Hindawi
Mathematical Problems in Engineering
Volume 2020, Article ID 2350627, 13 pages
https://doi.org/10.1155/2020/2350627

Research Article

Combined Accelerator for Attribute Reduction: A Sample Perspective

Yan Chen,[1] Jingjing Song,[1,2] Keyu Liu,[1] Yaojin Lin,[2,3] and Xibei Yang[1]

[1]*School of Computer, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu 212003, China*
[2]*Key Laboratory of Data Science and Intelligence Application, Fujian Province University, Zhangzhou, Fujian 363000, China*
[3]*School of Computer Science, Minnan Normal University, Zhangzhou, Fujian 363000, China*

Correspondence should be addressed to Jingjing Song; songjingjing108@163.com

Received 21 November 2019; Revised 12 January 2020; Accepted 21 January 2020; Published 19 February 2020

Academic Editor: Giuseppe D'Aniello

Copyright © 2020 Yan Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the field of neighborhood rough set, attribute reduction is considered as a key topic. Neighborhood relation and rough approximation play crucial roles in the process of obtaining the reduct. Presently, many strategies have been proposed to accelerate such process from the viewpoint of samples. However, these methods speed up the process of obtaining the reduct only from binary relation or rough approximation, and then the obtained results in time consumption may not be fully improved. To fill such a gap, a combined acceleration strategy based on compressing the scanning space of both neighborhood and lower approximation is proposed, which aims to further reduce the time consumption of obtaining the reduct. In addition, 15 UCI data sets have been selected, and the experimental results show us the following: (1) our proposed approach significantly reduces the elapsed time of obtaining the reduct; (2) compared with previous approaches, our combined acceleration strategy will not change the result of the reduct. This research suggests a new trend of attribute reduction using the multiple views.

1. Introduction

With the rapid development of technology, a large number of high-dimensional data come into being such as video, image, and text. The high dimensionality of data [1–5] may contribute to a great challenge for decision-making and data analysis. Attribute reduction [6–12], as an efficient tool for reducing the dimensionality of data [13–15], has been widely concerned by many researchers.

In the topic of attribute reduction, there are two important notions for the process of obtaining the reduct. One fundamental notion is binary relation [16], and the other is rough approximation.

Firstly, binary relation is effective to conduct information granulation [17–21]. In the classical rough set [22–25], binary relation is an equivalence relation, and each single equivalence class can be regarded as an information granule [26]. Nevertheless, the equivalence relation can only deal with categorical data, and this somewhat limits the applications of the rough set. Therefore, several generalized binary relations and the corresponding rough sets have been proposed [27–30, 31]. Among those generalized rough approximations, the neighborhood rough set offers us a useful mechanism to process continuous data. This is mainly because the neighborhood relation can be directly obtained via computing the distance between samples. From the viewpoint of Granular Computing (GrC), the neighborhood rough set groups the samples into different neighborhood granules [16].

To obtain the neighborhood granule for each sample, all the other samples in the data should be scanned by using traditional approaches [26, 32, 33]. With the increasing of the size of data, the elapsed time of computing neighborhood granules will be unacceptable. To save the time consumption of computing neighborhood granules, Liu et al. [34] have introduced the hash function into the calculating of neighborhood. Such hash function can map all the samples into a series of buckets. Therefore, only samples in the adjacent buckets should be considered to obtain a neighborhood granule. It should be emphasized that such a

method provides a novel approach for reducing the time consumption of obtaining the neighborhood granule from the perspective of the sample.

Secondly, rough approximation is a general concept represented by a pair of sets, which is always characterized via the so-called upper and lower approximations. Following the detailed description of the mechanism of bucket, it is not difficult to observe that the bucket only reduces the scanning space of obtaining the neighborhood granule from the viewpoint of the sample. However, the mechanism of bucket takes all the samples into consideration when computing the lower approximation. If the scale of data is becoming larger, then the elapsed time of deriving lower approximation will be significantly increased. For such a reason, Qian et al. [35] have proposed the local approximation which aims to remove the useless information in the process of obtaining the lower approximation. Such a strategy only uses the samples in the target concept instead of the samples in the whole universe for deriving the lower approximation. Therefore, such a method fully meets the requirements of big data analysis.

However, no matter which strategy is employed, it can only accelerate the process of obtaining the reduct from one and only one point of view. Motivated by further reducing the elapsed time of obtaining the reduct, a new approach which combines the above two crucial notions will be proposed from the perspective of the sample. Firstly, the mechanism of bucket is used, which aims to reduce the scanning space of deriving neighborhood granules. Secondly, by using the local approximation, the useless information granules will be removed for quickly computing the lower approximation. Finally, the appropriate attributes will be selected until the intended constraint defined in the attribute reduction is satisfied. Obviously, the main contribution of our approach is to take both advantages of bucket and local approximation, and then it may be more effective for reducing the elapsed time of obtaining the reduct. It should be noticed that though our approach pay much attention to the viewing of the sample, our objective is actually achieved by considering two different perspectives of the sample. One is the sample related to constructing information granule, and the other is the sample related to constructing lower approximation.

The remainder of this paper is organized as follows. In Section 2, the basic knowledge of the neighborhood rough set will be presented. In Section 3, the mechanism of bucket, local approximation, and our proposed strategy will be described in detail. In Section 4, comparative experimental results and the corresponding analyses will be shown over 15 UCI data sets. Finally, some remarks and perspectives for the future work conclude the paper in Section 5.

2. Preliminaries

2.1. Binary Relation. Without loss of generality, a decision system can be represented as $\mathrm{DS} = \langle U, AT \cup d \rangle$ in which U is the set of samples called the universe, AT is the set of conditional attributes, and d is the decision attribute. Furthermore, $\forall x_i \in U$, $a(x_i)$ indicates its value over the conditional attribute $a \in AT$, and $d(x_i)$ denotes its value of decision attribute.

With regard to a given decision system DS, to derive the result of information granulation over U, various forms of binary relations have been explored in terms of different requirements. Two typical forms will be analyzed as follows:

(i) One is the equivalence relation. If information granulation is required over the categorical data, then the binary relation can be given by the equivalence relation. For example, in supervised learning task, the values over decision attribute in a decision system are categorical. To derive the result of information granulation over the decision attribute, the equivalence relation $\mathrm{IND}_d = \{(x_i, x_j) \in U \times U \mid d(x_i) = d(x_j)\}$ can be employed. Then, the result of information granulation (i.e., partition) $(U/\mathrm{IND}_d) = \{X_1, X_2, \ldots, X_s\}$ can be derived, and the value of s is the number of equivalence classes. Specifically, each equivalence class in such partition is referred to an information granule.

(ii) The other is the parameterized binary relation. If information granulation is required over the continuous data, then the binary relation can be given by parameterized binary relation. Take the neighborhood relation as an example, and the values over conditional attribute in the decision system are continuous; to derive the results of information granulation over the conditional attribute, the neighborhood relation $\delta_A = \{(x_i, x_j) \in U \times U \mid \mathrm{dis}_A(x_i, x_j) \le \delta\}$ can be used, in which $\mathrm{dis}_A(x_i, x_j)$ is the distance between samples x_i and x_j over A, and δ is the radius. It follows that neighborhood system [36] can be derived. Specifically, each neighborhood obtained by $\delta_A(x_i) = \{x_j \in U \mid (x_i, x_j) \in \delta_A\}$ is referred to an information granule.

2.2. Neighborhood Rough Set

Definition 1. Given a decision system DS, $\forall A \subseteq AT$ and $\forall X_p \in (U/\mathrm{IND}_d)$ $(1 \le p \le s)$, the neighborhood lower and upper approximations of X_p with respect to A are defined as

$$\underline{\mathbb{N}_A}(X_p) = \{x_i \in U \mid \delta_A(x_i) \subseteq X_p\}; \tag{1}$$

$$\overline{\mathbb{N}_A}(X_p) = \{x_i \in U \mid \delta_A(x_i) \cap X_p \neq \varnothing\}. \tag{2}$$

The pair $[\underline{\mathbb{N}_A}(X_p), \overline{\mathbb{N}_A}(X_p)]$ is referred to as a neighborhood rough set of X_p in terms of A.

