Accelerator for multi-granularity attribute reduction☆

Zehua Jiang [a], Xibei Yang [a,*], Hualong Yu [a], Dun Liu [b], Pingxin Wang [c], Yuhua Qian [d]

[a] School of Computer, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu, 212003, PR China
[b] School of Economics and Management, Southwest Jiaotong University, Chengdu, Sichuan, 610031, PR China
[c] School of Science, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu, 212003, PR China
[d] Institute of Big Data Science and Industry, Shanxi University, Taiyuan, Shanxi, 030006, PR China

HIGHLIGHTS

- The multi-granularity is introduced into attribute reduction.
- An acceleration strategy is proposed to speed up the process of finding multi-granularity reduct.
- Our approach can reduce the time consumption of finding reduct.

ARTICLE INFO

ABSTRACT

By considering the information granulation in Granular Computing, the concept of the multi-granularity is important. It is mainly because different results of information granulation will imply different levels of granularity. Nevertheless, multi-granularity has been paid less attention to the problem of attribute reduction in rough set which is regarded as one of the most important mathematical tools in Granular Computing. Therefore, how to search the multi-granularity reduct will be mainly explored in this paper. Different from the previous studies which generate reduct by using one and only one granularity, multi-granularity reduct is actually a set of the reducts derived from multiple levels of granularity. A natural way for computing multi-granularity reduct is to repeat the process of searching reduct for each level of granularity. Obviously, such an approach is time-consuming. To fill such a gap, an acceleration strategy is introduced into the process of searching multi-granularity reduct. Our acceleration strategy can be respectively realized through considering two variations of granularity: 1) from a finer granularity to a coarser granularity; 2) from a coarser granularity to a finer granularity. Such two variations indicate that the reduct related to previous granularity may have guidance on the computation of reduct related to the present granularity. Consequently, two accelerators are designed for speeding up the process of finding multi-granularity reduct. The experimental results over 16 UCI data sets show that our accelerators can not only reduce the elapsed time of searching attributes significantly, but also select attributes which will not contribute to a poorer classification performance. This study suggests new trends concerning the problem of attribute reduction and the corresponding searching strategy.

1. Introduction

Presently, as one of the important tools for characterizing uncertainty, rough set [1–10] has been successfully applied to Feature Selection [11–15], Pattern Recognition [16–19], Data Mining [20] and so on. It should be emphasized that in the development of rough set, much attention has been paid to the attribute reduction [21–33]. The reason can be attributed to the following two aspects: (1) attribute reduction is useful in reducing the dimensionality of data; (2) the derived reduct, i.e., a subset of the raw condition attributes has clear explanation based on the used measure in attribute reduction.

Up to now, it is noticed that most of the results about attribute reduction focus on one and only one fixed structure. For example, considering the classical rough set for analyzing data with categorical values, such a fixed structure is actually the partition derived from the raw attributes. Moreover, considering

the neighborhood rough set [34–36] for analyzing data with continuous or even mixed values, such a fixed structure is actually the neighborhood system [37] derived from the raw attributes and the given parameter, i.e., radius. Immediately, how to quantitatively characterize those structures has gained a substantial amount of attention. Fortunately, from the viewpoint of Granular Computing (GrC) [38–45], the concept of the granularity offers us a simple and intuitive way to measure the structure. In view of this, attribute reduction based on one and only one fixed structure can be regarded as the single granularity [46,47] based attribute reduction.

Nevertheless, it should be noticed that the single granularity based attribute reduction may have some inherent limitations in practical application. Some challenges are as follows.

1. Single granularity based attribute reduction cannot be used for parameter selection. For instance, Gaussian kernel fuzzy rough set [48–50] and neighborhood rough set are two typical examples of parameterized rough set, different parameters [51] indicate that different levels of granularity have been used to construct rough sets. If one and only one parameter is taken into consideration in attribute reduction, then it is impossible for us to determine whether the reduct related to the given parameter can provide a satisfied generalization performance for us. Therefore, to find reduct with better generalization performance, multiple different parameters should be tested, that is, multi-granularity is required.

2. Single granularity based attribute reduction is lack of power in reflecting the variation tendency of generalization performance [52]. For instance, by using the heuristic algorithm, one and only one reduct is obtained over the given granularity, there is no comparison among the performances derived from multiple reducts in terms of multiple levels of granularity. From this point of view, multi-granularity should be paid much attention to the problem of attribute reduction.

3. The process of finding single granularity based reduct is not suitable for multiple levels of granularity. For instance, if s different parameters are required to construct rough set, then the process of finding single granularity based reduct should be executed s times repeatedly until the reducts in terms of all parameters have been obtained. Moreover, if the number of used parameters is further increased, then the elapsed time for finding all of the reducts will also be increased. From this point of view, how to speed up the process of finding reducts with respect to multi-granularity is another important topic to be addressed.

To overcome the limitations of single granularity based attribute reduction mentioned above, a set which contains different levels of granularity, i.e., multi-granularity will be introduced into the framework of attribute reduction. Consequently, a set of the reducts can be derived, such a set is referred to as the multi-granularity reduct in the context of this paper. For example, if the structure of multi-granularity [53–56] is constructed by using a set of parameters, then the generalization performances related to reducts which are derived from those parameterized granularity can be compared. From this point of view, not only the reduct which provides us with better performance can be found for realizing parameter selection, but also the variation tendency of the generalization performance in terms of the varied granularity can be observed.

Following the above discussions, the immediate problem is to find multi-granularity reduct. Generally speaking, a natural way of finding multi-granularity reduct is to compute reduct one by one in terms of each granularity by using a searching strategy. Nevertheless, it can be observed that the mechanism may be time-consuming if the number of used granularity is greater. Therefore, to further reduce the elapsed time of computing multi-granularity reduct, an acceleration strategy will be introduced into the process of finding multi-granularity reduct. Our acceleration strategy is mainly realized by considering two variations of granularity respectively: one variation is from a finer granularity to a coarser granularity, the other variation is from a coarser granularity to a finer granularity. With such two variations, the basic thinking of our acceleration strategy is that the reduct related to the previous granularity may offer guidance for searching reduct over the present granularity. This is the core of speeding up the process of finding multi-granularity reduct.

Obviously, multi-granularity attribute reduction can be considered as a generalization of single granularity based reduct. The foundation of multi-granularity attribute reduction is based on the consideration of some inherent limitations of single granularity based attribute reduction. Furthermore, to reduce the time consumption of finding multi-granularity reduct, an acceleration strategy should be designed. It must be noticed that such a strategy is based on the consideration of variation of granularity. It is mainly because the variation of granularity may guide the computation of granularity related reduct.

The rest of this paper is organized as follows. The basic notions of neighborhood rough set and neighborhood based granularity are introduced in Section 2, it is mainly because the neighborhood approach is useful in providing a simple way for us to understand multi-granularity. In Section 3, not only the concept of multi-granularity attribute reduction is proposed, but also two acceleration algorithms are designed for finding multi-granularity reduct. The experimental results and the corresponding analyses have been addressed in Section 4. We then conclude some remarks and perspectives for future research in Section 5.

2. Preliminary knowledge

2.1. Neighborhood rough set

Formally, a decision system can be denoted as DS= $\langle U, AT, d \rangle$: U is a nonempty finite set of samples such that $U = \{x_1, x_2, \ldots, x_n\}$, it is called the universe; AT is a nonempty finite set of the condition attributes; d is the attribute of decision. $\forall x_i \in U$ and $\forall a \in AT$, $a(x_i)$ is the value of x_i over condition attribute a, and $d(x_i)$ denotes the label of sample x_i, i.e., the decision value of x_i.

Given a decision system DS, assuming that the decision values of all samples are categorical, it is not difficult to define an equivalence relation over d such that:

$$\text{IND}(\{d\}) = \{(x_i, x_j) \in U \times U : d(x_i) = d(x_j)\}.$$

By using IND($\{d\}$), a partition $U/\text{IND}(\{d\}) = \{X_1, X_2, \ldots, X_k\}$ over the universe U can be derived. $\forall X_p \in U/\text{IND}(\{d\})$, X_p is regarded as the pth decision class which contains samples with the same label. Specially, $\forall x_i \in U$, $[x_i]_d$ denotes the decision class which contains the sample x_i.

Given a decision system DS, suppose that $B \subseteq AT$ is a subset of condition attributes, $\forall x_i, x_j \in U$, dis_{ij}^B is then used to denote the Euclidean distance between x_i and x_j through using B. Note that all of the Euclidean distances between samples used in this paper have been normalized. Furthermore, if a radius $\delta \in [0, 1]$ is given, then the neighborhood relation is shown in the following Eq. (1):

$$N_B^\delta = \{(x_i, x_j) \in U \times U : dis_{ij}^B \le \delta\}. \tag{1}$$

Correspondingly, $\forall x_i \in U$, the neighborhood of x_i determined by B is:

$$\delta_B(x_i) = \{x_j \in U : dis_{ij}^B \le \delta\}. \tag{2}$$

Definition 1 ([34]). Given a decision system DS, $\delta \in [0, 1]$, $\forall B \subseteq AT$, the neighborhood lower and upper approximations of d in terms of B are defined as:

$$\underline{N_B^\delta}(d) = \bigcup_{p=1}^{k} \underline{N_B^\delta}(X_p), \tag{3}$$

$$\overline{N_B^\delta}(d) = \bigcup_{p=1}^{k} \overline{N_B^\delta}(X_p); \tag{4}$$

in which for each $X_p \in U/IND(\{d\})$, the neighborhood lower and upper approximations of decision class X_p are:

$$\underline{N_B^\delta}(X_p) = \{x_i \in U : \delta_B(x_i) \subseteq X_p\}, \tag{5}$$

$$\overline{N_B^\delta}(X_p) = \{x_i \in U : \delta_B(x_i) \cap X_p \neq \emptyset\}. \tag{6}$$

Definition 2 ([34]). Given a decision system DS, $\delta \in [0, 1]$, $\forall B \subseteq AT$, the approximation quality of d in terms of B is defined as:

$$\gamma^\delta(B, d) = \frac{|\underline{N_B^\delta}(d)|}{|U|}, \tag{7}$$

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

From the viewpoint of rough set theory, the approximation quality is frequently used to evaluate the certainty of belongingness. The greater the value of the approximation quality, the higher the degree of the belongingness.

Proposition 1 ([34]). *Given a decision system* DS, $\delta \in [0, 1]$, $\forall B', B \subseteq AT$, *if* $B' \subseteq B$ *holds, then we have* $\gamma^\delta(B', d) \leq \gamma^\delta(B, d)$.

Proposition 1 implies that the value of approximation quality is monotonic, that is, using more attributes will not contribute to a lesser value of approximation quality.

Definition 3 ([57]). Given a decision system DS, $\delta \in [0, 1]$, $\forall B \subseteq AT$, the conditional entropy of d in terms of B is defined as:

$$CE^\delta(B, d) = -\frac{1}{|U|} \sum_{x_i \in U} |\delta_B(x_i) \cap [x_i]_d| \log \frac{|\delta_B(x_i) \cap [x_i]_d|}{|\delta_B(x_i)|}. \tag{8}$$

From the viewpoint of rough set theory, the conditional entropy reflects the discrimination ability of the set of condition attributes B for the decision attribute d. The lower the value of the conditional entropy, the stronger the ability of the discrimination.

