International Journal of Approximate Reasoning 164 (2024) 109062

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

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

A novel approach to discretizing information systems associated with neighborhood rough sets

Di Zhang [a], Ping Zhu [a,b,*]

[a] School of Science, Beijing University of Posts and Telecommunications, Beijing 100876, China
[b] Key Laboratory of Mathematics and Information Networks, Beijing University of Posts and Telecommunications, Ministry of Education, China

ARTICLE INFO

Keywords:
Neighborhood rough sets
Attribute reduction
Discretization
Granular computing

ABSTRACT

By effectively capturing the similarity relationship between objects, neighborhood rough set models offer a highly advantageous approach to analyzing numerical data. However, it is computationally demanding to construct granular structures by calculating the distance between all objects in the universe. In this paper, we propose a novel approach to discretizing information systems associated with neighborhood rough sets. Firstly, a unique bilayer, quasi discretization-based neighborhood, is developed with the aim of minimizing the time required for model generation. Most of the current evaluation indices for a rough set model primarily focus on the evaluation of the lower and upper approximations, rather than the granular structure of the model. We introduce two evaluation indices for binary relations, namely, the Gaussian balance index and the quality index. These indices aim to provide a comprehensive assessment of the granular structure in terms of quantity and quality. Moreover, we present two extension models for discretization neighborhood rough sets and devise three attribute reduction algorithms. In order to evaluate the performance of the models, three comparative experiments are conducted. The results obviously reveal the high efficiency and effectiveness of the proposed models.

1. Introduction

In recent years, there has been a notable surge in the volume of data, which has generated considerable attention towards the efficient extraction and exploration of the information embedded within the data [15]. Granular computing [26] is being increasingly acknowledged as an essential method for information processing, with rough sets [33] serving as a key representative. Rough sets have found extensive applications in various domains such as decision-making [24,59], attribute reduction [5,22], uncertain description [9], and other related areas [10,53]. Through the utilization of equivalence relations, rough sets have the ability to partition structured data (i.e., tabular data) into positive region, boundary region, and negative region. The samples located within the boundary region are considered to be uncertain data. Rough set theory can partition data efficiently, but it cannot deal with continuous data, such as human height and weight. Consequently, numerous scholars have conducted extensive research on rough sets. Advances in rough sets emerged chronologically later than fuzzy sets. To address the challenges related to partitioning in the fuzzy systems, Dubois et al. proposed fuzzy rough sets and rough fuzzy sets [11]. Based on the cost (risk) theory, Yao designed a decision-theoretic rough set model [54]. Ziarko et al. constructed the variable precision rough set model, which can tolerate some error rate in the data partitioning process [60]. Several novel rough set models have been developed to address the limitation of

* Corresponding author at: School of Science, Beijing University of Posts and Telecommunications, Beijing 100876, China.
E-mail addresses: zhangd@bupt.edu.cn (D. Zhang), pzhubupt@bupt.edu.cn (P. Zhu).

https://doi.org/10.1016/j.ijar.2023.109062
Received 31 August 2023; Received in revised form 7 October 2023; Accepted 15 October 2023

Available online 20 October 2023

rough sets in handling discrete data exclusively. These models have extended the traditional equivalence relations into tolerance relations [30], dominance relations [13,14], (α, β)-indiscernibility relations [55], and neighborhood relations [18,19] to enhance the applicability of rough sets in handling various types of data.

Neighborhood rough set models have gained considerable attention in academia because of their ability to approximate data and represent similarity among samples by utilizing neighborhood relations. During the past decade, many innovative models have been proposed and employed to address various practical issues. Wang et al. constructed a fuzzy neighborhood rough set model by defining the fuzzy decision of a sample based on fuzzy neighborhood [44]. Noting that uncertainty plays a central role in the field of rough set research [58], Chen et al. investigated several uncertainty measures of neighborhood granules [6] and Xu et al. analyzed two self-information uncertainty measures based on neighborhood information systems [48]. To resolve the limitation of rough sets in handling data with limited labels, Wang et al. conducted a study on the local neighborhood rough set model. This model establishes a connection between neighborhood rough sets and local rough sets [46]. Combining algebraic and information-theoretic perspectives, Sun et al. explored entropy measures utilizing neighborhood rough sets [41]. By exploring the superiorities of both neighborhood and k-nearest neighbor relations, Wang et al. proposed the concept of k-nearest neighborhood rough sets to process attribute reduction [45].

Neighborhood systems are one of the objects studied in the field of neighborhood rough set theory [26]. Yang et al. re-analyzed neighborhood systems from a granular computing perspective and demonstrated that the representation of the variable precision rough set model can be achieved through neighborhood systems [50]. Yang et al. investigated the general structure of neighborhood systems under incomplete information conditions [51]. D'eer et al. introduced fuzzy neighborhood systems based on a given fuzzy covering, and defined the concepts of fuzzy minimal and fuzzy maximal descriptions of objects [12]. In an information system, features may be derived from various levels or aspects, thus requiring the assignment of variable weights to different attributes [16]. Sun et al. carried out label correlations to determine the weight of every feature and then used the weights for feature selection [38]. Considering the divergent impacts of variable attributes on feature selection, Pan et al. proposed the weighted dominance-based neighborhood rough sets [32]. Xie et al. considered the weights of features through data binning method and information entropy theory and designed a weighted neighborhood probabilistic rough set model [47]. Some semi-supervised feature selection algorithms are designed to select attribute subsets for partially labeled mixed-type data [20,37]. As a tool to describe uncertainty, neighborhood rough sets have been widely used in gene selection [8,39], spectral analysis [56], credit scoring [35,43], green supplier development (GSD) [2], prediction of solar activity [1], and other fields.

We realize that determining the optimal radius in a neighborhood rough set has sparked considerable debate among scientists. The conventional approach of utilizing a fixed radius in the vanilla neighborhood method neglects to account for the impact of labels. Consequently, Yang et al. combined neighborhood rough set theory and pseudo-label strategy to divide samples by utilizing both distance and pseudo-label as criteria [52]. Another approach, presented by Zhang et al., involves estimating a surrounding function determined by the label distribution to automatically generate the radius of each object [57]. Barnali et al. [3] determined the suitable neighborhood parameter by observing the maximum rate of changes in the boundary region. In order to avoid the unwarranted selection of neighborhood radius by experiments or expert judgment, Sun et al. conducted an adaptive function to determine the appropriate neighborhood radius [40].

Attribute reduction is an essential research topic of rough set theory. With the advent of the big data era, the computing power strain caused by high complexity of attribute reduction based on neighborhood rough sets has garnered significant attention. To reduce the running time of the algorithms, several approaches have been designed to address this issue, including the fish swarm algorithms [7,61], parallel attribute reduction algorithm [4], and runner-root algorithm [21] based on neighborhood rough sets. However, these efforts have proven insufficient in addressing the core issue. Fundamentally altering the granular structure of the neighborhood is necessary, and hence many methods have been proposed to deal with this prevailing concern. Liu et al. divided the data into a series of buckets and then calculated distances between samples within the same or adjacent buckets [28]. Liu et al. proposed a simple scheme, called the granular cabin, which aims to significantly enhance the efficiency of neighborhood granular structures [27]. However, these approaches may divide regions too intricately, and some regions contain a lot of samples, resulting in a wastage of computational resources. Rough set models are not capable of directly handling numerical data [36,42]. In additional, the researchers utilized discretization to separate continuous data [23,31,49], leading to the potential loss of information between samples.

In this paper, we estimate several new discretization neighborhood rough set models aimed at reducing the time required for algorithm generation and improving the representation of information between samples by integrating discretization techniques with neighborhood rough set theory. Firstly, we discretize the data into block structures, called information blocks. Secondly, we calculate the distance between elements in the same block and adjacent blocks. Since the calculation of distances between samples only needs to be performed within a subset of the universe, rather than the entire universe, the generation time of the algorithm can be significantly reduced. Then, two evaluation methods for binary relations are provided. We compare our model with some existing models based on these evaluation indices. Subsequently, we extend our model and design two new models: the weighted discretization neighborhood rough set model (WDN) and the fast discretization neighborhood rough set model (FDN). These two models play essential roles in adapting to data set distribution and accelerating model running time. Finally, we propose three attribute reduction algorithms corresponding to each of the three proposed models.

Our contributions can be summarized as follows:

(1) By combining discretization theory with the neighborhood rough set model, we conduct an in-depth study of discretization neighborhood rough set model. This approach successfully decreases the time required for model generation.

(2) The two proposed evaluation indices for binary relations offer a novel approach to evaluating the effectiveness of rough set models.
(3) We study a weighted quasi discretization neighborhood rough set model, as well as a faster generation method for discretization neighborhood granular structures. These theories have the potential to offer a fresh outlook on the examination of other rough set models.

The structure of the paper is shown as follows: In Section 2, we introduce neighborhood rough sets and discuss a discretization scheme. In Section 3, we propose a discretization neighborhood rough set model and estimate an attribute reduction algorithm based on this model. We develop both the weighted discretization neighborhood rough set model and the fast discretization neighborhood rough set model in Section 4. We design a series of comparative experiments to evaluate comprehensively our models in Section 5 and conclude the paper in Section 6.

2. Preliminaries

Rough sets deal with data classification through equivalence relations, which are too strict to properly partition data. Consequently, similarity relations are employed as a substitute for equivalence relations. In this section, we first introduce similarity relations. Then, some commonly used discretization technologies and neighborhood rough set models are described. Finally, an analysis of the limitations of discretization technologies and neighborhood rough sets is conducted through the use of illustrative examples.

In various domains, such as data mining and knowledge discovery in Databases (KDD), information is frequently represented in the form of an information system.

Definition 2.1 *([29,34]).* An information system is a four tuple $IS = \langle U, A, V, f\rangle$, which satisfies the following assumptions:
(1) $U = \{x_i\}_{i=1}^{n}$ is a non-empty finite set, which is a family of objects. U is called the universe of IS and n is the number of objects.
(2) $A = C \bigcup D$ is the attribute set of an object, where C is the condition attribute set and D is the decision attribute.
(3) $V = \{V_a\}_{a=1}^{m}$ is a set of the value domain of attributes. If the value domain V_a is continuous, we say V_a is a numerical attribute value domain, otherwise, nominal attribute value domain. If the attribute set C contains both numerical and nominal attributes, then C is said to be a mixed attribute set.
(4) $f = \{f_a\}_{a=1}^{m}$ stands for a family of functions, where f_a is a mapping from an object x_i to an attribute value V_a.

Remark 2.1. Normalization on mixed data.

- For arbitrary numerical attribute $a \in C$ and object $x \in U$, if $V_a \subseteq [0,1]$, then $f_{a_i}(x)$ is called a fuzzy relation between U and a, and otherwise, V_a may be standardized by using max-min normalization, that is, $V_a = \frac{value - min}{max - min}$.
- For any nominal attribute $b \in C$, let V_b be converted into $|V_b|$ numbers evenly distributed within the interval [0,1], where $|V_b|$ represents the cardinality of V_b.

By Remark 2.1, the standardization of mixed data is performed. In rough set theory, an equivalence relation is equivalently expressed as an indiscernible relation.

Definition 2.2 *([33]).* Let $IS = \langle U, A, V, f\rangle$ be an information system and $B \subseteq A$. The indiscernible relation is defined as follows:

$$IND(B) = \{(x,y) \in U \times U \mid f_a(x) = f_a(y), \forall a \in B\}. \tag{1}$$

We say that x and y satisfy the indiscernible relation, if their corresponding attribute values are the same. If an attribute $a \in B$ is continuous, the cardinality of $IND(B)$ will be small, and hence the concept of similarity relations is being introduced.

Definition 2.3. A binary relation $SR \subseteq U \times U$ is called a similarity relation, if it satisfies the following:
(1) $(x,x) \in SR$ for any $x \in U$.
(2) for any $x, y \in U$, if $(x,y) \in SR$, then $(y,x) \in SR$.

There are many kinds of similarity relations based on an information system IS, such as neighborhood relations and tolerance relations. Next, let us recall neighborhood relations.

