MDPI

Article

An Accelerating Reduction Approach for Incomplete Decision Table Using Positive Approximation Set

Tao Yan [1,2,*], **Chongzhao Han** [1,2], **Kaitong Zhang** [1] and **Chengnan Wang** [1]

1 School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China; czhan@xjtu.edu.cn (C.H.); zkt0301@stu.xjtu.edu.cn (K.Z.); chengnanhz@stu.xjtu.edu.cn (C.W.)
2 MOE Key Lab for Intelligent Network and Network Security, Xi'an Jiaotong University, Xi'an 710049, China
* Correspondence: yantao@xjtu.edu.cn

Abstract: Due to the explosive growth of data collected by various sensors, it has become a difficult problem determining how to conduct feature selection more efficiently. To address this problem, we offer a fresh insight into rough set theory from the perspective of a positive approximation set. It is found that a granularity domain can be used to characterize the target knowledge, because of its form of a covering with respect to a tolerance relation. On the basis of this fact, a novel heuristic approach ARIPA is proposed to accelerate representative reduction algorithms for incomplete decision table. As a result, ARIPA in classical rough set model and ARIPA-IVPR in variable precision rough set model are realized respectively. Moreover, ARIPA is adopted to improve the computational efficiency of two existing state-of-the-art reduction algorithms. To demonstrate the effectiveness of the improved algorithms, a variety of experiments utilizing four UCI incomplete data sets are conducted. The performances of improved algorithms are compared with those of original ones as well. Numerical experiments justify that our accelerating approach enhances the existing algorithms to accomplish the reduction task more quickly. In some cases, they fulfill attribute reduction even more stably than the original algorithms do.

Keywords: rough set; incomplete decision table; variable precision model; attribute reduction; positive approximation set

1. Introduction

With the wide usage of a diversity of advanced sensors, heterogeneous information acquisition in real-world applications has become much more simple. It also brings the challenge of dealing with a huge amount of data collected by these sensors and generating useful information. To address this challenge, multiple intelligent computing approaches were proposed, e.g., fuzzy set theory, Dempster–Shafer evidence theory, and rough set theory. Rough set theory (RST) is considered as a generalization of set theory for analyzing and processing a variety of data sets consisting of incomplete, imprecise, inconsistent, or uncertain data. It originated from Zdzislaw I. Pawlak [1] and has been identified as a creative and innovative mathematical tool in the last two decades. The rough-set-based data mining approaches have superiority in that they need no prior information, in contrast with other widely utilized strategies, such as SVM, PCA, and DNN [2–6]. Attribute reduction, or feature selection, has become one of the hot spots in the research area of big data. In recent years, the number of objects and dimensions of data sets has been increasing exponentially, as well as the quantity of large-scale data sets. For example, hundreds of thousands of attributes, which reflect various characteristics of the corresponding objects in practice, are stored in various data-sets [7]. However, a large portion of them give no benefit to the subsequent pattern recognition at all, but only take up precious storage space and consume computing time in vain. Hence, it has become a research focus to overcome this obstacle.

All conventional attribute reduction approaches can be classified into three main strategies—filtering, packing, and embedding [8]. The first strategy picks up attribute subsets on the basis of a specific type of measure, e.g., distance [9], information gain [10], dependence [11], and consistency [12]. There exist two types among these measures, one is based on distance and the other is based on consistency [13]. The second strategy adopts a particular learning algorithm to evaluate and choose attribute subsets. The third strategy is a combined strategy of the above two. Generally, the ultimate goal of rough-set-based attribute reduction is to make sure that the chosen attribute subset with lower dimension owns exactly the same discriminability as the universal set of original attributes, but does not maximize the discriminability of classes blindly [14].

The problem of attribute reduction has received increasing attention in recent decades and efforts have been made by different researchers to address various drawbacks. One of the representative methods is proposed by Skowron, who employed the discriminability matrix to retrieve all potential reducts from a given data set [15]. To fulfill the reduction task for an incomplete decision table (IDT), Skowron's method was developed by Kryszkiewicz into a generalized approach utilizing discriminability matrix [16]. Shu et al. researched an incremental attribute selection approach for the data sets with dynamic incomplete data to improve the performances of other algorithms [17–19]. To evaluate candidate features in incomplete data, Qian and Shu studied a feature selection approach on the basis of mutual information criterion [20]. Jin and Li investigated in a reduction algorithm based on positive region, i.e., FPR algorithm, to reduce the computation load of attribute reduction [21]. Yan and Han presented an conditional entropy-based reduction algorithm for IDT to evaluate the uncertainty of condition attributes and eliminate redundant ones [22,23]. Xie and Qin investigated the inconsistency degree and demonstrated an incremental attribute reduction algorithm in dynamic data environments [24]. Ma et al. researched a general steg analysis attribute selection approach on the basis of α-positive region reduction [25]. Jing et al. introduced the incremental mechanisms of computing a reduct with a multi-granulation view and gave a method of updating reducts as the objects and attributes of DT change dynamically, or increasing simultaneously [26,27]. Sun et al. proposed an fuzzy neighborhood multi-granulation rough-set-based feature selection approach in neighborhood decision systems [28]. Unfortunately, some of the aforementioned methods and other reduction approaches can only deal with the issue of reduction for decision table, but not for IDT, because of the high complexity of the latter. Additionally, almost all of conventional reduction approaches for IDT would suffer from different degrees of long processing time due to large-scale computation when they process incomplete decision tables. To overcome this shortcoming, a variety of heuristic algorithms have been investigated, which can shorten computing time and reserve certain properties of corresponding IDT [29–34]. Nevertheless, their efficiencies for practical applications are still not satisfying. That is why we made our efforts to realize attribute reduction for IDT in a more intelligent and more efficient manner.

The aim of this paper is not to find a way of generating superior reducts, in contrast with most of existing attribute reduction approaches, but to study how to search for the same reduct in a more efficient way. Furthermore, the accelerating approaches for existing reduction algorithms in different rough set models, as well as their properties, are investigated in this paper. The major contributions of this research work are concluded as follows: (1) The concept of positive approximation set is constructed and one of its properties is investigated; (2) A novel heuristic accelerating approach of attribute reduction using positive approximation set for IDT is proposed; (3) The implementations of our accelerating approach are realized in different rough set models and tested by utilizing incomplete UCI data sets in the real world; (4) The performances of both computing time and stability of the proposed approach are exhibited and compared with some most recent reduction methods to verify its superiority. The simulations justify that our approach outputs precisely the same reducts as other reduction methods, while it consumes evidently less time and operates more stably in some cases. This paper is organized as follows. Some relevant

preliminaries and background concepts are briefly reviewed in Section 2. The details of positive-approximation-set-based reduction approach are provided in Section 3. Section 4 conducts a series of simulations utilizing UCI data sets and gives some analysis. Section 5 draws some conclusions.

2. Preliminaries

For the purpose of presenting our accelerating approach, it is of significance to review some concepts of rough set concerning our main subject at the very beginning. The rough set theory was firstly proposed by Z. Pawlak to describe and tackle imprecise, uncertain, and vague concepts [1]. Both classical and generalized rough set model contain a variety of mathematical concepts and definitions. To keep our research understandable, some preliminaries are presented in this section at first. Additional mathematical foundations of this paper, described in more detail, with some examples, can be found in [22].

2.1. Classical Rough Set Model

RST-based attribute reduction begins with a given data table, i.e., an information system (IS). It consists of all objects we are interested in, as well as their features which can be described by a finite attribute set. An IS containing non-empty attribute values is considered as a complete IS, otherwise it implies as an incomplete information system (IIS). Generally, meeting empty attribute value in data mining and other data processing is almost inevitable. These empty values commonly stand for unavailable feature or inaccessible data, which may be caused by the error in measurement, the impreciseness in data acquisition, the low level of belief in the obtained data, and other potential factors. Therefore, an IIS means the existence of unavailable data or missing value in the system [35]. If an IIS contains a decision attribute which is different from other condition attributes and can indicate the category of the corresponding object, then it stands for an incomplete decision table (IDT).

An IS can be described by a pair (U, A), where $U = \{x_1, \ldots, x_n\}$ indicates the universe of discourse which is actually a non-empty, finite set of objects, and $A = \{a_1, \ldots, a_m\}$ indicates a finite attribute set. There also exists a mapping $a : U \to V_a$ for any $a \in A$, where V_a denotes the domain of the attribute a.

A decision table (DT) with the form of $(U, C \cup \{d\})$ is actually a special information system, where C indicates the whole condition attribute set in DT which can reflect specific features of the target object, and $d \notin C$ indicates decision attribute which implies the object's category. Let V_d indicate the domain of decision attribute mapping $d(x)$. An attribute set is actually a feature set for pattern classification, and a training pattern set or its sign set can be represented by the universe of discourse.

Let $[x]_R$ denote an equivalence relation on U, and $\varnothing$ denote an empty set. It implies that relation R is reflexive, symmetric, and transitive. Hence it can generate a partition $U/R = \text{IND}(R) = \{[x]_R | x \in U\}$ on U, where $\text{IND}(R)$ indicates a equivalence class (i.e., an indiscernible class) which is generated by the relation R. In RST, it can also be considered as an elementary set of R. As for any target set $X \subseteq U$, the following two elementary sets of R can be used to approximate X.

$$R_-(X) = \{[x]_R | [x]_R \subseteq X\} \tag{1}$$

$$R^-(X) = \{[x]_R | [x]_R \cap X \neq \varnothing\} \tag{2}$$

They are defined as the lower and upper approximation sets of X, respectively. Furthermore, the equations of positive region, negative region, boundary region, and approximation measure are, respectively, presented as

$$\text{POS}(X) = R_-(X) \tag{3}$$

$$\text{NEG}(X) = U - R^-(X) \tag{4}$$

$$\text{BND}(X) = R^-(X) - R_-(X) \tag{5}$$

$$\alpha_R(X) = \frac{|R_-(X)|}{|R^-(X)|} \tag{6}$$

where $X \neq \varnothing$. The lower approximation is equivalent with the positive region of X, which denotes a subset consisting of the objects that can be undoubtedly classified as members of X. In contrast, the upper approximation consists of the objects that are possibly members of X. Moreover, the negative region consists of the objects that can be definitely ruled out as the members of X. Finally, the approximation measure $\alpha_R(X)$ is utilized to evaluate the completeness degree of our knowledge on X.

We use $*$ to denote empty attribute value, which means that the value of the corresponding condition attribute of the object is missing or unavailable. The IS and DT containing $*$ attribute value are, respectively, defined as incomplete information system (IIS) and incomplete decision table (IDT). Commonly, the process of attribute reduction for incomplete data set is starts with an IDT.

2.2. Incomplete Variable Precision Rough Set Model

In the latest decade, a variety of generalized rough-set-model-based reduction approaches have been proposed and developed. This subsection is dedicated to introducing some notations concerning incomplete variable precision model for use.

