Research Article

Positive Macroscopic Approximation for Fast Attribute Reduction

Zheng-Cai Lu,[1] Zheng Qin,[1,2] Qiao Jing,[1] and Lai-Xiang Shan[1]

[1] *Department of Computer Science & Technology, Tsinghua University, Beijing 100084, China*
[2] *School of Software, Tsinghua University, Beijing 100084, China*

Correspondence should be addressed to Zheng-Cai Lu; luzc09@mails.tsinghua.edu.cn

Attribute reduction is one of the challenging problems facing the effective application of computational intelligence technology for artificial intelligence. Its task is to eliminate dispensable attributes and search for a feature subset that possesses the same classification capacity as that of the original attribute set. To accomplish efficient attribute reduction, many heuristic search algorithms have been developed. Most of them are based on the model that the approximation of all the target concepts associated with a decision system is dividable into that of a single target concept represented by a pair of definable concepts known as lower and upper approximations. This paper proposes a novel model called macroscopic approximation, considering all the target concepts as an indivisible whole to be approximated by rough set boundary region derived from inconsistent tolerance blocks, as well as an efficient approximation framework called positive macroscopic approximation (PMA), addressing macroscopic approximations with respect to a series of attribute subsets. Based on PMA, a fast heuristic search algorithm for attribute reduction in incomplete decision systems is designed and achieves obviously better computational efficiency than other available algorithms, which is also demonstrated by the experimental results.

1. Introduction

Rough set theory (RST) [1] is a powerful mathematical tool for dealing with imprecision, uncertainty, and vagueness. As an extension of traditional set theory supporting approximation in decision making, RST provides a well-established model that the approximation of an indefinable target concept is represented by a pair of definable concepts known as lower and upper approximations. In recent years, more and more attention has been paid to RST, and much success of its applications has already covered a variety of fields such as artificial intelligence, machine learning, and knowledge discovery [2–6].

Attribute reduction is one of the key topics in RST, viewed as the strongest and most important result to distinguish RST from other theories [7]. Its task is just to eliminate reducible or dispensable attributes and search for a feature subset with the same classification capacity as that of the original attribute set. Much use has been made of attribute reduction as a preprocessing stage prior to classification of decision systems, making analysis algorithms more efficient and learned classifiers more compact.

In attribute reduction, we encounter four general search strategies. The most intuitive one is the exhaustive search, which checks all the possible candidate subsets and retrieves those that satisfy the given criteria. The exhaustive search results in high time complexity and has been proved to be an NP-hard problem [8]. Another alternative way characterized by the incomplete search applies mapping and pruning techniques to minimization. It is achieved by mapping pertinent elements to a structured model and pruning useless branches in the search space [9, 10]. Similar to the exhaustive search, the incomplete search finds the minimal reduct at the expense of great computational effort. The third strategy for attribute reduction conducts a random search using the techniques such as genetic algorithm [11], ant colony optimization [12], and particle swarm optimization [13]. The random search provides a robust solution but is also computationally very expensive. The fourth and most practical strategy discovers feature subsets by the heuristic search, where attributes with

high quality are preferred as heuristics according to an evaluation function [14–18]. The heuristic search has the ability to seek out an optimal or suboptimal reduct as well as acceptable computational complexity, playing an important role in the attribute reduction community.

To accomplish efficient attribute reduction, much work has been devoted to developing heuristic search algorithms. From the viewpoint of evaluation functions, they can be classified into three main categories: positive region-based reduction [19–25], combination-based reduction [26–30], and entropy-based reduction [31–33].

The positive region-based reduction takes the change of rough set positive region caused by the addition of an attribute as the significance of the attribute. Attributes with the highest significance are selected as heuristics to guide the search process. One of the classic instances is the quick reduct algorithm proposed by Chouchoulas and Shen [19], which can pick the best path to a reduct from the whole search space and receive many improved versions [20–22]. By using decomposition and sorting techniques to calculate positive region, Meng and Shi [23] put forward a fast positive region-based algorithm for feature selection from incomplete decision systems. Qian et al. [24, 25] constructed an efficient accelerator for heuristic search using a series of positive regions to approximate a given target decision on the gradually reduced universe. It can be incorporated into heuristic attribute reduction algorithms and make the modified versions capable of greatly reducing computational time.

Unlike the positive region-based reduction, the combination-based reduction considers positive region as well as other available information such as rule support and boundary region. An attribute is evaluated by combined measure generated by positive region and additional information. With consideration of the overall quality of the potential set of rules, Zhang and Yao [27] introduced rule support into the evaluation function and proposed a support-oriented algorithm called parameterized average support heuristic (PASH), which selects features causing high average support of rule over all decision classes. Parthaláin and Shen [28] used distance metric to qualify the objects in the boundary region with regard to their proximity to the lower approximation and presented the distance metric-assisted tolerance rough set attribute reduction algorithm, which employs a new evaluation measure created by combining the distance metric and the dependency degree.

Different from the above two categories, the entropy-based reduction gives the evaluation functions from the information view, such as combination entropy and rough entropy. Qian and Liang [32] presented the concept of combination entropy for describing the uncertainty of information systems and used its condition entropy to select a feature subset. Sun et al. [33] utilized rough entropy-based uncertainty measures to evaluate the roughness and accuracy of knowledge and then constructed a heuristic search algorithm with low computational complexity for attribute reduction in incomplete decision systems.

These investigations have offered interesting insights into attribute reduction. When dealing with large incomplete data,

however, they still suffer from computational inefficiency. A more efficient and feasible attribute reduction approach is really desirable. This paper just intends to provide such a solution.

One can observe that most of heuristic attribute reduction algorithms are based on the model that the approximation of all the target concepts from a decision system is dividable into that of a single target concept represented by lower and upper approximations. Little work has hitherto taken the approximation problem into account at the macroscopic level. In this paper, we propose a novel model called macroscopic approximation, considering all the target concepts of a decision system as an indivisible whole to be approximated by rough set boundary region derived from inconsistent blocks, as well as an efficient approximation framework called positive macroscopic approximation (PMA), addressing macroscopic approximations with respect to a series of attribute subsets. Based on PMA, a fast heuristic attribute reduction algorithm for incomplete decision systems is designed and achieves obviously better computational efficiency than other available algorithms, which is also demonstrated by the experimental results.

The remainder of this paper is organized as follows. In Section 2, we review some basic concepts of RST and outline the quick reduct algorithm. Section 3 investigates macroscopic approximation and positive macroscopic approximation. In Section 4, a fast heuristic attribute reduction algorithm based on positive macroscopic approximation is devised and illustrated by a worked example. In Section 5, some experiments are practiced to validate the time efficiency of the proposed algorithm. Finally, we give a concise conclusion in Section 6.

2. Preliminaries

In this section, we briefly recall some basic concepts, such as incomplete decision system, tolerance relation, tolerance block, tolerance class, positive region, and boundary region, together with the quick reduct algorithm, needed in the following sections.

2.1. Basic Concepts. A decision system is an information system with distinction between decision attributes and condition attributes. It is generally formulated by a data table where columns are referred to as attributes and rows as objects of interest. If there exist objects that contain missing data, the decision system is incomplete; otherwise, it is complete.

An incomplete decision system is a tuple $S = (U, A)$, where U, called the universe of discourse, is a nonempty finite object set and A is an attribute set that consists of a condition attribute subset C and a decision attribute subset D. For any $a \in A$, there is a mapping f, $f : U \rightarrow V_a$, where V_a is the value domain of a. $V_C = \{V_a \mid a \in C\} \cup \{*\}$ (where "$*$" stands for missing values) and $V_D = \{V_d \mid d \in D\}$ are the value domains of C and D, respectively. For convenience, the tuple $S = (U, A)$ is usually denoted as $S = (U, C \cup D)$.