Proposition 2 ([57]). *Given a decision system* DS, $\delta \in [0, 1]$, $\forall B', B \subseteq AT$, *if* $B' \subseteq B$ *holds, then we have* $CE^\delta(B', d) \geq CE^\delta(B, d)$.

Proposition 2 implies that the value of conditional entropy is monotonic, that is, using more attributes will not increase the value of conditional entropy.

2.2. Neighborhood based granularity

Presently, it is well-known that the concept of granularity plays a key role in the development of rough set theory. Different levels of granularity may lead to different results of rough approximations and the corresponding measures, e.g., approximation quality and conditional entropy.

Neighborhood rough set provides a natural way for us to research the topic which is related to granularity in rough set theory. It is mainly because the neighborhood relation used to construct neighborhood rough set is directly determined by the value of radius. A smaller value of radius indicates a finer granularity of neighborhood relation while a greater value of radius indicates a coarser granularity of neighborhood relation. To characterize the levels of the granularity, the following definition can be used.

Definition 4 ([58,59]). Given a decision system DS, $\delta \in [0, 1]$, $\forall B \subseteq AT$, the neighborhood granularity determined by the radius δ is:

$$\mathbb{G}_B^\delta = \frac{|N_B^\delta|}{|U|^2}. \tag{9}$$

Obviously, $\frac{1}{|U|} \leq \mathbb{G}_B^\delta \leq 1$ holds. Take for instance one extreme case, if the neighborhood relation is $\omega = \{(x_i, x_i) \in U \times U : \forall x_i \in U\}$, then the finest granularity will be derived, and the value of the granularity will be the minimal one $\frac{1}{|U|}$. It follows that the approximation quality achieves the maximal value 1, while the conditional entropy achieves the minimal value 0. And in another extreme case, if the neighborhood relation is $\eta = \{(x_i, x_j) \in U \times U : \forall x_i, x_j \in U\}$, then the coarsest granularity will be derived, and the value of the granularity will be the maximal one 1. It follows that the approximation quality achieves the minimal value 0, while the conditional entropy achieves the maximal value $-\log \frac{1}{|U|}$.

Following Definition 4, it can be observed that if the neighborhood relation varies, then the value of granularity will also vary. Moreover, following Eq. (1), it can be observed that the result of neighborhood relation is determined by the value of radius and then we can conclude that different values of radius will lead to different levels of granularity.

2.3. Attribute reduction

In recent years, with respect to different requirements, different types of attribute reduction [60,61] have been proposed. However, it can be found that those definitions of attribute reduction may have similar structure, and then to extract commonness, Yao et al. have presented the following general form of attribute reduction [62,63].

Definition 5. Given a decision system DS, $\delta \in [0, 1]$, $\forall B \subseteq AT$, ρ^δ is a constraint related to δ, B is referred to as a ρ^δ-reduct if and only if :

1. B meets the constraint ρ^δ;
2. $\forall B' \subset B$, B' does not meet the constraint ρ^δ.

Definition 5 shows us a general form of attribute reduction in rough set theory. The first condition implies that the attribute subset B should satisfy the given constraint ρ^δ, the second condition implies that the attribute subset B is the smallest one, i.e., no attributes can be deleted from B if the constraint ρ^δ is required to meet.

Example 1. Generally speaking, in rough set theory, different constraints can be constructed by using different measures. Since different measures may have different explanations (e.g., the explanations can be the performances of characterizing uncertainty [64,65], classification ability [66–68] and so on), the constraint ρ^δ may have various forms. The following are some simple examples.

1. From the perspective of approximation quality, it reflects the degree of certainty that samples belong to the decision classes. If the value of approximation quality is greater, then the degree of such certainty will be higher. Therefore, constraint ρ^δ can be set as the form such that $\gamma^\delta(B, d) \geq \gamma^\delta(AT, d)$, that is, ρ^δ-reduct is the minimal subset of condition attributes, which will not contribute to a lesser value of approximation quality.

2. From the perspective of conditional entropy, it reflects the discrimination ability of condition attributes for decision attribute. If the value of conditional entropy is lower, then the ability of the discrimination will be stronger. Therefore, constraint ρ^δ can be set as the form such that $CE^\delta(B, d) \leq CE^\delta(AT, d)$, that is, ρ^δ-reduct is the minimal subset of condition attributes, which will not increase the value of conditional entropy.

To find the ρ^δ-reduct shown in Definition 5, a lot of researchers have put forward various searching strategies [69–73]. Presently, it can be observed that the heuristic searching strategy has been widely used because of its lower time complexity. It should be noticed that in the framework of heuristic searching, the significance function is a key factor. This is mainly because the significance function can be used to evaluate the importance of the candidate attribute. A general form of significance function is then defined as follows.

Definition 6. Given a decision system *DS*, $\delta \in [0, 1]$, $\forall B \subseteq AT$, ρ^δ is a constraint related to δ, the importance of attribute a with respect to B may have the following two forms such that:

$$Sig^\delta(a, B, d) = \rho^\delta(B \cup \{a\}, d) - \rho^\delta(B, d), \tag{10}$$

$$Sig^\delta(a, B, d) = \rho^\delta(B, d) - \rho^\delta(B \cup \{a\}, d). \tag{11}$$

In Definition 6, $\rho^\delta(B, d)$ is the ρ^δ-value obtained by B in terms of d. The significance function shown in Eq. (10) indicates that the greater the $\rho^\delta(B \cup \{a\})$-value, the higher the performance if attribute a is added into B; the significance function shown in Eq. (11) indicates that the lower the $\rho^\delta(B \cup \{a\})$-value, the higher the performance if attribute a is added into B.

For example, if the concept of approximation quality is required to characterize the certainty of belongingness, then the greater value of approximation quality is what we want and Eq. (10) should be used to evaluate the significance of the candidate attribute; if the concept of the conditional entropy is required to characterize the ability of discrimination, then the smaller value of conditional entropy is what we want and Eq. (11) should be used to evaluate the significance of the candidate attribute.

In the framework of heuristic searching, through using the significance function shown in Definition 6, the most important attribute will be selected in each iteration until the obtained set of attributes satisfies the constraint. The detailed process of computing reduct is shown in the following Algorithm 1.

Algorithm 1. Heuristic searching for computing reduct.

Input: A decision system DS and radius δ.
output: A ρ^δ-reduct B.
1. $B = \emptyset$, compute $\rho^\delta(AT, d)$;
2.   **Do**
     (1) $\forall a \in AT - B$, compute $Sig^\delta(a, B, d)$;
     (2) Select an attribute $b \in AT - B$ based on the set of the importance $\{Sig^\delta(a, B, d) : \forall a \in AT - B\}$;
     (3) $B = B \cup \{b\}$;
    Until constraint ρ^δ is satisfied;
3.   **While** $|B| \geq 2$ and constraint ρ^δ is satisfied
    $\forall c \in B$, compute $\rho^\delta(B - \{c\}, d)$;
    If constraint ρ^δ is satisfied
      $B = B - \{c\}$;
    End
   End
4.   **Return** B.

In Algorithm 1, the time complexity of computing reduct is $O(|U|^2 \cdot |AT|^2)$. It is mainly because in the worst case: (1) the time

complexity of computing importance of candidate attributes is $O(|U|^2 \cdot |AT|)$; (2) no attributes are redundant in AT, i.e., all of the condition attributes should be tested, and then the times of iterations in both Steps 2 and 3 are $|AT|$.

In practical application, the process of computing reduct by using Algorithm 1 requires to be executed repeatedly in terms of different values of δ. The purpose is to select a reduct with better generalization performance. As what has been pointed out in Section 2.2, different radii will lead to different levels of granularity, it follows that Algorithm 1 may be used to search multi-granularity reduct.

3. Multi-granularity attribute reduction

Following the discussions above, it is obvious that multi-granularity reduct can be derived by executing the process of Algorithm 1 repeatedly. In the following, we will firstly present the formal definition of multi-granularity attribute reduction from the viewpoint of neighborhood granularity.

Definition 7. Given a decision system *DS*, $\Theta = \{\delta_1, \delta_2, \ldots, \delta_s\}$ is a set of the ordered radii such that $\delta_1 \leq \delta_2 \leq \cdots \leq \delta_s$, $\mathbb{B} = \{B_1, B_2, \ldots, B_s\}$ is referred to as the multi-granularity reduct if and only if for each neighborhood granularity $\mathbb{G}_{AT}^{\delta_t}(1 \leq t \leq s)$, B_t is the ρ^{δ_t}-reduct.

In Definition 7, ρ^{δ_t}-reduct is actually a reduct related to δ_t if the radius δ_t is used. The attribute reduction shown in Definition 5 is a single granularity case, it is mainly because one and only one radius is considered. Different from it, the attribute reduction shown in Definition 7 is a multi-granularity case because multiple different radii have been used. Note that if one and only one radius is used in Definition 7, then Definition 7 will degenerate into Definition 5.

Based on Definition 7, we observe that the multi-granularity reduct $\mathbb{B}$ can be easily obtained by repeatedly executing Algorithm 1. However, such a process may take a substantial amount of time if many different values of radius should be tested. For instance, based on Algorithm 1, if the number of the used granularity is s, then the time complexity of computing multi-granularity reduct will be $O(|U|^2 \cdot |AT|^2 \cdot s)$.

To further reduce the time consumption for computing multi-granularity reduct, the novel searching strategy has became a necessary. In the following, we will propose two novel searching algorithms. One is called the forward accelerator based searching and the other is called the backward accelerator based searching.

3.1. Forward accelerator for multi-granularity attribute reduction

Given a decision system DS, if we consider two radii such that $\delta_1 \leq \delta_2$, then the granularity ranges from a finer one to a coarser one. The main thinking of our forward accelerator is: the process of searching ρ^{δ_2}-reduct is based on the ρ^{δ_1}-reduct. That is, to find ρ^{δ_2}-reduct, we do not begin our searching with an empty set because the searching space is AT in that case. In the framework of our forward accelerator, only the attributes in $AT - B_1$ need to be searched for deriving B_2. From this point of view, the following key steps are in the forward accelerator.

1. Find the ρ^{δ_1}-reduct B_1 by using Algorithm 1.
2. If B_1 is also the ρ^{δ_2}-reduct, then we have $\mathbb{B} = \{B_1, B_1\}$.
3. If B_1 is not the ρ^{δ_2}-reduct, then select suitable attributes from $AT - B_1$ and add them into B_1 until the constraint ρ^{δ_2} is satisfied, then generate B_2 and $\mathbb{B} = \{B_1, B_2\}$.

For example, from the perspective of approximation quality, if $\gamma^{\delta_2}(B_1, d) < \gamma^{\delta_2}(AT, d)$, then add some attributes in $AT - B_1$ into B_1, the value of approximation quality will be increased by Proposition 1. From the perspective of conditional entropy, if $CE^{\delta_2}(B_1, d) > CE^{\delta_2}(AT, d)$, then add some attributes in $AT - B_1$ into B_1, the value of conditional entropy will be decreased by Proposition 2. The process of adding attributes will be terminated if the constraint over granularity $\mathbb{G}_{AT}^{\delta_2}$ is satisfied.

Obviously, the above strategy can actually reduce the searching space if multi-granularity is required, and then it is possible for us to reduce the elapsed time for computing multi-granularity reduct. Consequently, a detailed process of computing multi-granularity reduct by using forward accelerator is designed in the following Algorithm 2.

Algorithm 2. The forward accelerator based searching.

