International Journal of Machine Learning and Cybernetics
https://doi.org/10.1007/s13042-022-01695-4

ORIGINAL ARTICLE

An incremental approach to feature selection using the weighted dominance-based neighborhood rough sets

Yanzhou Pan[1] · Weihua Xu[1] · Qinwen Ran[1]

Received: 19 July 2022 / Accepted: 13 October 2022

Abstract
Dominance-based neighborhood rough set (DNRS) is capable to give qualitative and quantitative descriptions of the relations between ordered objects. In spite of its effectiveness in feature selection, DNRS ignores the various significance of features. In fact, different features exert different impacts on decision-making. Once we explore these differences in advance, it is easier to find out features with high correlation and dependency. Likewise, it is inevitable that in big-data era the objects may update from time to time, which calls for efficient attribute reduction. However, the existing approaches are inappropriate for the weighted and ordered data. Motivated by these two deficiencies, first, we assign different weights to conditional attributes and establish the weighted dominance-based neighborhood rough set (WDNRS). Then a kind of conditional entropy in matrix form and ensuing updating principles are put forward to evaluate the significance of the attributes. In addition, grounded on the entropy, we come up with the heuristic algorithm and corresponding incremental mechanism when objects increase. Finally, twelve experiments are carried out to verify that it is effective and efficient for the designed method to select features in dynamic datasets.

Keywords Dynamic ordered data · Feature selection · Incremental learning · Weighted dominance-based neighborhood rough sets

1 Introduction

Feature selection, also referred to as attribute reduction, is of great use in the era of big data. It aims to extract necessary features from high-dimensional data regarded as good reflection of the whole datasets. Namely, uncorrelated or insignificant attributes are to be removed on the condition of the maintenance of the knowledge bases ability to classify. Then we reach the targets of reducing dimensionality, improving interpretability and moreover saving time and computational space. With a lot advancement, attribute reduction still remains a vigorous and growing research area. For example, machine learning is an essential method to obtain knowledge from a mass of data [1–6]. Some scholars also connect attribute reduction with graphs [7, 8].

Rough set theory (RST), proposed by Pawlak [9], is also an important and effective theoretic tool aiming at accessing datasets with inconsistency and uncertainty. With no need for prior knowledge, RST completely depends on the existing data. Pawlak regarded the relation produced by the conditional sets and decision sets as equal relation, which, however, is not exact in real life. Then some researchers have introduced varieties of relations such as similarity relations, neighborhood relations, and dominance-based relations into the original rough sets. On that basis, scholars have proposed many representative models such as neighborhood rough set (NRS) [10], variable precision rough sets (VPRS) [11] and dominance-based rough set (DRS) [12]. And in this study, we pay close attention to the neighborhood relation and the dominance-based relation.

It is well-described when using neighborhood relations to extract the similarity among samples and easier to evaluate the information systems with real-valued attributes. Meanwhile, Wang et al. constructed approximation operators on neighborhood systems [13]. Relevant classifiers was

✉ Weihua Xu
chxuwh@gmail.com

Yanzhou Pan
panyzhou98@163.com

Qinwen Ran
2036978388@qq.com

[1] College of Artificial Intelligence, Southwest University, Chongqing 400715, People's Republic of China

Published online: 03 November 2022

the focus of research for Hu et al. in [14]. And for attribute reduction, Sun et al. applied entropy [15]. [16] considered feature selection with missing labels and introduced multi-label fuzzy neighborhood rough sets. What's more, in [17], for imbalanced data, the authors designed a novel adaptive fuzzy neighborhood-based feature selection method with adaptive synthetic over-sampling.

Nowadays, more and more scholars have kept an eye on the dominance relation, which is pretty common in daily life. For example, the grades of students can be reckoned as having internal ordered characteristic, as well as the credit level and ranks of school. Namely, we can obtain the dominance principle by ranking attribute-values. In order to present this peculiarity, Greco et al. created dominance-based rough set approach (DRSA) [12], and successfully extended it to multi-criteria decision analysis [18]. Afterwards, many scholars have been developing stretched DRSA under different conditions, including generalized dominance rough sets [19] and soft dominance based rough sets [20]. And when connected with neighborhood relation, DRSA also performed well. Chen et al. proposed Dominance-based Neighborhood Rough Sets (DNRS) and further investigated its feature selection [21]. Later, the authors came up with parallel attribute reduction algorithms [22]. And in [23], Wan et al. probed feature interaction for hybrid attribute reduction. In spite of an effective attribute reduction, these approaches mentioned above are all short of the consideration of two respects.

