



# Two-Layer Information Granulation: Mapping-Equivalence Neighborhood Rough Set and Its Attribute Reduction

Changshun Liu<sup>1</sup>, Yan Liu<sup>1</sup>, Jingjing Song<sup>1,\*</sup> and Taihua Xu<sup>1,2</sup>

<sup>1</sup>School of Computer, Jiangsu University of Science and Technology, Zhenjiang, 212100, China

<sup>2</sup>Key Laboratory of Oceanographic Big Data Mining & Application of Zhejiang Province, Zhejiang Ocean University, Zhoushan, 316022, China

\*Corresponding Author: Jingjing Song. Email: songjingjing108@163.com

Received: 07 February 2023; Accepted: 23 April 2023; Published: 23 June 2023

**Abstract:** Attribute reduction, as one of the essential applications of the rough set, has attracted extensive attention from scholars. Information granulation is a key step of attribute reduction, and its efficiency has a significant impact on the overall efficiency of attribute reduction. The information granulation of the existing neighborhood rough set models is usually a single layer, and the construction of each information granule needs to search all the samples in the universe, which is inefficient. To fill such gap, a new neighborhood rough set model is proposed, which aims to improve the efficiency of attribute reduction by means of two-layer information granulation. The first layer of information granulation constructs a mapping-equivalence relation that divides the universe into multiple mutually independent mapping-equivalence classes. The second layer of information granulation views each mapping-equivalence class as a sub-universe and then performs neighborhood information granulation. A model named mapping-equivalence neighborhood rough set model is derived from the strategy of two-layer information granulation. Experimental results show that compared with other neighborhood rough set models, this model can effectively improve the efficiency of attribute reduction and reduce the uncertainty of the system. The strategy provides a new thinking for the exploration of neighborhood rough set models and the study of attribute reduction acceleration problems.

**Keywords:** Attribute reduction; information granulation; mapping-equivalence relation; neighborhood rough set

## 1 Introduction

In rough set theory [1–5], the whole sample is called the universe. The universe can be divided into several subsets of samples by indistinguishable relations between the samples, this is a process known as information granulation, and the divided subsets of samples are called information granules. The information granules are the basic units of the universe, and any concept in the universe can be approximated by the concatenation of information granules. This approach simulates the human



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

learning and reasoning process and is widely valued as it is easy to understand and generalize. Unfortunately, the classical rough set model based on indistinguishable relation is only suitable for dealing with discrete data, but not directly for continuous data, which is widely available in real applications. To extend the application of rough sets [6,7], Hu et al. [8] proposed the neighborhood rough set based on the concept of  $\delta$ -neighborhood in the metric space. The core of the neighborhood rough set model is the neighborhood relation. The composition of the neighborhood information granules is defined by the neighborhood relation. The process of constructing all the neighborhood information granules in the universe is called neighborhood information granulation, which can also be abbreviated as information granulation. Therefore, different information granulation processes and results are determined by different neighborhood relations. Then different neighborhood rough set models are generated.

Attribute reduction [9–12], as a feature selection technique [13,14], is one of the core research directions in rough sets. The exhaustive algorithm [15,16] and the heuristic algorithm [17,18] are two commonly used mechanisms for deriving attribute reduction. Compared with exhaustive algorithm, heuristic algorithm based on the greedy search strategy is more favored by many scholars because of its fast speed advantage. In heuristic algorithm, multiple iterations are required, the unselected attributes need to be evaluated separately in each iteration, and information granulation is performed in each evaluation. Unfortunately, the neighborhood information granulation is exceptionally time-consuming, resulting in low efficiency in attribute reduction. Therefore, how to accelerate attribute reduction is a subject that worth discussing. This has been noted by relevant scholars, and some valuable acceleration strategies have been proposed [19–21]. These strategies can be generally defined as two types: one is to reduce the frequency of information granulation during iteration [22,23]; the other is to accelerate the process of information granulation [24,25]. The research in this paper focuses on the latter, exploring ways to further improve the speed of information granulation from neighborhood relations, then accelerate the process of attribute reduction.

Most of the existing studies on neighborhood relations focus on issues such as the distribution and differences of samples in the universe, with little consideration of the impact on the efficiency of information granulation. For example, Hu et al. [26] introduced the  $k$ -nearest-neighbor relation, which addressed the problem of uneven distribution of samples in the universe and proposed the  $k$ -nearest-neighbor rough set. Wang et al. [27] integrated and analyzed the  $\delta$ -neighborhood relation [8] and the  $k$ -nearest-neighbor relation, proposed the  $k$ -nearest neighborhood relation, and induced the  $k$ -nearest neighborhood rough set. Yang et al. [28] considered that it is not enough to define the binary relation between samples only by distances, the label information between samples is also should be considered, so the pseudo-label neighborhood relation and pseudo-label neighborhood rough set are proposed. Zhou et al. [29] regarded that all existing neighborhood relations need to specify a parameter in advance, but it is a challenge to select unified and optimal parameters before learning for all different types of datasets. To fill such gap, the adaptive GAP neighborhood relation [29] was proposed, and the GAP neighborhood rough set was generated. Yang et al. [30] further improved the formula for calculating the distances between samples by using the distance metric learning (DMN), then a novel neighborhood rough set model based on DMN is proposed. Ba et al. [31] reinterpreted information granulation from a multigranulation view, and the Triple-G multigranulation rough set model was proposed by combining the granular ball relation [32], GAP neighborhood relation, and  $\delta$ -neighborhood relation.

Although the above neighborhood relations significantly improve the performance of neighborhood rough sets, but the process of their information granulations is still not efficient. Therefore, to further improve the efficiency of information granulation, based on the divide-and-conquer strategy,

a two-layer information granulation neighborhood rough set model is proposed, which is called the mapping-equivalence neighborhood rough set model. Two kinds of binary relations are used for the two-layer information granulation. The first layer of information granulation is based on our proposed mapping-equivalence relation, which aims to divide the universe into multiple sub-universes as shown in Fig. 1. The  $\delta$ -neighborhood relation is used for the second layer of information granulation, which aims to perform neighborhood information granulation as shown in Fig. 2. In this paper, the  $\delta$ -neighborhood relation is chosen because the value of  $\delta$  can be set to obtain different results of information granulations. This model can be described below:

- (1) Construct a mapping relation to generate a discrete mapping sample for each continuous sample. In Fig. 1,  $x_i$  ( $1 \leq i \leq n$ ) represents continuous samples, and  $x'_i$  represents discrete mapping samples.
- (2) According to the mapping-equivalence relation, divide the universe into mutually independent mapping-equivalence classes.
- (3) Establish mapping-equivalence relation between the continuous samples according to the equivalence relation of the mapping samples.
- (4) Treat each mapping-equivalence class as a sub-universe, and perform neighborhood information granulation. In Fig. 2,  $\delta(x_i)$  ( $1 \leq i \leq n$ ) represents the neighborhood information granules represented by  $x_i$ .
- (5) Merge all neighborhood information granules to complete the final information granulation.