Definition 2. Given a decision system DS, the approximation quality of d over $A \subseteq AT$ is defined as

$$\gamma(A, d) = \frac{\left| \cup_{p=1}^{s} \underline{\mathbb{N}_A}(X_p) \right|}{|U|}, \tag{3}$$

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

From the viewpoint of rough set theory, approximation quality reflects the percentage of the samples which belongs to one of the decision classes by the explanation of neighborhood lower approximation. The greater the value of approximation quality, the greater the degree of such belongingness.

2.3. Attribute Reduction. Attribute reduction [32, 37–41] aims to find an attribute subset without reducing some considered properties which are related to the raw attributes. Take the approximation quality-based attribute reduction as an example; such an attribute reduction is to obtain a minimal attribute subset that will not decrease the degree of the belongingness shown in Definition 2. The definition of the approximation quality reduct can be presented as follows.

Definition 3. Given a decision system DS, $\forall A \subseteq AT$, A is referred to as an approximation quality-based reduct if and only if

(1) $\gamma(A, d) \geq \gamma(AT, d)$

(2) $\forall A' \subset A, \gamma(A', d) < \gamma(A, d)$

Following the detailed description, generating the approximation quality reduct should satisfy two conditions. The first condition implies that the value of approximation quality over AT should not be reduced, and the second condition indicates that each attribute in A is individually necessary.

To find the approximation quality reduct, a lot of researching results have been put forward using various searching strategies [7, 8, 42, 43]. The forward greedy searching strategy has been widely accepted for its lower time complexity. Notably, the significance function is a necessary factor in the framework of forward greedy searching. This is mainly because each candidate attribute should be evaluated by the significance function. The definition of significance function is shown in the following.

Definition 4. Given a decision system DS, $\forall A \subseteq AT$, the significance of the candidate attribute a with respect to A is defined as

$$\begin{aligned}\mathrm{Sig}(a, A, d) &= \gamma(A \cup \{a\}, d) - \gamma(A, d) \\ &= \frac{\left|\cup_{p=1}^{s} \underline{\mathbb{N}_{A\cup\{a\}}}(X_p)\right|}{|U|} - \frac{\left|\cup_{p=1}^{s} \underline{\mathbb{N}_{A}}(X_p)\right|}{|U|}.\end{aligned} \tag{4}$$

Following Definition 4, the detailed process of obtaining the reduct by using the forward greedy searching strategy can be shown in the following.

In Algorithm 1, the time complexity of the computing neighborhood relation is $O(|U|^2)$. This is mainly because the distance between any two samples should be computed when computing the neighborhoods of all the samples. In the worst case, no attribute is redundant, and then all the attributes in the decision system should be added into the reduct. Therefore, the time complexity of Algorithm 1 is $O(|U^2| \cdot |AT|^2)$.

3. Bucket and Local Approximation-Based Attribute Reductions

3.1. Bucket-Based Attribute Reduction. As what has been indicated in Section 2.1, the neighborhood of one sample can be represented as the granule. In Algorithm 1, to obtain the granule, it is necessary to scan all the samples in the decision system. However, if the number of samples, i.e., the value of $|U|$ increases to a large scale, then the time consumption of obtaining the granule can be unacceptable. To alleviate the phenomenon of excessive time consumption, Liu et al. [34] have proposed the mechanism of bucket which employs a hash function to map all the samples into a series of sequenced buckets, and then only samples in the adjacent buckets should be considered instead of the whole samples in the universe. The definition of the bucket is shown in the following.

Definition 5 (see [34]). Given a decision system DS, x_0 is a special sample constructed from U, where $\forall a \in AT$, $a(x_0) = \min\{a(x_i)\colon \forall x_i \in U\}$. Then, the samples in U can be divided into finite buckets $B_0, \ldots, B_t$. B_t can be denoted as

$$B_t = \left\{ x_i \in U \,\middle|\, \left\lceil \frac{\mathrm{dis}(x_0, x_i)}{\delta} \right\rceil = t \right\}. \tag{5}$$

If $x_0 \in U$, then $B_0 = \{x_0\}$; otherwise, $B_0 = \varnothing$.

Theorem 1 (see [34]). *Given a decision system DS, $B_0, \ldots, B_t$ are the buckets, then $\forall x_i \in B_q$ $(q = 1, 2, \ldots, t-1)$, and the neighborhood of x_i is only contained in $B_{q-1} \cup B_q \cup B_{q+1}$; if $x_i \in B_0$, then the neighborhood of x_i is only contained in $B_0 \cup B_1$; if $x_i \in B_t$, then the neighborhood of x_i is only contained in $B_{t-1} \cup B_t$.*

Notably, only samples in the adjacent buckets should be taken into consideration for obtaining the information granule through observing Theorem 1. Therefore, the scanning space of obtaining the information granule can be reduced. It follows that the elapsed time of calculating the information granule can be decreased. Following Definition 5 and Theorem 1, forward greedy searching based on the bucket is presented as follows.

Different from Algorithm 1, the number of samples which should be scanned is $|U'| = |B_{q-1}| + |B_q| + |B_{q+1}|$ in Algorithm 2. Obviously, $|U| \geq |U'|$. It follows that the mechanism of bucket can reduce the number of times to scan samples for obtaining an information granule. Therefore, the efficiency of obtaining the reduct can be improved by using Algorithm 2.

3.2. Local Approximation-Based Attribute Reduction. However, the mechanism of bucket shown in Section 3.1 only considers how to reduce the time consumption of obtaining the neighborhood of each sample. In the process of obtaining the reduct by considering the measure of approximation quality, the elapsed time of deriving the lower approximation is also time consuming. Following such

Inputs: Decision system DS = $\langle U, AT \cup d \rangle$, radius δ.
Outputs: An approximation quality based reduct A.
(1) Compute $\gamma(AT, d)$;
(2) $A \longleftarrow \varnothing$;
(3) **Do**
  (i) $\forall x_i \in U$, compute $\delta_A(x_i)$;
  (ii) Compute the lower approximation based on $\delta_A(x_i)$ by using equation (1);
  (iii) $\forall a \in AT - A$, compute $\mathrm{Sig}(a, A, d)$ based on the lower approximation, and then select b such that $\mathrm{Sig}(b, A, d) = \max\{\mathrm{Sig}(a, A, d)\colon\ \forall a \in AT - A\}$;
  (iv) $A = A \cup \{b\}$;
  (v) Compute $\gamma(A, d)$;
  While $\gamma(A, d) \leq \gamma(AT, d)$;
(4) **Do**
  (i) $\forall c \in A$, compute $\gamma(A - \{c\}, d)$;
  (ii) **If** $\gamma(A - \{c\}, d) \geq \gamma(AT, d)$
    $A = A - \{c\}$;
    End
  While A changes and $|A| > 1$;
(5) **Return** A.

Algorithm 1: Forward greedy searching strategy for obtaining the reduct.

Inputs: Decision system DS = $\langle U, AT \cup d \rangle$, radius δ.
Outputs: An approximation quality reduct A.
(1) Compute $\gamma(AT, d)$;
(2) $A \longleftarrow \varnothing$;
(3) **Do**
  (i) $\forall x_i \in U$, compute $\delta_A(x_i)$ by using Definition 5 and Theorem 1;
  (ii) Compute the lower approximation based on $\delta_A(x_i)$ by using equation (1);
  (iii) $\forall a \in AT - A$, compute $\mathrm{Sig}(a, A, d)$ based on lower approximation, and then select b such that $\mathrm{Sig}(b, A, d) = \max\{\mathrm{Sig}(a, A, d)\colon\ \forall a \in AT - A\}$;
  (iv) $A = A \cup \{b\}$;
  (v) Compute $\gamma(A, d)$;
  While $\gamma(A, d) \leq \gamma(AT, d)$;
(4) **Do**
  (i) $\forall c \in A$, compute $\gamma(A - \{c\}, d)$;
  (ii) **If** $\gamma(A - \{c\}, d) \geq \gamma(AT, d)$
    $A = A - \{c\}$;
    End
  While A changes and $|A| > 1$;
(5) **Return** A.