Definition 2.4 *([18]).* Given an information system IS and $B \subseteq C$, for any $x \in U$, the neighborhood of object x is defined as

$$\delta_B(x) = \{y \in U \mid d_B(x,y) \leq \delta\}, \tag{2}$$

where δ is a threshold and $d_B(x,y)$ expresses the distance between objects x and y. In the paper, we utilize Euclidean metric as a measure of distance, that is, $d_B(x,y) = \sqrt{\sum_{i=1}^{m}\left(f_{a_i}(x) - f_{a_i}(y)\right)^2}$. If $y \in \delta_B(x)$, then we say that x and y satisfy the neighborhood relation RN, namely, $RN = \{(x,y) \in U \times U \mid y \in \delta_B(x)\}$.

Proposition 2.1. *The neighborhood relation RN is a similarity relation.*

Proof. Let $RN = \{(x, y) \in U \times U \mid y \in \delta_B(x)\}$. For arbitrary $x \in U$, it follows from $d_B(x, x) = 0$ that $x \in \delta_B(x)$, and therefore, $(x, x) \in RN$. For arbitrary $x, y \in U$, if $(x, y) \in RN$, then we have $y \in \delta_B(x)$. We can get $x \in \delta_B(y)$, so $(y, x) \in RN$. □

Definition 2.5 *([18]).* Given an information system IS and $U/D = \{D_1, D_2, \ldots, D_r\}$, for $B \subseteq C$, the lower and upper approximations of D based on neighborhood relations are defined as follows:

$$\underline{NR}_B(D) = \bigcup_{i=1}^{r} \underline{NR}_B(D_i),$$

and

$$\overline{NR}_B(D) = \bigcup_{i=1}^{r} \overline{NR}_B(D_i).$$

For $i = 1, 2, \ldots, r$, the lower and upper approximations $\underline{NR}_B(D_i)$ and $\overline{NR}_B(D_i)$ are expressed as follows:

$$\underline{NR}_B(D_i) = \{x \in U \mid \delta_B(x) \subseteq D_i\},$$

$$\overline{NR}_B(D_i) = \{x \in U \mid \delta_B(x) \cap D_i \neq \emptyset\}.$$

By using $\underline{NR}_B(D_i)$ and $\overline{NR}_B(D_i)$, the universe U is divided into three parts as follows:

$$POS_B(D_i) = \underline{NR}_B(D_i);$$

$$NEG_B(D_i) = U - \overline{NR}_B(D_i);$$

$$BN_B(D_i) = \overline{NR}_B(D_i) - \underline{NR}_B(D_i).$$

The symbols $POS_B(D_i)$, $NEG_B(D_i)$, and $BN_B(D_i)$ are the positive region, negative region, and boundary region of D_i with respect to B, respectively.

Definition 2.6 *([25]).* Given an information system IS and a positive integer k, for $a \in C$, the discretization scheme $\mathcal{D}_a^k$ of a is defined as follows:

$$\mathcal{D}_a^k : \{[d_0, d_1), [d_1, d_2), \ldots, [d_{k-1}, d_k]\}, \tag{3}$$

where $d_0, d_1, \ldots, d_k$ satisfy that $0 = d_0 < d_1 < \cdots < d_{k-1} < d_k = 1$ and $|d_1 - d_0| = |d_2 - d_1| = \cdots = |d_k - d_{k-1}|$. For $x \in U$, there exists $[d_{i-1}, d_i) \in \mathcal{D}_a^k$, such that $x \in [d_{i-1}, d_i)$. The discretization class $DC_a^k(x)$ of object x is represented as follows:

$$DC_a^k(x) = \{y \in U \mid y \in [d_{i-1}, d_i)\}. \tag{4}$$

For $B \subseteq C$, the discretization class $DC_B^k(x)$ of x is defined as follows:

$$DC_B^k(x) = \bigcap_{a \in B} DC_a^k(x). \tag{5}$$

The subset $DC_B^k(x)$ is also called the information block of x. The discretization relation DR_B^k is represented as $DR_B^k = \{(x, y) \in U \times U \mid y \in DC_B^k(x)\}$. Obviously, the discretization relation DR_B^k is an equivalence relation, which gives rise to a partition of the universe U.

Definition 2.7. Given an information system IS and $U/D = \{D_1, D_2, \cdots, D_r\}$, for $B \subseteq C$, the lower and upper approximations of D based on discretization relations are defined as follows:

$$\underline{DR}_B(D) = \bigcup_{i=1}^{r} \underline{DR}_B(D_i),$$

$$\overline{DR}_B(D) = \bigcup_{i=1}^{r} \overline{DR}_B(D_i).$$

The lower and upper approximations $\underline{DR}_B(D_i)$ and $\overline{DR}_B(D_i)$ are expressed as follows:

$$\underline{DR}_B(D_i) = \{x \in U \mid DC_B^k(x) \subseteq D_i\},$$

$$\overline{DR}_B(D_i) = \{x \in U \mid DC_B^k(x) \cap D_i \neq \emptyset\}.$$

The universe U is divided into three parts by using $\underline{DR}_B(D_i)$ and $\overline{DR}_B(D_i)$. It follows that

$$POS_B^{DR}(D_i) = \underline{DR}_B(D_i);$$
$$NEG_B^{DR}(D_i) = U - \overline{DR}_B(D_i);$$
$$BN_B^{DR}(D_i) = \overline{DR}_B(D_i) - \underline{DR}_B(D_i).$$

By employing the discretization method, we can partition the universe into three parts. The symbols $POS_B^{DR}(D_i)$, $NEG_B^{DR}(D_i)$, and $BN_B^{DR}(D_i)$ are the positive region, negative region, and boundary region of D_i with respect to B, respectively.

In general, it is not feasible for rough sets to efficiently obtain equivalence classes for numerical data. Although both neighborhood rough sets and discretization methods can solve this problem, they also have drawbacks in other aspects. For neighborhood rough sets, it is necessary to compute the distance between all objects in order to establish the neighborhood relation. However, this procedure is known to be time-consuming. Therefore, it is not an efficient strategy to calculate the upper and lower approximations of a target concept. For discretization methods, the value domain of an attribute is divided in several intervals, which have the same length. However, the presence of adjacent attribute values that are divided into separate intervals leads to a loss of information. For example, let us consider a numerical attribute "age" with a value domain ranging from 0 to 100. Firstly, [0,100] is normalized to [0,1] by using max-min normalization. Then, according to Definition 2.6, let $k = 5$, and we can determine that $\mathcal{D}^k_{\text{age}} : \{[0, 0.2), [0.2, 0.4), [0.4, 0.6), [0.6, 0.8), [0.8, 1]\}$. It is apparent that "0.21" and "0.30" belong to the same interval [0.2,0.4), whereas "0.19" and "0.21" are assigned to two different intervals. However, based on previous experience, the distance between "0.21" and "0.30" is greater than the distance between "0.19" and "0.21".

3. Discretization neighborhood rough set model

In this section, we define two binary relations and two evaluation functions. Then, we propose the discretization neighborhood rough set (DNR) model. Finally, attribute reduction-based discretization neighborhood rough set is built.

3.1. Discretization neighborhood rough sets

In order to alleviate the shortcomings of neighborhood rough sets and discretization theory, we propose a discretization neighborhood rough set model, which can effectively avoid information loss and reduce time-consuming.

Definition 3.1. Suppose that $IS = \langle U, A, V, f\rangle$ is an information system and $\mathcal{D}_a^k$ is a discretization scheme, where k is a positive integer. For $a \in C$ and $x \in U$, there exists $[d_{i-1}, d_i) \in \mathcal{D}_a^k$ such that $x \in [d_{i-1}, d_i)$. The quasi-discretization class $QDC_a^k(x)$ of x is defined as follows:

$$QDC_a^k(x) = \begin{cases} \{y \in U \mid y \in [d_0, d_2)\} & i = 1, \\ \{y \in U \mid y \in [d_{i-2}, d_{i+1})\} & 1 < i < k, \\ \{y \in U \mid y \in [d_{k-2}, d_k]\} & i = k. \end{cases} \tag{6}$$

For $i = 1, 2, \ldots, k$, the ith storage $ST_a(i)$ of a is defined as follows:

$$ST_a(i) = \{QDC_a^k(x) \mid x \in [d_{i-1}, d_i)\}. \tag{7}$$

For $B \subseteq C$, the quasi-discretization class $QDC_B^k(x)$ of x is defined as follows:

$$QDC_B^k(x) = \bigcap_{a \in B} QDC_a^k(x). \tag{8}$$

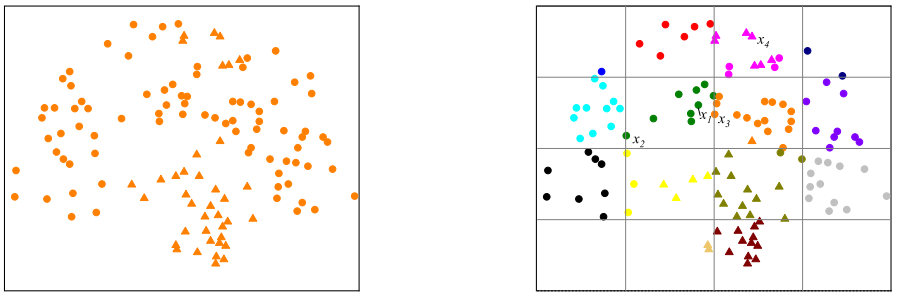
The quasi-discretization class $QDC_B^k(x)$ is also called a quasi-discretization information granule of x. The quasi-discretization relation $QDR_B^k = \{(x, y) \in U \times U \mid y \in QDR_B^k(x)\}$.

Proposition 3.1. *Let $IS = \langle U, A, V, f\rangle$ be an information system, $B, B_1, B_2 \subseteq C$, and $x \in U$. Then we have the following.*

(1) *If $B_1 \subseteq B_2$, then we have $QDC_{B_2}^k(x) \subseteq QDC_{B_1}^k(x)$ and $QDR_{B_2}^k \subseteq QDR_{B_1}^k$;*
(2) *$DC_B^k(x) \subseteq QDC_B^k(x)$;*
(3) *$DR_B^k \subseteq QDR_B^k$.*

Proof. (1) If $B_1 \subseteq B_2$, we set $B_1 = \{b_1, b_2, \ldots, b_{n_1}\}$ and $B_2 = \{b_1, b_2, \ldots, b_{n_1}, b_{n_1+1}, \ldots, b_{n_2}\}$. Consequently, we have

$$QDC_{B_2}^k(x) = \Big(\bigcap_{i=n_1+1}^{n_2} QDC_{b_i}(x)\Big) \bigcap QDC_{B_1}^k(x).$$



(a) (b)

Fig. 1. The scatter plot of the information system $\langle U, A, V, f\rangle$. The data comes from the UCI Machine Learning Repository, and some samples are deleted to ensure the visualization of the scatter plot [57]. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

By the properties of set operations, we can get that $QDC^k_{B_2}(x) \subseteq QDC^k_{B_1}(x)$. Then, we can compute that

$$\begin{aligned} QDR^k_{B_2} &= \{(x,y) \in U \times U \mid y \in QDR^k_{B_2}(x)\} \\ &\subseteq \{(x,y) \in U \times U \mid y \in QDR^k_{B_1}(x)\} \\ &= QDR^k_{B_1}. \end{aligned}$$

(2) By Definitions 3.1 and 2.6, we have

$$\begin{aligned} DC^k_B(x) &= \bigcap_{a\in B} DC^k_a(x) = \bigcap_{a\in B} \{y \in U \mid y \in [d_{i-1}, d_i)\} \\ &\subseteq \bigcap_{a\in B} \begin{cases} \{y \in U \mid y \in [d_0, d_2)\} & i = 1, \\ \{y \in U \mid y \in [d_{i-2}, d_{i+1})\} & 1 < i < k, \\ \{y \in U \mid y \in [d_{k-2}, d_k]\} & i = k, \end{cases} \\ &= \bigcap_{a\in B} QDC^k_a(x) = QDC^k_B(x). \end{aligned}$$

(3) From (2), for any $x \in U$, we see that $DC^k_B(x) \subseteq QDC^k_B(x)$. Then, we can find that

$$\begin{aligned} DR^k_B &= \{(x,y) \in U \times U \mid y \in DC^k_B(x)\} \\ &\subseteq \{(x,y) \in U \times U \mid y \in QDR^k_B(x)\} \\ &= QDR^k_B. \quad \square \end{aligned}$$

Remark 3.1. Discussion on the quasi-discretization relation.

