \{Q_1, Q_2, \ldots, Q_n\}$, we define a partial relation "$\preceq$" in information systems as follows: $P \preceq Q \Leftrightarrow \forall P_i \in U/P$, there is a $Q_j \in U/Q$ such that $P_i \subseteq Q_j$, which means that knowledge P is finer than knowledge Q (or knowledge Q is coarser than knowledge P). If $P \preceq Q$ and $P \neq Q$, then we say that knowledge P is strictly finer than knowledge Q, which is denoted as $P \prec Q$. It is clear that $Q \subseteq P \Rightarrow P \preceq Q$.

An information system $S = (U, A)$ is called a decision table when $A = C \cup D$, where C is a finite set of condition attributes, D is a finite set of decision attributes, and $C \cap D = \emptyset$. If x and y are discernible wrt the decision attribute D, $x, y \in U$, and they are indiscernible wrt all the condition attributes, x and y are considered inconsistent. Otherwise, they are considered consistent. Similarly, if x and y are consistent for any $x, y \in U$ where $x \neq y$, the decision table $S = (U, C, D)$ is considered consistent. Otherwise, $S = (U, C, D)$ is considered inconsistent.

For a decision table $S = (U, C, D)$, the positive region of D wrt the condition attribute set $P \subseteq C$ is $POS_P(D) = \cup_{D_i \in U/D} \underline{P}(D_i)$, $\underline{P}D_i = \cup\{Y \in U/P | Y \subseteq D_i\}$ is the P-lower approximation of D_i. In VPRS, $POS_P^{\beta}(D) = \cup_{D_i \in U/D} \underline{P}^{\beta}(D_i)$ is called the β-lower approximation, where $\underline{P}^{\beta}(D_i) = \cup\{Y \in U/P | \frac{|Y \cap D_i|}{|Y|} \geq \beta\}$ and $\beta \in (0.5, 1]$.

For a decision table $S = (U, C, D)$, S is consistent $\Leftrightarrow POS_C(D) = U \Leftrightarrow C \preceq D$.

2.2. Attribute reduction from the algebra viewpoint

Definition 1. Given a decision table $S = (U, C, D)$ and $P \subseteq C$, a dependency function involving the attribute sets P and D is defined as $\gamma_P(D) = \frac{|POS_P(D)|}{|U|}$. Using the dependency function, the significance measure of an attribute $a_i \in \{C - P\}$ for the condition attribute set P is expressed as $SIG_{dep}^{outer}(a_i, P, D) = \gamma_{P \cup \{a_i\}}(D) - \gamma_P(D)$. Correspondingly, the significance measure of an attribute r_i in P is expressed as $SIG_{dep}^{inner}(r_i, P, D) = \gamma_P(D) - \gamma_{P - \{r_i\}}(D)$.

Definition 2. Given a decision table $S = (U, C, D)$, if $a_i \in C$ and $SIG_{dep}^{inner}(a_i, C, D) = 0$, then a_i is reducible in C with reference to D from the algebra viewpoint, and we say that a_i is a dispensable attribute.

Definition 3. For a decision table $S = (U, C, D)$, $R \subseteq C$ is a relative reduction of C wrt the decision attribute set D if and only if $\gamma_R(D) = \gamma_C(D)$ and $\forall r_i \in R$, $\gamma_{R - \{r_i\}}(D) \neq \gamma_C(D)$.

This is the so-called attribute reduction of RST from the algebra viewpoint [33]. Commonly, there are more than one reducts in an information system. The intersection of all reducts is called the core of the information system. The core is denoted by $Core_{dep} = \{a_i | \gamma_{C - \{a_i\}}(D) \neq \gamma_C(D), a_i \in C\}$ in the algebra viewpoint. $Core_{dep}$ can be an empty set.

2.3. Attribute reduction from the information viewpoint

Definition 4. Given a decision table $S = (U, C, D)$ and $P \in C$, the conditional entropy of the decision attribute set D wrt another attribute set P is formulated as

$$H(D/P) = -\sum_{i=1}^{m} \frac{|P_i|}{|U|} \sum_{j=1}^{n} \frac{|D_j \cap P_i|}{|P_i|} \log_2 \frac{|D_j \cap P_i|}{|P_i|},$$

where $U/P = \{P_1, P_2, \ldots, P_m\}$ and $U/D = \{D_1, D_2, \ldots, D_n\}$. Using the conditional entropy, the significance measure of an attribute $a_i \in \{C - P\}$ for a condition attribute set P is expressed as

$$SIG_{con}^{outer}(a_i, P, D) = H(D/P) - H(D/(P \cup \{a_i\})).$$

Correspondingly, the significance measure of an attribute r_i in P is expressed as $SIG_{con}^{inner}(r_i, P, D) = H(D/(P - \{r_i\})) - H(D/P)$.

Definition 5. Given a decision table $S = (U, C, D)$, if $a_i \in C$ and $SIG_{con}^{inner}(a_i, C, D)$=0, then a_i is reducible in C wrt D from the information viewpoint. In this case, a_i is considered a dispensable attribute.

Definition 6. Let $S = (U, C, D)$ be a decision table. $R \subseteq C$ is a relative reduction of C wrt the decision attribute set D if and only if $H(D/R) = H(D/C)$ and $\forall r_i \in R$, $H(D/(R - \{r_i\})) \neq H(D/C)$.

This is the so-called attribute reduction of RST from the information viewpoint [33]. The core of the information viewpoint is denoted by $Core_{con} = \{a_i | H(D|(C - \{a_i\})) \neq H(D|C), a_i \in C\}$. $Core_{con}$ can be an empty set.

Algorithm1. IQRA
Input: Decision table $S=(U,C,D)$;
Output: One reduct *red*.
Step 1: Calculate $\gamma_C^1(D)$ and $red \leftarrow \emptyset$; //*red* is the pool to conserve the selected attributes.
Step 2: $Att \leftarrow C-red$, $\beta \leftarrow 1$, $\varepsilon \leftarrow 0.1$;
Step 3: For each $a_i \in Att$, compute $SIG_{dep}^{outer}(a_i, red, D, \beta)$;
Step 4: If $\max\{SIG_{dep}^{outer}(a_i, red, D, \beta), a_i \in Att\} > 0$, $SIG_{dep}^{outer}(a_k, red, D, \beta) = \max\{SIG_{dep}^{outer}(a_i, red, D, \beta)\}$, *then* $Att \leftarrow Att - \{a_k\}$, $red \leftarrow red \cup \{a_k\}$, go to step 6; else go to step 5;
Step 5: $\beta \leftarrow \beta - \varepsilon$. If $\beta \geq 0.5$, go to step 4; else, $a_k \leftarrow \{$First attribute in Att$\}$, $red \leftarrow red \cup a_k$, $Att \leftarrow Att - \{a_k\}$.
Step 6: $POS \leftarrow POS_{red}^1(D)$, $U \leftarrow U - POS_{red}^1(D)$; // removing redundant objects in calculating the positive region.
Step 7: If $\gamma_C^1(D) = \gamma_{red}^1(D)$, return *red*; else, go to step 3.

Fig. 1. The IQRA algorithm.

Note that we have already given two types of significance measures for heuristic functions in RST. They are the inner significance measure SIG^{inner} (including SIG_{dep}^{inner} and SIG_{con}^{inner}) and the outer significance measure SIG^{outer} (including SIG_{dep}^{outer} and SIG_{con}^{outer}). The inner significance measure is used to determine the redundancy of condition attributes, while the outer significance measure is used in a forward attribute reduction.

3. Limitations of existing attribute reduction algorithms

Measuring the significance of an attribute is a key issue for any heuristic attribute reduction algorithm. In the algebra viewpoint, the representative significance measure is the dependency degree. Correspondingly, conditional entropy is a representative significance measure in the information view [32]. Both of them are used frequently in attribute reduction algorithms. Each of the attribute reduction methods from the algebra viewpoint and information viewpoint can extract a single reduct from a given decision table. However, both of them have some limitations. For example, if the values of the dependency degree for all condition attributes are zero due to the case that no equivalence class is consistent in the first iteration in a reduction algorithm, we will have an empty set if the algorithm stops there [1]. Otherwise, if it makes an arbitrary choice and continues to the next iteration, it usually results in a super-reduct [30]. This is the problem that the algebra viewpoint more often faces. Prasad et al. [30] used VPRS to overcome these limitations. However, their algorithm still has these limitations in real data sets. Similarly, although the attribute reduct keeps the probabilistic distribution of the original datasets in the information view, it also has the same limitations as in the algebra view. In the rest of this paper, we reveal the limitations of the existing reduction methods under the algebra viewpoint and information viewpoint by using an illustrative example. Here, we only select IQRA [30] and CEBARKCC [32] as the representatives of the algebra viewpoint and information viewpoint, respectively. Figs. 1 and 2 give the main steps of IQRA and CEBARKCC, respectively. In Fig. 1, $SIG_{dep}^{outer}(a_i, red, D, \beta) = \gamma_{red\cup\{a_i\}}^{\beta}(D) - \gamma_{red}^{\beta}(D)$ and $\gamma_P^{\beta}(D) = \frac{|POS_P^{\beta}(D)|}{|U|}$.

Example 1. Let $S=(U,C,D)$ be a decision table, where $U=\{u_1,u_2,\ldots,u_{12}\}$, $C=\{a_1,a_2,\ldots,a_9\}$. We here use IQRA and CEBARKCC to compute the reduct of Table 1.

According to the IQRA and CEBARKCC algorithms, we can calculate the importance of each attribute in each iteration, as shown in Tables 2 and 3. The notation $*$ represents the attribute at the corresponding position that has been selected as an element of the reduct. The underlined texts are the values of the significance measure of the selected attribute in this stage. In Table 2, $\beta = 0.5$ means that the importance of each attribute is obtained by $SIG_{dep}^{outer}(a_i, red, D, \beta)$ when β is set to 0.5. According to Tables 2 and 3, it is clear that each attribute produces zero in the second iteration of computing the reduct. The reducts of

Algorithm2. CEBARKCC
Input: Decision table $S=(U,C,D)$;
Output: One reduct *red.*
Step 1: Compute $H(D|C)$;
Step 2: Compute $Core_{con}$, and $Att \leftarrow C - Core_{con}$;
Step 3: $red \leftarrow Core_{con}$; //*red* is the pool to conserve the selected attributes.
Step 4: If $|Core_{con}|>0$, compute $H(D|red)$, go to step 7;
Step 5: For each $a_i \in Att$, compute $SIG^{outer}_{con}(a_i, red, D)$;
Step 6: Select the attribute a_k which satisfies $SIG^{outer}_{con}(a_k, red, D) = \max\{SIG^{outer}_{con}(a_i, red, D)\}$. If the attribute like that is not only one, we select the one with the least number of combinations of values with *red*.
$Att \leftarrow Att - \{a_k\}$, $red \leftarrow red \cup a_k$;
Step 7: If $H(D|red) = H(D|C)$, return *red*; else, go to step 5.

Fig. 2. The CEBARKCC algorithm.

Table 1
A consistent decision table.

$U \times A$	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	D
u_1	1	1	1	1	1	1	1	1	1	1
u_2	1	1	1	1	1	1	1	1	1	1
u_3	1	2	1	1	2	2	2	1	2	1
u_4	1	2	1	1	2	2	2	1	2	1
u_5	1	1	1	1	2	2	2	1	1	2
u_6	1	1	1	1	2	2	2	1	1	2
u_7	1	2	1	1	1	1	1	1	2	2
u_8	1	2	1	1	1	1	1	1	2	2
u_9	2	3	3	1	3	3	3	2	3	3
u_{10}	2	3	2	1	4	4	3	3	4	3
u_{11}	2	3	2	1	4	3	3	2	3	4
u_{12}	2	3	3	1	3	4	3	3	4	4

Table 2
Attribute importance in each round according to IQRA.

