Applied Intelligence (2022) 52:17636–17651
https://doi.org/10.1007/s10489-022-03371-0

Accelerated multi-granularity reduction based on neighborhood rough sets

Yizhu Li[1] · Mingjie Cai[1] · Jie Zhou[2] · Qingguo Li[1]

Accepted: 9 February 2022 / Published online: 4 April 2022

Abstract

The notion of multi-granularity has been introduced into various mathematical models in granular computing. For example, neighborhood rough sets can derive a good multi-granularity structure by gradually changing the size of neighborhood radius. Attribute reduction is an important topic in neighborhood rough sets. However, in the case of multi-granularity, the challenge of high computational complexity and difficulty in synthesizing multi-granularity information when performing reduction algorithms always exists. To address such limitations, an accelerated algorithm for multi-granularity reduction is designed. Firstly, we construct a multi-granularity reduction structure with multiple different neighborhood radii to reduce the elapsed time of computing reducts. In this way, the consumed time of calculating the distance is similar to the one of single granularity reduction, and the elapsed time of computing multi-granularity reducts can be reduced. Secondly, multiple granularity information is integrated in each attribute evaluation. Finally, we evaluated the proposed method from multiple perspectives on 12 UCI datasets. Compared with other multi-granularity reduction algorithms, the proposed method not only generates reducts with relatively high quality, but also improves the time efficiency of multi-granularity reduction algorithm.

Keywords Neighborhood rough sets · Multi-granularity reduction · Attribute reduction

1 Introduction

A large number of high-dimensional, diverse and complex data have become a severe challenge to intelligent data processing [23]. Feature selection [6, 18, 23, 32, 35] is an important preorder step in machine learning and data mining. It focuses on dealing with the ever-increasing dimensions of data with limited computing power and extracting effective information from the complex high-dimensional data [25]. Specifically, feature selection methods need to be efficient and the selected features need to have a strong discrimination ability.

Feature selection based on granular computing is an emerging research field. Granular computing is a theory to process complex information and extracts knowledge from the data [11, 39]. It spans multiple disciplines and has multiple branch models [4, 45]. Information granulation adheres to the process of cognition and reasoning in which people observe and describe problems from different perspectives [41]. It divides complex data into some basic units. The multilevel granulation method can form multilevel information granules. Granules in the same layer can overlap and cooperate with each other. Granules in different layers form a hierarchical structure. This repetitive partitioning of data simplifies a complex problem [7, 16, 28, 38, 40, 44].

How to construct a granular structure is an essential problem in granular computing. According to the existing researches, there are usually two ways to construct granularity: data-based and model-based construction methods [21]. Data-based constructors usually include two aspects: feature

✉ Mingjie Cai
cmjlong@163.com

Yizhu Li
lyz96123@hnu.edu.cn

Jie Zhou
jie_jpu@163.com

Qingguo Li
liqingguoli@aliyun.com

[1] School of Mathematics, Hunan University, Changsha, 410082, China

[2] School of Computer Science and Software Engineering, Shenzhen University, Shenzhen, 518060, China

variation and sample variation. By increasing or decreasing the size of features or samples, different information granules can be formed [5, 29]. Model-based construction methods include model combination and model parameter variation. The latter is also a special case of model combination. The finer parameters often lead to more accurate models and can also represent finer granularity decision making methods. The difference of information granulation results can be reflected by the difference of parameter values [8, 24, 26].

Naturally, the neighborhood system uses the neighborhood radius to construct information granules. The smaller the neighborhood radius, the finer the granularity. Hu et al. [12] proposed neighborhood rough sets model based on neighborhood relations, which can deal with information systems with real values. Lin et al. [19] proposed a multi-granularity rough sets based on neighborhood rough sets. Then different measurements were applied to attribute reduction based on neighborhood rough sets [2, 10, 33, 42, 47], which can better deal with the data of different distributions. With the increasing size of data sets and increasing data dimensions, some acceleration strategies are also found in this context [25, 32, 35]. But for dynamically changing data, the traditional attribute reduction algorithms cannot dynamically update the reducts. In order to meet the challenges brought by constantly updated data sets, many scholars proposed dynamic reduction algorithms based on neighborhood rough sets [1, 15, 30, 43].

Previous studies have expanded neighborhood rough sets attribute reduction to other extent, but single-granularity reduction algorithms in the traditional neighborhood rough sets all have their inherent limitations [9, 21]. (1) The single-granularity attribute reduction of neighborhood rough sets cannot be used for neighborhood radius selection. (2) By using the single granularity based attribute reduction, only one reduct is obtained when giving a radius. There is no comparison among the performances derived from multiple reducts in terms of multiple radii. Multi-granularity reduction can generate multiple reducts according to multiple radii, which can reflect the variation tendency of discrimination performance [14]. (3) The reduction algorithms based on single granularity are not suitable for the reduction with multiple granularity levels. In neighborhood system, the multi-level granular structures can be constructed according to different neighborhood radii. Under the granularity structure constituted by different neighborhood radii, reducts that are suitable for different neighborhood radii can be generated.

In recent years, there are many studies on multi-granularity attribute reduction. Jiang [13] first proposed a multi-granularity reduction algorithm applicable to a variety of models, and carried out numerical experiments using neighborhood rough set model. Then he [14] added supervised neighborhood rough set model on the basis of reference [13]. However, this multi-granularity reduction algorithm does not fully consider the multi-level structure of neighborhood rough sets, that is, the reduction at the current granularity is only related to the reduction at the upper or lower granularity. So it cannot integrate the multi-granularity information effectively. Liu [21] uses a new multi-granularity reduction algorithm, which uses the The finest and coarsest grained information to carry out multi-granularity feature selection. The above studies all have the same shortcoming: it can not get multi-granularity reduction results quickly and effectively without losing all granularity information. In order to quickly obtain multi-granularity reducts, they all choose to ignore the information provided by most granularities.

To address the above problems, a novel multi-granularity reduction algorithm is designed in the context of neighborhood rough sets. It presents an accelerated strategy to conduct multi-granularity reduction while all granularity of information is calculated and well applied. Firstly, we designed the horizontal granulation structure. By observing the characteristics of the horizontal granulation structure, we found that it can reduce the number of sample distance calculation when adding attributes in multi-granularity reduction. Then, a comprehensive attribute evaluation strategy is designed by a voting method, and attributes suitable for multiple granularity are selected preferentially. Finally, the proposed algorithm is compared with the existing multi-granularity reduction algorithms. Experimental results demonstrate the superiority of the proposed algorithm in terms of the time consumption and classification accuracy.

The rest of study is organized as follows. In Section 2, the basic notions of neighborhood rough sets are introduced and the structure of neighborhood rough sets under multiple neighborhood radii is given. In this context, a novel multi-granularity reduction algorithm is proposed in Section 3. Section 4 demonstrates the effectiveness of the multi-granularity reduction algorithm through experimental results. Some conclusions are drawn in Section 5.

2 Single granularity reduction in neighborhood rough sets

2.1 Single granularity neighborhood rough sets

Formally, a decision system is denoted as a tuple $DS = \langle U, AT, d \rangle$, where U is a finite non-empty set $U = \{x_1, x_2, ..., x_n\}$, called the universe; AT is the set of all conditional attributes; for any $x_i \in U$, $\forall a \in AT$, $a(x_i)$ represents the value of sample x_i under attribute a; d is the decision attribute (this article mainly studies the problem

with a single decision attribute and d is an attribute whose value is nominal). For any $x_i \in U$, $d(x_i)$ describes the label of the sample x_i at the decision attribute. An equivalence relation with respect to d is denoted as $IND_d = \{(x, y) \in U \times U : d(x) = d(y)\}$. Through the equivalence relation IND_d, a partition of U can be obtained as $U/IND_d = \{X_1, X_2, ..., X_p\}$. Furthermore, $\forall X_i \in U/IND_d$ and $\forall x_j \in X_i$, we denote $X_i = [x_j]_d$.

Hu [12] has defined neighborhood rough set which was applied to attribute reduction.

Definition 2.1 Given a decision system $S = \langle U, AT, d\rangle$. For any $B \subseteq AT$, the distance matrix $D_B = (dis_{ij}^B)_{|U|\times|U|}$ in terms of B is defined as,

$$dis_{ij}^B = \sqrt[m]{\sum_{a\in B} \|a(x_i) - a(x_j)\|^m}, \tag{1}$$

where $|U|$ is the cardinal number of U and $x_i, x_j \in U$.

A neighborhood similarity relation N_B^δ can be derived from the distance matrix D_B and a given radius δ. The neighborhood simimarity relation N_B^δ could be writen as,

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

It is easy to notice that N_B^δ satisfy reflexivity and symmetry. By constructing the neighborhood relationship, points with high similarity can form a granule, $\delta_B(x_i) = \{x_j \mid x_j \in U, (x_i, x_j) \in N_B^\delta\}$. All the granules cover the whole universe.

