Knowledge-Based Systems 44 (2013) 90–100

Contents lists available at SciVerse ScienceDirect

Knowledge-Based Systems

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

An accelerator for attribute reduction based on perspective of objects and attributes

Jiye Liang [b], Junrong Mi [a], Wei Wei [b,*], Feng Wang [b]

[a] School of Management, Shanxi University, Taiyuan, 030006 Shanxi, China

[b] Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, School of Computer and Information Technology, Shanxi University, Taiyuan, 030006 Shanxi, China

ARTICLE INFO

Article history:
Received 3 January 2012
Received in revised form 28 December 2012
Accepted 25 January 2013
Available online 4 February 2013

Keywords:
Feature selection
Accelerating algorithm
Attribute reduction
Rough set
Large scale data

ABSTRACT

Feature selection is an active area of research in pattern recognition, machine learning and artificial intelligence, which greatly improves the performance of forecasting or classification. In rough set theory, attribute reduction, as a special form of feature selection, aims to retain the discernability of the original attribute set. To solve this problem, many heuristic attribute reduction algorithms have been proposed in the literature. However, these methods are computationally time-consuming for large scale datasets. Recently, an accelerator was introduced by computing reducts on gradually reducing the size of the universe. Although the accelerator can considerably shorten the computational time, it remains a challenging issue. To further enhance the efficiency of these algorithms, we develop a new accelerator for attribute reduction, which simultaneously reduces the size of the universe and the number of attributes at each iteration of the process of reduction. Based on the new accelerator, several representative heuristic attribute reduction algorithms are accelerated. Experiments show that these accelerated algorithms can significantly reduce computational time while maintaining their results the same as before.

1. Introduction

Feature selection is a preprocessing step in many applications including pattern recognition, machine learning, and data mining. Attribute reduction is regarded as a special form of feature selection in rough set theory and aims to retain the discriminatory power of the original attribute set [21,22,25,44]. In databases of practical applications (Image processing, Bioinformatics, Astronomy, Finance, etc.), the number of objects is very large and the dimension (the number of attributes) is very high as well [1,2,4,24]. It is well known that attributes irrelevant to recognition tasks may deteriorate the performance of learning algorithms [6,27]. In other words, storing and processing irrelevant attributes could be computationally very expensive. To address this issue, irrelevant attributes, as pointed out in [6,29], can be omitted, which will not severely affect the classification (recognition) accuracy. Therefore, the omission of some irrelevant attributes would be desirable relative to the costs involved [20].

According to how to combine the feature subset search with the construction of the classification model, feature selection techniques can be organized into three categories: wrapper strategy [11], filter strategy [4], and embedded strategy [30]. The wrapper strategy uses a classifier to assess feature subsets and train a learning machine for every feature subset considered. The interaction between feature subset search and classification model is its significant advantage. However, it thus is usually time-consuming[11,30]. The filter strategy employs another evaluation criterion different from the target classification scheme, and therefore usually does not involve any learning machine in the features selection process [4]. The embedded strategy generates candidate subsets by means of the methods used in filter strategy, and searches an optimal subset of features based on the classifier construction. The embedded methods combine the advantages of wrapper methods and filter methods, that they include the interaction with the classification model, while at the same time being far less computational than wrapper methods [30]. This paper focuses on the filter strategy in order to pursue both computational efficiency and solution quality regardless of a classification scheme. In filter methods, some common feature selection criteria are introduced as stopping conditions, which include information gain [12], consistency [2], and dependency [18]. These criteria can be divided into two main categories: distance-based and consistency-based [6]. For consistency-based feature selection, attribute reduction in rough set theory offers a systematic theoretic framework, which does not attempt to maximize the class separability but rather to retain the discernible ability of original attribute sets for the objects from the universe [9,36].

In recent years, many methods have been proposed and examined for finding reducts. Skowron [33] proposed an attribute reduction algorithm using a discernibility matrix, which can find

* Corresponding author. Tel./fax: +86 0351 7018176.
E-mail addresses: ljy@sxu.edu.cn (J. Liang), jrmi@sxu.edu.cn (J. Mi), weiwei@sxu.edu.cn (W. Wei), sxuwangfeng@126.com (F. Wang).

0950-7051/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2013.01.027

all reducts. Kryszkiewicz and Lasek [10] proposed an approach to the discovery of minimal sets of attributes functionally determining a decision attribute. Hu and Cercone [8] proposed a heuristic attribute reduction method, called positive-region reduction, which remains the positive region of target decision unchanged. Furthermore, many researchers introduced various information entropies (Shannon's entropy, complement entropy, combination entropy, etc.) to measure the uncertainty of an information table, and constructed the corresponding attribute reduction algorithms [13,14,16,17,28,38,39]. To handle hybrid data with numerical and categorical features, fuzzy rough set model and rough fuzzy set model were employed to obtain attribute reducts [5–7,31,32,37,42]. In addition, β-reduct proposed by Ziarko provides a suite of reduction methods in the variable precision rough set model [47]. By means of the tolerance rough set model, Parthaláin and Shen presented a new approach to deal with real-valued data, which can retaining dataset semantics [19]. Yao and Zhao introduced attribute reduction in decision-theoretic rough set models in the context of different classification properties, which provided a new insight into the problem of attribute reduction [45].

These attribute reduction algorithms mentioned above can be divided into two categories: finding all reducts (or an optimal reduct) and finding one reduct [3,46]. However, it has been proved to be an NP-hard problem to find all reducts [43]. By contrast, heuristic algorithms (finding one reduct) can efficiently lessen the computational burden of attribute reduction [5,6,8,13,14,28,34,41]. In this paper, we efforts to further improve the efficiency of heuristic algorithms. For convenience of our further development, we classify these attribute reduction methods in terms of heuristics into four categories: positive region reduction [8,21–23], Shannon's entropy reduction [34,35], complement entropy reduction [13,15] and combination entropy reduction [28]. Each of these heuristic methods can extract a single reduct from a given decision table and preserves the particular property of the decision table. Although these heuristic methods are much faster, attribute reduction still remains a computationally difficult problem when data sets are large. To overcome this difficulty, Qian and Liang [29] proposed an accelerator for attribute reduction based on positive approximation. The heuristic methods based on the accelerator can significantly decrease the time consuming and obtain the same attribute reduct as their original versions. In [26,40], this idea of accelerator was extended to incomplete data and hybrid data, and these corresponding accelerators can significantly improved the performance of attribute reduction algorithms. However, by means of the accelerator, only the insignificant objects are removed from datasets in each iteration of computing reducts. It has been observed that the number of attributes in datasets can also largely affect the efficiency of attribute reduction. This motivates the idea of this paper. In order to further improve the performance of the heuristic attribute reduction methods, we develop a new accelerator by gradually reducing not only the size of universe but also the number of attributes in each iteration of attribute reduction. By incorporating the new accelerator into each of the above four representative heuristic attribute reduction methods, we obtain their accelerating versions. Numerical experiments show that each of the improved methods can obtain the same attribute subset as its corresponding original method while greatly saving computational cost, especially for the large scale datasets.

The rest of study is organized as follows. A brief review of relative basic concepts in Section 2. In Section 3, through analyzing the rank preservation of four representative significant measures of attributes, we develop a new accelerator based on the perspective of objects and attributes. Experiments on ten datasets in UCI machine learning repository show that the four representative heuristic algorithms based on the proposed accelerator outperform their original counterparts in terms of time consuming in Section 4. Then, conclusion and future work come in Section 5.

2. Preliminaries

In this section, we review some basic concepts such as indiscernibility relation, partition, significance measures and forward attribute reduction algorithms.

2.1. Rough approximations

An information table is a 4-tuple $S = (U,A,V,f)$ (for short $S = (U,A)$), where U is a non-empty and finite set of objects, called a universe, and A is a non-empty and finite set of attributes, V_a is the domain of the attribute a, $V = \bigcup_{a \in A} V_a$ and $f: U \times A = V$ is a function $f(x,a) \in V_a$ for each $a \in A$ [21–23].

An indiscernibility relation $R_B = \{(x,y) \in U \times U | f(x,a) = f(y,a), \forall a \in B\}$ was determined by a non-empty subset $B \subseteq A$. $U/R_B = \{[x]_B | x \in U\}$ (just as U/B) indicates the partition of U resulted from R_B, where $[x]_B$ denotes the equivalence class determined by x with respect to B, i.e., $[x]_B = \{y \in U | (x,y) \in R_B\}$.

Furthermore, given an information table $S = (U,A)$ and an object subset $X \subseteq U$, $B \subseteq A$, one can construct a rough set of X on the universe by elemental information granules in the following definition:

$$\underline{B}X = \cup\{[x]_B | [x]_B \subseteq X\}, \text{ and } \overline{B}X = \cup\{[x]_B | [x]_B \cap X \neq \emptyset\},$$

where $\underline{B}X$ and $\overline{B}X$ are called B-lower approximation and B-upper approximation with respect to B, respectively. The order pair $\langle \underline{B}X, \overline{B}X \rangle$ is called a rough set of X.

There are two kinds of attributes for a classification problem, which can be characterized by a decision table $DT = (U, C \cup D)$ with $C \cap D = \emptyset$, where an element of C is called a condition attribute, C is called a condition attribute set, an element of D is called a decision attribute, and D is called a decision attribute set.

