Fusing attribute reduction accelerators

Yan Chen[a,f], Xibei Yang[a,b,f,g,*], Jinhai Li[c], Pingxin Wang[d,g], Yuhua Qian[e]

[a]*School of Computer, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu, 212100, China*
[b]*School of Economics and Management, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu, 212100, China*
[c]*Data Science Research Center and the Faculty of Science, Kunming University of Science and Technology, Kunming, Yunnan, 650500, China*
[d]*School of Science, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu, 212100, China*
[e]*Institute of Big Data Science and Industry, Shanxi University, Taiyuan, Shanxi, 030006, China*
[f]*Intelligent Information Processing Key Laboratory of Shanxi Province, Shanxi University, Taiyuan, Shanxi, 030006, China*
[g]*Key Laboratory of Oceanographic Big Data Mining & Application of Zhejiang Province, Zhejiang Ocean University, Zhoushan, Zhejiang, 316022, China*

Abstract

In the fields of rough set and machine learning, attribute reduction has been demonstrated to be effective in removing redundant attributes with clear explanations. Therefore, not only the generalization performances of the derived reducts, but also the efficiencies of searching reducts have drawn much attention. Immediately, various accelerators for quickly deriving reducts have been designed. However, most of the existing solutions merely speed up the procedure of searching reduct from one and only one perspective, it follows that the efficiencies of those accelerators may be further improved with a fusion view. For such a reason, a framework called Fusing Attribute Reduction Accelerators (FARA) is developed. Our framework is specifically characterized by the following three aspects: (1) sample based accelerator, which is realized by gradually reducing the volume of samples based on the mechanism of positive approximation; (2) attribute based accelerator, which is realized by adding multiple qualified attributes into the potential reduct for each iteration; (3) granularity based accelerator, which is realized by ignoring the candidate attributes within coarser granularity. By examining both the efficiencies of the searchings and the effectiveness of the searched reducts, comprehensive experiments over 20 public datasets fairly validated the superiorities of our framework against 5 popular accelerators.

Keywords: Accelerator; Attribute reduction; Granularity; Rough set

1. Introduction

Attribute reduction [2, 6, 8, 16, 17, 45], originally suggested by Pawlak [23] in the field of rough set [14, 35, 36, 38, 44], is a competent mechanism for performing dimensional reduction and feature selection. From the viewpoint of Granular Computing [11, 21, 24], attribute reduction is generally realized based on the results of information granulation [9, 10, 20, 30, 41]. Therefore, most of the constraints used in defining attribute reduction are closely related to the information granulation based measures, and then the form of attribute reduction can be given by considering the variations of those measures.

Besides the explorations of concepts related to attribute reduction, how to derive qualified reducts has also been widely addressed. Up to now, exhaustion and heuristics based searchings are two well-established strategies. Though exhaustion based searching can figure out all the reducts over a given data, it is quite time-consuming. Therefore, heuristics based searching has been paid much attention to for its lower complexity. Currently, most of the heuristics based searchings possess a similar structure: evaluate candidate attributes for selecting qualified attributes, and then add them into the potential reduct until the intended constraint in attribute reduction is satisfied.

*Corresponding author.
Email addresses: chenyan@stu.just.edu.cn (Yan Chen), jsjxy_yxb@just.edu.cn (Xibei Yang), jhlixjtu@163.com (Jinhai Li), pingxin_wang@hotmail.com (Pingxin Wang), jinchengqyh@126.com (Yuhua Qian)

Though heuristic searching, especially the forward greedy searching can be used to acquire reducts, the time consumption may also be higher with the dramatically increasing of the volume of data. For such a reason, many valuable accelerators have been designed [5, 15]. To the best of our knowledge, these accelerators can be categorized into the following three aspects: (1) accelerator based on sample; (2) accelerator based on attribute; (3) accelerator based on granularity. Obviously, these popular accelerators can only speed up the process of searching reduct based on one and only one local perspective. For example, sample based accelerators [12, 27] generally reduce the times of scanning samples or the frequencies of pairwise comparisons among samples; attribute based accelerators [5, 22] usually reduce the requirements of evaluating candidate attributes; granularity based accelerators [7, 17, 28] aim to reduce the searching space by grasping the essence of the granularity.

From the above discussions, it is not difficult to point out that those single-view accelerators may have their inherent limitations. Sample based accelerators do not take the times of evaluating candidate attributes into account, it follows that the iterations for completing evaluations and determining appropriate attributes in such type of accelerators are the same as those in conventional forward greedy searching. Attribute based accelerators do not take the volume of samples into consideration, it follows that the whole samples should also be scanned when evaluating the candidate attributes. Granularity based accelerators fail to consider the elapsed time of calculating the value of granularity, while such step may lead to extra computational cost in the process of deriving reducts.

Motivated by further improving the efficiencies of searching reducts, a framework called Fusing Attribute Reduction Accelerators (FARA) will be introduced into the procedure of the well-known forward greedy searching. Our FARA is actually constructed by considering the acceleration perspectives of sample, attribute and granularity, simultaneously. In what follows, such a framework will be elaborated.

(1) Accelerator based on sample. Qian et al. [26] proposed the mechanism called positive approximation, which is realized by gradually reducing the volume of samples without altering the determination of the appropriate attributes. In view of this, the efficiency of evaluating candidate attributes can be improved. We will introduce such material into our FARA invariantly.

(2) Accelerator based on attribute. Different from conventional forward greedy searching, two or more qualified attributes instead of only one attribute will be determined and added into the potential reduct. Such an action can be performed by considering the inherent characteristics of rough set. Immediately, the number of the iterations required for constructing reduct will be reduced, which may contribute to the improvement of the efficiency.

(3) Accelerator based on granularity. Candidate attributes with coarser granularity will be ignored in each iteration of evaluating attributes. From this point of view, through comparing the values of granularity over different attributes, the searching space of the candidate attributes can be further compressed, it follows that the elapsed time of deriving reduct can also be saved.

To sum up, the structure of our FARA will be described in the following Fig. 1.

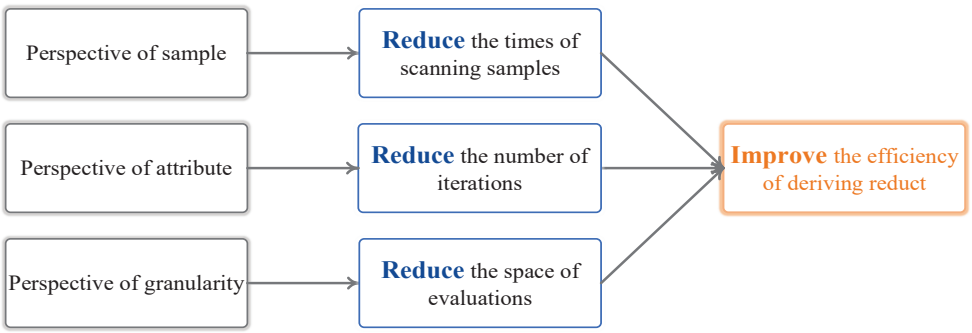


Figure 1: The structure of FARA.

The remainder of this paper is organized as follows. In Section 2, some preliminary knowledge about rough set, granularity and attribute reduction will be briefly described. In Section 3, our proposed framework of FARA is presented. In Section 4, comparative experimental results and the corresponding analyses are elaborated over both UCI and genetic datasets. Finally, this paper is ended with conclusions and future plans in Section 5.

2. Preliminaries

2.1. Rough set

In rough set theory [1, 3, 4, 19, 25, 31], a decision system can be formed as DS=$< U, AT, d >$, in which U is a nonempty finite set of samples, AT is a nonempty finite set of conditional attributes and d is the decision attribute. $\forall x \in U$ and $\forall a \in AT$, $a(x)$ indicates the value of x over conditional attribute a, $d(x)$ indicates the value of x over decision attribute, i.e., the label of x.

In most of the rough set approaches [33, 34, 37, 39, 43], the binary relation induced over the conditional attributes has been widely accepted as the basic method for performing information granulation. Some detailed examples are illustrated as follows.

(1) $\forall A \subseteq AT$, if the values of samples over attributes in A are categorical, then an indiscernibility relation can be defined as $\theta(A) = \{(x, y) \in U \times U : a(x) = a(y), \forall a \in A\}$, which is an equivalence relation.

(2) $\forall A \subseteq AT$, if the values of samples over attributes in A are numeric, then either the fuzzy relation or the neighborhood relation can be constructed, which are representative examples of the parameterized binary relation. For instance, neighborhood relation can be given by $\theta(A) = \{(x, y) \in U \times U : \triangle_A(x, y) \leq \delta\}$, in which δ is a parameter called radius, $\triangle_A(x, y)$ is the distance between samples x and y over A.

Based on the result of information granulation, the degree of information granulation can be characterized by the concept called granularity. Following different views of performing information granulation, granularity can also be equipped with different explanations.

(1) *Attribute based granularity.* For instance, if the information granulation is conducted by deriving the indiscernibility relation over the attributes subset, then the corresponding granularity implies the distinguishability among samples over such an attributes subset. Therefore, different attributes subsets with different distinguishabilities may contribute to different granularities.

(2) *Parameter based granularity.* For example, if the information granulation is conducted by deriving the parameterized binary relation, then the corresponding granularity indicates the distinguishability among samples over such a parameter. Therefore, different values of parameters with different distinguishabilities can also contribute to different granularities.

To sum up, a representative definition of granularity is shown as follows.

Definition 1. *Given a decision system* DS, $\forall A \subseteq AT$, *the granularity related to A is defined as:*

$$\mathbb{G}_A = \frac{|\theta(A)|}{|U|^2}, \tag{1}$$

in which $|X|$ denotes the cardinality of set X.

Notably, $\theta(A)$ can be applied to describe different binary relations. Following Eq. (1), it is not difficult to observe that such a granularity earnestly characterizes the distinguishability among samples. Immediately, the smaller the value of granularity, the stronger the discriminating ability among samples. Obviously, $1/|U| \leq \mathbb{G}(A) \leq 1$ holds.