Round/C	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9
1($\beta=0.5$)	1	1	1	0	1	1	1	1	1
2($\beta=0.5$)	*	0	0	0	0	0	0	0	0
3	*	*	0	0	0.67	0.67	0.67	0	0
4	*	*	0	0	*	0.33	0	0.33	0.33

Table 1 in the algebra view and information view are $\{a_1, a_2, a_5, a_6\}$ and $\{a_1, a_4, a_2, a_7, a_3, a_6\}$, respectively. However, the complete set of reducts of Table 1 is composed of $\{a_5, a_9\}$, $\{a_2, a_3, a_6\}$, $\{a_2, a_5, a_6\}$, $\{a_2, a_5, a_8\}$,$\{a_3, a_6, a_9\}$, $\{a_3, a_7, a_9\}$ and $\{a_2, a_3, a_7, a_8\}$. Clearly, neither IQRA nor CEBARKCC can obtain the reduct of Table 1. In addition, it is clear in Table 1 that attribute a_4 is redundant. Although the importance of attribute a_4 in each iteration is zero in Tables 2 and 3, neither of the two algorithms can find this and thus filter it out in the process of selecting attributes. That is because, although the significance

Table 3
Attribute importance in each round according to CEBARKCC.

Round/*C*	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9
1	0.92	0.92	0.92	0	0.92	0.92	0.92	0.92	0.92
2	*	0	0	0	0	0	0	0	0
3	*	0	0	*	0	0	0	0	0
4	*	*	0	*	0.67	0.67	0.67	0	0
5	*	*	0	*	0	0	*	0	0
6	*	*	*	*	0	0.33	*	0.33	0.33

measure of one attribute is equal to 0 in the algebra and information views, it does not mean the attribute is reducible. In contrast, the importance of attribute a_4 must be calculated in each iteration, and a_4 is selected as the element of the reduct in CEBARKCC.

It can be observed from Example 1 that the significance measures of attributes in the algebra and information views are not suitable for evaluating the attribute importance. It is clear that it is time-consuming when there are many redundant attributes. Therefore, these two viewpoints are not acceptable in dealing with large-scale datasets. However, in real applications, a number of datasets consist of a very large number of attributes, many of which are redundant. Therefore, it is necessary to investigate effective and efficient heuristic attribute reduction algorithms.

4. Attribute reduction from the discernibility view

Knowledge is based on the ability to classify objects in RST. In a sense, the knowledge content of an information system is the total number of ordered pairs $(x, y) \in U \times U$. Several uncertainty measures [39] have already been proposed based on discernibility, such as the discernibility and indiscernibility degree [40]. Based on these studies, in this section, we first introduce the discernibility degree and relative discernibility degree. Next, the properties of these measures are further analysed. Then, we propose a novel significance measure of attributes. The search strategies are also improved. Finally, a new efficient heuristic attribute reduction algorithm is presented.

4.1. Discernibility degree and relative discernibility degree

Knowledge is thought of as the discernibility capability of the attributes in the framework of the rough set methodology. In general, the objects in an equivalence class cannot be distinguished from each other, while those in different equivalence classes can be distinguished in RST. Therefore, in a broad sense, the knowledge content of a given attribute set can be characterized by the total number of pairs of the objects that can be distinguished from each other in the universe [40,41]. Considering this, we introduce several important definitions, including the discernibility degree [39] and relative discernibility degree [40] from the point of view of discernibility.

Definition 7. Let $S = (U, A)$ be an information system, $U = \{x_1, x_2, \ldots, x_{|U|}\}$, $P \subseteq A$, and a binary discernibility relation $\mathrm{DIS}(P)$ be defined as

$$\mathrm{DIS}(P) = \left\{ \left(x_i, x_j\right) \in U \times U | \exists a \in P, f(x_i, a) \neq f\left(x_j, a\right) \right\}.$$

Accordingly, the discernibility degree of P is denoted by $|\mathrm{DIS}(P)|$, which is equal to the number of ordered pairs in $\mathrm{DIS}(P)$.

$|\mathrm{DIS}(P)|$ can be regarded as the amount of knowledge contained in P. Thus, $|\mathrm{DIS}(P)|$ is a quantitative expression of P's discernibility. The greater $|\mathrm{DIS}(P)|$ is, the stronger the discernibility of P is.

Theorem 1. *Let $S = (U, A)$ be an information system, and $P, Q \subseteq A$. Then*

(1) *$P \preceq Q$, if and only if* $\mathrm{DIS}(Q) \subseteq \mathrm{DIS}(P)$.
(2) *If $P \preceq Q$, then* $|\mathrm{DIS}(Q)| \leq |\mathrm{DIS}(P)|$, *with equality if and only if* $\mathrm{DIS}(P) = \mathrm{DIS}(Q)$.

Proof. (1) ($\Rightarrow$) From the definition of a partial relation, we have that $P \preceq Q \Leftrightarrow \forall x_i \in U, [x_i]_P \subseteq [x_i]_Q$.

Therefore, we need only to prove that $\forall x_i \in U, [x_i]_P \subseteq [x_i]_Q \Rightarrow \mathrm{DIS}(Q) \subseteq \mathrm{DIS}(P)$. Assume that $\mathrm{DIS}(Q) \not\subseteq \mathrm{DIS}(P)$; then, $\exists x_i, x_j \in U$, $i \neq j$, s.t. $(x_i, x_j) \in \mathrm{DIS}(Q)$, but $(x_i, x_j) \notin \mathrm{DIS}(P)$, i.e., $x_j \notin [x_i]_Q$, but $x_j \in [x_i]_P$. Thus, $\exists x_i \in U, [x_i]_P \not\subseteq [x_i]_Q$ holds, which is contradictory to $\forall x_i \in U, [x_i]_P \subseteq [x_i]_Q$. Hence, $P \preceq Q \Rightarrow \mathrm{DIS}(Q) \subseteq \mathrm{DIS}(P)$ holds.

($\Leftarrow$) Because $\mathrm{DIS}(Q) \subseteq \mathrm{DIS}(P)$, we have that $\forall x_i, x_j \in U, i \neq j, (x_i, x_j) \in \mathrm{DIS}(Q) \Rightarrow (x_i, x_j) \in \mathrm{DIS}(P)$, i.e. $\forall x_j \notin [x_i]_Q \Rightarrow x_j \notin [x_i]_P$. Hence, $\forall x_i \in U, [x_i]_P \subseteq [x_i]_Q$ holds, i.e. $P \preceq Q$. Thus, (1) holds.

(2) It follows directly from Definition 7 and (1) of this theorem. □

From Theorem 1, it is clear that the discernibility degree increases monotonically as the information granule becomes smaller through finer classification. That is, the stronger the discernibility capability of attributes is, the larger the discernibility degree of attributes is, and the larger the amount of knowledge contained by P is. Note that the reverse relation of (2) in Theorem 1 cannot be established in general.

Generally, if $x_i \in [x_k]_P$, $x_i, x_k \in U$, and $P \subseteq A$, then $(x_i, x_k) \in \mathrm{IND}(P)$, i.e. x_i and x_k are indistinguishable from each other wrt P. In contrast, if x_i comes from the complement of the information granule, then x_i and x_k are distinguishable from each other wrt P, i.e. if $x_i \in \{U - [x_k]_P\}$, then $(x_i, x_k) \in \mathrm{DIS}(P)$. Considering this, we present the expression of the discernibility degree as follows.

Theorem 2. *Let $S = (U, A)$ be an information system, $P \subseteq A$, and $U/P = \{P_1, P_2, \ldots, P_m\}$. The discernibility degree $|\mathrm{DIS}(P)|$ is*

$$\left|\mathrm{DIS}(P)\right| = |U|^2 - \sum_{t=1}^{m} |P_t|^2. \tag{1}$$

Proof. It follows the definition of the discernibility relation that $(x_i, x_j) \in \mathrm{DIS}(P)$ for any $(x_i, x_j) \in P_t \times P_k$, $t \neq k$ and that $(x_r, x_q) \notin \mathrm{DIS}(P)$ for any $(x_r, x_q) \in P_t \times P_t$. Therefore, there are $|P_t|^2$ ordered pairs that do not belong to $\mathrm{DIS}(P)$ for any $P_t \in U/P$. In total, there are $|U|^2$ ordered pairs in the universe U, therefore $|\mathrm{DIS}(P)| = |U|^2 - \sum_{t=1}^{m} |P_t|^2$. □

Theorem 2 gives the discernibility measure for one attribute set, while, in many cases, we need to evaluate the relevance between two different attribute sets. Considering this, we extend the definition of discernibility degree to more than one attribute set and introduce the relative discernibility degree.

Definition 8. Let $S = (U, A)$ be an information system, $P, Q \subseteq A$, and the relative discernibility relation of Q wrt P be defined as $\mathrm{DIS}(Q/P) = \mathrm{DIS}(Q) - \mathrm{DIS}(Q) \cap \mathrm{DIS}(P)$. Accordingly, the relative discernibility degree of Q wrt P is denoted as $|\mathrm{DIS}(Q/P)|$.

From Definition 8, it is clear that $\mathrm{DIS}(Q/P)$ represents the ordered pairs whose elements can be distinguished by Q but cannot be distinguished by P. Accordingly, $|\mathrm{DIS}(Q/P)|$ is equal to the number of ordered pairs in $\mathrm{DIS}(Q/P)$. The relative discernibility degree represents the measure of the difference between knowledge P and knowledge Q on U. The greater $|\mathrm{DIS}(Q/P)|$ is, the larger the difference between knowledge P and knowledge Q is.

Theorem 3. *Given a decision table $S = (U, C, D)$ and $P, Q \subseteq C$, if $P \preceq Q$, then $|\mathrm{DIS}(D/Q)| \geq |\mathrm{DIS}(D/P)|$, with equality if and only if $\mathrm{DIS}(D) \cap \mathrm{DIS}(P) = \mathrm{DIS}(D) \cap \mathrm{DIS}(Q)$.*

Proof. Because$P \preceq Q$, we get from (1) of Theorem 1 that $\mathrm{DIS}(Q) \subseteq \mathrm{DIS}(P)$. Hence

$$\mathrm{DIS}(Q) \cap \mathrm{DIS}(D) \subseteq \mathrm{DIS}(P) \cap \mathrm{DIS}(D), \text{ i.e. } \left|\mathrm{DIS}(Q) \cap \mathrm{DIS}(D)\right| \leq \left|\mathrm{DIS}(P) \cap \mathrm{DIS}(D)\right|.$$

It follows from Definition 8 that

$$\left|\mathrm{DIS}(D/P)\right| = \left|\mathrm{DIS}(D) - \mathrm{DIS}(D) \cap \mathrm{DIS}(P)\right| = \left|\mathrm{DIS}(D)\right| - \left|\mathrm{DIS}(D) \cap \mathrm{DIS}(P)\right|, \text{ and thus}$$

$$\left|\mathrm{DIS}(D/Q)\right| - \left|\mathrm{DIS}(D/P)\right| = \left|\mathrm{DIS}(D) \cap \mathrm{DIS}(P)\right| - \left|\mathrm{DIS}(D) \cap \mathrm{DIS}(Q)\right| \geq 0.$$

Therefore, $|\mathrm{DIS}(D/Q)| \geq |\mathrm{DIS}(D/P)|$, with equality if and only if $\mathrm{DIS}(D) \cap \mathrm{DIS}(P) = \mathrm{DIS}(D) \cap \mathrm{DIS}(Q)$.

Note that the reverse relation of Theorem 3 is not true in general.

According to Definition 8, if $(x, y) \in \mathrm{DIS}(D/P)$ and $x, y \in U$, then x and y can be distinguished by decision attribute D but cannot be distinguished by condition attribute set P. Therefore, x and y are inconsistent, i.e. (x, y) is an inconsistent ordered pair. Intuitively, the more inconsistent the ordered pairs are, the higher the degree of the inconsistency of the decision table is. Theorem 3 shows that the relative discernibility degree measures the degree of inconsistency of a decision table. The finer the classification of P is, the smaller $|\mathrm{DIS}(D/P)|$ is, and the lower the degree of inconsistency of the decision table is. □

Corollary 1. *Let $S = (U, A)$ be an information system and $P, Q \subseteq A$. Then,*

(1) $0 \leq |\mathrm{DIS}(Q/P)|$, *with equality if and only if $P \preceq Q$;*
(2) $|\mathrm{DIS}(Q/P)| \leq |\mathrm{DIS}(Q)|$.