Let (U, A) be an IS which owns attribute subset $P \subseteq A$. The definition of a binary similarity relation on U can be expressed as

$$\text{SIM}(P) = \{(x,y) \in U \times U | \forall a \in P, [a(x) = a(y)] \cup [a(x) = *] \cup [a(y) = *]\} \tag{7}$$

As a matter of fact, $\text{SIM}(P)$ is essentially a tolerance relation on P. It can be simply obtained that $\text{SIM}(P) = \cap_{a \in P}\text{SIM}(\{a\})$.

Let $\text{SIM}(P) = \cap_{a \in P}\text{SIM}(\{a\})$ be an IIS, $P \subseteq A$ be a subset of condition attributes A, and X be a subset of the universe of discourse U. The target set X can be approximated by $\overline{\text{SIM}(P)}X$ and $\underline{\text{SIM}(P)}X$, i.e.,

$$\begin{cases} \underline{\text{SIM}(P)}X = \cup\{Y \in U/\text{SIM}(P) | Y \subseteq X\} \\ \overline{\text{SIM}(P)}X = \cup\{Y \in U/\text{SIM}(P) | Y \cap X \neq \varnothing\} \end{cases} \tag{8}$$

where $U/\text{SIM}(P)$ denotes a partition of the universe of discourse U with respect to $\text{SIM}(P)$.

A classification task for DT can be characterized by $\text{DT} = (U, C \cup D)$, where C indicates the universe of condition attributes, D indicates the decision attribute set, and there exists $C \cap D = \varnothing$. All objects are assumed to be partitioned by D into r disjoint sets, i.e., $\{X_1, X_2, \ldots, X_r\}$. Given a tolerance relation, SIM(P), generated from P, where P indicates a condition attribute subset $P \subseteq C$, then the lower and upper approximation set with respect to D can be defined, respectively, as

$$\begin{cases} \underline{\text{SIM}(P)}D = \left\{\underline{\text{SIM}(P)}X_1, \underline{\text{SIM}(P)}X_2, \ldots, \underline{\text{SIM}(P)}X_r\right\} \\ \overline{\text{SIM}(P)}D = \left\{\overline{\text{SIM}(P)}X_1, \overline{\text{SIM}(P)}X_2, \ldots, \overline{\text{SIM}(P)}X_r\right\} \end{cases} \tag{9}$$

Given $POS_P(D) = \bigcup_{i=1}^{r} \underline{\text{SIM}(P)}X_i$, i.e., the positive region of D with respect to P. The misclassification function c and the granularity-based approximation set have been proposed to construct variable precision rough set models [36]. This model can be further generalized for acquiring a more flexible algorithm for IDT attribute reduction.

Let the pair (U, A) be an IIS, $P \subseteq A$ be a subset of condition attributes, and X be a target subset of the universe of discourse U. The threshold β is given as $\beta \in [0, 0.5]$, then X can be approximated by $\underline{\mathrm{SIM}(P)}^{\beta}X$ and $\overline{\mathrm{SIM}(P)}^{\beta}X$, i.e.,

$$\begin{cases} \underline{\mathrm{SIM}(P)}^{\beta}X = \{x | D(S_P(x), S_P(x) \cap X) \leq \beta, x \in X\} \\ \overline{\mathrm{SIM}(P)}^{\beta}X = \{x | D(S_P(x), S_P(x) \cap X) \leq 1 - \beta, x \in U\} \end{cases} \tag{10}$$

where they satisfy $\underline{\mathrm{SIM}(P)}^{\beta}X \subseteq X \subseteq \overline{\mathrm{SIM}(P)}^{\beta}X$.

Let the pair $(\overline{U}, C \cup D)$ be a DT. All objects are assumed to be partitioned by D into r disjoint sets, i.e., $\{X_1, X_2, \ldots, X_r\}$. Given a tolerance relation $\mathrm{SIM}(P)$ generated from P, where P indicates a condition attribute subset $P \subseteq C$, then the lower and upper approximation set with respect to D in variable precision model can be defined, respectively, as

$$\begin{cases} \underline{\mathrm{SIM}(P)}^{\beta}D = \left\{\underline{\mathrm{SIM}(P)}^{\beta}X_1, \underline{\mathrm{SIM}(P)}^{\beta}X_2, \ldots, \underline{\mathrm{SIM}(P)}^{\beta}X_r\right\} \\ \overline{\mathrm{SIM}(P)}^{\beta}D = \left\{\overline{\mathrm{SIM}(P)}^{\beta}X_1, \overline{\mathrm{SIM}(P)}^{\beta}X_2, \ldots, \overline{\mathrm{SIM}(P)}^{\beta}X_r\right\} \end{cases} \tag{11}$$

The positive region of rough set in variable precision model can be obtained as $POS_P^{\beta}(D) = \bigcup_{i=1}^{r} \underline{\mathrm{SIM}(P)}^{\beta}X_i$, i.e., β-positive region of D with respect to P. According to the above framework, a novel algorithm can be demonstrated for attribute reduction in an incomplete variable precision model.

Let the pair (U, A) be an IIS. Given a partial order relation $\preceq$ on 2^A (power set of A) [36], if set P is crisper than set Q, in other words Q is rougher than P, then it is definite that $P \preceq Q$ satisfies (if $S_P(x_i) \subseteq S_Q(x_i)$ holds for any $i \in \{1, 2, \ldots, |U|\}$). If $P \neq Q$ and $P \preceq Q$ satisfy simultaneously, then we use the notation $P \prec Q$.

2.3. The Positive Approximation Set of IIS and IDT

An introduction of positive approximation set is demonstrated in this subsection as a preparation for proposing our algorithm. With regard to an incomplete data set, a granularity domain, which can be employed to describe target knowledge, is provided by a covering generated from a tolerance relation. Furthermore, a sequence of granularity domains ranging from rough to crisp is determined by a corresponding sequence of condition attribute subsets with granularity (same ranging as the domains) in the power set of condition attributes.

Let the pair (U, A) be an IIS, $X \subseteq U$ be a target subset, and $\mathbf{P} = \{P_1, P_2, \ldots, P_n\}$ be a subset family satisfying $P_1 \succeq P_2 \succeq \ldots \succeq P_n$, where $P_i \in 2^A$, $i = 1, 2, \ldots, n$. Given $\mathbf{P}_i = \{P_1, P_2, \ldots, P_i\}$, $\mathbf{P}_i$-lower and $\mathbf{P}_i$-upper approximation sets of X for IIS can be defined as

$$\begin{cases} \underline{\mathbf{P}_i}(X) = \bigcup_{k=1}^{i} \underline{\mathrm{SIM}(P_k)}X_k \\ \overline{\mathbf{P}_i}(X) = \overline{\mathrm{SIM}(P_i)}X \end{cases} \tag{12}$$

where $X_1 = X$. It can be obtained that $X_k = X - \bigcup_{j=1}^{k-1} \underline{\mathbf{P}_j}(X_j)$ for $k = 2, 3, \ldots, i$, where $i = 1, 2, \ldots, n$. This definition demonstrates the fact that X can be approximated by the corresponding approximation sets, i.e., $\underline{\mathbf{P}_i}(X)$ and $\overline{\mathbf{P}_i}(X)$. The $\mathbf{P}_i$-lower and $\mathbf{P}_i$-upper approximation sets of X for IIS in variable precision model can be defined, respectively, as

$$\begin{cases} \underline{\mathbf{P}_i}^{\beta}(X) = \bigcup_{k=1}^{i} \underline{\mathrm{SIM}(\mathbf{P}_k)}^{\beta}X_k \\ \overline{\mathbf{P}_i}^{\beta}(X) = \overline{\mathrm{SIM}(\mathbf{P}_i)}^{\beta}X \end{cases} \tag{13}$$

Let the pair (U, A) be an IIS, $X \subseteq U$ be a target subset, and $\mathbf{P} = \{P_1, P_2, \ldots, P_n\}$ be a subset family satisfying $P_1 \succeq P_2 \succeq \ldots \succeq P_n$, where $P_i \in 2^A$, $i = 1, 2, \ldots, n$. Given $\mathbf{P}_i = \{P_1, P_2, \ldots, P_i\}$, where $i = 1, 2, \ldots, n$, it can be obtained that

$$\mathrm{POS}^U_{\mathbf{P}_{i+1}}(D) = \mathrm{POS}^U_{\mathbf{P}_i}(D) \cup \mathrm{POS}^{U_{i+1}}_{\mathbf{P}_{i+1}}(D) \tag{14}$$

where $U_1 = U$, $U_{i+1} = U - \mathrm{POS}^U_{\mathbf{P}_i}(D)$. Since the positive approximation set of IIS is related to the structure of target concept X (i.e. it is related to the tolerance class in the lower approximation set of X with respect to $\mathbf{P}$), the tolerance class on U can be employed to redefine the $\mathbf{P}$-positive approximation set of X.

Let the pair $(U, C \cup D)$ be an IDT, $X \subseteq U$ be a target subset, $\mathbf{P} = \{P_1, P_2, \ldots, P_n\}$ be a subset family satisfying $P_1 \succeq P_2 \succeq \ldots \succeq P_n$, and $U/D = \{X_1, X_2, \ldots, X_r\}$ be a partition of the universe, U, with respect to D. The $\mathbf{P}$-lower and $\mathbf{P}$-upper approximation sets of D for IDT can be defined, respectively, as

$$\begin{cases} \underline{\mathbf{P}}D = \{\underline{\mathbf{P}}(X_1), \underline{\mathbf{P}}(X_2), \ldots, \underline{\mathbf{P}}(X_r)\} \\ \overline{\mathbf{P}}D = \{\overline{\mathbf{P}}(X_1), \overline{\mathbf{P}}(X_2), \ldots, \overline{\mathbf{P}}(X_r)\} \end{cases} \tag{15}$$

The $\mathbf{P}$-lower and $\mathbf{P}$-upper approximation sets of D for IDT in a variable precision model can be defined, respectively, as

$$\begin{cases} \underline{\mathbf{P}}^\beta D = \left\{\underline{\mathbf{P}}^\beta(X_1), \underline{\mathbf{P}}^\beta(X_2), \ldots, \underline{\mathbf{P}}^\beta(X_r)\right\} \\ \overline{\mathbf{P}}^\beta D = \left\{\overline{\mathbf{P}}^\beta(X_1), \overline{\mathbf{P}}^\beta(X_2), \ldots, \overline{\mathbf{P}}^\beta(X_r)\right\} \end{cases} \tag{16}$$

There exists a similar conclusion for IDT, which is $\mathrm{POS}^{\beta U}_{\mathbf{P}_{i+1}}(D) = \mathrm{POS}^{\beta U}_{\mathbf{P}_i}(D) \cup \mathrm{POS}^{\beta U_{i+1}}_{\mathbf{P}_{i+1}}(D)$, where $U_1 = U$, $U_{i+1} = U - \mathrm{POS}^{\beta U}_{\mathbf{P}_i}(D)$. This implies that the granularity sequence can be used to approximate the target knowledge D from positive direction. Our accelerating reduction algorithm for IDT was mainly inspired by this conclusion.

3. Accelerating Reduction Approach for IDT Using Positive Approximation Set

To achieve the ultimate goal of attribute reduction, it is necessary to obtain the specific attribute subset that contains least condition attributes and reserve the same discriminability as C. Three procedures should be taken into consideration for realizing a heuristic reduction algorithm—searching strategy, significance evaluation, and termination condition.