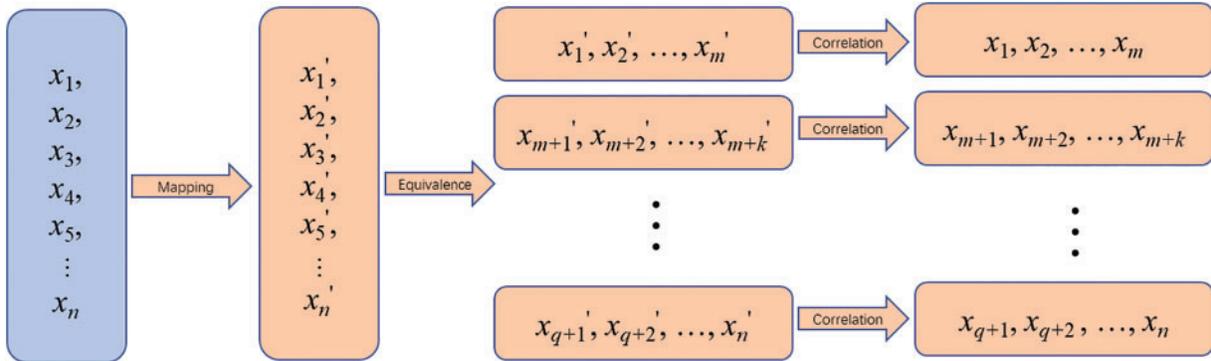


Figure 1: The first layer of information granulation

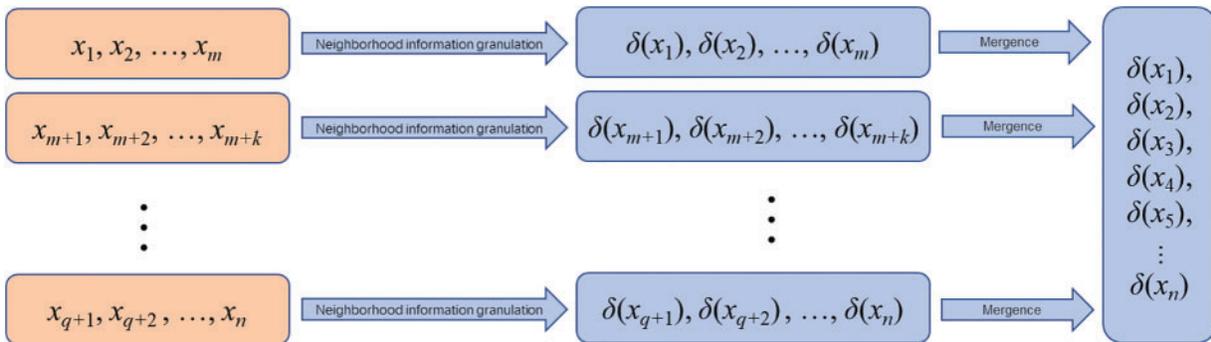
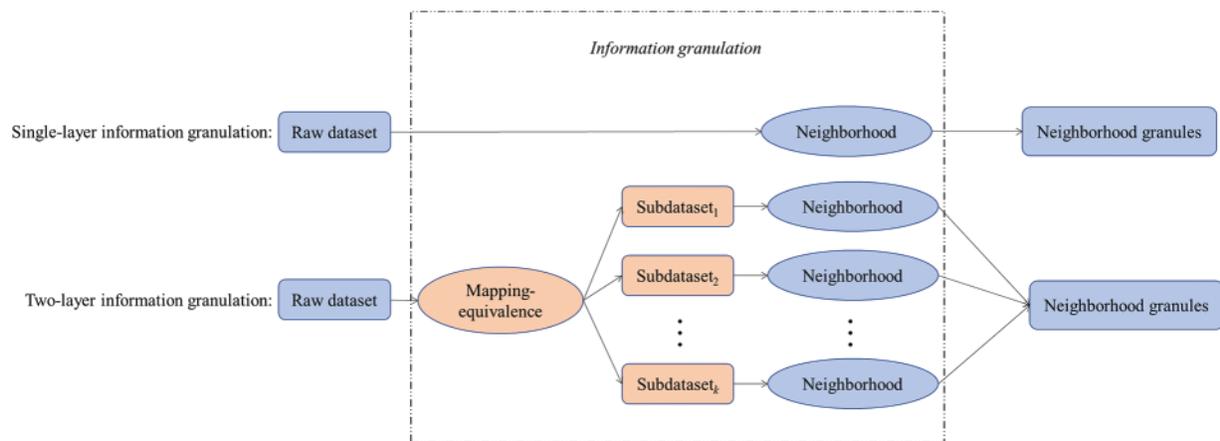


Figure 2: The second layer of information granulation

To further understand the whole process of two-layer information granulation more conveniently, a relevant illustration is given in Fig. 3. It is seen that the two-layer information granulation is more in line with the idea of divide-and-conquer in big data analysis than the single-layer information granulation which works directly on the raw dataset (universe). The raw dataset (universe) can be divided into multiple subdatasets (sub-universes) reasonably by the first layer of information granulation; the second layer of information granulation is performed on each divided subdataset (sub-universe) separately. Finally, all the information granules are merged to make up the final information granulation results. The universe with a large amount of data can reasonably be divided into several sub-universes to perform information granulation separately in this model. Thus, this model has higher efficiency of information granulation than other single-layer information granulation models. The higher the efficiency of information granulation, the less time consumption in each iteration of attribute reduction, and thus can accelerate attribute reduction.



**Figure 3:** Comparisons of different information granulation strategies

The remainder of this paper is organized as follows. In Section 2, we review the neighborhood rough set and discuss related concepts of reduct. In Section 3, we propose a two-layer information granulation strategy, from which the mapping-equivalence neighborhood rough set model is generated. Section 4 compares the proposed model with four commonly used models and gives experimental analyses to verify the effectiveness of the proposed model. Section 5 concludes the whole paper.

## 2 Preliminary Knowledge

This section will review some basic concepts of the neighborhood rough set and gives the definition of reduct.

Given a neighborhood decision system, it can be represented as  $NDS = \langle U, AT, d \rangle$ , in which  $U$  is a nonempty and finite set of samples, called the universe;  $AT$  is the set of condition attributes;  $d$  is a decision attribute.  $\forall x \in U$ ,  $d(x)$  denotes the decision attribute value of  $x$ ,  $\forall a \in AT$ ,  $a(x)$  denotes its value over condition attribute  $a$ .  $\forall x, y \in U$ , if  $d(x) = d(y)$ , then  $x$  and  $y$  satisfy the indistinguishable relation over  $d$ , referred to as  $IND_d = \{(x, y) \in U \times U \mid d(x) = d(y)\}$ . We can divide the  $U$  into decision classes by  $IND_d$ , denoted as  $U/IND_d = \{X_1, X_2, \dots, X_q\}$ , and generally abbreviated as  $U/d$ . If  $X_i \in U/d$ , then  $X_i$  is generally called the  $i$ th decision class. If sample  $x \in X_i$ , then  $X_i$  can be denoted by  $[x]_d$ . In a neighborhood rough set, the relation between two samples is determined by their distance. If

the distance is less than a certain threshold, we can say that the two samples satisfy the neighborhood relation. The formal definition of the neighborhood relation is as follows.

**Definition 1** [8] *Given an NDS,  $\delta$  is a given neighborhood radius,  $\forall A \subseteq AT$ , the neighborhood relation determined by  $A$  is defined as*

$$\delta_A = \{(x, y) \in U \times U \mid d_A(x, y) \leq \delta\}, \tag{1}$$

where  $d_A(x, y)$  is the distance between  $x$  and  $y$  over  $A$ . By  $\delta_A$ , the neighborhood of  $x$  over  $A$  is defined as

$$\delta_A(x) = \{y \in U \mid (x, y) \in \delta_A\}, \tag{2}$$

or

$$\delta_A(x) = \{y \in U \mid d_A(x, y) \leq \delta\}, \tag{3}$$

which is also called the neighborhood granule of  $x$ .

