Triple-**G**: a new MGRS and attribute reduction

Jing Ba[1] · Keyu Liu[1,2] · Hengrong Ju[3] · Suping Xu[4,5] · Taihua Xu[1] · Xibei Yang[1,6]

Abstract

Different from classical rough set, Multigranulation Rough Set (MGRS) is frequently designed for approximating target through using multiple results of information granulation. Presently, though many forms of MGRS have been intensively explored, most of them are constructed based on the homogeneous information granulation with respect to different scales or levels. They lack the multi-view which involves the results of heterogeneous information granulation. To fill such a gap, a Triple-**G** MGRS is developed. Such a Triple-**G** is composed of three different heterogeneous information granulations: (1) nei**G**hborhood based information granulation; (2) **G**ap based information granulation; (3) **G**ranular ball based information granulation. Neighborhood provides a parameterized mechanism while gap and granular ball offer two representative data-adaptive strategies for performing information granulation. Immediately, both optimistic and pessimistic MGRS can be re-constructed. Furthermore, the problem of attribute reduction is also addressed based on the proposed models. Not only the forward greedy searching is used for deriving the Triple-**G** MGRS related reducts, but also an attribute grouping based accelerator is reported for further speeding up the process of searching reducts. The experimental results over 20 UCI data sets demonstrate the follows: (1) from the viewpoint of the generalization performance, the reducts obtained by our Triple-**G** MGRS is superior to those obtained by previous researches; (2) attribute grouping does speed up the process of searching reducts.

Keywords Attribute reduction · Attribute grouping · Multigranulation rough set · Triple-**G**

1 Introduction

Up to now, in the field of Granular Computing (GrC) [14, 20, 41, 42, 49, 54], multigranulation has witnessed great success in knowledge representation [28], distributive information processing [63], groups of intelligent agents [29] and so on.

Following the basic principle of multigranulation, various studies have been reported and then most of them can be divided into the following two phases.

1. On the one hand, the structure related topics are explored in view of the multiple different granulations [15, 47, 59]. For example, by examining the relationships among different information granulations, Yang et al. [55] have systematically revealed the hierarchical structures of the multigranulation space; Qian et al. [30] have proposed

✉ Xibei Yang
  jsjxy_yxb@just.edu.cn

  Jing Ba
  B_Bajing@126.com

  Keyu Liu
  keyu_liu@my.swjtu.edu.cn

  Hengrong Ju
  juhengrong@ntu.edu.cn

  Suping Xu
  supingxu@yahoo.com

  Taihua Xu
  xth19890410@163.com

1 School of Computer, Jiangsu University of Science and Technology, Zhenjiang 212100, Jiangsu, China

2 School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China

3 School of Information Science and Technology, Nantong University, Jiangsu 226000, China

4 School of Computer Science and Technology, Nanjing University, Jiangsu 210000, China

5 Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6R 2V4, Canada

6 Key Laboratory of Oceanographic Big Data Mining and Application of Zhejiang Province, Zhoushan 316022, Zhejiang, China

a serious of measures for characterizing the coarser or finer relationships over the multigranulation space, immediately, the order structure of the multigranulation space can be quantitatively represented; Lin et al. [18] have investigated the uncertainty measures in multigranulation space based on the hierarchical structure.

2. On the other hand, the modelling and data analysis related topics are also investigated in the framework of multigranulation. For instance, Lin et al. [21] have introduced the neighborhood based multiple information granulations into the rough modelling; Sun et al. [35] have studied the fuzzy based multiple information granulations over two universes for decision making; Li et al. [22] have introduced the Dempster-Shafer evidence theory into the multigranulation framework for improving the effectiveness of clustering ensemble; Qian et al. [34] also proposed the local modelling in cost-sensitive [17] environment for realizing quick feature selection.

Though reviewing various modelling strategies in terms of multigranulation, it is not difficult to point out that Multigranulation Rough Set (MGRS) [16, 61, 62] aims to approximate the target through using multiple different results of information granulation [37, 57]. Up to now, though a lot of MGRS have been proposed and carefully studied, most of the used information granulations only come from the homogeneous mechanism. For example, through appointing different values of radius, different neighborhood relations [10] can be obtained and these neighborhood relations are considered as the results of multiple information granulation for constructing MGRS; by using different values of threshold, the multigranulation decision-theoretic rough set [33] can be constructed, in which the thresholds are used to rule the granularity of the information granulation; through using a set of the coverings [19] for representing the results of information granulation, the covering based MGRS and the corresponding properties can also be studied. Nevertheless, it must be emphasized that most of the these researches lack the multi-view [56] which is rooted in the heterogeneous information granulation.

From discussions above, a new MGRS called Triple-**G** MGRS (covers three different "**G**", i.e., nei**G**hborhood, **G**ap and **G**ranular ball) will be developed in the context of this paper. Different from previous MGRS, our Triple-**G** MGRS employs the following three different information granulations: (1) a nei**G**hborhood [25, 46, 53] based parameterized information granulation; (2) a **G**ap [65] based data-adaptive information granulation; (3) a **G**ranular ball [50, 51] based data-adaptive information granulation. Obviously, since both parameterized information granulation and data-adaptive information granulations have been used, our Triple-**G** MGRS does reflect the fundamental principle of heterogeneous information granulation. Therefore, the main

structure of such Triple-**G** MGRS model can be elaborated in the following Fig. 1.

Moreover, it must be pointed out that in the field of rough set, the neighborhood, gap and granular ball are three widely accepted techniques for performing information granulation when facing data with continuous values or mixed data with both continuous and categorical values. Not only the computations of these information granulations are straightforward because only the distances among samples are required, but also the semantic explanations of these information granulations are clear. For example, the neighborhood based information granulation characterizes the similarity or dissimilarity [38] between any two samples through using only one justifiable parameter, it follows that different parameters offer different scales of such similarity or dissimilarity; the granular ball based information granulation characterizes the similarity or dissimilarity between samples through using the labels of samples, it follows that such mechanism can figure the distribution of data in some degree.

As the core of the rough set [8, 23, 36, 43, 44, 48], the problem of attribute reduction [9, 39, 45, 52, 60] can also be further explored in terms of Triple-**G** MGRS. However, since obtaining reduct based on Triple-**G** MGRS requires three different types of the information granulation, the elapsed time of searching reduct may be huge. For such a reason, an acceleration mechanism [12, 24, 58] called attribute grouping [3] will be further introduced into the process of searching reduct for reducing the computation overhead. Without loss of generality, the essence of such an accelerator is to partition the raw attributes into different groupings and then attributes in those groupings which contain at least one attribute in the potential reduct will not be evaluated in the process of deriving reduct. This is why the searching space of candidate attributes can be compressed and the computational efficiency will be improved.

The main contribution of our research can be summarized as the following aspects: (1) the proposed Triple-**G** is a novel development of MGRS, which can characterize the
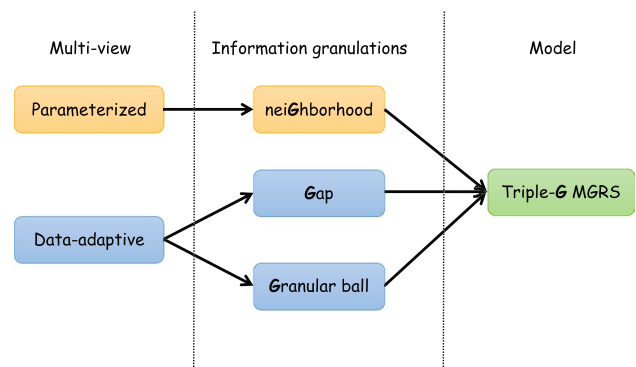

Fig. 1 The structure of Triple-**G** MGRS model

universe not only from the perspective of multigranulation, but also from the perspective of heterogeneity based multiview; (2) since three comprehensively different mechanisms of information granulation are employed, the proposed Triple-**G** model is more explanatory than other rough set models; (3) to further improve the efficiency of searching Triple-**G** related reduct, the attribute grouping [3] based acceleration mechanism is introduced into the procedure of searching reduct; (4) as a framework of accelerator, the attribute grouping is with strong expansibility and then different approaches for partitioning the attributes may generate different attribute grouping based accelerators.

The rest of this paper is organized as follows. In Sect. 2, basic notions related to neighborhood, gap, granular ball and multigranulation rough sets are briefly reviewed. Our proposed Triple-**G** MGRS and the corresponding attribute reduction will be addressed in Sect. 3. Comparative experimental results are reported in Sect. 4, as well as the corresponding analyses. This paper is ended in Sect. 5 by summarizing the novelties and scheduling some further perspectives.

2 Preliminaries

In this section, we will review some basic concepts such as neighborhood rough set, gap rough set, granular ball rough set, multigranulation rough set and the corresponding attribute reductions.

In researching rough set for the classification task, a decision system is denoted as $\mathcal{D} = \langle U, AT, d \rangle$, in which U is a finite non-empty set which consists of m samples, AT is the set of all condition attributes, d is the decision attribute and $d(x)$ indicates the label of sample x. In other words, the values over d are categorical and then such a decision attribute can be used to derive an equivalence relation such that IND $(d) = \{(x_i, x_j) \in U^2 : d(x_i) = d(x_j)\}$, it follows that the universe can be partitioned into serval disjoint decision classes such that $U/\text{IND}(d) = \{X_1, X_2, \cdots, X_n\}$.

2.1 Neighborhood rough set and attribute reduction

Definition 1 [10] Given a decision system $\mathcal{D}$, $\forall A \subseteq AT$, the neighborhood relation related to A over U can be defined as:

$$\mathcal{N}_A^\delta = \{(x_i, x_j) \in U \times U : \Delta_A(x_i, x_j) \le \delta\}, \tag{1}$$

in which $\Delta_A(x_i, x_j)$ is the distance between x_i and x_j by using the set of the attributes A, δ is the assigned radius which determines the scale of the neighborhood.

Furthermore, in the light of the information provided by neighborhood relation $\mathcal{N}_A^\delta$, $\forall x_i \in U$, the neighborhood of x_i related to A can be easily derived such that $\mathcal{N}_A^\delta(x_i) = \{x_j \in U : (x_i, x_j) \in \mathcal{N}_A^\delta\}$.

Definition 2 [10] Given a decision system $\mathcal{D}$, $\forall A \subseteq AT$ and $\forall X_p \in U/\text{IND}(d)$, the neighborhood lower and upper approximations of X_p related to A are defined as:

$$\underline{\text{N}}_A^\delta(X_p) = \{x_i \in U : \mathcal{N}_A^\delta(x_i) \subseteq X_p\}, \tag{2}$$

$$\overline{\text{N}}_A^\delta(X_p) = \{x_i \in U : \mathcal{N}_A^\delta(x_i) \cap X_p \ne \emptyset\}. \tag{3}$$

Following the basic notions of neighborhood and the corresponding rough set, various forms about attribute reduction have been proposed. Generally speaking, two representative trends of neighborhood rough set related measures are frequently used in defining attribute reductions. Therefore, a formal description of attribute reduction can be given as follows.