(1) Undoubtedly, the quasi-discretization relation QDR^k_B is a similarity relation rather than an equivalence relation. Moreover, the similarity relation of any object in the universe can be described accurately.
(2) The quasi-discretization relation QDR^k_B does not exhibit monotonicity with respect to k. In other words, if $k_1 \le k_2$, in general it is not possible to obtain $QDR^{k_2}_B \subseteq QDR^{k_1}_B$.
(3) If $k = 1, 2$, for all $x \in U$, it can be observed that $QDC^k_B(x) = U$, and any two objects have a quasi-discretization relation. If $k = 3$, there exists $x \in U$ with $x \in [d_1, d_2)$. We get $QDC^k_B(x) = U$.

Example 3.1. A scatter plot representing the information system $IS = \langle U, A, V, f\rangle$ is shown in Fig. 1(a), where U is $\{x_1, x_2, \ldots, x_{120}\}$, and $C = \{c_1, c_2\}$ with the horizontal axis c_1 and the vertical axis c_2. The decision attribute $D = \{d_1, d_2\}$ has two classes, where Class d_1 is represented by the triangles, and the dots represent Class d_2. All attribute values are standardized, and let $k = 4$. Then we can get the discretization scheme $\mathcal{D}^k_{c_i}$ of c_i is $\mathcal{D}^k_{c_i} : \{[0, 0.25), [0.25, 0.5), [0.5, 0.75), [0.75, 1]\}$. The discretization relation can be shown in Fig. 1(b), where various colors are utilized to represent different discretization classes.

From Fig. 1(b), we can get $(x_1, x_2) \in DR^k_C$ but $(x_1, x_3) \notin DR^k_C$. This is not consistent with our previous cognitive experience. By the definition of quasi-discretization relation, it can be concluded that $(x_1, x_3) \in QDR^k_C$. However, for (x_2, x_4), it is visually apparent that there exists a long "distance" between them. To avoid this tissue, we put forward a new concept: the quasi discretization-based neighborhood relations.

Definition 3.2. Let IS be an information system and D_a^k be a discretization scheme, where k is a positive integer. For $a \in C$ and $x \in U$, the quasi discretization-based neighborhood $\sigma_a^k(x)$ of x is defined as follows:

$$\begin{aligned}\sigma_a^k(x) &= \{y \in QDC_a^k(x) \mid d_a(x,y) \le \sigma\} \\ &= \{y \in ST_a(i) \mid d_a(x,y) \le \sigma, x \in [d_{i-1}, d_i)\},\end{aligned}$$

where σ is the neighborhood radius with $0 \le \sigma \le \frac{1}{k}$ and $QDC_a^k(x)$ is the quasi-discretization class of x. For $B \subseteq C$, the quasi discretization-based neighborhood σ_B^k of x is defined as follows:

$$\sigma_B^k(x) = \bigcap_{a \in B} \sigma_a^k(x). \tag{9}$$

The quasi discretization-based neighborhood $\sigma_B^k(x)$ is also called a quasi-discretization-based neighborhood information granule of x. The quasi-discretization-based neighborhood relation $\sigma_B^k = \{(x,y) \in U \times U \mid y \in \sigma_B^k(x)\}$.

Remark 3.2. Discussion on the quasi-discretization-based neighborhood relation.

(1) Obviously, the quasi-discretization-based neighborhood relation σ_B^k is a similarity relation.
(2) The reason of requiring that $\sigma \le \frac{1}{k}$ is to make $\{y \in U \mid d_a(x,y) \le \sigma\} = \sigma_a^k(x) = \{y \in QDC_a^k(x) \mid d_a(x,y) \le \sigma\}$.
(3) Different from the neighborhood $\delta_a(x)$ in Definition 2.4, the quasi discretization-based neighborhood $\sigma_a^k(x)$ only considers the elements of $QDC_a^k(x)$ instead of the entire universe U.

Proposition 3.2. *Let $IS = \langle U, A, V, f \rangle$ be an information system, $B, B_1, B_2 \subseteq C$, and $x \in U$. Then we have the following.*

(1) *If $B_1 \subseteq B_2$, then we get $\sigma_{B_2}^k(x) \subseteq \sigma_{B_1}^k(x)$ and $\sigma_{B_2}^k \subseteq \sigma_{B_1}^k$;*
(2) *If $\sigma_1 \le \sigma_2$, then we obtain ${\sigma_1}_a^k(x) \subseteq {\sigma_2}_a^k(x)$ and ${\sigma_1}_a^k \subseteq {\sigma_2}_a^k$;*
(3) *$\sigma_B^k(x) \subseteq QDC_B^k(x)$;*
(4) *$\sigma_B^k \subseteq QDR_B^k$.*

Proof. (1) If $B_1 \subseteq B_2$, we set $B_1 = \{b_1, b_2, \ldots, b_{n_1}\}$ and $B_2 = \{b_1, b_2, \ldots, b_{n_1}, b_{n_1+1}, \ldots, b_{n_2}\}$. So, we have

$$\sigma_{B_2}^k(x) = \Big(\bigcap_{i=n_1+1}^{n_2} \sigma_{b_i}(x)\Big) \bigcap \sigma_{B_1}^k(x).$$

By the properties of set operations, it is easy to obtain that $\sigma_{B_2}^k(x) \subseteq \sigma_{B_1}^k(x)$. Then, we can get

$$\begin{aligned}\sigma_{B_2}^k &= \{(x,y) \in U \times U \mid y \in \sigma_{B_2}^k(x)\} \\ &\subseteq \{(x,y) \in U \times U \mid y \in \sigma_{B_1}^k(x)\} \\ &= \sigma_{B_1}^k.\end{aligned}$$

(2) If $\sigma_1 \le \sigma_2$, then

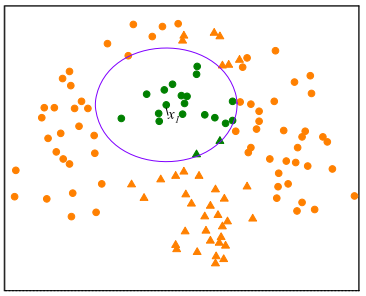
$$\begin{aligned}{\sigma_1}_a^k &= \{y \in ST_a(i) \mid d_a(x,y) \le \sigma_1, x \in [d_{i-1}, d_i)\} \\ &\subseteq \{y \in ST_a(i) \mid d_a(x,y) \le \sigma_2, x \in [d_{i-1}, d_i)\} \\ &= {\sigma_2}_a^k.\end{aligned}$$

Then, we can compute

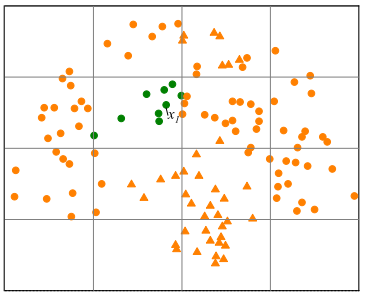
$$\begin{aligned}{\sigma_1}_a^k &= \{(x,y) \in U \times U \mid y \in {\sigma_1}_a^k(x)\} \\ &\subseteq \{(x,y) \in U \times U \mid y \in {\sigma_2}_a^k(x)\} \\ &= {\sigma_2}_a^k.\end{aligned}$$

(3) From Definitions 3.1 and 3.2,

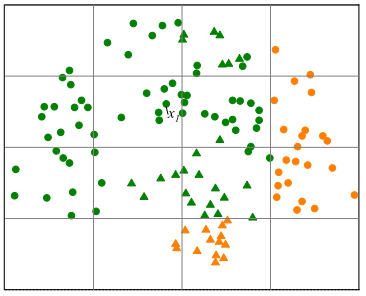
$$\begin{aligned}\sigma_B^k(x) &= \bigcap_{a \in B} \sigma_a^k(x) \\ &= \bigcap_{a \in B} \{y \in ST_a(i) \mid d_a(x,y) \le \sigma, x \in [d_{i-1}, d_i)\}\end{aligned}$$



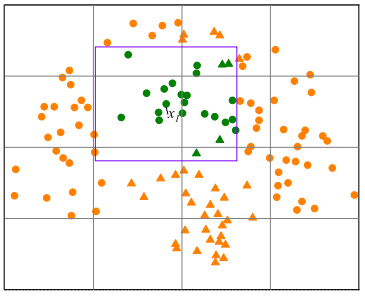
(a) The neighborhood $\delta_C(x_1)$ of x_1.



(b) The discretization class $DC_C^k(x_1)$ of x_1.



(c) The quasi-discretization class $QDC_C^k(x_1)$ of x_1.



(d) The quasi discretization-based neighborhood $\sigma_C^k(x_1)$ of x_1.

Fig. 2. The scatter plot of $\langle U, A, V, f\rangle$.

$$\subseteq \bigcap_{a\in B} \begin{cases} \{y\in U \mid y\in [d_0, d_2)\} & i=1, \\ \{y\in U \mid y\in [d_{i-2}, d_{i+1})\} & 1<i<k, \\ \{y\in U \mid y\in [d_{k-2}, d_k]\} & i=k, \end{cases}$$
$$= \bigcap_{a\in B} QDC_a^k(x) = QDR_B^k(x).$$

(4) For arbitrary $x\in U$, we have $\sigma_B^k \subseteq QDR_B^k$. Whence we have that

$$\begin{aligned} \sigma_B^k &= \{(x,y)\in U\times U \mid y\in \sigma_B^k(x)\} \\ &\subseteq \{(x,y)\in U\times U \mid y\in QDR_B^k(x) \\ &= QDR_B^k. \quad \square \end{aligned}$$

Example 3.2. Let us revisit the information system in Example 3.1. Let $k=4$ and $\sigma=0.2$. The neighborhood $\delta_C(x_1)$, the discretization class $DC_C^k(x_1)$, the quasi-discretization class $QDC_C^k(x_1)$, and the quasi discretization-based neighborhood $\sigma_C^k(x_1)$ of x_1 are demonstrated in Fig. 2(a), (b), (c), and (d), respectively. The solid lines show the ranges of $\delta_C^k(x_1)$ and $\sigma_C^k(x_1)$, and the green markers indicate the inclusions of these objects in $QDC_C^k(x_1)$ and $DC_C^k(x_1)$. From Fig. 2, we can find $|DC_C^k(x_1)|\subseteq|\delta_C^k(x_1)|\subseteq|\sigma_C^k(x_1)|\subseteq|QDC_C^k(x_1)|$, where $|\cdot|$ represents the cardinality of the set. In Fig. 2(c), it is evident that $QDC_C^k(x_1)$ contains a large number of elements so that the quasi-discretization classes lost the significance of relations, while we can only get little information in Fig. 2(b) because $|DC_C^k(x_1)|$ is too small. Consequently, it is imperative to establish a stable relation. Next, we propose two new definitions for evaluating the quality of relations.

Definition 3.3. Let $R\subseteq U\times U$ be a binary relation. Gaussian balance index (GBI) of R is defined as follows:

$$GBI_R = \exp\left(\frac{-(|R|/|U|^2-0.5)^2}{\eta}\right), \tag{10}$$

where $|\cdot|$ represents the cardinality of the set and η is a threshold.

Table 1
The decision information table $\langle U, C, D\rangle$.

	c_1	c_2	c_3	D
x_1	0.4	0.2	0.2	1
x_2	0.4	0.1	0.7	1
x_3	0.3	0.2	0.6	1
x_4	0	0.4	0.1	1
x_5	0.6	0.2	0.2	1
x_6	0.6	0.8	0.9	2
x_7	0.6	0.7	0.4	2
x_8	0.5	0.8	0.6	2
x_9	0.4	0.4	0.9	2
x_{10}	0.5	0.6	0.8	2

In the paper, we set $\eta = 0.1$. As a matter of fact, for relations $R_1 = \{(x, y) \in U \times U\}$ and $R_2 = \emptyset$, we obtain that $GBI_{R_1} = GBI_{R_2} = \min_R GBI_R = 0.08$.

Definition 3.4. Let $R \subseteq U \times U$ be a binary relation. Quality index (QI) of R is defined as follows:

$$QI_R = \frac{|GR_R|}{|R|}, \tag{11}$$

where the great relation $GR_R = \{(x, y) \in R \mid f_D(x) = f_D(y)\}$ represents the set of pairs of elements with the same decision attribute in R. Obviously, for any $(x, y) \in R$, if we can determine that x and y have the same label, then we say that R is an optimal relation in terms of quality.

