Information Sciences 505 (2019) 457–472

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

An efficient selector for multi-granularity attribute reduction

Keyu Liu[a,c], Xibei Yang[a,c,*], Hamido Fujita[b], Dun Liu[d], Xin Yang[e], Yuhua Qian[f]

[a] School of Computer, Jiangsu University of Science and Technology, Zhenjiang 212003, China
[b] Faculty of Information Technology, Ho Chi Minh City University of Technology (HUTECH), Ho Chi Minh City, Vietnam
[c] Key Laboratory of Data Science and Intelligence Application, Fujian Province University, Zhangzhou 363000, China
[d] School of Economics and Management, Southwest Jiaotong University, Chengdu 610031, China
[e] School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu 611130, China
[f] Institute of Big Data Science and Industry, Shanxi University, Taiyuan 030006, China

ARTICLE INFO

Article history:
Received 3 February 2019
Revised 11 July 2019
Accepted 14 July 2019
Available online 15 July 2019

Keywords:
Attribute reduction
Granular computing
Multi-granularity
Rough set

ABSTRACT

Presently, the mechanism of multi-granularity has been frequently realized by various mathematical tools in Granular Computing especially rough set. Nevertheless, as a key topic of rough set, attribute reduction has been rarely exploited by the concept of multi-granularity. To fill such a gap, Multi-Granularity Attribute Reduction is defined to characterize reduct which satisfies the intended multi-granularity constraint instead of one and only one granularity based constraint. Furthermore, to accelerate the searching process of reduct, Multi-Granularity Attribute Selector is introduced into the framework of heuristic algorithm. Its key procedure is twofold including: (1) fuse all the granularities based measure-values to construct the multi-granularity constraint; (2) integrate the suitable granularities based measure-values to evaluate the candidate attributes. Based on the multi-granularity structure formed by neighborhood rough set, the experimental results over 20 UCI data sets demonstrate that compared with single granularity attribute reduction, our selector can not only generate reducts which may not contribute to poorer classification performances, but also significantly reduce the elapsed time of computing reducts. This research suggests the new trend of attribute reduction in multi-granularity environment.

1. Introduction

Attribute reduction [2,14,21], a rough set based feature selection, aims to reduce the dimensionality of data by searching for a qualified reduct, i.e., a subset of conditional attributes. For such purpose, as reported by Yao et al. [43], a reduct is generally required to satisfy one intended constraint. Correspondingly, with respect to different constraints, various forms of attribute reduction have been explored [3,4,9,13,18].

With a careful reviewing of previous researches, from the viewpoint of Granular Computing (GrC) [11,26,36], it is not difficult to reveal that most of the attribute reduction approaches possess a similar mechanism: to form a suitable constraint for the corresponding attribute reduction, information granulation [12,22,37] is frequently conducted beforehand by using indiscernibility relation [21], distance function [47], clustering analysis [10], etc. It should be emphasized that different results of information granulation may contribute to different reducts, because with the varying results of information

* Corresponding author at: School of Computer, Jiangsu University of Science and Technology, Zhenjiang 212003, China.
E-mail addresses: liukeyu@stu.just.edu.cn (K. Liu), jsjxy_yxb@just.edu.cn (X. Yang), h.fujita@hutech.edu.vn (H. Fujita).

https://doi.org/10.1016/j.ins.2019.07.051
0020-0255/

granulation, the constraint related to the reduct may be stricter or looser. An immediate problem is how to distinguish different series of results, including information granulation, constraint and even reduct. Regrading such problem, fortunately, the concept of granularity [24] can be used to quantitatively characterize these results for revealing the difference.

Actually, the concept of granularity can be acquired by various approaches. To the best of our knowledge, most of previous researches related to granularity can be categorized into the following three strategies.

- **Parameter based granularity**. The granularity is closely related to the appointed parameter. For such case, to distinguish different results of information granulation, the immediate granularity is actually determined by the value of parameter, it follows that the difference between those results of information granulation can be reflected by the difference between the values of parameters. It should be noticed that smaller value of parameter generally contributes to a finer granularity. For instance, in neighborhood rough set [6], smaller value of parameter (radius for calculating neighborhoods of samples) may generate smaller size of the neighborhoods of samples, it follows that the finer granularity will be obtained. Similar explanation can also be observed in Gaussian kernel based fuzzy rough set [7].
- **Sample based granularity**. The granularity is closely related to the employed samples. For such case, to distinguish different results of information granulation, the immediate granularity is actually determined by the structure of samples, it follows that the difference between those results of information granulation can be reflected by the difference between the structures of samples. For instance, in *K*-fold cross-validation, *K* different sets of training and testing samples can be regrouped [8]. In view of GrC, *K* results of information granulation are derived. Immediately, the difference between these results can be reflected by structures of samples since for those regrouped sets, the contained training and testing samples are significantly different.
- **Attribute based granularity**. The granularity is closely related to the considered attributes. For such case, to distinguish different results of information granulation, the immediate granularity is actually determined by the distinguishability of attributes, it follows that the difference between those results of information granulation can be reflected by the difference between the distinguishabilities of attributes. As suggested by Xu et al. [35], different attributes with different distinguishabilities induce multiple equivalence relations, and then those relations may naturally construct multiple different granularities. Additionally, it should be noticed that the stronger distinguishability of attributes generally contributes to a finer granularity. For instance, as reported by Liao et al. [16], in the problem of feature selection, the considered attributes (conditional attributes, i.e., features) are supposed to have a feature-value granularity. Compared with those deleted attributes, the remained attributes may have stronger distinguishability since they may provide stronger relevances or better generalization performances, it follows that the finer granularity will be obtained.

Following these different forms of granularity, it is well known that most of the existing attribute reduction approaches are essentially single granularity based ones, which are referred to as single granularity attribute reductions in this paper. This is mainly because the construction of constraint focuses on one and only one fixed result of information granulation induced by single parameter, single sample space or single attribute set. Nevertheless, one single granularity attribute reduction may involve some limitations as follows.

1. Single granularity attribute reduction may result in the poor adaptability of the derived reduct to the problem of granularity diversity. As what has been reported by Yang and Yao [41], if a reduct is generated over one and only one considered granularity, it may not be still the qualified reduct over a little finer or coarser granularity which may be caused by slight variation of data.
2. Single granularity attribute reduction may increase the time consumption if multiple different granularities [44] are required. For example, given a set of multiple different granularities, a simple and direct method to design attribute reduction is: repeat the single granularity attribute reduction in terms of the number of considered granularities. Obviously, such process is too time-consuming.

To handle the limitations mentioned above, it is necessary to develop a novel thinking: re-consider attribute reduction in the case of multi-granularity. From this point of view, Multi-Granularity Attribute Reduction (MGAR) is proposed. Fig. 1(a) shows the framework of single granularity attribute reduction while the framework of our proposed MGAR is illustrated in Fig. 1(b).

Obviously, following Fig. 1, different from single granularity attribute reduction, with respect to multiple different granularities, two or more information granulations are required in MGAR. Moreover, based on these results of information granulation, a specific constraint, called multi-granularity constraint, is established and then attribute reduction can be conducted. Actually, as the most distinguished characteristic in MGAR, such multi-granularity constraint allows us to select attributes in a more versatile way. It follows that the derived reduct may offer us higher adaptability to granularity world.

Alluding to the definition of MGAR, how to realize it in detail becomes an interesting topic which deserves more investigations. Actually, the consideration of multiple different granularities in MGAR may easily result in greater complexity of the searching process of reduct. Immediately, an efficient selector for MGAR is proposed to alleviate such weakness, which is referred to as Multi-Granularity Attribute Selector (MGAS) in the context of this paper. Such selector is expected to speed up the searching process of multi-granularity reduct. The acceleration mechanism in MGAS is mainly embodied in two open problems as follows.

1. How to construct the multi-granularity constraint. To address such problem, a fused measure-value based multi-granularity constraint is designed. With such strategy, an undesired case can be avoided: the intended multi-granularity

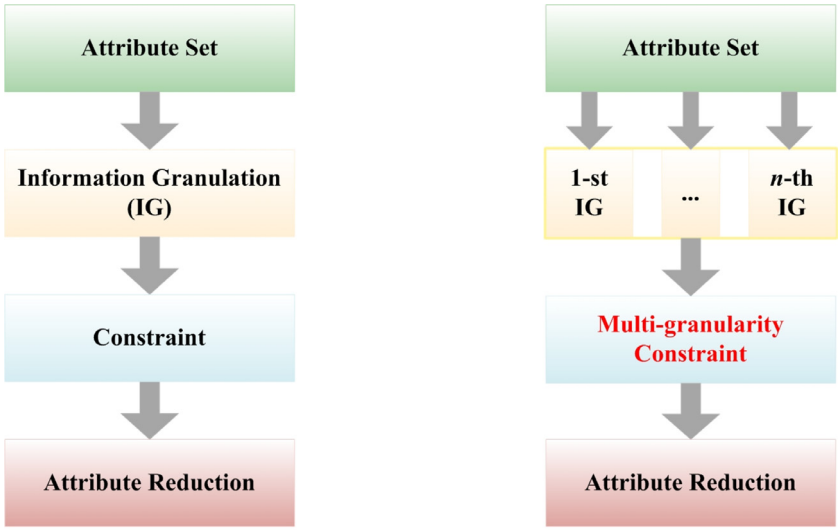


Fig. 1. The frameworks of single granularity attribute reduction and MGAR.

constraint may be so strict that no attribute can be deleted. Additionally, it can reduce the number of constraints with respect to multiple different granularities.