For one thing, there are less likely for each attribute to have the same weight as the others. Accordingly, it is better and necessary to assign different weights to different features. And the weight of an attribute stands for its importance when exploring the more significant conditional attributes for the decision one. Namely, we have already highlighted features highly-correlated to the decision in advance. Scholars have studied ways to assign weights of attributes under varieties of conditions. Wang et al. introduced a subclass-weighted classifier to cope with imbalanced data [24]. Vluymansa et al. came up with a weighted Parzen window function to access the probability from the weighted subsampled data [25]. And In Hu' s study, various features were endowed with different wights in neighborhood rough set [26]. However, the aforementioned papers were not involved with the dominance relation. That is the reason why we attempt to attach great importance to those attributes that are highly related to decision-making in feature selection under the background of the dominance relation. And this paper utilizes the correlation coefficients of attributes in accordance with decisions to acquire the weights which reflects the significance of attributes. For another, with the rapid update in the big era, static methods are no longer appropriate for dynamic information systems, which stimulate the rise of the incremental learning. Incremental learning is so conducive in the information age, a period flooded with massive changing data, for the reason that it is grounded on the existing results and keeps on acquiring knowledge from incoming data with no need to reconsider the original ones. In other words, if we use non-incremental approaches, aiming to access static datasets, it is a waste of time and calculation space. And nowadays, scholars have been developing numerous incremental methods, which can be divided into three aspects: objects-oriented, attribute-oriented and attribute values-oriented.

For the variation of objects, Liang et al. applied information entropy to incremental learning [27]. Sang et al. proposed matrix-based dominance conditional entropy when massive objects are the addition (IAR-A) or deletion (IAR-D) of an ordered system [28]. Furthermore, the authors combined the strengths of dominance-based neighborhood rough sets and fuzzy dominance rough sets to design fuzzy dominance neighborhood rough sets to access datasets with noise [29]. And in [30], Yuan et al. developed a dynamic algorithm based on progressive fuzzy three-way concept. For the variation of attributes, scholars have proposed several incremental algorithms related to rough sets for attribute reduction with dynamically adding attributes. Jing et al. introduced an matrix-grounded incremental algorithm to calculate knowledge granular under the variation in attributes [31]. Chen et al. refined concepts of the discernibility matrix and the discernible relation, which is considered as the criterion of the reduction rather than information granular or information entropy for dynamic-attribute datasets [32]. Except for variation of just one factor, furthermore, scholars have studied incremental mechanisms with at least two elements changing. For example, Dong et al. presented a unified approach for accessing dynamic datasets with objects and attributes increase at the same time [33]. In [34], Zhang et al. considered the simultaneous variation of information sources and attributes when fusing data.

Nowadays, scholars have been applying matrix form of information to dynamic datasets attribute reduction. For the reason that it is convenient to update the sub-matrix other than constructing the whole from scratch. It was initially applied in dynamic datasets with an individual variation [28, 29, 35]. Furthermore, Huang et al. utilized this approach in dynamic fuzzy decision systems on the simultaneous variation of samples and features [36]. Wang et al. proposed specific multi-dimensional algorithms when attribute values and objects vary at the same time [37], and the author stretched to the coinstantaneous variation of both attributes and objects [38]. It is apparent that most of the incremental attribute reduction approaches aforementioned are irrelevant to a dynamic information system combined the preference-order relation with different-weighted attributes. Therefore, the incremental mechanisms to attribute reduction in existence are poor to fit dynamic ordered datasets and easier to make samples misclassified, which are the motivation of this study. Whats more, in reality, chances are that the adding data change the weight of each feature, which brings about an upgrade of the weight. If we stick to the origin assignment of the weight, it is possible to reduce precision. And that is also the reason why we combine the weighted attributes with dynamic information system.

And uncertainty measure, used to evaluate the significance of attributes and quantify the inconsistency of data, attach great importance to feature selection. Information entropy, proposed by Shannon is the general measure [39]. Scholars have developed it such as the fuzzy information system designed by Hu et al. [40] . And for ordered datasets, the author proposed ascending and decreasing rank conditional entropy to access the consistency degree of the ranking data [41]. Sang et al. extended the ascending one to matrix form [28], which we are going to use in the subsequent study.

The contributions are summarized as follows: (1) We assign different weight to each attribute in dominance neighborhood rough sets(DNRS) and then define the weighted dominance neighborhood rough sets(WDNRS). (2) This paper applies the matrix to calculate the dominance conditional entropy. And on that basis, outer and inner significance measure can be obtained to find out the all the necessary attributes. Then dynamic updating mechanism is developed based on the static one with the variation of objects. (3) Experiments on twelve datasets downloaded from UCI are utilized to bear out that the designed model is conducive to promote the performance of classification. In addition, when objects increase, the ensuing incremental algorithms contribute to remarkably reduce the time and the space of calculation.

The rest of the paper is organized as follows: in Sect. 2, some related works are in retrospect, including the comparative approaches we consider in this study. Section 3 introduces the generation of the weight, which we connect with the dominance neighborhood rough sets. On the basis of Sect. 3, we come up with a matrix-based method to calculate the dominance conditional entropy in Sect. 4. Furthermore in Sect. 5, an updating mechanism is proposed and used to develop an incremental attribute reduction algorithm when adding objects from the information system. Extensive experiments are in presentation in Sect. 6. Eventually, we draw the conclusions in brief and picture future perspectives. And the brief framework of the paper can be seen in Fig. 1.