**Definition 2** [8] *Given an NDS,  $\delta$  is a given neighborhood radius,  $Uld = \{X_1, X_2, \dots, X_q\}$ ,  $\forall A \subseteq AT$ ,  $X_i \in Uld$ , the lower and upper approximations of  $X_i$  are defined as*

$$\underline{N}_A^\delta(X_i) = \{x \in U \mid \delta_A(x) \subseteq X_i\}, \tag{4}$$

$$\overline{N}_A^\delta(X_i) = \{x \in U \mid \delta_A(x) \cap X_i \neq \emptyset\}. \tag{5}$$

$\underline{N}_A^\delta(X_i)$  denotes the union of all neighborhood granules that can be contained by  $X_i$ ,  $\overline{N}_A^\delta(X_i)$  denotes the union of neighborhood granules that contain some elements in  $X_i$ .

To measure the contribution of condition attributes to the classification, many scholars have proposed multiple measures [33–36]. Among them, the approximation quality (AQ) [34] and the neighborhood mutual information (NMI) [35] are two commonly used measures, their formal definitions are given below.

**Definition 3** [34] *Given an NDS,  $\delta$  is a given neighborhood radius,  $Uld = \{X_1, X_2, \dots, X_q\}$ ,  $\forall A \subseteq AT$ , the approximation quality (AQ) is defined as*

$$AQ_\delta(A) = \frac{|\bigcup_{i=1}^q \underline{N}_A^\delta(X_i)|}{|U|}, \tag{6}$$

where  $|\cdot|$  represents the cardinality of set  $\cdot$ .

**Definition 4** [35] *Given an NDS,  $\delta$  is a given neighborhood radius,  $\forall A \subseteq AT$ , the neighborhood mutual information (NMI) about  $A$  and  $d$  is defined as*

$$NMI_\delta(A: d) = -\frac{1}{|U|} \sum_{x \in U} \log \frac{|\delta_A(x)| \cdot |[x]_d|}{|U| \cdot |\delta_A(x) \cap [x]_d|}. \tag{7}$$

If  $A$  refers to a subset of  $AT$ , both AQ and NMI can be used to measure the contribution of  $A$  to the classification, the greater the value of the measure, the greater the contribution of  $A$ .

Reduct is a special subset of  $AT$  with no weaker classification performance than  $AT$ . The process of acquiring a reduct is called attribute reduction, in which redundant attributes can be removed to obtain simplified decision rules.

**Definition 5** *Given an NDS, a measure  $\rho$ ,  $\forall A \subseteq AT$ , if the following conditions are satisfied*

$$(1) \rho(A) \geq \rho(AT),$$

$$(2) \forall a \in A, \rho(A - a) < \rho(A),$$

then  $A$  is called a reduct.

Based on the Def. 5, we can see that the reduct does not decrease the distinguishability of the system, and there are no redundant attributes in the reduct.

### 3 Mapping-Equivalence Neighborhood Rough Set Model

This section will describe the construction process of the mapping-equivalence neighborhood rough set model in detail. Because there are two layers of information granulation, it will be divided into two subsections for discussion. Subsection 3.1 discusses the mapping-equivalence relation. Subsection 3.2, the mapping-equivalence neighborhood rough set model is proposed, and then several related concepts concerning attribute reduction are also given.

#### 3.1 Mapping-Equivalence Relation

The first step in establishing the mapping-equivalence relation is to establish a mapping relation. Establishing the mapping relation depends on the relative distances between the samples and the center of each decision class. First of all, we will give the definition of the center of decision class.

**Definition 6** Given an NDS,  $Uld = \{X_1, X_2, \dots, X_q\}, \forall X_i \in Uld$ , the center of  $X_i$  is defined as

$$dc_i = \frac{1}{|X_i|} \sum_{x \in X_i} x. \quad (8)$$

$\forall x \in U$ , its mapping sample is defined as  $x' = (i, j)$ , where  $i = \operatorname{argmin} \left\{ \bigcup_{i=1}^q d(x, dc_i) \right\}$ ,  $j = \operatorname{argmax} \left\{ \bigcup_{j=1}^q d(x, dc_j) \right\}$ ,  $1 \leq i, j \leq q$ . That is,  $i$  represents the sequence number of the decision class closest to  $x$ , and  $j$  represents the sequence number of the decision class farthest from  $x$ .

All mapping samples are combined to form  $U'$ .  $\forall A \subseteq AT$ , if the mapping relation is established over  $A$ , then  $x'$  can be specially recorded as  $x'_A$ , and  $U'$  can be registered as  $U'_A$ . After all samples are mapped, the following mapping-equivalence relation can be defined.

**Definition 7** Given an NDS,  $\forall A \subseteq AT$ , the mapping-equivalence relation is represented as

$$IND_A^{ME} = \{(x, y) \in U \times U | x'_A = y'_A\}. \quad (9)$$

Obviously, the mapping-equivalence relation is transitive. That is,  $\forall x, y, z \in U$ , if  $x$  and  $y$  satisfy the mapping-equivalence relation,  $y$  and  $z$  also satisfy the mapping-equivalence relation, then  $x$  and  $z$  satisfy the mapping-equivalence relation. After the mapping-equivalence relation is defined, the universe can be granulated according to this mapping-equivalence relation, which is recorded as  $U/IND_A^{ME} = \{X_1^{ME}, X_2^{ME}, \dots, X_k^{ME}\}$ . The samples that satisfy the mapping-equivalence relation are in the same mapping-equivalence class, and  $X_i^{ME}$  represents the  $i$ th mapping-equivalence class. If a certain sample is used as a representative, the mapping-equivalence class can be expressed as follows.

**Definition 8** Given an NDS,  $\forall A \subseteq AT$ ,  $IND_A^{ME}, \forall x \in U$ , the mapping-equivalence class determined by  $x$  is

$$[x]_A^{ME} = \{y \in U | (x, y) \in IND_A^{ME}\}, \quad (10)$$

or

$$[x]_A^{ME} = \{y \in U | y'_A = x'_A\}. \quad (11)$$

*It is not difficult to draw the following conclusions:*

- (1)  $[x]_A^{ME} \neq \emptyset$ , because  $x \in [x]_A^{ME}$ ;
- (2)  $[x]_A^{ME} = [y]_A^{ME}$ , if  $y \in [x]_A^{ME}$ .

Assuming that there are  $n$  samples in  $U$ , the time complexity to calculate the center of the decision class is  $O(n)$ , and the time complexity from  $U$  to  $U'$  is also  $O(n)$ . Because the mapping-equivalence relation is transitive and the samples which satisfy the mapping-equivalence relation are in the same mapping-equivalence class, the time complexity of information granulation according to the mapping-equivalence relation is  $O(n \log n)$ . In summary, the time complexity of the first layer of information granulation is  $O(n \log n)$ .

### 3.2 Mapping-Equivalence Neighborhood Rough Set

After the first layer of information granulation is completed, the universe has been divided into multiple disjoint mapping-equivalence classes. If each mapping-equivalence class is regarded as a sub-universe, the second layer of information granulation is naturally induced.