2. How to evaluate the candidate attributes. To address such problem, the finest and coarsest granularities based evaluations of attributes are concerned. With such strategy, some redundant evaluations are pruned, it follows that the elapsed time of evaluating attributes is effectively reduced.

The main contribution of this work can be summarized as the following aspects: (1) through analyzing attribute reduction from the viewpoint of GrC, we offer the revelation that most of the previous attribute reduction approaches are essentially based on single granularity; (2) to solve the inherent limitations in single granularity attribute reduction, Multi-granularity Attribute Reduction is proposed; (3) to accelerate the searching process of multi-granularity reduct, Multi-granularity Attribute Selector is designed; (4) extensive experimental results are analyzed to demonstrate that our selector is effective in the classification-oriented attribute reduction and efficient in the computational issue of reduct.

The remainder of this paper is organized as follows. In Section 2, we briefly review related works of granularity. Preliminary knowledge is presented in Section 3. Multi-granularity Attribute Reduction (MGAR) and Multi-granularity Attribute Selector (MGAS) are introduced in Section 4. In Section 5, comparative experimental results over UCI data sets are shown, as well as the corresponding analyses. This paper is ended with conclusions and future perspectives in Section 6.

2. Preliminary knowledge

2.1. Granular computing and rough set

Up to now, by applying the mechanism of information granulation to problem solving, Granular Computing (GrC) has been developed as an umbrella which covers theories, methodologies and techniques related to the concepts of granule and granularity [17,38,39]. As basic elements in GrC, the concepts of granule and granularity have been thoroughly investigated [24,31]. In a wide sense, granule and granularity can be described as follows.

- Granule: a collection of entities drawn together by criteria of indistinguishability, similarity or functionality.
- Granularity: a discrimination ability of the result of information granulation.

Generally speaking, the concepts of granule and granularity are broad, and as one of the most advanced approaches to popularizing GrC, rough set [15] has been frequently employed for interpreting them specifically. In the following, we will briefly review some basic notions related to rough set and then provide the detailed interpretations of these concepts.

In the rough set theory, a decision system can be described by a pair such that $\mathcal{DS} =< U, AT \cup \{d\} >$: U is a nonempty and finite set of samples, called the universe; AT is a nonempty and finite set of conditional attributes; d is the decision attribute (note that in this paper, to simply our discussions, the individual word "attribute" denotes "conditional attribute" since our attribute reduction related research mainly focuses on the conditional attributes). Furthermore, $\forall x \in U$, $a(x)$ indicates its value over conditional attribute $a \in AT$ and $d(x)$ denotes its decision attribute value, i.e., the label.

In most of the rough set approaches, the binary relation is well accepted as an effective method to conduct information granulation. It is worth noting that the binary relation can be expressed in various ways, it follows that the detailed explanations of granule and granularity can be different. Some simple examples are given in the following example.

Example 1.

- If the binary relation is Pawlak's indiscernibility relation [21] such that $\mathrm{IND}_A = \{(x, y) \in U^2 : \forall a \in A, a(x) = a(y)\}$, in which $A \subseteq AT$ is a conditional attribute subset, then information granulation is to derive all the equivalence classes which constitute a partition over U through using IND_A. Correspondingly, each single equivalence class can be referred to as a granule. For one given data, the immediate granularity directly reflects the distinguishability of the considered conditional attribute subset.
- If the binary relation is neighborhood relation [6] such that $\mathrm{N}_A^\delta = \{(x, y) \in U^2 : \triangle_A(x, y) \leq \delta\}$, in which $\triangle_A(x, y)$ is the distance between samples x, y over A and δ is the radius, then information granulation is to derive all the neighborhoods of samples through using N_A^δ. Correspondingly, each neighborhood of one sample can be referred to as a granule. For one given data, the immediate granularity reflects the distinguishability based on both the considered conditional attribute subset and the appointed radius simultaneously.

Following Example 1, in view of rough set theory, various factors, e.g., different considered conditional attribute subsets and different appointed radii mentioned above may result in the variation of granularity.

2.2. Single granularity attribute reduction

As one of the key topics in rough set theory, attribute reduction has been widely concerned [21,33,34]. With respect to different requirements, various definitions of attribute reduction have been proposed as indicated in Refs. [20,27,28,32,40,41]. To facilitate the further understanding of attribute reduction, as reported by Yao et al. [43], a general form of attribute reduction from the viewpoint of GrC can be presented as follows.

Definition 1. Given a decision system $\mathcal{DS}$ and a granularity $\mathbb{G}$, assuming that $\rho^{\mathbb{G}}$-constraint is a constraint based on a considered measure ρ over $\mathbb{G}$, then any attribute subset in AT, A is referred to as a reduct in terms of $\mathbb{G}$ if and only if

(1) A satisfies the $\rho^{\mathbb{G}}$-constraint;
(2) $\forall B \subset A$, B does not satisfy the $\rho^{\mathbb{G}}$-constraint.

Following Definition 1, it is obvious that with one and only one considered granularity $\mathbb{G}$, the constraint presented above is a single granularity constraint. Consequently, attribute reduction defined in Definition 1 is actually explored based on one and only one granularity, which can be referred to as a single granularity attribute reduction. It follows that the generated reduct can be characterized as a single granularity reduct.

Moreover, in addition to the given granularity, the considered measure ρ also plays a crucial role in the construction of constraint. It should be pointed out that different measures possess different properties, and they may lead to various forms of $\rho^{\mathbb{G}}$-constraint as shown in the following Example 2.

Example 2.

- If the considered measure is positively correlated, that is, the measure-value is expected to be as high as possible, e.g., the measures of approximation quality [21] and classification accuracy, then the $\rho^{\mathbb{G}}$-constraint is usually expressed as "$\rho^{\mathbb{G}}(A) \geq \rho^{\mathbb{G}}(AT)$" where $\rho^{\mathbb{G}}(A)$ is the measure-value derived by the conditional attribute subset A over the considered granularity $\mathbb{G}$.
- If the considered measure is negatively correlated, that is, the measure-value is expected to be as low as possible, e.g., the measures of conditional entropy [45] and classification error rate, then the $\rho^{\mathbb{G}}$-constraint is usually expressed as "$\rho^{\mathbb{G}}(A) \leq \rho^{\mathbb{G}}(AT)$".

Following Example 2, it is easy to know that one qualified reduct is actually required to preserve, increase or decrease one considered measure-value. It further indicates that most of the attribute reductions are single granularity ones, because as a comparative index to check whether the reduct satisfies the intended constraint, the measure-value of the whole attribute set such that $\rho^{\mathbb{G}}(AT)$ is derived over one and only one granularity.

2.3. Heuristic algorithm of single granularity attribute reduction

Following Def. 1, to find the reduct, heuristic method is frequently employed due to its lower time consumption. Based on the framework of heuristic algorithm, Yao et al. [43] presented three types of searching strategies, i.e., "addition strategy", "deletion strategy" and "addition-deletion strategy". To simplify our discussions, only addition strategy is discussed in this paper. Through applying such strategy to realize single granularity attribute reduction, the detailed heuristic algorithm can be designed as shown in the following Algorithm 1.

Note that as one of the key steps in Algorithm 1, the selection mechanism of candidate attributes (see Step 3. (2)) is generally determined by the employed measure. For instance, if the employed measure is positively correlated, then the

Algorithm 1 Computing single granularity reduct.

Input: $\mathcal{DS} =< U, AT \cup \{d\} >$, a granularity $\mathbb{G}$ and $\rho^{\mathbb{G}}$-constraint.
Output: One single granularity reduct A.
Step 1. Calculate the measure-value over the raw attribute set AT over the single granularity $\rho^{\mathbb{G}}(AT)$;
Step 2. Set $A = \emptyset$;
Step 3. Do
(1) Evaluate each candidate attribute $a \in AT - A$ over $\mathbb{G}$ by calculating the corresponding measure-value $\rho^{\mathbb{G}}(A \cup \{a\})$;
(2) Select a qualified attribute $b \in AT - A$ with the satisfactory evaluation;
(3) Set $A = A \cup \{b\}$;
(4) Calculate $\rho^{\mathbb{G}}(A)$;
Until $\rho^{\mathbb{G}}$-constraint is satisfied;
Step 4. Return A.

higher measure-value, the better evaluation, the selected attribute b with the satisfactory evaluation can be regarded as the one which may increase the corresponding measure-value as much as possible such that $\rho^{\mathbb{G}}(A \cup \{b\}) = \max\{\rho^{\mathbb{G}}(A \cup \{a\}) : \forall a \in AT - A\}$; conversely, if the employed measure is negatively correlated, then the lower measure-value, the better evaluation, the selected attribute b with the satisfactory evaluation can be regarded as the one which may decrease the corresponding measure-value as much as possible such that $\rho^{\mathbb{G}}(A \cup \{b\}) = \min\{\rho^{\mathbb{G}}(A \cup \{a\}) : \forall a \in AT - A\}$.

Moreover, the total time complexity of Algorithm 1 is $\mathcal{O}(|U|^2 \times |AT|^2)$ where $|U|$ and $|AT|$ denote the numbers of samples and conditional attributes, respectively.

3. MGAR and MGAS

3.1. MGAR: multi-granularity attribute reduction

In retrospect, most of attribute reductions are single granularity based cases. However, as what has been pointed out in Section 1, single granularity attribute reduction may involve several inherent limitations: (1) it may result in the poor adaptability of the generated reduct to the problem of granularity diversity; (2) it may increase the time consumption if multiple different granularities are required. To solve these problems, Multi-Granularity Attribute Reduction (MGAR) is proposed in the following.