2 Related work

This section overviews some basic points of classical dominance-based neighborhood rough sets, which can be found in [9, 18, 22], and subsequent shortcomings.

2.1 Dominance-based neighborhood rough set

Definition 1 Given an information system with decision $IS = (U, AT \cup DT, h, g)$, where $U = \{x_1, x_2, \ldots, x_n\}$ is a nonempty and finite objects set; $AT = \{a_1, a_2, \ldots, a_p\}$ is a nonempty and finite set composed by conditional

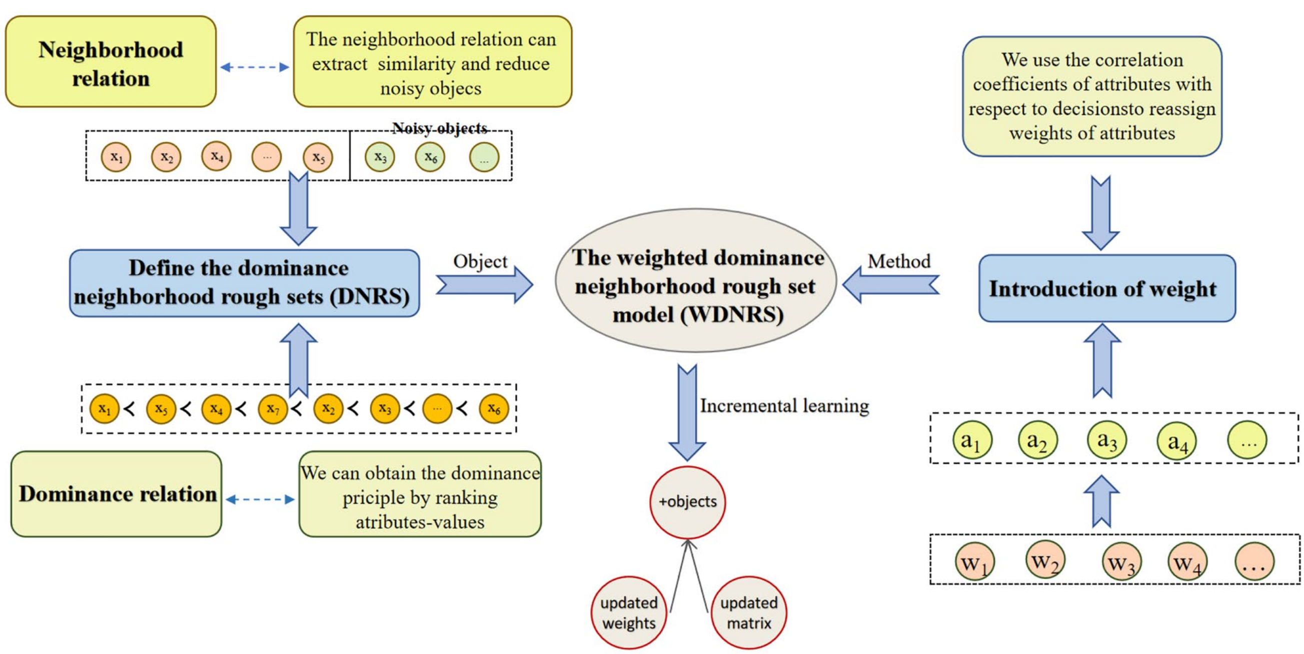

Fig. 1 The brief framework of the study

attributes, $DT = \{d_1, d_2, \dots, d_q\}$ is a nonempty and finite set constituted by decision attributes, and additionally, $AT \cap DT \neq \emptyset$; $h : U \times AT \longrightarrow V_a$ is the information function, and for $\forall a \in AT$, $\forall x \in U$, V_a is the finite domain of a; $g : U \times DT \longrightarrow V_d$ is the information function, and for $\forall d \in DT$, $\forall x \in U$, V_d is the finite domain of d. And this paper keeps an eye on the single-decision information system, namely, $DT = \{d\}$.

Definition 2 Given a decision information system ($ODIS$), $IS^{\succeq} = (U, AT \cup \{d\}, h, g)$, for any $\forall a \in AT$, all the domain of a is entirely pre-ordered by the relation $\succeq_a$. An increasing preference is defined as $\forall x, y \in U$, $y \succeq_a x \Leftrightarrow h(y, a) \geq h(x, a)$; similarly, a decreasing preference is defined as $y \succeq_a x \Leftrightarrow h(y, a) \leq h(x, a)$.

In real life, the decision-makers have already known the rank of criterion in accordance with prior knowledge. Considering simplicity and generality, this paper just includes criteria with increasing order.