Input: A decision system DS, set of the ordered radii $\Theta = \{\delta_1, \delta_2, \cdots, \delta_s\}$.
  output: A multi-granularity reduct $\mathbb{B} = \{B_1, B_2, \cdots, B_s\}$.
  1. Using Algorithm 1 to compute B_1;
  2. **For** $t = 2$ to s
    (1) $B_t = B_{t-1}$;
    (2) **If** B_t is the ρ^{δ_t}-reduct
      $t = t + 1$, go to **(1)**;
    Else
      Go to **(3)**;
    End
    (3) **Do**
      (i) $\forall a \in AT - B_t$, compute $Sig^{\delta_t}(a, B_t, d)$;
      (ii) Select an attribute $b \in AT - B_t$ based on the set of the importance $\{Sig^{\delta_t}(a, B_t, d) : \forall a \in AT - B_t\}$;
      (iii) $B_t = B_t \cup \{b\}$;
    Until constraint ρ^{δ_t} is satisfied;
    (4) **While** $|B_t| \geq 2$ and constraint ρ^{δ_t} is satisfied
      $\forall c \in B_t$, compute $\rho^{\delta_t}(B_t - \{c\}, d)$;
      If constraint ρ^{δ_t} is satisfied
        $B_t = B_t - \{c\}$;
      End
    End
  End
  4. **Return** $\mathbb{B} = \{B_1, B_2, \cdots, B_s\}$.

In Algorithm 2, the time complexity of computing ρ^{δ_1}-reduct is $O(|U|^2 \cdot |AT|^2)$, it is mainly because the process of finding such a reduct is the same to that of Algorithm 1. Furthermore, it should be noticed that the time complexity of computing ρ^{δ_2}-reduct is $O(|U|^2 \cdot |AT - B_1|^2)$, it is mainly because in the worst case, B_1 does not satisfy the constraint ρ^{δ_2}, more attributes in $AT - B_1$ should be searched and added into B_1, then the time complexity of computing importance of candidate attributes is $O(|U|^2 \cdot |AT - B_1|^2)$. Similarly, the time complexity of computing ρ^{δ_3}-reduct is $O(|U|^2 \cdot |AT - B_2|^2)$; $\cdots$; the time complexity of computing ρ^{δ_s}-reduct is $O(|U|^2 \cdot |AT - B_{s-1}|^2)$. Therefore, the whole time complexity of Algorithm 2 is $O(|U|^2 \cdot \sum_{t=2}^{s} |AT - B_{t-1}|^2 + |U|^2 \cdot |AT|^2)$. Obviously, such time complexity is much less than that of Algorithm 1 if multi-granularity reduct is required.

3.2. Backward accelerator for multi-granularity attribute reduction

Similar to the mentioned in forward accelerator, a backward accelerator can also be designed for finding multi-granularity reduct. If we consider two radii such that $\delta_1 \leq \delta_2$, then the main thinking of the backward accelerator is: the process of searching ρ^{δ_1}-reduct is based on ρ^{δ_2}-reduct. That is, to find ρ^{δ_1}-reduct, the following two aspects are considered: (1) only samples in $U - N_{B_2}^{\delta_2}(d)$ need to be scanned and then evaluate the importance

of the candidate attributes; (2) we do not begin our searching with an empty set, because the searching space is AT in that case. In the framework of our backward accelerator, only the attributes in $AT - B_2$ need to be searched for deriving B_1. Consequently, the following key steps are in the backward accelerator.

1. Find the ρ^{δ_2}-reduct B_2 by using Algorithm 1.
2. If B_2 is also the ρ^{δ_1}-reduct, then we have $\mathbb{B} = \{B_2, B_2\}$.
3. If B_2 is not the ρ^{δ_1}-reduct, then select suitable attributes from $AT - B_2$ and add them into B_2 until the constraint ρ^{δ_1} is satisfied, then generate B_1 and $\mathbb{B} = \{B_1, B_2\}$.

For example, from the perspective of approximation quality, if $\gamma^{\delta_1}(B_2, d) < \gamma^{\delta_1}(AT, d)$, then add some attributes in $AT - B_2$ into B_2, the value of approximation quality will be increased by Proposition 1. From the perspective of conditional entropy, if $CE^{\delta_1}(B_2, d) > CE^{\delta_1}(AT, d)$, then add some attributes in $AT - B_2$ into B_2, the value of conditional entropy will be decreased by Proposition 2. The process of adding attributes will be terminated if the constraint over granularity $\mathbb{G}_{AT}^{\delta_1}$ is satisfied.

Similar to what has been discussed in Section 3.1, the above strategy can also reduce the searching space if multi-granularity is required, and then it is possible for us to reduce the elapsed time of computing multi-granularity reduct. Consequently, a detailed process for computing multi-granularity reduct by using backward accelerator is designed in the following Algorithm 3.

Algorithm 3. The backward accelerator based searching.

Input: A decision system DS, set of the ordered radii $\Theta = \{\delta_1, \delta_2, \cdots, \delta_s\}$.
  output: A multi-granularity reduct $\mathbb{B} = \{B_1, B_2, \cdots, B_s\}$.
  1. Using Algorithm 1 to compute B_s;
  2. **For** $t = s - 1$ to 1
    (1) $B_t = B_{t+1}$, $U' = U - N_{B_{t+1}}^{\delta_{t+1}}(d)$;
    (2) **If** B_t is the ρ^{δ_t}-reduct
      $t = t - 1$, go to **(1)**;
    Else
      Go to **(3)**;
    End
    (3) **Do**
      (i) $\forall a \in AT - B_t$, compute $Sig^{\delta_t}(a, B_t, d)$ over U';
      (ii) Select an attribute $b \in AT - B_t$ based on the set of the importance $\{Sig^{\delta_t}(a, B_t, d) : \forall a \in AT - B_t\}$;
      (iii) $B_t = B_t \cup \{b\}$;
    Until constraint ρ^{δ_t} is satisfied;
    (4) **While** $|B_t| \geq 2$ and constraint ρ^{δ_t} is satisfied
      $\forall c \in B_t$, compute $\rho^{\delta_t}(B_t - \{c\}, d)$;
      If constraint ρ^{δ_t} is satisfied
        $B_t = B_t - \{c\}$;
      End
    End
  End
  4. **Return** $\mathbb{B} = \{B_1, B_2, \cdots, B_s\}$.

In Algorithm 3, the time complexity of computing ρ^{δ_s}-reduct is $O(|U|^2 \cdot |AT|^2)$, it is mainly because the process of finding such a reduct is the same to that of Algorithm 1. Furthermore, it should be noticed that the time complexity of computing $\rho^{\delta_{s-1}}$-reduct is $O(|U|^2 \cdot |AT - B_s|^2)$, it is mainly because in the worst case: (1) $N_{B_s}^{\delta_s}(d)$ is an empty set; (2) if B_s does not satisfy the constraint $\rho^{\delta_{s-1}}$, more attributes in $AT - B_s$ should be searched and added into B_s, then the time complexity of computing importance of candidate attributes is $O(|U|^2 \cdot |AT - B_s|^2)$. Similarly, the time complexity of computing $\rho^{\delta_{s-2}}$-reduct is $O(|U|^2 \cdot |AT - B_{s-1}|^2)$; $\cdots$; the time complexity of computing the ρ^{δ_1}-reduct is $O(|U|^2 \cdot |AT - B_2|^2)$. Therefore, the whole time complexity of Algorithm

3 is $O(|U|^2 \cdot \sum_{t=1}^{s-1} |AT - B_{t+1}|^2 + |U|^2 \cdot |AT|^2)$. Obviously, such time complexity is also much less than that of Algorithm 1 if multi-granularity reduct is required.

To understand the process of finding multi-granularity reduct by using our acceleration strategy clearly, an example will be shown as follows. This example not only reveals the mechanism of our proposed acceleration strategy, but also shows us how the searching space can be reduced. In the following Example 2, the approximation quality is used as the measure to derive multi-granularity reduct.

Example 2. Given a decision system DS=$\langle U, AT, d \rangle$, in which $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}$, $AT = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}\}$, and d is the decision attribute. We consider two radii δ_1=0.3, δ_2=0.4, that is, two levels of granularity will be used (see Table 1).

Based on Algorithm 2, firstly, the $\mathbb{G}_{AT}^{\delta_1}$ related reduct B_1 can be derived by using Algorithm 1 such that $B_1 = \{a_1, a_2, a_3, a_7\}$, that is, $\gamma^{\delta_1}(B_1, d) \geq \gamma^{\delta_1}(AT, d)$. Secondly, it is not difficult to validate that B_1 is also the reduct over $\mathbb{G}_{AT}^{\delta_2}$, that is, $\gamma^{\delta_2}(B_1, d) \geq \gamma^{\delta_2}(AT, d)$ also holds. Therefore, $B_2 = B_1 = \{a_1, a_2, a_3, a_7\}$ and then we have $\mathbb{B} = \{B_1, B_2\}$.

Based on Algorithm 3, firstly, the $\mathbb{G}_{AT}^{\delta_2}$ related reduct B_2 can be derived by using Algorithm 1 such that $B_2 = \{a_1, a_2, a_4, a_7\}$. Secondly, with calculation, it is not difficult to validate that B_2 is not the reduct over $\mathbb{G}_{AT}^{\delta_1}$, that is, $\gamma^{\delta_1}(B_2, d) \geq \gamma^{\delta_1}(AT, d)$ does not hold. Therefore, let $U' = U - N_{B_2}^{\delta_2}(d)$ and select important attributes in $AT - B_2 = \{a_3, a_5, a_6, a_8, \overline{a_9}, a_{10}\}$, until the constraint is satisfied, if follows that $B_1 = \{a_1, a_2, a_4, a_6, a_7\}$ and $\mathbb{B} = \{B_1, B_2\}$.

By the above example, it is not difficult to observe that the searching space of finding the present granularity related reduct is reduced by using our acceleration strategy. It follows that the time complexity will also be decreased if our forward and backward accelerators are used.

4. Experimental analysis

To demonstrate the effectiveness of our forward and backward accelerators, 16 UCI data sets have been selected to conduct the experiments. The details of used data sets are shown in the following Table 2.

In the context of this paper, not only 5-fold cross-validation is employed, but also 20 different radii have been used, i.e., $\Theta = \{0.025, \ldots, 0.500\}$. That is to say, 20 different levels of granularity have been considered in our experiments. The detailed process is as follows. Each data set is partitioned into 5 disjoint groups such that $U = \{U_1, U_2, U_3, U_4, U_5\}$. In the first round of calculation, $U_2 \cup U_3 \cup U_4 \cup U_5$ is regarded as the training set for deriving 20 reducts with respect to 20 radii, U_1 is the testing set for classification by using these 20 different reducts; $\cdots$; in the last round of calculation, $U_1 \cup U_2 \cup U_3 \cup U_4$ is regarded as the training set for deriving 20 reducts with respect to 20 radii, U_5 is the testing set for classification by using these 20 different reducts. Moreover, the measures related to both approximation quality and conditional entropy will be used in our experiments. In other words, it is expected to find reducts derived from different strategies which can offer greater values of approximation quality or lower values of conditional entropy.

Furthermore, the Kolmogorov–Smirnov (K–S) test will be employed for comparing the traditional strategy with our acceleration strategy. The purpose of our comparison is trying to reject the null-hypothesis that such two strategies perform equally well. Assuming that the significance level is set as 0.05 and if

the p-value of K–S test is lower than 0.05, we then reject the null-hypothesis.