Definition 3 Given a decision system $\mathcal{D}$, $\forall A \subseteq AT$, ε is a given non-zero measure,

1. if the measure ε is expected to be as higher as possible, then A is referred to as a reduct if and only if the following two conditions hold:

   (a) $\dfrac{\varepsilon(A)}{\varepsilon(AT)} \ge \zeta$,
   (b) $\forall A' \subset A, \dfrac{\varepsilon(A')}{\varepsilon(AT)} < \zeta$;

2. if the measure ε is expected to be as lower as possible, then A is referred to as a reduct if and only if the following two conditions hold:

   (a) $\dfrac{\varepsilon(AT)}{\varepsilon(A)} \ge \zeta$,
   (b) $\forall A' \subset A, \dfrac{\varepsilon(AT)}{\varepsilon(A')} < \zeta$;

in which ζ is a given threshold, $\varepsilon(A)$ is the value of the measure ε determined by A.

Following Def. 3, it can be observed that the reduct A is actually the minimum subset of the condition attributes which will not bring great changes to the considered measure ε. Up to now, in view of the neighborhood rough set, various measures have been proposed, e.g., approximation quality [10], neighborhood conditional entropy [64], neighborhood conditional discrimination index [40], neighborhood decision error rate [11] and so on. For example, the approximation quality is used to evaluate the degree of certainty in data and then such measure is expected to be as higher as possible; the neighborhood decision error rate is

used to estimate the ratio of misclassifications in data and then such measure is expected to be as lower as possible.

2.2 Gap rough set and attribute reduction

Though the neighborhood relation has been widely concerned for its flexibility, the determination or selection of the radius is an open problem. In view of this, Zhou et al. [65] have proposed the gap rough set which can also derive the information granules of samples without assigning the radii. In their approach, the gap based information granulation can automatically select the neighbors for each sample by the surrounding instances distribution. For such a reason, the process of deriving gap information granules is data-adaptive. The detailed notions related to gap information granulation will be shown as follows.

Definition 4 [65] Given a decision system $\mathcal{D}, \forall A \subseteq AT$, $\forall x_i \in U$, the rest of samples are sorted by ascending order based on the distances between these samples and x_i, they are denoted as a set such that $Dis_A(x_i)$, the width of gap is then denoted as $Wid_A(x_i)$:

$$Dis_A(x_i) = \{x_i^1, x_i^2, \cdots, x_i^{m-1}\}, \tag{4}$$

$$Wid_A(x_i) = \frac{\sum_{j=1}^{m-1} \Delta_A(x_i^j, x_i)}{m-1}, \tag{5}$$

in which $\Delta_A(x_i, x_i^1) \leq \Delta_A(x_i, x_i^2) \leq \cdots \leq \Delta_A(x_i, x_i^{m-1})$.

Following Def. 4, if the distance between two samples x_i^j and x_i^{j+1} is bigger than $Wid_A(x_i)$, it is called a Gap between x_i^j and x_i^{j+1} and denoted as $\mathcal{G}_A(x_i^j, x_i^{j+1})$.

Definition 5 [65] Given a decision system $\mathcal{D}, \forall A \subseteq AT$, $\forall x_i \in U$, the gap information granule of x_i over A is defined as:

$$\mathcal{G}_A(x_i) = \{x_i^k \in Dis_A(x_i) : \mathcal{G}_A(x_i^{k-1}, x_i^k) < Wid_A(x_i), 2 \leq k \leq w\} \cup \{x_i\}, \tag{6}$$

in which $w = \min\{t : \mathcal{G}_A(x_i^{t-1}, x_i^t) \geq Wid_A(x_i), 2 \leq t \leq (m-1)\}$.

Definition 6 [65] Given a decision system $\mathcal{D}, \forall A \subseteq AT$ and $\forall X_p \in U/\mathrm{IND}(d)$, the gap lower and upper approximations of X_p related to A are defined as:

$$\underline{G_A}(X_p) = \{x_i \in U : \mathcal{G}_A(x_i) \subseteq X_p\}, \tag{7}$$

$$\overline{G_A}(X_p) = \{x_i \in U : \mathcal{G}_A(x_i) \cap X_p \neq \emptyset\}. \tag{8}$$

Following Def. 6, Zhou et al. [65] have designed a forward greedy searching for deriving reduct based on gap rough set. The specific steps are shown as follows.

(1) The potential reduct *red* is initialed to be ∅.
(2) By the considered measure ε, calculate the measure value $\varepsilon(AT)$ over AT.
(3) $\forall a \in AT - red$, compute $\varepsilon(red \cup \{a\})$.
(4) Select an appropriate attribute b from $AT - red$ and add it into *red*.
(5) If the constraint in attribute reduction is fulfilled, then return *red*; otherwise, go to (3).

2.3 Granular ball rough set and attribute reduction

Besides the gap rough set, Xia et al. [50, 51] have also developed a novel concept called granular ball for adaptively acquiring the information granules of samples. Furthermore, it should be emphasized that the sizes of the information granules in terms of different samples may be different. Such a case also characterizes the internal structure of the samples in some degree.

For Xia et al.'s [50, 51] work, the concept of granular ball mainly involves two phases: (1) the center point; (2) the radius.

Definition 7 [50, 51] Given a decision system $\mathcal{D}, \forall A \subseteq AT$, $\forall \mathcal{B}_s \subseteq U, \mathcal{B}_s$ is a granular ball induced by A in which c is the center point of $\mathcal{B}_s$, r is the mean distance among all samples in $\mathcal{B}_s$ and the central point c, that is:

$$c = \frac{1}{|\mathcal{B}_s|} \sum_{i=1}^{|\mathcal{B}_s|} x_i, \tag{9}$$

$$r = \frac{1}{|\mathcal{B}_s|} \sum_{i=1}^{|\mathcal{B}_s|} \Delta_A(x_i, c), \tag{10}$$

in which $\Delta_A(x_i, c)$ is the distance between x_i and the center point c over A.

Following the concept of granular ball, the procedure of generating granular balls can be realized by using the unsupervised learning. For instance, the universe can be gradually partitioned by using 2-meaning clustering [5] until the purities of the granular balls reach the expected values. The details are elaborated as follows.

(1) The universe U is regarded as a grouping.
(2) Each grouping is divided by 2-means clustering [5].
(3) The center point c and the mean distance r are calculated.
(4) Obtain granular balls based on the centers and radii of grouping, calculate the purity [50, 51] of each granular ball.

(5) If the purity of each granular ball achieves the expected value, then all the granular balls are generated; otherwise, go to step (2).

Based on the above process, the set of all obtained granular balls induced by A is denoted as $\mathcal{B}_A$ in the context of this paper.

Definition 8 [50, 51] Given a decision system $\mathcal{D}, \forall A \subseteq AT$ and $\forall X_p \in U/\text{IND}(d)$, the granular ball lower and upper approximations of X_p related to A are defined as:

$$\underline{B_A}(X_p) = \{x_i : \exists \mathcal{B}_s \in \mathcal{B}_A, x_i \in \mathcal{B}_s, \mathcal{B}_s \subseteq X_p\}, \tag{11}$$

$$\overline{B_A}(X_p) = \{x_i : \exists \mathcal{B}_s \in \mathcal{B}_A, x_i \in \mathcal{B}_s, \mathcal{B}_s \cap X_p \neq \emptyset\}. \tag{12}$$

Following the granular ball rough set shown in Def. 8, an immediate problem is to calculate the corresponding reduct. Therefore, Xia et al. [50, 51] have designed a backward greedy searching for deriving granular ball rough set related reduct. The specific steps are shown as follows.

(1) The potential reduct *red* is initialized to be AT.
(2) By the considered measure ε, calculate the measure value $\varepsilon(AT)$ over AT.
(3) $\forall a \in AT - red$, compute $\varepsilon(red - \{a\})$.
(4) Select an appropriate attribute b from *red* and remove it from *red*.
(5) If the constraint in attribute reduction is fulfilled, then return *red*; otherwise, go to (3).

2.4 Multigranulation rough set and attribute reduction

From the perspective of GrC, the neighborhood rough set, the gap rough set and the granular ball rough set can be regarded as the similar model which is constructed based on one and only one strategy of information granulation. Nevertheless, in many real-world applications, the single information granulation is insufficient to characterize the multi-level or multi-view over data. For such reasons, Qian et al. [32] have developed a framework called MGRS which

aims to further expand the applications of rough set. Different from conventional researches, MGRS is constructed based on multiple different information granulations. The following definitions show us two representative models of MGRS.

Definition 9 [32] Given a decision system $\mathcal{D}, \forall A \subseteq AT$, $\forall X_p \in U/\text{IND}(d)$, the optimistic neighborhood multigranulation lower and upper approximations of X_p related to A are defined as:

$$\underline{\text{MG}^{Opt}_{\sum_{k=1}^{|A|} a_k}}(X_p) = \{x_i \in U : \mathcal{N}^\delta_{a_1}(x_i) \subseteq X_p \vee \cdots \vee \mathcal{N}^\delta_{a_{|A|}}(x_i) \subseteq X_p\}, \tag{13}$$

$$\overline{\text{MG}^{Opt}_{\sum_{k=1}^{|A|} a_k}}(X_p) = \sim \underline{\text{MG}^{Opt}_{\sum_{k=1}^{|A|} a_k}}(\sim X_p), \tag{14}$$

in which $\sim X_p$ is the complement of X_p.

Definition 10 [32] Given a decision system $\mathcal{D}, \forall A \subseteq AT$, $\forall X_p \in U/\text{IND}(d)$, the pessimistic neighborhood multigranulation lower and upper approximations of X_p related to A are defined as:

$$\underline{\text{MG}^{Pes}_{\sum_{k=1}^{|A|} a_k}}(X_p) = \{x_i \in U : \mathcal{N}^\delta_{a_1}(x_i) \subseteq X_p \wedge \cdots \wedge \mathcal{N}^\delta_{a_{|A|}}(x_i) \subseteq X_p\}, \tag{15}$$

$$\overline{\text{MG}^{Pes}_{\sum_{k=1}^{|A|} a_k}}(X_p) = \sim \underline{\text{MG}^{Pes}_{\sum_{k=1}^{|A|} a_k}}(\sim X_p). \tag{16}$$

Following the definitions of optimistic and pessimistic MGRS, some crucial measures related to rough set can also be generalized, e.g., approximation quality, conditional entropy, etc. Following these measures, the problem of attribute reduction then can be addressed over such two MGRS. The following Algorithm 1 shows us the detailed process of searching MGRS related reduct.

Algorithm 1: Searching Optimistic/Pessimistic MGRS Related Reducts

Input: Decision system $\mathcal{D} = \langle U, AT, d \rangle$, threshold ζ;
Output: One reduct red.