Most of conventional heuristic algorithms for attribute reduction have been suffering from huge amounts of computation in different degree. For addressing this disadvantage, our research does not intend to design a brand new reduction algorithm directly, but to utilize the aforementioned positive approximation set to optimize the existing heuristic strategies for reduction and improve their performances.

3.1. Definitions of Condition Attribute Significance

One of modern reduction approaches proposed by Xie et al. (abbreviation as IPR) [24] is adopted in the following section. It is essentially developed from Shu's algorithm [18,19]. To realize our accelerating reduction algorithm, several definitions of condition attribute significance should be presented at first. Each of the these definitions can be utilized for the subsequent reduction process.

Definition 1. *Let the pair $(U, C \cup D)$ be an IDT, and $B \subseteq C$ be a subset of condition attributes. As for $\forall a \in B$, the definition of the condition attribute significance of a inside B can be expressed as*

$$\mathrm{SIG}_1^{\mathrm{inner}}(a, B, D) = \gamma_B(D) - \gamma_{B-\{a\}}(D) \tag{17}$$

where $\gamma_B(D) = |\mathrm{POS}_B(D)|/|U|$.

Definition 2. *Let the pair $(U, C \cup D)$ be an IDT, and $B \subseteq C$ be a subset of condition attributes. As for $\forall a \in C - B$, the definition of the condition attribute significance of a outside B can be expressed as*

$$\text{SIG}_1^{\text{outer}}(a, B, D) = \gamma_{B \cup \{a\}}(D) - \gamma_B(D) \tag{18}$$

The above two definitions are provided by Qian and Liang et al. [37], and the following two come from Liang and Shi et al. [35].

Definition 3. *Let the pair $(U, C \cup D)$ be an IDT, and $B \subseteq C$ be a subset of condition attributes. As for $\forall a \in B$, the definition of the condition attribute significance of a inside B can be expressed as*

$$\text{SIG}_2^{\text{inner}}(a, B, D) = E(D|B - \{a\}) - E(D|B) \tag{19}$$

Definition 4. *Let the pair $(U, C \cup D)$ be an IDT, and $B \subseteq C$ be a subset of condition attributes. As for $\forall a \in C - B$, the definition of the condition attribute significance of a outside B can be expressed as*

$$\text{SIG}_2^{\text{outer}}(a, B, D) = E(D|B) - E(D|B \cup \{a\}) \tag{20}$$

On the basis of Definitions 1 and 2, the corresponding measures of significance can be utilized to construct a new algorithm in incomplete variable precision model, which is capable of reserving β-positive region with respect to the target knowledge D.

Definition 5. *Let the pair $(U, C \cup D)$ be an IDT, and $B \subseteq C$ be a subset of condition attributes. As for $\forall a \in B$, the definition of the condition attribute significance of a inside B can be expressed as*

$$\text{SIG}_3^{\text{inner}}(a, B, D) = \gamma_B^{\beta}(D) - \gamma_{B - \{a\}}^{\beta}(D) \tag{21}$$

where $\gamma_B^{\beta}(D) = \left| \text{POS}_B^{\beta}(D) \right| / |U|$.

Definition 6. *Let the pair $(U, C \cup D)$ be an IDT, and $B \subseteq C$ be a subset of condition attributes. As for $\forall a \in C - B$, the definition of the condition attribute significance of a outside B can be expressed as*

$$\text{SIG}_3^{\text{outer}}(a, B, D) = \gamma_{B \cup \{a\}}^{\beta}(D) - \gamma_B^{\beta}(D) \tag{22}$$

3.2. Rank Reservation Property of Attribute Significance

This subsection plans to give a discussion on rank reservation property of the condition attribute significance to provide a theory fundamental for proposing our accelerating reduction algorithm. For simplicity and clarity of the content, the notation $\text{SIG}_\lambda^{\text{outer}}(a, U, B, D)$ is adopted to indicate the condition attribute significance in previous subsection, where $\lambda \in (1, 2, 3)$. Additionally, $S_B^U(x)$ denotes a tolerance class generated from the object x, with respect to the attribute subset B, on the universe of discourse U. The detailed proofs of all lemmas and theorems appearing in this subsection are demonstrated in Appendixes A and B , respectively.

Firstly, two Lemmas are presented and proved aiming at investigating in the rank reservation property of the dependence based condition attribute significance for IDT.

Lemma 1. *Let A, B, C, A', B', C' be six finite set such that we have $A' = A \cup C$ and $B' = B \cup C'$ satisfied. If $A' \subseteq B'$ and $C' \cap (A \cup B) = \varnothing$ satisfy, then we have $A \subseteq B$.*

Lemma 2. *Let the pair $(U, C \cup D)$ be an IDT, such that we have $B \subseteq C$ and $U' = U - \text{POS}_B^U(D)$ satisfied. If $S_{B \cup \{a\}}^U(x') \subseteq S_D^U(x')$ and $x' \in U'$ satisfy, then we have $S_{B \cup \{a\}}^{U'}(x') \subseteq S_D^{U'}(x')$.*

Secondly, the theorem of rank reservation property can be proved as follows according to Lemmas 1 and 2.

Theorem 1. *Let the pair $(U, C \cup D)$ be an IDT, such that we have $B \subseteq C$ and $U' = U - \text{POS}_B^U(D)$ satisfied. As for $\forall a, b \in C - B$, if $\text{SIG}_1^{\text{outer}}(a, U, B, D) \geq \text{SIG}_1^{\text{outer}}(b, U, B, D)$ is satisfied, then we have*

$$\text{SIG}_1^{\text{outer}}(a, U', B, D) \geq \text{SIG}_1^{\text{outer}}(b, U', B, D) \tag{23}$$

Finally, to investigate in the rank reservation property of condition attribute significance in Yan's conditional entropy reduction approach for IDT [23], the following Lemma 3 is indispensable. Additionally, this property can be described by Theorems 2 and 3 in incomplete rough set model and incomplete variable precision model, respectively.

Lemma 3. *Let the pair $(U, C \cup D)$ be an IDT, such that we have $B \subseteq C$ and $U' = U - \text{POS}_B^U(D)$ satisfied. Then, we have $\left|S_B^U(x')\right| - \left|S_B^U(x') \cap S_D^U(x')\right| = \left|S_B^{U'}(x')\right| - \left|S_B^{U'}(x') \cap S_D^{U'}(x')\right|$, where $x' \in U'$.*

Theorem 2. *Let the pair $(U, C \cup D)$ be an IDT, such that we have $B \subseteq C$ and $U' = U - \text{POS}_B^U(D)$ satisfied. As for $\forall a, b \in C - B$, if $\text{SIG}_2^{\text{outer}}(a, U, B, D) \geq \text{SIG}_2^{\text{outer}}(b, U, B, D)$, then there exists $\text{SIG}_2^{\text{outer}}(a, U', B, D) \geq \text{SIG}_2^{\text{outer}}(b, U', B, D)$.*

Theorem 3. *Let the pair $(U, C \cup D)$ be an IDT, such that we have $B \subseteq C$, $\beta = 0$ and $U' = U - \text{POS}_B^U(D)$ satisfied. As for $\forall a, b \in C - B$, if $\text{SIG}_3^{\text{outer}}(a, U, B, D) \geq \text{SIG}_3^{\text{outer}}(b, U, B, D)$ satisfies, then there exists $\text{SIG}_3^{\text{outer}}(a, U', B, D) \geq \text{SIG}_3^{\text{outer}}(b, U', B, D)$.*

It can be concluded from the above theorems that the result of reduction would be unchanged as the object number of lower approximation set of positive approximation set for IDT is reduced. In other words, the significance rank of the selected reducts can be reserved when the positive region of positive approximation set for IDT narrows.

3.3. Accelerating Attribute Reduction Algorithms

Generally, all reduction approaches based on RST are designed to find a minimal subset consisting of no redundant attribute and reserving specific property, like the whole universe of condition attributes C. It is essentially NP-hard to seek out all potential reducts of an IDT, hence it is only necessary to search for any of them.

It is indispensable to achieve the tolerance class generated from the concerning attributes. Therefore, an accelerating algorithm of tolerance class acquisition for IDT reduction is proposed. The inspiration of this implementation partially comes from the method of radix sorting, and the computation complexity of the algorithm equates as follows:

$$O\left(|A|\,|U| + \sum_{j=1}^{|A|} \sum_{k=1}^{j-1} |*_{a_k}|\,|V_{a_k}|\right) \approx O\left(|A|\,|U| + |A|^2\,|U|\right) = O\left(|A|^2\,|U|\right) \tag{24}$$

where $|*_{a_k}|$ indicates the number of objects that own empty value in condition attribute a_k, and $|V_{a_k}|$ indicates the number of objects that own no empty value in a_k. A derived result of reduced computation complexity equates $O\left(|A|^2\,|U|\right)$.

The analysis of computation complexity reveals that the dimension of condition attributes has greater influence in the length of computing time, compared with the amount of target objects. Based on the above discussion, an accelerating reduction approach for IDT using positive approximation set (ARIPA) is proposed. In the framework of ARIPA, the evaluation function (or termination condition) can be expressed as $\text{EF}^U(B, D) = \text{EF}^U(C, D)$, which implies that the discernibility of condition attribute subset B is exactly the same as that of the universe of condition attributes C. The evaluation function can be chosen according to the original reduction algorithm we plan to accelerate. For an instance, if the original algorithm adopted is Yan's rough conditional entropy-based reduction algorithm in [23], then the corresponding evaluation function should be $\text{EN}^U(B, D) = \text{EN}^U(C, D)$, where EN denotes the rough conditional entropy. In other words, if $\text{EF}^U(B, D) = \text{EF}^U(C, D)$ satisfies, then B should be one of the reducts we search for. The detailed steps of ARIPA are

exhibited as follows. The outer significance $\text{SIG}^{\text{outer}}(a_k, red, D, U_i)$ and inner significance $\text{SIG}^{\text{inner}}(a_k, C, D, U)$ in Algorithm 1 can be either the pair of $\text{SIG}_1^{\text{outer}}$, $\text{SIG}_1^{\text{inner}}$ or the pair of $\text{SIG}_2^{\text{outer}}$, $\text{SIG}_2^{\text{inner}}$.

Algorithm 1: ARIPA.

Input: IDT $= (U, C \cup D)$

Output: Attribute reduct *red*

1: Initialize *red* as $\varnothing$, i.e., $red \leftarrow \varnothing$, where *red* indicates condition attribute subset which has been selected.

2: Evaluate $\text{SIG}^{\text{inner}}(a_k, C, D, U)$, where $k \leq |C|$.

3: **If** $\text{SIG}^{\text{inner}}(a_k, C, D, U) > 0$, **then** add a_k into *red*. IDT's kernel partly consists of condition attributes in *red* at this step.

4: $i \leftarrow 1, U_1 \leftarrow U, R_1 = red, \mathbf{P}_1 = \{R_1\}$.