Let $S = (U, C \cup D)$ be an incomplete decision system. For any subset $P \subseteq C$, P determines a binary relation, denoted by $\text{SIM}(P)$, which is defined as

$$\text{SIM}(P) = \{(u, v) \in U \times U \mid \forall a \in P, f(u, a)$$
$$= f(v, a) \text{ or } f(u, a) = * \text{ or } f(v, a) = *\}. \tag{1}$$

It is easily known that the binary relation is reflexive, symmetric, and intransitive, so it is a tolerance relation. Using P-tolerance relation, the universe can be divided into many tolerance blocks such that $\pi_P = \{X \mid X \in U/\text{SIM}(P)\}$, where X is a P-tolerance block and π_P is the family of all the P-tolerance blocks on U, called a P-approximation space. A P-tolerance block depicts the collection of objects which are possibly indiscernible from each other with respect to P. If there does not exist another P-tolerance block Y such that $X \subset Y$, then X is called a maximal P-tolerance block [34].

For any object $u \in U$, P-tolerance relation determines the tolerance class of u, denoted by $S_P(u)$, that is, $S_P(u) = \{v \in U \mid (u, v) \in \text{SIM}(P)\}$, describing the maximal set of objects which are probably indistinguishable to u with respect to P. There is a relationship between the tolerance class and the maximal tolerance block, shown as follows [34]:

$$S_P(u) = \bigcup_{X \in \pi_{mP}(u)} X, \tag{2}$$

where $\pi_{mP}(u)$ is the family of maximal P-tolerance blocks containing u. Then, it is easy to prove that

$$S_P(u) = \bigcup_{X \in \pi_P(u)} X, \tag{3}$$

where $\pi_P(u)$ is the family of P-tolerance blocks containing u.

Consider a partition $\pi_D = \{D_i \mid i = 1, 2, \ldots, j\}$ of U determined by D. π_D is the family of all the decision classes derived from the decision system. Each decision class can be viewed as a target concept approximated by a pair of precise concepts which are known as lower and upper approximations. The dual approximations of a target concept D_i are defined, respectively, as

$$\underline{P}(D_i) = \{u \in U \mid S_P(u) \subseteq D_i\},$$
$$\overline{P}(D_i) = \{u \in U \mid S_P(u) \cap D_i \neq \emptyset\}. \tag{4}$$

The low approximation of D_i is regarded as the maximal P-definable set contained in D_i, whereas the upper approximation of D_i is considered as the minimal P-definable set containing D_i. If $\underline{P}(D_i) = \overline{P}(D_i)$, then D_i is a P-exact set; otherwise, it is a P-rough set.

By the dual approximations, the universe of the decision system is partitioned into two mutually exclusive crisp regions: positive region and boundary region, defined, respectively, as

$$\text{POS}_P(\pi_D) = \bigcup_{i=1}^{j} \underline{P}(D_i), \tag{5}$$

$$\text{BND}_P(\pi_D) = \bigcup_{i=1}^{j} \left(\overline{P}(D_i) - \underline{P}(D_i)\right). \tag{6}$$

It can be perceived that the positive region is the collection of objects which are classified without any ambiguity into the target concepts using P-tolerance relation, while the boundary region is, in a sense, the undeterminable area of the universe, where none of the objects are classified with certainty into the target concepts as far as P is concerned. It is apparent that $\text{POS}_P(\pi_D) \cup \text{BND}_P(\pi_D) = U$ and $\text{POS}_P(\pi_D) \cap \text{BND}_P(\pi_D) = \emptyset$. If $\text{BND}_P(\pi_D) = \emptyset$ or $\text{POS}_P(\pi_D) = U$, we say that S is consistent; otherwise, it is inconsistent.

2.2. Rough Set Attribute Reduction. In RST environment, dependency degree of decision attribute set D to condition attribute set P ($P \subseteq C$) is definable in terms of positive region:

$$\gamma_P(D) = \frac{|\text{POS}_P(\pi_D)|}{|U|}, \tag{7}$$

where $|\text{POS}_P(\pi_D)|$ and $|U|$ are the cardinalities of $\text{POS}_P(\pi_D)$ and U, respectively. If $\gamma_P(D) = 1$, we say that D totally depends on P. If $\gamma_P(D) < 1$, we say that D partially depends on P. If $\gamma_P(D) = 0$, we say that D is completely independent of P. RST describes a variation of dependency degree caused by the addition of attribute a to P as the significance of a such that

$$\text{Sig}(a, P, D) = \gamma_{P \cup \{a\}}(D) - \gamma_P(D). \tag{8}$$

The bigger the significance value is, the more informative the attribute is. Accordingly, the quick reduct algorithm [19], regarded as a classical rough set attribute reduction algorithm, is constructed by iteratively adding the attribute with the highest significance to an attribute pool which begins with an empty set until the dependency value of the pool equals that of the set of the whole condition attributes. This process can be outlined by Algorithm 1.

3. Positive Macroscopic Approximation

It is well known that approximation is one of the core ideas of RST. A target concept is approximated by the low and upper approximations. Likewise, a decision system can be viewed as a super concept to be approximated by the low and upper super approximations, where the family of all the target concepts from the decision system is considered as the super concept, the positive region or the complementary set of the boundary region of the decision system acts as the low super approximation, and the universe of the decision system serves as the upper super approximation. If the low super approximation equals the upper super approximation,

```
Inputs:
U: the universe
C: C = {c_1, c_2, ..., c_n}, a condition attribute set
D: D = {d}, a decision attribute set
Output:
red : a reduct of C
Begin
Step 1. Let red= ∅ and R = C
Step 2. While γ_red(D) ≠ γ_C(D) and R ≠ ∅ do
            Select an attribute a ∈ R with Sig(a, red, D) = max(Sig(a, red, D)) where a_i ∈ R
            R = R − {a}
            red = red ∪ {a}
Step 3. Output red
End.
```

ALGORITHM 1: Quick reduct.

the decision system is consistent or exact; otherwise, it is inconsistent or rough. From the observation, it is easily understood that the approximation of a decision system can be represented by the positive region or the boundary region.

RST offers a feasible way to obtain the approximation of a decision system by means of that of a single target concept represented by lower and upper approximations as stated by (5) and (6). For convenience, this model is called microcosmic approximation. As opposed to microcosmic approximation, this section introduces a novel model called macroscopic approximation, where the approximation of a decision system is achieved by regarding all the target concepts as an inseparable entity to be approximated by the boundary region. Furthermore, we explore positive macroscopic approximation (PMA), which considers macroscopic approximations with respect to a series of attribute sets.

3.1. Macroscopic Approximation. Macrocosmic approximation is an alternative way to arrive at the approximation of a decision system by the boundary region. Due to an integral consideration of all the target concepts associated with a decision system, the low and upper approximations are unavailable. Affirmably, an attempt is deserved to pioneer a new avenue to macroscopic approximation. An inconsistent tolerance block, introduced in the following context, is capable of offering such a feasible solution.

Definition 1. Let $S = (U, C \cup D)$, $P \subseteq C$, and $X \in \pi_P$. Then, X is said an inconsistent tolerance block (IT-block) if $|\lambda(X)| > 1$ or a consistent tolerance block (CT-block), where $\lambda(X) = \{f(u, d) \mid u \in X\}$ and $|\lambda(X)|$ is the cardinality of $\lambda(X)$.

An IT-block describes a set of P-definable objects with diverse class labels, implying that a group of P-indistinguishable objects have the divergence of decision making, whereas a CT-block depicts a collection of P-definable objects with the same class label, indicating that a group of P-indiscernible objects share the same decision making. Accordingly, π_P are classified into two mutually exclusive crisp subfamilies. One is the consistent family,

denoted by π_P^{CT}, collecting all the CT-blocks from π_P such that $\pi_P^{CT} = \{Y \in \pi_P \mid |\lambda(Y)| = 1\}$. The other is the inconsistent family, denoted by π_P^{IT}, gathering all the IT-blocks from π_P such that $\pi_P^{IT} = \{Y \in \pi_P \mid |\lambda(Y)| > 1\}$. Obviously, $\pi_P^{CT} \cup \pi_P^{IT} = \pi_P$ and $\pi_P^{CT} \cap \pi_P^{IT} = \{\emptyset\}$.