Proof. It directly follows from Definition 8 and (1) of Theorem 1. □

Theorem 4. *Let $S = (U, A)$ be an information system, $P, Q \subseteq A$, $U/(P \cup Q) = \{M_1^{pq}, M_2^{pq}, \ldots, M_n^{pq}\}$, and $U/P = \{P_1, P_2, \ldots, P_m\}$; then, the relative discernibility degree $|\mathrm{DIS}(Q/P)|$ is:*

$$\left|\mathrm{DIS}(Q/P)\right| = \sum_{i=1}^{m} |P_i|^2 - \sum_{k=1}^{n} \left|M_k^{pq}\right|^2 \tag{2}$$

Proof. It directly follows from Definition 8 and Theorem 1. □

4.2. Attribute reduction from the discernibility view

To design a heuristic attribute reduction algorithm, the significance measure of attributes and the search strategy are crucial to the overall process of attribute reduction. In RST, two significance measures of attributes (i.e. the outer significance measure and inner significance measure) are widely used [36]. Accordingly, there are two search strategies for attribute reduction. One is forward search (FS), and the other is backward search (BS). One attribute with great outer significance measure at a time is added to the current attribute subset in FS until the outer significance measure does not increase. Although FS adds the most important

attribute to the reduct, it cannot reduce the redundant attribute in the process of computing the reduct. It is computationally time-consuming. Therefore, FS is not suitable for applications on large-scale datasets. In contrast, BS starts with all attributes and progressively removes the redundant attributes whose inner significance measures are zero until the inner significance measure decreases. Compared with FS, BS can reduce the redundant attributes; it, however, cannot guarantee that the most important attribute is added to the reduct. Therefore, the attribute reduct obtained by BS is often much longer, and it is also computationally very expensive when there are many redundant attributes. In this section, we propose a new significance measure based on the relative discernibility degree and then provide the definitions of a relative dispensable attribute and attribute reduct from the discernibility view. Finally, we propose a new efficient heuristic attribute reduction algorithm that shares the merits of both FS and BS.

Corollary 2. *Let $S = (U, C, D)$ be a decision table and $Q \subseteq P \subseteq C$. Then, $|\mathrm{DIS}(D/P)| \leq |\mathrm{DIS}(D/Q)|$, and the condition for equality is $\mathrm{DIS}(D) \cap \mathrm{DIS}(P) = \mathrm{DIS}(D) \cap \mathrm{DIS}(Q)$.*

Proof. It directly follows from Theorem 3. □

Corollary 2 shows that the relative discernibility degree decreases monotonically with increasing condition attributes. That is, the relative discernibility degree does not increase when adding a new condition attribute into the attribute subset. This property is very important for constructing a forward attribute reduction algorithm. As previously mentioned in Section 4.1, the relative discernibility degree can be used to measure the inconsistency of a decision table. According to Corollary 2, it can be observed that the number of inconsistent ordered pairs of the original subset will remain constant or decrease when adding any new attribute into the existing subset. The greater the decrease in the number of inconsistent ordered pairs, the more important the new attribute is. Therefore, the relative discernibility degree can be used to measure the importance of attributes in rough sets. According to Corollary 2, the significance measure of attributes can be defined as follows.

Definition 9. Let $S = (U, C, D)$ be a decision table and $Q \subseteq C$. $\forall c_i \in (C - Q)$, and the significance measure of attribute c_i relative to Q is defined as:

$$SIG^{outer}_{dis}(c_i, Q, D) = \left|\mathrm{DIS}(D/Q)\right| - \left|\mathrm{DIS}(D/(Q \cup \{c_i\}))\right| \tag{3}$$

Correspondingly, the significance measure of an attribute r_i in Q is:

$$SIG^{inner}_{dis}(r_i, Q, D) = \left|\mathrm{DIS}(D/(Q - \{r_i\}))\right| - \left|\mathrm{DIS}(D/Q)\right|. \tag{4}$$

From Definition 9, it can be seen that $SIG^{outer}_{dis}(c_i, Q, D)$ describes the decrease of inconsistent ordered pairs when adding the attribute c_i to Q. $SIG^{inner}_{dis}(r_i, Q, D)$ denotes the increase of inconsistent ordered pairs by deleting r_i from Q. The definition of attribute reduction from the discernibility viewpoint is given as follows.

Definition 10. Given a decision table $S = (U, C, D)$, an attribute set $R \subseteq C$ is a relative reduction of C wrt the decision attribute set D if it satisfies the following two conditions:

(1) $|\mathrm{DIS}(D/R)| = |\mathrm{DIS}(D/C)|$;
(2) $\forall r_i \in R$, $|\mathrm{DIS}(D/(R - \{r_i\}))| > |\mathrm{DIS}(D/C)|$.

This is the attribute reduction of RST from the discernibility viewpoint. The first condition keeps the number of inconsistent ordered pairs in the original data unchanged, which indicates the viewpoint of discernibility. The second one shows that the attribute subset R is minimal. From Definition 10, it can be seen that the purpose of attribute reduction in the discernibility view is to find a subset of attributes that has the same number of inconsistent ordered pairs as the original data without redundancy.

Definition 11. Given a decision table $S = (U, C, D)$, if $r_i \in C$ and $SIG^{inner}_{dis}(r_i, C, D) = 0$, then r_i is reducible in C with reference to D from the discernibility viewpoint. Here, r_i is referred to as a dispensable attribute.

To improve the time efficiency of a heuristic attribute reduction, several methods [30,36] have been proposed to sequentially reduce the universe. However, these methods are only aimed at the redundancy associated with the objects. They do not consider the redundancy associated with the attributes as the attribute reduction algorithm proceeds. Two strategies for accelerating the reduction procedure are introduced as follows.

For convenience, the significance measure of attribute c_i relative to P is denoted as $SIG^{outer}_{dis}(c_i, P, D, U)$. It represents the value of the significance measure on the universe U. Similarly, the relative discernibility degree on the universe U is denoted as $|\mathrm{DIS}(D/P, U)|$.

Theorem 5. *Let $S = (U, C, D)$ be a decision table, $P \subset C$, $M^P_1 \in U/P$, and $a_t \in \{C - P\}$. If there exists $D_j \in U/D$ or $C_k \in U/C$ satisfying $M^P_1 \subseteq D_j$ or $M^P_1 = C_k$, then*

$$SIG^{outer}_{dis}(a_t, P, D, U) = SIG^{outer}_{dis}(a_t, P, D, U'), \text{ where } U' = U - M^P_1.$$

Proof. Let $U/P = \{M^P_1, M^P_2, \ldots, M^P_H\}$, $U/(P \cup D) = \{M^{pd}_1, M^{pd}_2, \ldots, M^{pd}_L\}$, $U/(P \cup \{a_t\}) = \{M^{pa_t}_1, M^{pa_t}_2, \ldots, M^{pa_t}_W\}$, and $U/(P \cup \{a_t\} \cup D) = \{M^{pa_t d}_1, M^{pa_t d}_2, \ldots, M^{pa_t d}_E\}$. Without loss of generality, assume that $M^p_1/(P \cup D) = \{M^{pd}_1, M^{pd}_2, \ldots, M^{pd}_l\}$,

where $l \le L$; $M_1^p/(P \cup \{a_t\}) = \{M_1^{pa_t}, M_2^{pa_t}, \ldots, M_w^{pa_t}\}$, where $w \le W$; and $M_1^p/(P \cup \{a_t\} \cup D) = \{M_1^{pa_t d}, M_2^{pa_t d}, \ldots, M_e^{pa_t d}\}$, where $e \le E$. According to Theorem 4 and Definition 9, it can be derived that:

$$\begin{aligned} &SIG_{dis}^{outer}(a_t, P, D, U) - SIG_{dis}^{outer}(a_t, P, D, U') \\ &= \left|\mathrm{DIS}(D/P, U)\right| - \left|\mathrm{DIS}(D/(P\cup\{a_t\}), U)\right| - \left|\mathrm{DIS}(D/P, U')\right| + \left|\mathrm{DIS}(D/(P\cup\{a_t\}), U')\right| \\ &= \sum_{j=1}^{H}\left|M_j^p\right|^2 - \sum_{j=1}^{L}\left|M_j^{pd}\right|^2 - \sum_{j=1}^{W}\left|M_j^{pa_t}\right|^2 + \sum_{j=1}^{E}\left|M_j^{pa_t d}\right|^2 - \sum_{j=2}^{H}\left|M_j^p\right|^2 + \sum_{j=l+1}^{L}\left|M_j^{pd}\right|^2 + \sum_{j=w+1}^{W}\left|M_j^{pa_t}\right|^2 - \sum_{j=e+1}^{E}\left|M_j^{pa_t d}\right|^2 \\ &= \left|M_1^p\right|^2 - \sum_{j=1}^{l}\left|M_j^{pd}\right|^2 - \sum_{j=1}^{w}\left|M_j^{pa_t}\right|^2 + \sum_{j=1}^{e}\left|M_j^{pa_t d}\right|^2 \end{aligned}$$

(1) If there exists $D_j \in U/D$ such that $M_1^P \subseteq D_j$, then $M_1^P/D = M_1^P$, $M_1^P/(\{a_t\} \cup D) = M_1^P/\{a_t\}$, that is, $\sum_{j=1}^{l}|M_j^{pd}|^2 = |M_1^p|^2$, $\sum_{j=1}^{e}|M_j^{pa_t d}|^2 = \sum_{j=1}^{w}|M_j^{pa_t}|^2$. Therefore, $SIG_{dis}^{outer}(a_t, P, D, U) - SIG_{dis}^{outer}(a_t, P, D, U') = 0$.

(2) If there exists $C_k \in U/C$ such that $M_1^P = C_k$, then $M_1^P/\{a_t\} = M_1^P$, $M_1^P/(\{a_t\} \cup D) = M_1^P/D$. That is, $\sum_{j=1}^{w}|M_j^{pa_t}|^2 = |M_1^p|^2$, $\sum_{j=1}^{l}|M_j^{pd}|^2 = \sum_{j=1}^{e}|M_j^{pa_t d}|^2$. Therefore, $SIG_{dis}^{outer}(a_t, P, D, U) - SIG_{dis}^{outer}(a_t, P, D, U') = 0$.

Consequently, the theorem is proved completely. □

From Theorem 5, one can see that the significance measure of an attribute in the process of attribute reduction remains unchanged when reducing the objects that do not need to be distinguished any longer. Therefore, the objects in the universe will become less numerous as the attribute reduction proceeds, while the same selected attribute subset will be retained. This mechanism can be used to improve the computational performance of a heuristic attribute reduction algorithm in the discernibility view.

Zhang et al. [42] introduced the characterizations of three important types of attribute sets in a consistent decision table $S = (U, C, D)$, as follows:

1. the core attribute set *Core* of *S*: $Core = \cap_{i \le L} R_i$;
2. the relative indispensable attribute set *RI* of *S*: $RI = \cup_{i \le L} R_i - Core$;
3. the dispensable attribute set *DP* of *S*: $DP = C - (Core \cup RI)$,

where $(R_i : i \le L)$ is all the reducts of *S*. The core attribute set is commonly used in forward attribute reduction algorithms. However, the dispensable attribute set and relative indispensable attribute set are seldom employed. If the dispensable attribute set can be found and reduced in the forward attribute reduction process, the efficiency of computing the reduct will be improved. The definition of a relative dispensable attribute is given as follows. Further, a new method to determine the redundancy of a condition attribute in the process of computing the reduct is also proposed.

Definition 12. Let $S = (U, C, D)$ be a decision table and R_i be a reduct of *S*. $RDP = C - R_i$ is the dispensable attribute set relative to R_i. *RDP* represents the relative dispensable attribute set; meanwhile, $c_t \in RDP$ represents the relative dispensable attribute.

It is clear that the dispensable attribute set *DP* is included in the relative dispensable attribute set *RDP*. From Definition 12, it can be seen that, if the elements of *RDP* can be gradually reduced when computing the reduct R_i, the computational time of the reduction R_i can be significantly reduced. In the following part, the judgment theorem of a relative dispensable attribute is presented.