Definition 2. *Given a decision system* DS, $\forall A \subseteq AT$, $\forall X \subseteq U$, *the lower and upper approximations of X related to A are defined as:*

$$\underline{R}X = \{x \in U : \theta_A(x) \subseteq X\}; \tag{2}$$

$$\overline{R}X = \{x \in U : \theta_A(x) \cap X \neq \emptyset\}; \tag{3}$$

in which $\theta_A(x)$ is the information granule over x for each $x \in U$, i.e., $\theta_A(x) = \{y \in U : (x, y) \in \theta(A)\}$.

Lower approximation $\underline{R}X$ is the set in which samples can be classified with full certainty as members of X; upper approximation $\overline{R}X$ is the set in which samples can be possibly classified as members of X.

2.2. Attribute reduction

With the rapid developments of practical applications, not only the scale, but also the dimensionality of data has grown dramatically. Therefore, modern data analysis and decision making [32] are facing with the huge challenges if the number of attributes is enormous.

Fortunately, attribute reduction in the field of rough set has been demonstrated as an efficient way for reducing the dimensionality of data or removing the redundant attributes from data. Up to now, though various forms of attribute reduction have been proposed, most of them possess a similar structure. Therefore, Yao et al. [42] have presented a general form for extracting the commonness of attribute reductions, which is shown in the following Def. 3.

Definition 3. *Given a decision system* DS, C_ρ *is a constraint based on measure* ρ, $\forall A \subseteq AT$, A *is referred to as a* C_ρ*-reduct if and only if:*

(1) A *satisfies the constraint* C_ρ*;*

(2) $\forall A' \subset A$, A' *does not satisfy the constraint* C_ρ.

In Def. 3, ρ can be regarded as a function such that $\rho : 2^U \times 2^{AT} \to \mathbb{R}$, $\mathbb{R}$ is the set of all real numbers. For example, $\forall x \in U$ and $\forall a \in AT$, if the lower approximation is considered, then $\rho(x, a) = 1$ or 0 indicates that x is in or out of the lower approximation of one target X by the binary relation $\theta(\{a\})$. Furthermore, suppose that $A \subseteq AT$, $\rho(x, A) = 1$ or 0 indicates that x is in or out of the lower approximation of one target X by using the binary relation $\theta(A)$. Finally, suppose that $\forall U' \subseteq U$, $\rho(U', A)$ can be considered as a fusion of $\{\rho(x, A) : \forall x \in U'\}$ and it is denoted as

$$\rho(U', A) = \bigodot_{x \in U'} \rho(x, A)$$

in the context of this paper.

For example, if $\bigodot$ is specified by $\sum$, then $\rho(U', A) = \sum_{x \in U'} \rho(x, A)$. Such $\rho(U', A)$ can be used to acquire the value of approximation quality. Different forms of $\bigodot$, different explanations and values of $\rho(U', A)$ can be acquired.

It should be emphasized that the reduct shown in Def. 3 is the minimum attribute subset of AT, which can fulfill the requirement related to the given constraint C_ρ. The first condition indicates the joint sufficiency of the attributes set A while the second condition implies that each attribute in A is individually necessary.

Following Def. 3, how to quickly acquire a qualified reduct is a challenge, and then various searching strategies have been designed. Among these strategies, forward greedy searching has been favored by many researchers because the lower time complexity. The detailed process of such searching will be presented in the following Algorithm 1.

Algorithm 1: Forward greedy searching.

Input: Decision system DS, a constraint C_ρ.
Output: One C_ρ-reduct A.

```
A = ∅;
Calculate the measure-value such that ρ(U, AT) = ⊙_{x∈U} ρ(x, AT);
Repeat
    ∀a ∈ AT − A, evaluate a by calculating ρ(U, A ∪ {a});
    Select a qualified attribute b ∈ AT − A with a criterion;
    A = A ∪ {b};
    Calculate ρ(U, A);
Until C_ρ is satisfied;
Repeat
    ∀c ∈ A, compute ρ(U, A − {c});
    If C_ρ is satisfied
        A = A − {c};
    End
Until A does not change or |A| = 1;
Return A.
```

The procedure of Algorithm 1 includes two main steps. One step is the process of adding qualified attributes into the potential reduct until the established constraint is satisfied, and the other is the process of removing redundant attributes from the potential reduct. Obviously, such two steps are strictly consistent with the two conditions shown in Def. 3, respectively. Therefore, Algorithm 1 outputs a C_ρ-reduct.

Generally, in the worst case, each attribute in the conditional attributes set AT is necessary for constructing reduct. Therefore, the times required for evaluating attributes are $\frac{(|AT|+1)\cdot|AT|}{2}$, it follows that the time complexity of Algorithm 1 is $O(|U|^2 \cdot \frac{(|AT|+1)\cdot|AT|}{2})$, i.e., $O(|U|^2 \cdot |AT|^2)$.

3. Fusing accelerators for searching reducts

In Section 2.2, Algorithm 1 is a conventional forward greedy searching. However, the time consumption of Algorithm 1 may still be unacceptable if the volume of data is greater. To fill such a gap, some improved forward greedy searchings have also been developed for further speeding up the corresponding procedures. Generally speaking, most previous accelerators [17] can be categorized into the following three aspects.

(1) *Sample based accelerator*. The key of this type of accelerator is to reduce the times of scanning samples or comparing samples. For example, Liu et al. [18] introduced a hash function into the process of information granulation. By using such a function, all samples in the universe can be mapped into a series of different buckets. Following the inherent characteristics of such mapping, it can be proved that the pairwise comparison should only be conducted over the samples in the same or adjacent buckets instead of those in the whole universe. This is why such a mechanism can speed up the procedure of searching reduct from the perspective of sample.

(2) *Attribute based accelerator.* The key of this type of accelerator is to improve the efficiency of attribute evaluation or attribute searching. For example, Chen et al. [5] introduced the concept of attribute group into the procedure of searching and then presented a general framework for quickly deriving reduct. In such a framework, only the attributes out of those attribute groups which contain at least one attribute in the potential reduct should be evaluated. Immediately, the times which are required for evaluating the candidate attributes can be reduced, and then such an approach does accelerate the procedure of searching reduct from the perspective of attribute.

(3) *Granularity based accelerator.* The key of this type of accelerator is to grasp the essence of granularity: a smaller value of granularity indicates a stronger discriminating ability over samples, and then the attribute with smaller value of granularity may contribute much to the construction of reduct. For example, Rao et al. [28] introduced the dissimilarity between attributes into the process of searching reduct. In their method, the dissimilarity is reflected by the difference between attributes related to granularity, and then the attributes with greater dissimilarity are combined for achieving smaller value of granularity. That is, such combination is highly possible to be equipped with stronger discriminating ability, which may be more appropriate to be added into the potential reduct. Immediately, the elapsed time of searching reduct is expected to be saved through reducing the times of iterations in selecting attributes.

Obviously, the above three strategies only accelerate the procedure of searching from one and only one perspective, i.e., they are single-view accelerator for reduct acquisition. Therefore, some limitations may emerge.

(1) Single-view accelerator is lack of the multi-view based fusion for further improving the efficiency of searching reduct. For such reason, the single-view based approach is unsuitable for designing a general framework which can speed up the procedure of searching reduct universally.

(2) Furthermore, consider a single-view accelerator, it is not always effective for quickly computing reduct. Take Chen et al.'s [5] attribute group based approach as an example, if the structure of the groups of attributes is poor, then their accelerator may be inferior to the naive approach which should evaluate all the candidate attributes. From this point of view, it can be observed that the combination of different accelerators not only covers the shortages of single-view accelerators, but also is highly possible to further improve the efficiency of the searching.

To solve the problems shown in Section 3.1, a Fusing Attribute Reduction Accelerator (FARA) will be developed in the context of this paper. Our FARA simultaneously takes the sample, attribute and granularity based acceleration mechanisms into account, it can be regarded as a general framework of fused accelerators. The detailed illustration will be shown as follows.

3.1. Accelerator based on sample

Following the conventional forward greedy searching shown in Algorithm 1, it is not difficult to reveal: 1) the searching of reduct is closely related to the evaluations of candidate attributes; 2) the evaluations of candidate attributes are frequently realized through using all samples in the whole universe. Therefore, the size of the samples partially determines the efficiency of the evaluations and that of the computation of the reduct. For example, if the volume of samples is larger, then the time consumption of evaluations will also be higher. To fill such a gap, Qian et al. [26] have designed a framework called positive approximation, which aims to gradually reduce the volume of samples without altering the results of the selected attributes. The specific steps of such a framework can be elaborated as follows.

(1) The potential reduct A is initialed to be an empty set.
(2) $\forall a \in AT$, evaluate the candidate attribute a by calculating $\rho(U, A \cup \{a\})$.
(3) Select one qualified attribute $b \in AT$ based on the above evaluations, and then add b into A, go to (7).
(4) Set $U' = \{x \in U : \rho(x, A) \neq \rho(x, AT)\}$.
(5) $\forall a \in AT - A$, evaluate the candidate attribute a by calculating $\rho(U', A \cup \{a\})$.
(6) Select one qualified attribute $b \in AT - A$ based on the evaluations shown in (5), and then add b into A.
(7) If the intended constraint is satisfied, then output reduct A; otherwise, go to (4).

Therefore, by the above step (4), the redundant calculation can be removed because the size of the universe has been reduced as U', and then the time complexity of evaluating attributes can be reduced. This method provides a representative way to speed up the process of deriving reduct from the perspective of sample.

3.2. Accelerator based on attribute

By Algorithm 1, it is also not difficult to reveal: 1) the searching of reduct is closely related to the iterations for selecting qualified attributes; 2) for each iteration, one and only one qualified attribute can be fixed and then be added into the potential reduct. Therefore, the number of the raw attributes may determine the time consumption of the iterations and that of the deriving reduct to some extent. In other words, if the number of candidate attributes is greater, then the elapsed time of obtaining reduct may also be higher. From this point of view, new techniques for selecting the qualified attributes have become a necessity.

Notably, in Algorithm 1, one and only one qualified attribute is selected for each iteration, such case is an important factor for hindering the efficiency of computing reduct. For such a reason, a natural way is to select two or more attributes for each iteration, which may be useful for reducing the times of the required iterations. Immediately, the elapsed time of searching reduct is highly possible to be saved.