Definition 2.2 Giving universe U, N_B^δ is a neighborhood relation where B is the set of attributes and δ is the radius of the neighborhood. Neighborhood relation N_B^δ provides a feasible strategy to construct a family of neighborhood granules on U. Then we call $GS_B^\delta = \langle U, N_B^\delta\rangle$ a granular structure.

Definition 2.3 Given a decicion system $\langle U, AT, d\rangle$, U can be divided into p blocks by the decision attribute d, $U/IND_d = \{X_1, X_2, ..., X_p\}$. For any $B \subseteq AT$, the upper and lower approximations of d with respect to attributes B are defined as,

$$\underline{N_B}(d) = \bigcup_{i=1}^{p} \{x_j \mid \delta_B(x_j) \subseteq X_i, x_j \in U\}, \tag{2}$$

$$\overline{N_B}(d) = \bigcup_{i=1}^{p} \{x_j \mid \delta_B(x_j) \cap X_i \neq \emptyset, x_j \in U\}. \tag{3}$$

After obtaining the upper and lower approximations of d, we call $POS_B^\delta(d)$ positive region of B with repect to d if it satisfies that $\forall x \in POS_B^\delta(d)$, $\delta_B(x_i)$ are consistent belong to one decision class. It is obvious that $POS_B^\delta(d) = \underline{N_B}(d)$.

Definition 2.4 Given a decicion system $\langle U, AT, d\rangle$, the neighborhood approximation quality $\gamma_B^\delta(d)$ of d associated with B is formulated as,

$$\gamma_B^\delta(d) = |POS_B^\delta(d)|/|U|. \tag{4}$$

Obviously, the range of $\gamma_B^\delta(d)$ is between 0 and 1. It is the proportion of samples that can be correctly classified with respect to B. Therefore, the neighborhood approximation quality can be used to measure the discriminative power of attribute subsets.

Definition 2.5 [46] Given $DS = \langle U, AT, d\rangle$, $\delta \in [0, 1]$, $B \subseteq AT$, the conditional entropy of d in terms of B is formulated as:

$$CE^\delta(B, d) = -\frac{1}{|U|} \sum_{x_i \in U} |\delta_B(x_i) \cap [x_i]_d| \log \frac{|\delta_B(x_i) \cap [x_i]_d|}{\delta_B(x_i)} \tag{5}$$

2.2 Single granularity reduction

Attribute reduction not only reduces the cost of storing and processing data, but also improves the interpretability and generalization of the decision model. Up to now, a large number of attribute reduction algorithms have been proposed [3, 17, 22, 31]. From the perspective of granular computing, Yao proposed a generalized form of attribute reduction [14, 21].

Definition 2.6 Given a decision system $\langle U, AT, d\rangle$ and $\delta \ge 0$, ρ be a constraint function, $\forall B \subseteq AT$, we call B is a $\rho^\delta-$reduct if and only if
1. B meets ρ^δ constraint;
2. $\forall a \in B$, $B - \{a\}$ does not meet the ρ^δ constraint.

The constraint ρ^δ may have various forms. Approximation quality $\gamma_B^\delta(d)$ is an important index to measure the consistency of an attribute set. $\gamma_B^\delta(d)$ can be used as a measure of the importance of the attributes B in terms of d. The greater the approximate quality, the higher the consistency between attribute subset and decision attribute. Therefore, the increase of approximation quality after the addition of attributes can be used as an index to measure the importance of attributes. Some scholars also proposed many other functions to measure the importance of attributes [34, 36]. They can all be considered as different realization of ρ^δ constraints.

Definition 2.7 Given a decision system $\langle U, AT, d\rangle$ and $\delta > 0$. We have $B \in AT$, $\forall a \notin B$. The importance of attribute a with respect to B under radius δ is

$$Sig^\delta(a, B, d) = \rho^\delta(B \cup \{a\}, d) - \rho^\delta(B, d). \tag{6}$$

Algorithm 1 uses the greedy search strategy to obtain a reduct. It starts from the empty set $B = \emptyset$, and adds the most significant attribute into B each time until a subset of attributes satisfying the constraint is achieved.

Algorithm 1 Addition method for computing reduct.

Input: Decision system $\langle U, AT, d\rangle$, radius δ
Output: Reduct B
1: $B \leftarrow \emptyset$;
2: Compute D_{AT};
3: Using D_{AT} compute $\rho^{\delta}(AT, d)$;
4: **repeat**
5: **for** a_i in $AT - B$ **do**
6: Compute $D_{B\cup\{a_i\}}$;
7: Use $D_{B\cup\{a_i\}}$ to compute $\rho^{\delta}(B \cup \{a_i\}, d)$;
8: **end for**
9: Select a_k such that $\rho^{\delta}(B \cup \{a_k\}, d) = max\{\rho^{\delta}(B \cup \{a_i\}, d) : \forall a_i \in AT - B\}$;
10: $B = B \cup \{a\}$;
11: **until** B meets the given ρ^{δ}-constraint

Obviously, It's trivial to figure out that the time complexity of Algorithm 1 is $O(|U|^2 \cdot |AT|^2)$. This can be summed up in the following two aspects: (1) Calculate the distance matrix D_B time complexity is $O(|U|^2)$, because any two samples in U should be applied for computing distance; calculate the $\rho^{\delta}(B \cup \{a_i\}, d)$ time complexity is $O(|U|)$(The measurement should be approximation quality or conditional entropy). The overall time complexity of steps 6-7 is $O(|U|^2)$. (2) In the worst case, the reduct we find is AT, the original condition attribute. Then the program will execute the iteration $|AT|$ times in step 4 and execute the step 5 $|AT| - i$ times in the i^{th} iteration. Finally, the time complexity of Algorithm 1 is $O(|U|^2 \cdot |AT|^2)$.

3 Accelerator for multi-granularity reduction in neghborhood rough sets

In this section, we introduce neighborhood rough sets with multi-radii, and then a fast and effective multi-granularity reduction algorithm in neighborhood rough sets is proposed.

3.1 Neighborhood rough sets with multi-radii

Neighborhood is an effective tool to implement information granules in rough sets theory. It not only provides a simple way to interpret information granules, but also presents a natural idea for building multi-granularity concepts [13, 14, 20, 21, 37].

Variation of data and parameters are two factors associated with multi-granularity. On the one hand, the change of the number of attributes will affect the size of granularity. This implements a top-down or bottom-up granulation process as we gradually add or remove attributes to the data. We can consider it as a vertical grain structure. On the other hand, the granulation of the universe is related to parameters. Taking neighborhood rough set as an example, different radii form different relations and form different granule systems. With the increase of neighborhood radius, the size of granule will increase correspondingly. From this point of view, these different scales of the radii imply a structure of horizontal multilevel granularity [41].

Definition 3.1 Giving information system $\langle U, AT\rangle$ and $\Theta = \{\delta_1, \delta_1, ..., \delta_s\}$, $\{N_B^{\delta_i}\}_{i=1...s}$ is a set of neighborhood relation where $B \in AT$. $GS_B^{\delta_i} = \langle U, N_B^{\delta_i}\rangle$. Then the horizontal neighborhood multi-granularity structure is defined as,

$$\mathbb{GS}_B^{\Theta} = \{GS_B^{\delta_i}\}_{i=1...s}. \tag{7}$$

We will use the following conventions throughout this section when we study neighborhood rough sets under horizontal multi-granularity.

- N_B^{Θ} denotes a set of neighborhood similarity relation, that is $N_B^{\Theta} = \{N_B^{\delta_i}\}_{i=1...s}$;
- $\overline{N_B^{\Theta}}(d)$ and $\underline{N_B^{\Theta}}(d)$ denotes the multiple upper and lower approximations of d with respect to attributes B, that is $\overline{N_B^{\Theta}}(d) = \{\overline{N_B^{\delta_i}}(d)\}_{i=1...s}$, $\underline{N_B^{\Theta}}(d) = \{\underline{N_B^{\delta_i}}(d)\}_{i=1...s}$;
- $\gamma_B^{\Theta}(d)$ denotes the multiple approximation quality of d associated with B, that is $\gamma_B^{\Theta}(d) = \{\gamma_B^{\delta_i}(d)\}_{i=1...s}$;
- $CE^{\Theta}(B, d)$ denotes the multiple conditional entropy of d associated with B, that is $CE^{\Theta}(B, d) = \{CE^{\delta_i}(B, d)\}_{i=1...s}$.

Where $B \in AT$ and $\Theta = \{\delta_1, \delta_1, ..., \delta_s\}$, $\{N_B^{\delta_i}\}_{i=1...s}$ is a set of neighborhood relation.

After simple derivation, it can be found that given a single distance matrix D_B and a set of neighborhood set Θ, a set of neighborhood relation N_B^{Θ} can be obtained. Under this set of neighborhood relations, we can derive a set of upper and lower approximations and calculate a set of constraints $\rho^{\Theta} = \{\gamma_B^{\delta_i}(d)\}_{i=1...s}$.