**Definition 9** *Given an NDS,  $\delta$  is a given neighborhood radius,  $\forall A \subseteq AT$ , the mapping-equivalence neighborhood relation (ME-neighborhood relation) determined by  $A$  is defined as*

$$\delta_A^{ME} = \{(x, y) \in \text{IND}_A^{ME} \mid d_A(x, y) \leq \delta\}, \tag{12}$$

*$\forall X_i^{ME} \in U/\text{IND}_A^{ME}$ , if  $x \in X_i^{ME}$ , then the mapping-equivalence neighborhood of  $x$  over  $A$  is defined as*

$$\delta_A^{ME}(x) = \{y \in X_i^{ME} \mid (x, y) \in \delta_A^{ME}\}, \tag{13}$$

*or*

$$\delta_A^{ME}(x) = \{y \in X_i^{ME} \mid d_A(x, y) \leq \delta\}. \tag{14}$$

At this point, the proposed mapping-equivalence neighborhood rough set model completes the information granulation. Algorithm 1 presents this process in detail. Steps 1 to 12 of the algorithm are the first layer of information granulation. Through the analysis at the end of Subsection 3.1, we know that the time complexity is  $O(n \log n)$ ,  $n$  is the number of samples in the universe. Steps 13 to 19 are the second layer of information granulation. Suppose that step 12 divides the universe into  $k$  mapping-equivalence classes, and each mapping-equivalence class contains  $\frac{n}{k}$  samples on average. Then the time complexity of Steps 13 to 19 is  $k * O\left(\left(\frac{n}{k}\right)^2\right)$ . Therefore, the overall time complexity of Algorithm 1 is  $O(n \log n) + k * O\left(\left(\frac{n}{k}\right)^2\right)$ .

**Algorithm 1:** Two-Layer information granulation.

---

**Input:**  $\langle U, A \cup d \rangle, \delta$ ;  
**Output:** All information granules  $\delta_A^{ME}(U)$ ;

- 1  $U'_A \leftarrow \emptyset, \delta_A^{ME}(U) \leftarrow \emptyset$ ;
- 2  $U/d = \{X_1, X_2, \dots, X_q\}$ ;
- 3 **for**  $X_i \in U/d$  **do**
- 4    $dc_i = \frac{1}{|X_i|} \sum_{x \in X_i} x$
- 5 **end**
- 6 **for**  $x \in U$  **do**
- 7    $i = \operatorname{argmin} \left\{ \bigcup_{i=1}^q d(x, dc_i) \right\}$ ;
- 8    $j = \operatorname{argmax} \left\{ \bigcup_{j=1}^q d(x, dc_j) \right\}$ ;
- 9    $x'_A = (i, j)$ ;
- 10    $U'_A \leftarrow U'_A \cup x'_A$ ;
- 11 **end**
- 12  $U/\operatorname{IND}_A^{ME} = \{X_1^{ME}, X_2^{ME}, \dots, X_k^{ME}\}$ ;
- 13 **for**  $X_i^{ME} \in U/\operatorname{IND}_A^{ME}$  **do**
- 14   **for**  $\forall x \in X_i^{ME}$  **do**
- 15      $\delta_A^{ME}(x) = \{y \in X_i^{ME} | d_A(x, y) \leq \delta\}$ ;
- 16      $\delta_A^{ME}(U) \leftarrow \delta_A^{ME}(U) \cup \delta_A^{ME}(x)$ ;
- 17   **end**
- 18 **end**
- 19 **return**  $\delta_A^{ME}(U)$ .

---

After the information granulation is completed, the corresponding information granules can be obtained. Similar to the traditional neighborhood rough set model, this model also has the following definitions.

**Definition 10** Given an NDS,  $\delta$  is a given neighborhood radius,  $U/d = \{X_1, X_2, \dots, X_q\}, \forall A \subseteq AT, \forall X_i \in U/d$ , the lower and upper approximations of  $X_i$  based on the ME-neighborhood relation are defined as

$$\underline{N}_A^{ME}(X_i) = \{x \in U | \delta_A^{ME}(x) \subseteq X_i\}, \quad (15)$$

$$\overline{N}_A^{ME}(X_i) = \{x \in U | \delta_A^{ME}(x) \cap X_i \neq \emptyset\}. \quad (16)$$

**Definition 11** Given an NDS,  $\delta$  is a given neighborhood radius,  $U/d = \{X_1, X_2, \dots, X_q\}, \forall A \subseteq AT, \forall X_i \in U/d$ , the ME-approximation quality (AQ) is defined as

$$AQ_{ME}(A) = \frac{|\bigcup_{i=1}^q \underline{N}_A^{ME}(X_i)|}{|U|}. \quad (17)$$

**Definition 12** Given an NDS,  $\delta$  is a given neighborhood radius,  $\forall A \subseteq AT$ , the ME-neighborhood mutual information about  $A$  and  $d$  is defined as

$$NMI_{ME}(A: d) = -\frac{1}{|U|} \sum_{x \in U} \log \frac{|\delta_A^{ME}(x)| \cdot |[x]_d|}{|U| \cdot |\delta_A^{ME}(x) \cap [x]_d|}. \tag{18}$$

So far, the mapping-equivalence neighborhood rough set model has been constructed. To more accurately test the acceleration effect of the new proposed model on information granulation and attribute reduction, the forward greedy attribute reduction algorithm is adopted as one of the classical heuristics. A detailed description of the algorithm is given in Algorithm 2.

---

**Algorithm 2:** Searching the reduct of mapping-equivalence neighborhood rough set.

---

**Input:** NDS,  $\delta$ , measure  $\rho$ ;  
**Output:** A reduct  $red$ ;  
1  $red \leftarrow \emptyset, \rho(red) \leftarrow 0, R \leftarrow AT$ ;  
2 calculate  $\rho(AT)$ ;  
3 **while**  $\rho(red) < \rho(AT)$  **do**  
4  $\forall a \in R$ , perform information granulation over  $red \cup a$  by Algorithm 1 and calculate  $\rho(red \cup a)$ ;  
5 select an attribute  $b = argmax(\rho(red \cup a))$ ;  
6  $red = red \cup b, R = R - b$ ;  
7 **repeat**  
8  $\forall a \in red$ , perform information granulation over  $red - a$  by Algorithm 1 and calculate  $\rho(red - a)$ ;  
9 **if**  $\rho(red - a) \geq \rho(red)$   
10  $red = red - a$ ;  
11 **end**  
12 **until**  $red$  no longer changes  
13 **return**  $red$ .

---

## 4 Experiments

### 4.1 Description of the Datasets

To verify the effectiveness of the proposed strategy, 9 UCI datasets were selected to conduct the experiments. The basic information of the datasets is described in Table 1, in which ‘‘Sample’’ represents the cardinality of the universe, ‘‘Attribute’’ represents the number of all condition attributes, and ‘‘Class’’ represents the number of decision classes. It is worth mentioning that some datasets contain training samples and test samples, which we have merged, and some datasets have multiple decision attributes, which we only take the first one.

**Table 1:** Datasets description

ID	Dataset	Sample	Attribute	Class
1	Breast Cancer Wisconsin (Diagnostic)	569	31	2
2	Cardiotocography	2126	21	10
3	Car Evaluation	1728	6	4
4	Diabetic Retinopathy Debrecen Data Set	1151	19	2
5	Drug Consumption (quantified)	1885	12	7
6	QSAR Biodegradation	1055	41	2
7	Steel Plates Faults	1941	27	2

(Continued)

**Table 1:** Continued

ID	Dataset	Sample	Attribute	Class
8	Waveform Database Generator (Version 1)	5000	21	3
9	Yeast	1484	8	10

## 4.2 Experimental Comparisons

In this subsection, the mapping-equivalence neighborhood rough set (MENRS) is compared with other neighborhood rough sets. Because the  $\delta$ -neighborhood relation is used in proposed model, some neighborhood rough set models based on  $\delta$ -neighborhood relation are selected, including the  $k$ -nearest neighborhood rough set (KNRS) [27], pseudo-label neighborhood rough set (PLNRS) [28], Triple-G multigranulation rough set (TMGRS) [31] and  $\delta$ -neighborhood rough set (DNRS) [8]. The KNRS, PLNRS and TMGRS were introduced in Section 1, the DNRS was reviewed in Section 2. AQ and NMI will be employed as the measures for searching reducts. It should be noted that NMI cannot be applied to TMGRS because TMGRS does not have a definite neighborhood relation, so TMGRS is only involved in the comparison experiments using AQ as the measure.