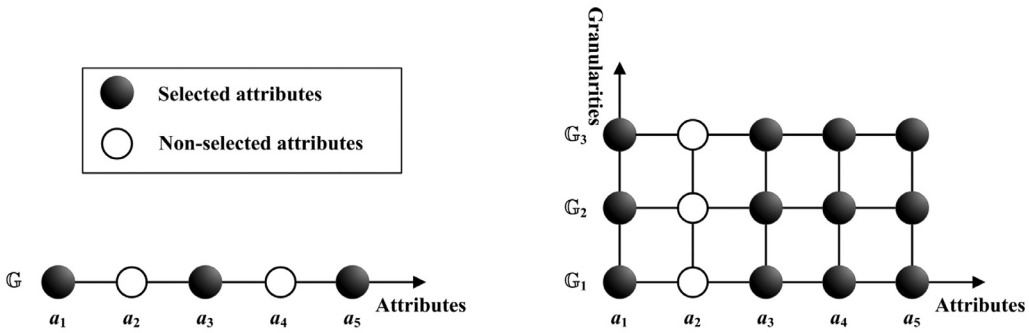
Definition 2. Given a decision system $\mathcal{DS}$ and a set of multiple different granularities such that $\mathbb{MG} = \{\mathbb{G}_1, \mathbb{G}_2, \ldots, \mathbb{G}_n\}$, assuming that $\rho^{\mathbb{MG}}$-constraint is a multi-granularity constraint based on a considered measure ρ, then any attribute subset in AT, A is referred to as a multi-granularity reduct if and only if

(1) A satisfies the $\rho^{\mathbb{MG}}$-constraint;
(2) $\forall B \subset A$, B does not satisfy the $\rho^{\mathbb{MG}}$-constraint.

Obviously, multi-granularity constraint designed in Definition 2 is the most representative portion for distinguishing between MGAR and single granularity attribute reduction defined in Definition 1. Therefore, how to construct a suitable multi-granularity constraint inevitably becomes an open problem in MGAR. In the following Example 3, some typical forms of multi-granularity constraint are given.

Example 3.

- Given a set of multiple different granularities $\mathbb{MG}$, the constraint defined by measure ρ over one granularity $\mathbb{G}_m \in \mathbb{MG}$ is denoted by $\rho^{\mathbb{G}_m}$-constraint, then $\rho^{\mathbb{MG}}$-constraint can be formed as a combination of all the constraints such that {$\rho^{\mathbb{G}_1}$-constraint, $\rho^{\mathbb{G}_2}$-constraint, ..., $\rho^{\mathbb{G}_n}$-constraint}. For such case, the multi-granularity reduct becomes the minimal subset of the conditional attributes which satisfies all of these constraints.
- Given a set of multiple different granularities $\mathbb{MG}$, through evaluating all of the granularities, the granularity with superior performance can be found, which is referred to as the justifiable granularity [23,25] denoted by $\mathbb{G}_{just}$, then $\rho^{\mathbb{MG}}$-constraint can be formed as $\rho^{\mathbb{G}_{just}}$-constraint. For such case, the multi-granularity reduct becomes the minimal subset of the conditional attributes which satisfies the justifiable granularity based constraint.
- Given a set of multiple different granularities $\mathbb{MG}$, with the considered measure ρ, the measure-value of A over one granularity $\mathbb{G}_m$ is denoted by $\rho^{\mathbb{G}_m}(A)$, then multi-granularity measure-value over conditional attribute subset denoted by $\rho^{\mathbb{MG}}(A)$ can be considered as a fusion of measure-values in $\{\rho^{\mathbb{G}_1}(A), \rho^{\mathbb{G}_2}(A), \cdots, \rho^{\mathbb{G}_n}(A)\}$. Correspondingly, $\rho^{\mathbb{MG}}$-constraint can be defined for preserving, increasing or decreasing such fused measure-value. For such case, the multi-granularity reduct becomes the minimal subset of the conditional attributes which satisfies such fused measure-value based constraint.



(a) Single granularity attribute reduction (b) Multi-granularity attribute reduction

Fig. 2. The general mechanisms of single granularity attribute reduction and MGAR.

Through observing these different forms of multi-granularity constraint, it is easily known that MGAR is so different from single granularity attribute reduction as presented in Definition 1. To further analyze the difference, the general mechanism of MGAR will be compared with that of single granularity attribute reduction as illustrated in the following Fig. 2.

Compared with the single granularity attribute reduction shown in Fig. 2 (a), MGAR shown in Fig. 2(b) can be regarded as a two dimensional structure with the consideration of multi-granularity, the *X*-axis indicates the candidate attributes while the *Y*-axis denotes the multiple different granularities. Following Fig. 2, it is not difficult to observe the following.

(1) In Fig. 2(a), only one granularity $\mathbb{G}$ is considered and three attributes a_1, a_3, a_5 are selected, it follows that the generated reduct is $\{a_1, a_3, a_5\}$. Following Definition 1, it is mainly because such single granularity reduct satisfies one intended single granularity constraint over one and only one granularity $\mathbb{G}$.

(2) In Fig. 2(b), three different granularities $\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_3$ are considered and four attributes a_1, a_3, a_4, a_5 are selected, it follows that the generated reduct is $\{a_1, a_3, a_4, a_5\}$. Following Def. 2, it is mainly because such multi-granularity reduct satisfies one intended multi-granularity constraint by considering three different granularities $\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_3$. For example, if the multi-granularity constraint is the first case in Example 3, then such reduct is required to satisfy all three constraints, i.e., $\mathbb{G}_1$ based constraint, $\mathbb{G}_2$ based constraint and $\mathbb{G}_3$ based constraint. Obviously, compared with single granularity attribute reduction, the immediate constraint is stricter. Consequently, to satisfy it, more attributes may be selected.

From discussions above, MGAR delivers a novel way to explore the attribute reduction in the problem of multi-granularity. This is mainly because the multi-granularity constraint for defining MGAR has been modified, it follows that such constraint can adapt successfully to the structure of multiple different granularities.

Note that if one and only one granularity is required, i.e., $n = 1$ in Definition 2, then the multi-granularity constraint becomes a single granularity one.

3.2. MGAS: multi-granularity attribute selector

Following the definition of MGAR, how to realize it becomes a challenging topic. For such purpose, two open problems should be mainly addressed: (1) how to construct the multi-granularity constraint; (2) how to evaluate the candidate attributes. However, to solve such two problems, it is trivial to observe that computational complexity may be increased due to the requirement of multiple different granularities. Therefore, to realize MGAR in a more feasible way, a Multi-Granularity Attribute Selector (MGAS) will be proposed in this section.

Before we delve into the algorithm of MGAS, the construction of multi-granularity constraint and the evaluation of candidate attributes are re-considered as follows.

3.2.1. The construction of multi-granularity constraint

With a critical reviewing of the multi-granularity constraints discussed in Example 3, some of them may involve various limitations. We will elaborate them in the following.

- For the first case in Example 3, the multi-granularity constraint is a combination of all the single granularity constraints in terms of multiple different granularities. It may be too strict and then it is highly possible that no attribute can be deleted from the raw attribute set. In other words, the corresponding attribute reduction may be meaningless and useless.
- For the second case in Example 3, the multi-granularity constraint is a constraint based on the justifiable granularity among all the considered granularities. It actually involves the issue of granularity selection [46,47], and there exist some difficulties. For instance, as reported by Yao [42], to measure the effectiveness of one considered granularity, attribute reduction can be viewed as a desirable method. In other words, to search a justifiable granularity, it is required to compute reducts over all the granularities in advance, such process may be repetitive and cost higher elapsed time especially the number of considered granularities is great.

To avoid the problems in expressing the multi-granularity constraint, the third case in Example 3 (use the fused measure-value to design multi-granularity constraint) is employed in our MGAS since it is superior to the other two cases: (1) compared with the first case in Example 3, such strategy avoids the undesired case discussed above, and reduces the number of constraints in terms of multiple different granularities; (2) compared with the second case in Example 3, such strategy avoids the tedious work of estimating all the considered granularities. Therefore, by applying the fusion of measure-value to the construction of multi-granularity constraint, it is expected to accelerate the searching process of a multi-granularity reduct.

3.2.2. The evaluation of candidate attributes

Different from the one and only one evaluation presented in Algorithm 1, more evaluations are required to select qualified attribute since multiple different granularities have been taken into account. A natural way is to evaluate each candidate attribute over all of the granularities. However, such strategy may result in several limitations as follows.

- It may result in higher elapsed time. For instance, if the number of considered granularities is n, then each candidate attribute will be evaluated by n times repeatedly. Obviously, such process is computationally expensive especially the number of considered granularities is great.
- It may result in redundant work. For instance, if there is no significant difference between two considered granularities, then the evaluations based on such two granularities may be indistinguishable, it follows that one of such two evaluations is unnecessary. Obviously, some of the evaluations are redundant to be pruned.

To address the problems mentioned above, two granularities which may be entirely different are focused in MGAS, i.e., the finest and coarsest granularities. Correspondingly, candidate attributes will be evaluated over such two distinguishable granularities. By substituting the combined evaluation over the finest and coarsest granularities for the evaluations over all the granularities, it is expected that the reduct may adapt to multi-granularity structure as well as possible. Moreover, such strategy may be effective in speeding up the searching process of a multi-granularity reduct.

3.2.3. The heuristic algorithm of MGAS

Through the usage of the re-designed multi-granularity constraint and combined evaluation, the framework of heuristic algorithm is employed to realize MGAS. The detailed algorithm is designed in the following Algorithm 2.