Let $G = \{(x, y) \mid x, y \in U\}$ be a set of ordered binary pairs.

- If for any $(x, y) \in G$, there exists a set G' such that $(y, x) \in G'$ and $|G| = |G'|$, then we say G' is a complementary set of G.
- We call $I = \{(x, x) \mid x \in U\}$ is an identity relation set.

Example 3.3. A decision information system $IS = \langle U, A, V, f\rangle$ is shown in Table 1. The universe U is $\{x_1, x_2, \ldots, x_{10}\}$ and the condition attribute set C is $\{c_1, c_2, c_3\}$, where all the data are discretized. Let $k = 2$ and $\delta = \sigma = 0.4$.

(1) The neighborhood relation RN_C can be computed as follows:

$$RN_C = R \cup R' \cup I,$$

where

$$R = \{(x_1, x_5), (x_2, x_3), (x_2, x_9), (x_3, x_9), (x_6, x_8), (x_6, x_{10}), (x_7, x_8), (x_8, x_{10}), (x_9, x_{10})\}.$$

(2) The discretization relation DR_C^k is obtained as follows:

$$DR_C^k = S \cup S' \cup I,$$

where

$$S = \{(x_1, x_4), (x_2, x_3), (x_2, x_9), (x_3, x_9), (x_6, x_8), (x_6, x_{10}), (x_8, x_{10})\}.$$

(3) The quasi-discretization relation QDR_C^k is obtained as follows:

$$QDR_C^k = \{(x, y) \mid x, y \in U\}.$$

(4) The neighborhood-based quasi discretization relation σ_C^k is calculated as follows:

$$\sigma_C^k = N \cup N' \cup I,$$

where

$$N = \{(x_1, x_3), (x_1, x_4), (x_1, x_5), (x_2, x_3), (x_2, x_9), (x_3, x_5), (x_3, x_9), (x_3, x_{10}),$$
$$(x_6, x_8), (x_6, x_9), (x_6, x_{10}), (x_7, x_8), (x_7, x_{10}), (x_8, x_9), (x_8, x_{10}), (x_9, x_{10})\}.$$

Through the above analysis, we can get detailed comparison results shown in Table 2. Obviously, we find that the neighborhood-based quasi discretization relation σ_C^k has the best results in terms of the Gaussian balance index and the quality index. This means that σ_C^k is a stable relation. Compared to other relations, the Gaussian balance index of quasi-discretization relation QDR_C^k is the lowest. This is due to $QDR_C^k = U \times U$.

Table 2
The Gaussian balance indices and quality indices of relations.

Index	RN_C	DR_C^k	QDR_C^k	σ_C^k
The number of relation |R|	28	24	100	42
The number of great relation $\|GR_R\|$	24	20	50	36
The Gaussian balance index GBI_R	0.62	0.51	0.08	0.94
The quality index QI_R	0.86	0.83	0.50	0.86

Definition 3.5. Given a decision information system $IS = \langle U, A, V, f\rangle$ and any $X \subseteq U$, the lower and upper approximations of X with respect to $B \subseteq A$ are defined as follows:

$$\underline{app}_B(X) = \{x \in U \mid \sigma_B^k(x) \subseteq X\},$$
$$\overline{app}_B(X) = \{x \in U \mid \sigma_B^k(x) \cap X \neq \emptyset\}.$$

If $\underline{app}_B(X) = \overline{app}_B(X)$, then we say X is a crisp set, otherwise, a rough set. The universe U is divided into three parts by using $\underline{app}_B(X)$ and $\overline{app}_B(X)$. It follows that

$$POS_B(X) = \underline{app}_B(X);$$
$$NEG_B(X) = U - \overline{app}_B(X);$$
$$BN_B(X) = \overline{app}_B(X) - \underline{app}_B(X).$$

The positive region $POS_B(X)$ indicates the set of elements that are identified certainly as belonging to X. The negative region $NEG_B(X)$ is the set of elements that are absolutely not in X. The boundary region $BN_B(X)$ is a collection of uncertain elements.

Definition 3.6. Let $IS = \langle U, A, V, f\rangle$ be a decision information system and $U/D = \{D_1, D_2, \ldots, D_r\}$. For $B \subseteq C$, the lower and upper approximations of D with respect to B are defined as follows:

$$\underline{app}_B(D) = \bigcup_{i=1}^{r} \underline{app}_B(D_i),$$

and

$$\overline{app}_B(D) = \bigcup_{i=1}^{r} \overline{app}_B(D_i).$$

Proposition 3.3. *Let $IS = \langle U, A, V, f\rangle$ be an information system, $B, B_1, B_2 \subseteq C$, and $X \subseteq U$. Then we have the following.*

(1) *If $B_1 \subseteq B_2$, then we can get $\underline{app}_{B_1}(X) \subseteq \underline{app}_{B_2}(X)$ and $\overline{app}_{B_2}(X) \subseteq \overline{app}_{B_1}(X)$.*
(2) *If $B_1 \subseteq B_2$, then we can find $\underline{app}_{B_1}(D) \subseteq \underline{app}_{B_2}(D)$ and $\overline{app}_{B_2}(D) \subseteq \overline{app}_{B_1}(D)$.*
(3) *$\underline{app}_B(X) = \sim \overline{app}_B(\sim X)$, $\overline{app}_B(X) = \sim \underline{app}_B(\sim X)$.*
(4) *If $\sigma_1 \preceq \sigma_2$, then we can obtain $\underline{app}_B^{\sigma_2}(X) \subseteq \underline{app}_B^{\sigma_1}(X)$ and $\overline{app}_B^{\sigma_1}(X) \subseteq \overline{app}_B^{\sigma_2}(X)$.*

Proof. (1) If $B_1 \subseteq B_2$, then we set $B_1 = \{b_1, b_2, \ldots, b_{n_1}\}$ and $B_2 = \{b_1, b_2, \ldots, b_{n_1}, b_{n_1+1}, \ldots, b_{n_2}\}$. By Proposition 3.2(1), we have $\sigma_{B_2} \subseteq \sigma_{B_1}$. So, we can compute that

$$\begin{aligned}\underline{app}_{B_1}(X) &= \{x \in U \mid \sigma_{B_1}^k(x) \subseteq X\}\\ &\subseteq \{x \in U \mid \sigma_{B_2}^k(x) \subseteq X\}\\ &= \underline{app}_{B_2}(X)\end{aligned}$$

and

$$\begin{aligned}\overline{app}_{B_2}(X) &= \{x \in U \mid \sigma_{B_2}^k(x) \cap X \neq \emptyset\}\\ &\subseteq \{x \in U \mid \sigma_{B_1}^k(x) \cap X \neq \emptyset\}\\ &= \overline{app}_{B_1}(X).\end{aligned}$$

(2) It follows from (1) that

$$\underline{app}_{B_1}(D) = \bigcup_{i=1}^{r} \underline{app}_{B_1}(D_i) \subseteq \bigcup_{i=1}^{r} \underline{app}_{B_2}(D_i) = \underline{app}_{B_2}(X),$$

$$\overline{app}_{B_2}(D) = \bigcup_{i=1}^{r} \overline{app}_{B_2}(D_i) \subseteq \bigcup_{i=1}^{r} \overline{app}_{B_1}(D_i) = \overline{app}_{B_1}(X).$$

(3) From the complementarity of sets,

$$\begin{aligned}\overline{app}_B(\sim X) &= \{x \in U \mid \sigma_B^k(x) \cap \sim X \neq \varnothing\} = \sim \{x \in U \mid \sigma_B^k(x) \cap \sim X = \varnothing\} \\ &= \sim \{x \in U \mid \sigma_B^k(x) \subseteq X\} = \sim \underline{app}_B(X).\end{aligned}$$

Similarly,

$$\begin{aligned}\underline{app}_B(\sim X) &= \{x \in U \mid \sigma_B^k(x) \subseteq \sim X\} = \sim \{x \in U \mid \sigma_B^k(x) \subseteq X\} \\ &= \sim \{x \in U \mid \sigma_B^k(x) \cap X = \varnothing\} = \sim \overline{app}_B(X).\end{aligned}$$

(4) If $\sigma_1 \leq \sigma_2$, then we know

$$\underline{app}_B^{\sigma_2}(X) = \{x \in U \mid {\sigma_2}_B^k(x) \subseteq X\} \subseteq \{x \in U \mid {\sigma_1}_B^k(x) \subseteq X\} = \underline{app}_B^{\sigma_1}(X),$$

and

$$\overline{app}_B^{\sigma_1}(X) = \{x \in U \mid {\sigma_1}_B^k(x) \cap X \neq \varnothing\} \subseteq \{x \in U \mid {\sigma_2}_B^k(x) \cap X \neq \varnothing\} = \overline{app}_B^{\sigma_2}(X). \quad \square$$

Definition 3.7. Let $IS = \langle U, A, V, f \rangle$ be a decision information system and $U/D = \{D_1, D_2, \ldots, D_r\}$. For $B \subseteq C$, the positive region and boundary region are respectively represented as follows:

$$POS_B(D) = \underline{app}_B(D);$$
$$BN_B(D) = \overline{app}_B(D) - \underline{app}_B(D).$$

The positive region $POS_B(D)$ of D is the union of the lower approximations of all decision classes. The boundary region $BN_B(D)$ is the upper approximation of D minus the lower approximation of D.

Proposition 3.4. *Let $IS = \langle U, A, V, f \rangle$ be an information system, and $B_1, B_2 \subseteq C$. If $B_1 \subseteq B_2$, then*

$$POS_{B_1}(D) \subseteq POS_{B_2}(D)$$
$$BN_{B_2}(D) \subseteq BN_{B_1}(D).$$

Proof. By Definition 3.7 and Proposition 3.3, if $B_1 \subseteq B_2$, we know

$$POS_{B_1}(D) = \underline{app}_{B_1}(D) \subseteq \underline{app}_{B_2}(D) = POS_{B_2}(D)$$

and

$$\begin{aligned}BN_{B_2}(D) &= \overline{app}_{B_2}(D) - \underline{app}_{B_2}(D) \\ &\subseteq \overline{app}_{B_1}(D) - \underline{app}_{B_1}(D) \\ &= BN_{B_1}(D). \quad \square\end{aligned}$$

3.2. Attribute reduction based on discretization neighborhood rough sets

We propose a novel algorithm for attribute reduction through neighborhood based-quasi discretization relations.

Definition 3.8. Let $IS = \langle U, A, V, f \rangle$ be a decision information system and $U/D = \{D_1, D_2, \ldots, D_r\}$. For $B \subseteq C$, the dependency degree is defined as follows:

$$\gamma_B(D) = \frac{|POS_B(D)|}{|U|}.$$

The dependency degree $\gamma_B(D)$, which satisfies $0 \leq \gamma_B(D) \leq 1$, reflects the degree of uncertainty in the data. The higher value of $\gamma_B(D)$, the lower level of uncertainty. If $POS_B(D) = U$, we obtain $\gamma_B(D) = 1$. If $POS_B(D) = \emptyset$, we get $\gamma_B(D) = 0$ and it means that the data is in a state of complete uncertainty.

Proposition 3.5. *Let $IS = \langle U, A, V, f \rangle$ be an information system, and $B_1, B_2 \subseteq C$. If $B_1 \subseteq B_2$, then*

$$\gamma_{B_1}(D) \leq \gamma_{B_2}(D).$$

As the number of attributes increases, the dependency degree $\gamma_B(D)$ gradually improves.

Definition 3.9. Let $IS = \langle U, A, V, f \rangle$ be a decision information system and $U/D = \{D_1, D_2, \ldots, D_r\}$. For $B \subseteq C$ and $a \in C - B$, the significance of a is defined as follows:

$$S(a, B, D) = \gamma_{B \cup a}(D) - \gamma_B(D).$$

If $B = \emptyset$, we set $S(a, B, D) = \gamma_a(D)$. If $S(a, B, D) = 0$, attribute a is insignificant for attribute subset B. If $S(a, B, D) < S(b, B, D)$, we consider b to be more crucial than a. Therefore, an attribute can be selected by comparing its significance.

Definition 3.10. Let $IS = \langle U, A, V, f \rangle$ be a decision information system and $B \subseteq C$. Attribute subset B is a reduct of C, if B satisfies:
(1) $\gamma_B(D) = \gamma_C(D)$;
(2) $\gamma_{B-\{b\}}(D) \neq \gamma_B(D)$, for any $b \in B$.