Both experiments are implemented by using MATLAB R2020b and conducted on a laptop with Intel Intel(R) Core(TM) i7-8750H CPU @ 2.20 GHz 2.21 GHz and 8 GB memory. To facilitate unified processing, all datasets are normalized to [0,1] in advance. Taking into account that when the value of  $\delta$  is large, all the rough set model reduction results tend to be homogeneous. Therefore, to facilitate experimental comparison, 10 different  $\delta$  (0.03, 0.06, ..., 0.3) have been selected. The 5-fold cross-validation strategy will be used to ensure the reliability of the experimental results. In the tables, the data with better performance will be bolded for easy observation.

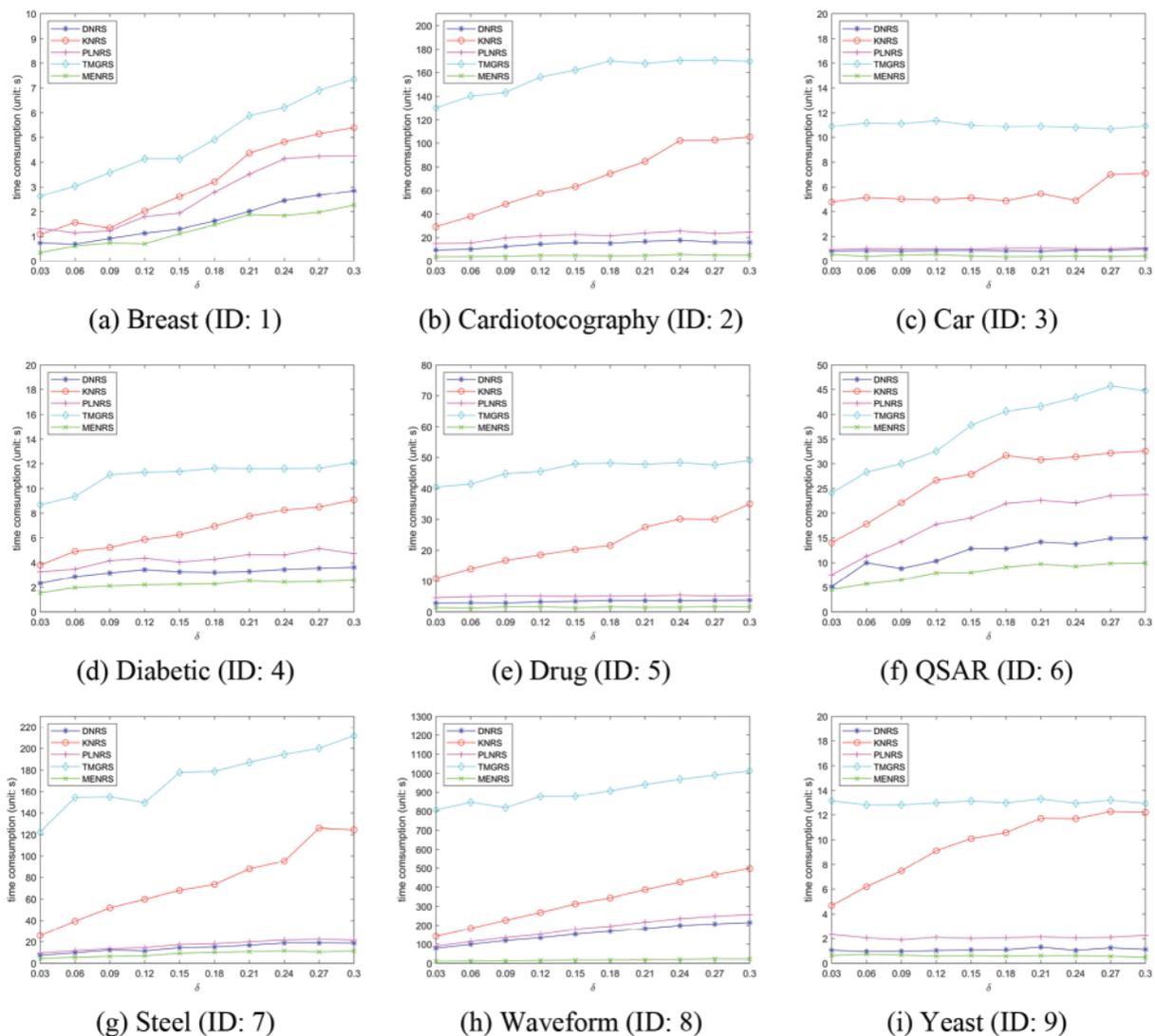
### 4.2.1 Comparisons of Time Consumption

In this subsection, we will take note of the time consumption of five rough set models for the two measures considered in this paper. Detailed results are recorded in Figs. 4 and 5. The abscissa is the selected 10 different  $\delta$ , and the ordinate is the time consumption of the attribute reduction. Taking a closer inspection and comparison, we can conclude the following.