Definition 3 Let an ODIS , $IS^{\succeq} = (U, AT \cup \{d\}, h, g)$, for $\forall A \subseteq AT$, $a \in A$, the dominance-based neighborhood relation $DN_A^{\succeq\lambda}$ on A is defined as

$$DN_A^{\succeq\lambda} = \{(x, y) \mid d_A(x, y) \leq \lambda \wedge h(x, a) \leq h(y, a)\}, \tag{1}$$

where $d_A(x, y) = \sqrt{\sum_{a \in A} [h(x, a) - h(y, a)]^2}$. d_A is a Euclidean distance function reflecting the distance between any two objects under attribute subset A and λ ($\lambda > 0$) is a threshold.

Moreover, d, as a decision attribute, the dominance relation on d is expressed as $D_d^{\lambda+} = \{(x, y) \mid g(x, d) \leq g(y, d)\}$ and $D_d^{\lambda-} = \{(x, y) \mid g(x, d) \geq g(y, d)\}$.

Definition 4 Let an ODIS, $IS^{\succeq} = (U, AT \cup \{d\}, h, g)$, for $\forall A \subseteq AT$, $\forall x \in U$, the dominating neighborhood sets and dominated neighborhood sets according to A are defined as

$$\begin{aligned} DN_A^{\lambda+}(x) &= \{y \in U \mid y\, DN_A^{\succeq\lambda}\, x\}, \\ DN_A^{\lambda-}(x) &= \{y \in U \mid x\, DN_A^{\succeq\lambda}\, y\}. \end{aligned} \tag{2}$$

In an ODIS, $\{d\}$ is a DT with a single decision attribute, $U/d = \{d^1, d^2, \dots, d^t\}(t \leq |U|)$, where preference-order relationship is suitable to d^t, and $d^t \succeq \dots d^{t'} \succeq d^1$.

Definition 5 Let an ODIS, $IS^{\succeq} = (U, AT \cup \{d\}, h, g)$, for $\forall A \subseteq AT$, $X \subseteq U$, the lower and upper approximations of the subset X are defined as

$$\begin{aligned} \underline{DN_A^{\succeq\lambda}}(X) &= \{x \in X \mid DN_A^{\lambda+}(x) \subseteq X\}, \\ \overline{DN_A^{\succeq\lambda}}(X) &= \{x \in X \mid DN_A^{\lambda+}(x) \cap X \neq \emptyset\}, \end{aligned} \tag{3}$$

where $\underline{DN_A^{\succeq\lambda}}(X)$ and $\overline{DN_A^{\succeq\lambda}}(X)$ are a pair of approximation operators. If $\underline{DN_A^{\succeq\lambda}}(X) = \overline{DN_A^{\succeq\lambda}}(X)$, then X with respect to $\underline{DN_A^{\succeq\lambda}}(X)$ and $\overline{DN_A^{\succeq\lambda}}(X)$ is accurate; otherwise, X is rough.

Property 1 *According to Definition 5, we can easily get that* $\underline{DN_A^{\succeq\lambda}}(X) \subseteq X \subseteq \overline{DN_A^{\succeq\lambda}}(X)$.

Definition 6 Let an ODIS, $IS^{\succeq} = (U, AT \cup \{d\}, h, g)$, $U/d = \{d^1, d^2, \dots, d^t\}$, for $\forall A \subseteq AT$, the upper and lower approximations of d with respect to A are defined as

$$\begin{aligned} \underline{DN_A^{\succeq\lambda}}(d) &= \bigcup_{i=1}^{t} \underline{DN_A^{\succeq\lambda}}(d^i), \\ \overline{DN_A^{\succeq\lambda}}(d) &= \bigcup_{i=1}^{t} \overline{DN_A^{\succeq\lambda}}(d^i). \end{aligned} \tag{4}$$

And the boundary and the positive regions of d with respect to A are defined as

$$\begin{aligned} BN_A^{\succeq\lambda}(d) &= \overline{DN_A^{\succeq\lambda}}(d) - \underline{DN_A^{\succeq\lambda}}(d), \\ POS_A^{\succeq\lambda}(d) &= \bigcup_{d^i \in U/d} \underline{DN_A^{\succeq\lambda}}(d^i). \end{aligned} \tag{5}$$

Definition 7 Let an ODIS, $IS^{\succeq} = (U, AT \cup \{d\}, h, g)$, for $\forall A \subseteq AT$, the dependency degree of d with respect to A is defined as

$$\gamma_A^{\succeq\lambda}(d) = \frac{|POS_A^{\succeq\lambda}(d)|}{|U|}, \tag{6}$$

where $|\cdot|$ is the representation of the cardinality of the set. $\gamma_A^{\succeq\lambda}(d)$ is the reflection of the ability to approximate $\{d\}$ of subset A. And a larger $\gamma_A^{\succeq\lambda}(d)$ corresponds to the subset's stronger ability to approximate.

Additionally, the value of the dependency degree depends on two factors. One is the value of λ, and the other is the range of the subset A. As for λ, the larger the value of λ is, the smaller the $\gamma_A^{\succeq\lambda}(d)$ is. While for A, with more attributes included in A, the dependency degree grows larger.