Next, we present an attribute reduction algorithm for discretization neighborhood rough sets, which is shown in Algorithm 1.

Algorithm 1 The greedy algorithm based on DNR model (GADNR).

Input: A decision information system $IS = \langle U, A, V, f \rangle$, and thresholds k, σ, and θ.
Output: A reduct B of A.

```
1: B ← ∅, V ← U, A ← C, and flag = 1.
2: for any a ∈ A do
3:     Calculate the discretization scheme D_a^k of a.
4:     Compute the ith storage ST_a(i) of a, where i = 1, 2, ..., k.
5:     for i = 1 : |U| do
6:         Get the distance of objects x_i and x_j, where x_i and x_j belong to the same storage.
7:         Calculate the neighborhood based- quasi discretization σ_a^k of x.
8:     end for
9: end for
10: while flag do
11:     for any a ∈ A − B do
12:         Compute the significance S(a, B, D) of a.
13:     end for
14:     Choose the attribute a' to maximize SIG(a, B, D).
15:     if SIG(a', B, D) ≤ θ then
16:         flag = 0.
17:     else
18:         B ← B ∪ {a'}
19:     end if
20: end while
21: return B.
```

In Algorithm 1, in Steps 2-4, in order to compute the discretization scheme D_a^k of a and the storage $ST_a(i)$, the time complexity of this algorithm is $O(mn)$, where m is the cardinality of the universe U, and n is the cardinality of the condition attribute set C. In Steps 5-8, we need to compute the distance between x_i and x_j in the same storage $ST_a(i)$ and the time complexity is $O(mnp)$, where $p = \max_{a \in A, i=1,\ldots,k} |ST_a(i)|$. We add at most n elements to B, so Steps 10-20 are performed at most n times. The time complexity of Steps 10-20 can be expressed as $O(mn^2)$. Consequently, the computational complexity of Algorithm 1 is $O(mnp + mn^2)$. Because $p \ll m$, we can improve the operation speed through Algorithm 1.

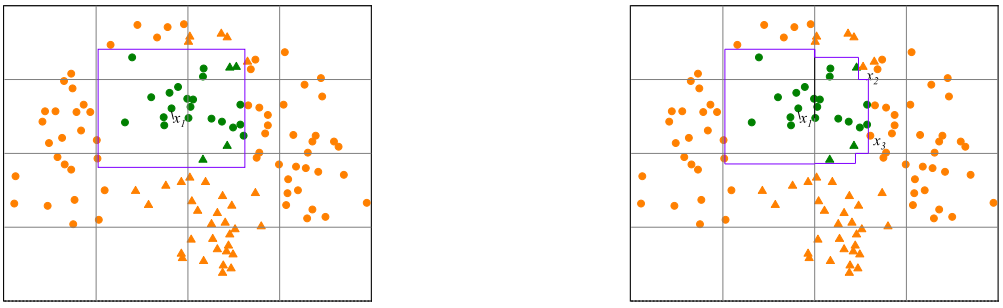
4. Extensions of the discretization neighborhood rough set model

Compared to the traditional neighborhood rough set model, our model has a distinctive bilayer structure. Firstly, the universe U is roughly divided by the theory of discretization. Then, based on the storage $ST_a(i)$, the local area is calculated accurately using neighborhood relation. This particular structure not only exhibits a reduced computational cost but also inspires us to extend the model to solve some other problems.

4.1. Weighted quasi discretization neighborhood rough sets

Various data sets have different distributions. However, most of existing rough set models fail to consider the distribution of data differentiation. Next, we propose a weighted quasi discretization neighborhood rough set model.

Definition 4.1. Let IS be an information system and D_a^k be a discretization scheme, where k is a positive integer. For $a \in C$ and $x \in U$, the weighted quasi discretization-based neighborhood $\psi_a^k(x)$ of x is defined as follows:



(a) The quasi discretization-based neighborhood $\sigma_C^k(x_1)$ of x_1.

(b) The weighted quasi discretization-based neighborhood $\psi_C^k(x_1)$ of x_1.

Fig. 3. The scatter plot of $\langle U, A, V, f\rangle$.

$$\psi_a^k(x) = \{y \in DC_a^k(x) \mid \ \lambda_1 d_a(x,y) \le \psi\} \bigcup \{y \in QDC_a^k(x)/DC_a^k(x) \mid \ \lambda_1 d_a(x, b_{(x,y)}) + \lambda_2 d_a(b_{(x,y)}, y) \le \psi\},$$

where λ_1 and λ_2 are the weights and ψ is the neighborhood radius with $0 \le \psi \le \frac{1}{k}$. For $(x,y) \in QDR_B^k$ and $(x,y) \notin DR_B^k$, we assume that the boundary $b_{(x,y)}$ created by the discretization scheme $\mathcal{D}_a^k$ between x and y is d_i, hence we can get $d_a(b_{(x,y)}, x) = d_a(x, d_i)$. For $B \subseteq C$, the weighted quasi discretization-based neighborhood ψ_B^k of x is defined as follows:

$$\psi_B^k(x) = \bigcap_{a \in B} \psi_a^k(x). \tag{12}$$

The weighed discretization neighborhood relation $\psi_B^k = \{(x,y) \in U \times U \mid y \in \psi_B^k(x)\}$. We can provide a new technique to describe the weights λ_1 and λ_2.

Definition 4.2. Let IS be an information system and $\mathcal{D}_a^k$ be a discretization scheme, where k is a positive integer. For $x \in U$, the information block $DC_B^k(x)$ is a subset of U. The distance weight λ in $DC_B^k(x)$ is defined as follows:

$$\lambda = \frac{2}{1 + QI_R}, \tag{13}$$

where QI_R is the quality index of R in Definition 3.4 with $R = DC_B^k(x)$. The distance weight satisfies $\lambda \in [1,2]$. In brief, the higher QI_R, the smaller the distance weight, and accordingly ψ_B^k contains more elements.

Example 4.1. Let us continue to analyze the information system in Example 3.2. Let $k = 4$ and $\sigma = \psi = 0.2$. Next, we only calculate the weighted quasi discretization-based neighborhood $\psi_C^k(x_1)$ of x_1 in Fig. 3(b). Suppose that λ_{ij} represents the weight of the block in row i and column j. Then, we can obtain:

$$\lambda_{11} = \frac{2}{1+\frac{1}{1}} = 1.00, \quad \lambda_{12} = \frac{2}{1+\frac{36}{36}} = 1.00, \quad \lambda_{13} = \frac{2}{1+\frac{65}{121}} = 1.30,$$
$$\lambda_{21} = \frac{2}{1+\frac{121}{121}} = 1.00, \quad \lambda_{22} = \frac{2}{1+\frac{81}{81}} = 1.00, \quad \lambda_{23} = \frac{2}{1+\frac{226}{256}} = 1.06,$$
$$\lambda_{31} = \frac{2}{1+\frac{64}{64}} = 1.00, \quad \lambda_{32} = \frac{2}{1+\frac{25}{49}} = 1.32, \quad \lambda_{33} = \frac{2}{1+\frac{148}{196}} = 1.14.$$

Through Definition 4.1, we can get the weighted quasi discretization-based neighborhood $\psi_C^k(x_1)$ of x_1, which is shown in Fig. 3(b). The solid lines show the ranges of $\sigma_C^k(x_1)$ and $\psi_C^k(x_1)$. We can find that $x_2, x_3 \in \sigma_C^k(x_1)$, but $x_2, x_3 \notin \psi_C^k(x_1)$ in Fig. 3.

In addition, the lower and upper approximates of the quasi discretization-based neighborhood can be defined using the same method.

Definition 4.3. Given a decision information system $IS = \langle U, A, V, f\rangle$ and any $X \subseteq U$, the lower and upper approximations of X with respect to $B \subseteq A$ are defined as follows:

$$\underline{app}_B(X) = \{x \in U \mid \psi_B^k(x) \subseteq X\}, \quad \overline{app}_B(X) = \{x \in U \mid \psi_B^k(x) \cap X \neq \emptyset\}.$$

If $\underline{app}_B(X) = \overline{app}_B(X)$, then we say X is a crisp set, otherwise, a rough set. The universe U is divided into three parts by using $\underline{app}_B(X)$ and $\overline{app}_B(X)$. The division of decision attribute D into the universe is denoted as $U/D = \{D_1, D_2, \ldots, D_r\}$. The lower and upper approximations of D with respect to B are defined as follows:

$$\underline{app}_B(D) = \bigcup_{i=1}^{r} \underline{app}_B(D_i), \quad \overline{app}_B(D) = \bigcup_{i=1}^{r} \overline{app}_B(D_i).$$

Algorithm 2 The greedy algorithm based on WDN model (GAWDN).

Input: A decision information system $IS = \langle U, A, V, f \rangle$, and thresholds k, σ, and θ.
Output: A reduct B of A.
1: $B \leftarrow \emptyset$, $V \leftarrow U$, $A \leftarrow C$, and $flag = 1$.
2: **for** any $a \in A$ **do**
3: Calculate the discretization scheme $\mathcal{D}_a^k$ of a.
4: Compute the ith storage $ST_a(i)$ of a, where $i = 1, 2, \ldots, k$.
5: Calculate the distance weight λ of ith storage $ST_a(i)$, where $i = 1, 2, \ldots, k$.
6: **for** $i = 1 : |U|$ **do**
7: Get the weight distance of objects x_i and x_j, where x_i and x_j belong to the same or adjacent storage.
8: Calculate the weighted quasi discretization-based neighborhood ψ_a^k of x.
9: **end for**
10: **end for**
11: **Steps 10-20** in Algorithm 1.
12: **return** B.

Since WDN has the same meaning and definition about the positive region and dependency degree as the discretization neighborhood rough sets, we use Definitions 3.7 and 3.8 to define the positive region and dependency degree of WDN, respectively. As a matter of fact, we only need to modify Step 7 in Algorithm 1.

4.2. Fast discretization neighborhood rough sets

One of the reasons for proposing the discretization neighborhood rough set model is that objects on the boundary of a block cannot find close samples easily. However, for objects near the center of a block, their similar samples can be perfectly represented by the discretization method. Obviously, we do not need to meaninglessly calculate the neighborhood of these objects.

Definition 4.4. Let IS be an information system. For $a \in C$, $\mathcal{D}_a^k : \{[d_0, d_1), [d_1, d_2), \ldots, [d_{k-1}, d_k]\}$ is a discretization scheme of IS. For $[d_{i-1}, d_i)$, where $i = 1, 2, \ldots, k$, we call $c_i = \dfrac{d_i - d_{i-1}}{2}$ is the center of $[d_{i-1}, d_i)$. The interior point set of $[d_{i-1}, d_i)$ is defined as follows:

$$I_a(i) = \{y \in [d_{i-1}, d_i) \mid d_a(c_i, y) \leq \alpha\}, \tag{14}$$

where $\alpha \in (0, \frac{1}{k})$ is a parameter. The boundary point set of $[d_{i-1}, d_i)$ is defined as follows:

$$B_a(i) = \{y \in [d_{i-1}, d_i)\} - I_a(i). \tag{15}$$

We divide the elements of $[d_{i-1}, d_i)$ into the interior point set and boundary point set by parameter α.