Given a decision table $DT = (U, C \cup D)$, $B \subseteq C$, $U/D = \{Y_1, Y_2, \ldots, Y_n\}$, the lower and upper approximations of the decision attribute set D are defined as

$$\underline{B}D = \{\underline{B}Y_1, \underline{B}Y_2, \ldots, \underline{B}Y_n\}, \text{ and } \overline{B}D = \{\overline{B}Y_1, \overline{B}Y_2, \ldots, \overline{B}Y_n\}.$$

Let $POS_B^{(U,C)}(D) = \bigcup_{i=1}^{n} \underline{B}Y_i$, which is called the positive region of D with respect to B in the decision table $DT = (U, C \cup D)$.

2.2. Four representative significance measures of attributes

In heuristic attribute reduction methods, attribute significance measure is a crucial factor. Therefore, we will introduce four representative significance measures here, which are based on positive region, Shannon's conditional entropy, complement conditional entropy and combination conditional entropy.

- Positive region (PR) was first employed in a heuristic attribute reduction algorithm, called positive region reduction, which keeps the positive region of target decision unchanged [8].
- Shannon's conditional entropy (SCE) was introduced to search reducts of a decision table [34,38]. This reduction algorithm calls Shannon's entropy reduction, which remains the conditional entropy of target decision. Shannon's conditional entropy of B with respect to D in $DT = (U, C \cup D)$ is denoted as

$$H^{(U,C)}(D|B) = -\sum_{i=1}^{m} p(X_i) \sum_{j=1}^{n} p(Y_j|X_i) \log(p(Y_j|X_i)),$$

where $p(X_i) = \frac{|X_i|}{|U|}$ and $p(Y_j|X_i) = \frac{|X_i \cap Y_j|}{|X_i|}$, and X is a non-empty set.

- Complement conditional entropy (PCE) was defined to measure the uncertainty and applied to reduce redundant attribute of a decision table [13,14]. The reduction method based on the entropy is called complement entropy reduction, which can preserve the conditional entropy of a given decision table. The conditional entropy of B with respect to D in $DT = (U, C \cup D)$ is denoted as

$$E^{(U,C)}(D|C) = \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{|Y_j \cap X_i|}{|U|} \frac{|Y_j^c - X_i^c|}{|U|},$$

where Y_j^c and X_i^c are the complements of Y_j and X_i, respectively.

- Combination conditional entropy (CCE) is based on the intuitionistic knowledge content nature of information gain, which can be used to obtain attribute reducts [28]. The reduction method can remain combination conditional entropy of a given decision table. The conditional entropy of B with respect to D in $DT = (U, C \cup D)$ is defined as

$$CE^{(U,C)}(D|C) = \sum_{i=1}^{m}\left(\frac{|X_i|}{|U|}\frac{C^2_{|X_i|}}{C^2_{|U|}} - \sum_{j=1}^{n}\frac{|X_i \cap Y_j|}{|U|}\frac{C^2_{|X_i \cap Y_j|}}{C^2_{|U|}}\right),$$

where $C^2_{|X_i|} = \frac{|X_i| \times (|X_i|-1)}{2}$ denotes the number of pairs of the objects which are not distinguishable from each other in the equivalence class X_i.

The corresponding significance measures based on the measures mentioned above are given as follows.

Let $DT = (U, C \cup D)$ be a decision table and $B \subseteq C$. For $\forall a \in B$, the inner significance measures of a based on positive region, Shannon's conditional entropy, complement conditional entropy and combination conditional entropy are respectively defined as

$$Sig_1^{inner}(a, B, C, D, U) = \gamma_B^{(U,C)}(D) - \gamma_{B-\{a\}}^{(U,C)}(D),$$
$$Sig_2^{inner}(a, B, C, D, U) = H^{(U,C)}(D|B - \{a\}) - H^{(U,C)}(D|B),$$
$$Sig_3^{inner}(a, B, C, D, U) = E^{(U,C)}(D|B - \{a\}) - E^{(U,C)}(D|B),$$
$$Sig_4^{inner}(a, B, C, D, U) = CE^{(U,C)}(D|B - \{a\}) - CE^{(U,C)}(D|B),$$

where $\gamma_B^{(U,C)}(D) = \frac{|POS_B^{(U,C)}(D)|}{|U|}$.

By means of the inner significant measures, the definition of core [13,21,28,38] can be denoted as follows:

Let $S = (U, C \cup D)$ be a decision table and $a \in C$. If $Sig_\Delta^{inner}(a, C, C, D, U) > 0 (\Delta = 1, 2, 3, 4)$, then a is a core attribute of S in the context of type Δ.

Furthermore, we suppose $S = (U, C \cup D)$ be a decision table and $B \subseteq C$. For $\forall a \in C - B$, the outer significance measures of a based on positive region, Shannon's conditional entropy, complement conditional entropy and combination conditional entropy are respectively defined as

$$Sig_1^{outer}(a, B, C, D, U) = \gamma_{B\cup\{a\}}^{(U,C)}(D) - \gamma_B^{(U,C)}(D),$$
$$Sig_2^{outer}(a, B, C, D, U) = H^{(U,C)}(D|B) - H^{(U,C)}(D|B \cup \{a\}),$$
$$Sig_3^{outer}(a, B, C, D, U) = E^{(U,C)}(D|B) - E^{(U,C)}(D|B \cup \{a\}),$$
$$Sig_4^{outer}(a, B, C, D, U) = CE^{(U,C)}(D|B) - CE^{(U,C)}(D|B \cup \{a\}),$$

where $\gamma_B^{(U,C)}(D) = \frac{|POS_B^{(U,C)}(D)|}{|U|}$.

2.3. Forward attribute reduction algorithms

In rough set theory, many heuristic attribute reduction algorithms have been designed to achieve efficiently attribute reducts, in which forward greedy search strategy is common [5,6,8,13,15,29,34]. In general, starting with an attribute with the maximal inner significance measure, a forward greedy attribute reduction approach takes an attribute with the maximal outer importance into the attribute reduct in each loop until this subset satisfies the stopping criterion, which yields an attribute reduct. Formally, a forward greedy attribute reduction algorithm can be written as follows.

Algorithm 1. [8,29,38]. General forward greedy attribute reduction algorithm

Input: Decision table $S = (U, C \cup D)$;
Output: One reduct *red*.
Step 1: $red \leftarrow \emptyset$;//*red* is the pool to conserve the selected attributes
Step 2: Compute $Sig^{inner}(a_k, C, C, D, U)$, $k \leqslant |C|$;
Step 3: Put a_k into *red*, where $Sig^{inner}(a_k, C, C, D, U) > 0$;
Step 4: While $EF^{(U,C)}(red, D) \neq EF^{(U,C)}(C, D)$ Do//This provides a stopping criterion.
$\{red \leftarrow red \cup \{a_0\}$, where $Sig^{outer}(a_0, red, C, D, U) = max\{Sig^{outer}(a_k, red, C, D, U), a_k \in C - red\}\}$;
Step 5: return *red* and end.

3. Rank preservation of significance measures of attributes

It is well known that each of the significance measures of attributes provides some heuristic information for forward attribute reduction algorithms. In this section, to further improve the performance of these attribute reduction algorithms, we will focus on the rank preservation of the four significance measures of attributes from the perspective of decreasing the number of objects and attributes simultaneously.

In order to prove the rank preservation of a significance measure of attributes, we need the following lemma.

Lemma 3.1. *Let* $0 \leqslant a_i, b_i \leqslant 1, i = 1, 2, \ldots, n, \sum_{i=1}^{n} a_i = 1,$ *and* $\sum_{i=1}^{n} b_i = 1$. *If* $\sum_{i=1}^{n} a_i \times b_i = 1$, *then* $\exists 1 \leqslant u \leqslant n$ *such that* $a_u = b_u = 1$ *and* $a_k = b_k = 0$ *for* $\forall k \neq u$.

Proof. By means of the existing conditions, we have that

$$\sum_{i=1}^{n} a_i \times b_i = \sum_{i=1}^{n}\left(a_i \times \left(1 - \sum_{j=1, j\neq i}^{n} b_j\right)\right) = \sum_{i=1}^{n} a_i - \sum_{i=1}^{n}\left(a_i \times \sum_{j=1, j\neq i}^{n} b_j\right)$$
$$= 1 - \sum_{i=1}^{n}\left(a_i \times \sum_{j=1, j\neq i}^{n} b_j\right).$$

Thus, one has

$$\sum_{i=1}^{n} a_i \times b_i = 1 \iff 1 - \sum_{i=1}^{n}\left(a_i \times \sum_{j=1, j\neq i}^{n} b_j\right) = 1$$
$$\iff \sum_{i=1}^{n}\left(a_i \times \sum_{j=1, j\neq i}^{n} b_j\right) = 0$$
$$\iff a_i \times \sum_{j=1, j\neq i}^{n} b_j = 0, \ for \ \forall i \leqslant n$$
$$\iff a_i = 0 \text{ or } \sum_{j=1, j\neq i}^{n} b_j = 0, \ for \ \forall i \leqslant n$$

Furthermore, because of $\sum_{i=1}^{n} a_i = 1$, there exists $u \leqslant n$ such that $a_u \neq 0$. Therefore $\sum_{j=1, j\neq u}^{n} b_j = 0$, i.e., $b_k = 0, for \forall k \neq u$. And because of $\sum_{j=1}^{n} b_j = 1$, we can obtain $b_u = 1$.