It is worth noting the distinction between the boundary region and the IT-block. The former consists of objects of which tolerance classes cannot be entirely contained in target concepts, while the latter is such an entity that overlaps two or more target concepts. The following lemmas are used to investigate the relationship between them.

Lemma 2. *Let $S = (U, C \cup D)$ and $P \subseteq C$. For any $u \in BND_P(\pi_D)$, there must exist at least one IT-block containing u.*

Proof. This proof is done by contradiction. For any $u \in BND_P(\pi_D)$, suppose that there does not exist any IT-block containing u. That is to say, any $X \in \pi_P(u)$ is a CT-block, which implies that there must exist some decision value d_0 ($d_0 \in V_D$) such that $\lambda(X) = d_0$. d_0 corresponds to a certain decision class D_0 ($D_0 \in \pi_D$) containing all the objects whose decision values are equal to d_0, and then $X \subseteq D_0$ and $\bigcup_{X \in \pi_P(u)} X \subseteq D_0$. So, $S_P(u) = \bigcup_{X \in \pi_P(u)} X \subseteq D_0$. This means that $u \in POS_P(\pi_D)$, contradicted with $u \in BND_P(\pi_D)$. Hence, for any $u \in BND_P(\pi_D)$, there must exist at least one IT-block containing u. The lemma holds. $\square$

Lemma 3. *Let $S = (U, C \cup D)$ and $P \subseteq C$. For any $X \in \pi_P^{IT}$, $X \subseteq BND_P(\pi_D)$.*

Proof. For any $u \in X$, we have $S_P(u) = S'_P(u) \cup X$, where $S'_P(u) = \bigcup_{X' \in \pi_P(u)} X'$. Since $X \in \pi_P^{IT}$, there is $X \not\subseteq D_0$ for any $D_0 \in \pi_D$ (or if $X \subseteq D_0$, then $|\lambda(X)| = |\lambda(D_0)| = 1$, and thereby $X \in \pi_P^{CT}$, contradicted with $X \in \pi_P^{IT}$). Hence, $S_P(u) = S'_P(u) \cup X \not\subseteq D_0$, which gives $u \notin POS_P(\pi_D)$. As $POS_P(\pi_D) \cup BND_P(\pi_D) = U$ and $POS_P(\pi_D) \cap BND_P(\pi_D) = \emptyset$ in S, $u \notin POS_P(\pi_D)$ implies that $u \in BND_P(\pi_D)$. Therefore, for any $X \in \pi_P^{IT}$ and $u \in X$, there is $u \in BND_P(\pi_D)$. This

means that $X \subseteq \mathrm{BND}_P(\pi_D)$ for any $X \in \pi_P^{\mathrm{IT}}$. The lemma holds. $\qquad\square$

Theorem 4. *Let* $S = (U, C \cup D)$ *and* $P \subseteq C$. *Then,* $\mathrm{BND}_P(\pi_D) = \bigcup_{X \in \pi_P^{\mathrm{IT}}} X$.

Proof. For any $u \in \mathrm{BND}_P(\pi_D)$, according to Lemma 2, there must exist at least one tolerance block X containing u such that $X \in \pi_P^{\mathrm{IT}}$, which yields $\mathrm{BND}_P(\pi_D) \subseteq \bigcup_{X \in \pi_P^{\mathrm{IT}}} X$. On the other hand, for any $X \in \pi_P^{\mathrm{IT}}$, by Lemma 3, we have $X \subseteq \mathrm{BND}_P(\pi_D)$, which gives $\bigcup_{X \in \pi_P^{\mathrm{IT}}} X \subseteq \mathrm{BND}_P(\pi_D)$. In summary, $\mathrm{BND}_P(\pi_D) = \bigcup_{X \in \pi_P^{\mathrm{IT}}} X$. The theorem holds. $\qquad\square$

Theorem 4 reveals the relationship that the boundary region is the union of IT-blocks, which is in essence the materialization of macroscopic approximation by integrally approximating all the target concepts using the boundary region derived from IT-blocks. Unlike the microscopic approximation stated by (6), the macroscopic approximation is directly constructed on the elementary members of the approximation space rather than the low and upper approximations, making the calculation of the approximation of the decision system more efficient.

3.2. PMA.
For a given decision system, we can build a series of attribute subsets according to the following rule. The first set has only one attribute selected from the set of the whole attributes, the union of the first set and another attribute chosen from the remaining attributes is regarded as the second set, and so on. The newly generated attribute subsets form an ascending order of the sequence, called a positive sequence. If a sequence is organized by a descending order, it is called a converse sequence. Assume that attribute subsets A and B are two arbitrary elements of a positive sequence. If A contains B, we say that the tolerance relation determined by A is finer than that determined by B. Conversely, if A is contained in B, we say that the tolerance relation determined by A is coarser than that determined by B. Accordingly, a positive sequence determines a train of tolerance relations stretching from coarse to fine.

In this subsection, we explore positive macroscopic approximation (PMA), which addresses macroscopic approximations with respect to a positive sequence. To construct an efficient PMA, the following relevant definitions and lemmas are needed.

Definition 5. Let $S = (U, C \cup D)$, $P \subseteq C$, $a \in C - P$, and $X \in \pi_P$. Then, the family of subblocks determined by a on X is defined as

$$\theta_a(X) = \{\{u \in X \mid f(u, a) = b \text{ or } f(u, a) = *\} \mid b \in V_a\}. \tag{9}$$

The consistent and inconsistent subfamilies of $\theta_a(X)$ are defined as $\theta_a^{\mathrm{CT}}(X) = \{Y \in \theta_a(X) \mid |\lambda(Y)| = 1\}$ and $\theta_a^{\mathrm{IT}}(X) = \{Y \in \theta_a(X) \mid |\lambda(Y)| > 1\}$, respectively. Apparently, $\theta_a^{\mathrm{CT}}(X) \cup \theta_a^{\mathrm{IT}}(X) = \theta_a(X)$ and $\theta_a^{\mathrm{CT}}(X) \cap \theta_a^{\mathrm{IT}}(X) = \{\emptyset\}$.

Lemma 6. *Let* $S = (U, C \cup D)$, $P \subseteq C$, $a \in C - P$, *and* X *a CT-block. Then, any* $Y \in \theta_a(X)$ *is a CT-block and* $\theta_a^{\mathrm{CT}}(X) = \theta_a(X)$, $\theta_a^{\mathrm{IT}}(X) = \{\emptyset\}$.

Proof. Since X is a CT-block, $|\lambda(X)| = 1$. Suppose that $\lambda(X) = d_0$ ($d_0 \in V_d$), then $f(u, d) = d_0$ for any $u \in X$. $Y \in \theta_a(X)$ means that $Y \subseteq X$, so we have $v \in X$ for any $v \in Y$ and $f(v, d) = d_0$. Hence, $\lambda(Y) = d_0$ and $|\lambda(Y)| = 1$, which yields that Y is a CT-block. Moreover, $\theta_a^{\mathrm{CT}}(X) = \{Y \in \theta_a(X) \mid |\lambda(Y)| = 1\} = \theta_a(X)$ and $\theta_a^{\mathrm{IT}}(X) = \{Y \in \theta_a(X) \mid |\lambda(Y)| > 1\} = \{\emptyset\}$. The lemma holds. $\qquad\square$

This lemma shows that any subblock derived from a CT-block is also a CT-block. In other words, inconsistent tolerance subblocks are derivable only from IT-blocks.