Note that in the following subsections, in terms of the measure of approximation quality, "γ-reduct" denotes the results related to reducts which are derived by Algorithm 1, while "$\uparrow$-γ-reduct" denotes the results related to reducts which are derived by Algorithm 2, and "$\downarrow$-γ-reduct" denotes the results related to reducts which are derived by Algorithm 3.

Similarly, in terms of the measure of conditional entropy, "CE-reduct" denotes the results related to reducts which are derived by Algorithm 1, while "$\uparrow$-CE-reduct" denotes the results related to reducts which are derived by Algorithm 2, and "$\downarrow$-CE-reduct" denotes the results related to reducts which are derived by Algorithm 3.

Moreover, it should be stressed that the reducts derived by using Algorithm 1, Algorithm 2 and Algorithm 3 are multi-granularity reducts, it is mainly because 20 different radii are used to find reducts and different radii indicate different levels of granularity.

4.1. Comparisons of elapsed time

In this section, the elapsed time for computing multi-granularity reducts by using different algorithms will be compared. The detailed results are shown in the following Fig. 1.

With a deep investigation of Fig. 1, it is not difficult to observe the following.

1. Compared with the process of computing γ-reducts, the process of searching $\uparrow$-γ-reducts is characterized with lower time consumption. Similarly, compared with the process of generating CE-reducts, the process of searching $\uparrow$-CE-reducts is also quick. Such results indicate that our forward accelerator shown in Algorithm 2 can significantly speed up the process of finding multi-granularity reduct. Moreover, it should be noticed that for the first radius 0.025, i.e., the first granularity, the time consumption of Algorithm 2 is similar to that of Algorithm 1. It is mainly because in Algorithm 2, the process of finding this reduct is the same to that of Algorithm 1.
2. Compared with the process of computing γ-reducts, the process of searching $\downarrow$-γ-reducts is characterized with lower time consumption. Similarly, compared with the process of generating CE-reducts, the process of searching $\downarrow$-CE-reducts is also quick. Such results indicate that our backward accelerator shown in Algorithm 3 can also significantly speed up the process of finding multi-granularity reduct. Moreover, it should be noticed that for the last radius 0.500, i.e., the last granularity, the time consumption of Algorithm 3 is also similar to that of Algorithm 1. It is mainly because in Algorithm 3, the process of finding this reduct is the same to that of Algorithm 1.

To further analyze the elapsed time of computing reducts by different strategies from the viewpoint of statistics, the K–S test will be employed. The purpose of K–S test is trying to reject the null-hypothesis that two compared methods perform equally well. The significance level is set as 0.05 in this paper, if the obtained p-value is lower than 0.05, then we reject the null-hypothesis. The detailed results of p-values are shown in Table 3.

Following the results shown in Table 3, it is not difficult to observe the following.

1. For the comparison of the elapsed time of computing $\uparrow$-γ-reduct and γ-reduct, all obtained p-values are lower than 0.05. Similarly, For the comparison of the elapsed time

Table 1
A toy example of data.

Samples	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	d
x_1	0.9949	0.7981	0.5359	0.0291	0.8866	0.6359	0.7325	0.9612	0.2998	0.7561	1
x_2	0.4906	0.2204	0.8291	0.1362	0.9703	0.7898	0.1629	0.4664	0.1591	0.2882	1
x_3	0.5035	0.8579	0.2674	0.6946	0.9425	0.5663	0.9211	0.7870	0.6653	0.6062	1
x_4	0.7688	0.9047	0.1762	0.5157	0.6381	0.3774	0.2222	0.4226	0.6842	0.7660	2
x_5	0.3881	0.2920	0.4312	0.5426	0.0906	0.8216	0.0836	0.9437	0.7924	0.8462	2
x_6	0.4533	0.7259	0.4757	0.8085	0.0747	0.3049	0.0737	0.0013	0.3486	0.9020	2
x_7	0.1329	0.3394	0.7852	0.7937	0.1825	0.3194	0.7696	0.9813	0.2501	0.5957	2
x_8	0.7585	0.2727	0.1307	0.5019	0.0317	0.7850	0.8177	0.5702	0.3450	0.0685	3
x_9	0.5652	0.1703	0.0514	0.2766	0.7249	0.5037	0.7404	0.3465	0.3286	0.2180	3
x_{10}	0.6486	0.6640	0.6275	0.1197	0.1442	0.2610	0.7582	0.5575	0.9275	0.8694	3

Table 2
Data sets description.

| ID	Data sets	Samples	Attributes	Decision classes
1	Breast cancer wisconsin (Diagnostic)	569	30	2
2	Cardiotocography	2126	21	10
3	Dermatology	366	34	6
4	Forest type mapping	523	27	4
5	Libras movement	360	90	15
6	Page blocks classification	5473	10	5
7	QSAR biodegradation	1055	41	2
8	Statlog (Image segmentation)	2310	18	7
9	Statlog (Landsat satellite)	6435	36	6
10	Steel plates faults	1941	33	2
11	Waveform database generator (Version 1)	5000	21	3
12	Waveform database generator (Version 2)	5000	40	3
13	Wall-following robot navigation	5456	24	4
14	Website phishing	1353	9	2
15	Wine quality	4898	11	7
16	Wireless indoor localization	2000	7	4
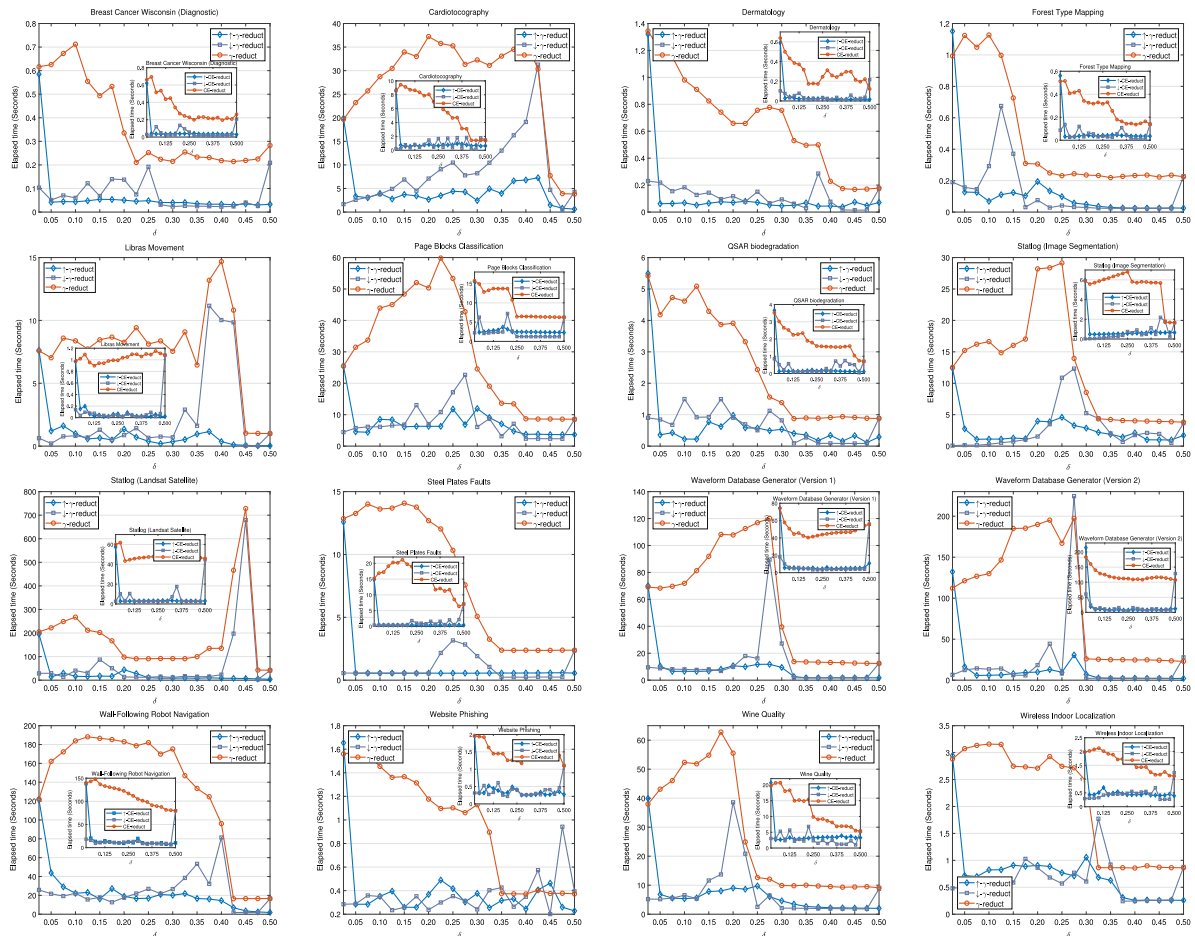
Fig. 1. Comparisons of elapsed time for computing different reducts.

Table 3
p-values for comparing elapsed time of deriving reducts.

ID	γ-reduct & ↑-γ-reduct	γ-reduct & ↓-γ-reduct	CE-reduct & ↑-CE-reduct	CE-reduct & ↓-CE-reduct
1	4.74E−09	5.55E−10	4.74E−09	4.74E−09
2	2.49E−07	1.53E−06	4.74E−09	1.53E−06
3	4.74E−09	8.42E−06	4.74E−09	4.74E−09
4	4.74E−09	1.53E−06	4.74E−09	4.74E−09
5	1.53E−06	4.15E−05	4.74E−09	4.74E−09
6	1.53E−06	8.42E−06	4.74E−09	3.63E−08
7	3.63E−08	7.25E−04	4.74E−09	3.63E−08
8	1.53E−06	1.53E−06	4.74E−09	2.49E−07
9	3.63E−08	1.53E−06	4.74E−09	4.74E−09
10	4.74E−09	3.63E−08	4.74E−09	4.74E−09
11	4.74E−09	8.42E−06	4.74E−09	4.74E−09
12	3.63E−08	2.49E−07	4.74E−09	4.74E−09
13	8.42E−06	1.53E−06	4.74E−09	5.55E−10
14	1.83E−04	8.42E−06	4.74E−09	4.74E−09
15	3.63E−08	1.53E−06	4.74E−09	3.63E−08
16	4.15E−05	8.42E−06	4.74E−09	4.74E−09

Table 4
Comparisons among stabilities of reducts (the greater values are in bold).

ID	↑-γ-reduct	↓-γ-reduct	γ-reduct	↑-CE-reduct	↓-CE-reduct	CE-reduct
1	**0.7390**	0.6044	0.3213	**0.9046**	0.6789	0.7224
2	**0.4141**	0.2944	0.2983	**0.7926**	0.4806	0.3857
3	**0.6414**	0.4376	0.3545	**0.7569**	0.6185	0.3538
4	0.4279	**0.5986**	0.2964	**0.9400**	0.6311	0.3665
5	**0.4580**	0.3580	0.1378	0.7565	**0.8036**	0.2015
6	**0.5416**	0.4296	0.3946	**0.8477**	0.7033	0.7072
7	**0.4141**	0.4126	0.2577	**0.7810**	0.5289	0.5243
8	**0.5292**	0.3042	0.3361	**0.9842**	0.7143	0.6864
9	**0.5185**	0.4217	0.2597	**0.9312**	0.7337	0.3540
10	**1.0000**	0.5315	0.5315	**1.0000**	0.5815	0.4666
11	**0.4563**	0.4105	0.3653	**0.8911**	0.8796	0.6254
12	**0.4867**	0.4347	0.3414	**0.9159**	0.7568	0.4578
13	**0.4719**	0.4174	0.4060	**0.8313**	0.6735	0.6050
14	**0.6310**	0.5816	0.4520	**0.7739**	0.6924	0.5476
15	**0.4871**	0.4767	0.3228	**0.9199**	0.6111	0.6123
16	**0.5191**	0.4677	0.4799	**0.8578**	0.6965	0.5932

Table 5
p-values for comparing classification accuracies based on different reducts (CART classifier).