Algorithm 2 Multi-Granularity Attribute Selector (MGAS).

Input: $\mathcal{DS} =< U, AT \cup \{d\} >$, set of considered granularities $\mathbb{MG} = \{\mathbb{G}_1, \mathbb{G}_2, \ldots, \mathbb{G}_n\}$ and $\rho^{\mathbb{MG}}$-constraint.
Output: One multi-granularity reduct A.
Step 1. Calculate the measure-values of the raw attribute set AT over all the considered granularities $\rho^{\mathbb{G}_1}(AT)$, $\rho^{\mathbb{G}_2}(AT)$, $\cdots$, $\rho^{\mathbb{G}_n}(AT)$;
Step 2. Calculate the fused measure-value $\rho^{\mathbb{MG}}(AT)$ based on the measure-values derived in Step 1;
`//Such fusion can be realized by using ''arithmetic mean'', ''harmonic mean'', etc.`
Step 3. Set $A = \emptyset$;
Step 4. Do
(1) Evaluate each candidate attribute $a \in AT - A$ over the finest and coarsest granularities $\mathbb{G}_f$ and $\mathbb{G}_c$ by calculating the corresponding measure-value $\rho^{\mathbb{G}_{f\&c}}(A \cup \{a\}) = \rho^{\mathbb{G}_f}(A \cup \{a\}) + \rho^{\mathbb{G}_c}(A \cup \{a\})$;
(2) Select a qualified attribute $b \in AT - A$ with the satisfactory evaluation;
(3) Set $A = A \cup \{b\}$;
(4) Calculate $\rho^{\mathbb{MG}}(A)$;
Until $\rho^{\mathbb{MG}}$-constraint is satisfied;
Step 5. Return A.

Following Algorithm 2, two key steps should be paid much attention as follows.

- The first key step is the calculation of the fused measure-value (see Step 2). Such fusion can be realized by various operations. For example, if the arithmetic mean is employed, then the fusion is obtained by $\frac{1}{n}\Sigma_{m=1}^{n}\rho^{\mathbb{G}_m}(A)$; if the harmonic mean is employed, then the fusion is derived by $\frac{1}{n}\Sigma_{m=1}^{n}(\frac{1}{\rho^{\mathbb{G}_m}(A)})^{-1}$. In summary, no matter which operation we adopt, the major purpose of such step is to reduce the complexity of constructing the multi-granularity constraint.
- The other key step is the organization of the selection mechanism (see Step 4). To evaluate candidate attributes and then select a qualified one in each iteration, the finest and coarsest granularities are concerned instead of all the granularities. Moreover, different from Algorithm 1, the immediate selection mechanism of candidate attributes is realized by comparing the sum of measure-values over such two granularities. Similarly, if the employed measure is positively correlated, the maximal sum indicates the satisfactory evaluation, it follows that attribute which leads to such maximal value will be selected; conversely, the minimum sum indicates the satisfactory evaluation, it follows that attribute which leads to such minimum value will be selected. In summary, no matter which evaluation we adopt, the major purpose of such step is to reduce the number of the evaluations of candidate attributes.

Table 1
Characteristics of the experimental data sets.

ID	Data sets	Samples	Attributes	Decision classes
1	Breast tissue	106	10	6
2	Cardiotocography	2126	23	10
3	Congressional voting	435	16	2
4	Crowdsourced mapping	10,845	28	6
5	Dermatology	366	33	6
6	Diabetic retinopathy debrecen	1151	20	2
7	Fertility	100	10	2
8	Gesture phase segmentation	9900	19	5
9	Libras movement	360	91	15
10	MAGIC gamma telescope	19,020	11	2
11	Page blocks classification	5473	10	5
12	Parkinson multiple sound recording	1040	26	2
13	QSAR biodegradation	1055	41	2
14	Statlog (Landsat Satellite)	6435	36	6
15	Statlog (Image Segmentation)	2310	19	7
16	Vertebral column	310	6	2
17	Wall-Following robot vavigation	5456	24	2
18	Waveform database generator	5000	20	3
19	Wine quality	4898	12	7
20	Wireless indoor localization	2000	7	4

Moreover, for Steps 1 and 2 in Algorithm 2, the time complexity is $\mathcal{O}(|U|^2 \times |AT| \times n)$ because n different granularities has been considered; for Step 4, the time complexity is $\mathcal{O}(|U|^2 \times |AT|^2)$. And then the total time complexity is $\mathcal{O}(|U|^2 \times |AT| \times n + |U|^2 \times |AT|^2)$. Note that if $|AT| > n$, then the time complexity of Algorithm 2 is $\mathcal{O}(|U|^2 \times |AT|^2)$. Immediately, the time complexity of Algorithm 2 is the same to that of Algorithm 1. Such result suggests that in view of time complexity, although more different granularities have been taken into account, our proposed algorithm may not increase the elapsed time of finding reduct.

4. Experiments

4.1. Data sets

To demonstrate the effectiveness of the proposed MGAS (Algorithm 2), 20 real-world data sets from UCI Machine Learning Repository have been employed in the context of this paper. Table 1 summarizes some detailed statistics of those data sets used in our experiments. Note that all of the data sets are used for classification task with numerical attribute values, and they have been normalized by column.

4.2. Experimental setup and configuration

All the experiments have been carried out on a personal computer with Windows 10, Intel Core 2 Duo T5800 CPU (2.60 GHz) and 16.00 GB memory. The programming language is Matlab R2014a.

In the following experiments, neighborhood rough set [6] is used to realize MGAS because neighborhood rough set can naturally form a multi-granularity structure if multiple different radii are considered. Correspondingly, the set of considered granularities $\mathbb{MG}$ in Algorithm 2 can be intuitively replaced by a set of radii $\mathbb{R}$. Note that we appoint a set of ascending ordered radii including 10 different radii such that $\mathbb{R} = \{\delta_1 = 0.03, \delta_2 = 0.06, \ldots, \delta_{10} = 0.3\}$. It should be noticed that: (1) to obtain the fused measure-value, the harmonic mean is employed, because it is not only a simple and quick way to realize the fusion, but also can mitigate the impact of greater measure-values and aggravate the impact of smaller ones; (2) to evaluate the candidate attributes, considering that a greater radius derives a finer granularity while a smaller radius derives a coarser granularity [47], the smallest and greatest radii δ_1 and δ_{10} are employed to reflect the finest and coarsest granularities $\mathbb{G}_f$ and $\mathbb{G}_c$ as shown in Algorithm 2.

The 10-fold cross-validation is used in our experiments to test the performances of different reducts. In other words, the set of raw data is randomly partitioned into 10 equal sized groups, the 9 groups compose the training samples for computing reducts and the rest of the 1 group is regarded as the testing samples to test the classification performance of the derived reduct. It should be noticed that with such 10 appointed radii, 10 different reducts will be obtained by Algorithm 1 while only 1 reduct will be obtained by MGAS. Therefore, after 10-fold cross-validation, 100 reducts will be derived by Algorithm 1 while 10 reducts will be obtained by MGAS. The averages of such two types of results are mainly compared.

Moreover, four widely used measures in neighborhood rough set are employed. The detailed formulas of these measures are presented in the following.

- **Neighborhood Decision Error Rate (NDER)** [6]: evaluates the classification performance of conditional attribute subset with respect to Neighborhood Classifier (NEC) [6]. Given a radius $\delta_m \in \mathbb{R}$, the corresponding measure-value of NDER is:

$$\mathrm{NDER}^{\delta_m}(A) = \frac{|\{x \in U : \mathrm{Pre}_A^{\delta_m}(x) \neq d(x)\}|}{|U|}, \tag{1}$$

where $\mathrm{Pre}_A^{\delta_m}(x)$ denotes the predicted label of sample x related to A in terms of δ_m by NEC, and $|\cdot|$ is the cardinality of a set.
The lower the $\mathrm{NDER}^{\delta_m}(A)$, the better the performance of conditional attribute subset A. From this point of view, the constraint is set to be "$\mathrm{NDER}^{\delta_m}(A) \leq \mathrm{NDER}^{\delta_m}(AT)$" in Algorithm 1 for deriving Neighborhood Decision Error Rate Reduct (NDERR).
Correspondingly, if the measure of NDER is introduced into Algorithm 2, then the fused measure-value is:

$$\mathrm{NDER}^{\mathbb{R}}(A) = \frac{1}{|\mathbb{R}|}\Big(\sum_{\delta_m \in \mathbb{R}} \frac{1}{\mathrm{NDER}^{\delta_m}(A)}\Big)^{-1}. \tag{2}$$

Similarly, the constraint is set to be "$\mathrm{NDER}^{\mathbb{R}}(A) \leq \mathrm{NDER}^{\mathbb{R}}(AT)$" in Algorithm 2 for deriving Multi-Granularity Neighborhood Decision Error Rate Reduct (MG-NDERR).