From discussions above, how to select two or more qualified attributes and then add them into the potential reduct is an interesting issue to be addressed. Some basic considerations can be suggested as follows.

(1) Randomly select multiple candidate attributes for each iteration and then add them into the potential reduct. Nevertheless, such mechanism is easy to bring us the instabilities of the reducts.
(2) Select multiple candidate attributes for each iteration based on a metric over the attributes and then add them into the potential reduct. However, such mechanism is highly possible to bring us the extra time consumption. This is mainly because the metric computation requires extra time.

To overcome the above limitations, we will propose a new approach which is useful for the selection of multiple appropriate attributes. Such an approach is actually designed by considering the union of lower approximation. The specific details of selecting multiple appropriate attributes will be shown in the following Fig. 2, together with the corresponding explanations.

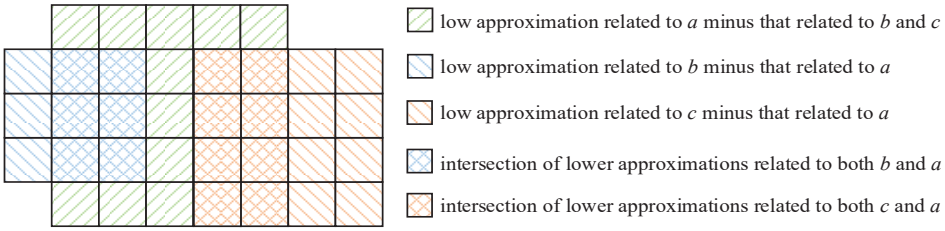


Figure 2: Relationships among lower approximations related to three attributes a, b and c.

In Fig. 2,

(1) assume that the selected attribute in the last round of iteration is a, and then the remained attributes are b and c, which will be evaluated in the current iteration;
(2) obviously, the union of lower approximations related to both attributes c and a is greater than that related to both attributes b and a, and then it is possible for us to achieve the objective of attribute reduction as quickly as possible if both c and a are added into the potential reduct simultaneously.

Therefore, two or more attributes may be selected and added into the potential reduct, it follows that the number of iterations used in forward greedy searching can be reduced, which will contribute to lower time consumption.

3.3. Accelerator based on granularity

In the field of Granular Computing, the discriminating abilities of the attributes have been widely addressed. It has also been accepted that the concept of Granularity is a simple and useful way for characterizing such a discriminating ability. Generally speaking, a finer granularity indicates that the corresponding attribute possesses the stronger discriminating ability while a coarser granularity implies that the corresponding attribute possesses the weaker discriminating ability.

Since the forward greedy searching strategy shown in Algorithm 1 is closely related to the variations of the attributes, it is then not difficult to reveal: 1) all the attributes which are out of the potential reduct should be evaluated in each iteration; 2) some attributes which are out of the potential reduct will possess coarser granularities, it follows that these attributes are highly possible to provide poor discriminating ability and the lower contribution to the objective of the attribute reduction. Therefore, the number of the candidate attributes with coarser granularities may adversely affect the time consumption of the evaluation and that of the computing reduct to some extent. From this point of view, how to reduce the times of evaluating candidate attributes by considering the granularities of those attributes is necessary.

Therefore, a naive way is to ignore some candidate attributes with coarser granularity. The specific steps of such method can be shown as follows.

(1) The potential reduct A is initialed to be an empty set.
(2) $\forall a \in AT - A$, evaluate the candidate attribute a by calculating $\rho(U, A \cup \{a\})$.
(3) Select one qualified attribute $b \in AT - A$ based on the above evaluations, and then add b into A.
(4) If the intended constraint is satisfied, then output reduct A; otherwise, go to (5).
(5) Add the potential candidate attributes into T such that $T = \{a : \forall a \in AT - A, \mathbb{G}_{\{a\}} \leq \mathbb{G}_{\{b\}}\}$.
(6) Select one qualified attribute $b \in T$, and then add attribute b into A.
(7) Repeatedly execute (5)-(6) until the intended constraint of attribute reduction is satisfied.

Following the above discussions, it is not difficult to observe that in step (5), the number of the candidate attributes can be considerably decreased through comparing the values of granularity over different attributes. Immediately, such a mechanism may also be helpful for reducing the elapsed time of deriving reduct.

3.4. Fusing accelerators

Following what have been addressed in Sections 3.1-3.3, a fusing attribute reduction accelerator for calculating reduct will be designed in this section. The detailed algorithm will be shown as follows.

Algorithm 2: Fusing attribute reduction accelerator (FARA).

Input: Decision system DS, a constraint C_ρ.
Output: One C_ρ-reduct A.

$A = \emptyset$;
$\forall a \in AT$, calculate the granularity $\mathbb{G}_{\{a\}}$;
Calculate the measure-value such that $\rho(U, AT) = \bigodot_{x \in U} \rho(x, AT)$;
$k = 1$;
$T = \emptyset$;
Repeat
  If $k = 1$ **then**
    $\forall a \in AT$, evaluate a by calculating $\rho(U, A \cup \{a\})$;
    Select a qualified attribute $b \in AT$ with a criterion;
    $A = A \cup \{b\}$;
    Set $U' = \{x \in U : \rho(x, A) \neq \rho(x, AT)\}$;
    `// Positive approximation approach`
  End
  Else
    Add the potential candidate attributes into T such that $T = \{a : \forall a \in AT - A, \mathbb{G}_{\{a\}} \leq \mathbb{G}_{\{b\}}\}$;
    `// Accelerator based on granularity`
    $\forall a \in T$, evaluate a by calculating $\rho(U', A \cup \{a\})$;
    `// Accelerator based on sample`
    Select a qualified attribute $b \in T$ with a criterion;
    $A = A \cup \{b\}$;
  End
  Sort the attributes in $AT - A$ by a descending order of $\rho(U', \{a\} \cup \{b\})$ where $a \in AT - A$, select the first t attributes and add them into A;
  `// Accelerator based on attribute`
  Calculate $\rho(U, A)$;
Until C_ρ *is satisfied*;
Repeat
  $\forall c \in A$, compute $\rho(U, A - \{c\})$;
  If C_ρ *is satisfied* **then**
    $A = A - \{c\}$;
  End
Until A *does not change or* $|A| = 1$;
Return A.

Following Algorithm 2, the procedure of deriving reduct has been speeded up from three different perspectives. To facilitate the discussion of time complexity, if only the positive approximation approach is used, then in the worst time, the time complexity of Algorithm 2 is $O\Big(|U|^2 \cdot |AT| + |U_1|^2 \cdot (|AT| - 1) + |U_2|^2 \cdot (|AT| - 2) + \cdots + |U_{|AT|-1}|^2 \cdot 1\Big)$, i.e., $O\Big(|U|^2 \cdot |AT| + \sum_{i=1}^{|AT|-1} |U_i|^2 \cdot (|AT| - i)\Big)$; if only the accelerator based the perspective of attribute is used, assuming $\lfloor \frac{|AT|}{t+1} \rfloor = s$, then in the worst time, the time complexity is $O\Big(|U|^2 \cdot (|AT| + (|AT| - (t+1)) + \cdots + (|AT| - s(t+1)))\Big)$; if only the accelerator based on the granularity is used, then in the worst time, the time complexity is $O\Big(|U|^2 \cdot \frac{(|AT|+1) \cdot |AT|}{2}\Big)$. Therefore, the time complexity of Algorithm 2 is $O\Big(|U|^2 \cdot |AT| + \sum_{i=1}^{s} |U_i|^2 \cdot (|AT| - i(t+1))\Big)$. However, the time complexity of Algorithm 1 is $O\Big(|U|^2 \cdot (|AT| + (|AT| - 1) + (|AT| - 2) + \cdots + 1)\Big)$. Obviously, the time complexity of Algorithm 2 is less than that of Algorithm 1. Immediately, our FARA can contribute to reducing the time consumption

of searching reduct significantly.

4. Experimental analysis

4.1. Datasets

To demonstrate the effectiveness of our proposed FARA, a total of 16 UCI datasets and 4 genetic datasets were used for conducting the comparative experiments. The details of these data are presented in the following Tab. 1. All of the experiments were carried out on a personal computer with Windows 10, Intel Core i7-9570H CPU (2.60 GHz) and 8.00 GB memory. The programming language was Matlab R2017b.

Table 1: Datasets description

ID	Datasets	#Samples	#Attributes	#Labels	Domains
1	Breast Cancer Wisconsin (Diagnostic)	569	30	2	UCI
2	Cardiotocography	2126	22	10	UCI
3	Dermatology	366	34	6	UCI
4	Diabetic Retinopathy Debrecen	1151	19	2	UCI
5	Forest Type Mapping	523	27	4	UCI
6	Ionosphere	351	32	2	UCI
7	LSVT Voice Rehabilitation	126	256	2	UCI
8	Musk (Version 1)	476	167	2	UCI
9	QSAR Biodegradation	1055	41	2	UCI
10	Sonar	208	60	2	UCI
11	Synthetic Control Chart Time Series	600	60	6	UCI
12	Statlog (Image Segmentation)	2310	18	7	UCI
13	Steel Plates Faults	1941	33	2	UCI
14	Statlog (Vehicle Silhouettes)	846	18	4	UCI
15	Urban Land Cover	675	147	9	UCI
16	Waveform Database Generator (Version 1)	5000	21	3	UCI
17	Brain Tumor	90	5920	5	Genetic
18	Lung Cancer	203	12600	5	Genetic
19	NCI 60	60	5726	9	Genetic
20	SRBCT	83	2308	4	Genetic

4.2. Experimental setup

In our experiments, both the neighborhood rough set [13] and pseudo-label neighborhood rough set [40] were employed to define the forms of attribute reduction. Note that such two attribute reductions aimed to remove irrelevant or redundant attributes with the constraint of more than 95% of the approximation quality related to the raw attributes [29]. In addition, 20 radii, i.e., 0.02, 0.04, $\cdots$, 0.40 were used in our experiments. These radii were practically used to express the size of the neighborhood. Moreover, 5-fold cross-validation was also used to calculate reducts, and the classification performances of the obtained reducts could be tested over the testing samples.