1   Compute the measure $\varepsilon(AT)$ related to MGRS over AT;
2   $red = \emptyset$;
3   **While** $\dfrac{\varepsilon(red)}{\varepsilon(AT)} < \zeta$ **do**

     `// Measure ε is expected to be as higher as possible`

4     $\forall a \in AT - red, \forall X_p \in U/IND(d)$, compute $\mathbf{MG}^{Opt}_{\sum_{k=1}^{|red \cup \{a\}|} a_k}(X_p)$

     or $\mathbf{MG}^{Pes}_{\sum_{k=1}^{|red \cup \{a\}|} a_k}(X_p)$;

5     $\forall a \in AT - red$, by the above approximations, calculate the measure value $\varepsilon(red \cup \{a\})$;
6   Select an appropriate attribute b from $AT - red$ based on the obtained measure values;
7     $red = red \cup \{b\}$;
8   Compute $\varepsilon(red)$;
9   **End**
10 **Repeat**

     `// Remove the redundant attributes`

11    $\forall c \in red, \varepsilon(red - \{c\})$;
12   **If** $\dfrac{\varepsilon(red - \{c\})}{\varepsilon(AT)} \geq \zeta$
13      $red = red - \{c\}$;
14   **End**
15 **Until** red does not change or $|red| = 1$;
16 **Return** red.

The above algorithm shows us the procedure of deriving reduct if the measure ε is expected to be as higher as possible. Similarly, if the measure is expected to be as lower as possible, then the terminal condition in Algorithm 1 should be replaced by $\dfrac{\varepsilon(AT)}{\varepsilon(red)} < \zeta$.

In Algorithm 1, it is not difficult to observe that each candidate attribute in $AT - red$ is required to be evaluated. In the worst case, all the conditional attributes must be added into the reduct. Therefore, the number of all iterations is $(|AT| + (|AT| - 1) + (|AT| - 2) + \cdots + 1)$. In addition, in the process of calculating the value of measure, the samples in the whole universe U should be scanned and any two samples in U should be compared. In summary, the number of iterations required in Algorithm 1 is $|U|^2 \times (|AT| + (|AT| - 1) + (|AT| - 2) + \cdots + 1)$, i.e., the time complexity of Algorithm 1 is $\mathcal{O}(|U|^2 \times |AT|^2)$.

3 Attribute reduction based on Triple-G MGRS

3.1 Triple-G MGRS

In Sect. 2.4, it is not difficult to observe that such two MGRS models take one and only one form of the information granulation into account, which can characterize the rough set related uncertainties. This is mainly because the process of deriving neighborhood relation is unique though different attributes have been employed. For such reason, we can observe that the MGRS shown in Defs. 9 and 10 do not comprehensively grasp the fundamental principle of multi-view which is the core in the theory of GrC.

Fortunately, following what have been addressed in Sect. 2, besides the neighborhood based information granulation which heavily depends on the value of the radius, two state-of-the-art data-adaptive information granulations have been presented. Therefore, to further characterize the uncertainty from multi-view in the framework of MGRS, both

the nei**G**hborhood rough set, **G**ap rough set and **G**ranular ball rough set will be simultaneously used for constructing MGRS, such type of the rough set will be referred to as the Triple-**G** MGRS. The following definition shows us two representative models of Triple-**G** MGRS.

Definition 11 Given a decision system $\mathcal{D}, \forall A \subseteq AT$, $\forall X_p \in U/\text{IND}(d)$, the Triple-**G** optimistic lower and upper approximations of X_p related to A are defined as:

$$\underline{\text{TG}}_A^{Opt}(X_p) = \{x_i \in U : \mathcal{N}_A^\delta(x_i) \subseteq X_p \vee \mathcal{G}_A(x_i) \subseteq X_p \vee \\ \mathcal{B}_s \subseteq X_p \text{ where } \mathcal{B}_s \in \mathcal{B}_A \text{ and } x_i \in \mathcal{B}_s\}, \quad (17)$$

$$\overline{\text{TG}}_A^{Opt}(X_p) = \sim \underline{\text{TG}}_A^{Opt}(\sim X_p). \quad (18)$$

Theorem 1 *Given a decision system* $\mathcal{D}, \forall A \subseteq AT$, $\forall X_p \in U/IND(d)$, *we have*

$$\overline{\text{TG}}_A^{Opt}(X_p) = \{x_i \in U : \mathcal{N}_A^\delta(x_i) \cap X_p \neq \emptyset \wedge \mathcal{G}_A(x_i) \cap X_p \neq \emptyset \\ \wedge \mathcal{B}_s \cap X_p \neq \emptyset \text{ where } \mathcal{B}_s \in \mathcal{B}_A \text{ and } x_i \in \mathcal{B}_s\}. \quad (19)$$

Proof By Def. 11, we have

$$x \in \overline{\text{TG}}_A^{Opt}(X_p) \Leftrightarrow x \notin \underline{\text{TG}}_A^{Opt}(\sim X_p)$$
$$\Leftrightarrow \mathcal{N}_A^\delta(x_i) \nsubseteq (\sim X_p) \wedge \mathcal{G}_A(x_i) \nsubseteq (\sim X_p) \wedge \mathcal{B}_s \nsubseteq (\sim X_p)$$
$$\quad \text{where } \mathcal{B}_s \in \mathcal{B}_A \text{ and } x_i \in \mathcal{B}_s$$
$$\Leftrightarrow \mathcal{N}_A^\delta(x_i) \cap X_p \neq \emptyset \wedge \mathcal{G}_A(x_i) \cap X_p \neq \emptyset \wedge \mathcal{B}_s \cap X_p \neq \emptyset$$
$$\quad \text{where } \mathcal{B}_s \in \mathcal{B}_A \text{ and } x_i \in \mathcal{B}_s.$$

$\square$

By Theorem 1, we can observe that though the Triple-**G** optimistic upper approximation is defined by the complement of the Triple-**G** optimistic lower approximation, it can also be considered as a set, in which samples have non-empty intersection with the target in terms of each used information granulation.

Definition 12 Given a decision system $\mathcal{D}, \forall A \subseteq AT$, $\forall X_p \in U/\text{IND}(d)$, the Triple-**G** pessimistic lower and upper approximations of X_p related to A are defined as:

$$\underline{\text{TG}}_A^{Pes}(X_p) = \{x_i \in U : \mathcal{N}_A^\delta(x_i) \subseteq X_p \wedge \mathcal{G}_A(x_i) \subseteq X_p \\ \wedge \mathcal{B}_s \subseteq X_p \text{ where } \mathcal{B}_s \in \mathcal{B}_A \text{ and } x_i \in \mathcal{B}_s\}, \quad (20)$$

$$\overline{\text{TG}}_A^{Pes}(X_p) = \sim \underline{\text{TG}}_A^{Pes}(\sim X_p). \quad (21)$$

Theorem 2 *Given a decision system* $\mathcal{D}, \forall A \subseteq AT$, $\forall X_p \in U/IND(d)$, *we have*

$$\overline{\text{TG}}_A^{Pes}(X_p) = \{x_i \in U : \mathcal{N}_A^\delta(x_i) \cap X_p \neq \emptyset \vee \mathcal{G}_A(x_i) \cap X_p \neq \emptyset \\ \vee \mathcal{B}_s \cap X_p \neq \emptyset \text{ where } \mathcal{B}_s \in \mathcal{B}_A \text{ and } x_i \in \mathcal{B}_s\}. \quad (22)$$

Proof The proof of Theorem 2 is similar to that of Theorem 1. $\square$

Similar to Defs. 9 and 10, the structures of MGRS shown in Defs. 11 and 12 are preserved, i.e., one sample belongs to the Triple-**G** optimistic lower approximation if and only if at least one type of the information granule related to such sample is contained in the target, one sample belongs to the Triple-**G** pessimistic lower approximation if and only if all the information granules related to such sample are contained in the target. Nevertheless, it must be pointed out that different from previous MGRS, our new Triple-**G** MGRS uses multiple different mechanisms for performing information granulation and then derive the information granules for constructing rough sets.

Similar to Algorithm 1, we can also design the following Algorithm 2 for searching reduct based on Triple-**G** MGRS.

Algorithm 2: Searching Optimistic/Pessimistic Triple-**G** MGRS Related Reducts

Input: Decision system $\mathcal{D} = \langle U, AT, d \rangle$, threshold ζ;
Output: One reduct red.

1 Compute the measure $\varepsilon(AT)$ related to Triple-**G** MGRS over AT;
2 $red = \emptyset$;
3 **While** $\dfrac{\varepsilon(red)}{\varepsilon(AT)} < \zeta$ **do**
       // Measure ε is expected to be as high as possible
4      $\forall a \in AT - red, \forall X_p \in U/IND(d)$, compute $\underline{\textbf{TG}}_A^{Opt}(X_p)$ or $\underline{\textbf{TG}}_A^{Pes}(X_p)$;
5      $\forall a \in AT - red$, by the above approximations, calculate the measure value $\varepsilon(red \cup \{a\})$;
6   Select an appropriate attribute b from $AT - red$ based on the obtained measure values;
7      $red = red \cup \{b\}$;
8   Compute $\varepsilon(red)$;
9 **End**
10 **Repeat**
       // Remove the redundant attributes
11     $\forall c \in red, \varepsilon(red - \{c\})$;
12   **If** $\dfrac{\varepsilon(red - \{c\})}{\varepsilon(AT)} \geq \zeta$
13         $red = red - \{c\}$;
14   **End**
15 **Until** red does not change or $|red| = 1$;
16 **Return** red.

Similar to Algorithm 1, if the measure is expected to be as lower as possible, then the terminal condition in Algorithm 1 should be replaced by $\dfrac{\varepsilon(AT)}{\varepsilon(red)} < \zeta$.

By observing Algorithm 2, in the worst case, all the conditional attributes must be added into the reduct. Therefore, the number of iterations is $(|AT| + (|AT| - 1) + (|AT| - 2) + \cdots + 1)$. In addition, the lower approximations should be obtained by the neighborhood, gap and granular ball based information granulations simultaneously, it follows that in the calculation of the value of measure, the required number of iterations is $(|U|^2 + |U| \times |\mathcal{B}_{AT}| + |U|^2)$. In summary, the number of iterations in Algorithm 2 is $(|U|^2 + |U| \times |\mathcal{B}_{AT}| + |U|^2) \times (|AT| + (|AT| - 1) + (|AT| - 2) + \cdots + 1)$, i.e., the time complex of Algorithm 2 is $\mathcal{O}(|U|^2 \times |AT|^2)$.

3.2 Attribute reduction based on Triple-G MGRS with attribute grouping

As what can be observed in Sect. 3.2, the number of iterations used in Algorithm 2 is greater than that is required in Algorithm 1, this is mainly because both the strategies of gap and granular ball are further introduced into the process of information granulation. From this point of view, Algorithm 2 may be very time-consuming and then the quick searching has become a necessary.