- **Conditional Discrimination Index (CDI)** [29]: represents the ability of conditional attribute subset to distinguish samples with different decision attribute values. Given a radius $\delta_m \in \mathbb{R}$, the corresponding measure-value of CDI is:

$$\mathrm{CDI}^{\delta_m}(A) = \log \frac{|\mathrm{N}_A^{\delta_m}|}{|\mathrm{N}_A^{\delta_m} \cap \mathrm{IND}_{\{d\}}|}, \tag{3}$$

where $\mathrm{IND}_{\{d\}} = \{(x, y) \in U^2 : d(x) = d(y)\}$ is an equivalence relation over d.
The lower the $\mathrm{CDI}^{\delta_m}(A)$, the better the performance of conditional attribute subset A. From this point of view, the constraint is set to be "$\mathrm{CDI}^{\delta_m}(A) \leq \mathrm{CDI}^{\delta_m}(AT)$" in Algorithm 1 for deriving Conditional Discrimination Index Reduct (CDIR).
Correspondingly, if the measure of CDI is introduced into Algorithm 2, then the fused measure-value is:

$$\mathrm{CDI}^{\mathbb{R}}(A) = \frac{1}{|\mathbb{R}|}\Big(\sum_{\delta_m \in \mathbb{R}} \frac{1}{\mathrm{CDI}^{\delta_m}(A)}\Big)^{-1}. \tag{4}$$

Similarly, the constraint is set to be "$\mathrm{CDI}^{\mathbb{R}}(A) \leq \mathrm{CDI}^{\mathbb{R}}(AT)$" in Algorithm 2 for deriving Multi-Granularity Conditional Discrimination Index Reduct (MG-CDIR).

- **Conditional Entropy (CE)** [45]: measures the discriminating ability of conditional attribute subset for different decision classes. Given a radius $\delta_m \in \mathbb{R}$, the corresponding measure-value of CE is:

$$\mathrm{CE}^{\delta_m}(A) = -\frac{1}{|U|}\sum_{x \in U} |\mathrm{N}_A^{\delta_m}(x) \cap [x]_d| \log \frac{|\mathrm{N}_A^{\delta_m}(x) \cap [x]_d|}{|\mathrm{N}_A^{\delta_m}(x)|}, \tag{5}$$

where $\mathrm{N}_A^{\delta_m}(x) = \{y \in U : \triangle_A(x, y) \leq \delta_m\}$ is the neighborhood of sample x related to A in terms of δ_m and $[x]_d = \{y \in U : d(y) = d(x)\}$ denotes the decision class with the same label including x.
The lower the $\mathrm{CE}^{\delta_m}(A)$, the better the performance of conditional attribute subset A. From this point of view, the constraint is set to be "$\mathrm{CE}^{\delta_m}(A) \leq \mathrm{CE}^{\delta_m}(AT)$" in Algorithm 1 for deriving Conditional Entropy Reduct (CER).
Correspondingly, if the measure CE is introduced into Algorithm 2, then the fused measure-value is:

$$\mathrm{CE}^{\mathbb{R}}(A) = \frac{1}{|\mathbb{R}|}\Big(\sum_{\delta_m \in \mathbb{R}} \frac{1}{\mathrm{CE}^{\delta_m}(A)}\Big)^{-1}. \tag{6}$$

Similarly, the constraint is set to be "$\mathrm{CE}^{\mathbb{R}}(A) \leq \mathrm{CE}^{\mathbb{R}}(AT)$" in Algorithm 2 for deriving Multi-Granularity Conditional Entropy Reduct (MG-CER).

- **Approximation Quality (AQ)** [21]: reflects the approximation ability of granule space induced by conditional attribute subset to characterize the decision attribute. Given a radius $\delta_m \in \mathbb{R}$, the corresponding measure-value of AQ is:

$$\mathrm{AQ}^{\delta_m}(A) = \frac{|\{x \in U : \mathrm{N}_A^{\delta_m}(x) \subseteq [x]_d\}|}{|U|}. \tag{7}$$

The higher the $\mathrm{AQ}^{\delta_m}(A)$, the better the performance of conditional attribute subset A. From this point of view, the constraint is set to be "$\mathrm{AQ}^{\delta_m}(A) \geq \mathrm{AQ}^{\delta_m}(AT)$" in Algorithm 1 for deriving Approximation Quality Reduct (AQR).
Correspondingly, if the measure of AQ is introduced into Algorithm 2, then the fused measure-value is

$$\mathrm{AQ}^{\mathbb{R}}(A) = \frac{1}{|\mathbb{R}|}\Big(\sum_{\delta_m \in \mathbb{R}} \frac{1}{\mathrm{AQ}^{\delta_m}(A)}\Big)^{-1}. \tag{8}$$

Similarly, the constraint is set to be "$\mathrm{AQ}^{\mathbb{R}}(A) \geq \mathrm{AQ}^{\mathbb{R}}(AT)$" in Algorithm 2 for deriving Multi-Granularity Approximation Quality Reduct (MG-AQR).

Table 2
Comparisons among lengths of the derived reducts (greater values are in bold).

ID	MG-NDERR	NDERR	MG-CDIR	CDIR	MG-CER	CER	MG-AQR	AQR
1	**2.0**	1.7	**2.0**	1.9	**2.2**	2	2.6	**2.6**
2	**5.6**	4.8	**6.2**	4.9	**7.4**	6.3	14.6	**15.9**
3	1.0	1.0	1.0	**1.0**	**2.0**	1.7	4.6	**5.2**
4	18.6	**19.4**	**4.2**	2.9	**19.4**	15.8	**25.0**	23.9
5	**10.8**	8.4	**6.6**	6.2	**7.2**	6.9	**16.8**	13.7
6	**2.0**	1.8	**3.0**	2.2	**2.6**	1.9	6.2	**10.3**
7	1.0	1.0	**6.6**	5.8	**7.0**	6.9	**7.0**	6.4
8	**4.4**	3.3	**4.0**	3.3	**5.0**	3.7	4.0	**7.5**
9	**10.6**	7.3	**9.2**	7.9	**10.2**	8.3	79.6	**80.3**
10	**3.2**	3.1	2.0	**2.1**	**5.0**	3.9	**9.0**	8.9
11	2.0	**2.5**	**3.2**	2.23	**4.0**	3.0	6.0	**6.6**
12	**7.2**	4.5	**2.0**	1.4	**5.8**	4.4	**18.0**	8.8
13	**7.0**	4.8	**2.2**	1.7	**6.4**	4.7	**31.2**	19.8
14	4.2	**5.4**	**7.8**	3.6	**8.6**	6.8	25.4	**26.3**
15	**4.8**	4.6	**4.0**	3.1	**5.4**	4.6	**18.0**	10.4
16	**4.6**	2.4	**2.6**	2.0	**3.0**	2.2	**5.2**	4.1
17	**23.6**	16.4	**23.8**	12.8	**24.0**	21.2	19.5	**19.7**
18	**12.5**	11.6	**8.2**	7.7	**12.1**	10.6	11.6	**14.0**
19	1.0	1.0	**3.4**	2.2	**6.0**	4.3	**10.0**	9.7
20	3.4	**3.9**	**4.2**	3.9	**7.0**	5.2	**6.0**	5.8

Table 3
Comparisons among elapsed time of obtaining the reducts (lower values are underlined).

ID	MG-NDERR	NDERR	MG-CDIR	CDIR	MG-CER	CER	MG-AQR	AQR
1	0.0437	0.1237	0.0360	0.1130	0.0834	0.2523	0.1017	0.3274
2	203.7074	674.3429	226.6137	680.9788	262.5895	903.9458	481.4911	1846.2440
3	0.3343	1.1238	0.3554	1.2504	0.7299	2.3158	1.6039	6.4719
4	2401.4142	10028.5679	402.9895	1142.4111	1469.3281	5288.2749	3147.6339	12237.7919
5	4.6687	17.6740	2.6125	11.6872	3.9290	17.3678	9.6175	36.2576
6	9.1443	36.0467	16.8256	43.8223	14.3409	42.4605	39.3834	227.2622
7	0.0200	0.0621	0.1073	0.2670	0.1399	0.4241	0.1262	0.4053
8	388.6939	1140.6194	199.0829	633.0064	263.7292	782.8295	330.6200	1898.5733
9	39.5482	132.6300	28.7590	119.1528	41.0761	160.3771	251.3016	1145.8268
10	669.1836	2074.4523	330.4471	1133.0415	772.7762	2114.0008	1395.6269	4469.4901
11	296.6381	1187.3030	508.0678	962.7241	636.3635	1386.3777	1019.4279	3475.1671
12	40.5201	107.6017	13.3698	34.9316	32.0365	112.2134	75.9286	187.6932
13	57.4567	210.4849	19.5454	66.8204	51.9015	187.5545	148.6019	527.4090
14	291.9864	1525.1897	290.1098	612.0272	1200.879	339.5559	1392.4134	5693.7849
15	204.3822	752.9989	165.3931	474.5692	225.8572	745.1726	606.2785	1445.5704
16	0.3948	0.5980	0.2557	0.5209	0.3267	0.7043	0.4943	1.1146
17	1829.8070	5608.2429	1843.2395	5198.7613	2790.7101	9992.1131	2440.9476	9582.7002
18	926.4208	3444.7131	722.7882	2742.1751	1357.0116	4873.4644	1230.2517	5415.1827
19	107.9122	238.8074	38.8363	86.5996	61.79256	155.1373	175.4844	580.4078
20	45.7066	136.5912	51.4159	146.0533	64.1834	167.8975	65.3743	184.6895

With these measures, 4 groups of multi-granularity reducts and single granularity reducts will be compared, i.e., MG-NDERR vs. NDERR, MG-CDIR vs. CDIR, MG-CER vs. CER and MG-AQR vs. AQR. Moreover, 3 types of comparisons will be made, mainly including: (1) the lengths of the reducts; (2) the elapsed time of obtaining the reducts; (3) the classification accuracies derived by reducts over the testing samples.

4.3. Comparisons among lengths of the derived reducts

In this section, the lengths of the derived reducts will be compared. The following Table 2 shows us the corresponding result.

Following Table 2, it is not difficult to observe that in most cases, the lengths of multi-granularity reducts derived from MGAS are greater than single granularity reducts derived from Algorithm 1. From this point of view, it may be concluded that to satisfy the multi-granularity constraint, MGAS intends to select more attributes. It is mainly because the multi-granularity constraint constructed in MGAS may be stricter so that more attributes are required to satisfy it.