Definition 7. Let $S = (U, C \cup D)$, $P \subseteq C$, $a \in C - P$, and W a family of P-tolerance blocks. Then, the family of subblocks determined by a on W is defined as

$$\omega_a(W) = \{\theta_a(X) \mid X \in W\}. \tag{10}$$

The consistent and inconsistent subfamilies of $\omega_a(W)$ are defined as $\omega_a^{\mathrm{CT}}(W) = \{Y \in \omega_a(W) \mid |\lambda(Y)| = 1\}$ and $\omega_a^{\mathrm{IT}}(W) = \{Y \in \omega_a(W) \mid |\lambda(Y)| > 1\}$, respectively. Clearly, $\omega_a^{\mathrm{CT}}(W) \cup \omega_a^{\mathrm{IT}}(W) = \omega_a(W)$ and $\omega_a^{\mathrm{CT}}(W) \cap \omega_a^{\mathrm{IT}}(W) = \{\emptyset\}$.

Lemma 8. *Let* $S = (U, C \cup D)$, $P \subseteq C$, $a \in C - P$, *and* W *a family of CT-blocks. Then, any* $Y \in \omega_a(W)$ *is a CT-block and* $\omega_a^{\mathrm{CT}}(W) = \omega_a(W)$, $\omega_a^{\mathrm{IT}}(W) = \{\emptyset\}$.

Proof. For any $X \in W$, X is a CT-block. By Lemma 6, any $Z \in \theta_a(X)$ is also a CT-block, which yields that any $Y \in \{\theta_a(X) \mid X \in W\} = \omega_a(W)$ is a CT-block. Since $|\lambda(Y)| = 1$, $\omega_a^{\mathrm{CT}}(W) = \{Y \in \omega_a(W) \mid |\lambda(Y)| = 1\} = \omega_a(W)$ and $\omega_a^{\mathrm{IT}}(W) = \{Y \in \omega_a(W) \mid |\lambda(Y)| > 1\} = \{\emptyset\}$. The lemma holds. $\qquad\square$

Theorem 9. *Let* $S = (U, C \cup D)$, $P \subseteq C$, *and* $a \in C - P$. *Then,* $\pi_{P \cup \{a\}}^{\mathrm{IT}} = \omega_a^{\mathrm{IT}}(\pi_P^{\mathrm{IT}})$.

Proof. Assume that $P = \{a_1, a_2, \ldots, a_t\}$. By (9) and (10), we have $\pi_P = \omega_{a_t}(\cdots(\omega_{a_2}(\theta_{a_1}(U)))\cdots)$ and $\pi_{P \cup \{a\}} = \omega_a(\omega_{a_t}(\cdots(\omega_{a_2}(\theta_{a_1}(U)))\cdots))$. Then, $\pi_{P \cup \{a\}} = \omega_a(\pi_P) = \omega_a(\pi_P^{\mathrm{IT}}) \cup \omega_a(\pi_P^{\mathrm{CT}})$.

$$
\begin{aligned}
\pi_{P \cup \{a\}}^{\mathrm{IT}} &= \{X \in \pi_{P \cup \{a\}} \mid |\lambda(X)| > 1\} \\
&= \{X \in \omega_a(\pi_P^{\mathrm{IT}}) \cup \omega_a(\pi_P^{\mathrm{CT}}) \mid |\lambda(X)| > 1\} \\
&= \{X \in \omega_a(\pi_P^{\mathrm{IT}}) \mid |\lambda(X)| > 1\} \\
&\quad \cup \{X \in \omega_a(\pi_P^{\mathrm{CT}}) \mid |\lambda(X)| > 1\} \\
&= \omega_a^{\mathrm{IT}}(\pi_P^{\mathrm{IT}}) \cup \omega_a^{\mathrm{IT}}(\pi_P^{\mathrm{CT}}).
\end{aligned}
\tag{11}
$$

By Lemma 8, $\omega_a^{\mathrm{IT}}(\pi_P^{\mathrm{CT}}) = \{\emptyset\}$, and then $\pi_{P \cup \{a\}}^{\mathrm{IT}} = \omega_a^{\mathrm{IT}}(\pi_P^{\mathrm{IT}})$. The theorem holds. $\qquad\square$

Inputs:
U: the universe
P: $P = \{a_1, a_2, \ldots, a_t\} \subseteq C$, a condition attribute subset
D: $D = \{d\}$, a decision attribute set
Output:
PMA(ρ_t): a family of boundary regions
Begin
Sept 1. PMA$(\rho_t) = \langle \emptyset \rangle$, $P_0 = \emptyset$, $\pi_{P_0}^{\text{IT}} = \{U\}$ // regard the universe as a root IT-block
Step 2. For $i = 1$ to t do
$\qquad\qquad P_i = P_{i-1} \cup \{a_i\}$ $\qquad\qquad\qquad$ // generate an element P_i of a positive sequence
$\qquad\qquad \pi_{P_i}^{\text{IT}} = \omega_{a_i}^{\text{IT}}(\pi_{P_{i-1}}^{\text{IT}})$ $\qquad\qquad\qquad$ // derive IT-blocks with respect to P_i from those with respect to P_{i-1}
$\qquad\qquad \text{BND}_{P_i}(\pi_D) = \cup\{X \in \pi_{P_i}^{\text{IT}}\}$ $\qquad$ // obtain the boundary region by means of IT-blocks
$\qquad\qquad \text{PMA}(\rho_t) = \text{PMA}(\rho_t) + \{\text{BND}_{P_i}(\pi_D)\}$ $\quad$ // add the boundary region to the sequence
Step 3. Output PMA(ρ_t)
End.

ALGORITHM 2: PMA.

Theorem 9 indicates that $\pi_{P\cup\{a\}}^{\text{IT}}$ can be deduced by π_P^{IT}, which implies that the IT-blocks determined by the finer tolerance relation can be achieved by successively decomposing the IT-blocks determined by the coarser tolerance relation, regardless of corresponding CT-blocks. By Theorem 9, we arrive at an efficient PMA, embodied by Definition 10.

Definition 10. Let $S = (U, C \cup D)$, $P \subseteq C$ with $P = \{a_1, a_2, \ldots, a_t\}$, and $\rho_t = \langle P_1, P_2, \ldots, P_t \rangle$ a positive sequence, where $P_1 = \{a_1\}$, $P_2 = \{a_1, a_2\}, \ldots,$ and $P_t = \{a_1, a_2, \ldots, a_t\}$. Then, PMA with respect to ρ_t can be defined as

$$\text{PMA}(\rho_t) = \left\langle \text{BND}_{P_t}(\pi_D) \mid 1 \le i \le t \right\rangle, \tag{12}$$

where $\text{BND}_{P_t}(\pi_D) = \bigcup_{X \in \pi_{P_t}^{\text{IT}}} X$, $\pi_{P_t}^{\text{IT}} = \omega_{a_t}^{\text{IT}}(\pi_{P_{t-1}}^{\text{IT}})$, and $\pi_{P_1}^{\text{IT}} = \omega_{a_1}^{\text{IT}}(\{U\}) = \theta_{a_1}^{\text{IT}}(U)$.

PMA is in fact a sequence of boundary regions, each of which denotes the approximation of the decision system with respect to some attribute subset. Since the tolerance relations determined by the positive sequence are from coarse to fine, it is easily proved that corresponding boundary regions stretch from broad to narrow. In other words, PMA is the sequence of gradually reduced boundary regions. When $P = C$, PMA portrays in detail the evolution of boundary region becoming narrower and narrower until reaching the boundary region with respect to the set of the whole condition attributes.

PMA provides an efficient approximation framework where a decision system is consecutively approximated by the boundary region derived from IT-blocks according to a positive sequence. This mechanism can be visualized in Figure 1.