Then, we have $\sum_{j=1, j\neq k}^{n} b_j = 1$, for $\forall k \neq u$, thus $a_k = 0$, for $\forall k \neq u$, i.e., $\sum_{i=1, i\neq u}^{n} a_i = 0$. So, it is obvious that $a_u = 1$.

That is to say, $\exists u \leqslant n$ such that $a_u = b_u = 1$ and $a_k = b_k = 0$ for $\forall k \neq u$. □

Based on Lemma 3.1, we give the following theorem.

Theorem 3.1. *Let $DT = (U, C \cup D)$ be a decision table, $B \subseteq C$. If $E^{(U,C)}(D|B) = E^{(U,C)}(D|C)$, then $POS_B^{(U,C)}(D) = POS_C^{(U,C)}(D)$ and $U'_B/B = U'_C/C$, where $U'_B = U - POS_B^{(U,C)}(D), U'_C = U - POS_C^{(U,C)}(D)$.*

Proof. By the existing condition $B \subseteq C$, it is obvious that $U/B \succeq U/C$. Without any loss of generalization, we suppose that $U/C = \{X_1, X_2, \ldots, X_m\}$, $U/B = \{X_1, X_2, \ldots, X_{u-1}, X_{u+1}, \ldots, X_{v-1}, X_{v+1}, \ldots, X_m, X_u, X_{u+1}, \ldots, X_{v-1}, X_{v+1}, \ldots, X_m, X_u \cup X_v\}$ and $U/D = \{Y_1, Y_2, \ldots, Y_n\}$, then

$$\begin{aligned}E^{(U,C)}(D|C) - E^{(U,C)}(D|B) &= \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{|X_i \cap Y_j|}{|U|}\frac{\left|Y_j^c - X_i^c\right|}{|U|} \\ &\quad - \sum_{i=1, i\neq u,v}^{m}\sum_{j=1}^{n}\frac{|X_i \cap Y_j|}{|U|}\frac{\left|Y_j^c - X_i^c\right|}{|U|} \\ &\quad - \sum_{j=1}^{n}\frac{|(X_u \cup X_v) \cap Y_j|}{|U|}\frac{\left|Y_j^c - (X_u \cup X_v)^c\right|}{|U|} \\ &= \sum_{j=1}^{n}\frac{|X_u \cap Y_j| + |X_v \cap Y_j|}{|U|} \\ &\quad \times \frac{|X_u - Y_j| + |X_v - Y_j|}{|U|} - \sum_{j=1}^{n}\frac{|X_u \cap Y_j|}{|U|} \\ &\quad \times \frac{|X_u - Y_j|}{|U|} - \sum_{j=1}^{n}\frac{|X_v \cap Y_j|}{|U|}\frac{|X_v - Y_j|}{|U|} \\ &= \sum_{j=1}^{n}\frac{|X_u \cap Y_j|}{|U|}\frac{|X_v| - |X_v \cap Y_j|}{|U|} \\ &\quad + \sum_{j=1}^{n}\frac{|X_v \cap Y_j|}{|U|}\frac{|X_u| - |X_u \cap Y_j|}{|U|} \\ &= \sum_{j=1}^{n}\frac{|X_u||X_v|(\mu_{uj} + \mu_{vj} - 2\mu_{uj} \times \mu_{vj})}{|U|^2},\end{aligned}$$

where $\mu_{ij} = \frac{|X_i \cap Y_j|}{|X_i|}, 0 \leqslant \mu_{ij} \leqslant 1$.

Furthermore, because $E^{(U,C)}(D|C) - E^{(U,C)}(D|B) = 0$, we have that

$$\begin{aligned}\sum_{j=1}^{n}(\mu_{uj} + \mu_{vj} - 2\mu_{uj} \times \mu_{vj}) = 0 &\iff \sum_{j=1}^{n}\mu_{uj} + \sum_{j=1}^{n}\mu_{vj} \\ &= 2\sum_{j=1}^{n}\mu_{uj} \times \mu_{vj} \iff 2 \\ &= 2\sum_{j=1}^{n}\mu_{uj} \times \mu_{vj} \iff \sum_{j=1}^{n}\mu_{uj} \times \mu_{vj} = 1.\end{aligned}$$

According to Lemma 3.1, if $\sum_{j=1}^{n}\mu_{uj} \times \mu_{vj} = 1$, then $\exists w \leqslant n$ such that $\mu_{uw} = \mu_{vw} = 1$ and $\mu_{uj} = \mu_{vj} = 0$ for $j \leqslant n$ $(j \neq w)$, that is to say, the equivalent classes X_u and X_v belong to the same decision class, i.e., X_u, $X_v \subseteq Y_w$. Thus, $X_u \cup X_v \subseteq POS_B^{(U,C)}(D)$ and $X_u, X_v \subseteq POS_C^{(U,C)}(D)$.

And because of $U/C = \{X_1, X_2, \ldots, X_m\}$ and $U/B = \{X_1, X_2, = \{X_1, X_2, \ldots, X_{u-1}, X_{u+1}, \ldots, X_{v-1}, X_{v+1}, \ldots, X_m, X_u \cup X_v\}$, the objects in $POS_B^{(U,C)}(D)$ is the same as the ones in $POS_C^{(U,C)}(D)$, and the equivalence classes in U'_C are identical with the ones in U'_B.

Therefore, if $U/B \succeq U/C$ and $E^{(U,C)}(D|B) = E^{(U,C)}(D|C)$, then $POS_B^{(U,C)}(D) = POS_C^{(U,C)}(D)$ and $U'_B/B = U'_C/C$. □

Theorem 3.1 states, for two different decision tables, the equivalence classes that are not in the positive regions of them are identical with each other if the partition derived from the condition attribute set in one decision table is coarser than the one in the other and the values of complement conditional entropy of these two tables are equal.

Theorem 3.2 [29]. *Let $DT = (U, C \cup D)$ be a decision table, $B \subseteq C$, then*

(1) $Sig_1^{outer}(a, B, C, D, U) = \dfrac{|U'_B|}{|U|}Sig_1^{outer}(a, B, C, D, U'_B)$,

(2) $H^{(U,C)}(D|B) = \dfrac{|U'_B|^2}{|U|^2}H^{(U'_B,C)}(D|B)$,

(3) $E^{(U,C)}(D|B) = \dfrac{|U'_B|^2}{|U|^2}E^{(U'_B,C)}(D|B)$,

(4) $CE^{(U,C)}(D|B) = \dfrac{|U'_B|^2}{|U|^2}CE^{(U'_B,C)}(D|B)$,

where $U'_B = U - POS_B^{(U,C)}(D)$.

By means of Theorem 3.2, the inherent relationships between the outer significance measures based on positive region, Shannon's conditional entropy, complement conditional entropy and combination conditional entropy in $(U, C \cup D)$ and in $(U'_B, C \cup D)$ were revealed.

Theorem 3.3. *Let $DT = (U, C \cup D)$ be a decision table, $B \subseteq B' \subseteq C$. If $Sig_3^{outer}(a, B, C, D, U) = 0$ for $\forall a \in C - B'$, then $Sig_3^{outer}(a, B', C, D, U) = 0$, where Δ=1, 2, 3, 4.*

Proof. By the existing condition $Sig_3^{outer}(a, B, C, D, U) = 0$, i.e. $E^{(U,C)}(D|B \cup \{a\}) = E^{(U,C)}(D|B)$ and Theorem 3.1, we have that $POS_B^{(U,C)}(D) = POS_{B\cup\{a\}}^{(U,C)}(D), U'_B = U'_{B\cup\{a\}}$ and $U'_B/B = U'_{B\cup\{a\}}/(B \cup \{a\})$.

For convenience, we suppose

$$U'_B/B' = \{X_1, X_2, \ldots, X_p\}, U'_{B\cup\{a\}}/(B' \cup \{a\}) = \{X'_1, X'_2, \ldots, X'_p\}$$

$$(X_i = X'_i \text{ for } \forall i \leqslant p),$$

$$POS_B^{(U,C)}(D)/B' = \{X_{p+1}, X_{p+2}, \ldots, X_m\},$$

$$POS_{B\cup\{a\}}^{(U,C)}(D)/(B' \cup \{a\}) = \{X'_{p+1}, X'_{p+2}, \ldots, X'_l\}(l \geqslant m).$$

By means of different values of Δ, four cases will be considered in the following proof.