Algorithm 2: Bucket-based forward greedy searching strategy for obtaining the reduct.

consideration, Qian et al. [35] have proposed a local approach which can reduce the time consumption of computing lower approximation.

Through observing the process of Algorithms 1 and 2, equation (1) is adapted to compute the lower approximation, and then the value of approximation quality can be derived from the obtained lower approximation. Through observing equation (1), it is not difficult to observe that calculating the lower approximation should use all the information granules obtained by scanning all the samples, which is exceedingly time consuming. In fact, only the set of information granules $\{\delta_A(x_i) \mid \delta_A(x_i) \cap X_p \neq \varnothing,\ x_i \in U\}$ is useful, while the set of information granules $\{\delta_A(x_i) \mid \delta_A(x_i) \cap X_p = \varnothing,\ x_i \in U\}$ is useless. Therefore, the local rough approximation is defined as follows.

Definition 6. Given a decision system DS, $\forall A \subseteq AT$ and $\forall X_p \in (U/\mathrm{IND}_d)$, INDd, the neighborhood lower and upper approximations of X_p with respect to A are defined as

$$\underline{\mathbb{N}_A}(X_p) = \left\{x_i \in X_p \,\middle|\, \delta_A(x_i) \subseteq X_p\right\}; \tag{6}$$

$$\overline{\mathbb{N}_A}(X_p) = \left\{x_i \in U \,\middle|\, \delta_A(x_i) \cap X_p \neq \varnothing\right\}. \tag{7}$$

By using equation (6), the elapsed time of obtaining the lower approximation can be decreased, and then the time consumption of obtaining the reduct can also be reduced. The detailed algorithm is shown as follows.

In Algorithm 3, to compute the lower approximation of X_p, only the samples in X_p instead of the whole samples in U should be used. Therefore, the time complexity of obtaining the lower approximation is $O(|U| \cdot |X_p|)$. However, in Algorithm 1, the time complexity of obtaining the lower approximation is $O(|U|^2)$. Therefore, by using the local approximation approach, the elapsed time of obtaining the reduct can be decreased.

3.3. Bucket and Local Approximation-Based Attribute Reduction. Based on the detailed descriptions of bucket and local approximation, we can observe that both of these methods reduce the time consumption from the perspective of the sample.

(i) For the mechanism of bucket, it employs a hash function to map all the samples into a series of sequenced buckets, and only samples in the adjacent buckets should be considered when obtaining an information granule. Then, the elapsed time of obtaining an information granule is decreased, and it follows that the time of obtaining the reduct can be saved.

(ii) For the approach of local approximation, it can reduce the scope of obtaining a lower approximation, and then useless information can be removed. The elapsed time of computing the lower approximation is decreased, and it follows that the time of obtaining the reduct can also be saved.

Nevertheless, no matter which strategy is employed, it can only improve the time efficiency of obtaining the reduct from one and only one point of view. On the basis of this, we will propose a strategy which combines the above two approaches for further reducing the elapsed time of obtaining the reduct. Firstly, the mechanism of bucket is employed to obtain the information granule, which can reduce the scanning space of samples. Secondly, the useless information granules can be removed for quickly obtaining the lower approximation. Finally, the appropriate attributes can be selected by the measure of approximation quality until the constraint defined in the attribute reduction is satisfied.

In Algorithm 4, the time complexity of obtaining the lower approximation is $O(|U'| \cdot |X_p|)$. However, in Algorithm 2, the time complexity of obtaining the lower approximation is $O(|U'| \cdot |U|)$. In Algorithm 3, the time complexity of obtaining the lower approximation is $O(|U| \cdot |X_p|)$. Therefore, the time consumption of obtaining the reduct can be further decreased by using our Algorithm 4.

4. Experiments

To demonstrate the effectiveness of our proposed method in this paper, 15 UCI data sets have been selected to conduct experiments. The basic descriptions of data sets are shown in Table 1. All the experiments have been carried out on a personal computer with Windows 10, dual-core 2.60 GHz CPU, 8 GB memory. The programming language is MATLAB R2017b.

In our experiments, 5-fold cross-validation has been adopted to test the performances of different reducts. It follows that data have been divided into 5 parts of the same size. In each round of calculation, 80% of the samples have been regarded as the training data set for deriving the reduct, and the rest have been denoted as testing data for classification. Furthermore, 20 radii, i.e., 0.01, 0.02, . . . , 0.2 have been selected, which are recommended in [34].

Note that, in the following subsections, "AQR," "BAQR," "LAQR," and "BLAQR" denote reducts which are derived by Algorithms 1–4, respectively.

4.1. Comparisons of Elapsed Time. In this experiment, the elapsed time of obtaining the reduct by using four different algorithms will be compared. Moreover, the standard deviation is used to characterize the stability of time.

With a deep investigation of Figure 1 and Tables 2 and 3, it is not difficult to observe the following:

(1) With the increasing value of δ, the increasing trends can be observed for the elapsed time of computing reduct by using four algorithms in most cases. This is mainly because the number of attributes in the reduct is possibly to be increased if the value of δ increases.

(2) Compared with Algorithm 1, both Algorithm 2 and Algorithm 3 emerge lower time consumption. Moreover, Algorithm 3 requires lower time consumption than Algorithm 2 does. Take "Cardiotocography" as an example, if δ is 0.1, then the elapsed time of obtaining reducts by using three different algorithms is 15.7143, 14.3278, and 14.0373 seconds, respectively.

(3) Compared with the process of computing BAQR and LAQR, the process of searching BLAQR requires lower elapsed time. Take "Diabetic Retinopathy Debrecen" as an example; if δ is 0.15, then the elapsed time of obtaining BLAQR is 2.9222 seconds. The elapsed time of obtaining BAQR and LAQR is 3.9703 and 3.1344 seconds, respectively. Such results indicate that our proposed strategy has speeded up the process of finding the reduct significantly.

(4) Through observing Tables 2 and 3, Algorithm 4 comes with a lower mean time. Furthermore, standard deviation is also lower than the ones produced

Inputs: Decision system DS = $\langle U, AT \cup d \rangle$, radius δ.
Outputs: An approximation quality reduct A.
(1) Compute $(U/\mathrm{IND}_d) = \{X_1, X_2, \ldots, X_s\}$;
(2) Compute $\gamma(AT, d)$;
(3) $A \longleftarrow \varnothing$;
(4) **Do**

(i) $\forall X_p \in (U/\mathrm{IND}_d)$, $\forall x_i \in X_p$, compute $\delta_A(x_i)$;
(ii) Compute the lower approximation based on $\delta_A(x_i)$ by using equation (7);
(iii) $\forall a \in AT - A$, compute $\mathrm{Sig}(a, A, d)$ based on the lower approximation, and then select b such that $\mathrm{Sig}(b, A, d) = \max\{\mathrm{Sig}(a, A, d) : \forall a \in AT - A\}$;
(iv) $A = A \cup \{b\}$;
(v) Compute $\gamma(A, d)$;
While $\gamma(A, d) \leq \gamma(AT, d)$;

(5) **Do**

(i) $\forall c \in A$, compute $\gamma(A - \{c\}, d)$;
(ii) **If** $\gamma(A - \{c\}, d) \geq \gamma(AT, d)$
$A = A - \{c\}$;
End
While A changes and $|A| > 1$;

(6) **Return** A.

Algorithm 3: Local approximation-based forward greedy searching strategy for obtaining the reduct.