Proposition 3.2 *In decision information system $\langle U, AT, d\rangle$, we have $B \in AT$. Given a single distance matrix D_B with a set of neighborhood radii $\Theta = \{\delta_1, \delta_2, ..., \delta_s\}$ and $U/IND_d = \{X_1, X_2, ..., X_p\}$, a set of constraints $\rho^{\Theta} = \{\gamma_B^{\delta_i}(d)\}_{i=1...s}$ can be obtained.*

In order to effectively utilize this property to reduce the repeated operation of D_B in the multi-granularity reduction algorithm, an appropriate algorithm structure needs to be designed.

3.2 Synthesize multi-granularity reduction

The available neighborhood related attribute reduction methods are almost based on a single granularity situation. However, attribute reduction of a single granularity may involve several limitations: (1) It may result in poor adaptability of the generated reduction under different granularity environments; (2) If the results of attribute reduction under different radii needs to be calculated, the time consumption may be greatly increasing [13, 14]. Jiang et al. [14] formalizes the multi-granularity attribute reduction from the perspective of neighborhood granularity.

Definition 3.3 Given a decision system $\langle U, AT, d\rangle$, $\Theta = \{\delta_1, \delta_2, ..., \delta_s\}$ is a monotonically increasing neighborhood radius set such that $\forall i \in \{1, 2, ..., s-1\}, \delta_i < \delta_{i+1}$. We call $\mathbb{B} = \{B_1, B_2, ..., B_s\}$ multi-granularity reducts if and only if each B_i is the ρ^{δ_i}-reduct under granularity $GS_{AT}^{\delta_i}$.

To get a multi-granularity reduction $\mathbb{B}$, Algorithm 1 can be exploited repeatedly with different δ to get a set of reducts. This intuitive idea is shown in Fig. 1a. However, this process will take a considerable amount of time. Jiang et al. proposed a forward (reverse) accelerated search algorithm, which greatly reduced the time to obtain multi-granulariy reducts under different neighborhood radii [14]. For convenience, the forward accelerated search algorithm proposed by Jiang et al. is called FABS and the reverse accelerated search algorithm is called RABS. The general idea of FABS is shown in Fig. 1b. B_1' and B_2' in Fig. 1 respectively represent the reduction under neighborhood radius δ_1, δ_2.

As shown in Fig. 1b, B_1' obtained for the current neighborhood radius δ_1 serve as the initial attribute subset under the next neighborhood δ_2; the B_2' obtained under the current neighborhood radius δ_2 is used as the initial attribute subset for the next neighborhood δ_3; the rest can be done in the same manner. As for RABS, it execute like FABS. The difference is that RABS chooses δ_s as the first neighborhood radius for feature selection, and then the reduct that was found under δ_s radius is the initial candidate reduct under δ_{s-1} neighborhood radius. And so on until the reduct under δ_1 neighborhood is found. This mechanism does not effectively synthesize information from multiple granularities, because the reduction under the current granularity is mainly affected by the reduction under the previous granularity. In addition, the multi-granularity reduction of neighborhood rough sets may lead to some redundant computations.

To overcome the drawbacks of FABS, we will use the special properties of neighborhood rough sets to propose an accelerated method. The general idea of the algorithm is shown in Fig. 2. As it depicted in Fig. 2, $\Theta = \{\delta_1, \delta_2, \delta_3, \delta_4\}$, $AT = \{a_1, a_2, ..., a_m\}$ is an attribute set, $\mathbb{B} = \{B_1'', B_2'', B_3'', B_4''\}$ is a multi-granularity reducts and $B \subseteq AT$ is a pre-selected subset of attributes under all radii, whose initial value is shown in Fig. 2a as $\emptyset$. Then, suppose a_1 is selected and added to B through multi-granularity attribute evaluation criteria (Specific attribute evaluation criteria will be discussed in Section 3.2.2.), and then a_2 is selected and added to B for the second time. As shown in Fig. 2d, when adding the k^{th} attribute, B exactly satisfies the ρ^{δ_1}constraint, so the reduction under the neighborhood radius of δ_1 is $B_1'' = \{a_1, a_2, ..., a_k\}$. The multigranularity reduction for the remaining radii continues. When a_{k+1} is added to B, $B = \{a_1, ..., a_k, a_{k+1}\}$ satisfies ρ^{δ_2} and ρ^{δ_4} constraints respectively, so δ_2 and δ_4 also stop and $B_2'' = B_4'' = \{a_1, ..., a_k, a_{k+1}\}$, as shown in Fig. 2e. When a_{k+2} is added to B, B satisfies ρ^{δ_3} constraint, so

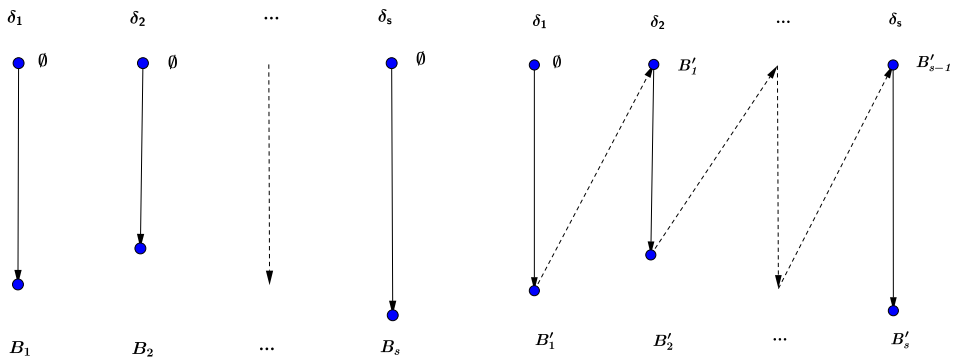

Fig. 1 Algorithm diagram

(*a*) Algorithm 1 with multi-granularity reduction (*b*) Forward accelerated search algorithm

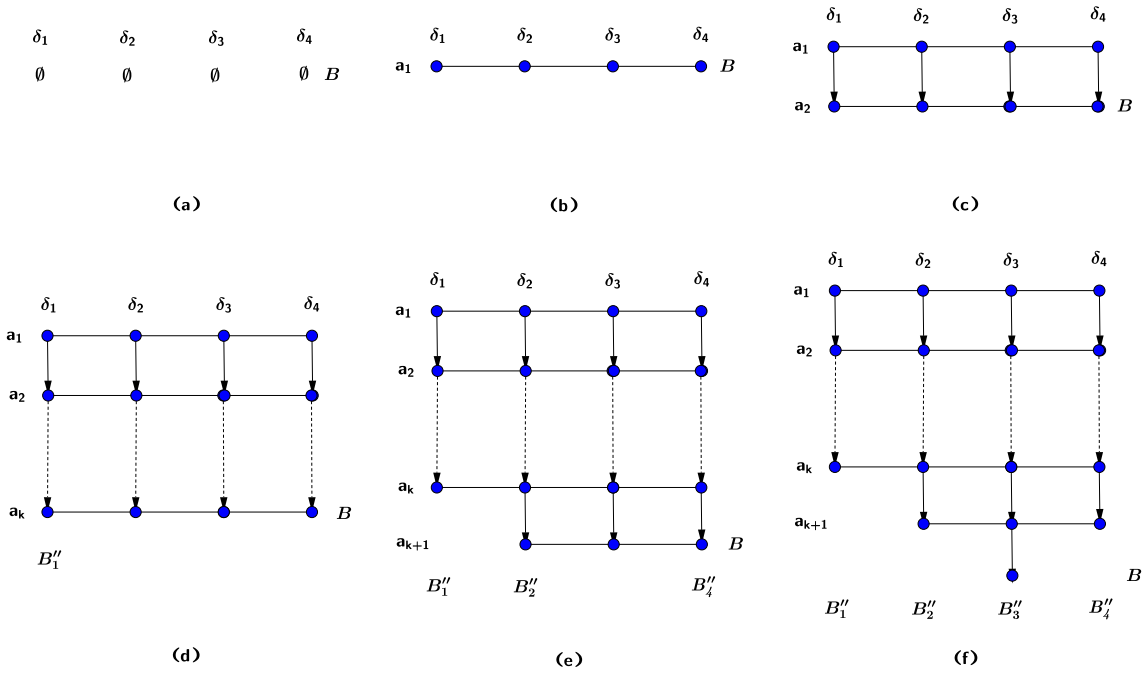

Fig. 2 Algorithm diagram

$B_3'' = \{a_1, ..., a_{k+2}\}$. Then all the reducts have been found, $\mathbb{B} = \{B_1'', B_2'', B_3'', B_4''\}$. It is worth noting that Algorithm 2 degenerates into Algorithm 1 when $\Theta = \{\delta\}$.