5: **While** $U_i \neq \varnothing$ and $\text{EF}^{U_i}(red, D) \neq \text{EF}^{U_i}(C, D)$, **do**

6:     {Evaluate the positive region of the positive approximation set $\text{POS}_{\mathbf{P}_i}^U(D)$,

7:     $U_i = U - \text{POS}_{\mathbf{P}_i}^U(D)$,

8:     $i \leftarrow i + 1$,

9:     $red \leftarrow red \cup \{a_0\}$, where $\text{SIG}^{\text{outer}}(a_0, red, D, U_i) = \max\{\text{SIG}^{\text{outer}}(a_k, red, D, U_i)\}$ $a_k \in C - red\}$, **End**.

10: $R_i \leftarrow R_i \cup \{a_0\}, \mathbf{P}_i \leftarrow \{R_1, R_2, \ldots, R_i\}$.

11: Return *red*.

To accelerate the reduction algorithm in the incomplete variable precision rough set (IVPR) model by ARIPA, it is remodeled on the basis of the β-positive approximation set. The IVPR-version of accelerating reduction, Algorithm 2 (ARIPA-IVPR), is illustrated as follows.

Algorithm 2: ARIPA-IVPR.

Input: IDT $= (U, C \cup D)$, threshold $\beta \leq 0.5$

Output: Attribute reduct *red*

1: Initialize *red* as $\varnothing$, i.e., $red \leftarrow \varnothing$, where *red* indicates condition attribute subset which has been selected.

2: Evaluate $\text{SIG}_3^{\text{inner}}(a_k, C, D, U)$, where $k \leq |C|$.

3: **If** $\text{SIG}_3^{\text{inner}}(a_k, C, D, U) > 0$, **then** add a_k into *red*. IDT's kernel partly consists of condition attributes in *red* at this step.

4: $i \leftarrow 1, U_1 \leftarrow U, R_1 = red, \mathbf{P}_1 = \{R_1\}$.

5: **While** $U_i \neq \varnothing$ and $\gamma_{red}^{\beta U_i}(D) \neq \gamma_C^{\beta U_i}(D)$, **do**

6:     {Evaluate the positive region of the positive approximation set $\text{POS}_{\mathbf{P}_i}^{\beta U}(D)$,

7:     $U_i = U - \text{POS}_{\mathbf{P}_i}^{\beta U}(D)$,

8:     $i \leftarrow i + 1$,

9:     $red \leftarrow red \cup \{a_0\}$, where $\text{SIG}_3^{\text{outer}}(a_0, red, D, U_i) = \max\{\text{SIG}_3^{\text{outer}}(a_k, red, D, U_i)\}$ $a_k \in C - red\}$, **End**.

10: $R_i \leftarrow R_i \cup \{a_0\}, \mathbf{P}_i \leftarrow \{R_1, R_2, \ldots, R_i\}$.

11: Return *red*.

---

4. Experiments

To investigate the efficiency and effectiveness of the proposed ARIPA and ARIPA-IVPR, four incomplete data sets are picked up from the UCI Machine Learning Database at University of California for experimental purposes. The performances of the proposed algorithms were analyzed and compared with those of other state-of-the-art algorithms to prove their superiority.

4.1. Experiments on ARIPA and ARIPA-IVPR

Due to the existence of continuous attribute values contained in the chosen incomplete data sets, Tsai's CACC discretization algorithm [38] is adopted as a preprocess before reduction to discretize continuous values into discrete ones. Another aim of this step is to reduce the computation load of subsequent steps and compress the data scale. The average CPU time of ARIPA, ARIPA-IVPR, and their competitors is counted in seconds as their running time. All simulation work is conducted on the PC with the configurations of 8GB RAM, Intel i5-8400 2.8GHz CPU, Matlab R2019a, Win10 (64 bit). The statistical results of the four incomplete data sets for simulations are summarized and analyzed, respectively, in Table 1.

Table 1. Summary of the experimental incomplete data sets.

Incomplete Data Sets	Objects	Condition Attributes	Empty Values	Decision Classes	Incomplete Rate (%)
Audiology standardized	226	69	291	24	1.87
Breast cancer Wisconsin	699	10	16	2	0.23
Dermatology	366	34	8	6	0.06
Soybean large	307	35	712	19	6.63

To compare our improved reduction algorithms with other competitors (Xie's IPR [24] and Yan's ILCE [23]), a modern approach is carried out for evaluating their computation complexities [39]. The same reduct would be obtained by each pair of the improved and original algorithm, thus we just have to make an comparison between their running times. The graphical illustrations of their performances are shown in Figures 1 and 2. In these figures, the *x*-axis indicates the number of data segments which increases from 1 to 20 (all objects of each incomplete data set are equally divided into 20 segments), and the *y*-axis indicates the corresponding running time. The experiments using incomplete data segments in different scales would make us aware of the trend of the computing time as the scale grows. Furthermore, the simulations indirectly prove that our accelerating algorithm would exhibit more outstanding performance when the incomplete data set contains tens of thousands of objects.

With regard to the framework of incomplete variable precision model, Kang's IVPR algorithm [36] is conducted as a competitor for our improved ARIPA-IVPR. The experiment results are illustrated in Figures 3–6.

4.2. Results and Discussions

It can be noticed from Figures 1–6 that the computing time of the improved algorithm increases more smoothly than that of the original algorithm as the number of data segments grow. Essentially, this consequence can be the result for the following three reasons. (1) The accelerated algorithm consumes much less computing time when the universe of discourse shrinks dramatically. (2) As for the same incomplete data segments, the original algorithms have to consume more time to evaluate the condition attribute significance of the potential reducts. (3) Our accelerating algorithm would encapsulate all concerning objects into the lower approximation set with respect to the decision attribute set during the reduction, hence it ensures that the improved reduction algorithm would consume less time to finish the reduction. These results are caused by the rank reservation property of the condition attribute significance, as discussed in Section 3.2. It provides a solution to the inefficiency

of the existing heuristic algorithms for IDT reduction. Since the reducts from different algorithms are identical, the same classification accuracy can be ensured in subsequent process, no matter what type of classifier is chosen, e.g., SVM, decision tree, etc. It is possible that the accelerating reduction algorithm we propose leads in the problem of over-fitting, in the perspective of classifier. However, discussion on this issue is not included in this paper.
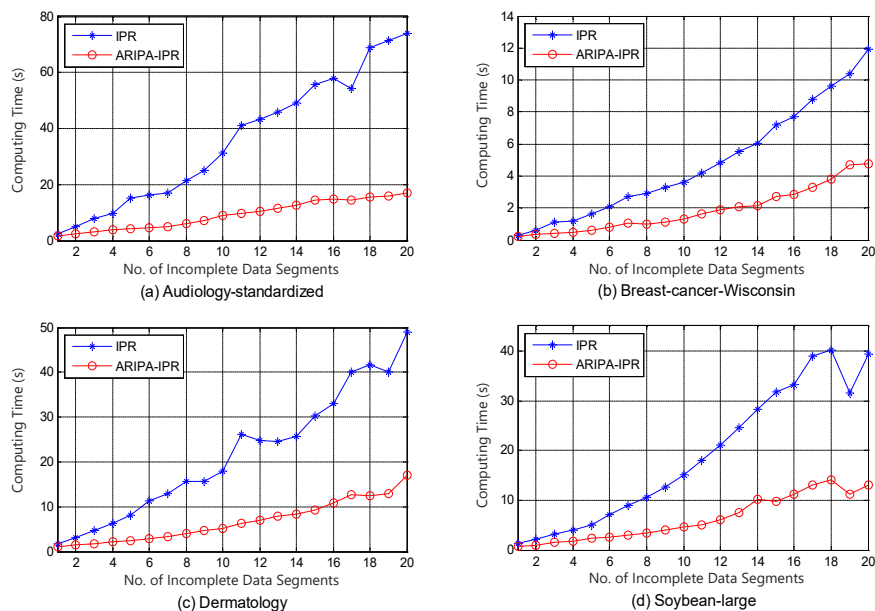

Figure 1. Computing time of IPR and ARIPA-IPR for (**a**) Audiology standardized, (**b**) Breast cancer Wisconsin, (**c**) Dermatology, and (**d**) Soybean large.
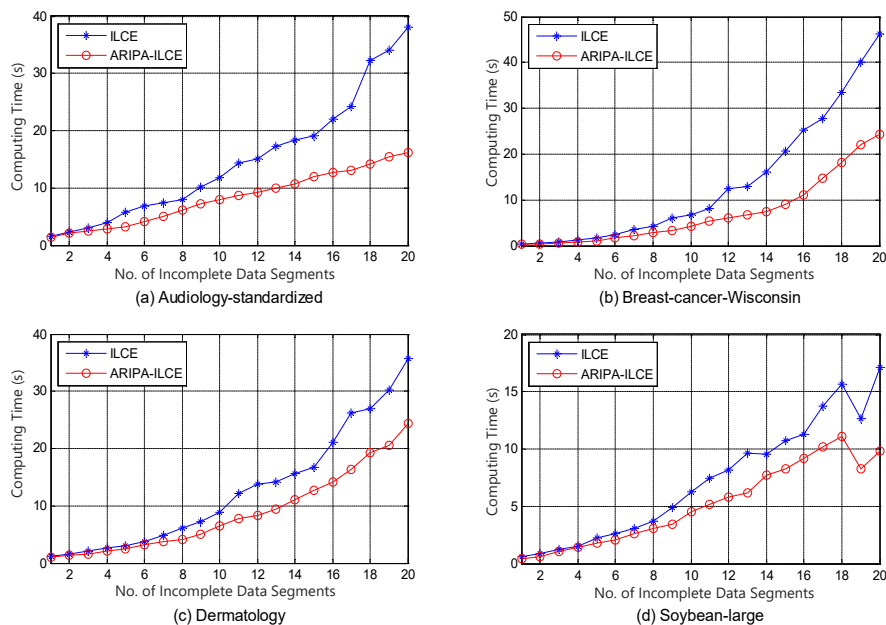