(1) Δ = 1

Because of $B' \supseteq B$, it is obvious that $U'_B/B' = U'_{B\cup\{a\}}/(B' \cup \{a\})$, and then $POS_{B'}^{(U,C)}(D) = POS_{B'\cup\{a\}}^{(U,C)}(D)$. Thus, we can obtain that

$$\begin{aligned}POS_{B'}^{(U,C)}(D) &= \cup\{X_i | X_i \subseteq Y_j, X_i \in U'_B/B', Y_j \in U/D\} \\ &\quad \cup POS_B^{(U,C)}(D) \\ &= \cup\left\{X'_i | X'_i \subseteq Y_j, X'_i \in U'_{B\cup\{a\}}/B', Y_j \in U/D\right\} \\ &\quad \cup POS_{B\cup\{a\}}^{(U,C)}(D) \\ &= POS_{B'\cup\{a\}}^{(U,C)}(D).\end{aligned}$$

Furthermore, we can obtain that

$$\begin{aligned}Sig_1^{outer}(a, B', C, D, U) &= \frac{1}{|U|} \times \left(|POS_{B'\cup\{a\}}^{(U,C)}(D)| - |POS_{B'}^{(U,C)}(D)|\right) \\ &= 0.\end{aligned}$$

(2) Δ = 2

By the condition $B' \supseteq B$, it is easy to obtain $U'_B/B' = U'_{B\cup\{a\}}/(B' \cup \{a\}), POS_B^{(U,C)}(D) \subseteq POS_{B'}^{(U,C)}(D)$ and $POS_{B\cup\{a\}}^{(U,C)}(D) \subseteq POS_{B'\cup\{a\}}^{(U,C)}(D)$. Therefore, we can obtain that

$$Sig_2^{outer}(a,B',C,D,U) = H^{(U,C)}(D|B') - H^{(U,C)}(D|B'\cup\{a\})$$
$$= -\sum_{i=1}^{m}\frac{|X_i|}{|U|}\sum_{j=1}^{n}\frac{|Y_j\cap X_i|}{|X_i|}\log\frac{|Y_j\cap X_i|}{|X_i|} + \sum_{k=1}^{l}\frac{|X'_k|}{|U|}\sum_{j=1}^{n}\frac{|Y_j\cap X'_k|}{|X'_k|}\log\frac{|Y_j\cap X'_k|}{|X'_k|}$$
$$= -\sum_{i=1}^{p}\frac{|X_i|}{|U|}\sum_{j=1}^{n}\frac{|Y_j\cap X_i|}{|X_i|}\log\frac{|Y_j\cap X_i|}{|X_i|} - \sum_{i=p+1}^{m}\frac{|X_i|}{|U|}\sum_{j=1}^{n}\frac{|Y_j\cap X_i|}{|X_i|}\log\frac{|Y_j\cap X_i|}{|X_i|} + \sum_{k=1}^{p}\frac{|X_k|}{|U|}\sum_{j=1}^{n}\frac{|Y_j\cap X_k|}{|X_k|}\log\frac{|Y_j\cap X_k|}{|X_k|} + \sum_{k=p+1}^{l}\frac{|X'_k|}{|U|}\sum_{j=1}^{n}\frac{|Y_j\cap X'_k|}{|X'_k|}\log\frac{|Y_j\cap X'_k|}{|X'_k|}$$
$$= 0.$$

(3) $\Delta = 3$
From the existing condition $B'\supseteq B$, we have that $U'_B/B' = U'_{B\cup\{a\}}/(B'\cup\{a\}), POS_B^{(U,C)}(D)\subseteq POS_{B'}^{(U,C)}(D)$ and $POS_{B\cup\{a\}}^{(U,C)}(D)\subseteq POS_{B'\cup\{a\}}^{(U,C)}(D)$. Therefore, we can obtain that

$$Sig_3^{outer}(a,B',C,D,U) = H^{(U'_B,C'_B)}(D|B') - H^{(U'_B,C'_B)}(D|B'\cup\{b\})$$
$$= \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{|Y_j\cap X_i|}{|U|}\frac{\left|Y_j^c - X_i^c\right|}{|U|} - \sum_{k=1}^{l}\sum_{j=1}^{n}\frac{|Y_j\cap X'_k|}{|U|}\frac{\left|Y_j^c - X_k'^c\right|}{|U|}$$
$$= \sum_{i=1}^{p}\sum_{j=1}^{n}\frac{|Y_j\cap X_i|}{|U|}\frac{\left|Y_j^c - X_i^c\right|}{|U|} + \sum_{i=p+1}^{m}\sum_{j=1}^{n}\frac{|Y_j\cap X_i|}{|U|}\frac{\left|Y_j^c - X_i^c\right|}{|U|} - \sum_{k=1}^{p}\sum_{j=1}^{n}\frac{|Y_j\cap X_k|}{|U|}\frac{\left|Y_j^c - X_k^c\right|}{|U|} - \sum_{k=p+1}^{l}\sum_{j=1}^{n}\frac{|Y_j\cap X'_k|}{|U|}\frac{\left|Y_j^c - X_k'^c\right|}{|U|}$$
$$= 0.$$

(4) $\Delta = 4$
By means of $B'\supseteq B$, it is obvious that $U'_B/B', POS_B^{(U,C)}(D)\subseteq POS_{B'}^{(U,C)}(D)$ and $POS_{B\cup\{a\}}^{(U,C)}(D)\subseteq POS_{B'\cup\{a\}}^{(U,C)}(D)$. Therefore, we have that

$$Sig_4^{outer}(a,B',C,D,U) = CE^{(U'_B,C'_B)}(D|B') - CE^{(U'_B,C'_B)}(D|B'\cup\{b\})$$
$$= \sum_{i=1}^{m}\left(\frac{|X_i|}{|U|}\frac{C^2_{|X_i|}}{C^2_{|U|}} - \sum_{j=1}^{n}\frac{|X_i\cap Y_j|}{|U|}\frac{C^2_{|X_i\cap Y_j|}}{C^2_{|U|}}\right) - \sum_{k=1}^{l}\left(\frac{|X'_k|}{|U|}\frac{C^2_{|X'_k|}}{C^2_{|U|}} - \sum_{j=1}^{n}\frac{|X'_k\cap Y_j|}{|U|}\frac{C^2_{|X'_k\cap Y_j|}}{C^2_{|U|}}\right)$$
$$= \sum_{i=1}^{p}\left(\frac{|X_i|}{|U|}\frac{C^2_{|X_i|}}{C^2_{|U|}} - \sum_{j=1}^{n}\frac{|X_i\cap Y_j|}{|U|}\frac{C^2_{|X_i\cap Y_j|}}{C^2_{|U|}}\right) + \sum_{i=p+1}^{m}\left(\frac{|X_i|}{|U|}\frac{C^2_{|X_i|}}{C^2_{|U|}} - \sum_{j=1}^{n}\frac{|X_i\cap Y_j|}{|U|}\frac{C^2_{|X_i\cap Y_j|}}{C^2_{|U|}}\right) - \sum_{k=1}^{p}\left(\frac{|X_k|}{|U|}\frac{C^2_{|X_k|}}{C^2_{|U|}} - \sum_{j=1}^{n}\frac{|X_k\cap Y_j|}{|U|}\frac{C^2_{|X_k\cap Y_j|}}{C^2_{|U|}}\right) - \sum_{k=p+1}^{l}\left(\frac{|X'_k|}{|U|}\frac{C^2_{|X'_k|}}{C^2_{|U|}} - \sum_{j=1}^{n}\frac{|X'_k\cap Y_j|}{|U|}\frac{C^2_{|X'_k\cap Y_j|}}{C^2_{|U|}}\right)$$
$$= 0.$$

In any case, for each of Δ, if $Sig_3^{outer}(a,B,C,D,U) = 0$ for $\forall a\in C-B'$, then $Sig_\Delta^{outer}(a,B',C,D,U) = 0$. □

Theorem 3.3 states that the significance measures based on positive region, Shannon's conditional entropy, complement conditional entropy and combination conditional entropy of a with respect to the attribute set B' (B' is a superset of B) are zero if the significance measure based on complement conditional entropy of a with respect to B is zero.

Corollary 3.1. *Let DT = (U, C ∪ D) be a decision table, $B\subseteq B'\subseteq C$. If* $B_3^* = \left\{a\middle|Sig_3^{outer}(a,B,C,D,U) = 0, a\in C-B'\right\}$,

$B_\Delta^{**} = \left\{a\middle|Sig_\Delta^{outer}(a,B',C,D,U) = 0, a\in C-B'\right\}$, *then*

$B_3^*\subseteq B_\Delta^{**}$,

where $\Delta = 1, 2, 3, 4$.

It is easy to prove this corollary by means of Theorem 3.3.

Theorem 3.4. *Let DT = (U, C ∪ D) be a decision table, $B\subseteq B'\subseteq C$. If $b,c\in C-B'-B_\Delta^{**}$ and $Sig_\Delta^{outer}(b,B',C,D,U) > Sig_\Delta^{outer}(c,B',C,D,U)$, then*

$$Sig_\Delta^{outer}(b,B',C'_B,D,U'_B) > Sig_\Delta^{outer}(c,B',C'_B,D,U'_B),$$

where $C'_B = C - B_3^*, B_3^* = \left\{a\middle|Sig_3^{outer}(a,B,C,D,U) = 0, a\in C-B'\right\}$, $U'_B = U - POS_B^{(U,C)}(D), B_\Delta^{**} = \left\{a\middle|Sig_\Delta^{outer}(a,B',C,D,U) = 0, a\in C-B'\right\}$, $\Delta = 1,2,3,4$.