4.4. Comparisons among elapsed time of obtaining the reducts

In this section, the elapsed time of obtaining reducts will be compared. The following Table 3 shows us the corresponding result.

Table 4
Comparison among SVM classification accuracies (greater values are in bold).

ID	MG-NDERR	NDERR	MG-CDIR	CDIR	MG-CER	CER	MG-AQR	AQR
1	**0.5563**	0.5033	0.5481	**0.5560**	**0.5576**	0.5475	0.4896	**0.5063**
2	**0.7441**	0.7255	**0.7606**	0.7226	0.7545	**0.7563**	0.7474	**0.7549**
3	0.9425	0.9425	0.9425	0.9425	**0.9448**	0.9430	**0.9425**	0.9395
4	**0.8883**	0.8877	**0.8256**	0.8128	**0.8867**	0.8787	**0.8957**	0.8928
5	**0.9645**	0.9473	**0.9509**	0.8515	**0.9563**	0.9443	0.9371	**0.9387**
6	0.6073	0.6073	**0.6160**	0.6127	**0.6151**	0.6082	**0.6168**	0.6073
7	0.8800	0.8800	0.8800	0.8800	**0.8800**	0.8690	0.8800	0.8800
8	**0.4932**	0.4915	**0.5044**	0.5019	**0.5001**	0.4982	0.4938	**0.5089**
9	**0.4694**	0.4333	**0.4611**	0.4525	0.4528	**0.4639**	0.4556	**0.4650**
10	**0.8134**	0.8125	0.7603	**0.7617**	**0.8338**	0.8190	0.8187	**0.8247**
11	**0.9361**	0.9267	**0.9351**	0.9206	**0.9331**	0.9222	**0.9377**	0.9376
12	**0.6449**	0.6344	**0.6167**	0.6076	**0.6531**	0.6439	**0.6258**	0.6194
13	**0.7744**	0.7734	**0.7659**	0.7291	**0.7972**	0.7938	**0.8038**	0.7963
14	**0.8395**	0.8281	**0.8347**	0.818	**0.8476**	0.8436	**0.8488**	0.8477
15	**0.9039**	0.8986	**0.8792**	0.8394	0.8797	**0.8905**	**0.8827**	0.8787
16	**0.7581**	0.7271	**0.7419**	0.7213	0.7355	**0.7361**	**0.7226**	0.7094
17	0.7693	**0.7700**	**0.7647**	0.7503	**0.7647**	0.7629	**0.7614**	0.7576
18	**0.8560**	0.8464	0.8112	**0.8138**	**0.8500**	0.8397	0.8336	**0.8528**
19	**0.4488**	0.4486	0.5020	**0.5053**	0.5206	**0.5174**	0.5188	**0.5285**
20	**0.9795**	0.9792	**0.9795**	0.9732	**0.9780**	0.9735	**0.9765**	0.9748

With 10 appointed radii, Algorithm 1 is repeated 10 times to obtain 10 single granularity reducts. However, MGAS is performed only 1 time for deriving 1 multi-granularity reduct. Consequently, as illustrated in Table 3, the elapsed time of deriving reducts by MGAS are significantly lower than that by Algorithm 1.

It should be noticed that although the number of using MGAS is only 10% of that of using Algorithm 1, the elapsed time of deriving multi-granularity reducts are still higher than 10% of that of deriving single granularity reducts. Take the result on data set "Breast Tissue (ID: 1)" as an example, deriving one MG-NDERR costs 0.0437 s while deriving ten NDERRs costs 0.1237 s, and 0.0437 is significantly greater than 0.01237 ($0.1237 \times 10\%$). The reason for such result may be attributed to two aspects: (1) as we can observe in Table 2, more attributes are selected by MGAS, it follows that the elapsed time of searching process is increased; (2) as shown in MGAS, each candidate attribute is evaluated over two granularities (the finest and coarsest granularities) instead of one and only one granularity in Algorithm 1, such process also costs more time consumption.

4.5. Comparisons among classification accuracies of the derived reducts

In this section, the classification performances of reducts will be compared. Note that SVM (LIBSVM [1]), CART and NEC [6] are employed to evaluate such performances.

4.5.1. Classification accuracies based on SVM

The comparisons among classification accuracies based on SVM and corresponding analysis will be presented. Table 4 shows the detailed result of SVM classification accuracies.

By Table 4, it is easily observed that in most cases, the SVM classification accuracies over testing data derived by multi-granularity reducts are higher.

Moreover, to further analyze the results of classification accuracies from the viewpoint of statistics, One-way Analysis of Variance [5] is used to test whether the average classification accuracies are all the same. If the average classification accuracies are similar, then the returned p-value is higher than or equal to the 5% significance level (0.05); conversely, the returned p-value is lower than 0.05. The statistical comparisons are shown in Table 5.

Following Table 5, we can observe two main cases: (1) for p-values ≥ 0.05, the SVM classification accuracies derived by multi-granularity reducts from MGAS are similar to those from Algorithm 1; (2) for p-values ≤ 0.05, the SVM classification accuracies derived by multi-granularity reducts from MGAS are significantly different from those from Algorithm 1, and in most cases, the SVM classification accuracies derived by multi-granularity reducts are significantly higher.

In view of such observation, it is easily concluded that the multi-granularity reducts generated by MGAS may offer the similar or even better SVM based classification performances.

4.5.2. Classification accuracies based on CART

In the following, the results of average classification accuracies based on CART are shown in Table 6.

By Table 6, it is easily observed that in most cases, the classification accuracies derived by multi-granularity reducts are higher.

Similarly, One-way Analysis of Variance [5] is used to analyze the results of such classification accuracies from the viewpoint of statistics. The detailed statistical analysis is shown in the following Table 7.

Table 5
Statistical analysis among SVM classification accuracies. In addition, •/◦ indicates whether the results by MGAS is significantly superior/inferior to those by compared methods (the significance level is 0.05).

ID	MG-NDERR &NDERR	MG-CDIR &CDIR	MG-CER &CER	MG-AQR &AQR
1	0.0090•	0.2723	0.2359	0.0274
2	0.0004•	0.0001•	0.4674	0.0136•
3	1.0000	1.0000	0.0008•	0.2244
4	0.6357	0.0695	0.0223•	0.0190•
5	0.0000•	0.0001•	0.1131	0.6697
6	0.9948	0.2537	0.0054•	0.0043•
7	1.0000	1.0000	0.0510	1.0000
8	0.8322	0.4405	0.6627	0.2010
9	0.0002•	0.1051	0.1141	0.0025◦
10	0.7506	0.6142	0.0006•	0.0000◦
11	0.0647	0.0025•	0.0053•	0.8276
12	0.0396•	0.0379•	0.0141•	0.2501
13	0.9083	0.0003•	0.4323	0.0704
14	0.1159	0.0063•	0.2193	0.5547
15	0.0617	0.0001•	0.0067◦	0.0558
16	0.0105•	0.1261	0.9428	0.0111•
17	0.2131	0.0115•	0.0361•	0.1052
18	0.0235•	0.5499	0.0181•	0.0001•
19	0.0372•	0.1413	0.2175	0.0078
20	0.6680	0.0579	0.0758	0.0672

Table 6
Comparisons among CART classification accuracies (greater values are in bold).

ID	MG-NDERR	NDERR	MG-CDIR	CDIR	MG-CER	CER	MG-AQR	AQR
1	**0.5948**	0.5619	**0.6043**	0.5894	**0.6043**	0.5892	**0.5567**	0.5473
2	**0.7756**	0.7465	**0.7747**	0.7210	**0.7916**	0.7773	0.7926	**0.7992**
3	0.9563	0.9563	**0.9563**	0.9561	**0.9563**	0.9533	**0.9540**	0.9522
4	0.8762	**0.8781**	**0.7997**	0.7696	**0.8843**	0.8777	0.8769	**0.8772**
5	**0.9426**	0.9358	**0.9398**	0.8687	**0.9509**	0.9421	0.9208	**0.9273**
6	0.6012	**0.6137**	0.5769	**0.6001**	**0.6464**	0.6167	**0.6724**	0.6177
7	0.8800	0.8800	0.7800	**0.8150**	0.7900	**0.7940**	0.8000	**0.8090**
8	**0.8464**	0.8020	0.7823	**0.8129**	0.7945	**0.8130**	0.7795	0.8061
9	**0.5639**	0.5411	0.5528	**0.5649**	**0.5778**	0.5636	**0.5694**	0.5558
10	**0.7759**	0.7752	0.6965	**0.7066**	**0.8101**	0.7856	**0.8195**	0.8187
11	**0.9463**	0.9388	**0.9596**	0.9302	**0.9594**	0.9341	**0.9657**	0.9643
12	**0.6325**	0.6262	**0.5753**	0.5748	**0.6341**	0.6210	**0.6374**	0.6249
13	**0.8019**	0.7651	**0.7536**	0.7244	**0.8095**	0.7955	0.8000	**0.8023**
14	**0.8457**	0.8289	**0.8466**	0.8091	**0.8511**	0.8459	**0.8527**	0.8513
15	0.9390	**0.9465**	**0.9506**	0.8965	**0.9654**	0.9506	**0.9584**	0.9573
16	**0.8065**	0.7332	**0.7677**	0.6968	**0.7839**	0.7335	**0.7774**	0.7490
17	**0.9923**	0.9812	**0.9923**	0.9629	**0.9923**	0.9892	**0.9846**	0.9834
18	**0.7568**	0.7504	**0.7468**	0.7376	**0.7588**	0.7506	0.7384	**0.7447**
19	0.4439	**0.4467**	**0.5331**	0.5077	**0.5490**	0.5389	0.5666	**0.5697**
20	**0.9735**	0.9723	**0.9735**	0.9645	**0.9700**	0.9650	**0.9710**	0.9682

Following Table 7, it is not difficult to observe that in most cases, MGAS offers us the reducts with the similar or even higher classification accuracies based on CART.