Definition 4.5. Given an information system IS and a discretization scheme $\mathcal{D}_a^k$, for $x \in [d_{i-1}, d_i)$, the fast discretization neighborhood class $FN_a^k(x)$ of x is defined as follows:

$$FN_a^k(x) = \begin{cases} \{y \in [d_{i-1}, d_i)\} & x \in I_a(i), \\ \{y \in QDC_a^k(x) \mid \ d_a(x, y) \leq \sigma\} & x \in B_a(i). \end{cases} \tag{16}$$

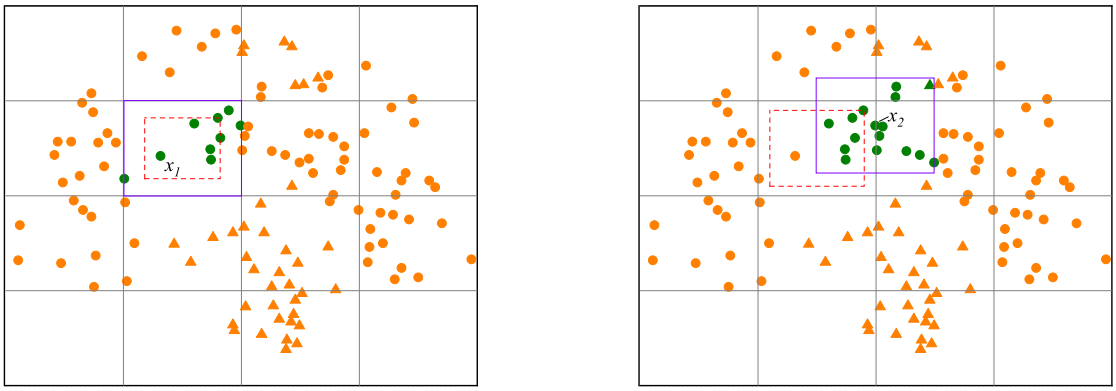
For $B \subseteq C$, the fast discretization neighborhood class $FN_a^k(x)$ of x is defined as follows:

$$FN_B^k(x) = \bigcap_{a \in B} FN_a^k(x). \tag{17}$$

For the boundary point $x \in B_a(i)$, the fast discretization neighborhood class $FN_a^k(x) = \sigma_a^k(x)$. If x is an interior point of $[d_{i-1}, d_i)$, then $FN_a^k(x)$ is the interval $[d_{i-1}, d_i)$.

Example 4.2. The information system in Example 3.2 is re-analyzed. Let $k = 4$ and $\sigma = \psi = 0.2$. The fast discretization neighborhood class $FN_C^k(x)$ is shown in Fig. 4. The ranges of $FN_a^k(x_1)$ and $FN_a^k(x_2)$ are described by solid lines.

In addition, the fast discretization neighborhood rough sets can be defined using the same method.



(a) The fast discretization neighborhood class $FN_B^k(x_1)$ of x_1. (b) The fast discretization neighborhood class $FN_B^k(x_2)$ of x_2.

Fig. 4. The scatter plot of $\langle U, A, V, f \rangle$.

Definition 4.6. Given a decision information system $IS = \langle U, A, V, f \rangle$ and any $X \subseteq U$, the lower and upper approximations of X with respect to $B \subseteq A$ are defined as follows:

$$\underline{app}_B(X) = \{x \in U \mid FN_B^k(x) \subseteq X\},$$
$$\overline{app}_B(X) = \{x \in U \mid FN_B^k(x) \cap X \neq \emptyset\}.$$

The division of decision attribute D into the universe is denoted as $U/D = \{D_1, D_2, \ldots, D_r\}$. The lower and upper approximations of D with respect to B are defined as follows:

$$\underline{app}_B(D) = \bigcup_{i=1}^{r} \underline{app}_B(D_i),$$

and

$$\overline{app}_B(D) = \bigcup_{i=1}^{r} \overline{app}_B(D_i).$$

The greedy algorithm based on the FDN model, shown in Algorithm 3, is proposed for attribute reduction.

Algorithm 3 The greedy algorithm based on FDN model (GAFDN).

Input: A decision information system $IS = \langle U, A, V, f \rangle$, and thresholds k, σ, and θ.
Output: A reduct B of A.
1: $B \leftarrow \emptyset$, $V \leftarrow U$, $A \leftarrow C$, and $flag = 1$.
2: **for** any $a \in A$ **do**
3: Calculate the discretization scheme D_a^k of a.
4: Compute the ith storage $ST_a(i)$ of a, where $i = 1, 2, \ldots, k$.
5: Calculate the distance weight λ of ith storage $ST_a(i)$, where $i = 1, 2, \ldots, k$.
6: **for** $i = 1 : |U|$ **do**
7: Get the weight distance of objects x_i and x_j, where x_i and x_j belong to the same or adjacent storage.
8: Calculate the neighborhood based- quasi discretization ψ_a^k of x.
9: **end for**
10: **end for**
11: **Steps 10-20** in Algorithm 1.
12: **return** B.

5. Experiments

In this section, three comparative experiments are designed to analyze the advantages of the three models using nine data sets, and the results are shown in Table 3. These data sets come from the UC Irvine Machine Learning Repository (http://archive.ics.uci.edu/ml/index.php). The experimental environment is Intel(R) Core i7-12650H CPU and 16 GB of memory. The software is MATLAB 2016a, and the MATLAB parallel calculation method is used in these experiments.

In the second and third experiments, the three proposed algorithms GADNR (Algorithm 1), GAWDN (Algorithm 2), and GAFDN (Algorithm 3) are compared with the following baselines on rough set models:

(1) Raw classifier (Raw): The data is classified without using any rough set model and without attribute reduction.

Table 3
The information of data sets.

No.	Name	Object	Attribute	Class	Type
1	Air	359	64	3	Numerical
2	Australian	690	14	2	Heterogeneous
3	Breast	277	9	2	Nominal
4	Bupa	345	5	2	Numerical
5	Glass	214	9	6	Numerical
6	Heart	270	13	2	Heterogeneous
7	Ionosphere	351	34	2	Numerical
8	Phoneme	5404	5	2	Numerical
9	Sonar	208	60	2	Numerical

(2) GANRS: Neighborhood rough set attribute reduction algorithm [18].
(3) NNRS: The k-nearest neighborhood rough set attribute reduction algorithm [45].
(4) FarKNNB: The k-nearest neighbor rough set attribute reduction algorithm [17].
(5) GAVRNR: Variable radius neighborhood rough set attribute reduction algorithm [57].
(6) DS: The rough set attribute reduction algorithm based on a discretization scheme [49].
(7) QDR-FarKNNB: The attribute reduction algorithm based on the quasi-discretization relation in Definition 3.1 associated with the k-nearest neighbor rough set [17].
(8) QDR-NNRS: The attribute reduction algorithm based on the quasi-discretization relation in Definition 3.1 associated with the k-nearest neighbor rough set [45].

We conduct an ablation study from model (6) to model (8) to analyze the impact of discretization on the rough set model.

5.1. Optimal parameter selection

In this subsection, parameter change analysis and optimal parameter search are carried out for the proposed models DNR, WDN, and FDN. The parameter k represents the number of elements in the discretization scheme. One of the current tasks is to find the optimal k so that the models can achieve the highest classification accuracy. We first set the initial k to be $0.015 * m$, where m is the number of objects in the data set. If $k < 1000$, then we stipulate that the step of k is 2, otherwise it is 4. The classifier used in the experiment is kNN with $k = 3$. We set $\sigma = \psi = \alpha = \frac{1}{k}$, where σ, ψ, and α are the neighborhood radii of the models DNR, WDN, and FDN, respectively.

The experiment results are shown in Fig. 5. We can draw some conclusions:

- The classification accuracy of the three models indicates a trend of initially increasing and then decreasing.
- For most data sets, the classification accuracy of the DNR model is the highest.
- In the data sets Breast and Bupa, the classification accuracy is almost unchanged with the change of parameter k, because the sensitivity of the model to the parameter is low.
- In the data sets Australian, Glass, and Heart, the classification accuracy corresponding to the initial parameters of some algorithms is usually low. The rough set model screens out a large number of uncertain samples and selects only a few certain samples, which ultimately hinders effective classification.

5.2. Assessment on running time

In the second experiment, we compare the running time of the algorithms. The greatest advantage of the discretization neighborhood rough set model is that it can efficiently reduce the computing time of the distances between objects. The running time of different algorithms is shown in Table 4, and the rankings for running time are exhibited in Table 5.

From Tables 4 and 5, we can find that:

- In each data set, the original classifier, which does not use any rough set model, takes the shortest time. Because other models carry out attribute reduction under the framework of the original classifier, the process of mining uncertain data is increased, resulting in the rough set models taking more time.
- The running time of algorithms 6 to 11 is much smaller than that of the other four rough set attribute reduction algorithms. Through ablation study, we can find that discretization technology can greatly reduce the running time in rough set models.
- The algorithm DS has the shortest running time among the rough set algorithms, because it does not compute the distances between objects. Moreover, algorithms 7 to 11 only incur a slight increase in running time compared to Algorithms 2 to 4. This result indicates that algorithms 7 to 10 only calculate the distances between a limited number of objects.
- From algorithm 7 to algorithm 11, algorithms 9, 10, and 11 have the smallest running time among these algorithms. In addition, algorithms 7 and 8 have more parameters and require more training time to find optimal values for these parameters. Therefore, we achieve the purpose of discretizing information systems associated with neighborhood rough sets.

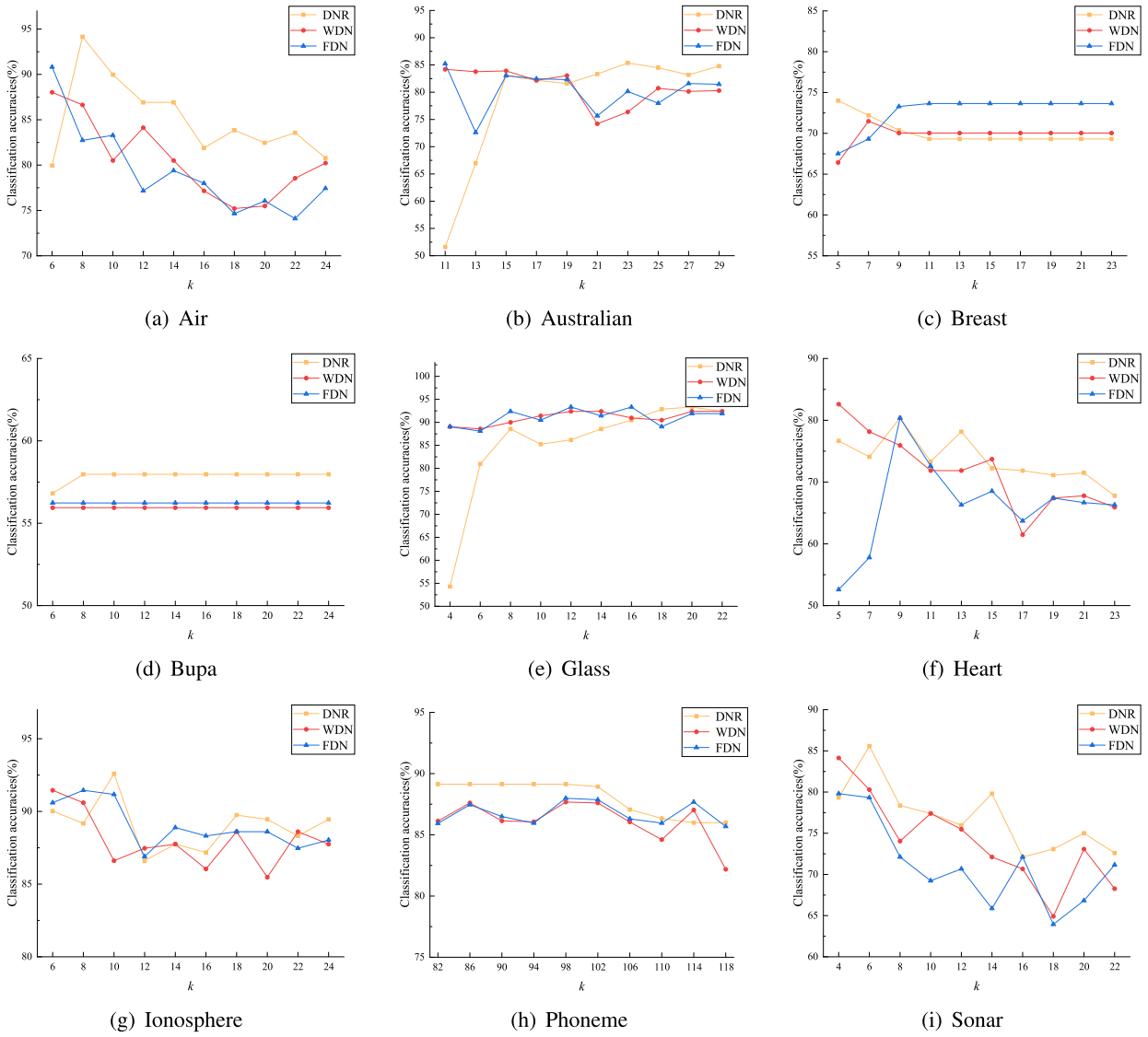


Fig. 5. The change of classification accuracy about parameter k.

Table 4
Running time of different algorithms using kNN.