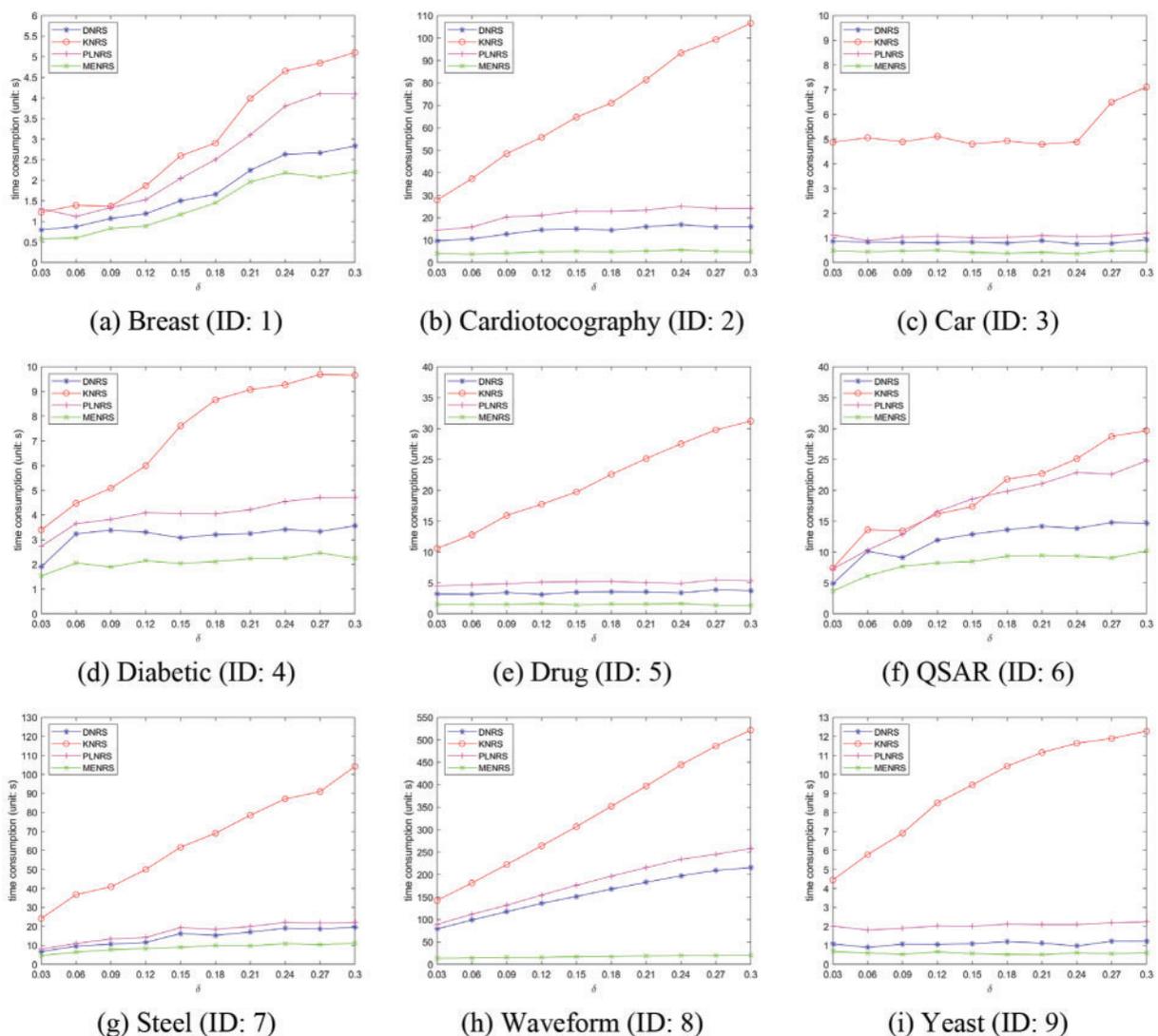
- (1) With the increase of  $\delta$ , the time consumption of attribute reduction also tends to increase. This is reflected in almost all datasets. The larger the  $\delta$ , the more condition attributes need to be selected, so the number of iterations of the algorithm increases and the time consumption increases.
- (2) MENRS does allow for faster information granulation. In almost all datasets, MENRS has obvious advantages in terms of time consumption. Further observation of Car (ID: 3), Waveform (ID: 8) and Yeast (ID: 9) shows that MENRS has more obvious advantages on datasets with larger amounts of data and more decision classes. Conversely, on a dataset with a smaller amount of data and fewer decision classes, the advantage of MENRS is less obvious. This is mainly because when the total number of samples is large, MENRS can effectively divide the universe into mapping-equivalence classes with a smaller total number of samples, so the efficiency is higher. The more decision classes, the more mapping-equivalence classes obtained by the first layer information granulation of MENRS, and the smaller the total number of samples in a single mapping-equivalence class, so the information granulation is faster.

### 4.2.2 Comparisons of Classification Accuracy

In this subsection, we will take the reducts from the five neighborhood rough sets respectively, use the CART [37] classifier and the KNN (K = 5) [38] classifier for classification. The column AQ represents the classification accuracy of the reducts obtained with the AQ as the measure, and the column NMI represents the classification accuracy of the reducts obtained with the NMI as the measure. 10  $\delta$  values were used to obtain the reducts and classifications respectively, and the final result was averaged. Detailed comparison results are reported in Tables 2 and 3.



**Figure 4:** Time consumption by using AQ



**Figure 5:** Time consumption by using NMI

**Table 2:** Classification accuracy by using the CART classifier

ID	AQ					NMI			
	DNRS	KNRS	PLNRS	TMGRS	MENRS	DNRS	KNRS	PLNRS	MENRS
1	0.9329	0.9269	0.9329	0.9307	<b>0.9340</b>	0.9341	0.9297	<b>0.9348</b>	0.9333
2	0.8085	0.8085	0.8055	<b>0.8113</b>	0.8109	0.8077	0.8051	0.8081	<b>0.8084</b>
3	0.9549	0.9549	0.9549	0.9549	0.9549	0.9549	0.9549	0.9549	0.9549
4	<b>0.6276</b>	0.6271	0.6271	0.6234	0.6249	0.6270	<b>0.6286</b>	0.6269	0.6262
5	<b>0.2946</b>	<b>0.2946</b>	0.2929	0.2933	0.2945	0.2940	0.2937	0.2934	<b>0.2965</b>

(Continued)

**Table 2:** Continued

ID	AQ					NMI			
	DNRS	KNRS	PLNRS	TMGRS	MENRS	DNRS	KNRS	PLNRS	MENRS
6	0.8001	0.8062	0.8051	<b>0.8067</b>	0.8058	0.7997	<b>0.8033</b>	0.8007	0.8015
7	0.9167	0.9174	0.9171	<b>0.9252</b>	0.9150	0.9131	0.9123	0.9132	<b>0.9161</b>
8	0.7278	0.7278	0.7294	<b>0.7546</b>	0.7265	0.7324	0.7330	0.7356	<b>0.7458</b>
9	<b>0.5262</b>	<b>0.5262</b>	0.5258	0.5244	0.5214	0.5274	0.5260	<b>0.5282</b>	0.5236

**Table 3:** Classification accuracy by using the KNN classifier

ID	AQ					NMI			
	DNRS	KNRS	PLNRS	TMGRS	MENRS	DNRS	KNRS	PLNRS	MENRS
1	0.9594	0.9608	0.9613	<b>0.9627</b>	0.9585	<b>0.9610</b>	0.9607	0.9584	0.9607
2	0.7564	0.7564	0.7573	<b>0.7706</b>	0.7547	0.7574	0.7567	<b>0.7590</b>	0.7548
3	0.9340	0.9340	0.9340	0.9340	0.9340	0.9340	0.9340	0.9340	0.9340
4	<b>0.6269</b>	<b>0.6269</b>	0.6259	0.6263	0.6265	0.6268	<b>0.6277</b>	0.6270	0.6270
5	<b>0.3326</b>	<b>0.3326</b>	0.3324	0.3285	0.3320	0.3317	0.3326	0.3315	<b>0.3340</b>
6	0.8444	0.8459	0.8427	<b>0.8487</b>	0.8444	0.8436	0.8441	0.8417	<b>0.8487</b>
7	0.9270	0.9278	0.9258	<b>0.9331</b>	0.9273	0.9269	0.9264	<b>0.9278</b>	0.9262
8	0.7798	0.7798	<b>0.7827</b>	<b>0.8126</b>	0.7797	0.7822	0.7816	0.7847	<b>0.8003</b>
9	<b>0.5596</b>	<b>0.5596</b>	0.5586	0.5545	0.5556	<b>0.5596</b>	<b>0.5596</b>	0.5588	0.5528

Taking a close observation of [Table 2](#), it can be drawn that the classification performance of the reducts taken from the five models is comparable. Across the 9 datasets, the models do not differ much in performance. For example, on dataset Breast (ID: 1), PLNRS and MENRS perform best, and on datasets such as Diabetic (ID: 4) and Yeast (ID: 9), DNRS, KNRS and TMGRS are better. This is because although the neighborhood rough sets are different, the condition attributes selected based on the attribute reduction algorithm are similar, so the obtained reducts are not much different. Looking at [Table 3](#), the same conclusion as that of [Table 2](#) can be drawn. Therefore, it can be concluded that MENRS performs well and is not weaker than other neighborhood rough sets in terms of classification accuracy.

#### 4.2.3 Comparisons of Information Granulation Results

In this subsection, we will use the five neighborhood rough sets for information granulation of the universe separately and use two measures, AQ and NMI, to evaluate the results of information granulation. The AQ column represents the results obtained by using AQ as the measure, and the NMI column represents the results measured by NMI. When  $\delta$  is small, the neighborhood members of the sample may be few, which is inconvenient for reflecting the differences between the models. Therefore, for convenience of observation, take the maximum value of 0.3 among the 10  $\delta$  for comparison.