Proof. In terms of the different values of Δ, we will give the proof from the following four cases.

(1) $\Delta = 1$
By the existing condition $Sig_1^{outer}(b,B',C,D,U) > Sig_1^{outer}(c,B',C,D,U)$ and Theorem 3.2, we have that

$$Sig_1^{outer}(b,B',C,D,U'_B) > Sig_1^{outer}(c,B',C,D,U'_B).$$

From the existing condition $b,c\in C-B'-B_1^{**}$ and Corollary 3.1, it is obvious that $b,c\notin B_3^*$, where $B_3^* = \left\{a\middle|Sig_3^{outer}(a,B,C,D,U) = 0, a\in C-B'.\right\}, B_1^{**} = \left\{a\middle|Sig_1^{outer}(a,B',C,D,U) = 0, a\in C-B'\right\}$. Therefore, we have that $B'\cap B_3^* = \emptyset$ and

$$Sig_1^{outer}(b,B',C'_B,D,U'_B) = \frac{1}{|U'_B|}\times\left(\left|POS_{B'\cup\{b\}}^{(U'_B,C'_B)}(D)\right| - \left|POS_{B'}^{(U'_B,C'_B)}(D)\right|\right)$$
$$= \frac{1}{|U'_B|}\times\left(\left|POS_{B'\cup\{b\}}^{(U'_B,C-B_3^*)}(D)\right| - \left|POS_{B'}^{(U'_B,C-B_3^*)}(D)\right|\right)$$
$$= \frac{1}{|U'_B|}\times\left(\left|POS_{B'\cup\{b\}}^{(U'_B,C)}(D)\right| - |POS_{B'}^{(U'_B,C)}(D)|\right)$$
$$= Sig_1^{outer}(b,B',C,D,U'_B).$$

In similarity,

$$Sig_1^{outer}(c,B',C'_B,D,U'_B) = Sig_1^{outer}(c,B',C,D,U'_B).$$

Therefore, one has

$$Sig_1^{outer}(b,B',C'_B,D,U'_B) > Sig_1^{outer}(c,B',C'_B,D,U'_B).$$

(2) $\Delta = 2$
By the existing condition $Sig_2^{outer}(b,B',C,D,U) > Sig_2^{outer}(c,B',C,D,U)$ and Theorem 3.2, we have that

$$Sig_2^{outer}(b,B',C,D,U'_B) > Sig_2^{outer}(c,B',C,D,U'_B).$$

From the existing condition $b,c \in C - B' - B_2^{**}$ and Corollary 3.1, it is easy to obtain that $b,c \notin B_3^*$, where $B_3^* = \left\{a | Sig_3^{outer}(a,B,C,D,U) = 0, a \in C - B'.\right\}, B_2^{**} = \left\{a | Sig_2^{outer}(a,B',C,D,U) = 0, a \in C - B'.\right\}$. Thus, one has that $B' \cap B_3^* = \emptyset$ and

$$\begin{aligned} Sig_2^{outer}(b,B',C'_B,D,U'_B) &= H^{(U'_B,C'_B)}(D|B') - H^{(U'_B,C'_B)}(D|B' \cup \{b\}) \\ &= H^{(U'_B,C-B_3^*)}(D|B') - H^{(U'_B,C-B_3^*)}(D|B' \cup \{b\}) \\ &= H^{(U'_B,C)}(D|B') - H^{(U'_B,C)}(D|B' \cup \{b\}) \\ &= Sig_2^{outer}(b,B',C,D,U'_B). \end{aligned}$$

In similarity,

$$Sig_2^{outer}(c,B',C'_B,D,U'_B) = Sig_2^{outer}(c,B',C,D,U'_B).$$

Therefore, one has

$$Sig_2^{outer}(b,B',C'_B,D,U'_B) > Sig_2^{outer}(c,B',C'_B,D,U'_B).$$

(3) $\Delta = 3$

By the existing condition $Sig_3^{outer}(b,B',C,D,U) > Sig_3^{outer}(c,B',C,D,U)$ and Theorem 3.2, we have that

$$Sig_3^{outer}(b,B',C,D,U'_B) > Sig_3^{outer}(c,B',C,D,U'_B).$$

Furthermore, by means of $b,c \in C - B' - B_3^{**}$ and Corollary 3.1, it is easy to obtain that $b,c \notin B_3^*$, where $B_3^* = \left\{a | Sig_3^{outer}(a,B,C,D,U) = 0, a \in C - B'\right\}, B_3^{**} = \left\{a | Sig_1^{outer}(a,B',C,D,U) = 0, a \in C - B'.\right\}$. Therefore, we have that $B' \cap B_3^* = \emptyset$ and

$$\begin{aligned} &Sig_3^{outer}(b,B',C'_B,D,U'_B) \\ &= E^{(U'_B,C'_B)}(D|B') - E^{(U'_B,C'_B)}(D|B' \cup \{b\}) \\ &= E^{(U'_B,C-B_3^*)}(D|B') - E^{(U'_B,C-B_3^*)}(D|B' \cup \{b\}) \\ &= E^{(U'_B,C)}(D|B') - E^{(U'_B,C)}(D|B' \cup \{b\}) \\ &= Sig_3^{outer}(b,B',C,D,U'_B). \end{aligned}$$

In similarity,

$$Sig_3^{outer}(c,B',C'_B,D,U'_B) = Sig_3^{outer}(c,B',C,D,U'_B).$$

Therefore, one has

$$Sig_3^{outer}(b,B',C'_B,D,U'_B) > Sig_3^{outer}(c,B',C'_B,D,U'_B).$$

(4) $\Delta = 4$

By the existing condition $Sig_4^{outer}(b,B',C,D,U) > Sig_4^{outer}(c,B',C,D,U)$ and Theorem 3.2, we have that

$$Sig_4^{outer}(b,B',C,D,U'_B) > Sig_4^{outer}(c,B',C,D,U'_B).$$

According to $b,c \in C - B' - B_4^{**}$ and Corollary 3.1, it is easy to obtain that $b,c \notin B_3^*$, where $B_3^* = \left\{a | Sig_3^{outer}(a,B,C,D,U) = 0, a \in C - B'.\right\}$, $B_1^{**} = \left\{a | Sig_1^{outer}(a,B',C,D,U) = 0, a \in C - B'\right\}$. Therefore, we have $B' \cap B_3^* = \emptyset$ and

$$\begin{aligned} &Sig_4^{outer}(b,B',C'_B,D,U'_B) \\ &= CE^{(U'_B,C'_B)}(D|B') - CE^{(U'_B,C'_B)}(D|B' \cup \{b\}) \\ &= CE^{(U'_B,C-B_3^*)}(D|B') - CE^{(U'_B,C-B_3^*)}(D|B' \cup \{b\}) \\ &= CE^{(U'_B,C)}(D|B') - CE^{(U'_B,C)}(D|B' \cup \{b\}) \\ &= Sig_4^{outer}(b,B',C,D,U'_B). \end{aligned}$$

In similarity,

$$Sig_4^{outer}(c,B',C'_B,D,U'_B) = Sig_4^{outer}(c,B',C,D,U'_B).$$

Therefore, one has

$$Sig_4^{outer}(b,B',C'_B,D,U'_B) > Sig_4^{outer}(c,B',C'_B,D,U'_B).$$

In a word, for each of Δ, if $b,c \in C - B_\Delta^{**}$ and $Sig_\Delta^{outer}(b,B',C,D,U) > Sig_\Delta^{outer}(c,B',C,D,U)$, then $Sig_\Delta^{outer}(b,B',C'_B,D,U'_B) > Sig_\Delta^{outer}(c,B',C'_B,D,U'_B)$. □

From Theorem 3.4, we can see that the rank of attribute significance measures can be preserved while the attributes that are insignificant for complement conditional entropy are removed and the useless objects for computing reducts are simultaneously deleted. It should be pointed out that Theorem 3.4 provides the key theoretical foundation of the accelerating attribute reduction algorithms in the next section.

4. Accelerator for attribute reduction and experimental analysis

In this section, we first review the accelerator for attribute reduction proposed in [29]. Furthermore, by means of the rank preservation of significance measures in Section 3, we introduce a novel accelerator from the perspective of objects and attributes. In order to better show the efficiency and effectiveness of the proposed accelerator, a comparison experiment with the accelerator in [29] will be given.

4.1. Attribute reduction accelerator

In paper [29], an accelerator for attribute reduction (an accelerating reduction algorithm) was proposed through gradually removing useless objects for computing reducts within each iteration, which is described as follows.