Figure 2. Computing time of ILCE and ARIPA-ILCE for (**a**) Audiology standardized, (**b**) Breast cancer Wisconsin, (**c**) Dermatology, and (**d**) Soybean large.
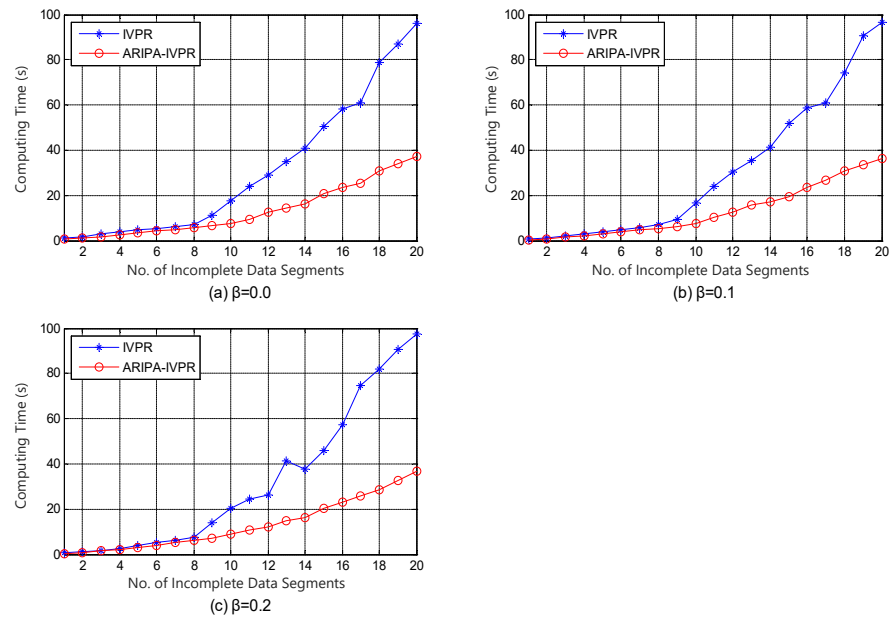
Figure 3. Computing time of IVPR and ARIPA-IVPR for Audiology standardized data set in (**a**) $\beta = 0.0$, (**b**) $\beta = 0.1$, and (**c**) $\beta = 0.2$.
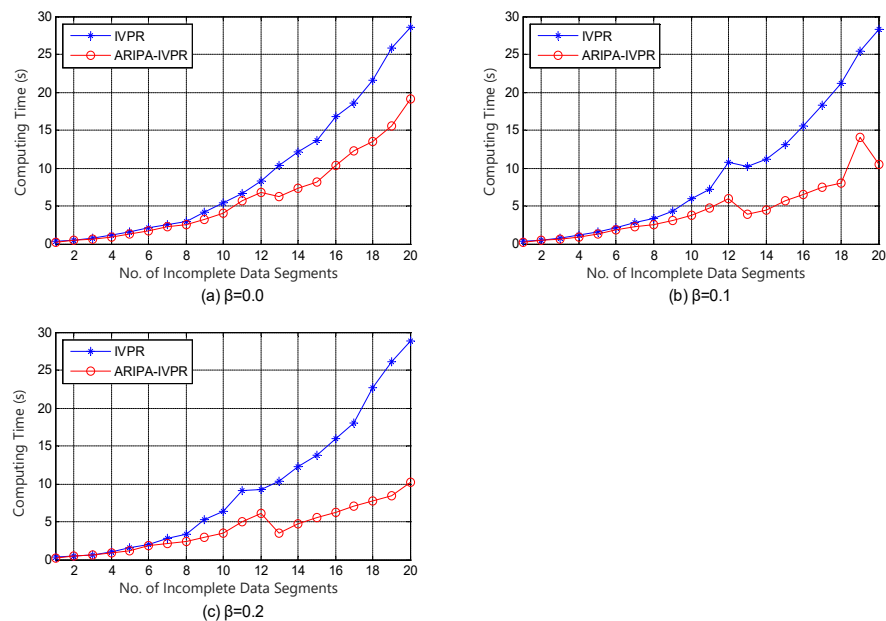

Figure 4. Computing time of IVPR and ARIPA-IVPR for Breast cancer Wisconsin data set in (**a**) $\beta = 0.0$, (**b**) $\beta = 0.1$, and (**c**) $\beta = 0.2$.

It also can be observed that the computing time rises up for most of time when the number of data segments increases in each experiment, no matter which incomplete data set, competitor algorithm, style of rough set model, or value of β we choose. However, not all the curves show a strictly monotone increasing function, and the opposite may take place in a few cases (e.g., in Figure 4). This phenomenon a result of the possibility that the new added data segment, in contrast to the existing ones, may contain specific knowledge that is more useful for attribute reduction as well as compressing the computation load.
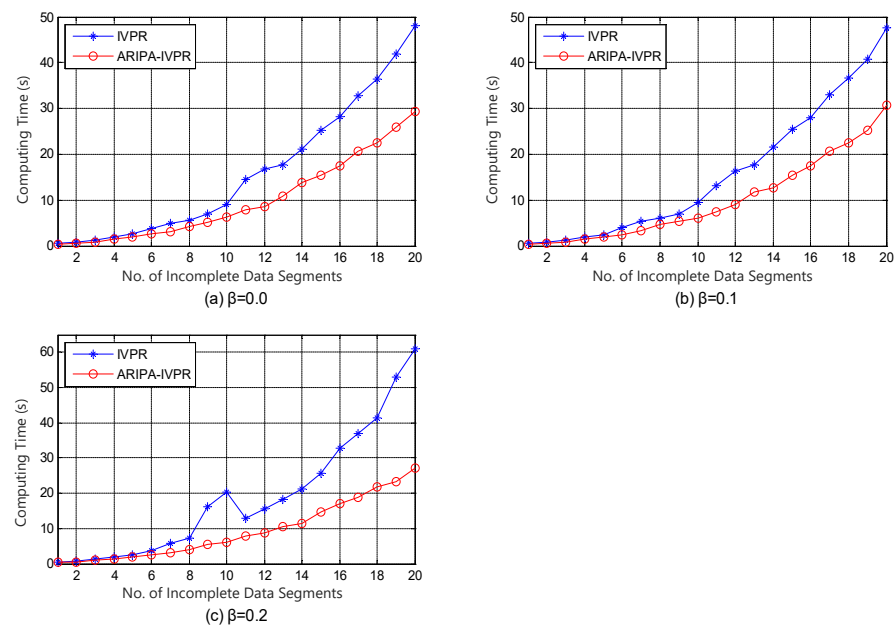
Figure 5. Computing time of IVPR and ARIPA-IVPR for Dermatology data set in (**a**) $\beta = 0.0$, (**b**) $\beta = 0.1$, and (**c**) $\beta = 0.2$.
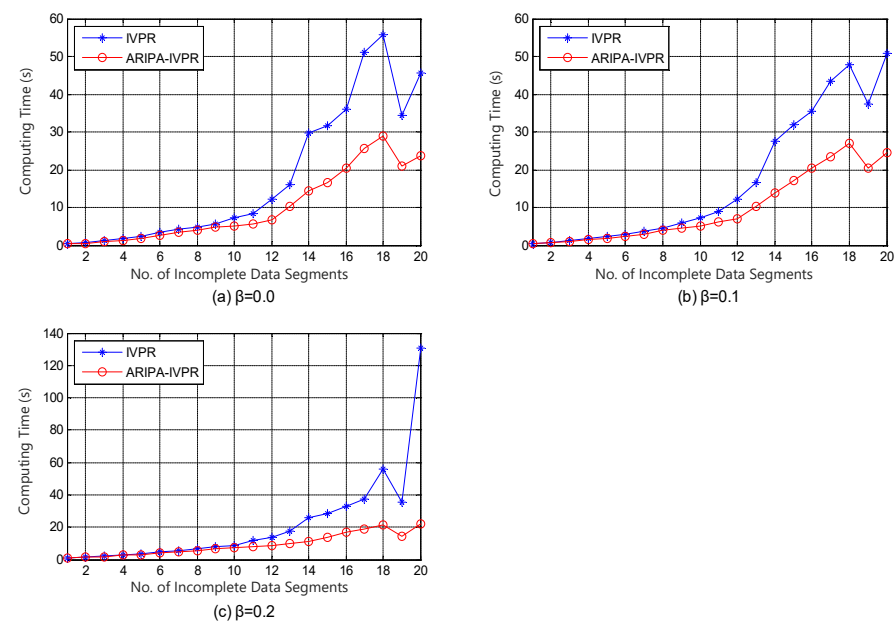

Figure 6. Computing time of IVPR and ARIPA-IVPR for Soybean large data set in (**a**) $\beta = 0.0$, (**b**) $\beta = 0.1$, and (**c**) $\beta = 0.2$.

The computation complexities of state-of-the-art [23] and improved algorithms are analyzed step by step in Table 2. It can be observed that the major difference in computation aspect is brought by step 2 and steps 5–9 of the algorithms. Among these steps, step 2 corresponds to the evaluation of the attribute significance of potential reducts, and steps 5–9 correspond to the loop which includes the evaluation of the positive region of positive approximation set and the heuristic search for real reducts. Moreover, Figures 7 and 8 indicate that our improved algorithms run more efficiently than the original algorithms, both in rough set model and variable precision model ($\beta = 0.0, 0.1, 0.2$). Hence, the experiment results justify the conclusion that the accelerated algorithms are more efficient for reduction in practical applications.

Table 2. Analysis on the computation complexity of existing and accelerated attribute reduction algorithm.

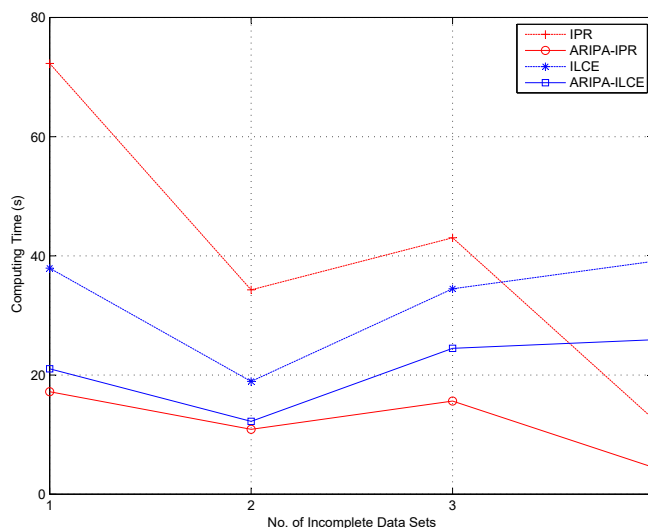| Algorithms	Step 2	Step 3	Steps 5–9	Other Steps
Existing algorithm	$O\left(\lvert C \rvert^2 \lvert U \rvert^2\right)$	$O(\lvert C \rvert)$	$O\left(\sum_{i=1}^{\lvert C \rvert} \left(\lvert C \rvert - i + 1\right)^2 \lvert U \rvert^2\right)$	Constant
Accelerated algorithm	$O\Big(\lvert C \rvert^2 \lvert U \rvert + \lvert C \rvert$ $\sum_{j=1}^{\lvert C \rvert} \sum_{k=1}^{j-1} \lvert {*}_{a_k} \rvert \lvert V_{a_k} \rvert\Big)$	$O(\lvert C \rvert)$	$O\Big(\sum_{i=1}^{\lvert C \rvert} \left(\lvert C \rvert - i + 1\right)^2 \lvert U_i \rvert +$ $\left(\lvert C \rvert - i + 1\right) \sum_{j=1}^{\lvert C \rvert - i + 1} \sum_{k=1}^{j-1} \lvert {*}_{a_k}^{U_i} \rvert \lvert V_{a_k}^{U_i} \rvert\Big)$	Constant

Figure 7. Computing time of IPR, ARIPA-IPR, ILCE, and ARIPA-ILCE for four incomplete data sets. (1—Audiology standardized; 2—Breast cancer Wisconsin; 3—Dermatology; 4—Soybean large.)
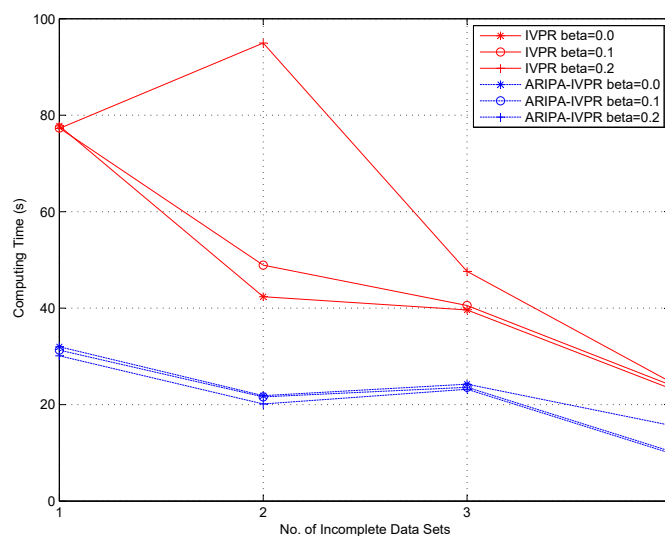

Figure 8. Computing time of IVPR and ARIPA-IVPR for four incomplete data sets ($\beta = 0.0, 0.1, 0.2$). (1—Audiology standardized; 2—Breast cancer Wisconsin; 3—Dermatology; 4—Soybean large.)