2.2 Shortcomings of dominance neighborhood rough sets

When calculating the dominance-based neighborhood classes of the DNRS, we generally regard the weight of each attribute as 1. In other words, each attribute has the same degree of importance in feature selection, which contributes to inadequate exploration of the internal relationship among conditions and decisions. Additionally, the attributes

Table 1 A decision information system

U	a_1	a_2	a_3	a_4	d
x_1	0.28	0.89	0.21	0.29	1
x_2	0.30	0.90	0.19	0.26	1
x_3	0.48	0.51	0.20	0.39	1
x_4	0.50	0.52	0.26	0.38	1
x_5	0.61	0.69	0.29	0.35	1
x_6	0.39	0.71	0.27	0.18	2
x_7	0.37	0.65	0.33	0.24	2
x_8	0.62	0.90	0.38	0.20	2
x_9	0.76	0.86	0.35	0.36	2
x_{10}	0.89	0.50	0.39	0.28	2

Table 2 Information granules

U	$DN_{A_1}^{\geq\lambda}(x)$	$DN_{A_2}^{\geq\lambda}(x)$
x_1	$\{x_1, x_2\}$	$\{x_1, x_4, x_5\}$
x_2	$\{x_2\}$	$\{x_1, x_2, x_4\}$
x_3	$\{x_3\}$	$\{x_3\}$
x_4	$\{x_4\}$	$\{x_4, x_5, x_7\}$
x_5	$\{x_5\}$	$\{x_5, x_9\}$
x_6	$\{x_6\}$	$\{x_6, x_7\}$
x_7	$\{x_6, x_7\}$	$\{x_7, x_{10}\}$
x_8	$\{x_8\}$	$\{x_8, x_{10}\}$
x_9	$\{x_9\}$	$\{x_9\}$
x_{10}	$\{x_{10}\}$	$\{x_{10}\}$

with larger attribute-values are more likely to be selected if assigned the same weights. Next, an instance will show the weakness of the classical DNRS and verify the necessity of the assignment of various weights.

Example 2.1 Given a decision information table $IS^{\geq} = (U, AT \cup \{d\}, h, g)$ is shown in Table 1, where sample set is $U = \{x_1, x_2, \cdots, x_{10}\}$, conditional attribute set is $AT = \{a_1, a_2, a_3, a_4\}$, decision attribute set is $D = \{d\}$. These samples are divided into two parts $d^1 = \{x_1, x_2, x_3, x_4, x_5\}$ and $d^2 = \{x_6, x_7, x_8, x_9, x_{10}\}$ by d. Given two attribute subsets $A_1 = \{a_1, a_2\}, A_2 = \{a_3, a_4\}$ and a neighborhood threshold $\lambda = 0.1$, the dominance neighborhood information granules are formed by A_1 and A_2 under $\lambda = 0.1$ are shown in Table 2.

From Table 2, we know that the granularity of neighborhood information granules induced by A_1 is finer than that induced by A_2. According to the definition of $POS_A^{\geq\lambda}$ and Table 2, we can get that $POS_{A_1}^{\geq\lambda} = U$ and $POS_{A_2}^{\geq\lambda} = \{x_1, x_2, x_3, x_6, x_7, x_8, x_9, x_{10}\}$. Therefore,we can get that $\gamma_{A_1}^{\geq\lambda}(d) = 1$ and $\gamma_{A_2}^{\geq\lambda}(d) = 0.8$. So from the above analysis, it is obvious that the ability of A_1 to approximate d is better than that of A_2. We then use KNN, LSVC and LR to classify unknown regions, and the classification results are shown in Fig. 2. As is shown in the picture, the separability of attribute subset A_2 is significantly higher than that of A_1. Under attribute subset A_1, some samples are misclassified by KNN, LSVC and LR (see Fig. 2a–c). Under attribute subset A_2, all samples can be correctly classified by KNN, LSVC and LR (see Fig. 2d–f). A_1 approximates d better than A_2, but A_1 is less separable with respect to d than A_2. Therefore, it is