Inputs: Decision system DS = $\langle U, AT \cup d \rangle$, radius δ.
Outputs: An approximation quality reduct A.
(1) Compute $(U/\mathrm{IND}_d) = \{X_1, X_2, \ldots, X_s\}$;
(2) Compute $\gamma(AT, d)$;
(3) $A \longleftarrow \varnothing$;
(4) $\forall X_p \in (U/\mathrm{IND}_d)(1 \leq p \leq s)$, $\forall x_i \in X_p$, Hash (x_i, B_k);
(5) **Do**

(i) **For each** $x_i \in X_p$
For each $x_j \in B_{k-1} \cup B_k \cup B_{k+1}$
If $\mathrm{dis}(x_i, x_j) \leq \delta$
x_j is the neighbor of x_i;
End
End
Compute $\delta_A(x_i)$;
End
(ii) Compute the lower approximation based on $\delta_A(x_i)$ by using equation (7);
(iii) $\forall a \in AT - A$, compute $\mathrm{Sig}(a, A, d)$ based on the lower approximation, and then select b such that $\mathrm{Sig}(b, A, d) = \max\{\mathrm{Sig}(a, A, d) : \forall a \in AT - A\}$;
(iv) $A = A \cup \{b\}$;
(v) Compute $\gamma(A, d)$;
While $\gamma(A, d) \leq \gamma(AT, d)$;

(6) **Do**

(i) $\forall c \in A$, compute $\gamma(A - \{c\}, d)$;
(ii) **If** $\gamma(A - \{c\}, d) \geq \gamma(AT, d)$
$A = A - \{c\}$;
End
While A changes and $|A| > 1$;

(7) **Return** A.

Algorithm 4: Bucket and local approximation for obtaining the reduct.

Table 1: Descriptions of data sets.

ID	Data sets	Samples	Attributes	Decision classes
1	Cardiotocography	2126	21	10
2	Contraceptive Method Choice	1473	10	3
3	Crowdsourced Mapping	10845	28	6
4	Diabetic Retinopathy Debrecen	1151	20	2
5	German	1000	24	2
6	Letter Recognition	20000	16	26
7	Magic Gamma Telescope	19020	10	2
8	Page Blocks Classification	5473	10	5
9	Parkinson Multiple Sound Recording	1208	27	2
10	QSAR Biodegradation	1055	41	2
11	Statlog (image segmentation)	2310	19	7
12	Statlog (Landsat satellite)	6435	36	6
13	Wall-Following Robot Navigation	5456	24	4
14	Waveform Database Generator (version 1)	5000	21	3
15	Waveform Database Generator (version 2)	5000	40	3

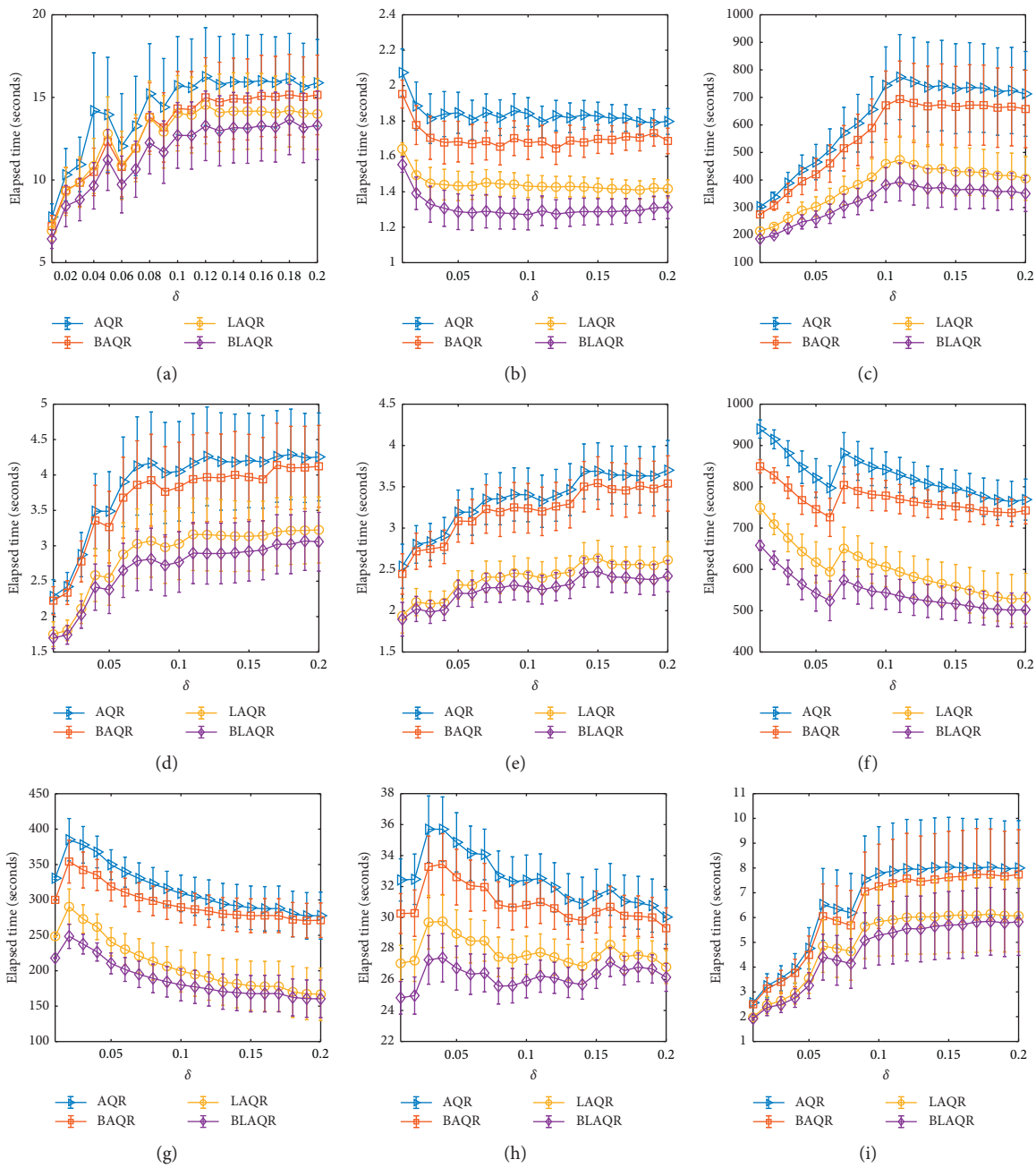

Figure 1: Continued.