ID	γ-reduct & ↑-γ-reduct	γ-reduct & ↓-γ-reduct	CE-reduct & ↑-CE-reduct	CE-reduct & ↓-CE-reduct
1	7.25E−04	0.0232	7.25E−04	*0.9655*
2	*0.9655*	*0.4973*	2.49E−07	*0.9655*
3	*0.0591*	*0.9655*	2.49E−07	0.0082
4	*0.2753*	*0.2753*	8.42E−06	0.0082
5	0.0026	*0.4973*	2.49E−07	0.0082
6	0.0232	*0.7710*	1.83E−04	*0.9999*
7	0.0232	*0.7710*	3.63E−08	0.0232
8	0.0232	*0.4973*	7.25E−04	*1.0000*
9	*0.1349*	0.0232	3.63E−08	*0.1349*
10	0.0026	*1.0000*	8.42E−06	*0.9999*
11	0.0232	*0.7710*	3.63E−08	1.53E−06
12	0.0232	*0.0591*	8.42E−06	1.53E−06
13	*0.7710*	*0.9999*	0.0026	*0.2753*
14	*0.2753*	*0.2753*	0.0232	*0.9655*
15	7.25E−04	*0.7710*	0.0232	0.0082
16	*0.7710*	0.0232	4.74E−09	*0.2753*

Table 6
p-values for comparing classification accuracies based on different reducts (KNN classifier).

ID	γ-reduct & ↑-γ-reduct	γ-reduct & ↓-γ-reduct	CE-reduct & ↑-CE-reduct	CE-reduct & ↓-CE-reduct
1	7.25E−04	0.0082	0.0026	*0.9999*
2	*0.4973*	*0.4973*	2.49E−07	0.0232
3	0.0232	*0.9999*	2.49E−07	0.0232
4	*0.0591*	0.0082	0.0026	0.0232
5	*0.9655*	*0.4973*	1.53E−06	*0.4973*
6	0.0232	*0.7710*	1.83E−04	*0.9999*
7	*0.0591*	*0.1349*	4.15E−05	*0.0591*
8	0.0232	*0.2753*	7.25E−04	*1.0000*
9	*0.1349*	0.0026	2.49E−07	*0.4973*
10	0.0026	*1.0000*	8.42E−06	*0.9999*
11	0.0232	*1.0000*	1.53E−06	1.53E−06
12	0.0232	*0.0591*	8.42E−06	1.53E−06
13	*0.7710*	*0.9655*	*0.1349*	*0.2753*
14	0.0232	*0.1349*	*0.1349*	*0.7710*
15	*0.1349*	*0.4973*	0.0232	*1.0000*
16	*0.4973*	*0.7710*	3.63E−08	0.0026

Table 7
p-values for comparing classification accuracies based on different reducts (SVM classifier).

ID	γ-reduct & ↑-γ-reduct	γ-reduct & ↓-γ-reduct	CE-reduct & ↑-CE-reduct	CE-reduct & ↓-CE-reduct
1	1.83E−04	*0.4973*	7.25E−04	*0.9999*
2	*0.9655*	*0.4973*	1.53E−06	0.0082
3	*0.0591*	*0.7710*	4.74E−09	0.0082
4	*0.1349*	0.0232	0.0026	0.0082
5	*0.1349*	*0.2753*	3.63E−08	7.25E−04
6	0.0232	*0.9655*	1.83E−04	*0.7710*
7	0.0232	*0.9999*	4.74E−09	0.0232
8	0.0232	*0.1349*	7.25E−04	*1.0000*
9	*0.1349*	0.0082	3.63E−08	*0.2753*
10	0.0026	*1.0000*	1.53E−06	2.49E−07
11	0.0232	*0.9655*	3.63E−08	1.53E−06
12	0.0232	*0.4973*	0.0082	1.53E−06
13	*0.4973*	*0.9655*	*0.7710*	*0.4973*
14	*0.4973*	*0.2753*	*0.9655*	*0.4973*
15	1.83E−04	*0.7710*	0.0082	*1.0000*
16	*0.2753*	*0.0591*	4.74E−09	0.0026

our backward accelerator and the traditional approach also perform significantly different in time consumption.

Moreover, considering the results shown in both Fig. 1 and Table 3, it is not difficult for us to conclude that compared with the traditional strategy, i.e., Algorithm 1, our forward accelerator and backward accelerator possess the significant advantage in speeding up the process of finding multi-granularity reduct.

4.2. Comparisons of lengths

In this section, the lengths of different reducts will be compared. The following Fig. 2 reports the detailed results.

Following Fig. 2, it is not difficult to observe the following.

1. For most of the data sets and radii we tested in our experiments, the lengths of ↑-γ-reducts and ↓-γ-reducts are greater than those of γ-reducts. Similarly, the lengths of ↑-CE-reducts and ↓-CE-reducts are also greater than those of CE-reducts. Therefore, we can conclude that Algorithm 2 and Algorithm 3 require more attributes to construct reducts. Moreover, considering both this conclusion and the results shown in Fig. 1, we can observe that though the reducts derived from our acceleration strategy contain more attributes, the elapsed time of our strategy is still lower than that of traditional strategy. Obviously, our acceleration strategy is superior to traditional strategy from the viewpoint of time consumption.
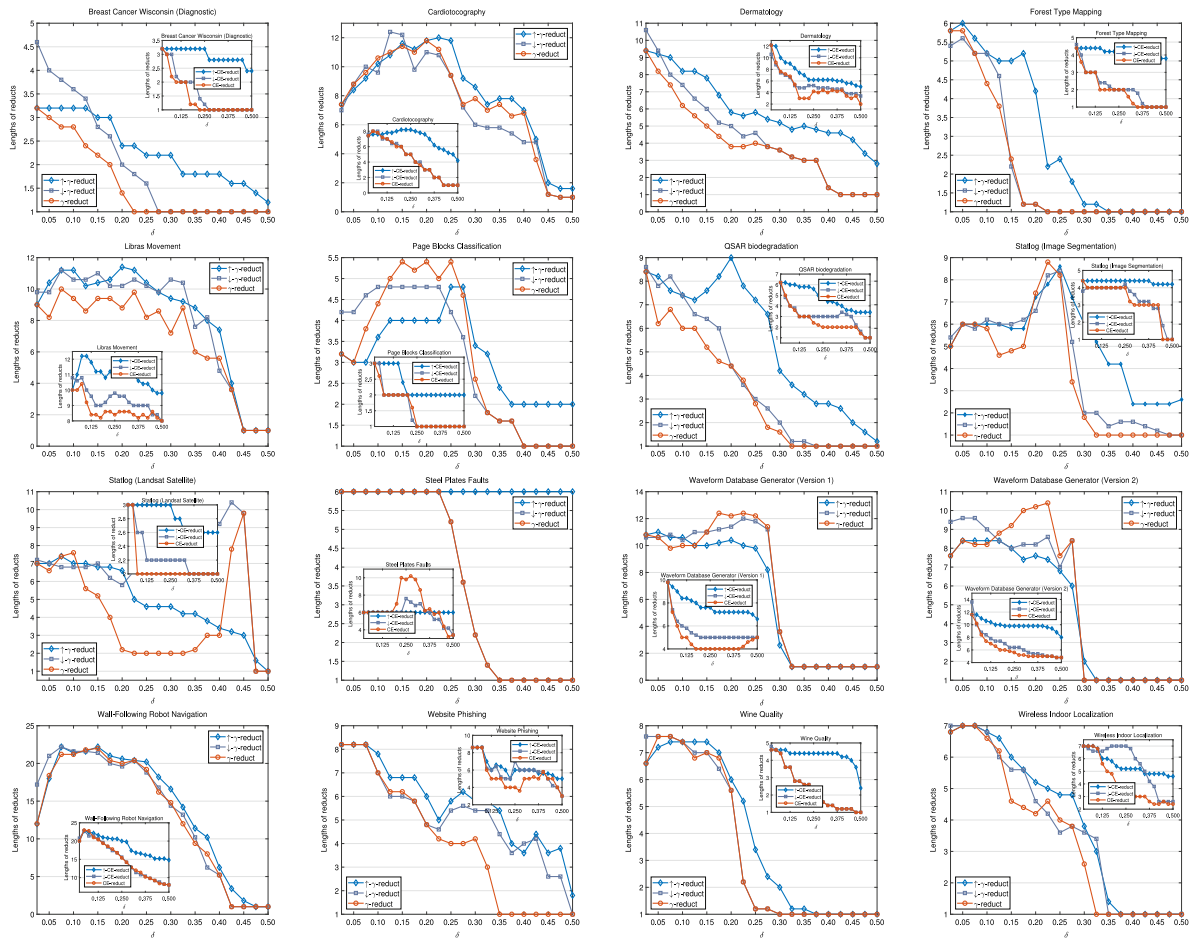
of computing ↑-CE-reduct and CE-reduct, all obtained p-values are also lower than 0.05. Such results imply that our forward accelerator and the traditional approach perform significantly different in time consumption.

2. For the comparison of the elapsed time of computing ↓-γ-reduct and γ-reduct, all obtained p-values are lower than 0.05. Similarly, For the comparison of the elapsed time of computing ↓-CE-reduct and CE-reduct, all obtained p-values are also lower than 0.05. Such results imply that

Fig. 2. Comparisons among different lengths w.r.t. different reducts.

2. Furthermore, it can be observed that in most of the cases, the lengths of $\uparrow$-γ-reducts are greater than those of $\downarrow$-γ-reducts, the lengths of $\uparrow$-CE-reducts are also greater than those of $\downarrow$-CE-reducts. Consequently, we can say that by comparing with Algorithm 3, Algorithm 2 requires more attributes to construct reducts.

4.3. Comparisons of stabilities

Recently, the stability of reduct has been paid much attention to by some researchers. Note that the stability of reduct they studied indicates that the degree of variation of reducts if the used samples have been changed [51,62,74]. Different from it, in the following, we will consider another type of stability of reduct, which is expected to reflect the variation of reducts if granularity has been changed. Since in our experiment, different radii have been used, and different radii imply different levels of granularity, the following Eq. (12) can be used to compute the stability of reduct.

$$Sta = \sum_{t=1}^{s-1} \frac{|B_t \cap B_{t+1}|}{|B_t \cup B_{t+1}|}, \tag{12}$$

in which B_t denotes the ρ^{δ_t}-reduct.

Obviously, $Sta \in [0,1]$ holds. For any two reducts related different radii, Sta achieves the minimal value 0 if and only if $B_t \cap B_{t+1} = \emptyset$, it implies that the variation of radii, i.e., granularity will result in a great effect on the results of reduct, and then the obtained multi-granularity reduct is completely unstable; Sta achieves the maximal value 1 if and only if $B_t = B_{t+1}$, it implies

that the variation of radii, i.e., granularity has no effect on the results of reduct, and then the obtained multi-granularity reduct is completely stable.