To compare with our proposed FARA, five acceleration algorithms were reproduced: fast positive region computation algorithm (FPRC) [18], feature selection algorithm based on positive approximation (FSPA) [26], attribute group for attribute reduction (AGAR) [5], searching single granularity reduct based on dissimilarity between attributes (SGDA) [28], and the forward greedy searching strategy (FGS), as shown in Algorithm 1.

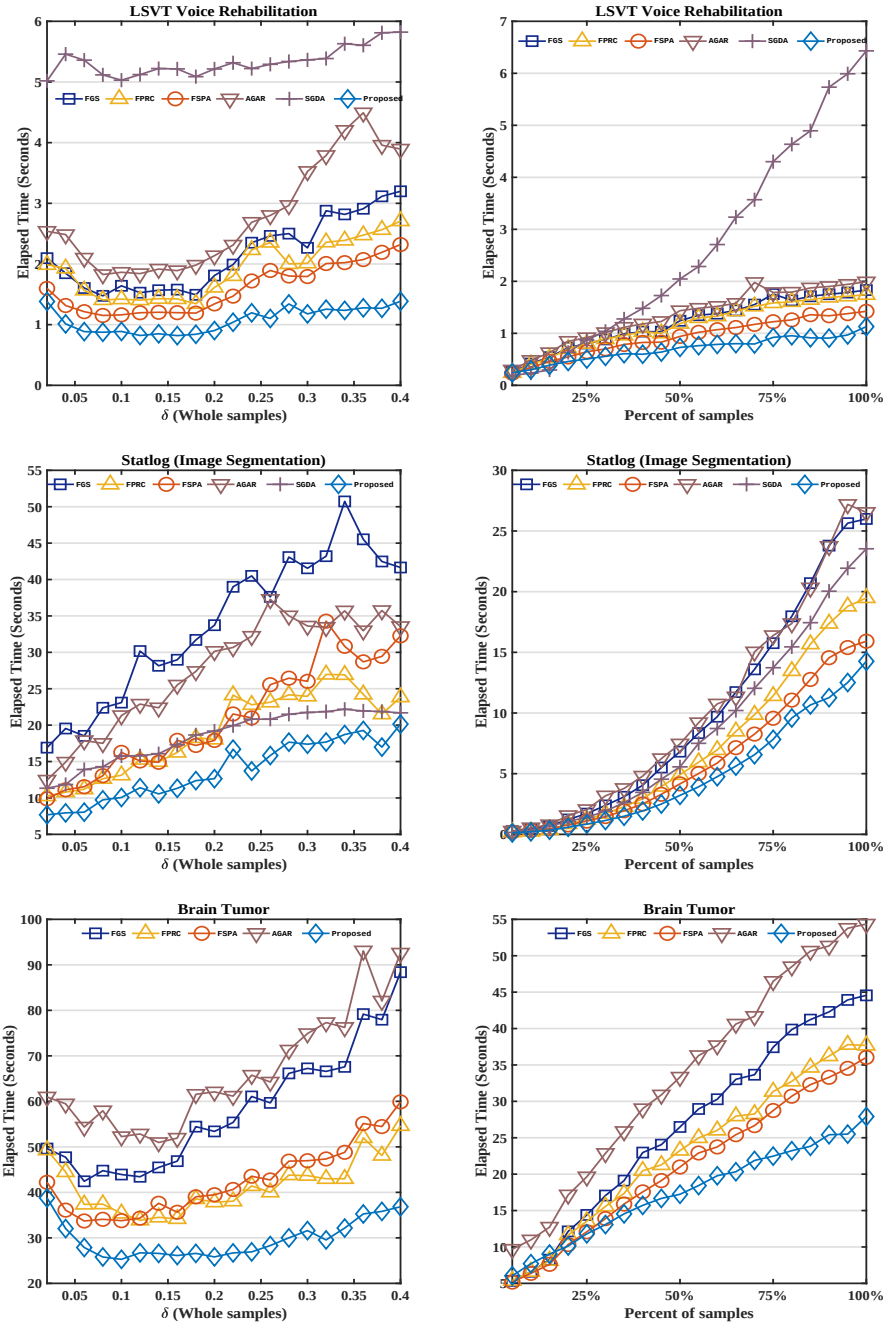
4.3. Comparisons of elapsed time

In this subsection, the elapsed time of obtaining reducts is shown on the left in Figs. 3-4. Note that the right figures display more detailed changing trends of six algorithms with the increasing of samples. For example, for "LSVT Voice Rehabilitation" dataset, the time consumptions of deriving reducts by using six algorithms are compared with the increasing of samples over radius = 0.1.

Additionally, the process of obtaining reduct by using SGDA in genetic datasets is quite time-consuming. For instance, take "Brain Tumor" dataset as an example, the time consumption of obtaining only one reduct is more than 1200 seconds. Correspondingly, the elapsed time of obtaining reduct by using SGDA is significantly greater than those of obtaining reducts by using other accelerators. In view of this, SGDA is not employed to compare with our

proposed FARA and other four strategies over genetic datasets. In other words, six algorithms are compared over 16 UCI datasets, and five algorithms except SGDA are compared over 4 genetic datasets.

To simplify our discussions, the results over four datasets, i.e., "LSVT Voice Rehabilitation", "Statlog (Image Segmentation)", "Brain Tumor" and "Lung Cancer" are taken as examples.



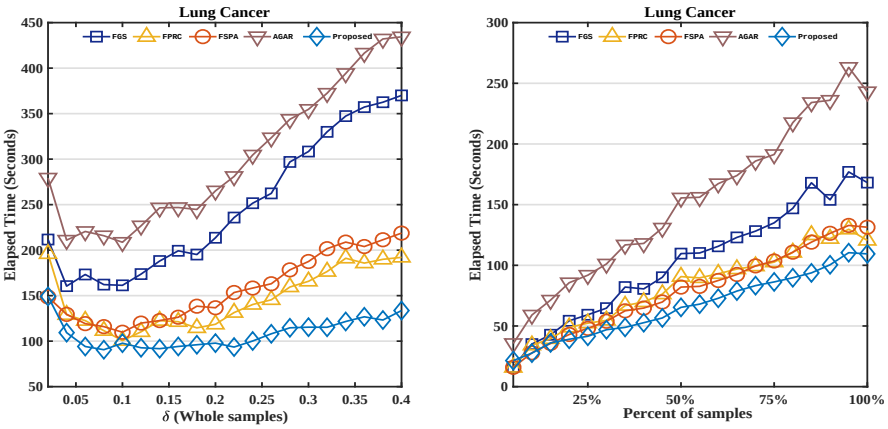


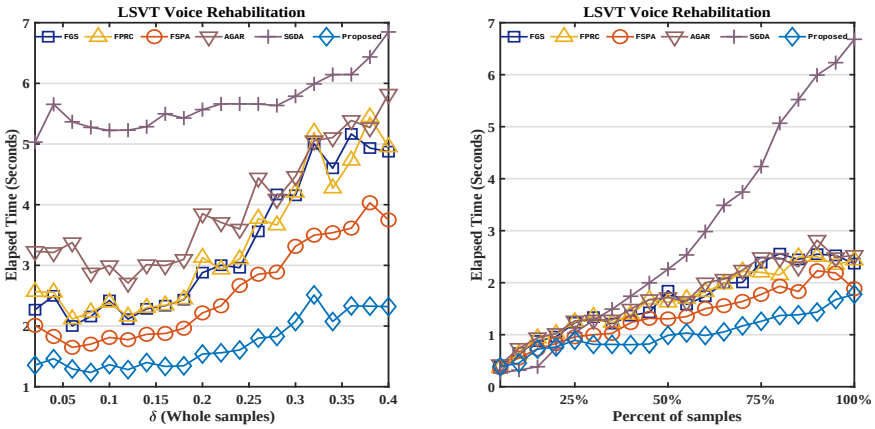
Figure 3: The elapsed time of deriving reducts over neighborhood rough set.

With a careful investigation of the above Fig. 3, it is not difficult to observe the follows.

(1) Through observing the left sub-figures, compared with the elapsed time of deriving reducts by using FGS, FPRC, FSPA and AGAR, the calculations of reducts by using our FARA obviously were less time-consuming. Take "Lung Cancer" dataset as an example, if $\delta = 0.1$, then generating reducts by using FGS, FPRC, FSPA and AGAR costed 213.6074, 118.5567, 136.6739, 265.2528 seconds, respectively. Notably, obtaining reduct by using our FARA only costed 97.8427 seconds. In view of this, our proposed strategy does speed up the process of obtaining reduct.

(2) With the increasing of the value of radius, the increasing trends could be observed for the elapsed time of obtaining reducts by using FGS, FPRC, FSPA, AGAR, SGDA and FARA. This is mainly because the required attributes in the reducts are possible to be increased if the value of radius increases, which may contribute to the requirements of more time.

(3) Through observing the right figures, the elapsed time of six algorithms increased with the increasing of the volumes of samples. Notably, our proposed approach was much more faster than other strategies. Furthermore, the differences were profoundly larger when the volumes of samples increased. Take "Statlog (Image Segmentation)" dataset as an example, if the percentage of the used samples was 5% over the whole universe, then the elapsed time of deriving reducts by using FGS, FPRC, FSPA, AGAR, SGDA and FARA were 0.2581, 0.1781, 0.1950, 0.2707, 0.1580 and 0.1205 seconds, respectively. If all of the samples were used, then the time consumptions of searching reducts by using FGS, FPRC, FSPA, AGAR, SGDA and FARA were 25.9927, 19.4823, 15.9159, 26.5532, 23.5324 and 14.2602 seconds, respectively.