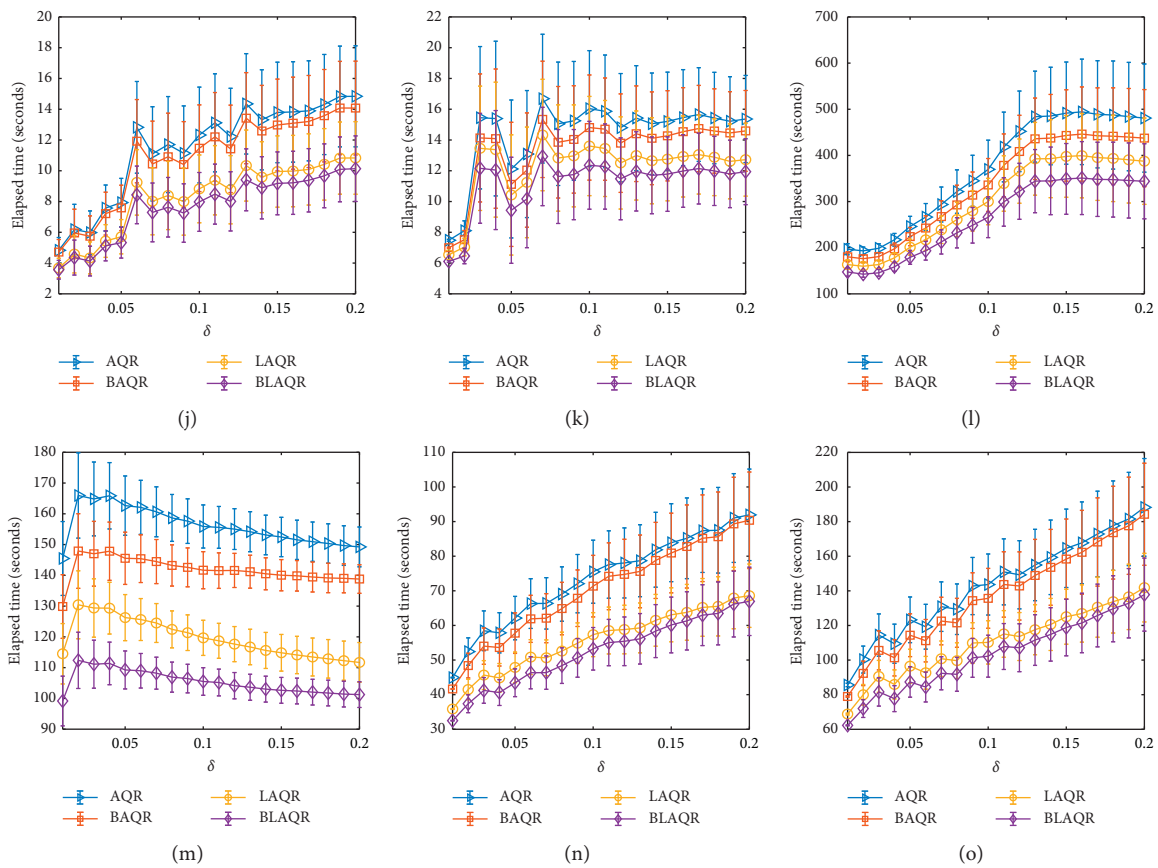

Figure 1: Comparisons of elapsed time for computing different reducts. (a) Cardiotocography. (b) Contraceptive Method Choice. (c) Crowdsourced Mapping. (d) Diabetic Retinopathy Debrecen. (e) German. (f) Letter Recognition. (g) Magic Gamma Telescope. (h) Page Blocks Classification. (i) Parkinson Multiple Sound Recording. (j) QSAR Biodegradation. (k) Statlog (image segmentation). (l) Statlog (Landsat satellite). (m) Wall-Following Robot Navigation. (n) Waveform Database Generator (version 1). (o) Waveform Database Generator (version 2).

Table 2: The elapsed time ($\delta = 0.01$).

ID	Algorithm 1 (seconds)	Algorithm 2 (seconds)	Algorithm 3 (seconds)	Algorithm 4 (seconds)
1	7.7614 ± 0.7933	7.7231 ± 0.6369	6.9335 ± 0.6232	6.4300 ± 0.5710
2	2.0736 ± 0.1333	1.9523 ± 0.0786	1.6427 ± 0.0341	1.5539 ± 0.0436
3	303.2941 ± 0.7575	275.1430 ± 0.7099	214.1942 ± 0.9268	186.1075 ± 0.4704
4	2.2841 ± 0.2375	2.2251 ± 0.1969	1.7519 ± 0.1755	1.6982 ± 0.1504
5	2.5383 ± 0.2688	2.4451 ± 0.2444	1.9373 ± 0.2082	1.8960 ± 0.2035
6	939.8496 ± 21.9583	849.6899 ± 16.0775	749.6158 ± 14.7290	657.9987 ± 12.0429
7	330.6784 ± 1.7847	300.1296 ± 0.9151	248.6087 ± 1.2150	217.9840 ± 0.7602
8	32.4269 ± 1.3612	30.2577 ± 1.2877	27.0394 ± 1.1620	24.8191 ± 1.0545
9	2.5644 ± 0.0962	2.5009 ± 0.0605	1.9736 ± 0.0375	1.9190 ± 0.0452
10	4.8290 ± 0.8321	4.7155 ± 0.8163	3.6787 ± 0.6328	3.5682 ± 0.6079
11	7.4820 ± 0.0606	6.9983 ± 0.0361	6.5404 ± 0.0259	6.1038 ± 0.0580
12	196.6553 ± 11.9935	180.2292 ± 10.8550	163.4637 ± 10.0040	147.4294 ± 9.0052
13	145.4798 ± 12.0058	129.8684 ± 10.3470	114.5147 ± 9.8129	99.1553 ± 8.0070
14	44.9296 ± 0.0999	41.6441 ± 0.0571	35.8451 ± 0.1155	32.4829 ± 0.0846
15	85.6206 ± 0.9990	78.9694 ± 0.3766	68.8183 ± 0.4956	62.2852 ± 0.2512

by AQR, BAQR, and LAQR for obtaining reducts in most cases.

4.2. Comparisons of Stability-Based Reducts. In this section, the stabilities of reducts will be compared using four different algorithms. For more details about computation of the stability, see [20]. Moreover, the standard deviation [44] is used to characterize the dispersion degree of stability. The lower the value of the standard deviation, the higher the stability of reducts.

Table 3: The elapsed time ($\delta = 0.2$).

ID	Algorithm 1 (seconds)	Algorithm 2 (seconds)	Algorithm 3 (seconds)	Algorithm 4 (seconds)
1	15.8818 ± 2.6218	15.1942 ± 2.4092	13.9933 ± 2.1455	13.3126 ± 2.0835
2	1.7975 ± 0.0730	1.6874 ± 0.0721	1.4172 ± 0.0507	1.3127 ± 0.0642
3	712.6823 ± 154.1671	658.0492 ± 140.5334	405.5043 ± 79.2855	350.7878 ± 65.0520
4	4.2553 ± 0.6203	4.1227 ± 0.5770	3.2243 ± 0.4658	3.0570 ± 0.4148
5	3.7009 ± 0.3600	3.5396 ± 0.3346	2.6141 ± 0.2216	2.4199 ± 0.1884
6	768.9334 ± 49.6549	742.6549 ± 32.5468	530.3251 ± 60.3075	502.3758 ± 41.4835
7	277.8121 ± 33.1599	271.3526 ± 24.4083	166.8020 ± 36.9960	160.3587 ± 26.4904
8	30.0322 ± 1.7625	29.3092 ± 1.3176	26.8057 ± 1.0691	26.1383 ± 0.9010
9	8.0012 ± 1.9094	7.7373 ± 1.8071	6.0683 ± 1.4523	5.8254 ± 1.3521
10	14.8391 ± 3.2831	14.0761 ± 3.0478	10.8337 ± 2.3596	10.1352 ± 2.1318
11	15.3631 ± 2.8326	14.5911 ± 2.6272	12.7235 ± 2.3736	11.9391 ± 2.1424
12	480.8548 ± 117.1881	437.4929 ± 105.1775	387.4126 ± 93.4885	343.9659 ± 81.3193
13	149.2419 ± 6.4690	138.7982 ± 4.6385	111.6941 ± 6.4658	101.2385 ± 4.1834
14	91.9483 ± 13.2181	90.3759 ± 13.9434	68.5839 ± 9.1511	66.9474 ± 9.8640
15	188.2401 ± 28.2218	184.2565 ± 29.4879	141.8513 ± 19.8201	137.8270 ± 21.1095