Algorithms	Air	Australian	Breast	Bupa	Glass	Heart	Ionosphere	Phoneme	Sonar	Average
1.Raw	0.07	0.07	0.04	0.04	0.04	0.04	0.06	0.97	0.07	0.16
2.GANRS	25.62	14.29	0.78	0.39	0.46	1.11	7.80	103.08	3.75	17.48
3.NNRS	5.63	7.02	0.70	0.32	0.12	0.60	1.77	128.29	0.91	16.15
4.FarKNNB	33.78	10.80	0.90	0.39	0.33	1.98	7.63	151.29	10.59	24.19
5.GAVRNR	9.18	5.57	0.52	0.43	0.12	0.72	1.48	85.10	1.11	11.58
6.DS	1.04	0.61	0.07	0.06	0.06	0.09	0.36	8.3	0.31	1.21
7.QDR-FarKNNB	1.81	0.63	0.11	0.10	0.09	0.13	0.39	8.53	0.44	1.36
8.QDR-NNRS	2.06	0.70	0.12	0.12	0.11	0.16	0.52	9.13	0.66	1.51
9.GADNR	1.51	0.63	0.08	0.06	0.06	0.10	0.49	8.00	0.39	1.26
10.GAWQD	1.57	0.71	0.09	0.07	0.07	0.11	0.30	8.81	0.41	1.35
11.GAFDQN	1.31	0.67	0.08	0.06	0.06	0.10	0.35	8.42	0.28	1.26

- In most data sets, the algorithm GAWDN consumes more time than GADNR. Compared with algorithm GADNR, algorithm GAWDN adds a weight-solving process.

Table 5
Rankings for running time of different algorithms using kNN.

Algorithms	Air	Australian	Breast	Bupa	Glass	Heart	Ionosphere	Phoneme	Sonar	Average
1.Raw	1	1	1	1	1	1	1	1	1	1.00
2.GANRS	10	11	10	9	11	10	11	9	10	10.11
3.NNRS	8	9	9	8	8	8	9	10	8	8.56
4.FarKNNB	11	10	11	9	10	11	10	11	11	10.44
5.GAVRNR	9	8	8	11	8	9	8	8	9	8.67
6.DS	2	2	2	2	2	2	4	3	3	2.44
7.QDR-FarKNNB	6	3	6	6	6	6	5	5	6	5.44
8.QDR-NNRS	7	6	7	7	7	7	7	7	7	6.89
9.GADNR	4	3	3	2	2	3	6	2	4	3.22
10.GAWQD	5	7	5	5	5	5	2	6	5	5.00
11.GAFDQN	3	5	3	2	2	3	3	4	2	3.00

Table 6
The numbers of reduced attributes.

Algorithms	Air	Australian	Breast	Bupa	Glass	Heart	Ionosphere	Phoneme	Sonar	Average
1.Raw	64.0	14.0	9.0	5.0	10.0	13.0	34.0	5.0	60.0	23.78
2.GANRS	9.3	8.2	7.0	5.0	4.5	4.8	4.0	5.0	4.9	5.86
3.NNRS	11.2	11.8	9.0	5.0	2.0	8.9	6.8	5.0	7.8	7.50
4.FarKNNB	6.5	5.1	5.9	1.9	1.3	7.1	5.4	4.1	6.9	4.91
5.GAVRNR	7.4	6.2	6.9	5.0	2.0	5.3	3.9	5.0	3.0	4.97
6.DS	13.3	9.5	8.7	5.0	3.8	5.9	7.1	4.0	7.6	7.21
7.QDR-FarKNNB	14.1	2.0	5.0	2.0	2.0	4.0	3.0	3.0	7.0	4.68
8.QDR-NNRS	14.4	2.0	2.0	2.0	2.0	5.1	2.9	3.0	9.6	4.78
9.GADNR	21.8	8.9	3.9	5.0	1.0	8.9	7.6	5.0	10.7	8.09
10.GAWQD	17.5	9.0	7.6	5.0	2.8	9.2	3.6	4.0	9.3	7.56
11.GAFDQN	19.4	9.5	7.0	5.0	4.2	6.0	7.0	4.0	7.5	7.73

Table 7
The classification performance based on kNN with $k = 3$.

Algorithms	Air	Australian	Breast	Bupa	Glass	Heart	Ionosphere	Phoneme	Sonar	Average
1.Raw	**95.56(3.26)**	81.45(2.54)	70.00(7.18)	50.00(9.06)	85.24(4.74)	78.52(4.20)	85.43(5.46)	88.67(0.93)	83.81(8.46)	79.85(5.09)
2.GANRS	86.35(6.96)	83.77(3.47)	69.68(8.18)	58.26(6.19)	92.38(7.17)	69.26(8.74)	89.74(6.43)	88.6(1.43)	75.48(10.85)	79.28(6.60)
3.NNRS	88.86(3.71)	82.17(4.53)	**75.45(8.19)**	59.71(6.84)	90.48(7.10)	81.11(3.68)	92.31(5.22)	75.63(1.06)	80.77(8.23)	80.72(5.40)
4.FarKNNB	85.79(6.86)	**85.65(5.09)**	71.12(7.83)	58.84(6.21)	93.81(6.37)	**82.59(6.54)**	88.32(4.45)	74.74(2.54)	77.88(8.53)	79.86(6.05)
5.GAVRNR	84.12(7.87)	82.03(2.83)	73.29(6.36)	58.55(9.93)	94.76(4.17)	69.26(15.02)	92.02(4.44)	88.68(1.25)	75.00(10.05)	79.75(6.88)
6.DS	87.74(6.15)	84.20(3.77)	70.04(7.46)	60.29(7.96)	94.29(4.92)	75.56(11.34)	90.88(4.83)	87.79(1.39)	79.33(6.56)	81.12(6.04)
7.QDR-FarKNNB	91.09(5.68)	83.77(3.54)	72.56(8.49)	59.13(8.41)	**97.62(4.17)**	77.04(10.45)	90.03(5.27)	84.66(1.06)	**86.54(10.28)**	82.49(6.37)
8.QDR-NNRS	90.25(6.37)	83.04(5.38)	72.56(8.56)	**60.87(10.87)**	97.14(4.02)	77.41(6.43)	88.89(5.69)	84.34(1.75)	85.10(7.85)	82.18(6.32)
9.GADNR	94.15(3.98)	85.36(4.52)	74.01(6.21)	57.97(8.04)	93.33(5.14)	80.37(6.77)	**92.59(4.67)**	**89.14(1.28)**	85.58(3.39)	**83.61(4.89)**
10.GAWQD	88.02(6.27)	84.20(5.29)	71.48(7.19)	55.94(8.70)	93.81(3.92)	**82.59(8.38)**	91.74(6.55)	87.90(1.13)	84.13(12.71)	82.20(6.68)
11.GAFDQN	90.81(3.79)	85.22(5.34)	73.65(9.80)	56.23(8.57)	93.33(5.12)	80.37(7.25)	92.31(5.26)	87.99(1.47)	84.62(5.34)	82.73(5.77)

Table 8
The classification performance based on SVM.

Algorithms	Air	Australian	Breast	Bupa	Glass	Heart	Ionosphere	Phoneme	Sonar	Average
1.Raw	84.69(10.23)	85.51(4.32)	72.90(10.79)	61.22(10.32)	89.52(4.38)	71.48(8.91)	91.17(2.51)	85.16(1.98)	81.33(8.72)	80.33(6.91)
2.GANRS	76.88(4.14)	85.51(4.58)	73.28(6.16)	60.90(5.93)	90.95(5.24)	74.44(9.95)	90.05(5.01)	82.85(1.64)	76.48(8.44)	79.04(5.68)
3.NNRS	84.68(5.10)	85.51(5.34)	74.37(7.85)	59.08(9.40)	88.57(8.16)	82.59(5.25)	92.29(4.06)	75.50(1.64)	82.19(7.17)	80.53(6.00)
4.FarKNNB	76.59(6.37)	85.51(3.48)	75.45(9.92)	57.06(5.93)	90.48(5.02)	82.59(5.25)	91.46(4.24)	76.07(1.47)	76.81(9.13)	79.11(5.65)
5.GAVRNR	79.66(8.41)	85.51(4.32)	75.19(6.36)	62.58(8.33)	89.52(7.03)	74.07(11.97)	93.17(4.88)	85.05(1.01)	72.17(8.48)	79.66(6.75)
6.DS	79.09(10.59)	85.80(6.69)	74.03(4.56)	**62.67(7.42)**	88.57(7.51)	81.11(8.63)	90.88(5.52)	83.03(1.65)	79.79(5.59)	80.55(6.46)
7.QDR-FarKNNB	83.83(4.78)	85.51(3.48)	75.07(6.60)	61.20(7.83)	87.14(6.37)	77.41(6.40)	91.44(4.48)	84.34(1.81)	**82.76(5.94)**	80.97(5.30)
8.QDR-NNRS	84.67(4.00)	85.51(3.62)	72.94(6.55)	61.43(6.21)	89.52(7.38)	77.78(9.07)	91.71(4.75)	84.59(1.55)	80.24(5.42)	80.93(5.39)
9.GADNR	**91.93(2.75)**	86.23(5.60)	73.65(4.75)	62.61(6.65)	90.95(6.90)	77.04(5.18)	**94.02(3.42)**	**85.34(1.79)**	82.71(5.04)	**82.72(4.68)**
10.GAWQD	85.52(6.37)	85.51(4.37)	**75.82(8.79)**	62.62(10.07)	**92.38(6.02)**	**82.59(7.21)**	92.87(4.91)	82.74(1.47)	82.19(7.55)	82.47(6.31)
11.GAFDQN	87.46(4.60)	**86.96(3.55)**	75.46(8.34)	61.97(7.32)	89.05(8.11)	80.74(9.53)	93.16(3.36)	82.94(1.78)	80.33(7.85)	82.01(6.05)

- In most data sets, the FDN model has a unique neighborhood granular structure, which is why algorithm GAFDN takes less time than algorithm GADNR.

5.3. Classification performance

The third experiment is the classification accuracy analysis. In the experiment, kNN with $k = 3$ and SVM are used as classifiers, respectively. Table 6 represents the number of attributes obtained by each algorithm after attribute reduction. Tables 7 and 8 represent the classification accuracy rates of attribute reduction algorithms. The optimal algorithm corresponding to each data set is represented in bold. From Tables 6 to 8, we can obtain some important information:

- In Table 6, we can find that compared with the original data set, the attribute reduction algorithms based on rough sets select fewer data attributes, complete the task of attribute reduction better, and reduce the redundancy of data.
- From Tables 7 and 8, we can find that the three models GADNR, GAWDN, and GAFDN, all show a significant improvement in classification accuracy compared to the Raw data. Remarkably, the GADNR algorithm improves by 3.75% in the kNN classifier.
- In most data sets, algorithms 7 to 11 have higher classification accuracy compared to Algorithm DS. Therefore, it is necessary to combine some rough set models with the discretization scheme of information systems.
- In most data sets, the algorithm GADNR achieves the best results. Our discretization neighborhood relation structure has been verified, which has a high ability to find uncertainty data.
- In particular, in SVM classifier, algorithm GAWDN achieves better results than GADNR on some data sets. Therefore, it is necessary to determine the weights of the distances between objects.

6. Conclusion

In the paper, we have proposed a new approach to discretizing information systems associated with neighborhood rough sets to reduce the running time for attribute reduction algorithms. This may be considered as a promising aspect to shorten the construction time of neighborhood granular structures. We have investigated three novel bilayer granular structures, namely, quasi discretization-based neighborhood, weighted quasi discretization-based neighborhood, and fast discretization neighborhood. We have also explored two evaluation indices of binary relations and estimated comprehensively our models by comparing them with other neighborhood rough set models. Furthermore, we have discussed the properties of DNR, WDN, and FDN models. Finally, based on the three models, we have designed three attribute reduction algorithms. The experimental results have demonstrated that the proposed models greatly reduce the running time of the algorithm and improve the classification ability of neighborhood rough sets.

In future research, it is important to carefully consider the potential effects of neighborhood rough sets. For example, most of the existing rough set models use dependency (uncertainty representation) as a criterion for attribute reduction. In the paper, we have proposed two evaluation indices of binary relations, namely, Gaussian balance index and quality index, which can be used as judgment conditions for determining uncertainty in attribute reduction. These potential applications may lead to further developments and improvements in neighborhood rough set models.