The detailed results of stabilities of different types of multi-granularity reduct are shown in Table 4.

With a deep investigation of Table 4, it is not difficult to observe that for most of the data sets, the values of stability of $\uparrow$-γ-reduct and $\downarrow$-γ-reduct are greater than those of γ-reduct, the values of stability of $\uparrow$-CE-reduct and $\downarrow$-CE-reduct are also greater than those of CE-reduct. The results indicate that the multi-granularity reducts derived by using our acceleration strategy are more stable than those derived by using traditional searching strategy, i.e., Algorithm 1.

Moreover, it must be emphasized that compared with the backward accelerator, our forward accelerator performs better from the viewpoint of stability. In other words, the multi-granula rity reducts derived from Algorithm 2 and Algorithm 3 are more stable.

4.4. Comparisons of classification performances

In this section, the classification performances of multi-gran ula rity reducts will be further compared. It should be noticed that CART, KNN and SVM (LIBSVM [75]) classifiers are employed for testing the classification performances. The detailed results are shown in the following Figs. 3–5.

With a careful investigation of Figs. 3–5, it is not difficult to observe that the classification accuracies derived from $\uparrow$-γ-reducts and $\downarrow$-γ-reducts are greater than or equal to those de-rived from γ-reducts, the classification accuracies derived from
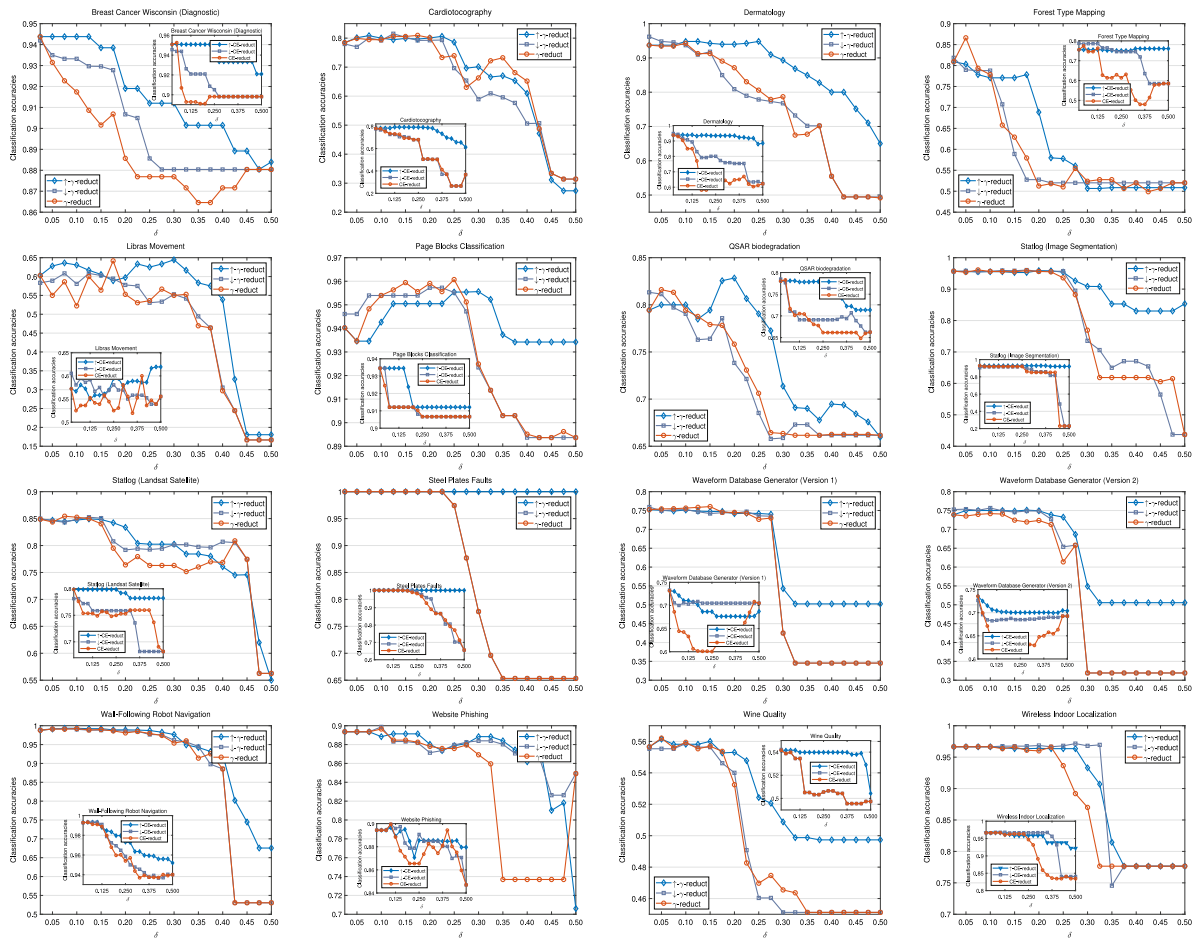
Fig. 3. Comparisons among classification accuracies w.r.t. different reducts (CART classifier).

Table 8
The values of radius w.r.t. the maximal classification accuracies (CART classifier).

ID	↑-γ-reduct	↓-γ-reduct	γ-reduct	↑-CE-reduct	↓-CE-reduct	CE-reduct
1	0.025	0.025	0.025	0.025	0.025	0.025
2	0.075	0.075	0.175	0.025	0.025	0.025
3	0.500	0.025	0.100	0.125	0.025	0.025
4	0.025	0.025	0.050	0.025	0.025	0.050
5	0.075	0.075	0.175	0.475	0.025	0.400
6	0.250	0.200	0.250	0.025	0.025	0.025
7	0.200	0.025	0.050	0.075	0.025	0.050
8	0.200	0.175	0.075	0.025	0.050	0.025
9	0.125	0.125	0.075	0.025	0.025	0.025
10	0.025	0.025	0.025	0.025	0.025	0.025
11	0.025	0.025	0.150	0.025	0.025	0.025
12	0.175	0.100	0.150	0.025	0.025	0.025
13	0.100	0.075	0.100	0.050	0.050	0.050
14	0.025	0.100	0.100	0.100	0.100	0.100
15	0.050	0.100	0.050	0.050	0.100	0.050
16	0.025	0.275	0.025	0.025	0.075	0.025

↑-CE-reducts and ↓-CE-reducts are also greater than or equal to those derived from CE-reducts. From this point of view, we can conclude that our forward and backward accelerators may select attributes with higher generalization performance.

Moreover, similar to Section 4.1, the K–S test is also employed for comparing the classification accuracies derived from different multi-granularity reducts. The detailed results are listed in Tables 5–7. Note that the *p*-values are greater than 0.05 are in italic.

Following the results shown in Tables 5–7, it is not difficult to observe the following.

1. For the comparison of the classification accuracies derived from ↑-γ-reducts and γ-reducts, the obtained *p*-values are less than 0.05 in some data sets, while the derived *p*-values are greater than 0.05 in some other data sets, such a result indicates that from the viewpoint of approximation quality, our forward accelerator and traditional approach do not perform equally well in some data sets, while they perform equally well in another data sets. For the comparison of the classification accuracies derived from ↓-γ-reducts and γ-reducts, most of the obtained *p*-values are greater than 0.05, such a result indicates that our backward accelerator
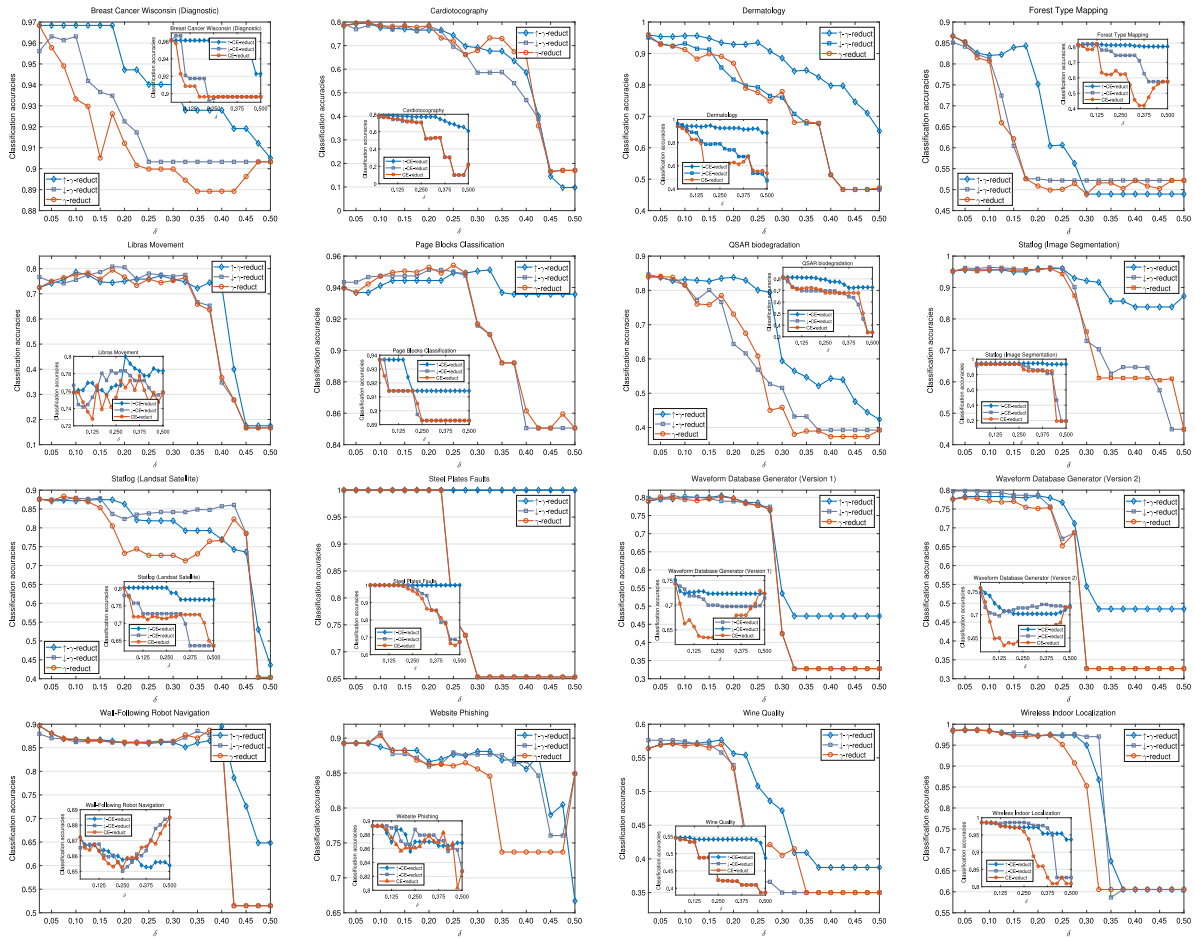
Fig. 4. Comparisons among classification accuracies w.r.t. different reducts (KNN classifier).

Table 9
The values of radius w.r.t. the maximal classification accuracies (KNN classifier).