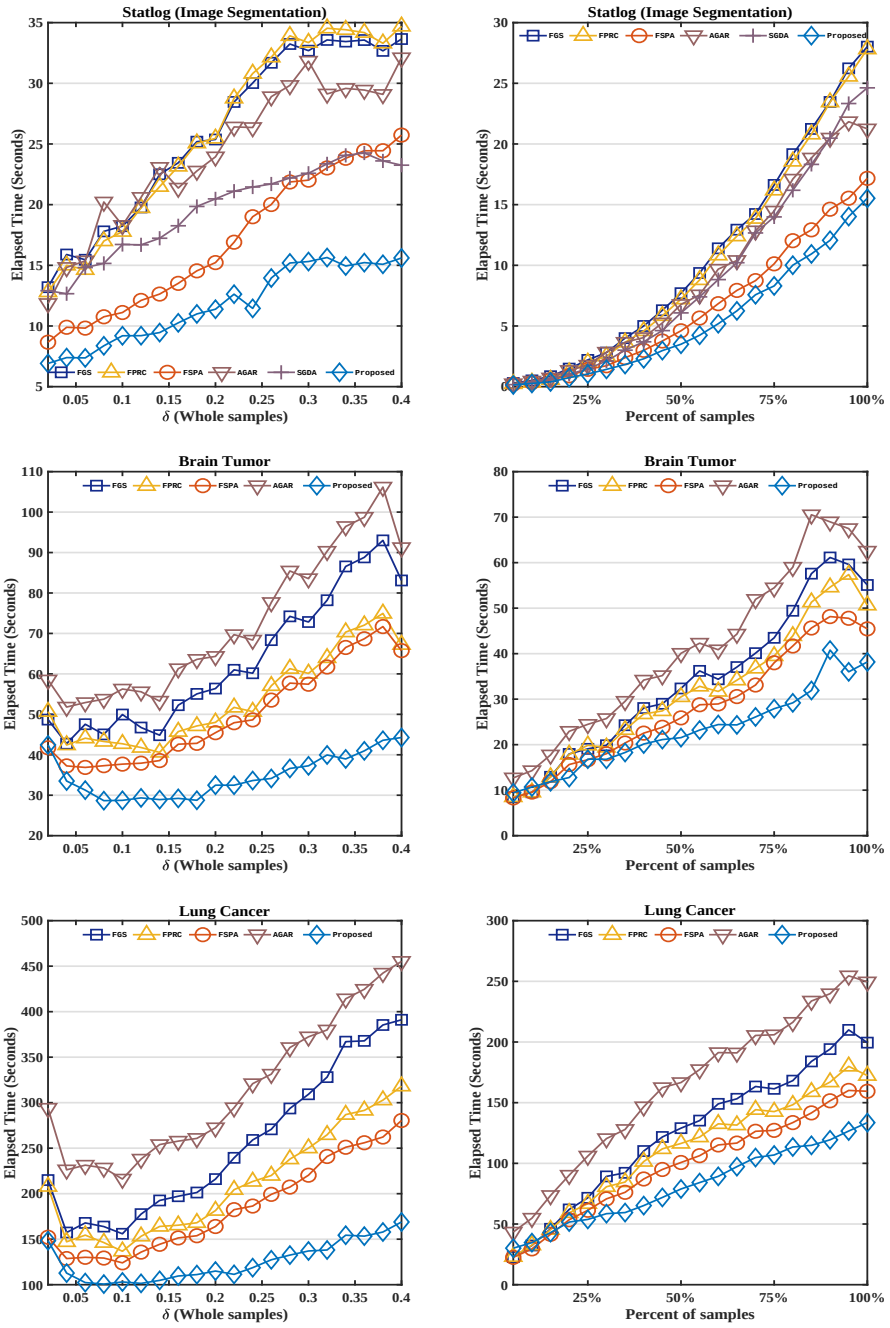


Figure 4: The elapsed time of deriving reducts over pseudo-label neighborhood rough set.

Similar to Fig. 3, Fig. 4 showed that the process of obtaining reduct by using our proposed approach was also superior to other accelerators in time consumption if the pseudo-label neighborhood rough set based attribute reduction was considered. Take the "Brain Tumor" dataset as an example, if $\delta = 0.12$, then to obtain the reducts by using FGS, FPRC, FSPA, AGAR and FARA, 61.0052, 51.7187, 47.9234, 69.7111 and 31.4827 seconds were required, respectively.

To clearly show the advantages of our proposed FARA in time efficiency, the values of speed-up ratio is also presented in the following Tab. 2.

Table 2: The values of speed-up ratio related to the elapsed time of deriving reducts

ID	Neighborhood rough set					Pseudo-label neighborhood rough set				
	FGS & Proposed	FPRC & Proposed	FSPA & Proposed	AGAR & Proposed	SGDA& Proposed	FGS & Proposed	FPRC & Proposed	FSPA & Proposed	AGAR & Proposed	SGDA& Proposed
1	2.5641	2.1590	1.4097	1.5280	2.0317	2.4131	2.6562	1.3353	1.9848	1.9508
2	1.7371	1.2145	1.4210	1.6315	1.1562	1.7445	1.6171	1.4175	1.2016	1.0963
3	3.0299	1.3742	1.3871	1.7945	1.2506	2.5972	1.9060	1.4351	1.5662	1.5856
4	1.7167	1.3900	1.3944	1.3827	*0.8868*	1.7001	1.7827	1.4049	1.2219	*0.9787*
5	2.0812	1.4914	1.4539	1.4576	1.1624	2.0769	2.0824	1.5001	1.6713	1.2807
6	2.2346	1.6753	1.3641	1.3447	1.6269	2.0904	2.0486	1.4122	2.2980	2.0726
7	2.0006	1.7824	1.4781	2.5608	4.9463	1.9333	1.9529	1.5035	2.3003	3.3346
8	2.7660	2.7950	1.5337	2.5473	8.5032	2.7965	3.2579	1.5391	2.1854	6.5000
9	2.4365	2.1435	1.4962	1.9882	2.0366	2.3967	2.5844	1.5228	1.5786	1.7529
10	2.0454	1.5936	1.5304	1.1561	1.5607	1.8394	1.8136	1.4865	1.2272	1.6123
11	2.3451	1.4228	1.5610	1.7833	2.2264	2.2185	2.1040	1.5231	1.4974	2.7110
12	2.4598	1.3865	1.5260	2.0031	1.3359	2.2069	2.2170	1.4414	2.0588	1.6656
13	2.4531	2.1814	1.2501	2.4591	3.9786	2.4222	2.6886	1.3039	2.2129	3.8860
14	1.7387	1.1629	1.3913	1.6700	1.1931	1.7287	1.6558	1.4242	1.1257	1.0058
15	2.2968	1.1053	1.6376	1.6654	3.0305	2.1838	1.8885	1.6070	1.7572	4.1737
16	1.4839	1.3479	1.1809	1.2453	1.1748	1.4843	1.6891	1.1913	1.0753	1.0556
17	1.9525	1.3965	1.4330	2.2244	——	1.8059	1.5475	1.4353	2.0697	——
18	2.2905	1.3504	1.4554	2.7806	——	2.0155	1.6799	1.4748	2.5012	——
19	1.9844	1.2298	1.4886	2.8280	——	1.8812	1.3785	1.5092	2.1291	——
20	1.8566	1.4591	1.4063	2.6095	——	1.7318	1.5860	1.4202	2.2324	——
Average	2.1737	1.5831	1.4399	1.9330	2.3813	2.0633	2.0068	1.4444	1.7948	2.2914

By Tab. 2, it is not difficult to observe that most of the values of speed-up ratio were greater than 1. In other words, our proposed FARA was superior to the other five algorithms in time efficiency in most cases.

To further compare six algorithms from the viewpoint of statistics, the Wilcoxon signed rank test was employed, in which the significance level was appointed as 0.05. That is, if the corresponding p-value was lower than 0.05, then it indicated that algorithms performed significantly different; otherwise, those algorithms performed similarly. The detailed results are shown in the following Tabs. 3-4.

Table 3: p-values for comparing elapsed time of deriving reducts over neighborhood rough set

ID	FGS & Proposed	FPRC & Proposed	FSPA & Proposed	AGAR & Proposed	SGDA & Proposed
1	$< 10^{-4}$	$< 10^{-3}$	*0.0679*	$< 10^{-2}$	$< 10^{-5}$
2	$< 10^{-3}$	*0.1075*	0.0193	$< 10^{-3}$	*0.2853*
3	$< 10^{-7}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-7}$	$< 10^{-3}$
4	$< 10^{-5}$	$< 10^{-3}$	$< 10^{-2}$	$< 10^{-3}$	*0.1199*
5	$< 10^{-5}$	$< 10^{-2}$	$< 10^{-1}$	$< 10^{-2}$	*0.2085*
6	$< 10^{-6}$	$< 10^{-4}$	$< 10^{-3}$	$< 10^{-1}$	$< 10^{-6}$
7	$< 10^{-7}$	$< 10^{-7}$	$< 10^{-3}$	$< 10^{-7}$	$< 10^{-7}$
8	$< 10^{-7}$	$< 10^{-6}$	$< 10^{-3}$	$< 10^{-7}$	$< 10^{-7}$
9	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-1}$	$< 10^{-3}$	$< 10^{-5}$
10	$< 10^{-6}$	$< 10^{-5}$	$< 10^{-4}$	$< 10^{-2}$	$< 10^{-6}$
11	$< 10^{-5}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-4}$	$< 10^{-7}$
12	$< 10^{-6}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-5}$	$< 10^{-2}$
13	$< 10^{-7}$	$< 10^{-7}$	$< 10^{-7}$	$< 10^{-7}$	$< 10^{-7}$
14	$< 10^{-3}$	*0.1895*	$< 10^{-1}$	$< 10^{-3}$	*0.0531*
15	$< 10^{-3}$	*0.5609*	$< 10^{-1}$	$< 10^{-2}$	$< 10^{-7}$
16	$< 10^{-2}$	$< 10^{-1}$	*0.2085*	*0.1075*	*0.1636*
17	$< 10^{-7}$	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-7}$	——
18	$< 10^{-7}$	$< 10^{-3}$	$< 10^{-5}$	$< 10^{-7}$	——
19	$< 10^{-7}$	$< 10^{-3}$	$< 10^{-5}$	$< 10^{-7}$	——
20	$< 10^{-6}$	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-7}$	——