Presently, in the field of rough set, various acceleration strategies [7, 13] have been proposed for speeding up the procedure of searching reduct. Most of them can be roughly grouped into two phases: (1) accelerator based on the perspective of sample; (2) accelerator based on the perspective of attribute. For example, the bucket [26] and positive approximation [31] are two typical representations of the sample based accelerators, the attribute grouping [3] is the typical representation of the attribute based accelerator. Through comparing these state-of-the-art accelerators, it can be observed that Chen et al.'s [3] attribute grouping based accelerator is superior to the sample based accelerators. This

is mainly because most of the sample based accelerators are vulnerable to the distribution of the samples. For instance, the bucket will be useless if the samples are extremely concentrated.

From discussions above, for further speeding up the searching of Triple-**G** MGRS related reduct, the attribute grouping will be introduced into Algorithm 2. Fundamentally, the essence of attribute grouping is to partition the raw set of the condition attributes into different groupings, it follows that attributes in those groupings which contain at least one attribute in the potential reduct will not be evaluated. Such a mechanism will significantly reduce the times of evaluating candidate attributes and then the elapsed time of calculating reduct can be saved. Immediately, the following Algorithm 3 will be designed.

Algorithm 3: Searching Optimistic/Pessimistic Triple-**G** MGRS Related Reducts with Attribute Grouping

---

Input: Decision system $\mathcal{D} = \langle U, AT, d \rangle$, threshold ζ;
Output: One reduct red.

1 Divide AT into different attribute groupings by using k-means clustering;

2 Initialize $red = \emptyset, T = AT, T' = \emptyset$;
    `// T is the set of candidate attributes, T' is used to record the`
    `   potential reduct in one iteration.`

3 Compute $\mathbf{TG}_A^{Opt}(X_p)$ or $\mathbf{TG}_A^{Pes}(X_p)$;

4 Compute $\varepsilon(AT)$;

5 **While** $\dfrac{\varepsilon(red)}{\varepsilon(AT)} < \zeta$ **do**
    `// Measure ε is expected to be as higher as possible`

6    $T = AT - red$;

7    $\forall a \in T$, if a and an attribute in T' belong to the same partition, then $T = AT - red - \{a\}$;

8   **If** $T = \emptyset$

9 | go to **14**;

10   **End**

11   **Else**

12 | go to **15**;

13   **End**

14   $T = AT - red, T' = \emptyset$;

15   Select an appropriate attribute b from $AT - red$ by the obtained measure values;

16   $red = red \cup \{b\}, T' = T' \cup \{b\}$;

17 **End**

18 **Repeat**
    `// Remove the redundant attributes`

19   $\forall c \in red, \varepsilon(red - \{c\})$;

20   **If** $\dfrac{\varepsilon(red - \{c\})}{\varepsilon(AT)} \geq \zeta$

21 | $red = red - \{c\}$;

22   **End**

23 **Until** red *does not change or* $|red| = 1$;

24 **Return** red.

Table 1 Data sets description

ID	Data sets	#Samples	#Attributes	#Labels	Domains
1	Banknote authentication	1372	4	2	Economics
2	Cardiotocography	2126	21	10	Medicine
3	Congressional voting	435	16	2	Politics
4	Contraceptive method	1473	9	3	Medicine
5	Dermatology	366	34	6	Medicine
6	Diabetes	372	8	2	Medicine
7	Ecoli	336	7	8	Medicine
8	Forest type mapping	523	27	4	Geography
9	Iris	150	4	3	Medicine
10	Leaf	340	15	36	Biology
11	Lymphography	98	18	4	Physics
12	Molecular biology1	106	57	2	Biology
13	Parkinson multiple sound recording	1208	26	2	Medicine
14	Seeds	210	7	3	Biology
15	Statlog (Heart)	270	13	2	Biology
16	Statlog (Image Segmentation)	2310	18	7	Computer
17	Ultrasonic flowmeter diagnostics-meter D	180	43	4	Medicine
18	Vertebral column	310	7	2	Biology
19	Waveform database generator(Version 1)	5000	21	3	Computer
20	Yeast	1482	8	10	Biology

Table 2 Maximal classification accuracies based on KNN classifier (higher accuracies are indicated in bold)

ID	Algorithm 1(O)	Algorithm 2(O)	Algorithm 3(O)	Algorithm 1(P)	Algorithm 2(P)	Algorithm 3(P)	Gap	GB
1	**1.00(0.22)**	**1.00(0.02)**	**1.00(0.02)**	0.82(0.02)	**1.00(0.02)**	**1.00(0.12)**	**1.00**	0.75
2	0.75(0.02)	0.80(0.22)	**0.80(0.24)**	0.26(0.04)	0.76(0.02)	**0.78(0.22)**	0.76	0.61
3	0.68(0.02)	**0.95(0.02)**	0.94(0.04)	0.66(0.02)	0.91(0.02)	**0.95(0.08)**	0.94	0.88
4	0.43(0.04)	**0.48(0.10)**	0.47(0.20)	0.43(0.02)	0.45(0.02)	**0.49(0.04)**	**0.50**	0.43
5	0.83(0.02)	0.94(0.36)	**0.94(0.40)**	0.45(0.02)	**0.97(0.02)**	0.79(0.40)	0.93	0.70
6	**0.71(0.04)**	0.69(0.30)	0.70(0.34)	0.62(0.08)	0.65(0.02)	**0.68(0.18)**	0.65	0.66
7	0.83(0.02)	0.83(0.10)	**0.84(0.18)**	0.63(0.02)	0.84(0.02)	**0.85(0.30)**	0.81	0.73
8	0.86(0.02)	**0.89(0.18)**	0.88(0.22)	0.53(0.02)	0.88(0.02)	**0.88(0.22)**	0.82	0.70
9	0.96(0.02)	**0.97(0.06)**	**0.97(0.06)**	0.96(0.02)	0.95(0.02)	**0.96(0.02)**	0.95	0.95
10	0.51(0.02)	**0.58(0.40)**	0.31(0.30)	0.18(0.02)	0.13(0.02)	**0.48(0.38)**	0.31	0.11
11	0.62(0.02)	**0.81(0.30)**	0.77(0.12)	0.47(0.16)	**0.82(0.02)**	0.79(0.40)	0.71	0.66
12	0.74(0.02)	0.75(0.34)	**0.76(0.26)**	0.66(0.02)	0.72(0.02)	**0.78(0.10)**	0.78	0.51
13	0.68(0.02)	**0.70(0.14)**	0.70(0.22)	0.55(0.04)	0.68(0.02)	**0.70(0.32)**	0.66	0.61
14	0.91(0.24)	0.92(0.14)	**0.93(0.26)**	0.86(0.02)	**0.92(0.02)**	0.91(0.26)	0.88	0.68
15	0.67(0.08)	0.81(0.14)	**0.81(0.18)**	0.66(0.14)	**0.81(0.02)**	0.80(0.22)	0.73	0.79
16	0.89(0.02)	**0.97(0.04)**	0.96(0.12)	0.62(0.02)	**0.96(0.02)**	0.96(0.22)	**0.97**	0.91
17	0.81(0.22)	**0.88(0.04)**	**0.88(0.08)**	0.67(0.32)	**0.86(0.02)**	0.69(0.32)	**0.81**	0.66
18	**0.84(0.10)**	0.82(0.40)	0.82(0.16)	0.73(0.02)	0.77(0.02)	**0.83(0.18)**	0.80	0.82
19	0.80(0.02)	**0.81(0.40)**	0.80(0.34)	0.47(0.02)	**0.81(0.02)**	0.80(0.38)	0.76	0.73
20	0.52(0.02)	**0.54(0.22)**	0.54(0.24)	0.35(0.18)	**0.54(0.02)**	**0.54(0.06)**	0.52	0.45

In Algorithm 3, if all condition attributes are divided into $q(1 < q < |AT|)$ groupings, then the numbers of attributes in those q groupings are $|AT_1|, |AT_2|, \cdots, |AT_q|$, respectively, i.e., $|AT_1| + |AT_2| + \cdots + |AT_q| = |AT|$. Therefore, in the "While" loop, the number of iterations is $(|AT| + (|AT| - |AT_1|) + (|AT| - |AT_1| - |AT_2|) + \cdots + 1)$. Moreover, both Algorithm 2 and Algorithm 3 require the same number of iterations in the step of calculating the value of measure, i.e., $(|U|^2 + |U| \times |\mathcal{B}_{AT}| + |U|^2)$. Finally, the number of iterations in Algorithm 3 is $(|U|^2 + |U| \times |\mathcal{B}_{AT}| + |U|^2) \times (|AT| + (|AT| - |AT_1|) + (|AT| - |AT_1| - |AT_2|) + \cdots + 1)$, i.e., the time complexity of Algorithm 3 is $\mathcal{O}(|U|^2 \times |AT|^2)$. Though the time complexity of Algorithm 3 is the same as that of Algorithm 2, it must be pointed out that such a case can only be observed if each grouping contains only one attribute. In most of the cases, we can set $|AT_q| \geq 1$, then $(|AT| + (|AT| - 1) + (|AT| - 2) + \cdots + 1) \geq (|AT| + (|AT| - |AT_1|) + (|AT| - |AT_1| - |AT_2|) + \cdots + 1)$ holds, it follows that the number of iterations used in Algorithm 3 is less than or equal to that used in Algorithm 2.

4 Experiments

4.1 Data sets

To demonstrate the effectiveness of our proposed algorithms, experimental comparisons will be conducted in the context of this section. All experiments will be carried out on a personal computer with Windows 10, Intel Core(TM) i5-4210U (2.60 GHz) and 8.00 GB memory. The programming language is Matlab R2018a. Moreover,

20 UCI data sets have been selected to conduct the experiments. The details of these data sets are shown in the following Table 1.

Besides Algorithms 1–3 which have been explored in this paper, the algorithms related to gap and granular ball rough sets for calculating reducts will also be tested in this section. Since Algorithms 1–3 should be performed by inputting the radius and then in our experiments, 20 different radii such as 0.02, 0.04, $\cdots$, 0.40 are selected with the step size of 0.02.

Notably, in our experiments, the K-means clustering is adopted to generate attribute groups. The value of K is set to be $K = \lceil |AT|/3 \rceil$ based on numerous previous testings.

4.2 Comparisons of classification accuracy

In this subsection, 10-fold cross-validation [6, 27] is used to calculate the reducts by using different algorithms. That is, we divide each data set into 10 parts with the same size, they are denoted by $U_1, U_2, \cdots, U_{10}$. For the first round of computation, $U_1 \cup \cdots \cup U_9$ is considered as the set of training

samples for deriving reduct and then U_{10} is regarded as the set of testing samples for obtaining the classification accuracy related to such obtained reduct; $\cdots$; for the last round of computation, $U_2 \cup \cdots \cup U_{10}$ is considered as the set of training samples for deriving reduct and then U_1 is regarded as the set of testing samples for obtaining the classification accuracy related to such obtained reduct. Immediately, the average classification accuracies related to different reducts can be obtained.

It must be pointed out that since 20 different radii have been selected for calculating reducts through using Algorithms 1–3, the following Tables 2 and 3 show us the maximal values of classification accuracies which are selected from the results related to such 20 radii. Tab. 2 is based on the KNN ($k = 3$) [2] classifier while Table 3 is based on the SVM (libSVM) classifier [1]. And in our experiments, $\zeta = 0.95$.