3.2.1 Acceleration of multi-granularity attribute evaluation

The feature selection algorithm based on rough sets can be divided into two parts: attribute evaluation and attribute searching [27]. In particular, the attribute evaluation can be divided into two steps in the neighborhood rough sets: (1) Let B be the attribute subset, calculating the distance matrix D_B in terms of B; (2) The attribute evaluation function $\rho^{\delta}(B, d)$ is calculated according to the distance matrix D_B. The first step often consumes the most of the computing time. Then how to calculate the distance matrix D_B without repeating in the context of multi-granularity reduction is the essential problem.

Reviewing the feature evaluation on single granularity reduction, suppose the candidate attributes be a_1, a_2, a_3, a_4 and the neighborhood radius be δ, while $B \subseteq AT$ is pre-selected reduction set. In order to select the best candidate attributes from $AT-B$, we need to calculate $\rho^{\delta}(B\cup\{a_1\}, d)$, $\rho^{\delta}(B \cup \{a_2\}, d)$, $\rho^{\delta}(B \cup \{a_3\}, d)$, and $\rho^{\delta}(B \cup \{a_4\}, d)$ respectively. Thus, it is necessary to calculate $D_{B\cup\{a_1\}}$, $D_{B\cup\{a_2\}}$, $D_{B\cup\{a_3\}}$ and $D_{B\cup\{a_4\}}$ respectively. The distance matrix D_B is calculated 4 times in total.

Suppose there are 4 candidate attributes $\{a_1, a_2, a_3, a_4\}$ and 10 neighborhood radii $\Theta = \{\delta_1, ..., \delta_{10}\}$, in order to select one optimal candidate attribute to add into the candidate reduct, it has to calculate the attribute evaluation function $\rho^{\delta_i}(B\cup\{a_j\}, d)$ 4×10 times. According to Section 2.1, we need to calculate the distance matrix $D_{B\cup\{a_j\}}$ 4×10 times when calculating the evaluation function $\rho^{\delta_i}(B \cup \{a_j\}, d)$. However, according to Proposition 3.2, $D_{B\cup\{a_j\}}$ only needs to be calculated once when 10 attribute evaluation functions $\rho^{\Theta}(B \cup \{a_j\})$ are calculated. So when you want to calculate the evaluation function 4×10 times, you only need to calculate D_B 4 times which is the same as the single granularity attribute reduction. In this way, the number of times Algorithm 2 calculating the distance matrix D_B in multi-granularity reduction is almost the same as Algorithm 1 calculates D_B in single-granularity reduction. This can effectively reduce the consumption of time.

3.2.2 Feature selection based on voting strategy

In single-granularity reduction process, the greedy strategy selects an attribute each time to add into the reduct. Then in the multi-granularity situation, how to select one optimum attribute to join the attribute subset? Generally, in some multi-granularity algorithms [27], they simply considered the finest and coarsest granularity in attribute selection. However, only the finest and coarser granularity does not fully capture the characteristics of multi-granularity structure. Therefore, the majority voting strategy is adopted to implement the comprehensive evaluation under the multi-granularity. Given $\langle U, AT, d\rangle$ and $B \subseteq AT$, we need to construct an evaluation matrix $E^{\Theta}_{AT-B} = (e_{ij})_{s\times|AT-B|}$ for

attribute selection. s represents the number of neighborhood radii, and B represents the candidate attributes.

$$e_{ij} = \rho^{\delta_i}(a_j, B, d), a_j \in AT - B. \tag{8}$$

One row of the evaluation matrix E^{Θ}_{AT-B} represents the importance of different candidate attributes in a certain neighborhood radius. One column of the evaluation matrix represents how important a candidate attribute is at different neighborhood radii. After obtaining the evaluation matrix E^{Θ}_{AT-B}, the optimal attributes in different rows can be found. The optimal attribute in each row represents the best candidate in the current neighborhood radius. The best attribute in each row gets one vote (if two attributes tie for first place within a neighborhood radius, both attribute get one vote each). The attribute with the most votes becomes the one selected in the round (if multiple attributes tie for first place, a random attribute for first place will be selected). Then the optimal attribute with the highest occurrence frequency is selected as the candidate attribute.

Example 1 For a decision system $\langle U, AT, d\rangle$, $AT = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7\}$, B is a subset of AT, $B = \{a_5, a_6, a_7\}$, and the neighborhood radii $\Theta = \{\delta_1, \delta_2, \delta_3, \delta_4\}$. In order to select the best attribute and add it to the candidate subset B, we calculate the attribute evaluation matrix E^{Θ}_{AT-B}. Assume that the values in Table 1 correspond to the values of the evaluation matrix E^{Θ}_{AT-B}, and that the larger the value in the matrix, the higher the importance of the attribute under this radius. Different rows represent the importance of attributes measured under different neighborhood radii. As shown in Table 1, the optimum attribute under different δ is then found. The best attribute with respect to δ_1 is a_2; the best attribute with respect to δ_2 is a_3; the best attribute with respect to δ_3 is a_2; the best attribute with respect to δ_4 is a_1. Summing up all the votes, we know that a_1 gets one vote, a_2 gets two votes, a_3 gets one vote, and a_4 gets none. Through the voting rules, a_2 is added to B. B will be updated to $\{a_2, a_5, a_6, a_7\}$.

Based on the above two optimization schemes, we improve the attribute reduction of multi-granularity neighborhood rough sets based on greedy policy search. By quickly calculating the constraint $\rho^{\Theta}(B, d)$, the times of computing the distance matrix D_B can be reduced to the same level as the attribute reduction of single-granularity rough sets. Feature selection based on voting strategy enables us to find a better attribute subset under multi-granularity. The proposed algorithm details are shown in Algorithm 2.

Table 1 E^{Θ}_{AT-B}

	a_1	a_2	a_3	a_4
δ_1	0.90	<u>0.91</u>	0.90	0.89
δ_2	0.89	0.89	<u>0.9</u>	0.89
δ_3	0.85	<u>0.86</u>	0.85	0.85
δ_4	<u>0.83</u>	0.82	0.82	0.80

Algorithm 2 Accelerated multi-granularity neighborhood rough sets attribute reduction.

Input: Decision system $\langle U, AT, d\rangle$, radius set $\Theta = \{\delta_1, \delta_2, ..., \delta_s\}$
Output: Multi-granularity reduct $\mathbb{B} = \{B_1, B_2, ..., B_s\}$
1: $B \leftarrow \emptyset$;
2: Compute D_{AT};
3: Use D_{AT} Compute $\rho^{\Theta}(AT, d)$;
4: **repeat**
5: **for** a_i in $AT - B$ **do**
6: Compute $D_{B\cup\{a_i\}}$;
7: Use $D_{B\cup\{a_i\}}$ to compute $\rho^{\Theta}(B \cup \{a_i\}, d)$;
8: $E^{\Theta}_{AT-B}[i, :] = \rho^{\Theta}(B \cup \{a_i\}, d)$;
9: **end for**
10: Use evaluation matrix E^{Θ}_{AT-B} to vote for each candidate attribute;
11: Attribute a_k was selected by voting strategy with respect to E^{Θ}_{AT-B};
12: $B = B \cup \{a_k\}$;
13: **for** ρ_i in $\rho^{\Theta}(AT, d)$ **do**
14: **if** $\rho_i = \rho^{\delta_i}(B, d))$ **then**
15: $B_i = B$;
16: $\Theta = \Theta - \{\delta_i\}$;
17: **end if**
18: **end for**
19: **until** $\Theta = \emptyset$

The iteration of the algorithm is mainly composed of three parts. The first part is the 5^{th} to 9^{th} steps of the algorithm, which construct the attribute evaluation matrix E^{Θ}_{AT-B}. The second part is the 10^{th} to the 12^{th} step of the algorithm. It selects the optimal attributes and adds them to the attribute subset by voting strategy according to the evaluation matrix E^{Θ}_{AT-B}. The third part is the 13-18 steps of the algorithm. For all δ belonging to Θ, we need to determine whether the attribute subset B satisfies the ρ^{δ}-constraint. If δ_t is satisfied, the reduction of B_t is determined: $B_t = B$, then the current δ_t is removed from Θ, and the algorithm no longer traverses the δ_t radius. When all δ in Θ are deleted, it means that all neighborhood radius reducts have been found. Then the algorithm is over. The following experiments will be used to prove the effectiveness of the algorithm.

Referring to Algorithm 1, it is not difficult to find that the time complexity of Algorithm 2 is also $O(|U|^2 \cdot |AT|^2)$. Similar to Algorithm 1, the time complexity of Algorithm 2 can also be divided into two parts: (1) calculate the distance

Table 2 Introduction to datasets

Data	#Samples	#Attributes	#Decision classes
wine	178	13	3
ionosphere	350	35	2
glass	214	10	6
data-banknote	1371	4	2
sonar	208	60	2
fertility-Diagnosis	100	10	2
accent	329	12	6
waveform	5000	21	3
plrx	182	13	2
wdbc	569	32	2
movement	360	91	10
BreastTissue	106	10	5

matrix D_B time complexity is $O(|U|^2)$, because any two samples in U should be applied for computing distance; calculate the $\rho^{\delta}(B \cup \{a_i\}, d)$ time complexity is $O(|\Theta| \cdot |U|)$(The measurement should be approximation quality or conditional entropy). The overall time complexity of steps 6-8 is $O(|U|^2) + O(|\Theta| \cdot |U|) = O(|U| \cdot (|U| + |\Theta|))$. (2) In the worst case, step 4 will run $|AT|$ times, the last iteration $B = AT$. In each iteration of step 4, we need to calculate distance matrix $|AT - B|$ time. This result is not affected even when the neighborhood radius increases. Finally, the time required for all calculations is accumulated, and the time complexity of Algorithm 1 is $O(|U| \cdot (|U| + |\Theta|) \cdot |AT|^2)$.