4.3. Algorithm Stability Analysis

To evaluate the stability of both original and improved algorithms, ten-fold cross-validation was applied. In this validation, a given data set is randomly parted into ten nearly equally sized subsets. Nine of them are treated as training sets, and one last subset is reserved as a testing set to evaluate the classification accuracy. The distance between

two different reducts C_i and C_j is evaluated in Equation (25), where C_0 and C_i indicate the reducts generated from U and the ith segment of U, respectively.

$$\text{Distance}(C_i, C_j) = 1 - \frac{|C_i \cap C_j|}{|C_i \cup C_j|} \tag{25}$$

Furthermore, by using the statistical method, mean (i.e., μ in Equation (26)) and standard deviation (i.e., σ in Equation (27)) of the above ten distances of the segments can be determined as well.

$$\mu = \frac{1}{10} \sum_{i=1}^{10} (1 - \frac{|C_i \cap C_0|}{|C_i \cup C_0|}) \tag{26}$$

$$\sigma = \sqrt{\frac{1}{10} \sum_{i=1}^{10} (\text{Distance}(C_i, C_0) - \mu)^2} \tag{27}$$

The stability of the reduct outputted from the heuristic reduction algorithm is characterized by standard deviation of those distances. More specifically, lower the standard deviation gets, more stably the corresponding reduction algorithm would run. The stability analysis of each pair of algorithms is carried out in Tables 3–5.

Table 3. Computing time and stability of IPR and ARIPA-IPR for four incomplete data sets.

Incomplete Data Sets	IPR's Computing Time (s)	ARIPA-IPR's Computing Time (s)	IPR's Stability	ARIPA-IPR's Stability
Audiology standardized	71.5631 ± 5.1558	17.0331 ± 1.1149	0.2624 ± 0.1380	0.2624 ± 0.1380
Breast-cancer-WI	13.3757 ± 1.6881	4.5208 ± 0.8269	0.0792 ± 0.1635	0.0792 ± 0.1635
Dermatology	42.5979 ± 1.9307	15.3719 ± 0.5125	0.2893 ± 0.2271	0.2893 ± 0.2271
Soybean large	35.9553 ± 3.9234	10.7832 ± 1.4693	0.2289 ± 0.2049	0.2289 ± 0.2049

Table 4. Computing time and stability of ILCE and ARIPA-ILCE for four incomplete data sets.

Incomplete Data Sets	ILCE's Computing Time (s)	ARIPA-ILCE's Computing Time (s)	ILCE's Stability	ARIPA-ILCE's Stability
Audiology standardized	37.5503 ± 2.7268	20.8380 ± 1.1990	0.1868 ± 0.1061	0.1868 ± 0.1061
Breast cancer Wisconsin	38.6970 ± 3.0960	25.6673 ± 2.4608	0.0727 ± 0.1160	0.0727 ± 0.1160
Dermatology	34.1226 ± 0.5987	24.2288 ± 0.5979	0.2537 ± 0.1784	0.2535 ± 0.1784
Soybean large	18.7405 ± 1.9096	12.1051 ± 0.7328	0.1754 ± 0.1349	0.1754 ± 0.1349

Table 5. Computing time and stability of IVPR and ARIPA-IVPR for four incomplete data sets.

Incomplete Data Sets	β	IVPR's Computing Time (s)	ARIPA-IVPR's Computing Time (s)	IVPR's Stability	ARIPA-IVPR's Stability
	0.0	76.4555 ± 3.0168	31.0325 ± 1.6524	0.2570 ± 0.1351	0.2570 ± 0.1351
Audiology standardized	0.1	76.0280 ± 3.5442	31.3581 ± 1.5407	0.2356 ± 0.1364	0.1782 ± 0.0895
	0.2	75.9417 ± 3.6175	29.2186 ± 1.0746	0.2329 ± 0.1705	0.1903 ± 0.1479
	0.0	22.6102 ± 3.1587	15.2200 ± 3.1775	0.0678 ± 0.1493	0.0678 ± 0.1493
Breast cancer Wisconsin	0.1	23.3089 ± 4.1228	9.9927 ± 3.5050	0.1167 ± 0.1779	0.0710 ± 0.1136
	0.2	23.9333 ± 3.6824	9.5895 ± 3.6140	0.2272 ± 0.2118	0.1389 ± 0.2826
	0.0	38.1813 ± 0.3769	23.5210 ± 0.4671	0.2209 ± 0.1930	0.2209 ± 0.1930
Dermatology	0.1	39.3397 ± 0.9589	22.8845 ± 0.4153	0.3329 ± 0.1241	0.2451 ± 0.1771
	0.2	46.1933 ± 5.9338	22.4904 ± 0.3651	0.4899 ± 0.2778	0.3640 ± 0.1962
	0.0	41.1054 ± 5.1171	21.1961 ± 2.9962	0.3520 ± 0.1881	0.3520 ± 0.1881
Soybean large	0.1	47.4468 ± 18.8684	20.9551 ± 2.4125	0.3752 ± 0.2183	0.3580 ± 0.2016
	0.2	92.1092 ± 23.7040	19.5275 ± 2.0451	0.4695 ± 0.1330	0.2348 ± 0.1209

In Table 3, it can be found that ARIPA-IPR consumes less computing time, and its lower standard deviation of computing time (in ten-fold cross-validation) implies better

robustness than that of the original IPR algorithm. On the other hand, they both own exactly the same stability, as well as the same standard deviation of stability. By borrowing the positive approximation set approach, ARIPA-IPR not only reduces the computation of IPR evidently and enhances its robustness simultaneously, but also holds the same stability as IPR by generating the identical reduct. Similarly, same conclusions can be drawn from Table 4 for the pair of ARIPA-ILCE and ILCE. With regard to the pair of ARIPA-IVPR and IVPR in Table 5, the former half of the above conclusion still holds, and the stability of them are identical if $\beta = 0.0$. This result can be explained reasonably by Theorem 3. While in case of $\beta = 0.1$ or 0.2, ARIPA-IVPR runs more stably than IVPR does. This is because in incomplete variable precision rough set model, the selected reduct (which is with respect to a nonzero β), would become closer to the reduct generated from the universe of condition attributes, when the norm of the lower approximation set of the positive approximation set decreases.

When β varies between 0.0 and 0.5, it can be noticed that the reducts output from our reduction algorithm may be diverse in different cases. This result can be explained through the definition of incomplete variable precision model, i.e., the concerning inclusion degree function is non-monotonic. Although this does not meet our expectation, it is still meaningful because of the following reasons. (1) When the improved reduction algorithm meets its termination condition, the output reduct would definitely contain all the condition attributes that are included in the reduct output from the original reduction algorithm, on the condition that the compressed subset of universe U_i is nonempty. Since the termination condition demands that $\gamma_{red}^{\beta U_i}(D) = \gamma_C^{\beta U_i}(D)$ and $\gamma_{red}^{\beta U}(D) = \gamma_C^{\beta U}(D)$ satisfy simultaneously, both of the reducts output from the original algorithm and the improved algorithm have the same approximation ability. (2) When the compressed subset of universe U_i is empty, the dependence of the selected subset outputted from the improved algorithm would be $\gamma_B^{\beta}(D) = \left(\left| POS_B^{\beta}(D) \right| \Big/ |U| \right) = 1$. Since all of the objects in the universe of discourse U are encapsulated into the lower approximation set with respect to the decision attribute in this case, the improved reduction algorithm, which provides us with a more satisfying option, would have a better approximation capability than the original one.

5. Conclusions

To address the disadvantage of conventional methods of attribute reduction for incomplete decision table in the aspect of computational efficiency, the concept of a positive approximation set based on a tolerance relation is introduced. Additionally, the rank reservation property of the condition attribute significance is discussed, and it is employed to accelerate other existing reduction algorithms under various heuristic strategies. As a result, a novel accelerating reduction approach for IDT using positive approximation set (ARIPA) is proposed. Several state-of-the-art reduction algorithms in different rough set models are accelerated by ARIPA. To assess the performances of both improved and original reduction algorithms, a series of experiments utilizing four real-world incomplete data sets are conducted. The results show that the ARIPA-improved algorithm would ensure the output of the same reduct as that from the original reduction algorithm. While the former can finish attribute reduction in a more efficient and maybe a more stable manner, in contrast with the latter. Average computing time of ARIPA-IPR, ARIPA-ILCE, and ARIPA-IVPR is cut to 33.32%, 55.21%, and 43.62%, respectively. The proposed approach has been verified distinctly effective for dealing with incomplete data sets with large amounts of objects. However, the question of how to ensure its high efficiency for incomplete data sets with hundreds of thousands of dimensions (condition attributes) is still an unresolved issue left for the future.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Simulation data sets supporting reported results can be found at the link to publicly archived UCI Machine Learning Repository http://archive.ics.uci.edu/ml/index.php (accessed on 6 January 2022).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

RST	rough set theory
IS	information system
IIS	incomplete information system
DT	decision table
IDT	incomplete decision table
POS	positive region
NEG	negative region
BND	boundary region
SIM	similarity relation
SIG	condition attribute significance

Appendix A

Lemma A1. *Let A, B, C, A', B', C' be six finite set, such that we have $A' = A \cup C$ and $B' = B \cup C'$ satisfied. If $A' \subseteq B'$ and $(A \cup B) \cap C' = \varnothing$ satisfy, then we have $A \subseteq B$.*

Proof. Let $a \in A$ satisfy, since $A' = A \cup C$ and $A \subseteq A'$, we have $a \in A'$. Since $A' \subseteq B'$, we can derive $a \in B'$. $(A \cup B) \cap C' = \varnothing$, thus we have $A \cap C' = \varnothing$; furthermore, $a \notin C'$. Since $B' = B \cup C'$, $a \in B'$ and $a \notin C'$, there exists $a \in B$. Finally, we can obtain $A \subseteq B$. QED. $\square$