ID	↑-γ-reduct	↓-γ-reduct	γ-reduct	↑-CE-reduct	↓-CE-reduct	CE-reduct
1	0.025	0.050	0.025	0.025	0.050	0.025
2	0.050	0.025	0.025	0.050	0.025	0.025
3	0.025	0.025	0.025	0.200	0.025	0.025
4	0.025	0.025	0.025	0.025	0.075	0.125
5	0.100	0.175	0.175	0.300	0.225	0.375
6	0.325	0.200	0.250	0.025	0.025	0.025
7	0.025	0.025	0.050	0.025	0.025	0.025
8	0.200	0.100	0.225	0.200	0.100	0.225
9	0.025	0.100	0.075	0.025	0.025	0.025
10	0.025	0.025	0.025	0.025	0.025	0.025
11	0.175	0.075	0.175	0.025	0.025	0.025
12	0.200	0.050	0.050	0.050	0.050	0.050
13	0.025	0.350	0.025	0.025	0.500	0.500
14	0.025	0.100	0.100	0.025	0.025	0.025
15	0.175	0.025	0.075	0.025	0.025	0.025
16	0.050	0.025	0.050	0.025	0.025	0.025

and traditional approach perform equally well from the viewpoint of approximation quality.

2. For the comparison of the classification accuracies derived from ↑-CE-reducts and CE-reducts, the obtained *p*-values are much less than 0.05, such a result implies that our forward accelerator and the traditional approach perform significantly different. For the comparison of the classification accuracies derived from ↓-CE-reducts and CE-reducts, the obtained *p*-values are less than 0.05 in some data sets, while the derived *p*-values are greater than 0.05 in some other data sets, the result implies that from the viewpoint

of conditional entropy, our backward accelerator and traditional approach do not perform equally well in some data sets, while they perform equally well in another data sets.

In summary, the above analyses imply that the multi-granularity reducts obtained by using our forward and backward accelerators will not contribute to a poorer classification performance. Specially, the multi-granularity reduct computed by using our forward accelerator can improve the classification performance from the perspective of conditional entropy.

Furthermore, to find the granularity with better generalization performance, the values of radii with respect to the maximal
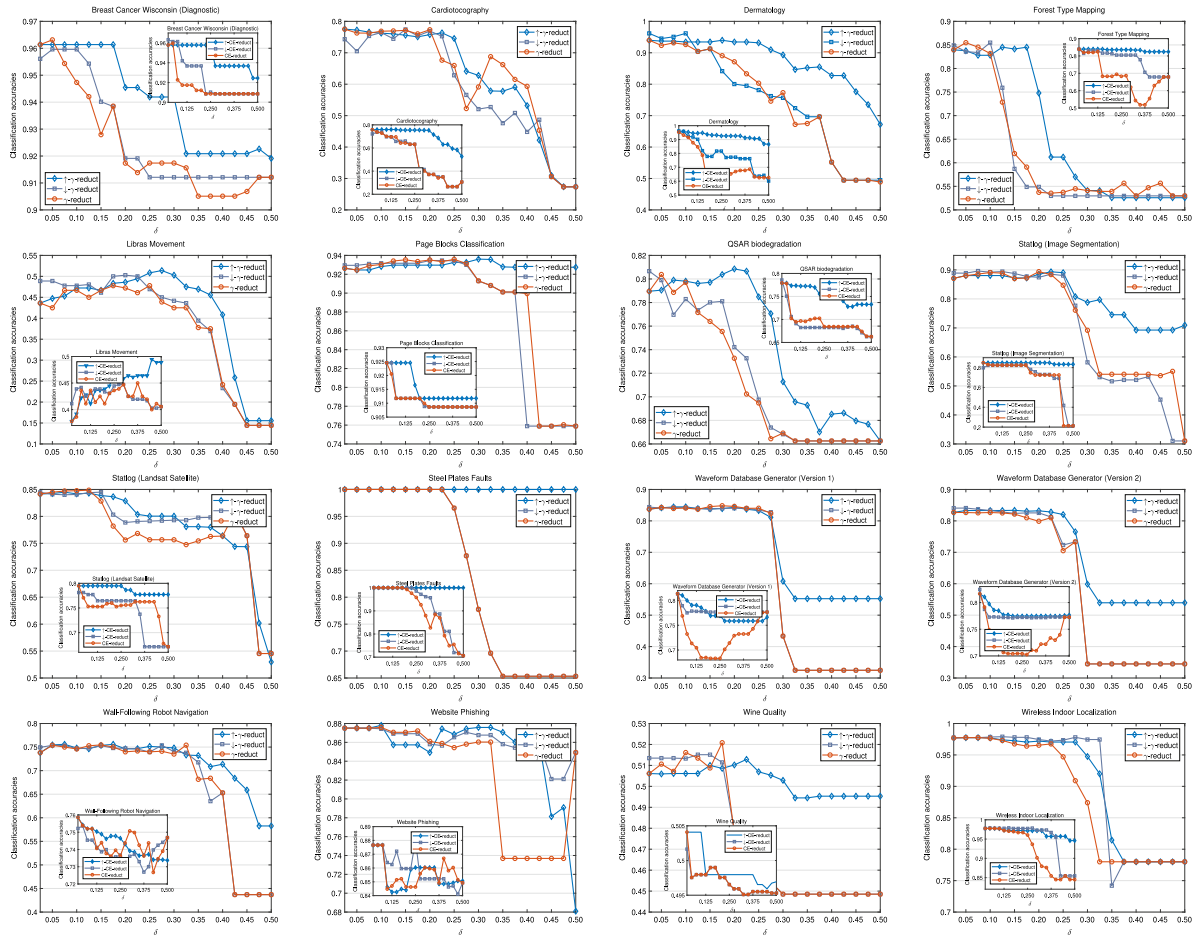
Fig. 5. Comparisons among classification accuracies w.r.t. different reducts (SVM classifier).

Table 10
The values of radius w.r.t. the maximal classification accuracies (SVM classifier).

ID	↑-γ-reduct	↓-γ-reduct	γ-reduct	↑-CE-reduct	↓-CE-reduct	CE-reduct
1	0.025	0.050	0.050	0.025	0.025	0.050
2	0.025	0.150	0.025	0.125	0.050	0.025
3	0.025	0.025	0.025	0.050	0.025	0.025
4	0.125	0.025	0.050	0.025	0.025	0.025
5	0.275	0.200	0.175	0.450	0.275	0.375
6	0.300	0.200	0.250	0.025	0.025	0.025
7	0.200	0.025	0.050	0.025	0.025	0.025
8	0.225	0.125	0.200	0.025	0.050	0.025
9	0.075	0.150	0.125	0.025	0.025	0.025
10	0.025	0.025	0.025	0.025	0.025	0.025
11	0.100	0.025	0.175	0.025	0.025	0.025
12	0.150	0.050	0.025	0.025	0.025	0.025
13	0.025	0.050	0.150	0.025	0.050	0.025
14	0.100	0.025	0.100	0.025	0.025	0.025
15	0.225	0.125	0.175	0.025	0.025	0.025
16	0.050	0.125	0.050	0.025	0.050	0.025

classification accuracies can be obtained. The detailed results are shown in Tables 8–10.

Tables 8–10 actually show us the results of selection of granularity, i.e., radii. It follows that the selected radii can generate reducts that provide us with higher classification performances.

5. Conclusions and future perspectives

In this paper, to find the multi-granularity reduct, an acceleration strategy is introduced into the process of searching attributes. Different from the traditional searching strategy which can only be executed on one and only one granularity, our proposed acceleration strategy can actually speed up the process of finding reducts in terms of multiple levels of granularity. This is mainly because the reduct derived from the previous granularity may guide the computation of reduct over the present granularity, it follows that the lower time complexity is required in our acceleration strategy. Furthermore, the experimental results have demonstrated that our acceleration strategy can not only reduce the elapsed time significantly, but also derive attributes which will not contribute to a poorer classification performance. In other words, our acceleration strategy is both efficiency and effectiveness.

The following topics are challenges for our further researches.

1. The proposed acceleration strategy can be realized through using other extended rough set models, e.g., fuzzy rough set.
2. More measures will be used in further exploration, e.g., conditional discrimination index and decision error rate.
3. More types of the expression of multi-granularity will be addressed, e.g., the multi-granularity derived from sample selecting or attribute clustering.

Acknowledgment

References

[1] 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, Knowl.-Based Syst. 91 (2016) 152–164.

[2] Z. Pawlak, Rough set, Internat. J. Comput. Inf. Sci. 11 (1982) 341–356.

[3] Y.H. She, X.L. He, H.X. Shi, Y.H. Qian, A multiple-valued logic approach for multigranulation rough set model, Int. Approx. Reason. 82 (2017) 270–284.

[4] Y.H. She, X.L. He, Algebraic structures related to nilpotent minimum algebras and rough sets, J. Intell. Fuzzy Syst. 29 (2015) 1367–1380.

[5] L.N. Wang, X. Yang, Y. Chen, L. Liu, S.Y. An, P. Zhuo, Dynamic composite decision-theoretic rough set under the change of attributes, Int. J. Comput. Intell. Syst. 11 (2018) 355–370.

[6] S. Wang, T.R. Li, C. Luo, H.M. Chen, H. Fujita, Domain-wise approaches for updating approximations with multi-dimensional variation of ordered information systems, Inform. Sci. 478 (2019) 100–124.

[7] W.H. Xu, W.X. Sun, Y.F. Liu, W.X. Zhang, Fuzzy rough set models over two universes, Int. J. Mach. Learn. Cybern. 4 (2013) 631–645.

[8] X. Yang, T.R. Li, H. Fujita, D. Liu, Y.Y. Yao, A unified model of sequential three-way decisions and multilevel incremental processing, Knowl.-Based Syst. 134 (2017) 172–188.

[9] X. Yang, T.R. Li, D. Liu, H.M. Chen, C. Luo, A unified framework of dynamic three-way probabilistic rough sets, Inform. Sci. 420 (2017) 126–147.

[10] C. Zhang, D.Y. Li, J.Y. Liang, Hesitant fuzzy linguistic rough set over two universes model and its applications, Int. J. Mach. Learn. Cybern. 9 (2018) 577–588.

[11] J.Y. Liang, F. Wang, C.Y. Dang, Y.H. Qian, A group incremental approach to feature selection applying rough set technique, IEEE Trans. Knowl. Data Eng. 26 (2014) 294–308.

[12] J.H. Liu, Y.J. Lin, Y.W. Li, W. Weng, S.X. Wu, Online multi-label streaming feature selection based on neighborhood rough set, Pattern Recognit. 84 (2018) 273–287.

[13] F. Min, J. Xu, Semi-greedy heuristics for feature selection with test cost constraints, Granul. Comput. 1 (2016) 199–211.

[14] R.W. Swiniarski, A. Skowron, Rough set mothods in feature selection and recognition, Pattern Recognit. Lett. 24 (2003) 833–849.

[15] C.Z. Wang, Q.H. Hu, X.Z. Wang, D.G. Chen, Y.H. Qian, Z. Dong, Feature selection based on neighborhood discrimination index, IEEE Tans. Neural Netw. Learn. Syst. 29 (2018) 2986–2999.

[16] Y.F. Chen, X.D. Yue, H. Fujita, S.Y. Fu, Three way decision support for diagnosis on focal liver lesions, Knowl.-Based Syst. 127 (2017) 85–99.

[17] J.H. Dai, W.T. Wang, Q. Xu, An uncertainy measure for incomplete decision tables and its applications, IEEE Trans. Cybern. 43 (2013) 1277–1289.

[18] J. Hu, T.R. Li, H.J. Wang, H. Fujita, Hierarchical cluster ensemble model based on knowledge granulation, Knowl.-Based Syst. 91 (2016) 179–188.