4 Experimental results

In order to verify that Algorithm 2 can effectively obtain multi-granularity reducts, this section compares Algorithm 2 with other multi-granularity reduction on 12 data sets. Details of the used datasets are shown in Table 2 below.

In our experiments, to simplify our discussions, FABS_D, RABS_D, PABS_D, Brute_D implies the following, respectively,

(1)FABS using approximation quality as constraint;
(2)RABS using approximation quality as constraint;
(3)PABS using approximation quality as constraint;
(4)Algorithm 1 using approximation quality as constraint.

FABS_CE, RABS_CE, PABS_CE, Brute_CE indicates the following, respectively,

(1)FABS using conditional entropy as constraint;
(2)RABS using conditional entropy as constraint;
(3)PABS using conditional entropy as constraint;
(4)Algorithm 1 using conditional entropy as constraint.

4.1 Experiments settings

All the multi-granularity reduction algorithms were implemented in Python 3.8 and were conducted on a dual processor, 24 core, 48 logic processor server with 64GB of memory.

We use the 5-fold cross validation to conduct the experiment, so as to reduce the deviation caused by the data and obtain more reliable and stable results. The data

Table 3 Comparisons among elapsed time of obtaining the multi-granularity reducts

	Approximation quality				conditional entropy			
	Brute_D	FABS_D	RABS_D	PABS_D	Brute_CE	FABS_CE	RABS_CE	PABS_CE
wine	3.3743	1.0082	0.8980	0.5674	3.3020	0.9707	0.9748	0.5671
ionosphere	63.2968	15.0055	14.0691	7.4364	58.3753	14.2442	13.3211	9.9344
glass	2.6545	1.0234	1.0508	0.3700	2.5987	0.8320	0.9679	0.3636
data banknote	8.9316	5.4687	5.7693	1.9743	9.0597	5.4463	5.7171	1.9721
sonar	37.7869	6.4985	7.2477	6.1832	32.3376	6.6177	6.3815	5.9979
fertility	0.9872	0.3680	0.3398	0.1336	0.9793	0.3461	0.3453	0.1362
accent	7.7219	2.3442	2.5105	1.1144	7.8092	2.2097	2.6954	1.2072
waveform	2256.8171	463.8463	486.2884	506.9659	2237.7814	520.9583	593.0367	518.1255
plrx	3.5947	1.0123	1.0194	0.4411	3.3576	0.9089	1.0774	0.5350
wdbc	113.0216	24.8895	25.5221	15.7413	107.9260	23.3899	24.3187	17.4738
movement	456.8637	131.3087	105.1430	70.2847	443.9714	128.4817	113.5914	85.9019
BreastTissue	1.1292	0.4412	0.4713	0.1746	1.1492	0.3872	0.4451	0.1450
Average rank	4.0000	2.1667	2.6667	1.1667	4.0000	2.3333	2.6667	1.0000

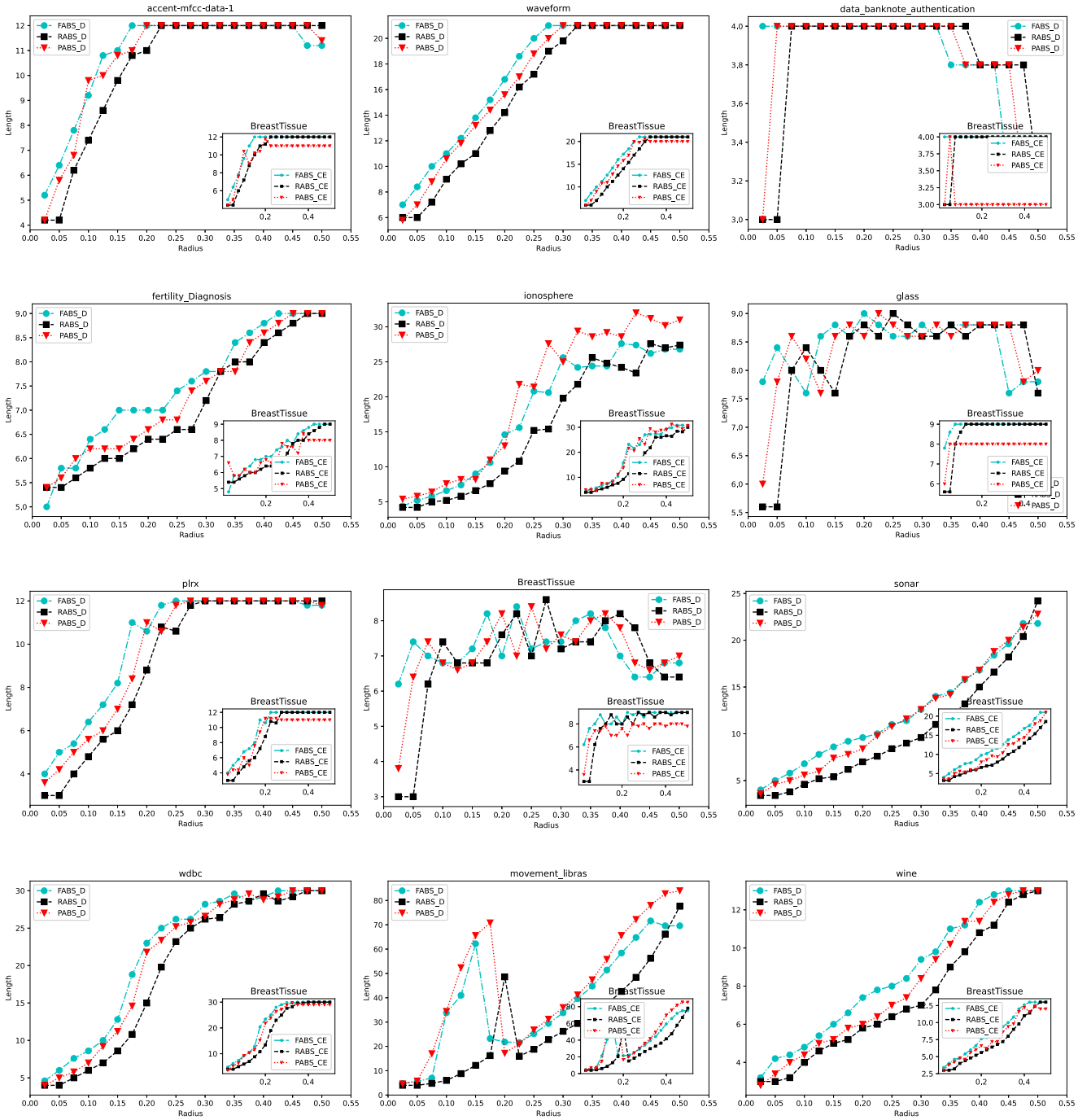

Fig. 3 Comparisons among different lengths w.r.t. different reducts

is divided into five sub-samples of the same size and disjoint. One single sub-sample is retained as the data of the validation model, and the other 5-1 subsamples are used to carry out training of the multi-granularity reduction algorithm. Cross validation is repeated five times, once for each subsample, averaging five results or using some other combination, resulting in a single estimate for the variable we want.

Twenty increasing neighborhood radius sets $\Theta = \{0.025, 0.05, ..., 0.5\}$ are used as the input variable Θ of FABS, PABS. Each neighborhood radius increases at a rate of 0.025, and different radius represent different granularity.

Table 4 Comparison of average length of multi-granularity reduction

	Approximation quality			conditional entropy		
	FABS_D	RABS_D	PABS_D	FABS_CE	RABS_CE	PABS_CE
wine	8.59	7.34	7.97	8.71	7.36	7.94
ionosphere	17.70	15.55	20.08	19.64	16.19	19.52
glass	8.43	8.22	8.38	8.92	8.59	7.90
data banknote	3.84	3.82	3.83	4.00	3.90	3.05
sonar	12.22	10.20	11.84	12.04	9.03	10.44
fertility	7.50	7.06	7.28	7.46	7.06	7.08
accent	10.84	10.31	10.69	10.99	10.35	10.04
waveform	17.15	15.83	16.60	17.26	15.70	16.22
plrx	10.06	9.18	9.65	10.12	9.19	9.21
wdbc	21.65	19.26	20.71	22.3	19.55	21.12
movement	38.96	28.96	45.49	38.83	29.49	45.04
BreastTissue	7.22	6.85	7.11	8.53	7.92	7.42
Average rank	2.83	1.00	2.16	2.91	1.33	1.75

Finally, reducts under 20 different granularity can be obtained. These reducts can be collectively referred to as multi-granularity reducts.