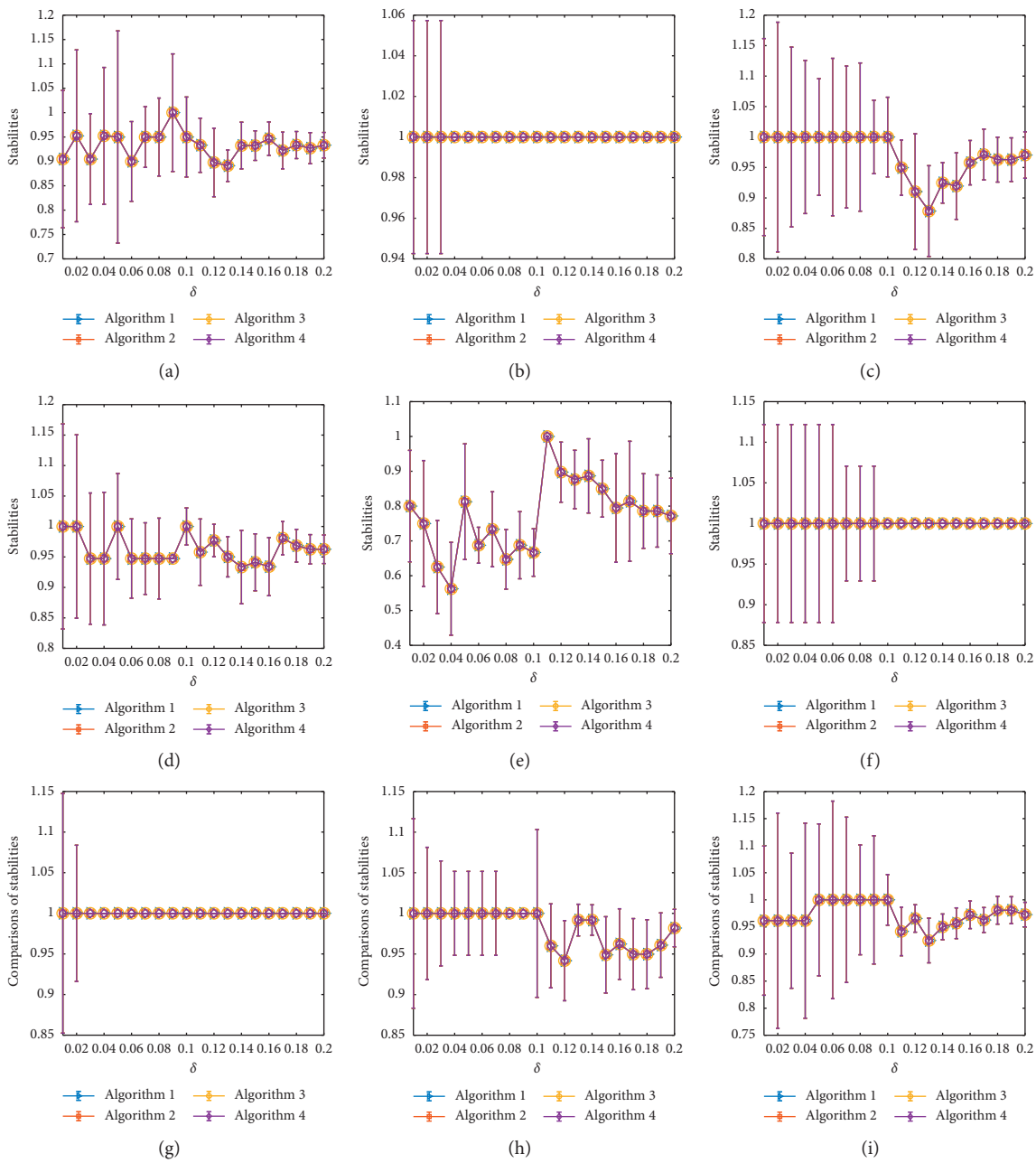

Figure 2: Continued.

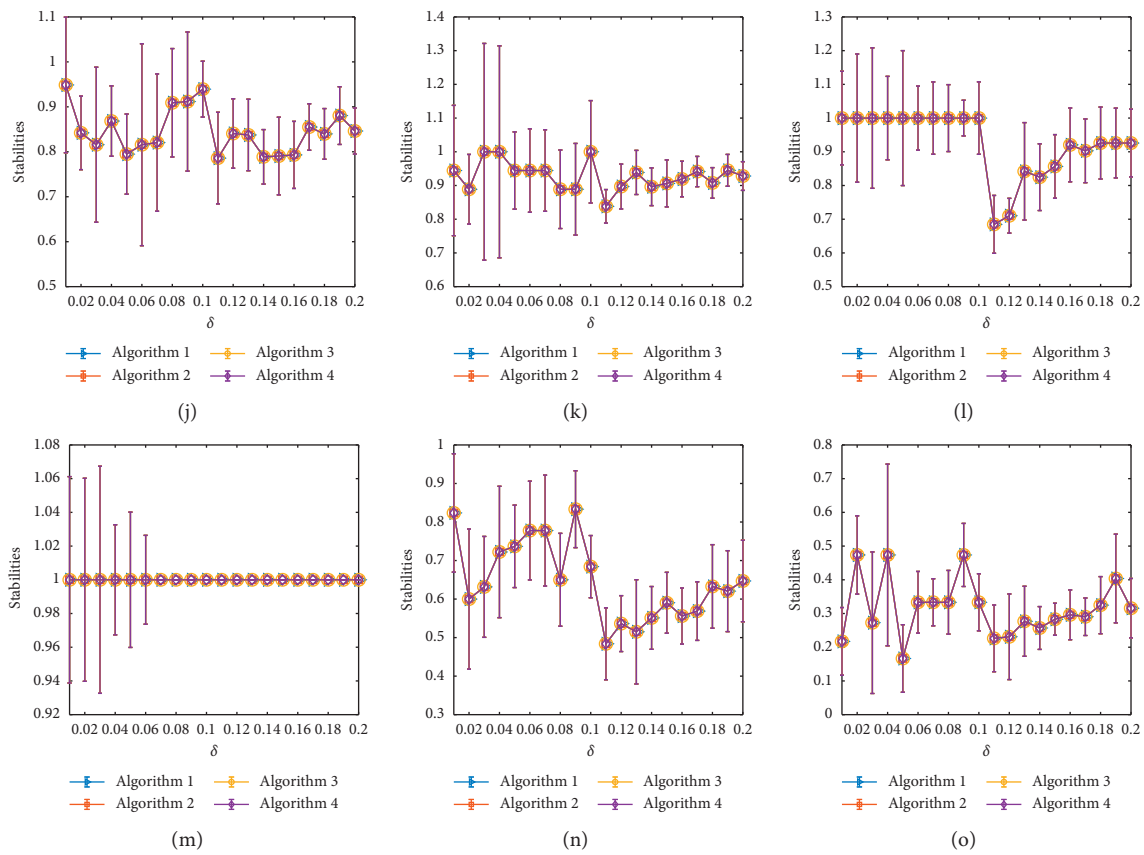

Figure 2: Comparisons of stability-based reducts. (a) Cardiotocography. (b) Contraceptive Method Choice. (c) Crowdsourced Mapping. (d) Diabetic Retinopathy Debrecen. (e) German. (f) Letter Recognition. (g) Magic Gamma Telescope. (h) Page Blocks Classification. (i) Parkinson Multiple Sound Recording. (j) QSAR Biodegradation. (k) Statlog (image segmentation). (l) Statlog (Landsat satellite). (m) Wall-Following Robot Navigation. (n) Waveform Database Generator (version 1). (o) Waveform Database Generator (version 2).

Following Figure 2, it is not difficult to observe the following:

(1) The stabilities of reducts by using four algorithms are the same. That is to say, the derived reducts by using four different algorithms are the same. This is mainly because the mechanism of bucket only reduces the scanning times of samples, and the obtained neighborhood relation does not change. Furthermore, the local approach only removes the useless granule, and the obtained lower approximation also does not change. Therefore, our proposed strategy which combines the mechanism of bucket and local approximation does not change the result of the reduct.

(2) The standard deviations of reducts are also the same. It follows that the dispersion degrees of stabilities for obtaining the reduct by using four algorithms are the same.

Because the stability-based reducts by using four algorithms are the same, the classification accuracies derived from AQR, BAQR, LAQR, and BLAQR are also the same. Therefore, the results of classification will not be shown in this paper.

4.3. Comparisons of Elapsed Time over Incremental Data. In this section, the time tendency with the incremental scale of data will be shown. The whole universe has been divided into 11 groups. Firstly, only the first group will be used to obtain the reduct and the corresponding elapsed time will be recorded; secondly, the second group will be added into the first group for computing reduct and the corresponding elapsed time will be recorded, and so on; and finally, all the groups will be combined for computing reduct and the corresponding elapsed time will be recorded.

Following Figure 3, it is not difficult to observe the followings:

(1) With the increasing of the volume of data, the elapsed time of obtaining the reduct by using four different algorithms is increasing. This is mainly because when the scale of the sample increases, the elapsed time of calculating the neighborhood relation and lower approximation increases. Therefore, the time consumption of obtaining the reduct will be increased.

(2) Though the data have been divided into a series of groups, it is also not difficult to observe that the elapsed time of obtaining the reduct by using our approach is less than that of other methods.