Algorithm 2. Accelerator for attribute reduction from the perspective of objects (ACC1)

Input: Decision table $DT = (U, C \cup D)$;
Output: One reduct *red*.
Step 1: $red \leftarrow \emptyset$;//*red* is the pool to conserve the selected attributes
Step 2: Compute $Sig^{inner}(a_k,C,C,D,U)$, $k \leqslant |C|$;
Step 3: Put a_k into *red*, where $Sig^{inner}(a_k,C,C,D,U) > 0$;// These attributes form the core of the given decision table
Step 4: $i \leftarrow 1$ and $U_1 \leftarrow U$;
Step 5: While $EF^{(U_i,C)}(red,D) \neq EF^{(U_i,C)}(C,D)$,
Do {Compute the positive region $POS_{red}^{(U_i,C)}(D)$,
$U_{i+1} = U_i - POS_{red}^{(U_i,C)}(D)$,
$red \leftarrow red \cup \{a_0\}$, where
$Sig^{outer}(a_0,red,C,D,U_{i+1}) = \max\{Sig^{outer}(a_k,red,C,D,U_{i+1}),$
$a_k \in C - red\}$,
$i \leftarrow i+1$;
Step 6: return *red* and end.

where $EF^{(U_i,C)}(B,D) = EF^{(U_i,C)}(C,D)$ is the stopping criterion. For example, while the positive region is employed as the evaluation function, we have that $EF^{(U_i,C)}(B,D) = POS_B^{(U_i,C)}(D)$ and $EF^{(U_i,C)}(C,D) = POS_C^{(U_i,C)}(D)$.

Comparison with Algorithm 1, the same attribute reducts can be obtained by using Algorithm 2 (ACC1) while the computational time is significantly reduced. However, in Algorithm 2 (ACC1), attribute reduction is accelerated only from the perspective of objects, which limits its performance. In order to further improve the efficiency of attribute reduction algorithm, a novel accelerator will be presented in this paper, which is based on the principle that the rank of attribute significant measures are preserved while the insignificant attributes to the process of attribute reduction are removed and the useless objects for computing reducts are simultaneously deleted within each iteration. In the following, the description of this accelerator is shown.

Algorithm 3. Accelerator for attribute reduction from the perspective of objects and attributes (ACC2)

Input: Decision table $DT = (U, C \cup D)$;
Output: One reduct *red*.
Step 1: $red \leftarrow \emptyset$;//*red* is the pool to conserve the selected attributes
Step 2: Compute $Sig^{inner}(a_k, C, C, D, U), k \leqslant |C|$;
Step 3: Put a_k into *red*, where $Sig^{inner}(a_k, C, C, D, U) > 0$;// These attributes form the core of the given decision table
Step 4: $i \leftarrow 1, U_1 \leftarrow U, C_1 \leftarrow C$ and $C_{insig} \leftarrow \emptyset$;
Step 5: While $EF^{(U_i,C_i)}(red, D) \neq EF^{(U_i,C_i)}(C, D)$, Do {Compute the positive region $POS_{red}^{(U_i,C_i)}(D)$,
$U_{i+1} = U_i - POS_{red}^{(U_i,C_i)}(D)$,
$red \leftarrow red \cup \{a_0\}$, where $Sig_{\Delta}^{outer}(a_0, red, C_i, D, U_{i+1})$
$= \max\left\{Sig_{\Delta}^{outer}(a_k, red, C_i, D, U_{i+1}), a_k \in C_i - red\right\}$,
compute C_{insig}, where
$C_{insig} = \left\{a \middle| Sig_3^{outer}(a, red, C_i, D, U_{i+1}) = 0, a \in C_i\right\}$,
$C_{i+1} = C_i - C_{insig}$,
$i \leftarrow i + 1$;
Step 6: return *red* and end.

where $EF^{(U_i,C_i)}(B, D) = EF^{(U_i,C_i)}(C, D)$ is stopping criterion. For example, while the positive region is employed as the evaluation function, we have that $EF^{(U_i,C_i)}(B, D) = POS_B^{(U_i,C_i)}(D)$ and $EF^{(U_i,C_i)}(C, D) = POS_C^{(U_i,C_i)}(D)$.

It is obvious that the time complexity of Algorithm 3 (ACC2) is the same as Algorithm 2 (ACC1). However, because both the size of universe and the cardinality of attribute set become smaller and smaller in the process of attribute reduction, the proposed accelerator can further reduce the computational time. Furthermore, we summarize three factors of the new accelerator as follows.

(1) The computational time of significance measure of every attribute is further decreased;
(2) The time consuming of computing the stopping criterion is also significantly reduced;
(3) The same attribute reducts can be obtained using the proposed algorithm as the original algorithm.

4.2. The analysis of algorithms' efficiency

In this subsection, by means of the proposed accelerator, four representative heuristic algorithms that employ positive region, Shannon's conditional entropy, complement conditional entropy and combination conditional entropy as heuristic information are accelerated. For convenience, these accelerated algorithms are denoted as ACC2-PR, ACC2-SCE, ACC2-PCE and ACC2-CCE. Furthermore, we will compare the performance of four accelerating attribute reduction algorithms (ACC1-PR, ACC1-SCE, ACC1-PCE and ACC1-CCE) in [29] with these accelerated algorithms in this paper. In the experiment, 10 datasets from Table 1 are employed. They are all numerical, and have been preprocessed by discretization.

Table 1
Description of 10 UCI data sets.

	Data sets	Number of objects	Number of attributes	Number of classes
1	KDDcup10per	494,021	42	13
2	Gisette	13,500	5000	5
3	Ticdate2000	5822	85	2
4	Sat.tst	4435	35	6
5	Final-general	10,104	71	5
6	Arcene train	100	10,000	6
7	Mushroom	5644	22	2
8	Optdigits	3820	64	3
9	Waveform ± noise	5000	24	2
10	Connect	67,557	42	3

To display the new algorithms' efficiency, we compare the computational time and reducts of each original accelerating algorithms with the corresponding new one on the datasets in Table 1. These algorithms are run on a personal computer with Windows XP and Intel Core2 Quad CPU Q9400 and 3 GB Memory. The software being used is Microsoft Visual Studio 2005 and Visual C#.

4.2.1. ACC1-PR and ACC2-PR

Table 2 shows that comparison of ACC1-PR with ACC2-PR using the ten datasets in Table 1, in which the comparisons of running time and reducts of these two algorithms. From Table 2, we can see that the running time of ACC2-PR is less than ACC1-PR on nine of 10 datasets, and the same attribute subset can be selected running these two algorithms ACC1-PR and ACC2-PR, which is determined by the rank preservation of significance measures in Section 3. The results shows that ACC2-PR is significant.

Furthermore, we take the datasets Gisette and Waveform ± noise for examples to explain the reason why attribute reduction algorithms are accelerated using the proposed accelerator, as shown in Tables 3 and 4. The tables indicate the number of objects and attributes within each loop of ACC2-PR. From Table 3, we can see that the number of objects and the number of attributes are 1079 and 4839 in the second loop respectively, and the number of objects and attributes are 1024 and 4610 within the third loop respectively. It is obvious that a lot of insignificant attributes are deleted in these loops, while the size of universe is still large. Therefore, compared with ACC1-PR, the computational time in the two loops is significantly reduced, which results in the running time computing the reducts of ACC2-PR is less than ACC1-PR as the dataset Gisette. Nevertheless, ACC2-PR are not ever faster than ACC1-PR as all of the datasets. Table 4 shows this case. From Ta-

Table 2
The running time and reducts of Algorithms ACC1-PR and ACC2-PR.

Data sets	Original attributes	ACC1-PR		ACC2-PR	
		Reducts	Time (s)	Reducts	Time (s)
KDDcup10per	42	24	736.743	24	647.447
Gisette	5000	13	2268.281	13	2102.720
Ticdate2000	85	24	1.429	24	1.218
Sat.tst	35	26	0.197	26	0.190
Final-general	71	21	9.961	21	9.516
Arcene train	10,000	4	109.332	4	98.323
Mushroom	22	3	0.360	3	0.324
Optdigits	64	6	1.051	6	1.039
Waveform ± noise	24	14	3.251	14	3.252
Connect	42	34	128.649	34	116.876

Table 3
The changes of objects and attributes of Dataset Gisette in each iteration of Algorithm ACC2-PR.

Loop no.	Number of objects	Number of attributes	Number of insignificant attributes
1	2323	4999	0
2	1079	4839	159
3	1024	4610	228
4	959	4599	10
5	880	4590	8
6	768	4578	11
7	633	4536	41
8	478	4500	35
9	309	4432	67
10	158	4357	74
11	71	4147	209
12	25	3903	243

Table 4
The changes of objects and attributes of Dataset waveform ± noise in each iteration of Algorithm ACC2-PR.

Loop no.	Number of objects	Number of attributes	Number of insignificant attributes
1	4999	23	0
2	4994	22	0
3	4958	21	0
4	4835	20	0
5	4466	19	0
6	3591	18	0
7	2377	17	0
8	1167	16	0
9	482	15	0
10	165	14	0
11	110	13	0
12	55	12	0
13	40	11	0

Table 5
The time consuming and reducts of running Algorithms ACC1-SCE and ACC2-SCE.

Data sets	Original attributes	ACC1-SCE		ACC2-SCE	
		Reducts	Time (s)	Reducts	Time (s)
KDDcup10per	42	24	339.527	24	266.593
Gisette	5000	13	2018.804	13	1890.478
Ticdate2000	85	24	2.268	24	1.752
Sat.tst	35	26	0.197	26	0.190
Final-general	71	21	11.429	21	10.516
Arcene train	10,000	5	140.108	5	133.710
Mushroom	22	4	0.462	4	0.428
Optdigits	64	6	1.261	6	1.252
Waveform ± noise	24	14	3.451	14	3.453
Connect	42	34	187.343	34	185.434

ble 4, we can see that the number of insignificant attributes is zero within each loop of ACC2-PR. That is to say, for the dataset Waveform ± noise, ACC2-PR is not superior to ACC1-PR.