4.2 Comparison of running time

Table 3 records the time consumption of attribute reduction of Algorithm 1, FABS and PABS in multi-granularity neighborhood rough sets under $\Theta = \{0.025, 0.05, ..., 0.5\}$. It should be noted that Algorithm 1 needs to be called several times with different neighborhood radius to obtain multi-granularity reducts. We underline the fastest algorithm to obtain multi-granularity reducts under the same constraint. According to the results shown in Table 3, it is not difficult to find that (Fig. 3):

(1). No matter what constraint is used, FABS, RABS and PABS are far better than Algorithm 1 in time efficiency. It shows that the fast multi-granularity reduction algorithm proposed in this paper can greatly save the time of multi-granularity reduction.

Table 5 Comparison of similarity of multi-granularity reduction

	Approximation quality			conditional entropy		
	FABS_D	RABS_D	PABS_D	FABS_CE	RABS_CE	PABS_CE
wine	0.6493	0.6636	0.7021	0.6588	0.6205	0.6250
ionosphere	0.6771	0.7478	0.7130	0.6546	0.5986	0.6556
glass	0.9635	0.9626	0.9900	0.9675	0.9800	0.9833
data banknote	0.9875	1.0000	1.0000	0.9875	1.0000	1.0000
sonar	0.2255	0.2170	0.2897	0.2576	0.3042	0.4102
fertility	0.8121	0.7471	0.7667	0.7505	0.6980	0.6918
accent	0.9049	0.8562	0.9195	0.9199	0.8351	0.9467
waveform	0.7883	0.7617	0.8189	0.7434	0.7724	0.8335
plrx	0.8362	0.8103	0.8958	0.9084	0.8911	0.8891
wdbc	0.8046	0.8765	0.8096	0.7750	0.6679	0.7849
movement	0.4767	0.6406	0.3340	0.3314	0.2763	0.3388
BreastTissue	0.8997	0.9100	0.9536	0.9428	0.9722	0.9854
Average rank	2.3333	2.1667	1.4167	2.1667	2.3333	1.4167

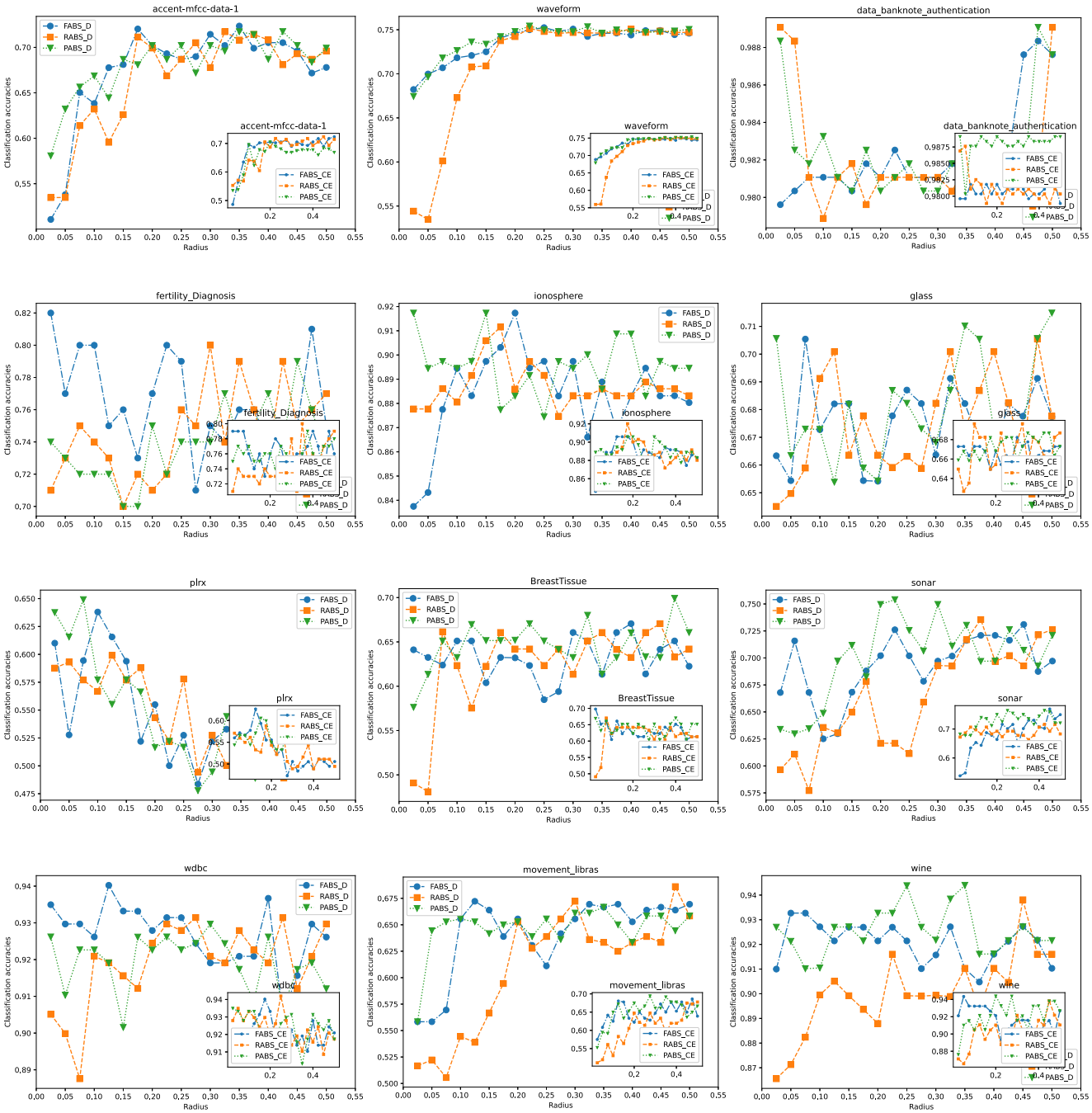

Fig. 4 Comparisons among classification accuracies w.r.t. different reducts (Decision tree)

(2). When using approximation quality, PABS is basically better than FABS and RABS in terms of time efficiency. Because PABS reduces the number of times to calculate the distance between samples. Occasionally, a special case occurs, such as FABS being faster than PABS on a data set waveform. It is found in the experiment that FABS may be better than PABS in time efficiency if the data sample is large and the number of attributes is small relative to the number of samples, while the overall multi-granularity reduction speed of PABS still maintains a high level.

4.3 Reduction attribute quality comparison

Table 3 exhibits the elapsed time of obtaining the multi-granularity reducts.. This section compares the average length of multi-granularity reducts obtained by different algorithm. After that, we calculat the similarity of multi-granularity reducts derived by PABS and Brute. Meanwhile the similarity between FABS with Brute and RABS with Brute is also calculated. Detailed results can be found in Tables 4 and 5.

Table 6 Average accuracy of KNN

	Approximation quality			conditional entropy		
	FABS_D	RABS_D	PABS_D	FABS_CE	RABS_CE	PABS_CE
wine	0.9593	0.9445	0.9583	0.9525	0.9365	0.9529
ionosphere	0.8367	0.8339	0.8389	0.8347	0.8341	0.8423
glass	0.6182	0.6190	0.6195	0.6215	0.6202	0.6592
data banknote	0.9977	0.9976	0.9977	0.9985	0.9981	0.9939
sonar	0.7269	0.7131	0.7439	0.7204	0.7343	0.7162
fertility	0.8795	0.8800	0.8800	0.8800	0.8800	0.8800
accent	0.7437	0.7360	0.7487	0.7492	0.7376	0.7309
waveform	0.8127	0.7887	0.8110	0.8155	0.7942	0.8142
plrx	0.6604	0.6590	0.6585	0.6577	0.6585	0.6644
wdbc	0.9652	0.9594	0.9657	0.9638	0.9638	0.9694
movement	0.6209	0.5589	0.6302	0.6134	0.5614	0.6265
BreastTissue	0.6457	0.6296	0.6454	0.6361	0.6236	0.6445
Average rank	1.7500	2.6667	1.5000	1.9167	2.4167	1.6667

4.3.1 Comparison of length

Table 4 and Fig. 3 show the average attribute length of multi-granularity reducts. The average attribute length of multi-granularity reducts can be written as:

$$\overline{B} = \frac{\sum\limits_{B\in\mathbb{B}} |B|}{|\mathbb{B}|}. \tag{9}$$

The underlined ones represent the algorithm with the shortest average attribute length under the same constraint. According to experience, reduct with too many attributes may result in redundancy of data, while too short reduct will lead to loss of information. The experimental results show that no matter which constraint is used, the multi-granularity reducts length obtained by PABS is mostly smaller than that obtained by FABS and larger than that obtained by RABS. It can be said that PABS algorithm adopts a more conservative strategy to obtain reduct. Therefor the reducts will not be too long or too short.