In Tables 2 and 3,

1. "Algorithm 1(O)" indicates the result of attribute reduction based on optimistic classical MGRS through using Algorithm 1;
2. "Algorithm 1(P)" indicates the result of attribute reduction based on pessimistic classical MGRS through using Algorithm 1;
3. "Algorithm 2(O)" indicates the result of attribute reduction based on optimistic Triple-**G** MGRS through using Algorithm 2;
4. "Algorithm 2(P)" indicates the result of attribute reduction based on pessimistic Triple-**G** MGRS through using Algorithm 2;
5. "Algorithm 3(O)" indicates the result of attribute reduction based on optimistic Triple-**G** MGRS through using Algorithm 3;
6. "Algorithm 3(P)" indicates the result of attribute reduction based on pessimistic Triple-**G** MGRS through using Algorithm 3;
7. "Gap" indicates the result of attribute reduction based on gap rough set;
8. "GB" indicates the result of attribute reduction based on granular ball rough set.

Therefore, it is not difficult to observe the following.

1. No matter optimistic or pessimistic case is considered, compared with classical MGRS, gap and granular ball rough sets, the reduct obtained by using our Triple-**G** MGRS can provide higher classification accuracies. Take the "Parkinson Multiple Sound Recording" data set as an example, in Table 2, the classification accuracy derived from the reduct related to optimistic Triple-**G** MGRS is 0.7003, it is superior to the classification accuracy derived from the reduct related to classical

Table 3 Maximal classification accuracies based on SVM classifier (higher accuracies are indicated in bold)

ID	Algorithm 1(O)	Algorithm 2(O)	Algorithm 3(O)	Algorithm 1(P)	Algorithm 2(P)	Algorithm 3(P)	Gap	GB
1	0.90(0.08)	**0.98(0.02)**	**0.98(0.02)**	0.85(0.02)	**0.98(0.02)**	0.98(0.12)	0.98	0.69
2	0.65(0.02)	0.75(0.20)	**0.77(0.20)**	0.33(0.22)	**0.78(0.02)**	0.77(0.40)	0.68	0.50
3	0.71(0.02)	0.94(0.02)	**0.95(0.24)**	0.71(0.02)	0.94(0.02)	**0.95(0.08)**	0.94	0.86
4	0.43(0.02)	0.51(0.22)	**0.51(0.06)**	0.43(0.02)	0.51(0.02)	**0.51(0.14)**	0.50	0.47
5	0.80(0.02)	0.92(0.26)	**0.93(0.20)**	0.49(0.34)	**0.97(0.02)**	0.76(0.38)	0.91	0.71
6	0.73(0.04)	**0.75(0.30)**	0.74(0.10)	0.68(0.02)	0.74(0.02)	**0.74(0.30)**	0.70	0.62
7	0.79(0.02)	0.81(0.08)	**0.82(0.08)**	0.64(0.02)	0.80(0.02)	**0.82(0.18)**	0.79	0.73
8	0.87(0.02)	**0.89(0.16)**	**0.89(0.16)**	0.49(0.02)	**0.89(0.02)**	0.88(0.26)	0.82	0.67
9	0.97(0.02)	**0.97(0.04)**	**0.97(0.04)**	0.97(0.02)	0.97(0.02)	**0.97(0.12)**	0.97	0.97
10	0.42(0.02)	**0.43(0.24)**	**0.43(0.32)**	0.20(0.02)	**0.43(0.02)**	0.39(0.40)	0.24	0.11
11	0.70(0.02)	**0.78(0.34)**	0.77(0.40)	0.65(0.02)	**0.83(0.02)**	0.77(0.40)	0.67	0.70
12	0.46(0.02)	0.71(0.34)	**0.72(0.24)**	0.46(0.02)	**0.73(0.02)**	0.66(0.20)	0.62	0.50
13	0.65(0.02)	**0.66(0.14)**	0.66(0.18)	0.57(0.26)	0.66(0.02)	**0.66(0.22)**	0.64	0.59
14	0.93(0.04)	**0.94(0.14)**	0.92(0.20)	0.86(0.02)	0.92(0.02)	**0.93(0.24)**	0.89	0.67
15	0.72(0.02)	0.83(0.16)	**0.83(0.16)**	0.67(0.14)	0.83(0.02)	**0.83(0.26)**	0.79	0.81
16	0.73(0.02)	0.82(0.02)	**0.85(0.18)**	0.44(0.10)	**0.93(0.02)**	0.88(0.22)	0.82	0.78
17	0.62(0.20)	**0.65(0.08)**	0.63(0.10)	0.59(0.02)	**0.68(0.02)**	0.61(0.06)	0.57	0.49
18	0.76(0.26)	**0.77(0.40)**	0.76(0.12)	0.68(0.02)	0.76(0.02)	**0.76(0.32)**	0.73	0.70
19	0.85(0.02)	0.87(0.40)	**0.87(0.38)**	0.51(0.02)	**0.87(0.02)**	0.86(0.40)	0.80	0.79
20	0.52(0.02)	**0.54(0.18)**	**0.54(0.18)**	0.39(0.18)	**0.54(0.02)**	0.54(0.20)	0.53	0.47

optimistic MGRS, i.e., 0.6847, it is also superior to the classification accuracies derived from the reducts related to both gap and granular ball rough sets, i.e., 0.6630 and 0.6084. Furthermore, take the "Forest Type Mapping" data set as an example, in Table 3, the classification accuracy derived from the reduct related to pessimistic Triple-**G** MGRS is 0.8929, it is superior to the classification accuracy derived from the reduct related to classical pessimistic MGRS, i.e., 0.4856, it is also superior to the classification accuracies derived from the reducts related to both gap and granular ball rough sets, i.e., 0.8164 and 0.6693.

2. By comparing with Algorithm 2, though Algorithm 3 is designed based on the consideration of accelerating of the process of calculating Triple-**G** based MGRS reduct, such improved algorithm can also offer well-matched classification performances. Take the "Cardiotocography" data set and the optimistic model as an example, in Table 2, the classification accuracy derived from the reduct related to Algorithm 3 is 0.8045 while the classification accuracy derived from the reduct related to Algorithm 2 is 0.8024. Furthermore, Take the "Statlog (Heart)" data set and the pessimistic model as an example, in Table 3, the classification accuracy derived from

the reduct related to Algorithm 3 is 0.8333 while the classification accuracy derived from the reduct related to Algorithm 2 is 0.8307.

3. From discussions above, we can conclude that the reducts derived by using our optimistic(pessimistic) Triple-**G** MGRS are better than the reducts derived by using optimistic(pessimistic) classical MGRS, gap and granular ball rough sets because higher classification abilities can be achieved.

Furthermore, to compare the above algorithms from the viewpoint of statistic, the Wilcoxon signed rank test [4] is employed. Note that the significant level is set as 0.05, that is, if the returned p-value is lower than 0.05, then it indicates that algorithms perform significantly different. The detailed results are shown in the following Table 4 and the italic portion is the data with a p-value greater than 0.05.

⁂ In Table 4, the value "NaN" indicates that the classification accuracies related to the two comparison algorithms over 20 different radii are exactly the same. Therefore, the value of Wilcoxon signed rank test is incalculable.

With a deep investigation of Table 4, it is not difficult to observe the follows.

Table 4 *p*-values for comparing classification accuracies based on different reducts

(a): *p*-values for comparing classification accuracies based on different reducts over optimistic MGRS

ID	Classifiers	Algorithm 2(O) vs. Algorithm 1(O)	Algorithm 2(O) vs. Gap	Algorithm 2(O) vs. GB	Algorithm 3(O) vs. Algorithm 1(O)	Algorithm 3(O) vs. Gap	Algorithm 3(O) vs. GB
1	KNN	2.86E-08	NaN	4.68E-10	2.86E-08	NaN	4.68E-10
	SVM	2.86E-08	NaN	4.68E-10	2.86E-08	NaN	4.68E-10
2	KNN	2.55E-07	0.0005	8.01E-09	1.91E-07	0.0040	8.01E-09
	SVM	2.21E-07	0.0040	2.10E-07	2.21E-07	0.0040	2.10E-07
3	KNN	4.68E-10	4.68E-10	4.68E-10	7.66E-09	2.24E-06	7.66E-09
	SVM	4.68E-10	NaN	4.68E-10	4.68E-10	NaN	4.68E-10
4	KNN	1.13E-08	8.01E-09	8.01E-09	1.12E-08	7.99E-09	7.99E-09
	SVM	4.68E-08	*1.0000*	7.98E-09	4.68E-08	*1.0000*	7.98E-09
5	KNN	2.38E-08	0.0040	7.92E-09	2.40E-08	3.99E-06	8.01E-09
	SVM	2.40E-08	0.0215	7.98E-09	2.40E-08	0.0215	7.98E-09
6	KNN	0.0005	0.0014	*0.0853*	0.0006	5.47E-05	*0.0852*
	SVM	4.07E-06	2.10E-07	7.99E-09	4.07E-06	2.10E-07	7.99E-09
7	KNN	3.34E-06	*0.7811*	7.92E-09	1.07E-06	0.0014	7.92E-09
	SVM	2.42E-07	*0.3881*	7.88E-09	2.42E-07	*0.3881*	7.88E-09
8	KNN	3.98E-06	3.99E-06	8.01E-09	3.49E-06	3.99E-06	7.99E-09
	SVM	2.59E-05	2.10E-07	8.01E-09	2.59E-05	2.10E-07	8.01E-09
9	KNN	0.0006	*0.3956*	4.76E-08	0.0038	*0.0853*	1.29E-08
	SVM	*0.0604*	0.1595	4.64E-05	*0.0604*	0.1595	4.64E-05
10	KNN	*0.5151*	0.0005	0.0215	*0.1884*	7.80E-09	0.0039
	SVM	0.0136	0.0040	7.96E-09	0.0136	0.0040	7.96E-09
11	KNN	4.79E-08	7.22E-09	7.22E-09	5.20E-08	0.0005	2.10E-07
	SVM	8.61E-05	7.19E-09	1.93E-07	8.61E-05	7.19E-09	1.93E-07
12	KNN	0.0027	2.09E-09	2.09E-09	*0.2688*	7.77E-09	7.77E-09
	SVM	7.44E-09	2.09E-09	2.09E-09	7.44E-09	2.09E-09	2.09E-09
13	KNN	0.0223	0.0040	8.01E-09	0.0026	3.99E-06	8.01E-09
	SVM	0.0021	*0.5729*	8.01E-09	0.0021	*0.5729*	8.01E-09
14	KNN	0.0054	3.23E-06	7.49E-09	0.0252	1.20E-05	7.65E-09
	SVM	0.0003	0.0305	7.58E-09	0.0002	0.0305	7.58E-09
15	KNN	1.08E-06	0.0039	*1.0000*	1.61E-07	0.0005	*0.5728*
	SVM	7.48E-07	0.0039	*0.5513*	7.48E-07	0.0039	*0.5513*
16	KNN	9.77E-09	*0.3421*	7.43E-10	6.04E-08	7.96E-09	7.96E-09
	SVM	6.65E-09	*0.3421*	7.43E-10	6.65E-09	*0.3421*	7.43E-10
17	KNN	1.29E-07	3.54E-05	7.79E-09	6.64E-08	2.94E-08	7.86E-09
	SVM	0.0040	*0.0853*	2.97E-08	0.0040	*0.0853*	2.97E-08
18	KNN	*0.8495*	0.0040	7.92E-09	*0.9892*	0.0005	7.95E-09
	SVM	*0.8707*	0.0185	2.08E-07	*0.8708*	0.0184	2.08E-07
19	KNN	2.35E-06	0.0215	3.99E-06	4.58E-06	0.0215	5.49E-05
	SVM	7.24E-06	0.0040	5.50E-05	7.24E-06	0.0040	5.50E-05
20	KNN	8.33E-08	2.08E-07	7.90E-09	7.05E-08	2.04E-07	7.73E-09
	SVM	7.13E-08	2.07E-07	7.85E-09	7.13E-08	2.07E-07	7.85E-09