Theorem 6. *Let $S = (U, C, D)$ be a decision table. R_i is a reduct of S, $P \subseteq R_i$, and $c_t \in \{C - P\}$. If $SIG_{dis}^{outer}(c_t, P, D) = 0$, then c_t is a relative dispensable attribute,* i.e. $c_t \notin R_i$.

Proof. Suppose $c_t \in R_i$, and let $Q = \{R_i - P - \{c_t\}\} \subseteq \{C - P\}$. According to Definition 10, we have $|\mathrm{DIS}(D/(R_i - \{c_t\}))| > |\mathrm{DIS}(D/R_i)|$. That is, $|\mathrm{DIS}(D/(P \cup Q))| > |\mathrm{DIS}(D/(P \cup \{c_t\} \cup Q))|$. According to Definition 8 and Theorem 3, it can be derived that $\mathrm{DIS}(D) \cap \mathrm{DIS}(P \cup Q) \subset \mathrm{DIS}(D) \cap \mathrm{DIS}(P \cup \{c_t\} \cup Q)$. According to Definition 7, $\mathrm{DIS}(P \cup Q) = \mathrm{DIS}(P) \cup \mathrm{DIS}(Q)$. Therefore,

$$\mathrm{DIS}(D) \cap \mathrm{DIS}(P \cup Q) = [\mathrm{DIS}(D) \cap \mathrm{DIS}(P)] \cup [\mathrm{DIS}(D) \cap \mathrm{DIS}(Q)] \subset [\mathrm{DIS}(D) \cap \mathrm{DIS}(P \cup \{c_t\})] \cup [\mathrm{DIS}(D) \cap \mathrm{DIS}(Q)],$$

i.e. $[\mathrm{DIS}(D) \cap \mathrm{DIS}(P)] \subset [\mathrm{DIS}(D) \cap \mathrm{DIS}(P \cup \{c_t\})]$.

From Definition 8, it can be derived that $|\mathrm{DIS}(D/P)| > |\mathrm{DIS}(D/(P \cup \{c_t\}))|$. In terms of formula (3), we have that $SIG_{dis}^{outer}(c_t, P, D) > 0$, which conflicts with the claim that $SIG_{dis}^{outer}(c_t, P, D) = 0$. Hence, $c_t \notin R_i$, and the theorem is proved completely. □

Theorem 6 shows that introducing a relative dispensable attribute does not lead to a decrease in inconsistent ordered pairs. Therefore, the relative dispensable attribute can be reduced in the process of computing the reduct. Note that all the existing methods to judge the redundancy of an attribute set are based on the inner significance measure, such as Definitions 2 and 5. However, the significance measures are outer significance measures in forward attribute reduction algorithms, such as QRA and FSPA-SCE. They cannot be used to estimate the redundancy of attributes, i.e. $SIG_{dep}^{outer}(q_i, Q, D) = 0$ or $SIG_{con}^{outer}(q_i, Q, D) = 0$ does

not mean q_i is reducible (see Example 1). Although Theorem 6 is also based on the outer significance measure, it can be used to judge the relative dispensable attributes that contain the dispensable attributes. It is clear that formula (3) is applicable to FS, while, at the same time, we can eliminate the relative dispensable attributes from the current attribute subset one-by-one in the process of computing the reduct. Therefore, Theorem 6 provides an approach to reduce the relative dispensable attribute in FS, which further accelerates the reduction procedure.

Note that Theorems 5 and 6 provide the key theoretical foundation for the efficient attribute reduction algorithm proposed in this paper. Thus, we give the efficient forward greedy search algorithm for computing the attribute reduct as follows:

Algorithm 3. Efficient forward attribute reduction algorithm from the discernibility view (FAR-DV)

Input: Decision table $S = (U, C, D)$.
Output: One reduct *red*.

Step 1: Compute $|\mathrm{DIS}(D/C)|$;
Step 2: $j \leftarrow 1, U^j \leftarrow U, U' \leftarrow \emptyset, A^j \leftarrow C, A' \leftarrow \emptyset, red \leftarrow \emptyset$; //*red* is the pool to conserve the selected attributes.
Step 3: $\forall a_t \in A^j$, compute $SIG^{outer}_{dis}(a_t, red, D, U^j)$; if $SIG^{outer}_{dis}(a_t, red, D, U^j) = 0$, then $A' \leftarrow A' \cup \{a_t\}$;
Step 4: Select the attribute a_k that satisfies $SIG^{outer}_{dis}(a_k, red, D, U^j) = \max\{SIG^{outer}_{dis}(a_t, red, D, U^j), a_t \in A^j\}$; if the attribute like that is not only one, we select the one whose discernibility degree value is the smallest. $A' \leftarrow A' \cup \{a_k\}$, $red \leftarrow red \cup \{a_k\}$;
Step 5: For each $M_r \in U^j/red$, if there exists $D_t \in U^j/D$ or $C_k \in U^j/C$ such that $M_r \subseteq D_t$ or $M_r = C_k$, then $U' \leftarrow U' \cup M_r$;
Step 6: $j \leftarrow j + 1, U^j \leftarrow U - U', A^j \leftarrow C - A'$; // removing redundant objects and the relative dispensable attribute in the process of calculating the reduct.
Step 7: If $|\mathrm{DIS}(D/red)| = |\mathrm{DIS}(D/C)|$, return *red*; else, go to Step 3.

In FAR-DV, we begin with an empty set *red* of attributes, and add one attribute that makes the significance measure maximal to the set *red* in each iteration until a reduct is found. This is the forward reduction algorithm. The steps that reduce the redundant attributes and objects are embedded in the algorithm. These steps include Steps 3, 5 and 6, which reduce the number of objects and attributes in the decision table during the attribute reduction procedure. Therefore, FAR-DV not only considers the redundancy associated with the objects but also the redundancy associated with the attributes in the process of calculating the reduct. In practice, most of the samples and attributes are reduced at the beginning of the reduction in most cases. Therefore, the computational expense is greatly reduced.

In this algorithm, we use the radix sorting approach [43] to compute the partition of an attribute set. Therefore, the computational complexity of Step 1 is $O(|C||U|)$ according to Formula (2). The time complexity for computing the significance measure of an attribute is $O(|C||U|)$ according to Formula (3). The overall time complexity of Step 3 through Step 6 is $O(\sum_{j=1}^{|red|} |A^j||U^j|)$, where *red* is a reduct obtained by FAR-DV. In summary, the computational complexity for the FAR-DV algorithm is $O(|C||U|) + O(\sum_{j=1}^{|red|} |A^j||U^j|)$. Therefore, FAR-DV is a relatively efficient algorithm for calculating the reduct of decision tables in comparison to the methods in the information view and algebra view (with time complexity of $O(|C||U|) + O(\sum_{j=1}^{|red|} |U_j|(||C| - j + 1|))$, where $|U^j| \le |U_j|$, $|A^j| \le ||C| - j + 1|$). This is because, only FAR-DV considers the sequentially reduced universe and attribute set. Note that we here suppose that the same number of selected attributes is obtained by the three viewpoints. However, more compact reducts are obtained by the discernibility view compared to the other two views. This further accelerates the speed of the FAR-DV algorithm.

5. The relationship between the attribute reductions from the viewpoints of algebra, information and discernibility

Attribute reduction has already been studied from the algebra viewpoint and information viewpoint. Wang et al. [33] noted that the relationship between the definitions of attribute reduction from the two viewpoints is an inclusion rather than an equivalence. Similarly, in this section, a comparative study on the quantitative relationships between the definitions of attribute reduction in the algebra view, information view and discernibility view is presented.

5.1. Equivalent relationship in consistent decision tables

For convenience, the following lemmas are introduced [44].

Lemma 1. *Let $S = (U, C, D)$ be a consistent decision table; then, $\gamma_C(D) = 1$ and $H(D|C) = 0$.*

Lemma 2. *Let $S = (U, C, D)$ be a consistent decision table. For any $c_i \in C$, $\gamma_{C-\{c_i\}}(D) = \gamma_C(D)$ if and only if $H(D|(C - \{c_i\})) = H(D|C)$.*

Lemma 3. *Let $S = (U, C, D)$ be a consistent decision table, and $R \subseteq C$. For any $r_i \in R$, $\gamma_{R-\{r_i\}}(D) \neq \gamma_C(D)$ and $\gamma_R(D) = \gamma_C(D)$ if and only if $H(D|(R - \{r_i\})) \neq H(D|C)$ and $H(D|R) = H(D|C)$.*

Based on Lemmas 2 and 3, Wang et al. [44] disclosed the equivalence in consistent decision tables between the concepts of attribute reduction defined from the algebra viewpoint and the information viewpoint. In the following, we discuss the relationship between the definition in the discernibility viewpoint and the definitions in the algebra view and information view.

Theorem 7. *Let $S = (U, C, D)$ be a consistent decision table. $\forall c_i \in C$, $|\mathrm{DIS}(D/(C - \{c_i\}))| = |\mathrm{DIS}(D/C)|$ if and only if $\gamma_{C-\{c_i\}}(D) = \gamma_C(D)$ or $H(D|(C - \{c_i\})) = H(D|C)$.*

Proof. According to Lemma 2, we only need to prove that

$$\left|\mathrm{DIS}(D/(C - \{c_i\}))\right| = \left|\mathrm{DIS}(D/C)\right| \Leftrightarrow \gamma_{C-\{c_i\}}(D) = \gamma_C(D).$$

($\Rightarrow$) Because $S = (U, C, D)$ is consistent, $C \preceq D$ and $POS_C(D) = U$. From (1) of Corollary 1, we have that $|\mathrm{DIS}(D/C)| = 0$ and thus that $\forall c_i \in C$, and we have that $|\mathrm{DIS}(D/(C - \{c_i\}))| = 0$, i.e. $\{C - \{c_i\}\} \preceq D$, $POS_{C-\{c_i\}}(D) = U$. Therefore, $\gamma_{C-\{c_i\}}(D) = \gamma_C(D) = 1$, i.e. $|\mathrm{DIS}(D/(C - \{c_i\}))| = |\mathrm{DIS}(D/C)| \Rightarrow \gamma_{C-\{c_i\}}(D) = \gamma_C(D)$.

($\Leftarrow$) Similarly, the reverse is true. □

Theorem 8. *Let $S = (U, C, D)$ be a consistent decision table. $\forall c_i \in C$, $|\mathrm{DIS}(D/(C - \{c_i\}))| \neq |\mathrm{DIS}(D/C)|$ if and only if $\gamma_{C-\{c_i\}}(D) \neq \gamma_C(D)$ or $H(D|(C - \{c_i\})) \neq H(D|C)$.*

Proof. It directly follows from Theorem 7. □

It can be found from Theorems 7 and 8 that whether or not an attribute is reducible in a consistent decision table is equivalent in the discernibility view, algebra view and information view of RST.

Theorem 9. *Let $S = (U, C, D)$ be a consistent decision table. $\forall r_i \in R \subseteq C$, $|\mathrm{DIS}(D/(R - \{r_i\}))| > |\mathrm{DIS}(D/C)|$ and $|\mathrm{DIS}(D/R)| = |\mathrm{DIS}(D/C)|$ if and only if it satisfies any of the following two conditions:*

(1) $\forall r_i \in R$, $\gamma_{R-\{r_i\}}(D) \neq \gamma_C(D)$ *and* $\gamma_R(D) = \gamma_C(D)$.
(2) $\forall r_i \in R$, $H(D|(R - \{r_i\})) \neq H(D|C)$ *and* $H(D|R) = H(D|C)$.

Proof. $S = (U, C, D)$ is a consistent decision table; from (1) of Corollary 1 and Lemma 1, we have that $|\mathrm{DIS}(D/C)| = H(D|C) = 0$ and $\gamma_C(D) = 1$. According to Lemma 3, we only need to prove (1) of this theorem.