PMA considers the universe U as a root IT-block and evaluates the boundary regions by repeatedly dividing the IT-blocks into the smaller ones with the positive sequence. For the attribute set P_1, the attribute a_1 works on the root IT-block and then outputs the consistent family $\pi_{P_1}^{\text{CT}}$ and the inconsistent family $\pi_{P_1}^{\text{IT}}$. The former is pruned, while the latter is used to produce the boundary region $\text{BND}_{P_1}(\pi_D)$

FIGURE 1: Mechanism of PMA.

and serves as father IT-blocks for next operation. As for P_2, $\pi_{P_2}^{\text{CT}}$ and $\pi_{P_2}^{\text{IT}}$ are generated by employing a_2 to operate on $\pi_{P_1}^{\text{IT}}$, and then $\text{BND}_{P_2}(\pi_D)$ is derived from $\pi_{P_2}^{\text{IT}}$. Likewise, we can obtain $\pi_{P_t}^{\text{CT}}$ and $\pi_{P_t}^{\text{IT}}$ calculated by using a_t to split $\pi_{P_{t-1}}^{\text{IT}}$ along with $\text{BND}_{P_t}(\pi_D)$ induced by $\pi_{P_t}^{\text{IT}}$. The detailed steps of this process are shown in Algorithm 2. PMA starts with an empty sequence; boundary regions with respect to a positive sequence are added incrementally. This process continues until all the attributes are traversed. For each loop, the boundary region is derivable from the corresponding IT-blocks obtained by operating on their predecessors with a single attribute.

There are several highlights decorating PMA. First, the inconsistent family with respect to some attribute subsets

Car	Price	Mileage	Size	Max speed	Acceleration
u_1	High	High	Full	Low	Good
u_2	*	*	Full	Low	Good
u_3	*	*	Compact	Low	Poor
u_4	Low	*	Full	High	Good
u_5	*	*	Full	High	Excellent
u_6	High	High	Full	*	Good

inherits from its predecessor rather than starting all over again, which makes efficient use of the intermediate results. Second, the cardinality of the boundary region with respect to a fine tolerance relation is not more than the cardinality of the boundary region with respect to a coarse tolerance relation, indicating that PMA works on a gradually reduced universe. Finally, obtaining a series of boundary regions needs to traverse the entire condition attribute set just only once. All the advantages are beneficial for PMA to achieve better computational efficiency, which can be visible in the process of attribute reduction addressed in the next section. The following example is employed to illustrate this idea.

Example 11. Consider an incomplete decision system $S = (U, C \cup D)$ showed in Table 1, where $U = \{u_1, u_2, \ldots, u_6\}$, $C = \{a_1, a_2, a_3, a_4\}$, and $D = \{d\}$. The attributes a_1, a_2, a_3, a_4, and d stand for price, mileage, size, max-speed, and acceleration, respectively. Note that Table 1 is somewhat different from that in the literature [23, 34].

Let $P = \{a_1, a_2, a_3, a_4\}$. We can build a positive sequence ρ_4 such that $\rho_4 = \{P_1, P_2, P_3, P_4\}$, where $P_1 = \{a_1\}$, $P_2 = \{a_1, a_2\}$, $P_3 = \{a_1, a_2, a_3\}$, and $P_4 = \{a_1, a_2, a_3, a_4\}$. Applying a_1 to the universe, we can get $\pi_{P_1}^{\text{IT}}$ and then $\text{BND}_{P_1}(\pi_D)$:

$$\pi_{P_1}^{\text{IT}} = \{\{u_1, u_2, u_3, u_5, u_6\}, \{u_2, u_3, u_4, u_5\}\},$$
$$\text{BND}_{P_1}(\pi_D) = \{u_1, u_2, u_3, u_4, u_5, u_6\}. \tag{13}$$

For P_2, the attribute a_2 is used to operate on $\pi_{P_1}^{\text{IT}}$ for $\pi_{P_2}^{\text{IT}}$. The corresponding results are generated:

$$\pi_{P_2}^{\text{IT}} = \{\{u_1, u_2, u_3, u_5, u_6\}, \{u_2, u_3, u_4, u_5\}\},$$
$$\text{BND}_{P_2}(\pi_D) = \{u_1, u_2, u_3, u_4, u_5, u_6\}. \tag{14}$$

Similarly, the results with respect to P_3 and P_4 are produced as follows:

$$\pi_{P_3}^{\text{IT}} = \{\{u_1, u_2, u_5, u_6\}, \{u_2, u_4, u_5\}\},$$
$$\text{BND}_{P_3}(\pi_D) = \{u_1, u_2, u_4, u_5, u_6\},$$
$$\pi_{P_4}^{\text{IT}} = \{\{u_4, u_5\}, \{u_5, u_6\}\}, \tag{15}$$
$$\text{BND}_{P_4}(\pi_D) = \{u_4, u_5, u_6\}.$$

Therefore, PMA with respect to ρ_4 is achieved:

$$\text{PMA}(\rho_4) = \langle \text{BND}_{P_1}(\pi_D), \text{BND}_{P_2}(\pi_D),$$
$$\text{BND}_{P_3}(\pi_D), \text{BND}_{P_4}(\pi_D) \rangle$$
$$= \langle \{u_1, u_2, u_3, u_4, u_5, u_6\}, \{u_1, u_2, u_3, u_4, u_5, u_6\},$$
$$\{u_1, u_2, u_4, u_5, u_6\}, \{u_4, u_5, u_6\} \rangle. \tag{16}$$

It is clear that $\text{BND}_{P_1}(\pi_D) \supseteq \text{BND}_{P_2}(\pi_D) \supseteq \text{BND}_{P_3}(\pi_D) \supseteq \text{BND}_{P_4}(\pi_D)$, which confirms the fact that the boundary region of PMA is gradually reduced in accordance with the positive sequence and finally catches up to the minimal boundary region ($\text{BND}_C(\pi_D) = \{u_4, u_5, u_6\}$).

4. PMA-Based Attribute Reduction

As mentioned previously, PMA offers a sequence of the boundary regions in descending order. If each selected attribute is so informative that it makes the boundary region narrowed remarkably, the boundary region with respect to some attribute subsets can keep up with the boundary region with respect to the set of all the attributes. In other words, a reduct is the attribute subset with the minimal number that creates the same approximation of the decision system as the original attribute set. Following the observation, we design a heuristic attribute reduction algorithm based on PMA, called PMA-AR. Before elaborating PMA-AR, we give a redefinition of the attribute dependency.

In (7), the dependency degree is definable in terms of positive region. Since positive region and boundary region are complementary within the universe of a decision system, the dependency degree can also be defined in terms of boundary region.

Definition 12. Let $S = (U, C \cup D)$, $P \subseteq C$, and $a \in C - P$, and the dependency degree of decision attribute set D to condition attribute set P is defined as

$$\gamma_P(D) = 1 - \frac{|\text{BND}_P(\pi_D)|}{|U|}. \tag{17}$$

Then, the significance of the attribute a is also equivalently redefined as

$$\text{Sig}(a, P, D) = \frac{|\text{BND}_P(\pi_D)| - |\text{BND}_{P \cup \{a\}}(\pi_D)|}{|U|}. \tag{18}$$