(b): *p*-values for comparing classification accuracies based on different reducts over pessimistic MGRS

ID	Classifiers	Algorithm 2(P) vs. Algorithm 1(P)	Algorithm 2(P) vs. Gap	Algorithm 2(P) vs. GB	Algorithm 3(P) vs. Algorithm 1(P)	Algorithm 3(P) vs. Gap	Algorithm 3(P) vs. GB
1	KNN	4.75E-09	4.68E-10	4.68E-10	4.32E-08	0.0015	7.49E-09
	SVM	4.75E-09	4.68E-10	4.68E-10	2.47E-07	6.67E-05	6.84E-09

Table 4 (continued)

(b): *p*-values for comparing classification accuracies based on different reducts over pessimistic MGRS

ID	Classifiers	Algorithm 2(P) vs. Algorithm 1(P)	Algorithm 2(P) vs. Gap	Algorithm 2(P) vs. GB	Algorithm 3(P) vs. Algorithm 1(P)	Algorithm 3(P) vs. Gap	Algorithm 3(P) vs. GB
2	KNN	7.73E-09	4.68E-10	4.68E-10	6.62E-08	*0.5728*	8.01E-09
	SVM	7.73E-09	4.68E-10	4.68E-10	6.62E-08	5.50E-04	8.01E-09
3	KNN	4.68E-10	4.68E-10	4.68E-10	7.98E-09	1.23E-05	1.23E-05
	SVM	4.68E-10	4.68E-10	4.68E-10	7.76E-09	2.14E-02	7.76E-09
4	KNN	4.68E-10	4.68E-10	4.68E-10	8.01E-09	8.01E-09	8.01E-09
	SVM	4.91E-09	4.68E-10	4.68E-10	4.67E-08	8.53E-02	7.99E-09
5	KNN	2.04E-09	4.68E-10	4.68E-10	1.84E-08	5.88E-09	4.90E-04
	SVM	4.91E-09	4.68E-10	4.68E-10	4.67E-08	8.53E-02	7.99E-09
6	KNN	6.38E-09	4.68E-10	4.68E-10	5.52E-08	2.03E-07	*0.2528*
	SVM	5.20E-09	4.68E-10	4.68E-10	5.57E-08	2.06E-07	7.79E-09
7	KNN	4.75E-09	4.68E-10	4.68E-10	4.50E-08	0.7751	7.88E-09
	SVM	4.75E-09	4.68E-10	4.68E-10	4.52E-08	2.53E-01	7.92E-09
8	KNN	7.79E-09	4.68E-10	4.68E-10	6.64E-08	7.99E-09	7.99E-09
	SVM	6.50E-09	4.68E-10	4.68E-10	5.78E-08	2.10E-07	7.99E-09
9	KNN	4.61E-06	4.68E-10	NaN	*0.6179*	0.0025	6.29E-07
	SVM	3.31E-09	4.68E-10	NaN	5.30E-01	1.29E-03	6.32E-06
10	KNN	4.68E-10	4.68E-10	4.68E-10	7.51E-09	7.51E-09	7.51E-09
	SVM	4.68E-10	4.68E-10	4.68E-10	7.47E-09	7.47E-09	7.47E-09
11	KNN	5.65E-09	4.68E-10	4.68E-10	5.20E-08	*0.0853*	2.10E-07
	SVM	5.65E-09	4.68E-10	4.68E-10	1.11E-06	5.49E-04	8.53E-02
12	KNN	2.04E-09	4.68E-10	4.68E-10	2.24E-05	2.94E-08	7.86E-09
	SVM	2.04E-09	4.68E-10	4.68E-10	2.34E-08	5.41E-05	3.91E-06
13	KNN	7.53E-09	4.68E-10	4.68E-10	6.49E-08	5.50E-05	8.01E-09
	SVM	2.04E-09	4.68E-10	4.68E-10	2.34E-08	5.41E-05	3.91E-06
14	KNN	6.66E-09	4.68E-10	4.68E-10	5.79E-08	7.12E-03	7.82E-09
	SVM	2.08E-09	4.68E-10	4.68E-10	2.36E-08	1.36E-01	7.70E-09
15	KNN	7.55E-09	4.68E-10	4.68E-10	6.45E-08	0.0005	*0.7810*
	SVM	7.44E-09	4.68E-10	4.68E-10	6.39E-08	2.53E-01	5.73E-01
16	KNN	2.67E-09	4.68E-10	4.68E-10	2.93E-08	7.99E-09	1.00E+00
	SVM	2.67E-09	4.68E-10	4.68E-10	2.94E-08	4.02E-03	1.00E+00
17	KNN	6.16E-09	4.68E-10	4.68E-10	1.64E-03	7.88E-09	7.78E-02
	SVM	2.67E-09	4.68E-10	4.68E-10	2.94E-08	4.02E-03	1.00E+00
18	KNN	7.49E-09	4.68E-10	4.68E-10	6.36E-08	0.2531	3.94E-06
	SVM	4.68E-10	4.68E-10	4.68E-10	3.71E-06	5.19E-05	5.19E-05
19	KNN	3.99E-09	4.68E-10	4.68E-10	3.99E-08	*1.0000*	4.02E-03
	SVM	3.99E-09	4.68E-10	4.68E-10	3.99E-08	5.73E-01	8.54E-02
20	KNN	6.31E-09	4.68E-10	4.68E-10	5.19E-08	3.72E-06	7.17E-09
	SVM	6.31E-09	4.68E-10	4.68E-10	5.47E-08	8.50E-02	7.66E-09

(1) No matter optimistic or pessimistic case is considered, there is a great difference between the classification accuracies obtained by reducts related to our Algorithms 2–3 and those obtained by reducts related to conventional MGRS, gap and granular ball rough sets. Take "Yeast" data set as an example, whether the optimistic or pessimistic case is used, all of the cor-

responding *p*-values are lower than 0.05. Furthermore, by combing with the observations over Tables 2–3, we can conclude that our proposed model does provide reducts which can significantly improve the classification performances.