CRediT authorship contribution statement

Di Zhang: Writing - original draft, Conceptualization, Formal analysis, Methodology, Data curation, Investigation, Validation, Visualization. **Ping Zhu:** Conceptualization, Formal analysis, Writing - review & editing, Supervision, Funding acquisition, Project administration.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Ping Zhu reports financial support was provided by National Natural Science Foundation of China under Grant 62172048.

Data availability

The authors do not have permission to share data.

Acknowledgements

The author would like to thank the anonymous reviewers for their helpful comments. This work was supported by the National Natural Science Foundation of China under Grant 62172048.

References

[1] A.F. Attia, M. Abd Elaziz, A.E. Hassanien, R.A. El-Sehiemy, Prediction of solar activity using hybrid artificial bee colony with neighborhood rough sets, IEEE Trans. Comput. Soc. Syst. 7 (5) (2020) 1123–1130.
[2] C. Bai, K. Govindan, A. Satir, H. Yan, A novel fuzzy reference-neighborhood rough set approach for green supplier development practices, Ann. Oper. Res. (2019) 1–35.
[3] B. Barman, S. Patra, A novel technique to detect a suboptimal threshold of neighborhood rough sets for hyperspectral band selection, Soft Comput. 23 (24) (2019) 13709–13719.
[4] H. Chen, T. Li, Y. Cai, C. Luo, H. Fujita, Parallel attribute reduction in dominance-based neighborhood rough set, Inf. Sci. 373 (2016) 351–368.
[5] J. Chen, P. Zhu, A variable precision multigranulation rough set model and attribute reduction, Soft Comput. 27 (1) (2023) 85–106.
[6] Y. Chen, Y. Xue, Y. Ma, F. Xu, Measures of uncertainty for neighborhood rough sets, Knowl.-Based Syst. 120 (2017) 226–235.
[7] Y. Chen, Z. Zeng, J. Lu, Neighborhood rough set reduction with fish swarm algorithm, Soft Comput. 21 (2017) 6907–6918.
[8] Y. Chen, Z. Zhang, J. Zheng, Y. Ma, Y. Xue, Gene selection for tumor classification using neighborhood rough sets and entropy measures, J. Biomed. Inform. 67 (2017) 59–68.
[9] Z. Chen, X. Ming, T. Zhou, Y. Chang, Sustainable supplier selection for smart supply chain considering internal and external uncertainty: an integrated rough-fuzzy approach, Appl. Soft Comput. 87 (2020) 106004.

[10] Y. Cheng, F. Zhao, Q. Zhang, G. Wang, A survey on granular computing and its uncertainty measure from the perspective of rough set theory, Granul. Comput. 6 (2021) 3–17.
[11] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, Int. J. Gen. Syst. 17 (2–3) (1990) 191–209.
[12] L. D'eer, C. Cornelis, L. Godo, Fuzzy neighborhood operators based on fuzzy coverings, Fuzzy Sets Syst. 312 (2017) 17–35.
[13] S. Greco, B. Matarazzo, R. Slowinski, Rough approximation of a preference relation by dominance relations, Eur. J. Oper. Res. 117 (1) (1999) 63–83.
[14] S. Greco, B. Matarazzo, R. Slowinski, Rough approximation by dominance relations, Int. J. Intell. Syst. 17 (2) (2002) 153–171.
[15] M.K. Gupta, P. Chandra, A comprehensive survey of data mining, Int. J. Inf. Technol. 12 (4) (2020) 1243–1257.
[16] M. Hu, E.C. Tsang, Y. Guo, D. Chen, W. Xu, A novel approach to attribute reduction based on weighted neighborhood rough sets, Knowl.-Based Syst. 220 (2021) 106908.
[17] Q. Hu, J. Liu, D. Yu, Mixed feature selection based on granulation and approximation, Knowl.-Based Syst. 21 (4) (2008) 294–304.
[18] Q. Hu, D. Yu, J. Liu, C. Wu, Neighborhood rough set based heterogeneous feature subset selection, Inf. Sci. 178 (18) (2008) 3577–3594.
[19] Q. Hu, D. Yu, Z. Xie, Neighborhood classifiers, Expert Syst. Appl. 34 (2) (2008) 866–876.
[20] K. Liu, T. Li, X. Yang, H. Chen, J. Wang, Z. Deng, SemiFREE: semi-supervised feature selection with fuzzy relevance and redundancy, IEEE Trans. Fuzzy Syst. 31 (10) (2023) 3384–3396.
[21] R.A. Ibrahim, M. Abd Elaziz, D. Oliva, S. Lu, An improved runner-root algorithm for solving feature selection problems based on rough sets and neighborhood rough sets, Appl. Soft Comput. 97 (2020) 105517.
[22] P. Jain, T. Som, Multigranular rough set model based on robust intuitionistic fuzzy covering with application to feature selection, Int. J. Approx. Reason. 156 (2023) 16–37.
[23] F. Jiang, Y. Sui, A novel approach for discretization of continuous attributes in rough set theory, Knowl.-Based Syst. 73 (2015) 324–334.
[24] H. Jiang, J. Zhan, D. Chen, Covering-based variable precision $(\mathcal{I},\mathcal{T})$-fuzzy rough sets with applications to multiattribute decision-making, IEEE Trans. Fuzzy Syst. 27 (8) (2018) 1558–1572.
[25] S. Kotsiantis, D. Kanellopoulos, Discretization techniques: a recent survey, GESTS Int. Trans. Comput. Sci. Eng. 32 (1) (2006) 47–58.
[26] T.Y. Lin, et al., Granular computing on binary relations I: data mining and neighborhood systems, in: Rough Sets in Knowledge Discovery, 1998, pp. 107–121.
[27] K. Liu, T. Li, X. Yang, X. Yang, D. Liu, P. Zhang, J. Wang, Granular cabin: an efficient solution to neighborhood learning in big data, Inf. Sci. 583 (2022) 189–201.
[28] Y. Liu, W. Huang, Y. Jiang, Z. Zeng, Quick attribute reduct algorithm for neighborhood rough set model, Inf. Sci. 271 (2014) 65–81.
[29] S. Luo, D. Miao, Z. Zhang, Y. Zhang, S. Hu, A neighborhood rough set model with nominal metric embedding, Inf. Sci. 520 (2020) 373–388.
[30] N. Mac Parthalain, Q. Shen, Exploring the boundary region of tolerance rough sets for feature selection, Pattern Recognit. 42 (5) (2009) 655–667.
[31] H.S. Nguyen, Discretization problem for rough sets methods, in: Rough Sets and Current Trends in Computing, Springer, Berlin/Heidelberg, 1998, pp. 545–552.
[32] Y. Pan, W. Xu, Q. Ran, An incremental approach to feature selection using the weighted dominance-based neighborhood rough sets, Int. J. Mach. Learn. Cybern. 14 (4) (2023) 1217–1233.
[33] Z. Pawlak, Rough sets, Int. J. Comput. Inf. Sci. 11 (1982) 341–356.
[34] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning About Data, vol. 9, Springer Science & Business Media, 1991.
[35] Y. Ping, L. Yongheng, Neighborhood rough set and SVM based hybrid credit scoring classifier, Expert Syst. Appl. 38 (9) (2011) 11300–11304.
[36] L. Shen, F.E. Tay, L. Qu, Y. Shen, Fault diagnosis using rough sets theory, Comput. Ind. 43 (1) (2000) 61–72.
[37] W. Shu, J. Yu, Z. Yan, W. Qian, Semi-supervised feature selection for partially labeled mixed-type data based on multi-criteria measure approach, Int. J. Approx. Reason. 153 (2023) 258–279.
[38] L. Sun, T. Wang, W. Ding, J. Xu, Y. Lin, Feature selection using Fisher score and multilabel neighborhood rough sets for multilabel classification, Inf. Sci. 578 (2021) 887–912.
[39] L. Sun, W. Wang, J. Xu, S. Zhang, Improved LLE and neighborhood rough sets-based gene selection using Lebesgue measure for cancer classification on gene expression data, J. Intell. Fuzzy Syst. 37 (4) (2019) 5731–5742.
[40] L. Sun, J. Zhang, W. Ding, J. Xu, Mixed measure-based feature selection using the Fisher score and neighborhood rough sets, Appl. Intell. (2022) 1–25.
[41] L. Sun, X. Zhang, J. Xu, S. Zhang, An attribute reduction method using neighborhood entropy measures in neighborhood rough sets, Entropy 21 (2) (2019) 155.
[42] D. Tian, X. Zeng, J. Keane, Core-generating approximate minimum entropy discretization for rough set feature selection in pattern classification, Int. J. Approx. Reason. 52 (6) (2011) 863–880.
[43] D. Tripathi, D.R. Edla, R. Cheruku, Hybrid credit scoring model using neighborhood rough set and multi-layer ensemble classification, J. Intell. Fuzzy Syst. 34 (3) (2018) 1543–1549.
[44] C. Wang, M. Shao, Q. He, Y. Qian, Y. Qi, Feature subset selection based on fuzzy neighborhood rough sets, Knowl.-Based Syst. 111 (2016) 173–179.
[45] C. Wang, Y. Shi, X. Fan, M. Shao, Attribute reduction based on k-nearest neighborhood rough sets, Int. J. Approx. Reason. 106 (2019) 18–31.
[46] Q. Wang, Y. Qian, X. Liang, Q. Guo, J. Liang, Local neighborhood rough set, Knowl.-Based Syst. 153 (2018) 53–64.
[47] J. Xie, B.Q. Hu, H. Jiang, A novel method to attribute reduction based on weighted neighborhood probabilistic rough sets, Int. J. Approx. Reason. 144 (2022) 1–17.
[48] J. Xu, K. Qu, Y. Sun, J. Yang, Feature selection using self-information uncertainty measures in neighborhood information systems, Appl. Intell. 53 (4) (2023) 4524–4540.
[49] D. Yan, D. Liu, Y. Sang, A new approach for discretizing continuous attributes in learning systems, Neurocomputing 133 (2014) 507–511.
[50] X. Yang, X. Li, T.Y. Lin, First GrC model-neighborhood systems the most general rough set models, in: 2009 IEEE International Conference on Granular Computing, Man & Cybernetics, 2003, pp. 3188–3193.
[51] X. Yang, M. Zhang, H. Dou, J. Yang, Neighborhood systems-based rough sets in incomplete information system, Knowl.-Based Syst. 24 (6) (2011) 858–867.
[52] X. Yang, S. Liang, H. Yu, S. Gao, Y. Qian, Pseudo-label neighborhood rough set: measures and attribute reductions, Int. J. Approx. Reason. 105 (2019) 112–129.
[53] Y. Yao, Information granulation and rough set approximation, Int. J. Intell. Syst. 16 (1) (2001) 87–104.
[54] Y. Yao, Decision-theoretic rough set models, in: Rough Sets and Knowledge Technology: Second International Conference, RSKT 2007, Toronto, Canada, May 14–16, 2007, in: Proceedings, vol. 2, Springer, 2007, pp. 1–12.
[55] B. Yu, Y. Hu, Y. Kang, M. Cai, A novel variable precision rough set attribute reduction algorithm based on local attribute significance, Int. J. Approx. Reason. 157 (2023) 88–104.
[56] W. Yu, M. Zhang, Y. Shen, Learning a local manifold representation based on improved neighborhood rough set and LLE for hyperspectral dimensionality reduction, Signal Process. 164 (2019) 20–29.
[57] D. Zhang, P. Zhu, Variable radius neighborhood rough sets and attribute reduction, Int. J. Approx. Reason. 150 (2022) 98–121.
[58] H. Zhang, Q. Sun, K. Dong, Information-theoretic partially labeled heterogeneous feature selection based on neighborhood rough sets, Int. J. Approx. Reason. 154 (2023) 200–217.
[59] L. Zhang, P. Zhu, Generalized fuzzy variable precision rough sets based on bisimulations and the corresponding decision-making, Int. J. Mach. Learn. Cybern. 13 (8) (2022) 2313–2344.
[60] W. Ziarko, Variable precision rough set model, J. Comput. Syst. Sci. 46 (1) (1993) 39–59.
[61] L. Zou, H. Li, W. Jiang, X. Yang, An improved fish swarm algorithm for neighborhood rough set reduction and its application, IEEE Access 7 (2019) 90277–90288.