4.2.2. ACC1-SCE and ACC2-SCE

Table 5 shows the running time and reducts of ACC1-SCE and ACC2-SCE on the 10 datasets in Table 1. From Table 5, we can see that ACC2-SCE is faster than ACC1-SCE on nine of 10 datasets, and the attribute subset obtained by ACC2-SCE is the same as ACC1-SCE, which is determined by the rank preservation of significance measures in Section 3. The results shows that ACC2-SCE is efficient.

Furthermore, we take the datasets Gisette and Waveform ± noise for examples to explain the reason why attribute reduction algorithms are accelerated using the proposed accelerator, as shown in Tables 6 and 7. From Table 6, we can see that the number of objects and the number of attributes are 1079 and 4839 within the second iteration of ACC2-SCE, respectively, and the number of objects and attributes are 1060 and 4610 within its third iteration. It is obvious that compared with ACC1-SCE, the running time of ACC2-SCE is evidently saved within these two loops. That is because that the number of objects is still very large, while a lot of insignificant attributes are deleted from the dataset Gisette in the process of reduction. Nevertheless, ACC2-SCE is not more efficient than ACC1-SCE as all of the datasets. Table 7 shows this case. From Table 7, we can see that the number of insignificant attributes is zero within each loop. That is to say, as dataset Waveform ± noise, ACC2-SCE is not better than ACC1-SCE.

Table 6
The changes of objects and attributes of Dataset Gisette in each iteration of Algorithm ACC2-SCE.

Loop no.	Number of objects	Number of attributes	Number of insignificant attributes
1	2323	4999	0
2	1079	4839	159
3	1060	4610	228
4	1016	4607	2
5	833	4602	4
6	556	4576	25
7	300	4517	58
8	112	4383	133
9	36	4040	342
10	4	3531	508

Table 7
The changes of objects and attributes of Dataset waveform ± noise in each iteration of Algorithm ACC2-SCE.

Loop no.	Number of objects	Number of attributes	Number of insignificant attributes
1	4999	23	0
2	4998	22	0
3	4984	21	0
4	4830	20	0
5	4413	19	0
6	3433	18	0
7	2050	17	0
8	1167	16	0
9	598	15	0
10	358	14	0
11	259	13	0
12	233	12	0

4.2.3. ACC1-PCE and ACC2-PCE

Table 8 shows the running time and reducts of ACC1-PCE and ACC2-PCE on the 10 datasets in Table 1. From Table 8, we can

Table 8
The time consuming and reducts of Algorithms ACC1-PCE and ACC2-PCE.

Data sets	Original attributes	ACC1-PCE		ACC2-PCE	
		Reducts	Time (s)	Reducts	Time (s)
KDDcup10per	42	24	383.571	24	313.696
Gisette	5000	13	2018.804	13	1890.478
Ticdate2000	85	24	2.074	24	1.591
Sat. tst	35	26	0.198	26	0.190
Final-general	71	20	9.919	20	9.731
Arcene train	10,000	4	144.378	4	135.352
Mushroom	22	4	0.446	4	0.407
Optdigits	64	6	1.195	6	1.162
Waveform ± noise	24	13	3.427	13	3.429
Connect	42	34	202.452	34	191.318

Table 9
The changes of objects and attributes of Dataset Gisette in each iteration of Algorithm ACC2-PCE.

Loop no.	Number of objects	Number of attributes	Number of insignificant attributes
1	6000	4999	0
2	2283	4953	45
3	1060	4834	118
4	1013	4607	226
5	826	4601	5
6	592	4540	60
7	307	4485	54
8	129	4318	166
9	37	4098	219
10	5	3576	521

Table 10
The changes of objects and attributes of Dataset waveform ± noise within each iteration of Algorithm ACC2-PCE.

Loop no.	Number of objects	Number of attributes	Number of insignificant attributes
1	4999	23	0
2	4998	22	0
3	4974	21	0
4	4786	20	0
5	4257	19	0
6	3366	18	0
7	2104	17	0
8	1029	16	0
9	412	15	0
10	141	14	0
11	49	13	0
12	10	12	0

see that ACC2-PCE is faster than ACC1-PCE on nine of 10 datasets, and the reducts obtained by ACC2-PCE is the same as ACC1-PCE, which is determined by the rank preservation of significance measures in Section 3. The results shows that ACC2-PCE is significantly efficient.

Furthermore, we take the datasets Gisette and Waveform ± noise for examples to explain the reason why attribute reduction algorithms are accelerated using the proposed accelerator, as shown in Tables 9 and 10. From Table 9, we can find that the number of objects and attributes are 2283 and 4953 within the second iteration, and number of objects and attributes are 1060 and 4834 within the third iteration respectively. Because the number of objects is still very large and a lot of insignificant attributes are deleted from these datasets, the running time of ACC2-PCE in these two iterations is much less than ACC1-PCE. However, ACC2-SCE is not ever efficient for all datasets. Table 10 shows this case. From Table 10, we can see that the number of insignificant

Table 11
The time consuming and reducts of running Algorithms ACC1-CCE and ACC2-CCE.

Data sets	Original attributes	ACC1-CCE		ACC2-CCE	
		Reducts	Time (s)	Reducts	Time (s)
KDDcup10per	42	24	342.511	24	282.013
Gisette	5000	13	4095.627	13	3721.957
Ticdate2000	85	24	2.295	24	1.869
Sat.tst	35	26	0.197	26	0.190
Final-general	71	21	10.830	21	9.735
Arcene train	10,000	5	147.122	5	140.522
Mushroom	22	4	0.473	4	0.432
Optdigits	64	6	1.202	6	1.17x2
Waveform ± noise	24	13	3.455	13	3.458
Connect	42	34	206.137	34	199.657

Table 12
The changes of objects and attributes of Dataset Gisette in each iteration of Algorithm ACC2-CCE.

Loop no.	Number of objects	Number of attributes	Number of insignificant attributes
1	6000	4999	0
2	5999	4953	45
3	5980	4952	0
4	5823	4950	1
5	2186	4947	2
6	982	4815	131
7	803	4589	225
8	560	4542	46
9	304	4462	79
10	123	4320	141
11	40	4067	252
12	6	3669	397

Table 13
The changes of objects and attributes of Dataset waveform ± noise in each iteration of Algorithm ACC2-CCE.

Loop no.	Number of objects	Number of attributes	Number of insignificant attributes
1	4999	23	0
2	4998	22	0
3	4974	21	0
4	4786	20	0
5	4257	19	0
6	3366	18	0
7	2132	17	0
8	1029	16	0
9	423	15	0
10	141	14	0
11	50	13	0
12	10	12	0

attributes is zero within each iteration. That is to say, as dataset Waveform ± noise, ACC2-SCE is not better than ACC1-SCE.

4.2.4. ACC1-CCE and ACC2-CCE

Table 11 shows the running time and reducts of ACC1-CCE and ACC2-CCE on the 10 datasets in Table 1. From Table 11, we can see that ACC2-CCE is more timesaving than ACC1-CCE on nine of ten datasets, and the reducts obtained by ACC2-CCE is the same as ACC1-CCE, which is determined by the rank preservation of significance measures in Section 3. The results shows that ACC2-CCE is significantly efficient.

Furthermore, we take the datasets Gisette and Waveform ± noise for examples to explain the reason why attribute reduction algorithms ACC2-CCE are accelerated using the proposed accelerator, as shown in Tables 12 and 13. From Table 12, we can see that, as dataset Gisette, the number of objects and attributes are 5999 and 4953 within the second iteration respectively, and the size of universe is 982 and the dimension is 4815 within the sixth iteration. Therefore, the running time of the two iterations are significantly saved using ACC2-CCE. That is because that the size of universe is still very big, while numerous insignificant attributes are deleted from the dataset Gisette. However, ACC2-SCE is not ever efficient for all of the datasets. From Table 13, we can see that the number of insignificant attributes is zero within each iteration in the process of reduction on the dataset Waveform ± noise. Therefore, we does not save time of computing reduct using ACC2-CCE for the dataset Waveform ± noise. That is to say, as dataset Waveform ± noise, ACC2-CCE is not significantly superior to ACC1-CCE.

In conclusion, based on the experimental analysis, it should be stressed that the new accelerating attribute reduction algorithms (ACC2-PR, ACC2-SCE, ACC2-PCE and ACC2-CCE) are all more efficient than the original accelerating algorithms in most of datasets, except for the datasets in which there are few insignificant attributes.

5. Conclusions

A new accelerator for attribute reduction has been proposed in this paper. We first find that there exist some insignificant attributes in the process of computing reducts, and proof that the significance of each attribute remain the same after deleting these insignificant attributes. We present a general accelerator based on perspective of objects and attributes. Comparison with the existing accelerator, the new one can simultaneously decrease the size of universe and the number of attributes within each iteration of the process of attribute reduction, which is the key point of further accelerating attribute reduction. Finally, we introduce four representative heuristic algorithms embedded the new accelerator based on the positive region, Shannon's entropy and complement entropy. Experimental results show that the heuristic algorithms embedded the proposed accelerator can significantly reduce the computational time of attribute reduction.