(2) In our Triple-**G** model, there is a great difference between the classification accuracies obtained by
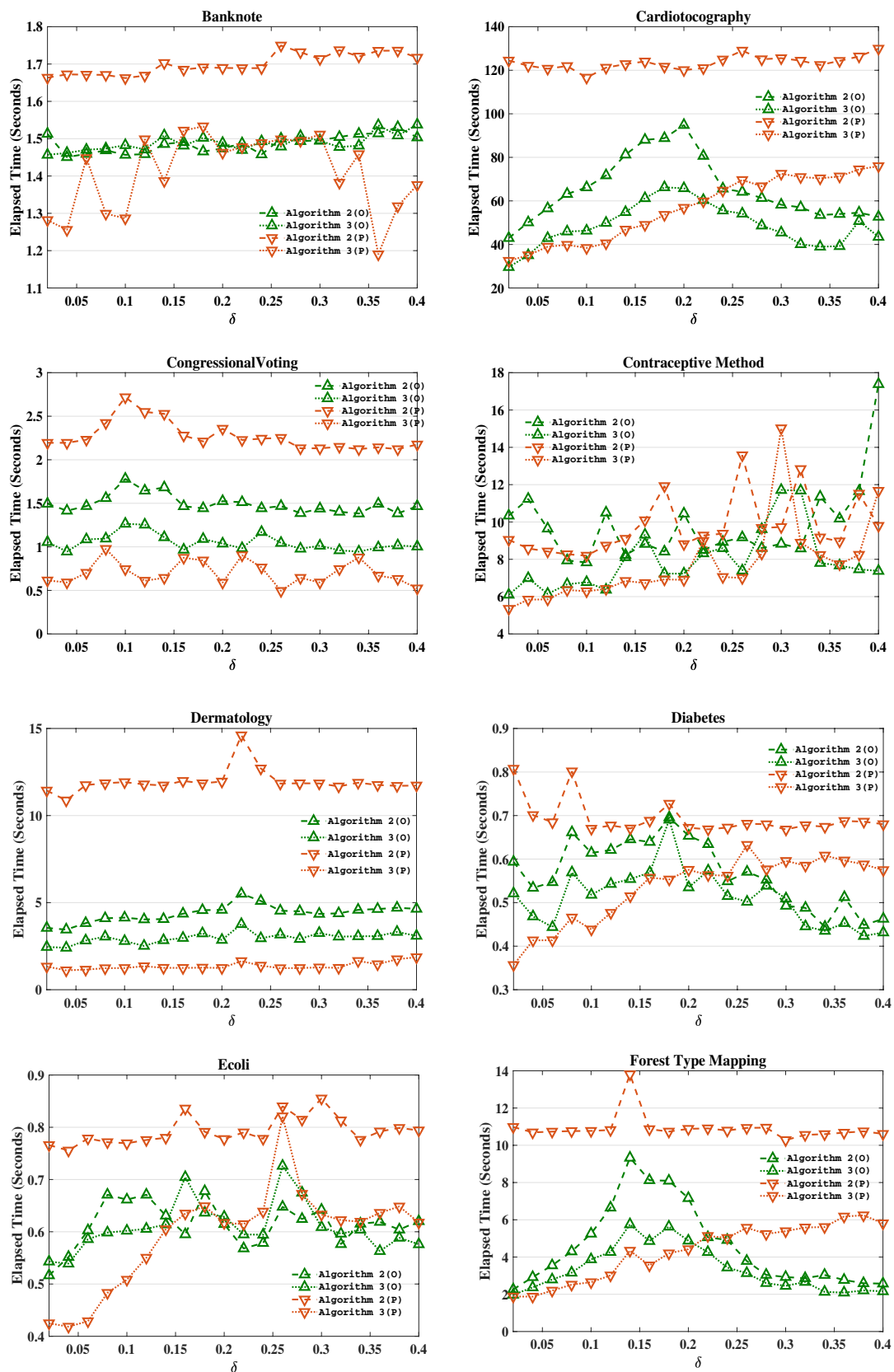
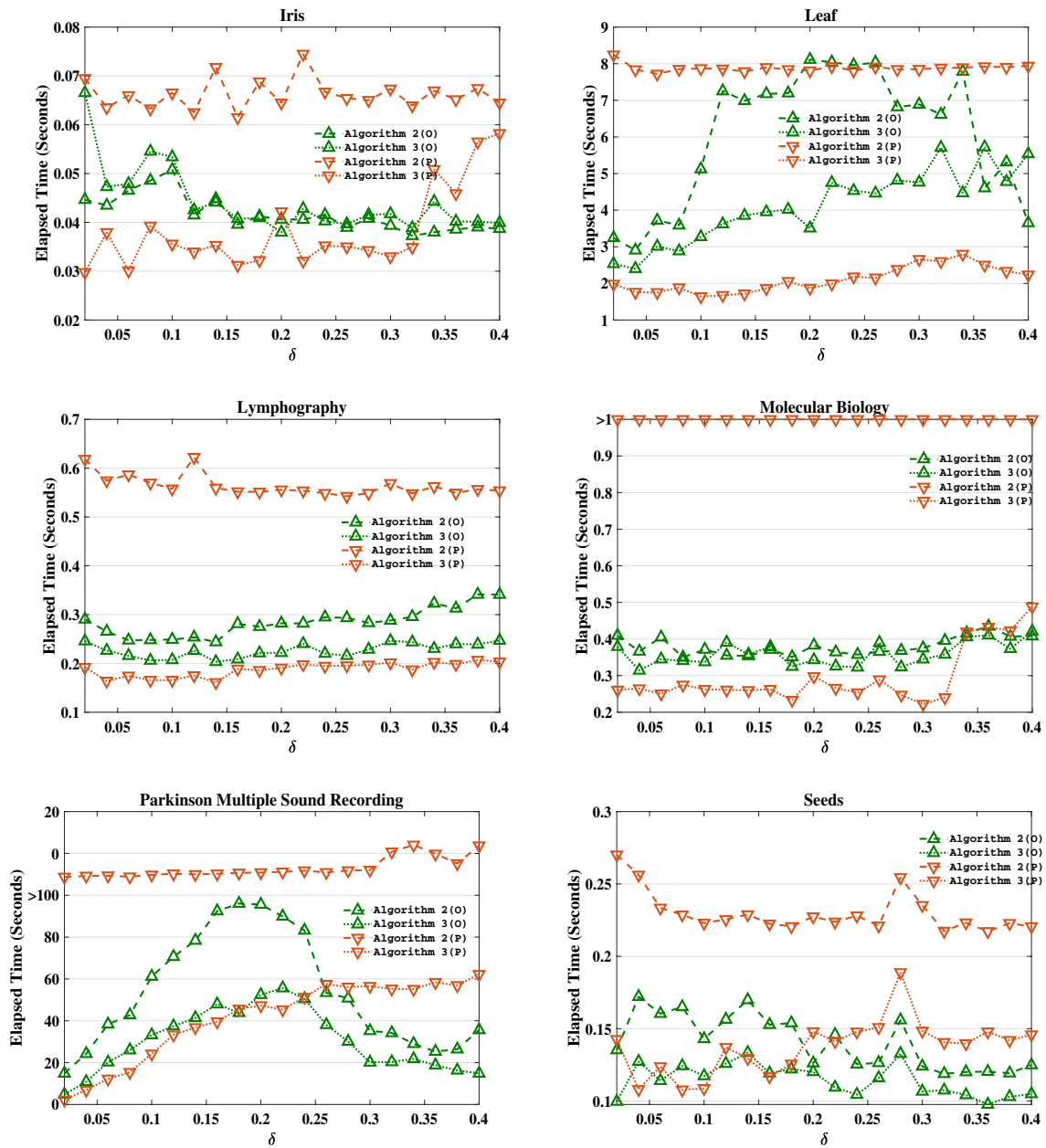**Fig. 2** The elapsed time of deriving reducts
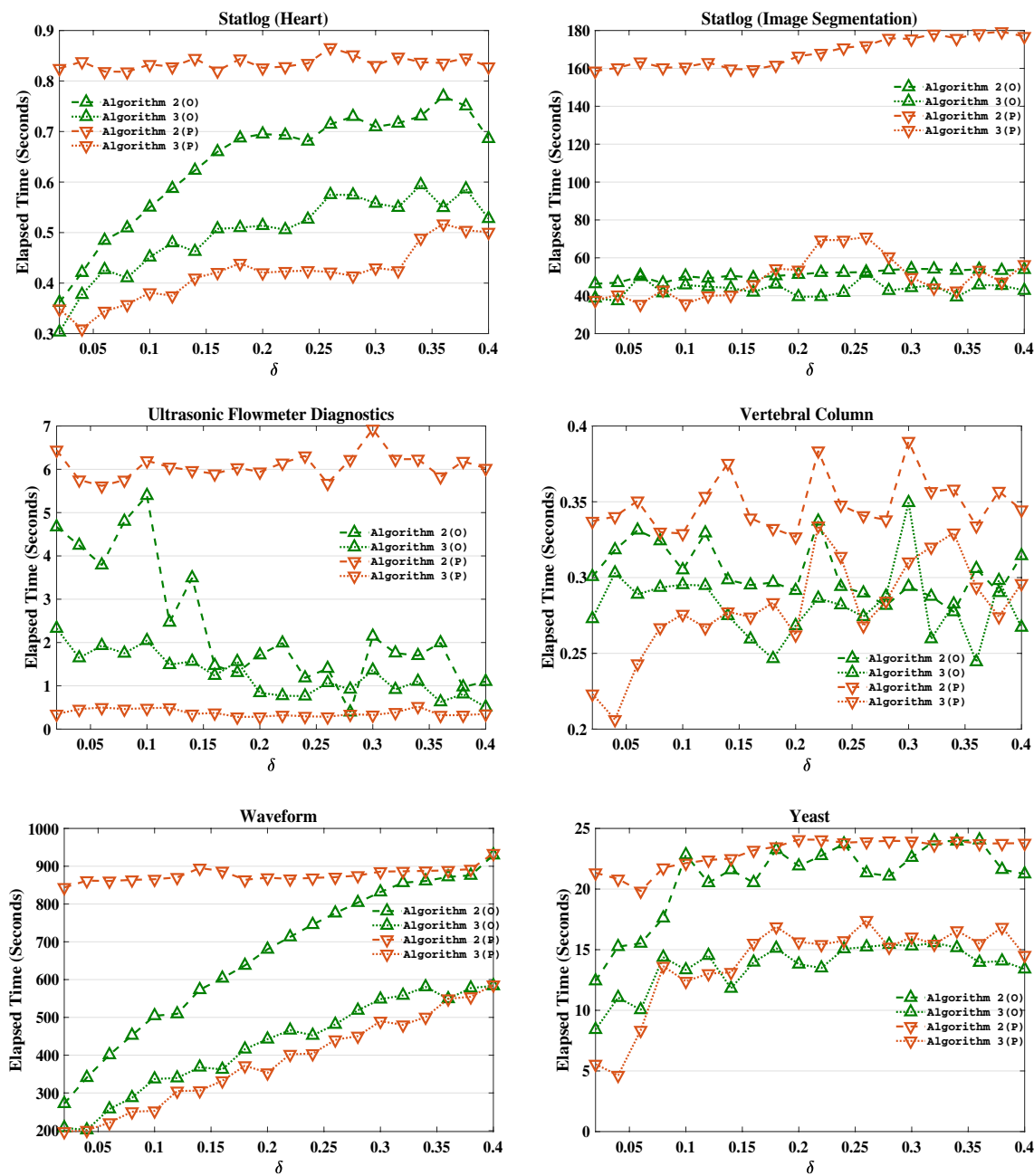
Fig. 2 (continued)

Fig. 2 (continued)

reduct related to our pessimistic case and that obtained by reduct related to optimistic case. For example, when the Algorithm 2 and Algorithm 1 are compared, over 20 data sets, the *p*-values of 4 data sets are greater than 0.05 in optimistic case. Similarly, the *p*-value of all the data sets are lower than 0.05. Furthermore, by combing with the observations over Tables 2–3, we can conclude that our proposed model in the pessimistic case can improve the classification performance better.

4.3 Comparisons of elapsed time

In this experiment, the elapsed time of obtaining reduct by using two different algorithms in optimistic and pessimistic cases will be compared.

With a deep investigation of Fig. 2, it is not difficult to observe the follows.

(1) In most cases, by comparing with optimistic Triple-**G** MGRS, the pessimistic Triple-**G** MGRS requires more

elapsed time for deriving reduct. Take "Dermatology" data set as an example, in Fig. 2, if $\delta = 0.2$, the elapsed time of obtaining reduct related to pessimistic Triple-**G** MGRS is 11.9529 s, it is superior to the elapsed time of obtaining reduct related to optimistic Triple-**G** MGRS, i.e., 4.5848 s.

(2) By comparing with Algorithm 2, Algorithm 3 is designed based on the consideration of accelerating the process of obtaining reduct by Triple-**G** MGRS. Therefore, such improved algorithm can significantly reduce the elapsed time of obtaining reduct. Take "Waveform Database Generator(Version 1)" data set as an example, in Fig. 2, if $\delta = 0.2$, the elapsed time of obtaining reducts related to optimistic and pessimistic Triple-**G** MGRS, i.e., 680.0876 and 869.5392 s, it is superior to the elapsed time of obtaining reducts related to optimistic and pessimistic Triple-**G** MGRS with attribute grouping, i.e., 442.3297 and 353.2678 s.

(3) From discussions above, we can conclude that the time consumption of obtaining reduct will be effectively reduced by introducing the acceleration mechanism. Furthermore, through combing the analyses shown in both Sects. 4.2 and 4.3, either optimistic or pessimistic case is considered, the improved Algorithm 3 can effectively reduce the time consumption of obtaining reduct without reducing the classification performance.

5 Conclusions and future perspectives

In the field of researching MGRS, to characterize the uncertainty of the target in terms of multi-view based heterogeneous information granulation, a Triple-**G** MGRS is developed. Different from previous models, our proposed model is actually constructed by using both the parameterized and data-adaptive information granulations. Therefore, our Triple-**G** MGRS presents not only the principle of multi-view, but also the mechanism of heterogeneous information granulation. Additionally, by considering the efficiency of searching Triple-**G** MGRS based reduct, an acceleration strategy called attribute grouping is introduced into the procedure of forward greedy searching. The experimental results have not only demonstrated the effectiveness of our new model in terms of the classification ability, but also verified the effectiveness of our accelerator for quickly deriving reduct. The following topics deserve our further investigations.

(1) In this paper, we have only used three neighborhood related techniques for realizing heterogeneous information granulation, diversified forms for obtaining information granulations will be further selected and used in our future study.

(2) Though the attribute grouping strategy can significantly reduce the time consumption of obtaining our Triple-**G** MGRS based reduct, such accelerator is only designed based on the perspective of sample, it can be further improved by combining both the sample and attribute based perspectives.