($\Rightarrow$) In terms of Theorem 7, we can get that $|\mathrm{DIS}(D/R)| = |\mathrm{DIS}(D/C)| \Rightarrow \gamma_R(D) = \gamma_C(D)$. Because $|\mathrm{DIS}(D/(R - \{r_i\}))| > |\mathrm{DIS}(D/C)|$, we have that $|\mathrm{DIS}(D/(R - \{r_i\}))| > 0$. It follows from (1) of Corollary 1 that $\{R - \{r_i\}\} \preceq D$ does not hold. Therefore, $S = (U, \{R - \{r_i\}\}, D)$ is an inconsistent decision table, i.e. the decision table will become inconsistent after deleting any attribute from R. Therefore, $POS_{R-\{r_i\}}(D) \subset U$, i.e. $\gamma_{R-\{r_i\}}(D) < \gamma_C(D) = 1$. Consequently, $|\mathrm{DIS}(D/(R - \{r_i\}))| > |\mathrm{DIS}(D/C)|$ and $|\mathrm{DIS}(D/R)| = |\mathrm{DIS}(D/C)| \Rightarrow \gamma_{R-\{r_i\}}(D) \neq \gamma_C(D)$ and $\gamma_R(D) = \gamma_C(D)$.

($\Leftarrow$) In a similar fashion, the reverse is true. □

From Theorem 9, it can be concluded that the definitions of attribute reduction based on these three different viewpoints are equivalent in consistent decision tables. Note that different reduction results may be obtained due to the use of different importances of attributes.

5.2. Inequivalent relationship in inconsistent decision tables

The significance measure of an attribute is a key problem in designing a heuristic attribute reduction algorithm. As we can see from Definitions 1, 4 and 9, there are three different definitions in the algebra view, information view and discernibility view. The significance measure in the algebra view only focuses on the consistent part of a decision table and completely ignores the inconsistent objects. In contrast, the significance measures in the information view and discernibility view take the decision table as a whole. The significance measure in the discernibility view is based on a stricter definition than that of the information view. Therefore, after taking consideration of a new condition attribute, the significance measure of the new attribute in the algebra and information views would be zero. However, it is not always zero in the discernibility view. This conclusion is demonstrated by the following example.

Example 2. Table 4 gives an inconsistent decision table $S = (U, C, D)$, where $U = \{u_1, u_2, \ldots, u_8\}$ and $C = \{a_1, a_2, \ldots, a_4\}$. Let $Q = \{a_1, a_2, a_3\}$. According to Table 4, we can calculate the significance measure of attribute a_4 relative to Q as follows:

1. In the algebra view, $SIG^{outer}_{dep}(a_4, Q, D) = \gamma_{Q\cup\{a_4\}}(D) - \gamma_Q(D) = 0.25 - 0.25 = 0$;

Table 4
An inconsistent decision table.

$U \times A$	a_1	a_2	a_3	a_4	D
u_1	0	1	0	0	1
u_2	0	1	0	1	1
u_3	0	1	0	0	0
u_4	0	1	0	1	0
u_5	0	0	0	1	2
u_6	1	1	0	1	2
u_7	0	1	1	1	2
u_8	0	1	1	1	3

2. In the information view, $SIG^{outer}_{con}(a_4, Q, D) = H(D|Q) - H(D|(Q \cup \{a_4\})) = 0.75 - 0.75 = 0$;
3. In the discernibility view, $SIG^{outer}_{dis}(a_4, Q, D) = |\mathrm{DIS}(D/Q)| - |\mathrm{DIS}(D/(Q \cup \{a_4\}))| = 10 - 6 = 4$.

If the significance measure of an attribute is zero in the discernibility view, then it would also be zero in the algebra view and information view. We have the following theorem for this conclusion.

For convenience, the following lemma is first introduced.

Lemma 4. [44]. *Let $S = (U, C, D)$ be a decision table, and $P, Q \subseteq C$. $U/Q = \{Q_1, Q_2, \ldots, Q_m\}$, and $U/P = \{Q_1, Q_2, \ldots, Q_{i-1}, Q_{i+1}, \ldots, Q_{j-1}, Q_{j+1}, \ldots, Q_m, Q_i \cup Q_j\}$ is another partition generated by merging the two equivalence classes Q_i and Q_j,* i.e. *$Q_i \cup Q_j$. Then,*

(1) *$H(D|P) = H(D|Q)$ if and only if $\forall D_k \in U/D$, $\frac{|Q_i \cap D_k|}{|Q_i|} = \frac{|Q_j \cap D_k|}{|Q_j|}$.*
(2) *If $H(D|P) = H(D|Q)$, then $POS_P(D) = POS_Q(D)$.*

Theorem 10. *Let $S = (U, C, D)$ be a decision table, $Q \subseteq C$, and $q_i \in \{C - Q\}$. If $|\mathrm{DIS}(D/(Q \cup \{q_i\}))| = |\mathrm{DIS}(D/Q)|$, then $H(D|(Q \cup \{q_i\})) = H(D|Q)$ and $\gamma_{Q \cup \{q_i\}}(D) = \gamma_Q(D)$.*

Proof. If U/Q=$U/(Q \cup \{q_i\})$, it is clear that this theorem holds. □

Suppose $U/Q \neq U/(Q \cup \{q_i\})$. From $Q \subset Q \cup \{q_i\}$ we have that there exists $X_m, X_n \in U/(Q \cup \{q_i\})$ such that $X_m \cup X_n \in U/Q$. According to Definition 1, we have that $\forall x_i \in X_m$and $\forall x_j \in X_n$, there is $(x_i, x_j) \in \mathrm{DIS}(Q \cup \{q_i\})$, but $(x_i, x_j) \notin \mathrm{DIS}(Q)$. Because $|\mathrm{DIS}(D/(Q \cup \{q_i\}))| = |\mathrm{DIS}(D/Q)|$, it follows from Corollary 2 that $\mathrm{DIS}(D) \cap \mathrm{DIS}(Q \cup \{q_i\}) = \mathrm{DIS}(D) \cap \mathrm{DIS}(Q)$, which means that $\forall (x_i, x_j) \in \mathrm{DIS}(Q \cup \{q_i\})$ and $(x_i, x_j) \in \mathrm{DIS}(D)$, we have that $(x_i, x_j) \in \mathrm{DIS}(Q)$.

(1) Suppose $H(D|(Q \cup \{q_i\})) \neq H(D|Q)$. According to (1) of Lemma 4, we obtain that there is $X_m, X_n \in U/(Q \cup \{q_i\})$ and $X_m \cup X_n \in U/Q$, which satisfies $\frac{|X_m \cap D_k|}{|X_m|} \neq \frac{|X_n \cap D_k|}{|X_n|}$, where $D_k \in U/D$. Therefore, there is $x_i, x_l \in X_m \cup X_n$ such that $x_i \notin D_k$ and $x_l \in D_k$. Without loss of generality, assume that $x_i \in X_m$ and $x_i \in D_t \in U/D$, where $t \neq k$. Then, we have the following two cases:

① if $\exists x_j \in X_n$ and $x_j \in D_k$, then $(x_i, x_j) \in \mathrm{DIS}(D)$ and $(x_i, x_j) \in \mathrm{DIS}(Q \cup \{q_i\})$, but $(x_i, x_j) \notin \mathrm{DIS}(Q)$, which contradicts the claim that $\mathrm{DIS}(D) \cap \mathrm{DIS}(Q \cup \{q_i\}) = \mathrm{DIS}(D) \cap \mathrm{DIS}(Q)$.
② else, $\forall x_j \in X_n$ and $x_j \notin D_k$, then $x_l \in X_m$, $(x_l, x_j) \in \mathrm{DIS}(D)$ and $(x_l, x_j) \in \mathrm{DIS}(Q \cup \{q_i\})$, but $(x_l, x_j) \notin \mathrm{DIS}(Q)$. This also conflicts with the claim that $\mathrm{DIS}(D) \cap \mathrm{DIS}(Q \cup \{q_i\}) = \mathrm{DIS}(D) \cap \mathrm{DIS}(Q)$.
Consequently, $H(D|(Q \cup \{q_i\})) = H(D|Q)$ holds.

(2) It follows directly from (2) of Lemma 4 and Definition 1 that $\gamma_{Q \cup \{q_i\}}(D) = \gamma_Q(D)$.

However, the reverse relation of this theorem cannot be derived generally. That is, $\forall q_i \in \{C - Q\}$, $|\mathrm{DIS}(D/(Q \cup \{q_i\}))| \neq |\mathrm{DIS}(D/Q)|$ does not always lead to $H(D|(Q \cup \{q_i\})) \neq H(D|Q)$ or $\gamma_{Q \cup \{q_i\}}(D) \neq \gamma_Q(D)$ (see Example 2). $|\mathrm{DIS}(D/(Q \cup \{q_i\}))| \neq |\mathrm{DIS}(D/Q)|$ is due to the change of the number of inconsistent ordered pairs, and this change does not always affect the positive region and conditional entropy. In contrast, $|\mathrm{DIS}(D/(Q \cup \{q_i\}))| = |\mathrm{DIS}(D/Q)|$ may not always hold when $H(D|(Q \cup \{q_i\})) = H(D|Q)$ or $\gamma_{Q \cup \{q_i\}}(D) = \gamma_Q(D)$. That is, although the positive region and the probabilistic distribution of the original data set are not changed, the number of inconsistent ordered pairs may be changed. Therefore, the relative discernibility degree may be changed.

Theorem 11. *Let $S = (U, C, D)$ be a decision table, and $Q \subseteq C$. If $\forall q_i \in \{C - Q\}$ and $SIG^{outer}_{dis}(q_i, Q, D) = 0$, then*

(1) *$SIG^{outer}_{con}(q_i, Q, D) = 0$ and $SIG^{outer}_{dep}(q_i, Q, D) = 0$.*
(2) *$SIG^{inner}_{con}(q_i, C, D) = 0$ and $SIG^{inner}_{dep}(q_i, C, D) = 0$.*

Proof. (1) It directly follows from formula (3) and Theorem 10.

(2) Let $T = C - \{q_i\} - Q$. Because $SIG^{outer}_{dis}(q_i, Q, D) = 0$, according to formula (3), it can be derived that $|\mathrm{DIS}(D/(Q \cup \{q_i\}))| = |\mathrm{DIS}(D/Q)|$. It follows from zero that $\mathrm{DIS}(D) \cap \mathrm{DIS}(Q \cup \{q_i\}) = \mathrm{DIS}(D) \cap \mathrm{DIS}(Q)$, i.e. $\mathrm{DIS}(D) \cap [\mathrm{DIS}(Q \cup \{q_i\}) \cup \mathrm{DIS}(T)] = \mathrm{DIS}(D) \cap [\mathrm{DIS}(Q) \cup \mathrm{DIS}(T)]$. According to Definition 7, it can be obtained that $\mathrm{DIS}(Q \cup T) = \mathrm{DIS}(Q) \cup \mathrm{DIS}(T)$ and $\mathrm{DIS}(D) \cap \mathrm{DIS}(Q \cup \{q_i\} \cup T) = \mathrm{DIS}(D) \cap \mathrm{DIS}(Q \cup T)$. Therefore, from Corollary 2, we have that $|\mathrm{DIS}(D/(Q \cup \{q_i\} \cup T))| = |\mathrm{DIS}(D/(Q \cup T))|$. According to Theorem 10, $H(D/(Q \cup \{q_i\} \cup T)) = H(D/(Q \cup T))$ and $\gamma_{Q \cup \{q_i\} \cup T}(D) = \gamma_{Q \cup T}(D)$. That is, $H(D/C) = H(D/(C - \{q_i\}))$ and $\gamma_C(D) = \gamma_{(C - \{q_i\})}(D)$. From Definitions 1 and 4, $SIG^{inner}_{con}(q_i, C, D) = 0$ and $SIG^{inner}_{dep}(q_i, C, D) = 0$. □

Note that the inverse of this theorem does not hold. From Theorem 11, it can be seen that, if q_i is reducible relative to Q in the discernibility view (i.e. $SIG^{outer}_{dis}(q_i, Q, D) = 0$), it must be reducible in the algebra view and information view (i.e. $SIG^{inner}_{dep}(q_i, C, D) = 0$ and $SIG^{inner}_{con}(q_i, C, D) = 0$). However, if condition attribute q_i is reducible in the algebra view or information view (i.e. $SIG^{inner}_{dep}(q_i, C, D) = 0$ or $SIG^{inner}_{con}(q_i, C, D) = 0$), it is not always reducible relative to Q in the discernibility view. That is because $SIG^{outer}_{dis}(q_i, Q, D)$ is not always equal to zero. Therefore, different condition attributes of an inconsistent decision table will be reduced in the algebra view, information view and discernibility view.