4.3.2 Comparison of similarity

This section compares the similarity of multi-granularity reducts derived by FABS, PABS to the multi-granularity reducts obtained by Algorithm 1. The higher the similarity

Table 7 Average accuracy of Decision tree

	Approximation quality			conditional entropy		
	FABS_D	RABS_D	PABS_D	FABS_CE	RABS_CE	PABS_CE
wine	0.9201	0.9019	0.9247	0.9181	0.9029	0.9195
ionosphere	0.8841	0.8868	0.8946	0.8869	0.8861	0.8900
glass	0.6746	0.6745	0.6795	0.6674	0.6662	0.6730
data banknote	0.9821	0.9822	0.9824	0.9808	0.9813	0.9875
sonar	0.6941	0.6657	0.7017	0.6851	0.6914	0.7286
fertility	0.7652	0.7467	0.7395	0.7638	0.7471	0.7614
accent	0.6749	0.6672	0.6829	0.6818	0.6726	0.6604
waveform	0.7360	0.7128	0.7388	0.7358	0.7157	0.7379
plrx	0.5379	0.5380	0.5412	0.5310	0.5298	0.5610
wdbc	0.9262	0.9191	0.9190	0.9206	0.9237	0.9247
movement	0.6422	0.6112	0.6476	0.6464	0.6095	0.6520
BreastTissue	0.6336	0.6230	0.6451	0.6324	0.6185	0.6389
Average rank	2.0833	2.5833	1.3333	2.0833	2.6667	1.2500

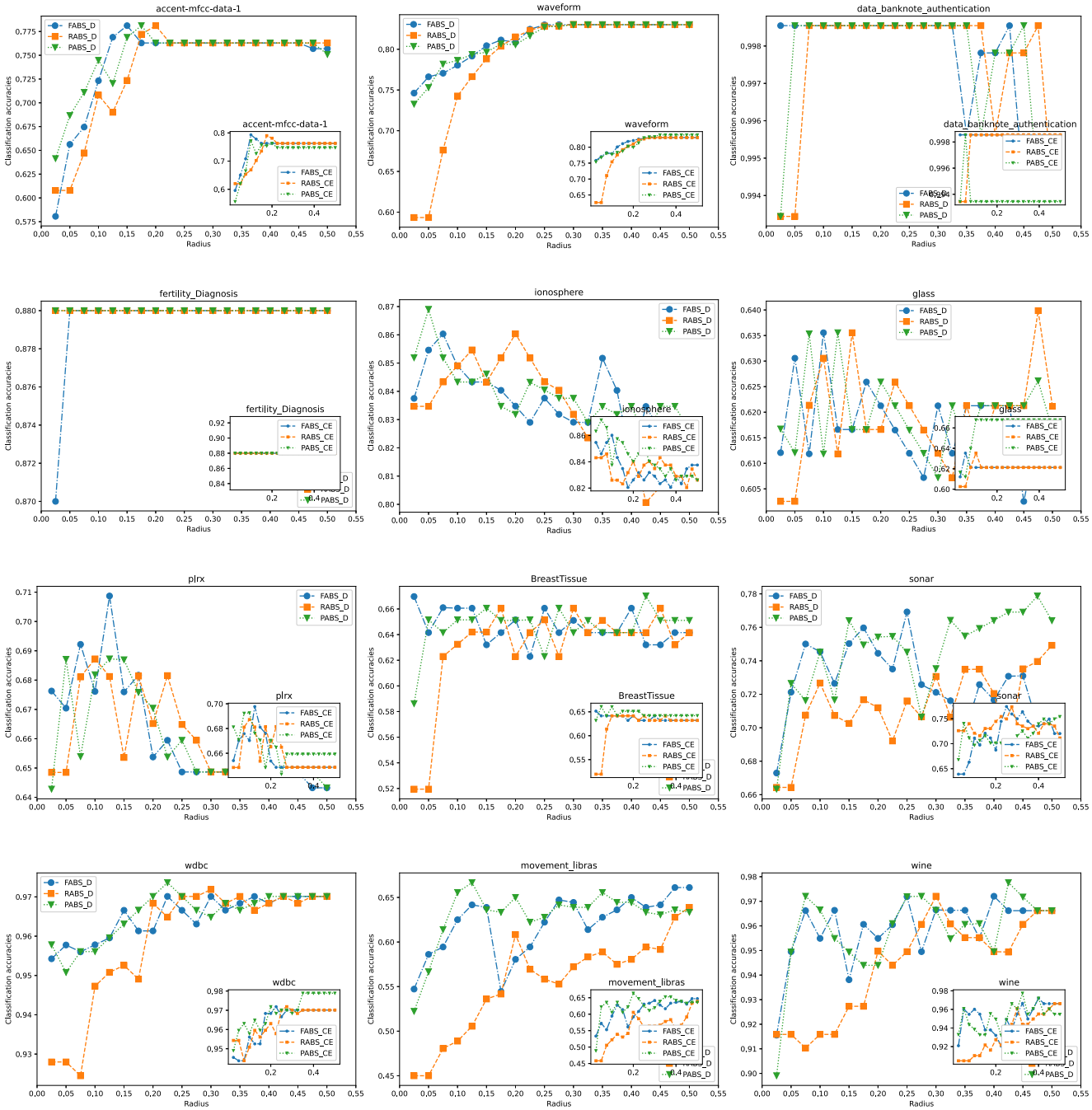

Fig. 5 Comparisons among classification accuracies w.r.t. different reducts (KNN)

with Algorithm 1, the smaller the difference with the original algorithm, and the more it can reflect the trend of the original multi-granularity reduction with the increase of the neighborhood, which is more beneficial to the selection of neighborhood radius under different data sets. So we define attribute similarity as: $Sim(B_1, B_2) = \frac{|B_1 \cap B_2|}{|B_1 \cup B_2|}$. Naturally, the similarity of multi-granularity reducts is

$$Sim(\mathbb{B}_1, \mathbb{B}_2) = \sum_{B_i \in \mathbb{B}_1, B_j \in \mathbb{B}_2} Sim(B_i, B_j). \tag{10}$$

Where $\mathbb{B}_1, \mathbb{B}_2$ are multi-granularity reducts while B_i, B_j are actually a reduct with respect to same δ.

As it depited in Table 5, the similarity between PABS and Algorithm 1 is higher than that between FABS and Algorithm 1 in most cases. It can be concluded that the multi-granularity reducts generated by PABS is more excellent than the multi-granularity reducts obtained by FABS. The Conclusion is true when comparing RABS with PABS.

4.4 Comparison of classification accuracy

Finally, we choose some classical classifiers to carry out experiments, which can get the classification accuracy under the multi-granularity reduction. Figure 4 displays the classification accuracy of different multi-granularity reduction algorithms using the classifier decision tree, and Table 6 demonstrates the average accuracy of multi-granularity reducts under the KNNclassifier. Figure 5 shows the accuracy rate obtained by using KNN classifier to classify different multi-granularity reduction algorithms. Table 7 demonstrates the average accuracy rate of multi-granularity reducts under decision tree classifier.

According to Tables 6 and 7, it is not difficult to find that the average accuracy of PABS is often higher than that of FABS in different data sets. However, according to Section 3, we know that the average length of the reduction obtained by PABS is smaller than that of FABS. This means that PABS uses fewer attributes to achieve a higher classification accuracy than FABS.

By comparing PABS with RABS, it can be inferred that the average accuracy of PABS is obviously better than that of RABS on most data sets. Therefore, it can be proved that PABS is more effective than FABS and RABS in multi-granularity reduction.

5 Conclusions

Compared with single-granularity reduction, multi-granularity reduction based on neighborhood rough set enable us to find the appropriate neighborhood radius. However, traditional multi-granularity reduction often takes a considerable amount of time. In this paper, we explored a multi-granularity reduction algorithm based on neighborhood rough sets. By reducing duplicate calculation of the sample distance matrix, the computation time of the algorithm will be greatly reduced. Furthermore, to select an attribute over multi-granularity attribute selection, the voting strategy is applied. Experiments on 12 UCI data sets demonstrate the effectiveness and efficiency of the algorithm. Based on the overall results, we can conclude that the proposed algorithm can not only synthesize information of multiple granularity to search for reduction, but also greatly reduce the time consumption required for obtaining multi-granularity reducts.

The following topics are challenges for further researches.

1. The acceleration strategy can be applied to other multi-granularity attribute reduction algorithms.
2. How to find the optimal reduct among multi-granularity reducts.

Acknowledgments The authors thanks editors and reviewers for their constructive comments and valuable suggestions. This work was supported by the Guangdong Basic and Applied Basic Research Foundation (No. 2021A1515011861), and Shenzhen Science and Technology Program (No. JCYJ20210324094601005).