References

1. Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2(3):1–39
2. Cover TM, Hart PE (1967) Nearest neighbor pattern classification. IEEE Trans Inform Theory 13(1):21–27
3. Chen Y, Liu KY, Song JJ, Fujita H, Yang XB, Qian YH (2020) Attribute group for attribute reduction. Inform Sci 535:64–80
4. Demšar J (2006) Statistical comparisons of classifiers over multiple data sets. J Mach. Learn. Res 7(1):1–30
5. Jain AK (2010) Data clustering: 50 years beyond K-means. Pattern Recognit Lett 31(8):651–666
6. Jiang GX, Wang WJ (2017) Markov cross-validation for time series model evaluations. Inform Sci 375:219–233
7. Jiang ZH, Liu KY, Yang XB, Yu HL, Fujita H, Qian YH (2020) Accelerator for supervised neighborhood based attribute reduction. Int J Approx Reasoning 119:122–150
8. Kong QZ, Xu WH (2019) The comparative study of covering rough sets and multi-granulation rough sets. Soft Comput 23:3237–3251
9. Kong QZ, Zhang XW, Xu WH, Xie ST (2020) Attribute reducts of multi-granulation information system. Artif Intell Rev 53:1353–1371
10. Hu QH, Yu DR, Xie ZX (2008) Neighborhood classifiers. Expert Syst Appl 34(2):866–876
11. Hu QH, Pedrycz W, Yu DR, Lang J (2010) Selecting discrete and continuous features based on neighborhood decision error minimization. IEEE Trans Syst Man Cybernet Part B-Cybernet 40(1):137–150
12. Jiang ZH, Yang XB, Yu HL, Liu D, Wang PX, Qian YH (2019) Accelerator for multi-granularity attribute reduction. Knowl-Based Syst 177:145–158
13. Jiang ZH, Liu KY, Song JJ, Yang XB, Li JH, Qian YH (2021) Accelerator for crosswise computing reduct. Appl Soft Comput 98:106740
14. Ju HR, Ding WP, Yang XB, Fujita H, Xu SP (2021) Robust supervised rough granular description model with the principle of justifiable granularity. Appl Soft Comput 110:107612
15. Ju HR, Pedrycz W, Li HX, Ding WP, Yang XB, Zhou XZ (2019) Sequential three-way classifier with justifiable granularity. Knowl-Based Syst 163:103–119
16. Ju HR, Yang XB, Song XN, Qi YS (2014) Dynamic updating multigranulation fuzzy rough set: approximations and reducts. Int J Mach Learn Cybernet 5(6):981–990
17. Ju HR, Yang XB, Yu HL, Li TJ, Yu DJ, Yang JY (2016) Cost-sensitive rough set approach. Inform Sci 355–356:282–298

18. Lin GP, Liang JY, Qian YH (2015) Uncertainty measures for multigranulation approximation space. Int J Uncert Fuzziness Knowl-Based Syst 23(3):443–457

19. Lin GP, Liang JY, Qian YH (2013) Multigranulation rough sets: from partition to covering. Inform Sci 241:101–118

20. Li JH, Liu ZM (2020) Granule description in knowledge granularity and representation. Knowledge-Based Syst 203:106160

21. Lin GP, Qian YH, Li JJ (2012) NMGRS: neighborhood-based multigranulation rough sets. Int J Approx Reason 53(7):1080–1093

22. Li FJ, Qian YH, Wang JT, Liang JY (2017) Multigranulation information fusion: a Dempster–Shafer evidence theory-based clustering ensemble method. Inform Sci 378:389–409

23. Li JH, Ren Y, Mei CL, Qian YH, Yang XB (2016) A comparative study of multi granulation rough sets and concept lattices via rule acquisition. Knowledge-Based Systems 91:152–164

24. Liu KY, Yang XB, Fujita H, Liu D, Yang X, Qian YH (2019) An efficient selector for multi-granularity attribute reduction. Inform Sci 505:457–472

25. Li JZ, Yang XB, Song XN, Li JH, Wang PX, Yu DJ (2019) Neighborhood attribute reduction: a multi-criterion approach. Int J Mach Learn Cybernet 10(4):731–742

26. Liu Y, Huang WL, Jiang YL, Zeng ZY (2014) Quick attribute reduct algorithm for neighborhood rough set model. Inform Sci 271:65–81

27. Min F, He HP, Qian YH, Zhu W (2011) Test-cost-sensitive attribute reduction. Inform Sci 181(22):4928–4942

28. Pawlak Z (1991) Rough sets: theoretical aspects of reasoning about data. Kluwer Academic Publishers

29. Qian YH, Cheng HH, Wang JT, Liang JY (2017) Grouping granular structures in human granulation intelligence. Inform Sci 382:150–169

30. Qian YH, Liang JY, Wang F (2009) A new method for measuring the uncertainty in incomplete information systems. Int J Uncertainty Fuzziness Knowl-Based Syst 17(6):855–880

31. Qian YH, Liang JY, Pedrycz W, Dang CY (2010) Positive approximation: an accelerator for attribute reduction in rough set theory. Artif Intell 174(9–10):597–618

32. Qian YH, Liang JY, Yao YY, Dang CY (2010) MGRS: a multigranulation rough set. Inform Sci 180(6):949–970

33. Qian YH, Liang XY, Lin GP, Guo Q, Liang JY (2017) Local multigranulation decision-theoretic rough sets. Int J Approx Reason 82:119–137

34. Qian YH, Liang XY, Wang Q (2018) Local rough set: a solution to rough data analysis in big data. Int J Approx Reason 97:38–63

35. Sun BZ, Ma WM, Qian YH (2017) Multigranulation fuzzy rough set over two universes and its application to decision making. Knowl-Based Syst 123:61–74

36. Sang BB, Yang L, Chen HM, Xu WH, Guo YT, Yuan Z (2019) Generalized multi-granulation double-quantitative decision-theoretic rough set of multi-source information system. Int J Approx Reason 115:157–179

37. She YH, He XL, Qian T, Wang QQ, Zeng WL (2019) A theoretical study on object-oriented and property-oriented multi-scale formal concept analysis. Int J Mach Learn Cybernet 10(2):3263–3271

38. Rao XS, Yang XB, Yang X, Chen XJ, Liu D, Qian YH (2020) Quickly calculating reduct: an attribute relationship based approach. Knowledge-Based Syst 200:106014

39. Tsang ECC, Song JJ, Chen DG, Yang XB (2019) Order based hierarchies on hesitant fuzzy approximation space. Int J Mach Learn Cybernet 10:1407–1422

40. Wang CZ, Hu QH, Wang XZ, Chen DG, Qian YH, Dong Z (2018) Feature selection based on neighborhood discrimination index. IEEE Trans Neural Netw Learn Syst 29(7):2986–2999

41. Wang XZ, Li JH (2020) New advances in three way decision, granular computing and concept lattice. Int J Mach Learn Cybernet 11(5):945–946

42. Wang XZ, Li JH (2020) New advances in three-way decision, granular computing and concept lattice. Int J Mach Learn Cybernet 11(5):945–946

43. Wang XZ, Tsang ECC, Zhao SY, Chen DG, Yeung DS (2007) Learning fuzzy rules from fuzzy samples based on rough set technique. Inform Sci 177(20):4493–4514

44. Wei W, Liang JY (2019) Information fusion in rough set theory?: an overview. Inform Fus 48:107–118

45. Wei W, Wang JH, Liang JY, Mi X, Dang CY (2015) Compacted decision tables based attribute reduction. Knowl-Based Syst 86:261–277

46. Wang CZ, Shi YP, Fan XD, Shao MW (2018) Attribute reduction based on k-nearest neighborhood rough sets. Int J Approx Reason 106:18–31

47. Wang X, Wang PX, Yang XB, Yao YY (2021) Attribution reduction based on sequential three-way search of granularity. Int J Mach Learn Cybernet 12(5):1439–1458

48. Xu WH, Guo YT (2016) Generalized multigranulation double-quantitative decision-theoretic rough set. Knowl-Based Syst 105:190–205

49. Xu WH, Li WT (2016) Granular computing approach to two-way learning based on formal concept analysis in fuzzy datasets. IEEE Trans Cybernet 46(2):166–179

50. Xia SY, Liu YS, Ding X, Wang GY, Yu H, Lu YG (2019) Granular ball computing classifiers for efficient, scalable and robust learning. Inform Sci 483:136–152

51. Xia SY, Zhang Z, Li WH, Wang GY, Giem E, Chen ZZ (2020) GBNRS: a novel rough set algorithm for fast adaptive attribute reduction in classification. IEEE Trans Knowl Data Eng. https://doi.org/10.1109/TKDE.2020.2997039

52. Xu SP, Yang XB, Yu HL, Yu DJ, Yang JY, Tsang ECC (2016) Multi-label learning with label-specific feature reduction. Knowl-Based Syst 104:52–61

53. Yang XB, Liang SC, Yu HL, Gao S, Qian YH (2019) Pseudo-label neighborhood rough set: measures and attribute reductions. Int J Approx Reason 105:112–129

54. Yang X, Li TR, Liu D, Fujita H (2019) A temporal-spatial composite sequential approach of three-way granular computing. Inform Sci 486:171–189

55. Yang XB, Qian YH, Yang JY (2013) On characterizing hierarchies of granulation structures. Fundamenta Informaticae 123(3):365–380

56. Yao YY (2016) A triarchic theory of granular computing. Granular Comput 1(2):145–157

57. Yang L, Xu WH, Zhang XY, Sang BB (2020) Multi-granulation method for information fusion in multi-source decision information system. Int J Approx Reason 122(5):47–65

58. Yao YY, Zhang XY (2017) Class-specfic attribute reducts in rough set theory. Inform Sci 418–419:601–618

59. Yang XB, Qi Y, Yu HL, Song XN, Yang JY (2014) Updating multigranulation rough approximations with increasing of granular structures. Knowl-Based Syst 64:59–69

60. Yang XB, Yao YY (2018) Ensemble selector for attribute reduction. Appl Soft Comput 70:1–11

61. Yang XB, Zhang YQ, Yang JY (2012) Local and global measurements of MGRS rules. Int J Comput Intell Syst 5(6):1010–1024

62. Yang XB, Qi YS, Song XN, Yang JY (2013) Test cost sensitive multigranulation rough set: model and minimal cost selection. Inform Sci 250:184–199

63. Zhang PF, Li TR, Wang GQ, Luo C, Chen HM, Zhang JB, Wang DX, Yu Z (2021) Multi-source information fusion based on rough set theory: a review. Inform Fus 68:85–117

64. Zhang X, Mei CL, Chen DG, Li JH (2016) Feature selection in mixed data: a method using a novel fuzzy rough set-based information entropy. Pattern Recognit 56:1–15

65. Zhou P, Hua XG, Li PP, Wu XD (2019) Online streaming feature selection using adapted neighborhood rough set. Inform Sci 481:258–279