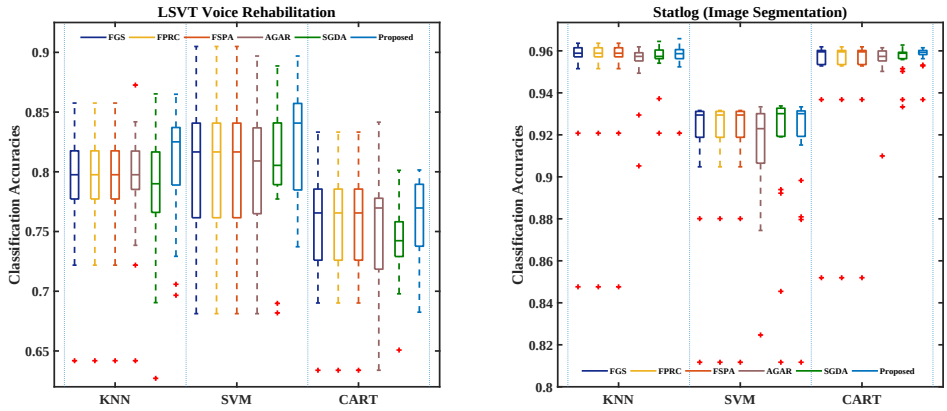
Table 4: p-values for comparing elapsed time of deriving reducts over pseudo-label neighborhood rough set

ID	FGS & Proposed	FPRC & Proposed	FSPA & Proposed	AGAR & Proposed	SGDA & Proposed
1	$< 10^{-4}$	$< 10^{-4}$	*0.0810*	$< 10^{-3}$	$< 10^{-5}$
2	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-1}$	*0.0909*	*0.4570*
3	$< 10^{-7}$	$< 10^{-7}$	$< 10^{-5}$	$< 10^{-6}$	$< 10^{-6}$
4	$< 10^{-4}$	$< 10^{-5}$	$< 10^{-2}$	$< 10^{-2}$	*0.2393*
5	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-1}$	$< 10^{-3}$	$< 10^{-1}$
6	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-3}$	$< 10^{-7}$	$< 10^{-7}$
7	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-3}$	$< 10^{-7}$	$< 10^{-7}$
8	$< 10^{-6}$	$< 10^{-7}$	$< 10^{-3}$	$< 10^{-6}$	$< 10^{-7}$
9	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-1}$	$< 10^{-2}$	$< 10^{-4}$
10	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-5}$	$< 10^{-2}$	$< 10^{-6}$
11	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-2}$	$< 10^{-4}$	$< 10^{-7}$
12	$< 10^{-6}$	$< 10^{-5}$	$< 10^{-2}$	$< 10^{-6}$	$< 10^{-5}$
13	$< 10^{-7}$	$< 10^{-7}$	$< 10^{-7}$	$< 10^{-7}$	$< 10^{-7}$
14	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-1}$	*0.2503*	*0.7150*
15	$< 10^{-3}$	$< 10^{-2}$	$< 10^{-1}$	$< 10^{-2}$	$< 10^{-7}$
16	$< 10^{-2}$	$< 10^{-3}$	*0.1806*	*0.5792*	*0.5979*
17	$< 10^{-7}$	$< 10^{-6}$	$< 10^{-4}$	$< 10^{-7}$	——
18	$< 10^{-6}$	$< 10^{-5}$	$< 10^{-3}$	$< 10^{-7}$	——
19	$< 10^{-5}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-7}$	——
20	$< 10^{-7}$	$< 10^{-7}$	$< 10^{-6}$	$< 10^{-7}$	——

With a deep investigation of Tabs. 3-4, we can draw the following conclusions. For the comparisons of time consumption in obtaining reducts between our approach and other acceleration strategies, the derived p-values were much lower than 0.05 in most cases. Such result implied that whichever neighborhood rough set was constructed, our approach performed significantly different from other accelerators.

4.4. Comparisons of classification performances

In this section, the classification accuracies derived by different types of reducts will be compared. To clearly reveal the results of classification performances of reducts in terms of SVM, KNN and CART classifiers, Box Plot, a statistical chart illustrating dispersion of datasets, was employed. The corresponding results are shown in the following Figs. 5-6.



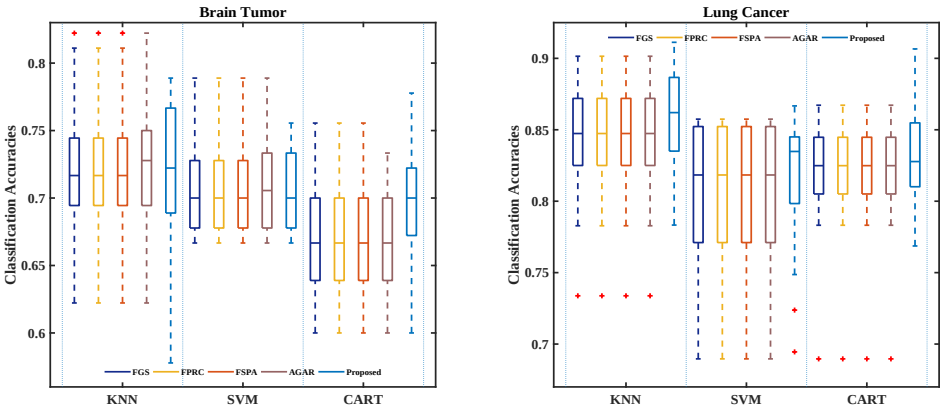


Figure 5: Classification accuracies related to different reducts over neighborhood rough set.

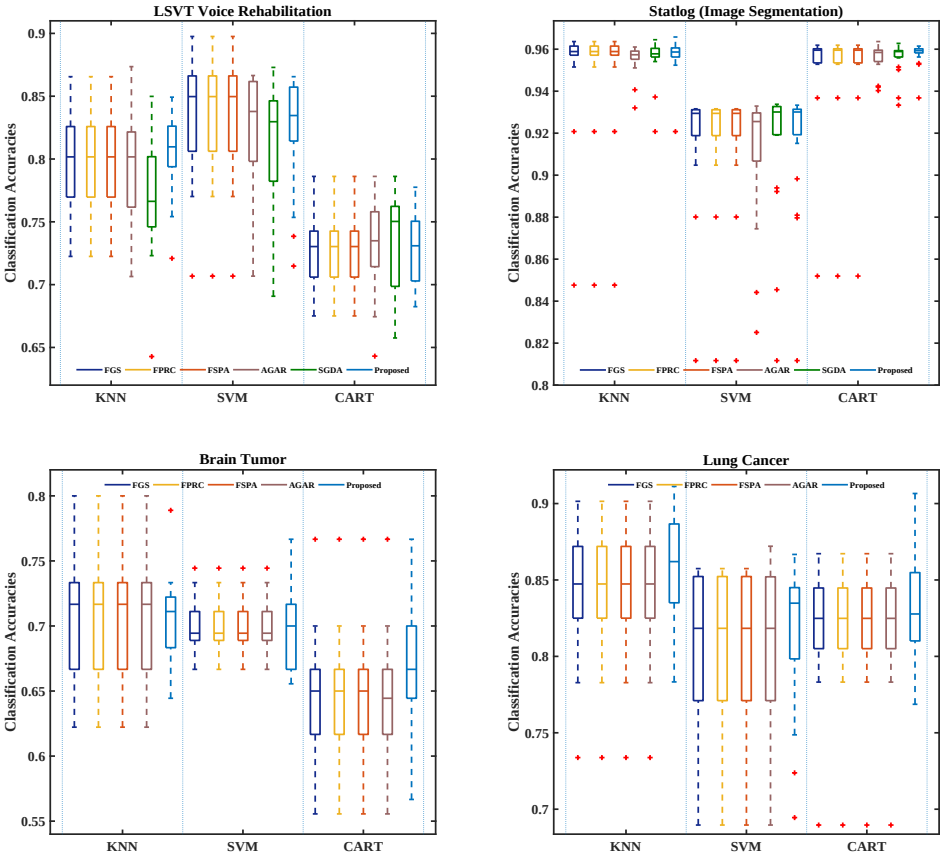


Figure 6: Classification accuracies related to different reducts over pseudo-label neighborhood rough set.

Through observing Figs. 5-6, it was not difficult to observe that whichever classifier was employed, the reducts derived by our proposed strategy offered well-matched classification accuracies. For example, in "Lung Cancer" dataset, the median value of classification accuracies over the reducts related to our FARA was similar to those over the reducts related to FGS, FPRC, FSPA, AGAR and SGDA.

In this experiment, the Wilcoxon signed rank test was also employed to compare the classification accuracies. The detailed results are shown in the following Tabs. 5-7.

Table 5: p-values for comparing classification accuracies based on SVM classifier

ID	Neighborhood rough set					Pseudo-label neighborhood rough set				
	FGS & Proposed	FPRC & Proposed	FSPA & Proposed	AGAR & Proposed	SGDA & Proposed	FGS & Proposed	FPRC & Proposed	FSPA & Proposed	AGAR & Proposed	SGDA & Proposed
1	0.9030	0.9030	0.9030	0.9245	0.6947	0.7971	0.7971	0.7971	0.4015	0.8498
2	0.5792	0.5792	0.5792	0.7972	0.8604	0.3368	0.3368	0.3368	0.9569	0.3720
3	0.5021	0.5021	0.5021	0.0333	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	0.1231	$< 10^{-3}$
4	0.4984	0.4984	0.4984	0.6641	1.0000	0.2913	0.2913	0.2913	0.6750	0.4898
5	0.7454	0.7454	0.7454	0.7972	0.1761	0.8604	0.8604	0.8604	0.3300	0.1717
6	0.9353	0.9353	0.9353	*0.0026*	0.4651	*0.0482*	*0.0482*	*0.0482*	0.6167	0.7454
7	0.1262	0.1262	0.1262	0.1804	0.0741	0.6165	0.6165	0.6165	0.3790	*0.0065*
8	0.9892	0.9892	0.9892	0.8285	0.1986	0.5154	0.5154	0.5154	0.7657	0.4161
9	0.5075	0.5075	0.5075	0.7764	0.1404	0.5428	0.5428	0.5428	0.7251	0.1333
10	0.3577	0.3577	0.3577	0.5792	0.2184	0.6359	0.6359	0.6359	0.4488	0.9031
11	0.0600	0.0600	0.0600	0.5699	0.1895	*0.0114*	*0.0114*	*0.0114*	0.1074	0.1367
12	0.9568	0.9568	0.9568	0.1156	0.3027	0.9568	0.9568	0.9568	0.1012	0.3090
13	0.9716	0.9716	0.9716	*0.0014*	*0.0014*	0.9716	0.9716	0.9716	$< 10^{-3}$	$< 10^{-3}$
14	0.6167	0.6167	0.6167	0.9892	0.4320	0.9892	0.9892	0.9892	0.7556	0.3781
15	0.4092	0.4092	0.4092	0.2912	0.0601	0.9568	0.9568	0.9568	0.3367	*0.0200*
16	0.4325	0.4325	0.4325	0.5607	0.0638	0.4901	0.4901	0.4901	0.4486	0.1195
17	0.8601	0.8601	0.8601	0.9568	——	0.7652	0.7652	0.7652	0.7447	——
18	0.2792	0.2792	0.2792	0.2792	——	0.2792	0.2792	0.2792	0.2673	——
19	0.2331	0.2331	0.2331	0.2027	——	0.6641	0.6641	0.6641	0.6061	——
20	0.1261	0.1261	0.1261	0.1261	——	0.2731	0.2731	0.2731	0.2233	——