[Table 4](#) is the results of the information granulation on the raw data, that is, the universe before attribute reduction. With a careful investigation of [Table 4](#), it is not difficult to observe that

information granulation in MENRS is more likely to reduce system uncertainty, and it performs the best in almost all datasets, followed by PLNRS, DNRS, KNRS and TMGRS. Take Cardiotocography (ID: 2) as an example, using AQ as a measure, MENRS = 0.6326, PLNRS = 0.6008, DNRS = KNRS = 0.5667, TMGRS = 0.5031. This is because the first layer of information granulation of MENRS divides the more similar samples into the same mapping-equivalence class. Therefore, in the second layer of information granulation, the samples in the neighborhood granule are more relevant. The pseudo-label strategy of PLNRS also screens the samples, but the pseudo-labels are generated by clustering, the clustering results are unstable, and the relations between samples are also unstable. Therefore, the quality of its information granulation is weaker than MENRS, but better than that of DNRS, KNRS and TMGRS.

**Table 4:** Measure of raw data

ID	AQ					NMI			
	DNRS	KNRS	PLNRS	TMGRS	MENRS	DNRS	KNRS	PLNRS	MENRS
1	0.9429	0.9429	0.9429	0.9253	<b>0.9473</b>	0.6429	0.6429	0.6429	<b>0.6445</b>
2	0.5667	0.5667	0.6008	0.5031	<b>0.6326</b>	1.7361	1.7361	1.7514	<b>1.7786</b>
3	0.7337	0.7337	0.7808	0.7377	<b>0.8300</b>	0.7596	0.7596	0.7727	<b>0.7858</b>
4	0.1010	0.1010	0.1010	0.0934	<b>0.1129</b>	0.1324	0.1383	0.1324	<b>0.1373</b>
5	0.4682	0.4682	0.4847	0.1875	<b>0.5245</b>	1.0284	1.0284	1.0584	<b>1.0953</b>
6	0.5829	0.5829	0.5829	0.5265	<b>0.6126</b>	0.4581	0.4594	0.4581	<b>0.4710</b>
7	0.9395	0.9395	0.9395	0.8925	<b>0.9420</b>	0.2481	0.2481	0.2481	<b>0.2489</b>
8	0.9970	0.9970	0.9970	0.7703	<b>0.9975</b>	1.0966	1.0966	1.0966	<b>1.0970</b>
9	0.0185	0.0185	0.0194	0.0116	<b>0.0219</b>	0.5827	0.7149	0.6633	<b>0.7171</b>

Table 5 is the results of the information granulation on the universe after attribute reduction. With a careful investigation of Table 5, it is not difficult to observe that information granulation in MENRS is more likely to reduce system uncertainty, performing optimally in almost all datasets, followed by PLNRS, and then DNRS, KNRS and TMGRS in that order, which is consistent with the observations in Table 4. Combining Tables 4 and 5, we can draw the conclusion that MENRS can better information granulation than other neighborhood rough set models.

**Table 5:** Measure of reduction

ID	AQ					NMI			
	DNRS	KNRS	PLNRS	TMGRS	MENRS	DNRS	KNRS	PLNRS	MENRS
1	0.9429	0.9429	0.9429	0.9253	<b>0.9495</b>	0.643	0.6429	0.6429	<b>0.6522</b>
2	0.5667	0.5667	0.6008	0.5038	<b>0.6326</b>	1.7361	1.7361	1.7514	<b>1.7786</b>
3	0.7337	0.7337	0.7808	0.7377	<b>0.8300</b>	0.7596	0.7596	0.7727	<b>0.7858</b>
4	0.1010	0.1010	0.1010	0.0934	<b>0.1129</b>	0.1324	0.1383	0.1324	<b>0.1373</b>
5	0.4682	0.4682	0.4847	0.1875	<b>0.5245</b>	1.0284	1.0284	1.0584	<b>1.0953</b>
6	0.5829	0.5829	0.5829	0.5273	<b>0.6149</b>	0.4584	0.4598	0.4584	<b>0.4710</b>

(Continued)

**Table 5:** Continued

ID	AQ					NMI			
	DNRS	KNRS	PLNRS	TMGRS	MENRS	DNRS	KNRS	PLNRS	MENRS
7	0.9395	0.9395	0.9395	0.8927	<b>0.9420</b>	0.2481	0.2481	0.2488	<b>0.2498</b>
8	0.9970	0.9970	0.9970	0.7732	<b>0.9975</b>	1.0966	1.0966	1.0966	<b>1.0970</b>
9	0.0185	0.0185	0.0194	0.0116	<b>0.0219</b>	0.5827	0.7149	0.6633	<b>0.7171</b>

## 5 Conclusions

In the field of neighborhood rough set research, the speed of attribute reduction is largely influenced by the efficiency of information granulation. The way of information granulation was determined by neighborhood relations. Therefore, a new research has been done on neighborhood relations and the mapping-equivalence neighborhood rough set model is generated.

The main contributions of this paper are as follows: first, the mapping-equivalence relation is proposed, which is established between continuous data. Second, the two-layer information granulation strategy is proposed and generates the mapping-equivalence neighborhood relation by combining the existing neighborhood relations. Finally, the mapping-equivalence neighborhood rough set model is naturally deduced from the mapping-equivalence neighborhood relation, which provides a new thinking for the expansion of neighborhood rough set models. The experimental results demonstrate the effectiveness of the proposed model.

Albeit the good performance of the proposed model, we acknowledge its limitations which motivate our more investigations in this direction. First, when the amount of data is relatively small, the advantages of the proposed model are not very obvious. Besides, the second layer of information granulation in this paper only uses the  $\delta$ -neighborhood relation, while do not take other neighborhood relations into consideration. Finally, there are some more intensive studies that were not conducted. Thus, the following topics deserve further research:

- (1) Improve the mapping-equivalence relation for better information granulation performance.
- (2) Enhance the extensibility of the proposed model by using other neighborhood relations.
- (3) Extend two-layer information granulation strategy to multiple layers should also be explored.

**Funding Statement:** This work is supported by the National Natural Science Foundation of China (Nos. 62006099, 62076111), and the Key Laboratory of Oceanographic Big Data Mining & Application of Zhejiang Province (No. OBDMA202104).

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] Z. Pawlak, "Rough sets," *International Journal of Computer & Information Sciences*, vol. 11, pp. 341–356, 1982.
- [2] D. Q. Miao, Q. G. Duan, H. Y. Zhang and N. Jiao, "Rough set based hybrid algorithm for text classification," *Expert Systems with Applications*, vol. 36, no. 5, pp. 9168–9174, 2009.
- [3] L. Ezhilarasi, A. P. Shanthi and V. Uma Maheswari, "Rough set based rule approximation and application on uncertain datasets," *Intelligent Automation & Soft Computing*, vol. 26, no. 3, pp. 465–478, 2020.