Lemma A2. *Let the pair $(U, C \cup D)$ be an IDT, such that we have $B \subseteq C$ and $U' = U - \text{POS}_B^U(D)$ satisfied. If $S_{B \cup \{a\}}^U(x') \subseteq S_D^U(x')$ and $x' \in U'$ satisfy, then we have $S_{B \cup \{a\}}^{U'}(x') \subseteq S_D^{U'}(x')$.*

Proof. Since $U' = U - \text{POS}_B^U(D)$ and $x' \in U'$ satisfy, two notations X and Y can be defined as follows.

$$X = \left\{ x \middle| \left(x \in S_{B \cup \{a\}}^U(x') \right) \cap \left(x \in \text{POS}_B^U(D) \right) \right\}$$

$$Y = \left\{ y \middle| \left(y \in S_D^U(x') \right) \cap \left(y \in \text{POS}_B^U(D) \right) \right\}$$

Therefore, we can obtain $S_{B \cup \{a\}}^U(x') = S_{B \cup \{a\}}^{U'}(x') \cup X$ and $S_D^U(x') = S_D^{U'}(x') \cup Y$. According to Y's formula, it can be derived that $Y \subseteq \text{POS}_B^U(D)$. Thus, there exists $Y \cap S_{B \cup \{a\}}^{U'}(x') = \varnothing$ and $Y \cap S_D^{U'}(x') = \varnothing$, i.e., $Y \cap \left(S_{B \cup \{a\}}^{U'}(x') \cup S_D^{U'}(x') \right) = \varnothing$. Then, according to $S_{B \cup \{a\}}^U(x') \subseteq S_D^U(x')$ and Lemma 1, it can be derived that $S_{B \cup \{a\}}^{U'}(x') \subseteq S_D^{U'}(x')$. $\square$

Lemma A3. *Let the pair $(U, C \cup D)$ be an IDT, such that we have $B \subseteq C$ and $U' = U - \text{POS}_B^U(D)$ satisfied. Then, we have $\left|S_B^U(x')\right| - \left|S_B^U(x') \cap S_D^U(x')\right| = \left|S_B^{U'}(x')\right| - \left|S_B^{U'}(x') \cap S_D^{U'}(x')\right|$, where $x' \in U'$.*

Proof. Since $U' = U - \text{POS}_B^U(D)$ and $x' \in U'$, two notations X and Y can be defined as follows.

$$X = \left\{ x \middle| \left(x \in S_B^U(x')\right) \cap \left(x \in \text{POS}_B^U(D)\right) \right\}$$
$$Y = \left\{ y \middle| \left(y \in S_D^U(x')\right) \cap \left(y \in \text{POS}_B^U(D)\right) \right\}$$

Then it can be obtained that $S_B^U(x') = S_B^{U'}(x') \cup X$ and $S_D^U(x') = S_D^{U'}(x') \cup Y$. According to the definitions of X and Y, we can derive that $X \subseteq \text{POS}_B^U(D)$ and $Y \subseteq \text{POS}_B^U(D)$. Thus, there exists $Y \cap S_B^{U'}(x') = \varnothing$ and $X \cap S_D^{U'}(x') = \varnothing$. As for $\forall x \in X$, it can be obtained that $x \in S_B^U(x')$. It can be derived that $x' \in S_B^U(x)$ on the basis of the symmetry of tolerance relation. Furthermore, it can be derived that $S_B^U(x) \subseteq S_D^U(x)$ on the basis of the definition of positive region, hence $x' \in S_D^U(x)$. Similarly, it can be obtained that $x \in S_D^U(x')$. Since for $\forall x \in X$ and $\forall x \in \text{POS}_B^U(D)$, there exists $x \in Y$, i.e., $X \subseteq Y$. Therefore, the following formula can be derived.

$$
\begin{aligned}
S_B^U(x') \cap S_D^U(x') &= \left(S_B^{U'}(x') \cup X\right) \cap \left(S_D^{U'}(x') \cup Y\right) \\
&= \left(S_B^{U'}(x') \cap S_D^{U'}(x')\right) \cup \left(S_B^{U'}(x') \cap Y\right) \cup \left(X \cap S_D^{U'}(x')\right) \cup (X \cap Y) \\
&= \left(S_B^{U'}(x') \cap S_D^{U'}(x')\right) \cup X
\end{aligned}
$$

And since $X \subseteq \text{POS}_B^U(D)$, we can obtain $\left(S_B^{U'}(x') \cap S_D^{U'}(x')\right) \cap X = \varnothing$ and $\left|S_B^U(x') \cap S_D^U(x')\right| = \left|S_B^{U'}(x') \cap S_D^{U'}(x')\right| + |X|$. Therefore, the following formula can be derived.

$$
\begin{aligned}
\left|S_B^{U'}(x')\right| - \left|S_B^{U'}(x') \cap S_D^{U'}(x')\right| &= \left(\left|S_B^U(x')\right| - |X|\right) - \left(\left|S_B^U(x') \cap S_D^U(x')\right| - |X|\right) \\
&= \left|S_B^U(x')\right| - |X| - \left|S_B^U(x') \cap S_D^U(x')\right| + |X| \\
&= \left|S_B^U(x')\right| - \left|S_B^U(x') \cap S_D^U(x')\right|
\end{aligned}
$$

i.e., $\left|S_B^U(x')\right| - \left|S_B^U(x') \cap S_D^U(x')\right| = \left|S_B^{U'}(x')\right| - \left|S_B^{U'}(x') \cap S_D^{U'}(x')\right|$, where $x' \in U'$. $\quad\square$

Appendix B

Theorem A1. *Let the pair $(U, C \cup D)$ be an IDT, such that we have $B \subseteq C$ and $U' = U - \text{POS}_B^U(D)$ satisfied. As for $\forall a, b \in C - B$, if $\text{SIG}_1^{\text{outer}}(a, U, B, D) \geq \text{SIG}_1^{\text{outer}}(b, U, B, D)$ satisfies, then we have*

$$\text{SIG}_1^{\text{outer}}(a, U', B, D) \geq \text{SIG}_1^{\text{outer}}(b, U', B, D).$$

Proof. According to the definition $\text{SIG}_1^{\text{outer}}(a, B, D) = \gamma_{B \cup \{a\}}(D) - \gamma_B(D)$, it is definite that the value of $\text{SIG}_1^{\text{outer}}(a, B, D)$ relies on the dependence function $\gamma_B(D) = |\text{POS}_B(D)|/|U|$. $U' = U - \text{POS}_B^U(D)$, hence, we can obtain $\text{POS}_B^{U'}(D) = \varnothing$. According to Lemma 2, we can obtain that if $S_{B \cup \{a\}}^U(x') \subseteq S_D^U(x')$, $x' \in U'$ satisfies, then there exists $S_{B \cup \{a\}}^{U'}(x') \subseteq S_D^{U'}(x')$. Therefore, it can be derived that $\text{POS}_{B \cup \{a\}}^{U'}(D) = \text{POS}_{B \cup \{a\}}^U(D) - \text{POS}_B^U(D)$. Furthermore, we can the following:

$$\frac{\mathrm{SIG}_1^{\mathrm{outer}}(a, U, B, D)}{\mathrm{SIG}_1^{\mathrm{outer}}(a, U', B, D)} = \frac{\gamma_{B \cup \{a\}}^U(D) - \gamma_B^U(D)}{\gamma_{B \cup \{a\}}^{U'}(D) - \gamma_B^{U'}(D)}$$

$$= \frac{|U'|}{|U|} \frac{\left|\mathrm{POS}_{B \cup \{a\}}^U(D)\right| - \left|\mathrm{POS}_B^U(D)\right|}{\left|\mathrm{POS}_{B \cup \{a\}}^{U'}(D)\right| - \left|\mathrm{POS}_B^{U'}(D)\right|}$$

$$= \frac{|U'|}{|U|} \frac{\left|\mathrm{POS}_{B \cup \{a\}}^U(D)\right| - \left|\mathrm{POS}_B^U(D)\right|}{\left|\mathrm{POS}_{B \cup \{a\}}^U(D)\right| - \left|\mathrm{POS}_B^U(D)\right|}$$

$$= \frac{|U'|}{|U|}.$$

Since $1 \geq |U'|/|U| \geq 0$, then, if $\mathrm{SIG}_1^{\mathrm{outer}}(a, U, B, D) > \mathrm{SIG}_1^{\mathrm{outer}}(b, U, B, D)$ satisfies, then there exists $\mathrm{SIG}_1^{\mathrm{outer}}(a, U', B, D) > \mathrm{SIG}_1^{\mathrm{outer}}(b, U', B, D)$. □

Theorem A2. *Let the pair* $(U, C \cup D)$ *be an IDT, such that we have* $B \subseteq C$ *and* $U' = U - \mathrm{POS}_B^U(D)$ *satisfied. As for* $\forall a, b \in C - B$, *if* $\mathrm{SIG}_2^{\mathrm{outer}}(a, U, B, D) \geq \mathrm{SIG}_2^{\mathrm{outer}}(b, U, B, D)$, *then there exists*

$$\mathrm{SIG}_2^{\mathrm{outer}}(a, U', B, D) \geq \mathrm{SIG}_2^{\mathrm{outer}}(b, U', B, D).$$

Proof. Let $U/\mathrm{SIM}(B) = \left\{S_B^U(x_1), S_B^U(x_2), \ldots, S_B^U(x_q), S_B^U(x_{q+1}), \ldots, S_B^U\left(x_{|U|}\right)\right\}$ and $U/SIM(D) = \left\{S_D^U(x_1), S_D^U(x_2), \ldots, S_D^U(x_q), S_D^U(x_{q+1}), \ldots, S_D^U\left(x_{|U|}\right)\right\}$ be true, where $x_i \in \mathrm{POS}_B^U(D)$, $i = 1, 2, \ldots, q$. The notation $\mathrm{EN}^U(D|B)$ is used to indicate the rough conditional entropy on the universe of discourse U in Yan's approach [23].

$$\mathrm{EN}^U(D|B) = \frac{1}{|U|^2} \sum_{i=1}^{|U|} \left(\left|S_B^U(x_i)\right| - \left|S_B^U(x_i) \cap S_D^U(x_i)\right|\right)$$

$$= \frac{1}{|U|^2} \sum_{i=1}^{q} \left(\left|S_B^U(x_i)\right| - \left|S_B^U(x_i) \cap S_D^U(x_i)\right|\right) + \frac{1}{|U|^2} \sum_{i=q+1}^{|U|} \left(\left|S_B^U(x_i)\right| - \left|S_B^U(x_i) \cap S_D^U(x_i)\right|\right)$$

$$= \frac{1}{|U|^2} \sum_{i=1}^{q} \left(\left|S_B^U(x_i)\right| - \left|S_B^U(x_i)\right|\right) + \frac{1}{|U|^2} \sum_{i=q+1}^{|U|} \left(\left|S_B^U(x_i)\right| - \left|S_B^U(x_i) \cap S_D^U(x_i)\right|\right)$$

$$= \frac{1}{|U|^2} \sum_{i=q+1}^{|U|} \left(\left|S_B^U(x_i)\right| - \left|S_B^U(x_i) \cap S_D^U(x_i)\right|\right)$$