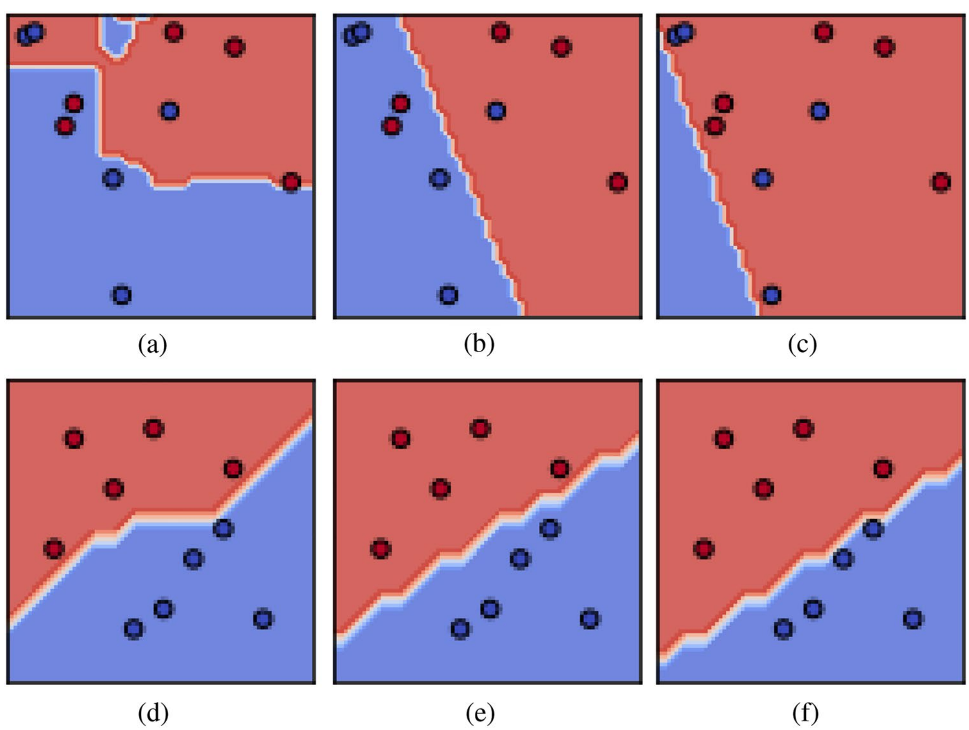

Fig. 2 Classification results under three classifiers on A_1 and A_2

limited to use the dominance neighborhood rough set model to measure the importance of attribute subsets. Next, we will introduce an effective model to measure the importance of attribute subsets.

3 The weighted dominance-based neighborhood rough sets

As illustrated in the above example, one of the disadvantages of DNRS is that each conditional attribute has the same weight when selecting necessary and non-redundant features. However, different weights are conducive to reveal how important each attribute is to decision-making. Then we are to propose a new rough set model named the weighted dominance-based neighborhood rough set(WDNRS) to assess the importance of feature subsets.

3.1 The generation of the *w*

Next, we give the generation process of w.

Let an ODIS , $IS^{\succeq} = (U, AT \cup \{d\}, h, g)$, for $\forall a \in AT$, $\forall x \in U$, $h(x, a)$ is the value of object with respect to a. Let the coefficient matrix be

$$M_A = \begin{pmatrix} h(x_1, a_1) & h(x_1, a_2) & \cdots & h(x_1, a_p) \\ h(x_2, a_1) & h(x_2, a_2) & \cdots & h(x_2, a_p) \\ \vdots & \vdots & \vdots & \vdots \\ h(x_n, a_1) & h(x_n, a_2) & \cdots & h(x_n, a_p) \end{pmatrix},$$

the vector of the decision attribute d be $M_d = \left(g(x_1, d), g(x_2, d), \ldots, g(x_n, d)\right)^T$, and the partition coefficients of attributes be $\eta = \left(\eta(a_1), \eta(a_2), \ldots, \eta(a_p)\right)^T$. In order to find out the optimum, in turn, we are going to solve the optimization problem

$$\eta^\star = argmin\|M_A\eta - M_d\|, \tag{7}$$

where $\|\bullet\|^2$ is the representation the 2-norm of a vector. To solve this, first assuming $M_A\eta = M_d$, then $(M_A)^T$ are multiplied to both sides to get $(M_A)^T M_A \eta = (M_A)^T M_d$. Finally, we can get $\eta = \left((M_A)^T M_A\right)^{-1} (M_A)^T M_d$.

Particularly, when the matrix $(M_A)^T M_A$ is not invertible, or a penalty term is in need in the optimum function, we transform equation to $J(\eta) = \|M_A\eta - M_d\|^2 + \|\eta\|^2$. Since $J(\eta)$ is a convex function, the minimum of $J(\eta)$ can be attained when $J'(\eta) = 0$. And that is to say $J'(\eta) = 2(M_A)^T (M_A\eta - M_d) + 2\eta = 0$. Hence, there is $\left(M_A^T M_A + E\right)\eta = M_A^T M_d$, where E is an identity matrix. Then in this situation, $\eta = \left(M_A^T M_A + E\right)^{-1} M_A^T M_d$. And if $M_A^T M_A$ or $M_A^T M_A + E$ is high dimensional or near to ill-conditioned, subfunction ”np.linalg.solve” in Numpy , is used to directly solve $M_A^T M_A \eta = M_A^T M_d$ or $\left(M_A^T M_A + E\right)\eta = M_A^T M_d$, rather than the inverse matrix.

Additionally, $|\eta(a)|$ is the absolute value of $\eta(a)$, and is the reflection of the relation between condition attribute a and decision attribute d. The larger the absolute value is, the stronger the internal correlation is.