Table 6: p-values for comparing classification accuracies based on KNN classifier

ID	Neighborhood rough set					Pseudo-label neighborhood rough set				
	FGS & Proposed	FPRC & Proposed	FSPA & Proposed	AGAR & Proposed	SGDA & Proposed	FGS & Proposed	FPRC & Proposed	FSPA & Proposed	AGAR & Proposed	SGDA & Proposed
1	0.8600	0.8600	0.8600	0.7554	0.1186	0.5694	0.5694	0.5694	0.9031	0.1259
2	0.7971	0.7971	0.7971	0.9353	0.7972	0.9353	0.9353	0.9353	0.9031	0.9676
3	0.5745	0.5745	0.5745	0.1192	*0.0011*	*0.0095*	*0.0095*	*0.0095*	*0.0267*	$< 10^{-3}$
4	0.3575	0.3575	0.3575	0.7659	0.9030	0.7971	0.7971	0.7971	0.4986	0.6354
5	0.5071	0.5071	0.5071	0.4732	0.0985	0.9568	0.9568	0.9568	0.9568	0.2183
6	0.6263	0.6263	0.6263	0.6073	0.1478	0.4406	0.4406	0.4406	0.9892	0.8817
7	0.1717	0.1717	0.1717	0.1439	0.3431	0.2731	0.2731	0.2731	0.7762	0.4650
8	0.5250	0.5250	0.5250	0.6750	0.0360	0.5428	0.5428	0.5428	0.6168	*0.0360*
9	0.8179	0.8179	0.8179	0.5975	0.9784	0.6165	0.6165	0.6165	0.3502	0.3788
10	0.6750	0.6750	0.6750	0.8392	0.8924	0.1045	0.1045	0.1045	0.2674	0.1719
11	0.5791	0.5791	0.5791	0.8286	0.2558	0.2731	0.2731	0.2731	1.0000	0.2182
12	0.6352	0.6352	0.6352	0.2131	0.4729	0.6352	0.6352	0.6352	0.0986	0.4729
13	1.0000	1.0000	1.0000	$< 10^{-3}$	0.1484	1.0000	1.0000	1.0000	$< 10^{-3}$	0.0664
14	1.0000	1.0000	1.0000	0.8181	1.0000	0.9892	0.9892	0.9892	0.7352	0.9783
15	0.1894	0.1894	0.1894	0.1515	0.0325	0.6553	0.6553	0.6553	0.6749	0.1594
16	0.7867	0.7867	0.7867	0.7351	0.2730	0.6166	0.6166	0.6166	0.6749	0.2285
17	0.8381	0.8381	0.8381	0.7336	——	0.9782	0.9782	0.9782	0.9239	——
18	0.9676	0.9676	0.9676	0.9676	——	0.9676	0.9676	0.9676	0.9461	——
19	0.8813	0.8813	0.8813	0.7965	——	0.8385	0.8385	0.8385	0.8812	——
20	0.5978	0.5978	0.5978	0.5978	——	0.4650	0.4650	0.4650	0.3940	——

Table 7: p-values for comparing classification accuracies based on CART classifier

ID	Neighborhood rough set					Pseudo-label neighborhood rough set				
	FGS & Proposed	FPRC & Proposed	FSPA & Proposed	AGAR & Proposed	SGDA & Proposed	FGS & Proposed	FPRC & Proposed	FSPA & Proposed	AGAR & Proposed	SGDA & Proposed
1	0.0546	0.0546	0.0546	*0.0154*	$< 10^{-3}$	0.4983	0.4983	0.4983	0.1593	0.9138
2	0.9892	0.9892	0.9892	0.7557	0.9892	0.5075	0.5075	0.5075	0.4249	0.9246
3	*0.0016*	*0.0016*	*0.0016*	*0.0025*	$< 10^{-6}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-6}$
4	0.3104	0.3104	0.3104	0.3507	0.3233	0.4903	0.4903	0.4903	0.2503	*0.0043*
5	0.7557	0.7557	0.7557	0.9031	0.9892	0.9568	0.9568	0.9568	0.8817	0.7139
6	0.3793	0.3793	0.3793	0.3234	0.8711	0.8498	0.8498	0.8498	0.9031	0.0810
7	0.9246	0.9246	0.9246	0.7557	0.0619	0.6456	0.6456	0.6456	0.7149	0.4818
8	0.6750	0.6750	0.6750	0.3369	0.0787	0.7660	0.7660	0.7660	0.7150	*0.0167*
9	0.6262	0.6262	0.6262	0.4246	0.6650	1.0000	1.0000	1.0000	0.3644	0.2666
10	0.1595	0.1595	0.1595	0.2674	0.1805	0.1368	0.1368	0.1368	0.9892	0.5075
11	0.3035	0.3035	0.3035	0.8497	*0.0033*	0.1847	0.1847	0.1847	0.8710	*0.0239*
12	0.7041	0.7041	0.7041	0.1840	0.2023	0.7041	0.7041	0.7041	0.2125	0.2023
13	1.0000	1.0000	1.0000	$< 10^{-2}$	0.1441	1.0000	1.0000	1.0000	$< 10^{-3}$	0.0788
14	0.7049	0.5885	0.5885	0.5885	1.0000	0.1595	0.1595	0.1595	0.3301	0.1888
15	0.5160	0.5160	0.5160	0.0547	0.4406	0.5699	0.5699	0.5699	0.4015	0.8180
16	0.6552	0.6552	0.6552	0.5426	0.1331	0.4405	0.4405	0.4405	0.4405	0.1553
17	*0.0475*	*0.0475*	*0.0475*	*0.0403*	——	0.0802	0.0802	0.0802	0.0757	——
18	0.3506	0.3506	0.3506	0.3506	——	0.3506	0.3506	0.3506	0.3234	——
19	0.2606	0.2606	0.2606	0.1508	——	0.5508	0.5508	0.5508	0.4641	——
20	0.4405	0.4405	0.4405	0.4405	——	0.0547	0.0547	0.0547	0.0882	——

Through observing Tabs. 5-7, no matter which algorithm was compared with our proposed approach, the returned p-values were greater than 0.05 in most cases. Combing with the results of Figs. 5-6, note that our proposed approach performed similarly with other compared algorithms in classification ability.

5. Conclusions and future plans

In this paper, we developed a framework called Fusing Attribute Reduction Accelerator for further speeding up the procedure of deriving reduct. Different from the previous popular accelerators, which can only speed up the process of deriving reduct from one and only one perspective, our device was actually a general framework for quickly searching reduct with respect to 3 different perspectives. The experimental results over 16 UCI datasets and 4 genetic datasets demonstrated the following superiorities: (1) our proposed strategy can significantly improve the time efficiency in searching qualified reduct; (2) the attributes in our obtained reduct provided well-matched classification performances.

The following topics will deserve our further investigations.

(1) In this study, our proposed strategy was only applied to a fixed parameter based granularity. Moreover, the concept of multi-granularity can be further introduced into our proposed strategy.

(2) Other measures such as conditional entropy, neighborhood decision error rate will be further explored for verifying the effectiveness of our proposed strategy.

Acknowledgment

This work is supported by the Natural Science Foundation of China (Nos. 62076111, 11971211, 62006099, 62006128, 61906078), Nature Science Foundation of Jiangsu, China (No. BK20191457), Open Project Foundation of Intelligent Information Processing Key Laboratory of Shanxi Province (No. CICIP2020004) and the Key Laboratory of Oceanographic Big Data Mining & Application of Zhejiang Province (No. OBDMA202002).