- [4] W. T. Li and W. H. Xu, "Double-quantitative decision-theoretic rough set," *Information Sciences*, vol. 316, pp. 54–67, 2015.
- [5] B. J. Fan, E. C. C. Tsang, W. H. Xu and J. H. Yu, "Double-quantitative rough fuzzy set based decisions: A logical operations method," *Information Sciences*, vol. 378, pp. 264–281, 2017.
- [6] X. Zhang, J. G. Luo, X. M. Sun and J. C. Xie, "Optimal reservoir flood operation using a decomposition-based multi-objective evolutionary algorithm," *Engineering Optimization*, vol. 51, no. 1, pp. 42–62, 2019.
- [7] K. H. Yuan, W. H. Xu, W. T. Li and W. P. Ding, "An incremental learning mechanism for object classification based on progressive fuzzy three-way concept," *Information Sciences*, vol. 584, pp. 127–147, 2022.
- [8] Q. H. Hu, D. R. Yu and Z. X. Xie, "Numerical attribute reduction based on neighborhood granulation and rough approximation," *Journal of Software*, vol. 19, no. 3, pp. 640–649, 2008.
- [9] Y. Y. Yang, D. G. Chen and H. Wang, "Active sample selection based incremental algorithm for attribute reduction with rough sets," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 4, pp. 825–838, 2017.
- [10] E. C. C. Tsang, D. G. Chen, D. S. Yeung, X. Z. Wang and J. W. T. Lee, "Attributes reduction using fuzzy rough sets," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 5, pp. 1130–1141, 2008.
- [11] M. Hu, E. C. C. Tsang, Y. T. Guo and W. H. Xu, "Fast and robust attribute reduction based on the separability in fuzzy decision systems," *IEEE Transactions on Cybernetics*, vol. 52, no. 6, pp. 5559–5572, 2022.
- [12] H. R. Ju, W. P. Ding, Z. Q. Shi, J. S. Huang, J. Yang *et al.*, "Attribute reduction with personalized information granularity of nearest mutual neighbors," *Information Sciences*, vol. 613, pp. 114–138, 2022.
- [13] U. Kanimozhi and D. Manjula, "An intelligent incremental filtering feature selection and clustering algorithm for effective classification," *Intelligent Automation & Soft Computing*, vol. 24, no. 4, pp. 701–709, 2018.
- [14] B. B. Sang, H. M. Chen, L. Yang, T. R. Li and W. H. Xu, "Incremental feature selection using a conditional entropy based on fuzzy dominance neighborhood rough sets," *IEEE Transactions on Fuzzy Systems*, vol. 30, no. 6, pp. 1683–1697, 2022.
- [15] Y. Y. Yao and Y. Zhao, "Discernibility matrix simplification for constructing attribute reducts," *Information Sciences*, vol. 179, no. 7, pp. 867–882, 2009.
- [16] X. B. Yang, Y. S. Qi, X. N. Song and J. Y. Yang, "Test cost sensitive multigranulation rough set: Model and minimal cost selection," *Information Sciences*, vol. 250, pp. 184–199, 2013.
- [17] X. Y. Zhang, H. Yao, Z. Y. Lv and D. Q. Miao, "Class-specific information measures and attribute reducts for hierarchy and systematicness," *Information Sciences*, vol. 563, pp. 196–225, 2021.
- [18] W. S. Du and B. Q. Hu, "A fast heuristic attribute reduction approach to ordered decision systems," *European Journal of Operational Research*, vol. 264, no. 2, pp. 440–452, 2018.
- [19] Y. Chen, X. B. Yang, J. H. Li, P. X. Wang and Y. H. Qian, "Fusing attribute reduction accelerators," *Information Sciences*, vol. 587, pp. 354–370, 2022.
- [20] Z. H. Jiang, X. B. Yang, H. L. Yu, D. Liu, P. X. Wang *et al.*, "Accelerator for multi-granularity attribute reduction," *Knowledge-Based Systems*, vol. 177, pp. 145–158, 2019.
- [21] Z. H. Jiang, K. Y. Liu, X. B. Yang, H. L. Yu, H. Fujita *et al.*, "Accelerator for supervised neighborhood based attribute reduction," *International Journal of Approximate Reasoning*, vol. 119, pp. 122–150, 2020.
- [22] Y. Chen, K. Y. Liu, J. J. Song, H. Fujita, X. B. Yang *et al.*, "Attribute group for attribute reduction," *Information Sciences*, vol. 535, pp. 64–80, 2020.
- [23] X. S. Rao, X. B. Yang, X. Yang, X. J. Chen, D. Liu *et al.*, "Quickly calculating reduct: An attribute relationship based approach," *Knowledge-Based Systems*, vol. 200, pp. 106014, 2020.
- [24] Y. Liu, W. L. Huang, Y. L. Jiang and Z. Y. Zeng, "Quick attribute reduct algorithm for neighborhood rough set model," *Information Sciences*, vol. 271, pp. 65–81, 2014.
- [25] Z. Chen, K. Y. Liu, X. B. Yang and H. Fujita, "Random sampling accelerator for attribute reduction," *International Journal of Approximate Reasoning*, vol. 140, pp. 75–91, 2022.
- [26] Q. H. Hu, J. F. Liu and D. R. Yu, "Mixed feature selection based on granulation and approximation," *Knowledge-Based Systems*, vol. 21, no. 4, pp. 294–304, 2008.

- [27] C. Z. Wang, Y. P. Shi, X. D. Fan and M. W. Shao, "Attribute reduction based on k-nearest neighborhood rough sets," *International Journal of Approximate Reasoning*, vol. 106, pp. 18–31, 2019.
- [28] X. B. Yang, S. C. Liang, H. L. Yu, S. Gao and Y. H. Qian, "Pseudo-label neighborhood rough set: Measures and attribute reductions," *International Journal of Approximate Reasoning*, vol. 105, pp. 112–129, 2019.
- [29] P. Zhou, X. G. Hu, P. P. Li and X. D. Wu, "Online streaming feature selection using adapted neighborhood rough set," *Information Sciences*, vol. 481, pp. 258–279, 2019.
- [30] X. L. Yang, H. M. Chen, T. R. Li, J. H. Wan and B. B. Sang, "Neighborhood rough sets with distance metric learning for feature selection," *Knowledge-Based Systems*, vol. 224, pp. 107076, 2021.
- [31] J. Ba, K. Y. Liu, H. R. Ju, S. P. Xu, T. H. Xu *et al.*, "Triple-G: A new MGRS and attribute reduction," *International Journal of Machine Learning and Cybernetics*, vol. 13, pp. 337–356, 2022.
- [32] S. Y. Xia, Y. S. Liu, X. Ding, G. Y. Wang, H. Yu *et al.*, "Granular ball computing classifiers for efficient, scalable and robust learning," *Information Sciences*, vol. 483, pp. 136–152, 2019.
- [33] Y. B. Wang, X. J. Chen and K. Dong, "Attribute reduction via local conditional entropy," *International Journal of Machine Learning and Cybernetics*, vol. 10, pp. 3619–3634, 2019.
- [34] Q. H. Hu, D. R. Yu, J. F. Liu and C. X. Wu, "Neighborhood rough set based heterogeneous feature subset selection," *Information Sciences*, vol. 178, no. 18, pp. 3577–3594, 2008.
- [35] Q. H. Hu, L. Zhang, D. Zhang, W. Pan, S. An *et al.*, "Measuring relevance between discrete and continuous features based on neighborhood mutual information," *Expert Systems with Applications*, vol. 38, no. 9, pp. 10737–10750, 2011.
- [36] Q. Wan, J. H. Li, L. Wei and T. Qian, "Optimal granule level selection: A granule description accuracy viewpoint," *International Journal of Approximate Reasoning*, vol. 116, pp. 85–105, 2020.
- [37] C. Iorio, M. Aria, A. D'Ambrosio and R. Siciliano, "Informative trees by visual pruning," *Expert Systems with Applications*, vol. 127, pp. 228–240, 2019.
- [38] S. C. Zhang, X. L. Li, M. Zong, X. F. Zhu and R. L. Wang, "Efficient kNN classification with different numbers of nearest neighbors," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 5, pp. 1774–1785, 2018.