Definition 8 Let an ODIS , $IS^{\succeq} = (U, AT \cup \{d\}, h, g)$, $\forall a \in AT$, the weight of a can be defined as

$$\omega(a) = \frac{|AT||\eta(a)|}{\sum_{a_i \in AT} \left|\eta(a_i)\right|}. \tag{8}$$

Property 2 *Let an ODIS* , $IS^{\succeq} = (U, AT \cup \{d\}, h, g)$, *for* $\forall a \in AT$, *the weight vector with features* $\omega = \left(\omega(a_1), \omega(a_2), \ldots, \omega(a_p)\right)^T$, *then*

(1) $\omega(a) \geq 0$;
(2) $\sum_{a_i \in AT} \omega(a_i) = |AT|$.

Proof (1)–(2) can be directly demonstrated by Definition 8.

From Property 2, we can get that the weights of conditional attributes are calculated by the partition coefficients between the conditions and decisions. The higher ω reflects the higher relevance between the condition attributes and the decision.

□

Definition 9 Let an ODIS , $IS^{\succeq} = (U, AT \cup \{d\}, h, g)$, $\omega = \left(\omega(a_1), \omega(a_2), \ldots, \omega(a_p)\right)^T$ is the weight vector of attribute, for attribute subset $A \subseteq AT$ and neighborhood threshold λ, the distance function between two objects is defined as

$$\begin{aligned} d_A(x, y) &= \sqrt{\sum_{a \in A} (\omega(a)(h(x, a) - h(y, a)))^2} \\ &= \sqrt{\sum_{a \in A} \omega^2(a)(h(x, a) - h(y, a))^2}. \end{aligned} \tag{9}$$

And the weighted dominance-based neighborhood relation and relative dominating and dominated sets are defined as

$$\mathcal{W}_A^{\succeq} = \{(x, y) \mid d_A(x, y) \le \lambda \ \wedge \ h(x, a) - \sigma \le h(y, a)\};$$
$$\mathcal{W}_A^{+} = \{y \in U \mid y \,\mathcal{W}_A^{\succeq}\, x\}, \quad \mathcal{W}_A^{-} = \{y \in U \mid x \,\mathcal{W}_A^{\succeq}\, y\}, \tag{10}$$

where $\omega(a)$ represents the weight of attribute a and it satisfies that $\omega(a) \ge 0$, $\sum_{a \in AT} |\omega(a)| = |AT|$. When $\omega(a) > 1$, the significance of attribute a is added in feature selection; otherwise, when $0 < \omega(a) < 1$, reduced. What's more, when $\omega(a) = 1$, the degree of importance of a remains still. Especially, if $\omega(a) = 0$, then this attribute can be pre-removed before selecting features. And we set a degree of tolerance σ for the dominance relation, since it is severe for an object to meet the need that all the conditional attributes' values must be larger than that of another object, especially when the distance requirement has been satisfied. And once setting $\sigma = 0$, it is a strictly dominance-based relation.

$\mathcal{W}_A^{\succeq}$ represents a weighted dominance-based neighborhood relation. When $\omega = 1$ and $\sigma = 0$, $\mathcal{W}_A^{\succeq}$ is a generalization of dominance-based neighborhood relation. That is to say, the dominance-based neighborhood relation is a special case of the weighted dominance-based neighborhood relation.

Let an ODIS , $IS^{\succeq} = (U, AT \cup \{d\}, h, g)$, $W_A^{\succeq}$ is a weighted dominance-based neighborhood relation, $\forall x \in U$, then we can get the reflexivity from Definition 9:

$$Reflexivity: (x, x) \in \mathcal{W}_A^{\succeq}. \tag{11}$$

Example 3.1 (Continuing from Example 2.1) In order to manifest the effectiveness of the weights, we review the example. The η and ω can be calculated through Eqs. 7 and 8, then $\eta = (-0.4634, 0.2867, 6.6948, -1.3231)$ and $\omega = (0.2114, 0.1308, 3.0542, 0.6036)$. Since $\omega(a_1)$, $\omega(a_2)$ and $\omega(a_4)$ are less than 1, the influence of a_1, a_2 and a_4 in decision making needs to be reduced. While a_3, corresponding to the weight more than 1, is paid more attention when making decision. And we still assume $\lambda = 0.1$ in this example. The weighted dominance neighborhood information granules generated by A_1 and A_2 are shown in Table 3.

Table 3 Information granules

U	$DN_{A_1}^{\succeq\lambda}(x)$	$DN_{A_2}^{\succeq\lambda}(x)$
x_1	$\{x_1, x_2, x_8\}$	$\{x_1\}$
x_2	$\{x_2, x_8\}$	$\{x_1, x_2, x_3\}$
x_3	$\{x_3, x_5, x_8, x_9\}$	$\{x_3\}$
x_4	$\{x_4, x_5, x_8, x_{10}\}$	$\{x_4, x_5\}$
x_5	$\{x_5, x_8, x_9\}$	$\{x_5\}$
x_6	$\{x_6, x_8, x_9\}$	$\{x_6\}$
x_7	$\{x_5, x_6, x_7, x_8, x_9\}$	$\{x_7, x_9\}$
x_8	$\{x_8\}$	$\{x_8, x_{10}\}$
x_9	$\{x_9\}$	$\{x_9\}$
x_{10}	$\{x_{10}\}$	$\{x_{10}\}$