Inputs:
U: the universe
C: $C = \{a_1, a_2, \ldots, a_n\}$, a condition attribute set
D: $D = \{d\}$, a decision attribute set
Output:
red: a reduct of C
Begin
Sept 1. $P = \emptyset$, $\pi_P^{\text{IT}} = \{U\}$, $\text{BND}_P(\pi_D) = U$, $R = C$
Sept 2. Compute $\text{BND}_C(\pi_D)$ by Algorithm 2
Step 3. While $\text{BND}_P(\pi_D) \neq \text{BND}_C(\pi_D)$ and $R \neq \emptyset$ do {
      $a_{\max} = \text{null}$, $\pi_{\max}^{\text{IT}} = \emptyset$, $\text{BND}_{\max}(\pi_D) = \emptyset$, $\text{Sig}(a_{\max}, P, D) = -1$
      For $i = 1$ to $|R|$ do {
        $\pi_{P \cup \{r_i\}}^{\text{IT}} = \omega_{r_i}^{\text{IT}}(\pi_P^{\text{IT}})$ // r_i denotes the No. i attribute in R
        $\text{BND}_{P \cup \{r_i\}}(\pi_D) = \cup\{X \in \pi_{P \cup \{r_i\}}^{\text{IT}}\}$
        $\text{Sig}(r_i, P, D) = \left(\left|\text{BND}_P(\pi_D)\right| - \left|\text{BND}_{P \cup \{r_i\}}(\pi_D)\right|\right) / |U|$
        If $\text{Sig}(r_i, P, D) > \text{Sig}(a_{\max}, P, D)$ then {
          $a_{\max} = r_i$, $\pi_{\max}^{\text{IT}} = \pi_{P \cup \{r_i\}}^{\text{IT}}$, $\text{BND}_{\max}(\pi_D) = \text{BND}_{P \cup \{r_i\}}(\pi_D)$
          $\text{Sig}(a_{\max}, P, D) = \text{Sig}(r_i, P, D)$
        } //end If
      }//end For
      $P = P \cup \{a_{\max}\}$, $\pi_P^{\text{IT}} = \pi_{\max}^{\text{IT}}$, $\text{BND}_P(\pi_D) = \text{BND}_{\max}(\pi_D)$, $R = R - \{a_{\max}\}$
    }//end While
Step 4. Let red = P

ALGORITHM 3: PMA-AR.

The significance expresses how the boundary region will be affected if the attribute a is added to the set P. In general, an attribute with a maximal significance value is preferentially selected to guide the search process.

PMA-AR is in essence an extension of the quick reduct algorithm indicated previously. It marries PMA and the boundary region-based significance. The former provides an efficient way to compute the boundary region, and the latter acts as a router to determine the optimal search path. This effective combination allows PMA-AR to have the ability to locate a reduct efficiently. Algorithm 3 gives the detailed description of PMA-AR.

PMA-AR is constructed on PMA and employs the boundary region-based significance as the evaluation function to determine the positive sequence. A new attribute subset is achieved by this way. Each of the unselected attributes is used to work on the IT-blocks generated by the current attribute subset, and then corresponding boundary region is derived from the resulted IT-blocks. By evaluating the boundary region-based significance, the attribute with the biggest significance value is selected and added to the current attribute subset, which creates the expected attribute subset. This process continues until the boundary region with respect to the newly generated attribute subset equals the boundary region with respect to the set of all the attributes; equivalently, the dependency degree of the decision attribute set to the newly generated attribute subset is equal to that of the decision attribute set to the original attribute set.

PMA-AR produces the shortest positive sequence together with the fastest evolution of the boundary regions from maximal to minimal. A hidden reduct is uncovered which makes use of the minimal attributes to describe the approximation of the decision system with respect to the set of all the attributes.

Note that the time complexity of PMA-AR is $O(|C||U| + \sum_{i=1}^{\text{red}}(|C|-i)\cdot|\text{BND}_{P_i}(\pi_D)|)$, which is dependent on Step 2 with $O(|C||U|)$ and Step 3 with $O(\sum_{i=1}^{\text{red}}(|C|-i)\cdot|\text{BND}_{P_i}(\pi_D)|)$, where $|C|$ and $|U|$ are the cardinalities of C and U, respectively. In the worst case where $\text{BND}_{P_i}(\pi_D) = U$ and red $= C$, it is theoretically possible that PMA-AR takes time $O(|C|^2|U|)$. In fact, according to Algorithm 3, if $\text{BND}_P(\pi_D) = U$ or red $= C$, PMA-AR costs time $O(|C||U|)$, which implies that the practical time consumption is much less than $O(|C|^2|U|)$. Compared with the existing attribute reduction algorithms for incomplete decision systems with time complexities not less than $O(|C|^2|U|\log|U|)$ [23, 33], PMA-AR achieves obviously lower time complexity. The following example is employed to illustrate the working process of PMA-AR.

Example 13. Consider $S = (U, C \cup D)$ described in Table 1, where $U = \{u_1, u_2, \ldots, u_6\}$, $C = \{a_1, a_2, a_3, a_4\}$, and $D = \{d\}$. The attributes a_1, a_2, a_3, a_4, and d stand for price, mileage, size, max-speed, and acceleration, respectively.

Unlike Example 11, the positive sequence associated with PMA-AR is dynamically generated by iteratively selecting the

attribute with the maximal significance value. To this end, the following results are produced:

$$\pi^{\text{IT}}_{P_0 \cup \{a_1\}} = \{\{u_1, u_2, u_3, u_5, u_6\}, \{u_2, u_3, u_4, u_5\}\},$$

$$\text{BND}_{P_0 \cup \{a_1\}}(\pi_D) = \{u_1, u_2, u_3, u_4, u_5, u_6\},$$

$$\text{Sig}(a_1, P_0, D) = 0,$$

$$\pi^{\text{IT}}_{P_0 \cup \{a_2\}} = \{\{u_1, u_2, u_3, u_4, u_5, u_6\}\},$$

$$\text{BND}_{P_0 \cup \{a_2\}}(\pi_D) = \{u_1, u_2, u_3, u_4, u_5, u_6\},$$

$$\text{Sig}(a_2, P_0, D) = 0,$$

$$\pi^{\text{IT}}_{P_0 \cup \{a_3\}} = \{\{u_1, u_2, u_4, u_5, u_6\}\},$$

$$\text{BND}_{P_0 \cup \{a_3\}}(\pi_D) = \{u_1, u_2, u_4, u_5, u_6\},$$

$$\text{Sig}(a_3, P_0, D) = \frac{1}{6},$$

$$\pi^{\text{IT}}_{P_0 \cup \{a_4\}} = \{\{u_1, u_2, u_3, u_6\}, \{u_4, u_5, u_6\}\},$$

$$\text{BND}_{P_0 \cup \{a_4\}}(\pi_D) = \{u_1, u_2, u_3, u_4, u_5, u_6\},$$

$$\text{Sig}(a_4, P_0, D) = 0.$$

(19)

It is evident that the attribute a_3 with the maximal significance value is available and can be used to create the attribute subset $P_1 = \{a_3\}$. Then, $\pi^{\text{IT}}_{P_1} = \{\{u_1, u_2, u_4, u_5, u_6\}\}$ and $\text{BND}_{P_1}(\pi_D) = \{u_1, u_2, u_4, u_5, u_6\}$. Now, $\pi^{\text{IT}}_{P_1}$ is regarded as father IT-blocks to breed son IT-blocks by adding one of the remaining attributes to P_1. From the newly generated IT-blocks, the boundary region is derivable. By evaluating the boundary region-based significance, the next expected attribute can be selected from the set of remaining attributes $\{a_1, a_2, a_4\}$. The corresponding results are exhibited as follows:

$$\pi^{\text{IT}}_{P_1 \cup \{a_1\}} = \{\{u_1, u_2, u_5, u_6\}, \{u_2, u_4, u_5\}\},$$

$$\text{BND}_{P_1 \cup \{a_1\}}(\pi_D) = \{u_1, u_2, u_4, u_5, u_6\},$$

$$\text{Sig}(a_1, P_1, D) = 0,$$

$$\pi^{\text{IT}}_{P_1 \cup \{a_2\}} = \{\{u_1, u_2, u_4, u_5, u_6\}\},$$

$$\text{BND}_{P_1 \cup \{a_2\}}(\pi_D) = \{u_1, u_2, u_4, u_5, u_6\},$$

$$\text{Sig}(a_2, P_1, D) = 0,$$

$$\pi^{\text{IT}}_{P_1 \cup \{a_4\}} = \{\{u_4, u_5\}, \{u_5, u_6\}\},$$

$$\text{BND}_{P_1 \cup \{a_4\}}(\pi_D) = \{u_4, u_5, u_6\},$$

$$\text{Sig}(a_4, P_1, D) = \frac{1}{3}.$$

(20)

Thus, the attribute a_4 with the maximal significance value is preferably selected and used to generate another attribute subset $P_2 = \{a_3, a_4\}$. Then, $\pi^{\text{IT}}_{P_2} = \{\{u_4, u_5\}, \{u_5, u_6\}\}$ and

Table 2: Description of datasets.