[19] D. Liu, D.C. Liang, C.C. Wang, A novel three-way decision model based on incomplete information system, Knowl.-Based Syst. 91 (2016) 32–45.

[20] H.M. Chen, T.R. Li, C. Luo, S.J. Horng, G.Y. Wang, A decision-theoretic rough set approach for dynamic data mining, IEEE Trans. Fuzzy Syst. 23 (2015) 1958–1970.

[21] D.G. Chen, Y.Y. Yang, Attribute reduction for heterogeneous data on the combination of classical and fuzzy rough set models, IEEE Trans. Fuzzy Syst. 22 (2014) 1325–1334.

[22] M. Dash, H. Liu, Consistency-based search in feature selection, Artificial Intelligence 151 (2003) 155–176.

[23] H.R. Ju, H.X. Li, X.B. Yang, X.Z. Zhou, B. Huang, Cost-sensitive rough set: a multi-granulation approach, Knowl.-Based Syst. 123 (2017) 137–153.

[24] N. Kwak, C.H. Choi, Input feature selection for classification problems, IEEE Trans. Neural Netw. 13 (2002) 143–159.

[25] G.M. Lang, M.J. Cai, H. Fujita, Q.M. Xiao, Related families-based attribute reduction of dynamic covering decision information systems, Knowl.-Based Syst. 162 (2018) 161–173.

[26] C. Lee, D.A. Landgrebe, Feature extraction based on decision boundaries, IEEE Trans. Pattern Anal. Mach. Intell. 15 (1993) 388–400.

[27] Y.H. Qian, J.Y. Liang, W. Wei, Consistency-preserving attibute reduction in fuzzy rough set framework, Int. J. Mach. Learn. Cybern. 4 (2013) 287–299.

[28] J.J. Song, E.C.C. Tsang, D.G. Chen, X.B. Yang, Minimal decision cost reduct in fuzzy decision-theoretic rough set model, Knowl.-Based Syst. 126 (2017) 104–112.

[29] C.Z. Wang, Q. He, M.W. Shao, Q.H. Hu, Feature selection based on maximal neighborhood discernibility, Int. J. Mach. Learn. Cybern. 9 (2018) 1929–1940.

[30] W. Wei, J.H. Wang, J.Y. Liang, X. Mi, C.Y. Dang, Compacted decision tables based attribute reduction, Knowl.-Based Syst. 86 (2015) 261–277.

[31] C.Z. Wang, Y. Huang, M.W. Shao, X.D. Fan, Fuzzy rough set-based attribute reduction using distance measures, Knowl.-Based Syst. 164 (2019) 205–212.

[32] C.Z. Wang, M.W. Shao, Q. He, Y.H. Qian, Y.L. Qi, Feature subset selection based on fuzzy neighborhood rough sets, Knowl.-Based Syst. 111 (2016) 173–179.

[33] Y.Y. Yao, X.Y. Zhang, Class-specific attribute reducts in rough set theory, Inf. Sci. 418–419 (2017) 601–618.

[34] Q.H. Hu, D.R. Yu, Z.X. Xie, Neighborhood classifiers, Expert Syst. Appl. 34 (2008) 866–876.

[35] Q.H. Hu, D.R. Yu, J.F. Liu, C.X. Wu, Neighborhood rough set based heterogeneous feature subset selection, Inform. Sci. 178 (2008) 3577–3594.

[36] X.B. Yang, S.C. Liang, H.L. Yu, S. Gao, Y.H. Qian, Pseudo-label neighborhood rough set: measures and attribute reductions, Internat. J. Approx. Reason. 105 (2019) 112–129.

[37] X.B. Yang, M. Zhang, H.L. Dou, J.Y. Yang, Neighborhood systems-based rough sets in incomplete information system, Knowl.-Based Syst. 24 (2011) 858–867.

[38] H. Fujita, A. Gaeta, V. Loia, F. Orciuoli, Improving awareness in early stages of security analysis: a zone partition method based on GrC, Appl. Intell. 49 (2019) 1063–1077.

[39] H. Fujita, A. Gaeta, V. Loia, F. Orciuoli, Resilience analysis of critical infrastructures: a cognitive approach based on granular computing, IEEE Trans. Cybern. 49 (2019) 1835–1848.

[40] A. Gacek, Granular modelling of signals: a framework of granular computing, Inform. Sci. 221 (2013) 1–11.

[41] J.H. Li, C.L. Mei, W.H. Xu, Y.H. Qian, Concept learning via granular computing: a cognitive viewpoint, Inform. Sci. 298 (2015) 447–467.

[42] Y.H. Qian, J.Y. Liang, W. Pedrycz, C.Y. Dang, An efficient accelerator for attribute reduction from incomplete data in rough set framework, Pattern Recognit. 44 (2011) 1658–1670.

[43] W.H. Xu, J.H. Yu, A novel aproach to information fusion in multi-source datasets: a granular computing viewpoint, Inform. Sci. 378 (2017) 410–423.

[44] W.H. Xu, W.T. Li, Granular computing approach to two-way learning based on formal concept analysis in fuzzy datasets, IEEE Trans. Cybern. 46 (2016) 366–379.

[45] X. Yang, T.R. Li, D. Liu, H. Fujita, A temporal-spatial composite sequential approach of three-way granular computing, Inform. Sci. 486 (2019) 171–189.

[46] X.B. Yang, S.P. Xu, H.L. Dou, X.N. Song, H.L. Yu, J.Y. Yang, Multigranulation rough set: a multiset based strategy, Int. J. Comput. Intell. Syst. 10 (2017) 277–292.

[47] X.B. Yang, Y. Qi, H.L. Yu, X.N. Song, J.Y. Yang, Updating multigranulation rough approximations with increasing of granular structures, Knowl.-Based Syst. 64 (2014) 59–69.

[48] Q.H. Hu, L.J. Zhang, Y.C. Zhou, W. Pedrycz, Large-scale multi-modality attribute reduction with multi-kernel fuzzy rough sets, IEEE Trans. Fuzzy Syst. 26 (2017) 226–238.

[49] Q.H. Hu, L. Zhang, D.G. Chen, W. Pedrycz, D.R. Yu, Gaussian kernel based fuzzy roughsets: model, uncertainty measures and applications, J. Approx. Reason. 51 (2010) 453–471.

[50] H.R. Ju, X.B. Yang, X.N. Song, Y.S. Qi, Dynamic updating multigranulation fuzzy rough set: approximations and reducts, Int. J. Mach. Learn. Cybern. 6 (2014) 981–990.

[51] S.P. Xu, P.X. Wang, J.H. Li, X.B. Yang, X.J. Chen, Attribute reduction: an ensemble strategy, in: Proc. of the 2017 International Joint Conference on Rough Sets, Olsztyn, 2017, pp. 362–375.

[52] Q.H. Hu, W. Pedrycz, D.R. Yu, J. Lang, Selecting discrete and continuous features based on neighborhood decision error minimization, IEEE Trans. Syst. Man Cybern. B 40 (2010) 137–150.

[53] Y.H. Chen, Y.Y. Yao, Multiview intelligent data analysis based on granular computing, in: Proc. of the 2006 IEEE International Conference on Granular Computing, Atlanta, 2006, pp. 281–286.

[54] J.H. Dai, S.C. Gao, G.J. Zheng, Generalized rough set models determined by multiple neighborhoods generated from a similarity relation, Soft Comput. 22 (2018) 2081–2094.

[55] J.Y. Liang, F. Wang, C.Y. Dang, Y.H. Qian, An efficient rough feature selection algorithm with a multi-granulation view, Internat. J. Approx. Reason. 53 (2012) 912–926.

[56] Y.J. Lin, J.J. Li, P.R. Lin, G.P. Lin, J.K. Chen, Feature selection via neighborhood multi-granulation fusion, Knowl.-Based Syst. 67 (2014) 162–168.

[57] 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 Recognit. 56 (2016) 1–15.

[58] J.Y. Liang, K.S. Chin, C.Y. Dang, C.M. Yam Richard, A new method for measuring uncertainty and fuzziness in rough set theory, Int. J. Gen. Syst. 31 (2002) 331–342.

[59] J.Y. Liang, Z.Z. Shi, The information entropy, rough entropy and knowledge granulation in rough set theory, Internat. J. Uncertain. Fuzziness Knowledge-Based Systems 12 (2004) 37–46.

[60] S.P. Xu, X.B. Yang, E.C.C. Tsang, A. Mantey Eric, Neighborhood collaborative classifiers, in: Proc. of the 2016 International Conference on Machine Learning and Cybernetics, Jeju, 2016, pp. 470–476.

[61] S.P. Xu, X.B. Yang, X.N. Song, H.L. Yu, Prediction of protein structural classed by decreasing nearest neighbor error rate, in: Proc. of the 2015 International Conference on Machine Learning and Cybernetics, Guangzhou, 2015, pp. 7–13.

[62] X.B. Yang, Y.Y. Yao, Ensemble selector for attribute reduction, Appl. Soft Comput. 70 (2018) 1–11.

[63] Y.Y. Yao, Y. Zhao, J. Wang, On reduct construction algorithms, Trans. Comput. Sci. II (2008) 100–117.

[64] J.H. Dai, Q. Xu, Attribute selection based on information gain ratio in fuzzy rough set theory with application to turnor classification, Appl. Soft Comput. 13 (2013) 211–221.

[65] W. Wei, J.Y. Liang, J.H. Wang, Y.H. Qian, Decision-relative discernibility matrices in the sense of entropies, Int. J. Gen. Syst. 42 (2013) 721–738.

[66] E.C.C. Tsang, Q.H. Hu, D.G. Chen, Feature and instance reduction for PNN classifiers based on fuzzy rough sets, Int. J. Mach. Learn. Cybern. 7 (2016) 1–11.

[67] J.Z. Li, X.B. Yang, X.N. Song, J.H. Li, P.X. Wang, D.J. Yu, Neighborhood attriute reduction: a multi-criterion approach, Int. J. Mach. Learn. Cybern. (2017) http://dx.doi.org/10.1007/s13042-017-0758-5.

[68] S.Y. Zhao, H. Chen, C.P. Li, X.Y. Du, H. Sun, A novel approach to building a robust fuzzy rough classifier, IEEE Trans. Fuzzy Syst. 23 (2015) 769–786.

[69] D.G. Chen, Y.Y. Yang, Z. Dong, An incremental algorithm for attribute reduction with variable precision rough sets, Appl. Soft Comput. 45 (2016) 129–149.

[70] J. Fan, Y.L. Jiang, Y. Liu, Quick attribute reduction with generalized indiscernibility models, Inform. Sci. 397–398 (2017) 15–36.

[71] F. Min, H.P. He, Y.H. Qian, W. Zhu, Test-cost-sensitive attribute reduction, Inform. Sci. 181 (2011) 4928–4942.

[72] X.B. Yang, Y.S. Qi, X.N. Song, J.Y. Yang, Test cost sensitive multigranulation rough set: model and minimal cost selection, Inform. Sci. 250 (2013) 184–199.

[73] Y.Y. Yao, Y. Zhao, Discernibility matrix simplification for constructing attribute reducts, Inform. Sci. 179 (2009) 867–882.

[74] F.J. Li, Y.H. Qian, J.T. Wang, C.Y. Dang, L.P. Jing, Clustering ensemble based on sample's stability, Artificial Intelligence 273 (2019) 37–55.

[75] C.C. Chang, C.J. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (2011) 1–27.