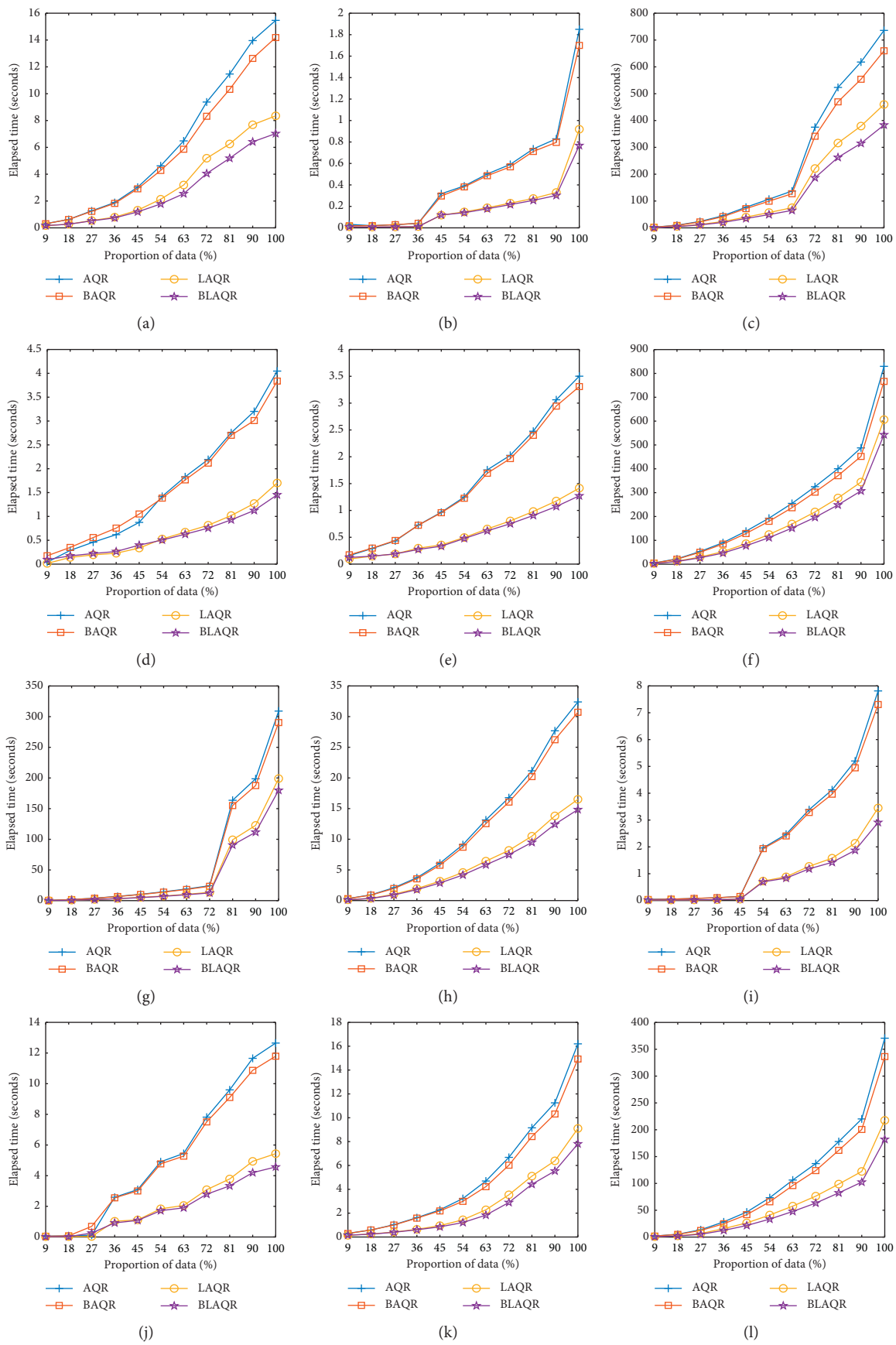

Figure 3: Continued.

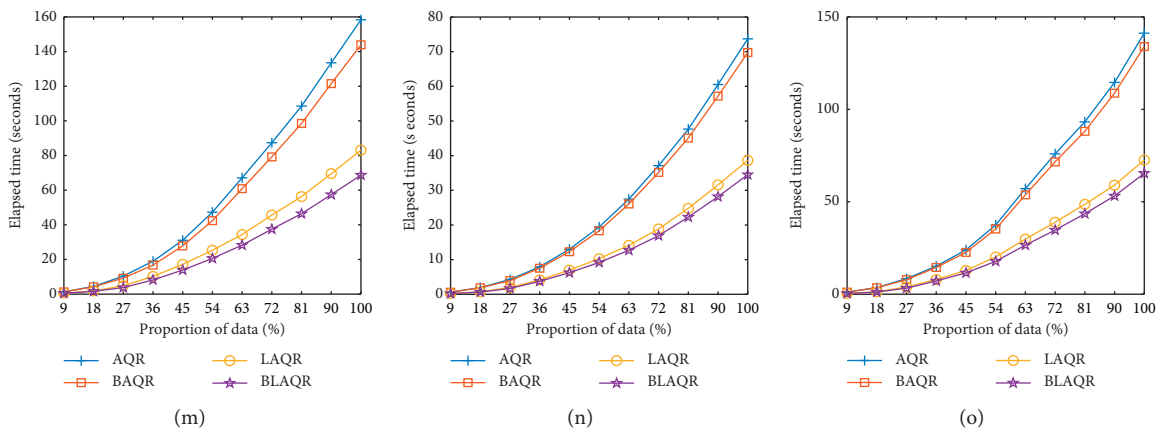

Figure 3: Comparisons of elapsed time over incremental data ($\delta = 0.1$). (a) Cardiotocography. (b) Contraceptive Method Choice. (c) Crowdsourced Mapping. (d) Diabetic Retinopathy Debrecen. (e) German. (f) Letter Recognition. (g) Magic Gamma Telescope. (h) Page Blocks Classification. (i) Parkinson Multiple Sound Recording. (j) QSAR Biodegradation. (k) Statlog (image segmentation). (l) Statlog (Landsat satellite). (m) Wall-Following Robot Navigation. (n) Waveform Database Generator (version 1). (o) Waveform Database Generator (version 2).

Furthermore, with the size of data becoming large, our approach has obvious advantages in time consumption compared with other algorithms.

5. Conclusions and Future Perspectives

In this paper, we have designed an accelerator for calculating the reduct using the concepts of both bucket and local approximation. Our proposed strategy not only reduces the scanning space of obtaining information granules but also removes the useless granules for computing the lower approximations. Furthermore, the experimental results show that our proposed approach can significantly reduce the time consumption of obtaining the reduct. In addition, compared with the previous research studies, the stability-based reducts are the same which implies that the obtained reducts are unchanged by our strategy.

The following topics deserve our further investigations:

(1) The proposed strategy is only realized from the viewpoint of sample; some other acceleration strategies which take the attribute into account can be further studied.

(2) Single granularity is applied to our proposed strategy; multiple granularities [16] can be further considered in the accelerator.

Data Availability

The (UCI) data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Natural Science Foundation of China (nos. 61906078 and 61572242) and the Key Laboratory of Data Science and Intelligence Application, Fujian Province University (no. D1901).

References

[1] M. Dowlatshahi, V. Derhami, and H. Nezamabadi-pour, "Ensemble of filter-based rankers to guide an epsilon-greedy swarm optimizer for high-dimensional feature subset selection," *Information*, vol. 8, no. 4, p. 152, 2017.

[2] J. F. Luo, H. Fujita, Y. Y. Yao, and K. Y. Qin, "On modeling similarity and three-way decision under incomplete information in rough set theory," *Knowledge-Based Systems*, 2019, In press.

[3] Z. H. Liu, J. J. Wang, G. Liu, and L. Zhang, "Discriminative low-rank preserving projection for dimensionality reduction," *Applied Soft Computing*, vol. 85, Article ID 105768, 2019.

[4] L. Sun, S. S. Chen, J. C. Xu, and Y. Tian, "Improved monarch butterfly optimization algorithm based on opposition-based learning and random local perturbation," *Complexity*, vol. 2019, Article ID 4182148, 20 pages, 2019.

[5] L. Sun, J. Xu, S. Liu, S. Zhang, Y. Li, and C. A. Shen, "A robust image watermarking scheme using Arnold transform and BP neural network," *Neural Computing and Applications*, vol. 30, no. 8, pp. 2425–2440, 2018.

[6] D. Chen and Y. Yang, "Attribute reduction for heterogeneous data based on the combination of classical and fuzzy rough set models," *IEEE Transactions on Fuzzy Systems*, vol. 22, no. 5, pp. 1325–1334, 2014.