Some future works are planed along the following directions. First, it would be interesting to investigate how our method can be extended to obtain attribute reducts from data with missing value and hybrid data. Second, since our current method requires continuous values of attribute be discretized, which motivate us to investigate how different discretization methods affect the performance of the proposed accelerator. Another direction is to extend our accelerator to the algorithms that deal with regression problems in which the class label is continuous values. Moreover, we will make effort to experiment our accelerated algorithms on genomic microarray data for effectively obtaining informative gene.

Acknowledgements

The research was supported by the National Natural Science Foundation of China (No. 71031006), Science and Technology Basic Condition Platform Construction Project of Shanxi Province (2012091002-0101), Key Project of Science and Technology of Shanxi Province (No. 20110321027-01), the Special Prophase Project for the National Key Basic Research and Development Program of China (973) (No. 2011CB311805), the National Key Basic Research and Development Program of China (973) (No. 2013CB329404) and the Natural Science Foundation of Shanxi Province (Nos. 2009021017-1, 2010021017-3).

References

[1] V.S. Babu, P. Viswanath, Rough-fuzzy weighted k-nearest leader classifier for large datasets, Pattern Recognition 42 (2009) 1719–1731.
[2] M. Dash, H. Liu, Consistency-based search in feature selection, Artificial Intelligence 151 (2003) 155–176.
[3] L. Feng, T.R. Li, D. Ruan, S.R. Gou, A vague-rough set approach for uncertain knowledge acquisition, Knowledge-Based Systems 24 (6) (2011) 837–843.
[4] I. Guyon, A. Elisseeff, An introduction to variable feature selection, Journal of Machine Learning Research 3 (2003) 1157–1182.
[5] Q.H. Hu, D.R. Yu, Z.X. Xie, Information-preserving hybrid data reduction based on fuzzy-rough techniques, Pattern Recognition Letters 27 (5) (2006) 414–423.
[6] Q.H. Hu, Z.X. Xie, D.R. Yu, Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation, Pattern Recognition 40 (2007) 3509–3521.
[7] Q.H. Hu, J.F. Liu, D.R. Yu, Mixed feature selection based on granulation and approximation, Knowledge-Based Systems 21 (4) (2008) 294–304.
[8] X.H. Hu, N. Cercone, Learning in relational databases: a rough set approach, International Journal of Computational Intelligence 11 (2) (1995) 323–338.
[9] R. Jensen, Q. Shen, Semantics-preserving dimensionality reduction: rough and fuzzy-rough-based approaches, IEEE Transactions on Knowledge and Data Engineering 16 (12) (2004) 1457–1471.
[10] M. Kryszkiewicz, P. Lasek, FUN: fast discovery of minimal sets of attributes functionally determining a decision attribute, Transactions on Rough Sets 9 (2008) 76–95.
[11] R. Kohavi, G.H. John, Wrappers for feature subset selection, Artificial Intelligence 97 (1–2) (1997) 273–324.
[12] C.K. Lee, G.G. Lee, Information gain and divergence-based feature selection for machine learning-based text categorization, Information Processing and Management 42 (2006) 155–165.
[13] J.Y. Liang, Z.B. Xu, The algorithm on knowledge reduction in incomplete information systems, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10 (1) (2002) 95–103.
[14] J.Y. Liang, K.S. Chin, C.Y. Dang, C.M. Yam Richid, A new method for measuring uncertainty and fuzziness in rough set theory, International of Journal General Systems 31 (4) (2002) 331–342.
[15] J.Y. Liang, Z.Z. Shi, D.Y. Li, M.J. Wierman, The information entropy, rough entropy and knowledge granulation in incomplete information systems, International Journal of General Systems 35 (6) (2006) 641–654.
[16] J.Y. Liang, J.H. Wang, Y.H. Qian, A new measure of uncertainty based on knowledge granulation for rough sets, Information Sciences 179 (4) (2009) 458–470.
[17] Z.Q. Meng, Z.Z. Shi, A fast approach to attribute reduction in incomplete decision systems with tolerance relation-based rough sets, Information Sciences 179 (2009) 2774–2793.
[18] M. Modrzejewski, Feature selection using rough set theory, Proceedings of European Conference on Machine Learning (1993) 213–226.
[19] N.M. Parthaláin, Q. Shen, Exploring the boundary region of tolerance rough sets for feature selection, Pattern Recognition 42 (2009) 655–667.
[20] T. Pavlenko, On feature selection, curse-of-dimensionality and error probability in discriminant analysis, Journal of Statistical Planning and Inference 115 (2003) 565–584.
[21] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Boston, 1991.
[22] Z. Pawlak, A. Skowron, Rudiments of rough sets, Information Sciences 177 (1) (2007) 3–27.
[23] Z. Pawlak, A. Skowron, Rough sets and boolean reasoning, Information Sciences 177 (1) (2007) 41–73.
[24] W. Pedrycz, G. Vukovich, Feature analysis through information granulation and fuzzy sets, Pattern Recognition 35 (2002) 825–834.
[25] H. Peng, F. Long, C. Ding, Features selection based on mutual information: criteria ofmax-dependency, max-relevance, and min-redundance, IEEE transactions on pattern analysis machine intellegence 27 (8) (2005) 1226–1238.
[26] Y.H. Qian, J.Y. Liang, W. Pedrycz, C.Y. Dang, An efficient accelerator for attribute reduction from incomplete data in rough set framework, Pattern Recognition 44 (2011) 1658–1670.
[27] Y.H. Qian, J.Y. Liang, D.Y. Li, H.Y. Zhang, C.Y. Dang, Measures for evaluating the decision performance of a decision table in rough set theory, Information Sciences 178 (2008) 181–202.
[28] Y.H. Qian, J.Y. Liang, Combination entropy and combination granulation in rough set theory, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 16 (2) (2008) 179–193.
[29] Y.H. Qian, J.Y. Liang, W. Pedrycz, C.Y. Dang, Positive approximation: an accelerator for attribute reduction in rough set theory, Artificial Intelligences 174 (9-10) (2010) 597–618.
[30] Y. Saeys, I. Inza, P. Larrañaga, A review of feature selection techniques in bioinformatics, Bioinformatics 23 (19) (2007) 2507–2517.
[31] Q. Shen, A. Chouchoulas, A rough-fuzzy approach for generating classification rules, Pattern Recognition 35 (2002) 2425–2438.
[32] Q. Shen, R. Jensen, Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring, Pattern Recognition 37 (2004) 1351–1363.
[33] A. Skowron, C. Rauszer, The discernibility matrices and functions in information tables, Intelligent Decision Support: Handbook of Applications and Advances of Rough Set Theory (1992) 331–362.
[34] D. Slezak, Approximate entropy reducts, Fundamenta Informaticae 53 (3–4) (2002) 365–390.
[35] D. Slezak, Approximate Reducts in Decision Tables, Research report, Institute of Computer Science, Warsaw University of Technology, 1995.
[36] R.W. Swiniarski, A. Skowron, Rough set methods in feature selection and recognition, Pattern Recognition Letters 24 (2003) 833–849.
[37] E.C.C. Tsang, D.G. Chen, D.S. Yeung, X.Z. Wang, J. Lee, Attributes Reduction using fuzzy rough sets, IEEE Transactions on Fuzzy Systems 16 (5) (2008) 1130–1141.
[38] G.Y. Wang, H. Yu, D.C. Yang, Decision table reduction based on conditional information entropy, Chinese Journal of Computer 25 (7) (2002) 759–766.
[39] G.Y. Wang, J. Zhao, J.J. An, A comparative study of algebra viewpoint and information viewpoint in attribute reduction, Fundamenta Informaticae 68 (3) (2005) 289–301.
[40] W. Wei, J.Y. Liang, Y.H. Qian, F. Wang, An attribute reduction approach and its accelerated version for hybrid data, in: The 8th IEEE International Conference on Cognitive Informatics, 2009, pp. 167–173.
[41] W. Wei, J.Y. Liang, Y.H. Qian, F. Wang, C.Y. Dang, Comparative study of decision performance of decision tables induced by attribute reductions, International Journal of General Systems 39 (8) (2010) 813–838.

[42] W. Wei, J.Y. Liang, Y.H. Qian, A comparative study of rough sets for hybrid data, Information Sciences (2011), doi:http://dx.doi.org/10.1016/j.ins.2011.12.006.
[43] S.K.M. Wong, W. Ziako, On optimal decision rules in decision tables, Bulletin of Polish Academy of Sciences 33 (1985) 693–696.
[44] X.B. Yang, M. Zhang, H.L. Dou, J.Y. Yang, Neighborhood systems-based rough sets in incomplete information system, Knowledge-Based Systems 24 (2011) 858–867.
[45] Y.Y. Yao, Y. Zhao, Attribute reduction in decision-theoretic rough set models, Information Sciences 178 (17) (2008) 3356–3373.
[46] Y.Y. Yao, Y. Zhao, Discernility matrix simplification for constructing attribute reducts, Information sciences 179 (2009) 867–882.
[47] W. Ziarko, Variable precision rough set model, Journal of Computer and System Sciences 46 (1993) 39–59.