No.	Dataset	Objects	Attributes	Classes
1	Lung cancer	32	56	3
2	Standardized audiology	200	69	24
3	Congressional voting	435	16	2
4	Balance scale weight and distance	625	4	3
5	Tic-tac-toe end game	958	9	2
6	Car evaluation	1728	6	4
7	Chess end game	3196	36	2
8	Nursery	12960	8	5

$\text{BND}_{P_2}(\pi_D) = \{u_4, u_5, u_6\}$. Since $\text{BND}_{P_2}(\pi_D)$ is equal to $\text{BND}_C(\pi_D)$, P_2 is just a reduct.

As a result, the reduct of the decision system is $\{a_3, a_4\}$.

5. Experimental Evaluation

In the following, we carry out several experiments on a personal computer with Windows XP, 2.53 GHZ CPU, and 2.0 G memory so as to evaluate PMA-AR in terms of the number of selected features and running time.

There are many heuristic search algorithms for attribute reduction in incomplete decision systems [17, 20, 23–25, 28, 33], of which three state-of-the-art algorithms are appropriate for comparison with PMA-AR. They are positive region-based algorithm (PRA) [23], distance metric-assisted algorithm (DMA) [28], and rough entropy-based algorithm (REA) [33], qualified as the representatives of positive region based reduction, combination-based reduction, and entropy-based reduction, respectively.

Our experiments employ eight publicly accessible datasets from UCI repository of machine learning databases [35]. Each of them is a discrete dataset with only one decision attribute. Since PMA-AR is designed to deal with incomplete data, five complete datasets, such as balance scale weight and distance, tic-tac-toe end game, car evaluation, chess end game, and nursery, are all turned into incomplete ones by randomly replacing some known attribute values with missing ones. In addition, an identifier attribute of standardized audiology is removed. The characteristics of these datasets are described in Table 2.

The experiments are performed by applying the four algorithms (PRA, DMA, REA, and PMA-AR) to the eight datasets shown in Table 2. The resulted number of selected features and running time expressed in seconds is exhibited in Tables 3 and 4, respectively.

From Table 3, it can be observed that the number of features selected by PMA-AR is the same as that by PRA but not less than those by DMA and REA. On the whole, the numbers by the four algorithms are relatively approximate. This indicates that the performances of the four algorithms are very close, though DMA and REA perform a little better than PMA-AR and PRA.

On the other hand, Table 4 shows that for each dataset, DMA needs the most time, PRA and REA get the second and

TABLE 3: Number of features selected by four algorithms.

No.	Dataset	DMA	REA	PRA	PMA-AR
1	Lung cancer	4	4	4	4
2	Standardized audiology	21	21	22	22
3	Congressional voting	8	8	9	9
4	Balance scale weight and distance	4	4	4	4
5	Tic-tac-toe end game	7	8	8	8
6	Car evaluation	6	6	6	6
7	Chess end game	26	28	29	29
8	Nursery	8	8	8	8

TABLE 4: Running time of four algorithms.

No.	Dataset	DMA	REA	PRA	PMA-AR
1	Lung cancer	1.054	0.817	0.849	0.808
2	Standardized audiology	7.132	3.167	3.363	1.564
3	Congressional voting	2.657	1.301	1.365	1.107
4	Balance scale weight and distance	2.548	1.211	1.306	0.910
5	Tic-tac-toe end game	14.745	6.694	7.703	1.834
6	Car evaluation	46.889	15.660	17.503	2.303
7	Chess end game	384.416	105.985	128.153	8.445
8	Nursery	569.490	139.974	179.430	9.224

third place, respectively, and PMA-AR need the lest. Moreover, the running time of DMA, PRA, and REA increases much more rapidly than that of PMA-AR. The differences can be illustrated by plotting the ratios of DMA, PRA, and REA to PMA-AR, respectively, as shown in Figure 2.

From Figure 2, we can find that the curve corresponding to DMA/PMA-AR increases most rapidly, and the curve corresponding to PRA/PMA-AR increases slightly rapidly than that corresponding to REA/PMA-AR. The experimental result coincides with the theoretical analysis that the time complexity of DMA is not less than $O(|C|^2|U|^2)$, which is the highest among all four algorithms; the second one is PRA, of which the time complexity is $O(|C|^2|U|\log|U|)$. Following PRA, REA has the time complexity not more than $O(|C|^2|U|\log|U|)$, and the most efficient one is PMA-AR with the time complexity much less than $O(|C|^2|U|)$. It is verified that PMA-AR achieves the best performance in terms of time efficiency.

One can also observe that although each curve tends to increase with size of datasets, it is not strictly monotonic, namely, the curves fluctuate significantly. This can be seen from the case that the ratio of DMA to PMA-AR on Dataset 2 is higher than that on Dataset 3. The main reason is that the attribute number of datasets is different, and more importantly, the number of selected features is also different. For example, Dataset 2 has 69 attributes, of which 21 features are selected, while Dataset 3 has 16 attributes, and 8 features are selected. Furthermore, the curves also indicate that when the number of attributes is far less than that of objects, the
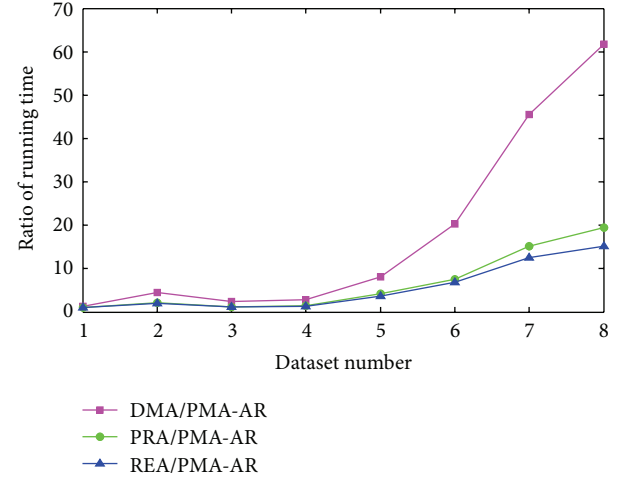

FIGURE 2: Ratios of DMA, PRA, and REA to PMA-AR.

running time mainly relies on the latter. This supports the conclusion that PMA-AR is more suitable for feature selection from large data than the other three algorithms because it is proportional to $|U|$.

6. Conclusions

Attribute reduction in incomplete decision systems has been a hotspot of rough set-based data analysis. To efficiently obtain a feature subset, many heuristic attribute reduction algorithms have been studied. Unfortunately, these algorithms are still computationally costly for large data. This paper has developed PMA-AR, which has the ability to find a feature subset as well as obviously better computational efficiency.

Unlike other algorithms featured by microcosmic approximation, PMA-AR adopts a novel model called macrocosmic approximation, which considers all the target concepts of a decision system as an indivisible whole to be approximated by rough set boundary region derived from IT-blocks. Constructed on PMA which serves as an accelerator for calculation of boundary region, PMA-AR is capable of efficiently identifying a reduct by using the boundary region-based significance as the evaluation function. Both theoretical analysis and experimental results demonstrate that PMA-AR indeed outperforms other available algorithms with regard to time efficiency.