[7] D. Chen, Y. Yang, and Z. Dong, "An incremental algorithm for attribute reduction with variable precision rough sets," *Applied Soft Computing*, vol. 45, pp. 129–149, 2016.

[8] J. Fan, Y. L. Jiang, and Y. Liu, "Quick attribute reduction with generalized indiscernibility models," *Information Sciences*, vol. 397-398, pp. 15–36, 2017.

[9] Y. Lin, Q. Hu, J. Liu, J. Li, and X. Wu, "Streaming feature selection for multilabel learning based on fuzzy mutual information," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 6, pp. 1491–1507, 2017.

[10] K. Qin, B. Li, and 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*, vol. 10, no. 10, pp. 2837–2850, 2019.

[11] Y. Yao and Y. Zhao, "Attribute reduction in decision-theoretic rough set models," *Information Sciences*, vol. 178, no. 17, pp. 3356–3373, 2008.

[12] H. Zhao and K. Qin, "Mixed feature selection in incomplete decision table," *Knowledge-Based Systems*, vol. 57, pp. 181–190, 2014.

[13] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: a new perspective," *Neurocomputing*, vol. 300, pp. 70–79, 2018.

[14] Y.-C. Ko and H. Fujita, "An evidential analytics for buried information in big data samples: case study of semiconductor manufacturing," *Information Sciences*, vol. 486, pp. 190–203, 2019.

[15] L. Sun, R. Liu, J. Xu, S. Zhang, and Y. Tian, "An affinity propagation clustering method using hybrid kernel function with LLE," *IEEE Access*, vol. 6, pp. 68892–68909, 2018.

[16] K. Liu, X. Yang, H. Fujita, D. Liu, X. Yang, and Y. Qian, "An efficient selector for multi-granularity attribute reduction," *Information Sciences*, vol. 505, pp. 457–472, 2019.

[17] J. Liang, J. Wang, and Y. Qian, "A new measure of uncertainty based on knowledge granulation for rough sets," *Information Sciences*, vol. 179, no. 4, pp. 458–470, 2009.

[18] H. Ju, X. Yang, H. Yu, T. Li, D.-J. Yu, and J. Yang, "Cost-sensitive rough set approach," *Information Sciences*, vol. 355-356, pp. 282–298, 2016.

[19] L. Sun, X.-Y. Zhang, Y.-H. Qian, J.-C. Xu, S.-G. Zhang, and Y. Tian, "Joint neighborhood entropy-based gene selection method with Fisher score for tumor classification," *Applied Intelligence*, vol. 49, no. 4, pp. 1245–1259, 2019.

[20] X. Yang, Y. Qi, H. Yu, X. Song, and J. Yang, "Updating multigranulation rough approximations with increasing of granular structures," *Knowledge-Based Systems*, vol. 64, pp. 59–69, 2014.

[21] X. Yang, S. Xu, H. Dou, X. Song, H. Yu, and J. Yang, "Multigranulation rough set: a multiset based strategy," *International Journal of Computational Intelligence Systems*, vol. 10, no. 1, pp. 277–292, 2017.

[22] Q. H. Hu, D. R. Yu, Z. X. Xie, and J. F. Liu, "Fuzzy probabilistic approximation spaces and their information measures," *IEEE Transactions on Fuzzy Systems*, vol. 14, no. 2, pp. 191–201, 2006.

[23] H. Ju, X. Yang, X. Song, and Y. Qi, "Dynamic updating multigranulation fuzzy rough set: approximations and reducts," *International Journal of Machine Learning and Cybernetics*, vol. 5, no. 6, pp. 981–990, 2014.

[24] Z. Pawlak and A. Skowron, "Rough sets: some extensions," *Information Sciences*, vol. 177, no. 1, pp. 28–40, 2007.

[25] E. C. C. Tsang, J. Song, D. Chen, and X. Yang, "Order based hierarchies on hesitant fuzzy approximation space," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 6, pp. 1407–1422, 2019.

[26] Z. Jiang, X. Yang, H. Yu, D. Liu, P. Wang, and Y. Qian, "Accelerator for multi-granularity attribute reduction," *Knowledge-Based Systems*, vol. 177, pp. 145–158, 2019.

[27] Q. Hu, D. Yu, J. Liu, and C. Wu, "Neighborhood rough set based heterogeneous feature subset selection," *Information Sciences*, vol. 178, no. 18, pp. 3577–3594, 2008.

[28] K. Liu, X. Yang, H. Yu, J. Mi, P. Wang, and X. Chen, "Rough set based semi-supervised feature selection via ensemble selector," *Knowledge-Based Systems*, vol. 165, pp. 282–296, 2019.

[29] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Alphen aan den Rijn, Netherlands, 1992.

[30] L. Sun, X. Zhang, Y. Qian, J. Xu, and S. Zhang, "Feature selection using neighborhood entropy-based uncertainty measures for gene expression data classification," *Information Sciences*, vol. 502, pp. 18–41, 2019.

[31] X. Yang, S. Liang, H. Yu, S. Gao, and Y. Qian, "Pseudo-label neighborhood rough set: measures and attribute reductions," *International Journal of Approximate Reasoning*, vol. 105, pp. 112–129, 2019.

[32] J. Li, X. Yang, X. Song, J. Li, P. Wang, and D.-J. Yu, "Neighborhood attribute reduction: a multi-criterion approach," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 4, pp. 731–742, 2019.

[33] L. Sun, L. Y. Wang, W. P. Ding, Y. H. Qian, and J. C. Xu, "Neighborhood multi-granulation rough sets-based attribute reduction using Lebesgue and entropy measures in incomplete neighborhood decision systems," *Knowledge-Based Systems*, p. 105373, 2019, In press.

[34] Y. Liu, W. L. Huang, Y. L. Jiang, and Z. Y. Zeng, "Quick attribute reduct algorithm for neighborhood rough set model," *Information Sciences*, vol. 271, pp. 65–81, 2014.

[35] Y. Qian, X. Liang, Q. Wang et al., "Local rough set: a solution to rough data analysis in big data," *International Journal of Approximate Reasoning*, vol. 97, pp. 38–63, 2018.

[36] X. Yang, M. Zhang, H. Dou, and J. Yang, "Neighborhood systems-based rough sets in incomplete information system," *Knowledge-Based Systems*, vol. 24, no. 6, pp. 858–867, 2011.

[37] Y. Gao, X. J. Chen, X. B. Yang, P. X. Wang, and J. S. Mi, "Ensemble-based neighborhood attribute reduction: a multigranularity view," *Complexity*, vol. 2019, Article ID 2048934, 17 pages, 2019.

[38] C. Gao, Z. Lai, J. Zhou, J. Wen, and W. K. Wong, "Granular maximum decision entropy-based monotonic uncertainty measure for attribute reduction," *International Journal of Approximate Reasoning*, vol. 104, pp. 9–24, 2019.

[39] Z. H. Jiang, K. Y. Liu, X. B. Yang, H. L. Yu, H. Fujita, and Y. H. Qian, "Accelerator for supervised neighborhood based attribute reduction," *International Journal of Approximate Reasoning*, vol. 119, pp. 122–150, 2020.

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

[41] E. C. C. Tsang, Q. Hu, and D. Chen, "Feature and instance reduction for pnn classifiers based on fuzzy rough sets," *International Journal of Machine Learning and Cybernetics*, vol. 7, no. 1, pp. 1–11, 2016.

[42] J. Song, E. C. C. Tsang, D. Chen, and X. Yang, "Minimal decision cost reduct in fuzzy decision-theoretic rough set model," *Knowledge-Based Systems*, vol. 126, pp. 104–112, 2017.

[43] C. Z. Wang, Q. H. Hu, X. Z. Wang, D. G. Chen, Y. H. Qian, and Z. Dong, "Feature selection based on neighborhood discrimination index," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 7, pp. 2986–2999, 2017.

[44] X. Yang and Y. Yao, "Ensemble selector for attribute reduction," *Applied Soft Computing*, vol. 70, pp. 1–11, 2018.