References

1. Chen H, Li T, Cai Y, Luo C, Fujita H (2016) Parallel attribute reduction in dominance-based neighborhood rough set. Inf Sci 373:351–368
2. Chen H, Li T, Fan X, Luo C (2019) Feature selection for imbalanced data based on neighborhood rough sets. Inf Sci 483:1–20
3. Chen Y, Liu K, Song J, Fujita H, Yang X, Qian Y (2020) Attribute group for attribute reduction. Inf Sci 535:64–80
4. Chen Y, Qin N, Li W, Xu F (2019) Granule structures, distances and measures in neighborhood systems. Knowl-Based Syst 165:268–281
5. Cheng Y, Zhang Q, Wang G, Qing B (2020) Optimal scale selection and attribute reduction in multi-scale decision tables based on three-way decision. Inf Sci 541:36–59
6. Dai J, Hu Q, Hu H, Huang D (2018) Neighbor inconsistent pair selection for attribute reduction by rough set approach. IEEE Trans Fuzzy Syst 26(2):937–950
7. Fan J, Zhao T, Kuang Z, Zheng Y, Zhang J, Yu J, Peng J (2017) HD-MTL Hierarchical deep multi-task learning for large-scale visual recognition. IEEE Trans Image Process 26(4):1923–1938
8. Fang Y, Gao C, Yao Y (2020) Granularity-driven sequential three-way decisions: A cost-sensitive approach to classification. Inf Sci 507:644–664
9. Gao Y, Chen X, Yang X, Wang P, Mi J (2019) Ensemble-based neighborhood attribute reduction: A multigranularity view. Complexity 2019:1–17
10. García-torres M, Gómez-vela F, Melián-batista B, Moreno-vega JM (2016) High-dimensional feature selection via feature grouping: A variable neighborhood search approach. Inf Sci 326:102–118
11. Hu C, Zhang L, Wang B, Zhang Z, Li F (2019) Incremental updating knowledge in neighborhood multigranulation rough sets under dynamic granular structures. Knowl-Based Syst 163:811–829
12. Hu Q, Yu D, Xie Z (2008) Neighborhood classifiers. Expert Syst Appl 34(2):866–876
13. Jiang Z, Liu K, Yang X, Yu H, Fujita H, Qian Y (2020) Accelerator for supervised neighborhood based attribute reduction. Int J Approx Reason 119:122–150
14. Jiang Z, Yang X, Yu H, Liu D, Wang P, Qian Y (2019) Accelerator for multi-granularity attribute reduction. Knowl-Based Syst 177:145–158
15. Jing Y, Li T, Fujita H, Wang B, Cheng N (2018) An incremental attribute reduction method for dynamic data mining. Inf Sci 465:202–218
16. Ju H, Ding W, Yang X, Fujita H, Xu S (2021) Robust supervised rough granular description model with the principle of justifiable granularity. Appl Soft Comput 110:107612
17. Li J, Yang X, Song X, Li J, Wang P, Jun D (2019) Neighborhood attribute reduction: A multi criterion approach. International Journal of Machine Learning and Cybernetics 10(4):731–742
18. Liang J, Wang F, Dang C, Qian Y (2012) An efficient rough feature selection algorithm with a multi-granulation view. Int J Approx Reason 53(6):912–926

19. Lin G, Qian Y, Li J (2012) NMGRS: Neighborhood-based multigranulation rough sets. Int J Approx Reason 53(7):1080–1093
20. Lin Y, Li J, Lin P, Lin G, Chen J (2014) Feature selection via neighborhood multi-granulation fusion. Knowl-Based Syst 67:162–168
21. Liu K, Yang X, Fujita H, Liu D, Yang X, Qian Y (2019) An efficient selector for multi-granularity attribute reduction. Inf Sci 505:457–472
22. Liu K, Yang X, Yu H, Mi J, Wang P, Chen X (2019) Rough set based semi-supervised feature selection via ensemble selector. Knowl-Based Syst 165:282–296
23. Ni P, Zhao S, Wang X, Chen H, Li C (2019) PARA: A positive-region based attribute reduction accelerator. Inf Sci 503:533–550
24. Qian J, Liu C, Miao D, Yue X (2020) Sequential three-way decisions via multi-granularity. Inf Sci 507:606–629
25. Qian Y, Liang J, Pedrycz W, Dang C (2010) Positive approximation: An accelerator for attribute reduction in rough set theory. Artif Intell 174(9-10):597–618
26. Qian Y, Liang J, Yao Y, Dang C (2010) MGRS: A multi-granulation rough set. Inf Sci 180(6):949–970
27. Rao X, Yang X, Yang X, Chen X, Liu D, Qian Y (2020) Quickly calculating reduct: An attribute relationship based approach. Knowl-Based Syst 200:106014
28. She Y, Li J, Yang H (2015) A local approach to rule induction in multi-scale decision tables. Knowl-Based Syst 89:398–410
29. She Y, Qian Z, He X, Wang J, Qian T (2021) On generalization reducts in multi-scale decision tables. Inf Sci 555:104–124
30. Shu W, Qian W, Xie Y (2020) Incremental feature selection for dynamic hybrid data using neighborhood rough set. Knowl-Based Syst 194:105516
31. Sun L, Wang L, Ding W, Qian Y, Xu J (2020) Neighborhood multi-granulation rough sets-based attribute reduction using lebesgue and entropy measures in incomplete neighborhood decision systems. Knowl-Based Syst 192:105373
32. Urbanowicz RJ, Meeker M, La Cava W, Olson RS, Moore JH (2018) Relief-based feature selection: Introduction and review. J Biomed Inform 85(July):189–203
33. Wan Q, Li J, Wei L, Qian T (2020) Optimal granule level selection: A granule description accuracy viewpoint. Int J Approx Reason 116:85–105
34. Wang C, Hu Q, Wang X, Chen D, Qian Y, Dong Z (2018) Feature selection based on neighborhood discrimination index. IEEE Transactions on Neural Networks and Learning Systems 29(7):2986–2999
35. Wang C, Huang Y, Shao M, Fan X (2019) Fuzzy rough set-based attribute reduction using distance measures. Knowl-Based Syst 164:205–212
36. Wang C, Huang Y, Shao M, Hu Q, Chen D (2020) Feature selection based on neighborhood self-information. IEEE Trans Cybern 50(9):4031–4042
37. Wu J, Song J, Cheng F, Wang P, Yang X (2020) Research on multi-granularity attribute reduction method for continuous parameters. Journal of Frontiers of Computer Science and Technology 61906078:1–10
38. Xu Y (2019) Multigranulation rough set model based on granulation of attributes and granulation of attribute values. Inf Sci 484:1–13
39. Xu K, Pedrycz W, Li Z (2021) Granular computing: An augmented scheme of degranulation through a modified partition matrix. Fuzzy Sets Syst 1:1–18
40. Yang T, Zhong X, Lang G, Qian Y, Dai J (2020) Granular matrix: A new approach for granular structure reduction and redundancy evaluation. Transactions on Fuzzy Systems 28(12):3133–3144
41. Yang X, Li T, Liu D, Fujita H (2020) A multilevel neighborhood sequential decision approach of three-way granular computing. Inf Sci 538:119–141
42. Yang X, Chen H, Li T, Wan J, Sang B (2021) Neighborhood rough sets with distance metric learning for feature selection. Knowl-Based Syst 224:107076
43. Yang Y, Chen D, Wang H (2017) Active sample selection based incremental algorithm for attribute reduction with rough sets. Transactions on Fuzzy Systems 25(4):825–838
44. Yao MX (2019) Granularity measures and complexity measures of partition-based granular structures. Knowl-Based Syst 163:885–897
45. Ye J, Zhan J, Ding W, Fujita H (2021) A novel fuzzy rough set model with fuzzy neighborhood operators. Inf Sci 544:266–297
46. Zhang X, Mei C, Chen D (2016) J. Li. Feature selection in mixed data A method using a novel fuzzy rough set-based information entropy. Pattern Recogn 56:1–15
47. Zhao H, Wang P, Hu Q (2016) Cost-sensitive feature selection based on adaptive neighborhood granularity with multi-level confidence. Inf Sci 366:134–149

Yizhu Li received his bachelor's degree from Dalian Maritime University in 2015. He is now a postgraduate student in Hunan University. His current research interests include granular computing and feature selection.

Mingjie Cai received the Ph.D. degree from Hunan University in 2016. Currently, he is an associate professor at School of Mathematics, Hunan University, China. His current research interests include granular computing and machine learning.

Jie Zhou received the Ph.D. degree from the Tongji University in 2011. He is now an associate researcher in the college of computer science and software engineering, Shenzhen University, China. His current major research interests include uncertainty analysis, pattern recognition, data mining, and intelligent systems.

Qingguo Li received the Ph.D. degree from Hunan University in 1997. Currently, he is a professor at School of Mathematics, Hunan University, China. His current research interests include granular computing, lattice topology and concept lattice.