Acknowledgment

References

[1] Z. Pawlak, "Rough sets," *International Journal of Computer and Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982.

[2] A. An, N. Shan, C. Chan, N. Cercone, and W. Ziarko, "Discovering rules for water demand prediction: an enhanced rough-set approach," *Engineering Applications of Artificial Intelligence*, vol. 9, no. 6, pp. 645–653, 1996.

[3] K. Li and Y.-S. Liu, "Rough set based attribute reduction approach in data mining," in *Proceedings of 2002 International Conference on Machine Learning and Cybernetics*, pp. 60–63, Beijing, China, November 2002.

[4] R. Li and Z.-O. Wang, "Mining classification rules using rough sets and neural networks," *The European Journal of Operational Research*, vol. 157, no. 2, pp. 439–448, 2004.

[5] H. Wang and S. Wang, "Discovering patterns of missing data in survey databases: an application of rough sets," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6256–6260, 2009.

[6] L. Shi, X. Ma, L. Xi, Q. Duan, and J. Zhao, "Rough set and ensemble learning based semi-supervised algorithm for text classification," *Expert Systems with Applications*, vol. 38, no. 5, pp. 6300–6306, 2011.

[7] Z. H. Wang, L. Shu, and X. Y. Ding, "Reduction of neighborhood-based generalized rough sets," *Journal of Applied Mathematics*, vol. 2011, Article ID 409181, 22 pages, 2011.

[8] A. Skowron and J. Stepaniuk, "Towards an approximation theory of discrete problems," *Fundamenta Informaticae*, vol. 15, no. 2, pp. 187–208, 1991.

[9] A. A. Bakar, M. N. Sulaiman, M. Othman, and M. H. Selamat, "Finding minimal reduct with binary integer programming in data mining," in *Proceedings of the TENCON*, pp. 141–146, September 2000.

[10] Y. Chen, D. Miao, R. Wang, and K. Wu, "A rough set approach to feature selection based on power set tree," *Knowledge-Based Systems*, vol. 24, no. 2, pp. 275–281, 2011.

[11] M. E. ElAlami, "A filter model for feature subset selection based on genetic algorithm," *Knowledge-Based Systems*, vol. 22, no. 5, pp. 356–362, 2009.

[12] R. Jensen and Q. Shen, "Finding rough set reducts with ant colony optimization," in *Proceeding of UK Workshop Computational Intelligence*, pp. 15–22, 2003.

[13] X. Y. Wang, J. Yang, X. L. Teng, W. J. Xia, and R. Jensen, "Feature selection based on rough sets and particle swarm optimization," *Pattern Recognition Letters*, vol. 28, no. 4, pp. 459–471, 2007.

[14] J. W. Guan and D. A. Bell, "Rough computational methods for information systems," *Artificial Intelligence*, vol. 105, no. 1-2, pp. 77–103, 1998.

[15] M. Beynon, "Reducts within the variable precision rough sets model: a further investigation," *The European Journal of Operational Research*, vol. 134, no. 3, pp. 592–605, 2001.

[16] M. Dash and H. Liu, "Consistency-based search in feature selection," *Artificial Intelligence*, vol. 151, no. 1-2, pp. 155–176, 2003.

[17] Y. Caballero, D. Alvarez, R. Bello, and M. M. Garcia, "Feature selection algorithms using rough set theory," in *Proceedings of the 7th International Conference on Intelligent Systems Design and Applications (ISDA '07)*, pp. 407–411, Rio de Janeiro, Brazil, October 2007.

[18] Y. H. Qian, J. Y. Liang, and C. Y. Dang, "Consistency measure, inclusion degree and fuzzy measure in decision tables," *Fuzzy Sets and Systems*, vol. 159, no. 18, pp. 2353–2377, 2008.

[19] A. Chouchoulas and Q. Shen, "Rough set-aided keyword reduction for text categorization," *Applied Artificial Intelligence*, vol. 15, no. 9, pp. 843–873, 2001.

[20] K. Thangavel, A. Pethalakshmi, and P. Jaganathan, "A novel reduct algorithm for dimensionality reduction with missing values based on rough set theory," *International Journal of Soft Computing*, vol. 1, no. 2, pp. 111–117, 2006.

[21] A. Pethalakshmi and K. Thangavel, "Performance analysis of accelerated quickreduct algorithm," in *Proceedings of the International Conference on Computational Intelligence and Multimedia Applications (ICCIMA '07)*, pp. 318–322, Sivakasi, Tamil Nadu, India, December 2007.

[22] P. S. V. S. S. Prasad and C. R. Rao, "IQuickReduct: an improvement to Quick Reduct algorithm," in *Rough Sets, Fuzzy Sets, Data Mining and Granular Computing*, H. Sakai, Ed., vol. 5908, pp. 152–159, Heidelberg, Germany, 2009.

[23] Z. Q. Meng and Z. Z. Shi, "A fast approach to attribute reduction in incomplete decision systems with tolerance relation-based rough sets," *Information Sciences*, vol. 179, no. 16, pp. 2774–2793, 2009.

[24] Y. Qian, J. Y. Liang, W. Pedrycz, and C. Dang, "Positive approximation: an accelerator for attribute reduction in rough set theory," *Artificial Intelligence*, vol. 174, no. 9-10, pp. 597–618, 2010.

[25] Y. H. Qian, J. Y. Liang, W. Pedrycz, and C. Y. Dang, "An efficient accelerator for attribute reduction from incomplete data in rough set framework," *Pattern Recognition*, vol. 44, no. 8, pp. 1658–1670, 2011.

[26] N. Zhong, J. Dong, and S. Ohsuga, "Using rough sets with heuristics for feature selection," *Journal of Intelligent Information Systems*, vol. 16, no. 3, pp. 199–214, 2001.

[27] M. Zhang and J. T. Yao, "A rough sets based approach to feature selection," in *Proceedings of the IEEE Annual Meeting of the North American Fuzzy Information Processing Society: Fuzzy Sets in the Heart of the Canadian Rockies (NAFIPS '04)*, pp. 434–439, Saskatchewan, Canada, June 2004.

[28] N. M. Parthaláin and Q. Shen, "Exploring the boundary region of tolerance rough sets for feature selection," *Pattern Recognition*, vol. 42, no. 5, pp. 655–667, 2009.

[29] N. M. Parthaláin, Q. Shen, and R. Jensen, "A distance measure approach to exploring the rough set boundary region for attribute reduction," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 3, pp. 305–317, 2010.

[30] J. Qian, D. Q. Miao, Z. H. Zhang, and W. Li, "Hybrid approaches to attribute reduction based on indiscernibility and discernibility relation," *International Journal of Approximate Reasoning*, vol. 52, no. 2, pp. 212–230, 2011.

[31] G. Y. Wang, J. Zhao, J. J. An, and Y. Wu, "A comparative study of algebra viewpoint and information viewpoint in attribute reduction," *Fundamenta Informaticae*, vol. 68, no. 3, pp. 289–301, 2005.

[32] Y. H. Qian and J. Y. Liang, "Combination entropy and combination granulation in rough set theory," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 16, no. 2, pp. 179–193, 2008.

[33] L. Sun, J. C. Xu, and Y. Tian, "Feature selection using rough entropy-based uncertainty measures in incomplete decision systems," *Knowledge-Based Systems*, vol. 36, pp. 206–216, 2012.

[34] Y. Leung and D. Y. Li, "Maximal consistent block technique for rule acquisition in incomplete information systems," *Information Sciences*, vol. 153, pp. 85–106, 2003.

[35] A. Asuncion and D. J. Newman, *UCI Machine Learning Repos-Itory*, Department of Information and Computer Science, University of California, Irvine, CA, USA, 2007, http://www.ics.uci.edu/mlearn/MLRepository.html.