4.5.3. Classification accuracies based on NEC

Different from the results of classification accuracies based on SVM and CART, with 10 appointed radii, 10 classification accuracies based on NEC can be derived by 1 reduct. From this point of view, to compared the results of NEC based classification accuracies clearly, Box Plot [19], a statistical chart illustrating dispersion of data sets, is employed. The corresponding results based on the Box Plots are shown in the following Fig. 3.

With a thorough investigation of Fig. 3, it is not difficult to observe that compared with single granularity reducts, the derived classification accuracies by multi-granularity reducts may not be lower. Take the result of data set "Breast Tissue (ID: 1)" as an example, the median values of derived classification accuracies by multi-granularity reducts (the central mark highlighted by red line in each Box Plot) are similar or even higher than those by single granularity reducts.

Moreover, to compare such results from a statistical point of view, Wilcoxon Rank Sum Test [30] is employed to test whether the distributions of the classification accuracies are consistent. With respect to Wilcoxon Rank Sum Test, the re-

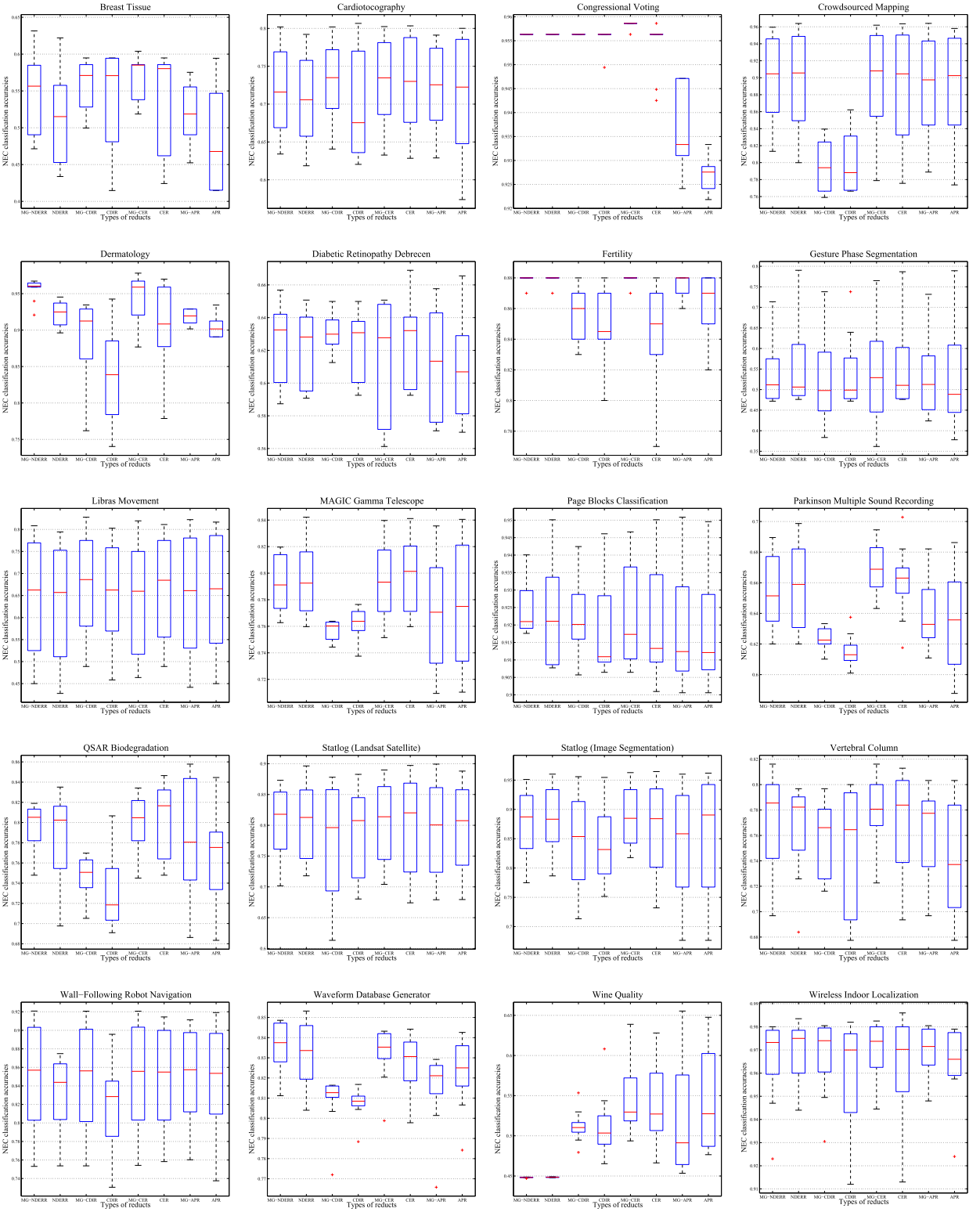


Fig. 3. Comparisons among NEC classification accuracies.

Table 7
Statistical analysis among CART classification accuracies. In addition, •/◦ indicates whether the results by MGAS is significantly superior/inferior to those by compared methods (the significance level is 0.05).

ID	MG-NDERR &NDERR	MG-CDIR &CDIR	MG-CER &CER	MG-AQR &AQR
1	0.1208	0.1916	0.0453•	0.5199
2	0.0047•	0.0006•	0.0003•	0.0430◦
3	1.0000	0.3306	0.1636	0.3539
4	0.0644	0.0078•	0.0120•	0.5860
5	0.1063	0.0001•	0.2341	0.0494◦
6	0.0541	0.0000◦	0.0000•	0.0000•
7	1.0000	0.0000◦	0.3611	0.1539
8	0.0346•	0.0554	0.1622	0.0623
9	0.0079•	0.0500◦	0.1306	0.0372•
10	0.8962	0.0681	0.0006•	0.7232
11	0.3375	0.0022•	0.0024•	0.0357•
12	0.5562	0.9151	0.0669	0.0228
13	0.0219•	0.0030•	0.1023	0.3915
14	0.0300•	0.0002•	0.2373	0.4908
15	0.1155	0.0000•	0.0087•	0.4071
16	0.0000•	0.0047•	0.0003•	0.0739
17	0.0112•	0.0000•	0.0330•	0.7079
18	0.0014•	0.0005•	0.0003•	0.0113◦
19	0.0832	0.0014•	0.1732	0.2826
20	0.0015•	0.0499•	0.1507	0.0343•

Table 8
Statistical analysis among NEC classification accuracies. In addition, •/◦ indicates whether the results by MGAS is significantly superior/inferior to those by compared methods (the significance level is 0.05).

ID	MG-NDERR &NDERR	MG-CDIR &CDIR	MG-CER &CER	MG-AQR &AQR
1	0.3254	0.8495	0.5449	0.2894
2	0.6776	0.2413	0.9698	0.7913
3	1.0000	0.3681	0.0364	0.0095•
4	0.9698	0.7054	0.9698	1.0000
5	0.0022•	0.1618	0.7334	0.0539
6	0.5708	0.4494	0.4274	0.6232
7	1.0000	0.7595	0.6438	0.2627
8	0.8501	0.5966	0.9097	0.7912
9	0.5708	0.6776	0.5966	0.8501
10	0.8044	0.2281	0.7570	0.7738
11	0.5966	0.4961	0.6500	0.8501
12	0.6776	0.0452•	0.3447	0.7913
13	0.6776	0.2121	0.7619	0.5205
14	0.9698	0.8501	0.9698	0.9097
15	0.9698	0.8501	0.6232	0.7501
16	0.5450	0.9698	0.8204	0.2712
17	0.4727	0.2730	0.9698	0.8501
18	0.6228	0.1508	0.4495	0.3847
19	0.2146	0.9539	0.7762	0.4928
20	0.7911	0.5705	0.7333	0.4725

turned p-value which is smaller than the default 5% significance level (0.05) indicates the rejection of the null hypothesis, i.e., the distributions of the classification accuracies are significantly different; conversely, the p-value higher than 0.05 indicates that there is not enough evidence to reject the null hypothesis and then we can conclude that the distributions of the classification accuracies are similar. The detailed statistical results are presented in the following Table 8.

Following Table 8, it is not difficult to observe that in most cases, the returned p-values are higher than 0.05, it may imply that NEC based classification accuracies derived by multi-granularity reducts are similar to those by single granularity reducts. Such observation is also consistent with the representation of Fig. 3. Take the result of the data set "Libras Movement (ID: 9)" as an example, in Fig. 3, the Box Plots formed by NEC classification accuracies related to multi-granularity reducts are similar to those derived from single granularity reducts, and the corresponding p-values shown in Table 8 are 0.5708, 0.6776, 0.5966 and 0.8501 all of which are higher than 0.05.

To sum up, following all of the experimental results, it is not difficult to conclude that if the problem of multi-granularity is explored in attribute reduction, MGAS can not only generate reducts which may not contribute to a poorer classification performance, but also significantly reduce the elapsed time of computing reducts.