From the above discussion of this section, it can be observed that the definition of a significance measure of an attribute in the discernibility view is not equivalent to its definitions in the algebra and information views.

Table 5
Attribute importance in each round according to FAR-DV.

Round/*C*	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9
1	64	80	68	0	84	84	80	68	84
2	0	16	0	×	*	4	0	4	20

Theorem 12. *Let $S = (U, C, D)$ be an inconsistent decision table, and $Q \subseteq C$. If $\forall q_i \in Q$, $|\mathrm{DIS}(D/(Q - \{q_i\}))| > |\mathrm{DIS}(D/C)|$ and $|\mathrm{DIS}(D/Q)| = |\mathrm{DIS}(D/C)|$, then $H(D|Q) = H(D|C)$ and $\gamma_Q(D) = \gamma_C(D)$.*

Note that $\forall q_i \in Q$ and $|\mathrm{DIS}(D/(Q - \{q_i\}))| > |\mathrm{DIS}(D/C)|$ do not always lead to $\gamma_{Q-\{q_i\}}(D) \neq \gamma_C(D)$ or $H(D|(Q - \{q_i\})) \neq H(D|C)$. This means that, for an inconsistent decision table, a reduct obtained in the discernibility view may not be a reduct in the algebra or information view. There may be some redundant condition attributes. Therefore, different reduct results will be obtained in the three different views. For example, the reducts of Table 4 are $\{a_1, a_2\}$ and $\{a_1, a_2, a_3\}$ in the algebra and information views, respectively. In contrast, it is irreducible in the discernibility view. That is, its reduction is $\{a_1, a_2, a_3, a_4\}$ in the discernibility view.

Remark. : As we know, the algebra view keeps the positive region of the original data set unchanged, and the information view keeps the probabilistic distribution of the original data set unchanged. Meanwhile, the discernibility view possesses the same number of inconsistent ordered pairs as the original data set. In this section, a comparative study on the quantitative relationships between these three different views is presented. The results show that the three views are not always equivalent. In fact, this is due to the inconsistency of a decision table, i.e. the three views address the inconsistency in different ways. The discernibility view considers more details, therefore it puts more restrictions on reducible attributes. Consequently, the three different views are inequivalent to each other in inconsistent decision tables. That is, the reduct from the algebra view is included in that from the discernibility and information views, and the reduct from the information view is included in that from the discernibility view; while they are equivalent in consistent decision tables. Note that, although the attribute reduct from the discernibility viewpoint gives the strictest definition, it does not imply that the reduct has the largest number of selected attributes in most cases. On the contrary, the relative discernibility degree considers more details, the most important attribute can therefore be added to the reduct, and thus the heuristic process of attribute reduction is accelerated. Consequently, the reduct of FAR-DV is more compact (see Section 6 below for more details).

6. Experimental analysis

The objective of the following experiments is to show the effectiveness and efficiency of the proposed attribute reduction algorithm. We first use FAR-DV to calculate the reduction of Table 1 to show the advantages of FAR-DV. Then, we compare the performance of SPSA, QRA, IQRA, FSPA-SCE and FAR-DV, where SPSA [22] is an accelerating algorithm of the discernibility matrix method and can find a suboptimal reduct, QRA and IQRA are reduct methods from the viewpoint of algebra, FSPA-SCE is the accelerative algorithm from the viewpoint of information theory proposed in [33], and FAR-DV is from the viewpoint of discernibility.

Example 3. (Continued from Example 1). Similar to Example 1, the importance of each attribute in each iteration based on the FAR-DV algorithm is presented in Table 5, where the notation "×" represents the attribute at the corresponding position that has been deleted. According to Table 5, the reduction of Table 1 is $\{a_5, a_9\}$. If a_6 is selected as the element of reduction in the first round, then the reduction $\{a_6, a_2, a_3\}$ is obtained. Comparing Example 3 with Example 1, it can be seen that different reduction algorithms obtain different reduction results. Specifically, FAR-DV not only finds the relative dispensable attribute a_4 and thus filters it out in the process of computing the reduction but also obtains a more compact reduct. It is clear that the computation time of FAR-DV will be greatly reduced by gradually reducing the dispensable attributes when dealing with larger data sets, which we will discuss in detail below.

In the following, the reduct results, the computational efficiency and the classified performance of SPSA, QRA, IQRA, FSPA-SCE and FAR-DV are presented and compared. The algorithms are terminated if the running time exceeds 3.6×10^5 s. FSPA-SCE starts from the core attribute set, while other algorithms do not include the process of computing core attributes. The UCI Machine Learning Repository is used in this experiment. More details of the dataset are listed in Table 6, where the names with ♣ indicate that some attribute values are missing in the data set. In Table 6, three data sets are associated with categorical attributes, such as Connect_4, Chess and Vote, and the rest of the data sets are associated with numerical attributes. Because these five algorithms only work on complete and categorical data, we use ROSETTA software to complete the incomplete data sets via the Mean/mode method and to discretize the numerical features with equal frequency binning in preprocessing. The experiments are conducted in MATLAB 7.4 using a PC with Windows XP, a Pentium(R) E6500 dual-core CPU running at 2.93 GHz, and 2.0 GB of RAM.

In the following experiments, we first increased the number of objects and compared the variation of computations time and reduct results of different reduction algorithms. We divided the data set connect_4 into 15 parts according to the number of objects. The first 1000 objects were viewed as the first data set, the first 5000 objects were viewed as the second data set, …, and all objects were viewed as the fifteenth data set (see the *x*-axis of Fig. 3). Figs. 3 and 4 present the variation of computational

Table 6
Data description.

Data set	Abbreviation	Samples	Numerical attributes	Categorical attributes	Class
Connect_4	Connect_4	67557	0	42	3
Arcene_train_valid	Arcene	200	10000	0	2
MAGIC gamma telescope data 2004	MAGIC	19020	10	0	2
Internet_ads	Internet_ads	3279	1558	0	2
Chess	Chess	3196	0	36	2
Ozone Level Detection ♣	Ozone	2536	72	0	2
Madelon_Train	Madelon	2000	500	0	2
Secom ♣	Secom	1567	590	0	2
Congressional Votes ♣	Vote	435	0	16	2
Ionosphere	Iono	351	34	0	2
Sonar	Sonar	208	60	0	2
Wine recognition	Wine	178	13	0	3

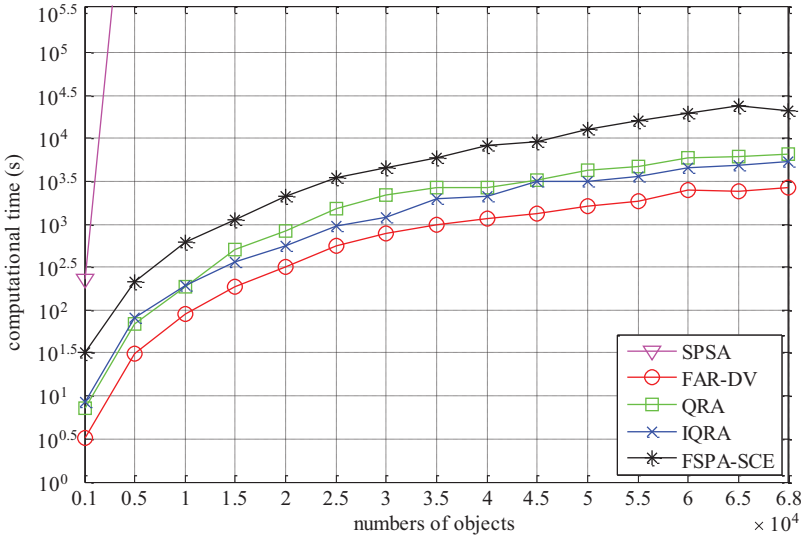

Fig. 3. Computational time with respect to the numbers of objects (Connect_4).

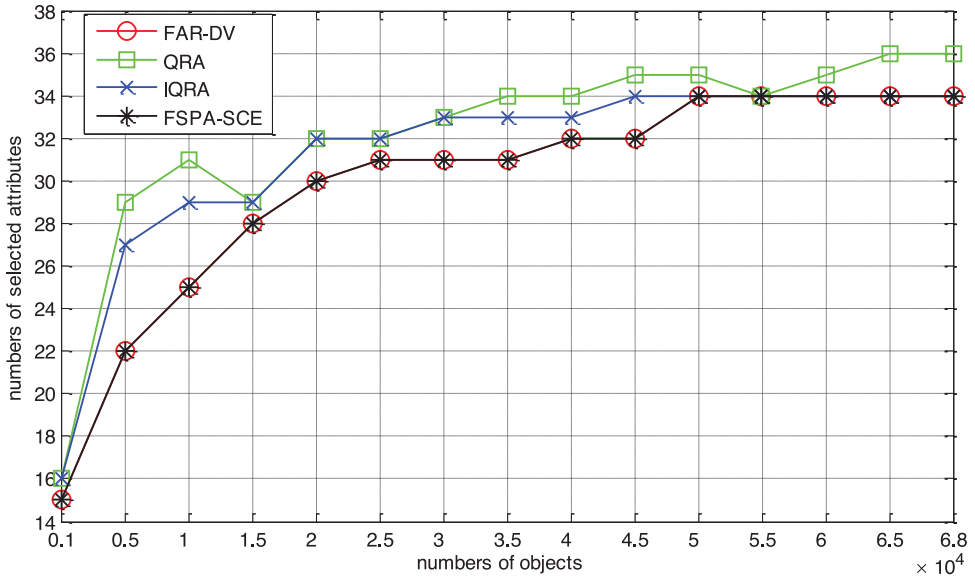

Fig. 4. Numbers of selected attributes with respect to the numbers of objects (Connect_4).

Table 7
The average numbers of selected attributes and computational times.

Data set	SPSA		QRA		IQRA		FSPA-SCE		FAR-DV	
	ave-T	*ave-R*	*ave-T*	*ave-R*	*ave-T*	*ave-R*	*ave-T*	*ave-R*	*ave-T*	*ave-R*
Connect_4	–	–	2761 ± 2208	29.7 ± 9.00	2139 ± 1792	31.2 ± 4.59	8649 ± 7990	27.6 ± 8.75	1102 ± 905	27.6 ± 8.75
Arcene	558 ± 839	17.4 ± 7.5	204 ± 223	25.9 ± 8.61	153 ± 156	26.2 ± 8.54	29200 ± 44817	10.2 ± 1.83	62 ± 67	9.3 ± 1.65

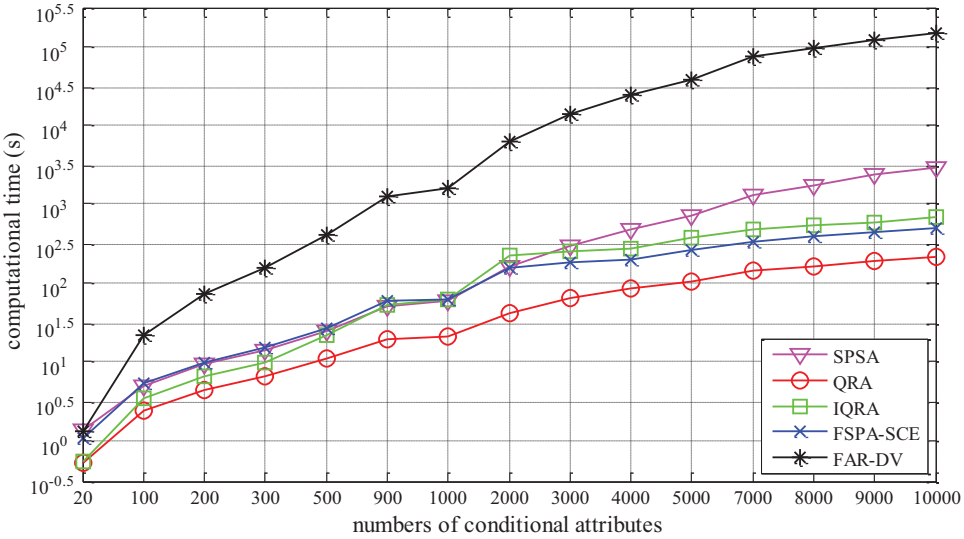

Fig. 5. Computational time with respect to the number of condition attributes (Arcene).

times and selected attributes with respect to the increasing number of objects. In Figs. 3 and 4, the *x*-axis denotes the number of objects, and the *y*-axis represents the computational time and the number of selected attributes. Overall, the computational time and the numbers of selected attributes of the five algorithms increase with the number of objects. SPSA is not able to give results within 3.6×10^5 s when the number of objects is greater than 2000. Therefore, the reduction results of SPSA are not included in Fig. 4. This shows that SPSA achieves the worst performance in terms of computational efficiency, especially for cases with a large number of objects. FAR-DV and FSPA-SCE selected a smaller number of attributes compared to QRA and IQRA, especially when the number of redundant attributes in the data set is large (e.g. the second and third data sets shown in Fig. 4). FAR-DV achieves the fastest performance. In contrast, FSPA-SCE is slow due to its time-consuming part for core attribute calculation. Compared to QRA, the attributes selected by the IQRA algorithm are the same or more compact for each data set (see Fig. 4). However, IQRA is not consistently faster than QRA. In contrast, IQRA is occasionally slower, such as for the 1st and 2nd data sets (see Fig. 3). This may be due to the fact that, VPRS is used as the heuristic information for the selection of attributes in IQRA, which increases the runtime of each iteration. From the results presented in Figs. 3 and 4, it can be concluded that FAR-DV is able to find the reduct with the smallest number of selected attributes in the shortest amount of time for each data set.