[1] R.K. Bania and A. Halder, R-Ensembler: a greedy rough set based ensemble attribute selection algorithm with kNN imputation for classification of medical data, Computer Methods and Programs in Biomedicine, 184, Article: 105122, (2020), DOI: 10.1016/j.cmpb.2019.105122

[2] M.J. Benítez-Caballero, J. Medina, E. Ramírez-Poussa, D. Ślęzak, Bireducts with tolerance relations, Information Sciences, 435, 26-39, (2018), DOI: 10.1016/j.ins.2017.12.037

[3] M.J. Benítez-Caballero, J. Medina, E. Ramírez-Poussa, D. Ślęzak, Rough-set-driven approach for attribute reduction in fuzzy formal concept analysis, Fuzzy Sets and Systems, 391, 117-138, (2020), DOI: 10.1016/j.fss.2019.11.009

[4] M.J. Benítez-Caballero, J. Medina, E. Ramírez-Poussa, D. Ślęzak, A computational procedure for variable selection preserving different initial conditions, International Journal of Computer Mathematics, 97, 387-404, (2020), DOI: 10.1080/00207160.2019.1613530
[5] Y. Chen, K.Y. Liu, J.J. Song, H. Fujita, X.B. Yang, Y.H. Qian, Attribute group for attribute reduction, Information Sciences, 535, 64-80, (2020), DOI: 10.1016/j.ins.2020.05.010
[6] J. Cai, J.W. Luo, S.L. Wang, Y. Sheng, Feature selection in machine learning: a new perspective, Neurocomputing, 300, 70-79, (2018), DOI: 10.1016/j.neucom.2017.11.077
[7] Y. Chen, P.X. Wang, X.B. Yang, J.S. Mi, D. Liu, Granular ball guided selector for attribute reduction, Knowledge-Based Systems, 229, Article: 107326, (2021), DOI: 10.1016/j.knosys.2021.107326
[8] D.G. Chen, Y.Y. Yang, Z. Dong, An incremental algorithm for attribute reduction with variable precision rough sets, Applied Soft Computing, 45, 129-149, (2016), DOI: 10.1016/j.asoc.2016.04.003
[9] J.H. Dai and H.W. Tian, Entropy measures and granularity measures for set-valued information systems, Information Sciences, 240, 72-82, (2013), DOI: 10.1016/j.ins.2013.03.045
[10] W.P. Ding, I. Triguero, C.T. Lin, Coevolutionary fuzzy attribute rrder reduction with complete attribute-value space tree, IEEE Transactions on Emerging Topics in Computational Intelligence, 5, 130-142, (2021), DOI: 10.1109/TETCI.2018.2869919
[11] H. Fujita, A. Gaeta, V. Loia, F. Orciuoli, Resilience analysis of critical infrastructures: a cognitive approach based on granular computing, IEEE Transactions on Cybernetics, 49,1835-1848, (2019), DOI: 10.1109/TCYB.2018.2815178
[12] J. Fan, Y.L. Jiang, Y. Liu, Quick attribute reduction with generalized indiscernibility models, Information Scineces, 397-398, 15-36, (2017), DOI: 10.1016/j.ins.2017.02.032
[13] Q.H. Hu, D.R. Yu, Z.X. Xie, Neighborhood classifiers, Expert Systems with Applications, 34, 866-876, (2008), DOI: 10.1016/j.eswa.2006.10.043
[14] M. Grzegorowski and D. Ślęzak, On resilient feature selection: computational foundations of r-C-reducts, Information Sciences, 499, 25-44, (2019), DOI: 10.1016/j.ins.2019.05.041
[15] Z.H. Jiang, K.Y. Liu, X.B. Yang, H.L. Yu, H. Fujita, Y.H. Qian, Accelerator for supervised neighborhood based attribute reduction, International Journal of Approximate Reasoning, 119, 122-150, (2020), DOI: 10.1016/j.ijar.2019.12.013
[16] A. Janusz and D. Ślęzak, Rough set methods for attribute clustering and selection, Applied Artificial Intelligence, 28, 220-242, (2014), DOI: 10.1080/08839514.2014.883902
[17] Z.H. Jiang, X.B. Yang, H.L. Yu, D. Liu, P.X. Wang, Y.H. Qian, Accelerator for multi-granularity attribute reduction, Knowledge-Based Systems, 177, 145-158, (2019), DOI: 10.1016/j.knosys.2019.04.014
[18] Y. Liu, W.L. Huang, Y.L. Jiang, Z.Y. Zeng, Quick attribute reduct algorithm for neighborhood rough set model, Information Sciences, 271, 65-81, (2014), DOI: 10.1016/j.ins.2014.02.093
[19] D. Liu, T.R. Li, J.B. Zhang, Incremental updating approximations in probabilistic rough sets under the variation of attributes, Knowledge-Based Systems, 73, 81-96, (2015), DOI: 10.1016/j.knosys.2014.09.008
[20] J.H. Li, Y. Ren, C.L. Mei, Y.H. Qian, X.B. Yang, A comparative study of multigranulation rough sets and concept lettices via rule acquisition, Knowledge-Based Systems, 91, 152-164, (2016), DOI: 10.1016/j.knosys.2015.07.024
[21] J.Y. Liang and Z.Z. Shi, The information entropy, rough entropy and knowledge granulation in rough set theory, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 12, 37-46, (2004), DOI: 10.1142/s0218488504002631
[22] F. Liu, B. Zhang, D. Ciucci, W. Wu, F. Min, A comparison study of similarity measures for covering-based neighborhood classifiers, Information Sciences, 448, 1-17, (2018), DOI: 10.1016/j.ins.2018.03.030
[23] Z. Pawlak, Rough sets: theoretical aspects of reasoning about data, Kluwer, Dordrecht, (1991), DOI: 10.1007/978-94-011-3534-4
[24] W. Pedrycz, G. Succi, A. Sillitti, J. Iljazi, Data description: a general framework of information granules, Knowledge-Based System, 80, 98-108, (2015), DOI: 10.1016/j.knosys.2014.12.030
[25] K.Y. Qin, B. Li, Z. Pei, Attribute reduction and rule acquisition of formal decision context based on object (property) oriented concept lattices, International Journal of Machine Learning and Cybernetics, 10, 2837-2850, (2019), DOI: 10.1007/s13042-018-00907-0
[26] Y.H. Qian, J.Y. Liang, W. Pedrycz, C.Y. Dang, Positive approximation: an accelerator for attribute reduction in rough set theory, Artificial Intelligence, 174, 597-618, (2010), DOI: 10.1016/j.artint.2010.04.018
[27] Y.H. Qian, X.Y. Liang, Q. Wang, J.Y. Liang, B. Liu, A. Skowron, Y.Y. Yao, J.M. Ma, C.Y. Dang, Local rough set: a solution to rough data analysis in big data, International Journal of Approximate Reasoning, 97, 38-63, (2018), DOI: 10.1016/j.ijar.2018.01.008
[28] X.S. Rao, X.B. Yang, X. Yang, X.J. Chen, D. Liu, Y.H. Qian, Quickly calculating reduct: an attribute relationship based approach, Knowledge-Based Systems, 200, Article: 106014, (2020), DOI: 10.1016/j.knosys.2020.106014
[29] D. Ślęzak, Approximate entropy reducts, Fundamenta Informaticae, 53, 365-390, (2002), DOI: 10.5555/780834.780843
[30] B.B. Sang, H.M. Chen, T.R. Li, W.H. Xu, H. Yu, Incremental approaches for heterogeneous feature selection in dynamic ordered data, Information Sciences, 541, 475-501, (2020), DOI: 10.1016/j.ins.2020.06.051
[31] Y.H. She, X.L. He, T. Qian, Q.Q. Wang, W.L. Zeng, A theoretical study on object-oriented and property-oriented multi-scale formal concept analysis, International Journal of Machine Learning and Cybernetics, 10, 3263-3271, (2019), DOI: 10.1007/s13042-019-01015-3
[32] B.Z. Sun, W.M. Ma, H.Y. Zhao, An approach to emergency decision making based on decision-theoretic rough set over two universes, Soft Computing, 20, 3617-3628, (2016), DOI: 10.1007/s00500-015-1721-6
[33] P. Sowkuntla and P.S.V.S. S. Prasad, MapReduce based improved quick reduct algorithm with granular refinement using vertical partitioning scheme, Knowledge-Based Systems, 189, Article: 105104, (2020), DOI: 10.1016/j.knosys.2019.105104
[34] S. Stawicki, D. Ślęzak, A. Janusz, S. Widz, Decision bireducts and decision reducts-a comparison, International Journal of Approximate Reasoning, 84, 75-109, (2017), DOI: 10.1016/j.ijar.2017.02.007
[35] E.C.C. Tsang, Q.H. Hu, D.G. Chen, Feature and instance reduction for pnn classifiers based on fuzzy rough sets, International Journal of Machine Learning and Cybernetics, 7, 1-11, (2016), DOI: 10.1007/s13042-014-0232-6
[36] E.C.C. Tsang, J.J. Song, D.G. Chen, X.B. Yang, Order based hierarchies on hesitant fuzzy approximation space, International Journal of Machine Learning and Cybernetics, 10, 1407-1422, (2019), DOI: 10.1007/s13042-018-0822-9
[37] Y.B. Wang, X.J. Chen, K. Dong, Attribute reduction via local conditional entropy, International Journal of Machine Learning and Cybernetics,

10, 3619-3634, (2019), DOI: 10.1007/s13042-019-00948-z
[38] W. Wei and J.Y. Liang, Information fusion in rough set theory: an overview, Information Fusion, 48, 107-118, (2018), DOI: 10.1016/j.inffus.2018.08.007
[39] S.P. Xu, X.B. Yang, H.L. Yu, D.J. Yu, J.Y. Yang, E.C.C. Tsang, Multi-label learning with label-specific feature reduction, Knowledge-Based Systems, 104: 52-61, (2016), DOI: 10.1016/j.knosys.2016.04.012
[40] X.B. Yang, S.C. Liang, H.L. Yu, S. Gao, Y.H. Qian, Pseudo-label neighborhood rough set: measures and attribute reductions, International Journal of Approximate Reasoning, 105, 112-129, (2019), DOI: 10.1016/j.ijar.2018.11.010
[41] L. Yang, W.H. Xu, X.Y. Zhang, B.B. Sang, Multi-granulation method for information fusion in multi-source decision information system, International Journal of Approximate Reasoning, 122, 47-65, (2020), DOI: 10.1016/j.ijar.2020.04.003
[42] Y.Y. Yao, Y. Zhao, J. Wang, On reduct construction algorithms, in: 2006 International Conference on Rough Sets and Knowledge Technology, 297-304, (2006), DOI: 10.1007/978-3-642-10649-1
[43] X. Zhang, C.L. Mei, D.G. Chen, J.H. Li, Feature selection in mixed data: a method using a novel fuzzy rough set-based information entropy, Pattern Recognition, 56, 1-15, (2016), DOI: 10.1016/j.patcog.2016.02.013
[44] X.Q. Zhang, Q.H. Zhang, Y.L. Cheng, G.Y. Wang, Optimal scale selection by integrating uncertainty and cost-sensitive learning in multi-scale decision tables, International Journal of Machine Learning and Cybernetics, 11, 1095-1114, (2020), DOI: 10.1007/s13042-020-01101-x
[45] P.F. Zhu, W.C. Zhu, Q.H. Hu, C.Q. Zhang, W.M. Zuo, Subspace clustering guided unsupervised feature selection, Pattern Recognition, 66, 364-374, (2017), DOI: 10.1016/j.patcog.2017.01.016