5. Conclusions and future perspectives

Different from previous researches, attribute reduction is explored with the consideration of multi-granularity in this paper. In view of such novel thinking, Multi-Granularity Attribute Reduction (MGAR) is defined, and then to realize MGAR feasibly, an efficient selector referred to as Multi-Granularity Attribute Selector (MGAS) is designed. Based on the neighborhood rough set which can naturally form multi-granularity structure by different radii, the experimental results over 20 UCI data sets demonstrate that our MGAS can offer the reducts which may not lead to poor classification performances and reduce the elapsed time of searching reducts simultaneously. The following topics deserve our further investigations.

(1) We have only employed neighborhood rough set to realize our MGAS, some other rough set models such as fuzzy rough set will be further applied to our framework.
(2) Since the harmonic mean is one of the expressions of average index to derive the fused measure in MGAS, some other operations will be further explored to realize the fusion.
(3) Incremental calculation will be applied to our MGAS to further accelerate the searching process of multi-granularity reduct.

Declaration of interest statement

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Acknowledgments

The authors would like to thank the editors and anonymous reviewers for their constructive comments. This work is supported by the Natural Science Foundation of China (No. 61572242), the Postgraduate Research & Practice Innovation Program of Jiangsu Province (No. KYCX19_1715) and the Key Laboratory of Data Science and Intelligence Application, Fujian Province University (No. D1901).

References

[1] C. Chang, C. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst.Technol. 2 (2011) 1–27.
[2] D. Chen, Y. Yang, Attribute reduction for heterogeneous data based on the combination of classical and fuzzy rough set models, IEEE Trans. Fuzzy Syst. 22 (2014) 1325–1334.
[3] J. Dai, H. Hu, W. Wu, Y. Qian, D. Huang, Maximal-discernibility-pair-based approach to attribute reduction in fuzzy rough sets, IEEE Trans. Fuzzy Syst. 26 (2018) 2174–2187.
[4] J. Dai, Q. Hu, J. Zhang, H. Hu, N. Zheng, Attribute selection for partially labeled categorical data by rough set approach, IEEE Trans. Cybern. 47 (2017) 2460–2471.
[5] R. Fisher, On the “probable error” of a coefficient of correlation, Metron 1 (1921) 3–32.
[6] Q. Hu, D. Yu, Z. Xie, Neighborhood classifiers, Expert Syst. Appl. 34 (2008) 866–876.
[7] Q. Hu, L. Zhang, D. Chen, W. Pedrycz, D. Yu, Gaussian kernel based fuzzy rough sets: model, uncertainty measures and applications, Int. J. Approx. Reason. 51 (2010) 453–471.
[8] G. Jiang, W. Wang, Markov cross-validation for time series model evaluations, Inf. Sci. 375 (2017) 219–233.
[9] H. Ju, X. Yang, H. Yu, T. Li, D. Yu, J. Yang, Cost-sensitive rough set approach, Inf. Sci. 355–356 (2016) 282–298.
[10] F. Li, Y. Qian, J. Wang, C. Dang, L. Jing, Clustering ensemble based on sample's stability, Artif. Intell. 273 (2019) 37–55.
[11] J. Li, C. Huang, J. Qi, Y. Qian, W. Liu, Three-way cognitive concept learning via multi-granularity, Inf. Sci. 378 (2017) 244–263.
[12] J. Li, C. Mei, C. Kumar, X. Zhang, On rule acquisition in decision formal contexts, Int. J. Mach. Learn. Cybern. 4 (2013) 721–731.
[13] J. Li, X. Yang, X. Song, J. Li, P. Wang, D. Yu, Neighborhood attribute reduction: a multi-criterion approach, Int. J. Mach. Learn.Cybern. 10 (2019) 731–742.
[14] J. Liang, F. Wang, C. Dang, Y. Qian, A group incremental approach to feature selection applying rough set technique, IEEE Trans. Knowl. Data Eng. 26 (2014) 294–308.
[15] J. Liang, J. Wang, Y. Qian, A new measure of uncertainty based on knowledge granulation for rough sets, Inf. Sci. 179 (2009) 458–470.
[16] S. Liao, Q. Zhu, Y. Qian, G. Lin, Multi-granularity feature selection on cost-sensitive data with measurement errors and variable costs, Knowl.-Based Syst. 158 (2018) 25–42.
[17] D. Liu, T. Li, D. Ruan, Probabilistic model criteria with decision-theoretic rough sets, Inf. Sci. 181 (2011) 3709–3722.
[18] K. Liu, X. Yang, H. Yu, J. Mi, P. Wang, X. Chen, Rough set based semi-supervised feature selection via ensemble selector, Knowl.-Based Syst. 165 (2019) 282–296.
[19] R. McGill, J. Tukey, W. Larsen, Variations of box plots, Am. Stat. 32 (1978) 12–16.
[20] F. Min, F. Liu, L. Wen, Z. Zhang, Tri-partition cost-sensitive active learning through knn, Soft Comput. 23 (2019) 1557–1572.
[21] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning About Data, Kluwer Academic Publishers, 1992.
[22] W. Pedrycz, From fuzzy data analysis and fuzzy regression to granular fuzzy data analysis, Fuzzy Sets Syst. 274 (2015) 12–17.
[23] W. Pedrycz, W. Homendad, Building the fundamentals of granular computing: a principle of justifiable granularity, Appl. Soft Comput. 13 (2013) 4209C4218.
[24] W. Pedrycz, G. Succi, A. Sillitti, J. Iljazi, Data description: a general framework of information granules, Knowl.-Based Syst. 80 (2015) 98–108.
[25] W. Pedrycz, X. Wang, Designing fuzzy sets with the use of the parametric principle of justifiable granularity, IEEE Trans. Fuzzy Syst. 24 (2016) 489–496.
[26] Y. Qian, H. Cheng, J. Wang, J. Liang, W. Pedrycz, C. Dang, Grouping granular structures in human granulation intelligence, Inf. Sci. 382–383 (2017) 150–169.
[27] J. Song, E. Tsang, D. Chen, X. Yang, Minimal decision cost reduct in fuzzy decision-theoretic rough set model, Knowl.-Based Syst. 126 (2017) 104–112.

[28] C. Wang, Q. He, M. Shao, Q. Hu, Feature selection based on maximal neighborhood discernibility, Int. J. Mach. Learn.Cybern. 9 (2018) 1929–1940.
[29] C. Wang, Q. Hu, X. Wang, D. Chen, Y. Qian, Z. Dong, Feature selection based on neighborhood discrimination index, IEEE Trans. Neural Netw. Learn.Syst. 29 (2018) 2986–2999.
[30] F. Wilconxon, Individual comparisons by ranking methods, Biometrics 1 (1945) 80–83.
[31] W. Wu, Y. Leung, A comparison study of optimal scale combination selection in generalized multi-scale decision tables, Int. J. Mach. Learn.Cybern. (2019), doi:10.1007/s13042-019-00954-1.
[32] S. Xu, X. Yang, H. Yu, D. Yu, J. Yang, E. Tsang, Multi-label learning with label-specific feature reduction, Knowl.-Based Syst. 104 (2016) 52–61.
[33] W. Xu, Y. Guo, Generalized multigranulation double-quantitative decision-theoretic rough set, Knowl.-Based Syst. 105 (2016) 190–205.
[34] W. Xu, W. Li, Granular computing approach to two-way learning based on formal concept analysis in fuzzy datasets, IEEE Trans. Cybern. 46 (2016) 366–379.
[35] W. Xu, W. Li, X. Zhang, Generalized multigranulation rough sets and optimal granularity selection, Granul. Comput. 2 (2017) 271–288.
[36] W. Xu, W. Sun, Y. Liu, W. Zhang, Fuzzy rough set models over two universes, Int. J. Mach. Learn.Cybern. 4 (2013) 631–645.
[37] W. Xu, J. Yu, A novel approach to information fusion in multi-source datasets: a granular computing viewpoint, Inf. Sci. 378 (2017) 410–423.
[38] X. Yang, T. Li, H. Fujita, D. Liu, A sequential three-way approach to multi-class decision, Int. J. Approx. Reason. 104 (2019) 108–125.
[39] X. Yang, T. Li, D. Liu, H. Fujita, A temporal-spatial composite sequential approach of three-way granular computing, Inf. Sci. 486 (2019) 171–189.
[40] X. Yang, Y. Qi, X. Song, J. Yang, Test cost sensitive multigranulation rough set: model and minimal cost selection, Inf. Sci. 250 (2013) 184–199.
[41] X. Yang, Y. Yao, Ensemble selector for attribute reduction, Appl. Soft Comput. 70 (2018) 1–11.
[42] Y. Yao, Three-way decision and granular computing, Int. J. Approx. Reason. 103 (2018) 107–123.
[43] Y. Yao, Y. Zhao, J. Wang, On reduct construction algorithms, Trans. Comput. Sci. II 5150 (2008) 100–117.
[44] J. Zhan, W. Xu, Two types of coverings based multigranulation rough fuzzy sets and applications to decision making, Artif. Intell. Rev. (2018), doi:10.1007/s10462-018-9649-8.
[45] X. Zhang, C. Mei, D. Chen, J. Li, Feature selection in mixed data: a method using a novel fuzzy rough set-based information entropy, Pattern Recognit. 56 (2016) 1–15.
[46] H. Zhao, P. Wang, Q. Hu, Cost-sensitive feature selection based on adaptive neighborhood granularity with multi-level confidence, Inf. Sci. 366 (2016) 134–149.
[47] P. Zhu, Q. Hu, W. Zuo, M. Yang, Multi-granularity distance metric learning via neighborhood granule margin maximization, Inf. Sci. 282 (2014) 321–331.