In [33], the standard deviation is used to characterize the robustness of a heuristic attribute reduction algorithm. A small value of the standard deviation means the stability of the algorithm is high. Similarly, we compute the average $\mu = \sum_{i=1}^{15} x_i$ and standard deviation $\sigma = \sqrt{\frac{1}{15}\sum_{i=1}^{15}(x_i - \mu)^2}$ of the computation times and the numbers of selected attributes on the 15 data sets originated from Connect_4. The results are shown in the first row of Table 7, where x_i represents the computational time or the number of selected attributes on the ith data set. In Table 7, ave-*T* represents the average time, and ave-*R* represents the average number of selected attributes on the 15 data sets. It can be observed that the average number of selected attributes of FAR-DV and FSPA-SCE is at least 2 less than that of IQRA and QRA, while the average reduced time of FAR-DV is only half that of IQRA and QRA and is one-tenth that of FSPA-SCE. In addition, according to the standard deviation in the first row of Table 7, it can be seen that FAR-DV has a smaller variation in the computational times, while FSPA-SCE has a larger variation. The variation of the reduct results of FAR-DV and FSPA-SCE is smaller than that of QRA and IQRA. Therefore, FAR-DV exhibits far better robustness than the other four algorithms to the number of objects increasing.

Next, we increased the number of condition attributes and compared the variation of the computational times and selected attributes of the five algorithms. Similarly, we divided the data set Arcene into 15 parts according to the number of condition attributes. The first 20 condition attributes were viewed as the first data set, the first 100 condition attributes were viewed as the second data set, …, and all condition attributes were viewed as the fifteenth data set (see the *x*-axis of Fig. 5). Figs. 5 and 6 display a more detailed change trend of the computational time and the number of selected attributes as the condition attribute number

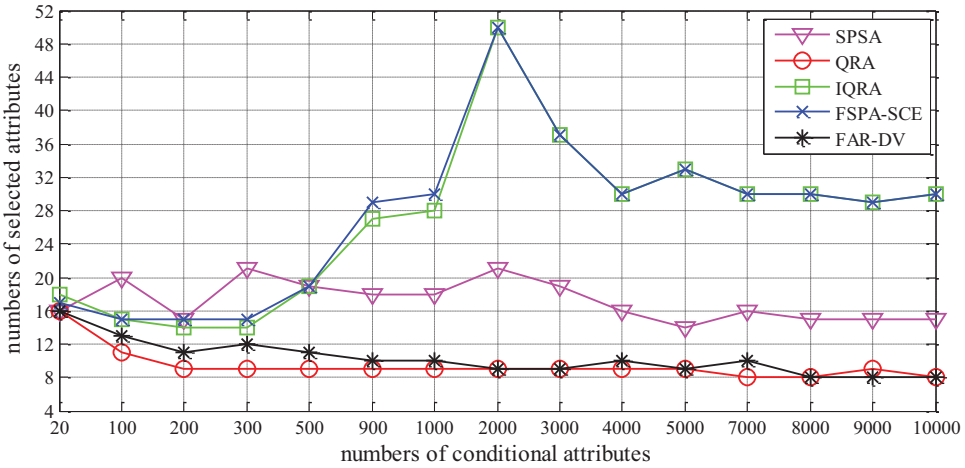

Fig. 6. Numbers of selected attributes with respect to the number of condition attributes (Arcene).

Table 8
Comparison of the numbers of selected attributes and computation times of the different algorithms.

Data	Raw	SPSA		QRA		IQRA			FSPA-SCE		FAR-DV	
		R	*T*(s)	*R*	*T*(s)	*R*	*T*(s)	Core	*R*	*T*(s)	*R*	*T*(s)
MAGIC	10	–	–	**10**	39.07	**10**	51.22	10	**10**	73.84	**10**	35.65
Internet_ads	1558	–	–	199	12889.34	199	4507.6	22	**106**	83583.92	**106**	1225.28
Chess	36	–	–	31	33.99	31	13.8	27	**29**	64.72	**29**	14.89
Ozone	72	–	–	15	93.26	14	62.80	1	**13**	599.88	**13**	34.65
Madelon	500	–	–	11	155.08	**10**	308.35	0	12	4411.82	**10**	113.56
Secom	590	8	15976	7	90.32	7	102.94	0	7	4551.72	**6**	62.44
Vote	16	**9**	27.22	10	1.90	10	2.32	7	**9**	9.62	10	1.07
Iono	34	**3**	19.94	5	0.84	5	0.61	0	**3**	5.06	**3**	0.34
Sonar	60	**2**	9.92	**2**	1.55	**2**	4.26	0	**2**	34.87	**2**	0.82
Wine	13	**5**	1.19	**5**	0.17	**5**	0.37	0	**5**	0.92	**5**	0.26
Average	288.9	–	–	29.5	1330.60	29.3	505.40	6.7	19.6	9333.6	**19.4**	148.90

increases. As a whole, the computational times of these five algorithms increased with the number of condition attributes. As seen in Fig. 5, the computational time of FSPA-SCE significantly increased with the number of condition attributes. However, the computational time of SPSA, IQRA, QRA and FAR-DV is robust to the number of condition attributes compared to FSPA-SCE. In a sense, the numbers of selected attributes should decrease with the increase of the number of condition attributes. FAR-DV and FSPA-SCE adhered to this assumption. However, IQRA and QRA did not (see Fig. 6). Further, FAR-DV obtained the most compact reducts on each data set except the 14th data set. FSPA-SCE achieved the second most compact result (see Fig. 6), and next is SPSA. Although VPRS was used as the heuristic information in IQRA to improve the limitations of QRA, it was outperformed by QRA. This means that IQRA cannot resolve the problem of QRA in certain cases. Both QRA and IQRA obtained a larger number of selected attributes compared to FAR-DV, FSPA-SCE and SPSA.

The average and standard deviation of computational time and number of selected attributes on the data set Arcene are shown in the second row of Table 7. FAR-DV achieved the lowest mean and standard deviation of computation times and the smallest number of selected attributes of the five algorithms. Thus, FAR-DV exhibits far better robustness than the other four algorithms to the increase of the number of condition attributes.

In addition, to further demonstrate the effectiveness of the proposed algorithm, we implemented and evaluated the five algorithms on the last ten UCI data sets of Table 6. The number of selected attributes and computation time with these five algorithms are shown in Table 8. The columns of raw, *R*, Core and *T* are the number of condition attributes in the raw data sets, the number of attributes in the reducts, the number of core attributes, and the computation time, respectively. The attribute subsets with minimal value among the five algorithms are shown in bold font. It can be observed from Table 8 that SPSA can obtain more compact reductions, but it is not applicable to data sets whose number of objects is larger than 2000, such as the first five data sets in Table 8. QRA is efficient for small data sets, such as Wine. However, it is very time-consuming for large data sets including many objects and attributes, such as Internet_ads. Although VPRS is integrated with IQRA, the reduction results of IQRA are not necessarily compact, and IQRA is still computationally expensive. For example, IQRA achieves the reduced time of 102.94 s on the data set Secom, while the number of selected attributes is the same as that of QRA. Moreover, the performance of FSPA-SCE is very dependent on the number of core attributes. The efficiency of FSPA-SCE is acceptable when the core attributes make up the great majority of selected attributes, and it can obtain the minimal reduct, such as for the data sets MAGIC and Chess. However,

Table 9
Comparison of the classification performance of different algorithms with the RBF-SVM classifier (Acc ± Std).

Data set	Raw data	SPSA	QRA	IQRA	FSPA-SCE	FAR-DV
MAGIC	69.87 ± 0.59	–	69.87 ± 0.59	69.87 ± 0.59	69.87 ± 0.59	69.87 ± 0.59
Internet_ads	86.98 ± 0.61	–	91.67 ± 1.90	91.67 ± 1.90	93.78 ± 2.25	93.53 ± 2.34
Chess	90.89 ± 6.18	–	92.83 ± 5.45	92.83 ± 5.45	92.83 ± 5.33	92.83 ± 5.33
Ozone	88.51 ± 9.05	–	94.39 ± 8.39	94.39 ± 8.39	94.39 ± 8.39	94.39 ± 8.39
Madelon	50.60 ± 1.61	–	50.75 ± 4.08	66.20 ± 3.47	67.30 ± 2.31	64.80 ± 4.30
Secom	93.37 ± 0.69	93.37 ± 0.69	93.50 ± 0.78	93.50 ± 0.78	93.50 ± 0.78	93.37 ± 0.69
Vote	86.89 ± 2.67	94.48 ± 2.92	93.54 ± 3.26	94.46 ± 3.49	94.48 ± 2.92	93.54 ± 3.26
Iono	64.70 ± 2.06	81.02 ± 7.97	75.05 ± 8.16	75.05 ± 8.16	81.04 ± 6.26	81.04 ± 6.26
Sonar	53.40 ± 2.46	55.33 ± 6.04	73.10 ± 4.99	74.07 ± 6.68	75.50 ± 10.87	73.10 ± 3.13
Wine	62.50 ± 9.37	93.89 ± 6.65	93.89 ± 6.65	93.89 ± 6.65	95.56 ± 4.38	95.56 ± 5.74
Average	74.77 ± 3.53	–	82.86 ± 4.43	84.59 ± 4.56	85.82 ± 4.41	85.20 ± 4.00

Table 10
Comparison of the classification performance of different algorithms with the CART classifier (Acc ± Std).