In addition, the following equation can be derived according to Lemma A3

$$\frac{1}{|U|^2} \sum_{i=q+1}^{|U|} \left(\left|S_B^U(x_i)\right| - \left|S_B^U(x_i) \cap S_D^U(x_i)\right|\right) = \frac{|U'|^2}{|U|^2} \frac{1}{|U'|^2} \sum_{i=q+1}^{|U|} \left(\left|S_B^U(x_i)\right| - \left|S_B^U(x_i) \cap S_D^U(x_i)\right|\right)$$

$$= \frac{|U'|^2}{|U|^2} \frac{1}{|U'|^2} \sum_{j=1}^{|U'|} \left(\left|S_B^{U'}(x_j)\right| - \left|S_B^{U'}(x_j) \cap S_D^{U'}(x_j)\right|\right)$$

$$= \frac{|U'|^2}{|U|^2} \mathrm{EN}^{U'}(D|B)$$

Therefore, there exists

$$\frac{\mathrm{SIG}_2^{\mathrm{outer}}(a, U, B, D)}{\mathrm{SIG}_2^{\mathrm{outer}}(a, U', B, D)} = \frac{|U'|^2}{|U|^2}.$$

Hence, for $\forall a, b \in C - B$, if $\text{SIG}_2^{\text{outer}}(a, U, B, D) \geq \text{SIG}_2^{\text{outer}}(b, U, B, D)$ satisfies, then there exists $\text{SIG}_2^{\text{outer}}(a, U', B, D) \geq \text{SIG}_2^{\text{outer}}(b, U', B, D)$. □

Theorem A3. *Let the pair $(U, C \cup D)$ be an IDT, such that we have $B \subseteq C$, $U' = U - \text{POS}_B^U(D)$ and $\beta = 0$ satisfied. As for $\forall a, b \in C - B$, if $\text{SIG}_3^{\text{outer}}(a, U, B, D) \geq \text{SIG}_3^{\text{outer}}(b, U, B, D)$ satisfies, there exists $\text{SIG}_3^{\text{outer}}(a, U', B, D) \geq \text{SIG}_3^{\text{outer}}(b, U', B, D)$.*

Proof. Omitted because of its similarity to the proof of Theorem A1. □

References

1. Pawlak, Z. Rough sets. *Int. J. Comput. Inf. Sci.* **1982**, *11*, 341–356. [CrossRef]
2. Zhu, L.Z.; Zhang, S.N.; Ma, Q.; Zhao, H.C.; Chen, S.; Wei, D.X. Classification of UAV-to-ground targets based on enhanced micro-doppler features extracted via PCA and compressed sensing. *IEEE Sens. J.* **2020**, *20*, 14360–14368. [CrossRef]
3. Rizvi, D.; Ahmad, S.; Khan, K.; Hasan, A.; Masood, A. A deep learning approach for fixed and rotary-wing target detection and classification in radars. *IEEE Aerosp. Electron. Syst. Mag.* 2022. *Early Access*. [CrossRef]
4. Maldonado, S.; Merigó, J.; Miranda, J. IOWA-SVM: A density-based weighting strategy for SVM classification via OWA operators. *IEEE Trans. Fuzzy Syst.* **2020**, *28*, 2143–2150. [CrossRef]
5. Kumar, A.; Prasad, P.S. Scalable fuzzy rough set reduct computation using fuzzy min–max neural network preprocessing. *IEEE Trans. Fuzzy Syst.* **2020**, *28*, 953–964. [CrossRef]
6. Xia, S.Y.; Zhang, H.; Li, W.H.; Wang, G.Y.; Giem, E.; Chen, Z.Z. GBNRS: A novel rough set algorithm for fast adaptive attribute reduction in classification. *IEEE Trans. Fuzzy Syst.* **2022**, *34*, 1231–1242. [CrossRef]
7. Hu, Q.H.; Zhang, L.J.; Zhou, Y.C.; Pedrycz, W. Large-scale multimodality attribute reduction with multi-kernel fuzzy rough sets. *IEEE Trans. Fuzzy Syst.* **2018**, *26*, 226–238. [CrossRef]
8. Kohavi, R.; John, G.H. Wrappers for feature subset selection. *Artif. Intell.* **1997**, *97*, 273–324. [CrossRef]
9. Yang, J.; Wang, G.Y.; Zhang, Q.H.; Wang, H.M. Knowledge distance measure for the multigranularity rough approximations of a fuzzy concept. *IEEE Trans. Fuzzy Syst.* **2020**, *28*, 706–717. [CrossRef]
10. Lee, C.; Lee, G. Information gain and divergence-based feature selection for machine learning-based text categorization. *Inf. Process. Manag.* **2006**, *42*, 155–165. [CrossRef]
11. Slezak, D. Degrees of conditional (in) dependence: A framework for approximate Bayesian networks and examples related to the rough set-based feature selection. *Int. J. Approx. Reason.* **2009**, *179*, 197–209. [CrossRef]
12. Szelag, M.; Greco, S.; Slowinski, R. Variable consistency dominance-based rough set approach to preference learning in multicriteria ranking. *Inform. Sci.* **2014**, *277*, 525–552. [CrossRef]
13. Dai, J.H.; Xu, Q. Attribute selection based on information gain ratio in fuzzy rough set theory with application to tumor classification. *Appl. Soft. Comput.* **2013**, *13*, 211–221. [CrossRef]
14. Jensen, R.; Shen, Q. Semantics-preserving dimensionality reduction: Rough and fuzzy-rough-based approaches. *IEEE Trans. Knowl. Data Eng.* **2004**, *16*, 1457–1471. [CrossRef]
15. Skowron, A. Extracting laws from decision tables—A rough set approach. *Comput. Intell.* **1995**, *11*, 371–388. [CrossRef]
16. Kryszkiewicz, M. Rough set approach to incomplete information systems. *Inform. Sci.* **1997**, *112*, 39–49. [CrossRef]
17. Shu, W.H.; Shen, H. A Rough-Set Based Incremental Approach for Updating Attribute Reduction Under Dynamic Incomplete Decision Systems. In Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Hyderabad, India, 7–10 July 2013.
18. Shu, W.H.; Shen, H. Incremental feature selection based on rough set in dynamic incomplete data. *Pattern Recogn.* **2014**, *47*, 3890–3906. [CrossRef]
19. Shu, W.H.; Qian, W.B. An incremental approach to attribute reduction from dynamic incomplete decision systems in rough set theory. *Data Knowl. Eng.* **2015**, *100*, 116–132. [CrossRef]
20. Qian, W.B.; Shu, W.H. Mutual information criterion for feature selection from incomplete data. *Neurocomputing* **2015**, *168*, 210–220. [CrossRef]
21. Jin, Y.; Li, Y.; He, Q. A Fast Positive-Region Reduction Method Based on Dominance-Equivalence Relations. In Proceedings of the International Conference on Machine Learning and Cybernetics (ICMLC), Jeju, Korea, 10–13 July 2016.
22. Yan, T.; Han, C.Z. A novel approach of rough conditional entropy-based attribute selection for incomplete decision system. *Math. Probl. Eng.* **2014**, *14*, 1–15. [CrossRef]
23. Yan, T.; Han, C.Z. Entropy Based Attribute Reduction Approach for Incomplete Decision Table. In Proceedings of the International Conference on Information Fusion (FUSION), Xi'an, China, 10–13 July 2017; pp. 947–954.
24. Xie, X.J.; Qin, X.L. A novel incremental attribute reduction approach for dynamic incomplete decision systems. *Int. J. Approx. Reason.* **2018**, *93*, 443–462. [CrossRef]
25. Ma, Y.Y.; Luo, X.Y.; Li, X.L.; Bao, Z.K.; Zhang, Y. Selection of rich model steganalysis features based on decision rough set alpha-positive region reduction. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *29*, 336–350. [CrossRef]

26. Jing, Y.; Li, T.; Fujita, H.; Yu, Z.; Wang, B. An incremental attribute reduction approach based on knowledge granularity with a multi-granulation view. *Inf. Sci.* **2017**, *411*, 23–38. [CrossRef]
27. Jing, Y.G.; Li, T.R.; Fujita, H.; Wang, B.L.; Cheng, N. An incremental attribute reduction method for dynamic data mining. *Inf. Sci.* **2018**, *465*, 202–218. [CrossRef]
28. Sun, L.; Wang, L.Y.; Ding, W.P.; Qian, Y.H.; Xu, J.C. Feature selection using fuzzy neighborhood entropy-based uncertainty measures for fuzzy neighborhood multigranulation rough sets. *IEEE Trans. Fuzzy Syst.* **2021**, *29*, 19–33. [CrossRef]
29. Li, M.Z.; Wang, G.Y. Approximate concept construction with three-way decisions and attribute reduction in incomplete contexts. *Knowl.-Based Syst.* **2016**, *91*, 165–178. [CrossRef]
30. Ni, P.; Zhao, S.Y.; Wang, X.Z.; Chen, H.; Li, C.P. PARA: A positive-region based attribute reduction accelerator. *Inf. Sci.* **2019**, *503*, 533–550. [CrossRef]
31. Liu, K.Y.; Yang, X.B.; Fujita, H.; Liu, D.; Yang, X.; Qian, Y.H. An efficient selector for multi-granularity attribute reduction. *Inf. Sci.* **2019**, *505*, 457–472. [CrossRef]
32. Jiang, Z.H.; Yang, X.B.; Yu, H.L.; Liu, D.; Wang, P.X.; Qian, Y.H. Accelerator for multi-granularity attribute reduction. *Knowl.-Based Syst.* **2019**, *177*, 145–158. [CrossRef]
33. Thuy, N.N.; Wongthanavasu, S. An efficient stripped cover-based accelerator for reduction of attributes in incomplete decision tables. *Expert Syst. Appl.* **2020**, *143*, 1–15. [CrossRef]
34. Ding, W.P.; Pedrycz, W.; Triguero, I.; Cao, Z.H.; Lin, C.T. Multigranulation supertrust model for attribute reduction. *IEEE Trans. Fuzzy Syst.* **2021**, *29*, 1395–1408. [CrossRef]
35. Liang, J.Y.; Shi, Z.; Li, D.; Wierman, M.J. Information entropy, rough entropy and knowledge granulation in incomplete information systems. *Int. J. Gen. Syst.* **2006**, *35*, 641–654. [CrossRef]
36. Kang, X.P.; Miao, D.Q. A variable precision rough set model based on the granularity of tolerance relation. *Knowl.-Based Syst.* **2016**, *102*, 103–115. [CrossRef]
37. Qian, Y.H.; Liang, J.Y.; Dang, C.Y. Knowledge structure, knowledge granulation and knowledge distance in a knowledge base. *Int. J. Approx. Reason.* **2009**, *50*, 174–188. [CrossRef]
38. Tsai, C.J.; Lee, C.I.; Yang, W.P. A discretization algorithm based on class-attribute contingency coefficient. *Inf. Sci.* **2008**, *178*, 714–731. [CrossRef]
39. Sanjeev, A.; Boaz, B. *Computational Complexity: A Modern Approach*; Cambridge University Press: New York, NY, USA, 2009; pp. 154–196.