Data set	Raw data	SPSA	QRA	IQRA	FSPA-SCE	FAR-DV
MAGIC	80.66 ± 0.91	–	80.66 ± 0.91	80.66 ± 0.91	80.66 ± 0.91	80.66 ± 0.91
Internet_ads	95.61 ± 2.75	–	96.07 ± 2.52	96.07 ± 2.52	95.03 ± 2.47	95.09 ± 2.62
Chess	97.78 ± 2.62	–	97.27 ± 3.06	97.27 ± 3.06	97.30 ± 2.84	97.34 ± 2.86
Ozone	89.04 ± 5.86	–	90.26 ± 8.26	90.26 ± 8.26	93.39 ± 8.24	91.39 ± 8.86
Madelon	64.20 ± 3.33	–	51.15 ± 5.09	66.30 ± 3.74	71.15 ± 2.96	68.30 ± 3.03
Secom	79.07 ± 12.70	89.21 ± 4.78	89.02 ± 4.73	89.02 ± 4.73	90.43 ± 2.89	90.77 ± 2.38
Vote	95.85 ± 3.91	95.85 ± 3.91	95.85 ± 3.91	95.85 ± 3.91	95.85 ± 3.91	95.85 ± 3.91
Iono	84.53 ± 12.98	73.54 ± 15.87	86.12 ± 6.49	86.12 ± 6.49	86.23 ± 9.92	86.23 ± 9.92
Sonar	65.88 ± 10.38	72.62 ± 14.17	65.48 ± 10.00	78.29 ± 6.44	76.45 ± 6.48	75.90 ± 6.26
Wine	91.04 ± 8.34	93.82 ± 6.11	93.82 ± 6.11	93.82 ± 6.11	92.22 ± 6.52	93.26 ± 6.82
Average	84.37 ± 6.38	–	84.57 ± 5.11	87.37 ± 4.62	87.87 ± 4.71	87.48 ± 4.76

if the core attributes are in the minority of selected attributes, FSPA-SCE will be very expensive for high-dimensional data sets, such as Internet_ads, Secom and Madelon. Compared with the other four algorithms, FAR-DV reduced the number of selected attributes by a large margin. That is, FAR-DV achieved the minimal reduct in most cases. Meanwhile, the time consumption is significantly reduced, which benefitted from filtering out the redundant objects and the relative redundant attributes.

Finally, to evaluate the classification performance of different reduction algorithms, two learning mechanisms were used to create classifiers. The results of 10 runs with 10-fold cross-validation for the RBF-SVM and CART algorithms are shown in Tables 9 and 10, respectively. From Tables 8, 9 and 10, it is clear that all of the attribute reduction algorithms can reduce some candidate attributes while maintaining or improving the classification accuracy in most of the cases. FSPA-SCE achieved the best classification performance of the algorithms tested. The classification accuracy for FAR-DV is comparable to that of the other three algorithms, but it is inferior to that of FSPA-SCE. That is, although FSPA-SCE and FAR-DV produced a smaller set of attributes, they achieved a slightly better classification accuracy compared to the other algorithms with both classifiers, such as on the data sets Chess and Iono, which more directly reflect the information contained in the selected attributes. Because VPRS is used as the heuristic information for the selection of attributes in IQRA, the average classification accuracy of IQRA is 2% higher than that of QRA.

The experimental results demonstrate that the proposed algorithm obtained the smallest number of selected attributes with the shortest computational time in most cases, and it maintained the highest classification accuracy. Therefore, the proposed algorithm is effective and efficient and can be applied to large-scale data sets with a large number of attributes or objects.

7. Conclusion

Attribute reduction is one of the major advantages of rough set analysis. It has already been studied from the algebra viewpoint and information viewpoint of RST. However, the attribute reduction algorithms from both the algebra and information viewpoints select attributes in a random manner. They are computationally expensive when tested on large-scale data sets. To overcome these limitations, a new heuristic attribute reduction algorithm from the discernibility viewpoint is proposed in this paper. The proposed algorithm takes into account both the redundant objects and the relative redundant attributes in the process of selecting attributes, which is the key to further accelerating attribute reduction. Therefore, the computation time of this algorithm is largely reduced, and the reduct is more compact. Moreover, a comparative study on the quantitative relationships between the concepts of attribute reduction from the algebra viewpoint, information viewpoint, and discernibility viewpoint is presented. It is also concluded that the identity of the three viewpoints holds in consistent decision tables, while the relationship among the reductions from these three viewpoints is an inclusion in inconsistent decision tables. The experimental results show that the proposed method significantly reduces the computational time and that it is an effective and efficient heuristic attribute reduction algorithm, especially when dealing with larger data sets. It should be noted that data sets with numerical or fuzzy

attribute values are more common in the real world; therefore, we will extend the proposed algorithm to deal with numerical or fuzzy data in our next study.

Acknowledgements

The authors would like to thank Dr. Yongjian Nian and Dr. Yulan Guo owing to their assist in the revision of this paper. Moreover, the authors would like to thank the anonymous reviewers and the Associate Editor for their valuable comments on this paper. This work was supported by the National Natural Science Foundation of China (No. 41301397, No. 61471371), the Natural Science Foundation of Hunan Province of China (No. 2015jj3022) and China Postdoctoral Science Foundation (No. 2012M512168).

References

[1] Q. Hu, W. Pedrycz, D. Yu, J. Lang, Selecting discrete and continuous features based on neighborhood decision error minimization, IEEE Syst. Man Cybern. B Cybern. 40 (2010) 137–150.
[2] W. Pedrycz, Granular Computing: Analysis and Design of Intelligent Systems, CRC Press/Francis Taylor, Boca Raton, 2013.
[3] Z. Meng, Z. Shi, Extended rough set-based attribute reduction in inconsistent incomplete decision systems, Inf. Sci. 204 (2012) 44–69.
[4] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning About Data, Kluwer Academic Publisher, London, 1991.
[5] Q. Hu, J. Liu, D. Yu, Mixed feature selection based on granulation and approximation, Knowledge-Based Syst. 21 (2008) 294–304.
[6] Y. Qian, J. Liang, W. Pedrycz, C.Y. Dang, An efficient accelerator for attribute reduction from incomplete data in rough set framework, Pattern Recogn. 44 (2011) 1658–1670.
[7] R. Swiniarski, A. Skowron, Rough set methods in feature selection and recognition, Pattern Recogn. Lett. 24 (2003) 833–849.
[8] N.M. Parthalain, Q. Shen, Exploring the boundary region of tolerance rough sets for feature selection, Pattern Recogn. 42 (2009) 655–667.
[9] T. Deng, C. Yang, X. Wang, A reduct derived from feature selection, Pattern Recogn. Lett. 33 (2012) 1638–1646.
[10] Z. Pawlak, A. Skowron, Rudiments of rough sets, Inf. Sci. 177 (2007) 3–27.
[11] R. Jensen, Q. Shen, Semantics-preserving dimensionality reduction: rough and fuzzy-rough-based approaches, IEEE Trans. Knowl. Data Eng. 16 (2004) 1457–1471.
[12] J. Qian, D.Q. Miao, Z.H. Zhang, W. Li, Hybrid approaches to attribute reduction based on indiscernibility and discernibility relation, Int. J. Approx. Reason. 52 (2011) 212–230.
[13] J.Y. Liang, J.R. Mi, W. Wei, F. Wang, An accelerator for attribute reduction based on perspective of objects and attributes, Knowledge-Based Syst. 44 (2013) 90–100.
[14] D.G. Chen, L. Zhang, S.Y. Zhao, Q.H. Hu, P.F. Zhu, A novel algorithm for finding reducts with fuzzy rough sets, IEEE Trans. Fuzzy Syst. 20 (2012) 385–389.
[15] Y.Y. Yao, Y. Zhao, Discernibility matrix simplification for constructing attribute reducts, Inf. Sci. 179 (2009) 867–882.
[16] T. Yang, Q. Li, B. Zhou, Related family: a new method for attribute reduction of covering information systems, Inf. Sci. 228 (2013) 175–191.
[17] G. Wang, J. Hu, Attribute reduction using extension of covering approximation space, Fund. Inform. 115 (2012) 219–232.
[18] E.C.C. Tsang, D.G. Chen, D.S. Yeung, X.Z. Wang, J.W.T. Lee, Attributes reduction using fuzzy rough sets, IEEE Trans. Fuzzy Syst. 16 (2008) 1130–1141.
[19] A. Skowron, C. Rauszer, The discernibility matrices and functions in information systems, Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory, Kluwer, Dordrecht, Netherlands, 1992, pp. 331–362.
[20] D.Q. Miao, Y. Zhao, Y.Y. Yao, H.X. Li, F.F. Xu, Relative reducts in consistent and inconsistent decision tables of the Pawlak rough set model, Inf. Sci. 179 (2009) 4140–4150.
[21] G. Lang, Q. Li, L. Guo, Discernibility matrix simplification with new attribute dependency functions for incomplete information systems, Knowl. Inf. Syst. 37 (2013) 611–638.
[22] D.G. Chen, S.Y. Zhao, L. Zhang, Y.P. Yang, X. Zhang, Sample pair selection for attribute reduction with rough set, IEEE Trans. Knowl. Data Eng. 24 (2012) 2080–2093.
[23] X.T. Hu, T.Y. Lin, J.C. Han, A new rough sets model based on database systems, Fund. Inform. 59 (2004) 135–152.
[24] F. Jiang, Y. Sui, L. Zhou, A relative decision entropy-based feature selection approach, Pattern Recogn. 48 (2015) 2151–2163.
[25] M. Li, C. Shang, S. Feng, J. Fan, Quick attribute reduction in inconsistent decision tables, Inf. Sci. 254 (2014) 155–180.
[26] M. Li, C.X. Shang, S.Z. Feng, J.P. Fan, Quick attribute reduction in inconsistent decision tables, Inf. Sci. 254 (2014) 155–180.
[27] Q.H. Hu, Z.X. Xie, D.R. Yu, Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation, Pattern Recogn. 40 (2007) 3509–3521.
[28] A. Chouchoulas, Q. Shen, Rough set-aided keyword reduction for text categorization, Appl. Artif. Intell. 15 (2001) 843–873.
[29] R. Jensen, Q. Shen, Fuzzy-rough attribute reduction with application to web categorization, Fuzzy Sets Syst. 141 (2004) 469–485.
[30] P.S.V.S.S. Prasad, C.R. Rao, IQuickReduct: an improvement to quick reduct algorithm, in: Proceedings of the 12th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing, Delhi, India, 2009, pp. 152–159.
[31] D. Yamaguchi, Attribute dependency functions considering data efficiency, Int. J. Approx. Reason. 51 (2009) 89–98.
[32] G. Wang, H. Yu, D. Yang, Decision table reduction based on conditional information entropy, Chin. J. Comput. 25 (2002) 759–766.
[33] G.Y. Wang, J. Zhao, J.J. An, Y. Wu, A comparative study of algebra viewpoint and information viewpoint in attribute reduction, Fund. Inform. 68 (2005) 289–301.
[34] J. Komorowski, Z. Pawlak, L. Polkowski, L. Skowron, Rough sets: a tutorial, in: S.K. Pal, A. Skowron (Eds.), Rough Fuzzy Hybridization: A New Trend in Decision-making, Springer, Singapore, 1999, pp. 3–98.
[35] J. Qian, P. Lv, X. Yue, C. Liu, Z. Jing, Hierarchical attribute reduction algorithms for big data using MapReduce, Knowledge-Based Syst. 73 (2015) 18–31.
[36] Y.H. Qian, J.Y. Liang, W. Pedrycz, C.Y. Dang, Positive approximation: an accelerator for attribute reduction in rough set theory, Artif. Intell. 174 (2010) 597–618.
[37] R. Susmaga, Reducts and constructs in attribute reduction, Fund. Inform. 61 (2004) 159–181.
[38] Y. Zhao, Y. Yao, F. Luo, Data analysis based on discernibility and indiscernibility, Inf. Sci. 177 (2007) 4959–4976.
[39] S.H. Teng, D. Zhang, L. Cui, et al., A new uncertainty measure of rough sets, in: Proceedings of the IEEE International Conference on Robotics and Biomimetics, Guilin, China, 2009, pp. 1189–1193.
[40] S.H. Teng, J.W. Wu, J.X. Sun, et al., An efficient attribute reduction algorithm, in: Proceedings of the IEEE International Conference on Advanced Computer Control, Shenyang, China, 2010, pp. 471–475.
[41] Y. Qian, J. Liang, Combination entropy and combination granulation in rough set theory, Int. J. Uncertainty Fuzziness Knowledge-Based Syst. 16 (2008) 179–193.
[42] W.X. Zhang, G.F. Qiu, W.Z. Wu, A general approach to attribute reduction in rough set theory, Sci. China Inf. Sci. 50 (2007) 188–197.
[43] Z.Y. Xu, Z.P. Liu, B.R. Yang, W. Song, A quick attribute reduction algorithm with complexity max {O(|C||U|), O(|C^2||U/C|)}, Chin. J. Comput. 29 (2006) 391–399.
[44] G.Y. Wang, Relationship between the algebra view and information view of rough set, Proc. SPIE 5098 (2003) 103–113.