From Table 3, we can get that $WPOS_{A_1}^{\lambda} = \{x_6, x_8, x_9, x_{10}\}$ and $WPOS_{A_2}^{\lambda} = U$. Hence, corresponding dependency degree are $\gamma_{A_1}^{\succeq\lambda}(d) = 0.4$ and $\gamma_{A_2}^{\succeq\lambda}(d) = 1$. Connected with Fig. 2, it is evident that the results obtained from the two perspectives are consistent. That is to say the ability of A_2 to approximate D in the dominance neighborhood rough set is better than that of A_1. Therefore,the weighted dominance neighborhood rough set model can make up the deficiency of the dominance neighborhood rough set model.

3.2 The weighted dominance-based neighborhood conditional entropy

In [41], Hu et al. proposed dominance-based condition entropy (DCE) to access the degree of ranking consistency of samples in an ODIS. It is obvious that DCE only obeys the dominance-based relation, which is not equivalent to the weighted dominance-based neighborhood relation. Since it ignores the influence of the noise and regards every attribute equally. In order to make up the above weaknesses, we propose the weighted dominance-based neighborhood conditional entropy (WDNCE) next.

Definition 10 Let an ODIS , $IS^{\succeq} = (U, AT \cup \{d\}, h, g)$, $\forall A \subseteq AT$, the WDNCE of A with respect to d is defined as

$$\mathcal{WE}_{d|A}^{\succeq}(U) = -\frac{1}{|U|} \sum_{i=1}^{n} log \frac{\left|\mathcal{W}_A^{+}(x_i) \cap D_d^{+}(x_i)\right|}{\left|\mathcal{W}_A^{+}(x_i)\right|}. \tag{12}$$

In Eq. 12 , $\frac{|\mathcal{W}_A^{+}(x_i) \cap D_d^{+}(x_i)|}{|\mathcal{W}_A^{+}(x_i)|}$ can be considered as a variable, which is the main element of WDNCE. And this variable is a reflection of the degree of ranking consistency of samples in accordance with the conditional feature subset A and the decision feature d. And it is evident that this variable is in inverse proportional to the value of WDNCE, which shows WDNCE is non-negative. What's more, chances that the rank produced by the reduced attribute subset is closer to the actual decision. That is to say, the smaller value of $\mathcal{WE}_{d|A}^{\succeq}(U)$, the more necessary of conditional attribute subset A, and vice versa.

In the following, we are going to give the definition of the attribute significance measure, which is conducive to acquire necessary and informative attributes when selecting features.

3.3 Attribute reduction

Definition 11 Let an ODIS , $IS^{\succeq} = (U, AT \cup \{d\}, h, g)$, $\forall A \subseteq AT$, $\forall a \in A$, the WDNCE-based inner significance measure of a for A is defined as

$$sim_{in}^{\succeq U}(a, A, d) = \mathcal{WE}_{d|A-\{a\}}^{\succeq}(U) - \mathcal{WE}_{d|A}^{\succeq}(U). \tag{13}$$

According to the explanation of WDNCE, a higher inner significance measure indicates the greater importance of the conditional attribute subset. We can select essential features from the sets and get the core feature set A, which is denoted as $Core_A = \{a \in A \mid sim_{in}^{\succeq}(a, A, d) > 0\}$.

Definition 12 Let an ODIS , $IS^{\succeq} = (U, AT \cup \{d\}, h, g)$, $\forall A \subseteq AT$, $\forall a \in (AT - A)$, the WDNCE-based outer significance measure of a for A is defined as

$$sim_{out}^{\succeq U}(a, A, d) = \mathcal{WE}_{d|A}^{\succeq}(U) - \mathcal{WE}_{d|A\cup\{a\}}^{\succeq}(U). \tag{14}$$

Similarly, the outer significance measure is also conducive to find out necessary features and delete redundant ones. Given an ODIS, $IS^{\succeq} = (U, AT \cup \{d\}, h, g)$ and $\forall a \in AT$, we have $a \in Core_A$ as long as $sim_{in}^{\succeq}(a, A, d) > 0$. And a can be viewed as an indispensable attribute. Then a reduct can be gained by gradually adding single attribute from the attributes not among the core and finding the subset with the highest outer significance measure.

Next, we are going to introduce the conditions of feature selection based on WDNCE.

Definition 13 Let an ODIS , $IS^{\succeq} = (U, AT \cup \{d\}, h, g)$, $\forall A \subseteq AT$, the selected attribute subset A can be a reduct of $IS^{\succeq}$ only if it meets the following requirements:

$$\begin{aligned} &(1)\ \mathcal{WE}_{d|A}^{\succeq}(U) = \mathcal{WE}_{d|AT}^{\succeq}(U); \\ &(2)\ \forall a \in A, \mathcal{WE}_{d|A-\{a\}}^{\succeq}(U) \neq \mathcal{WE}_{d|AT}^{\succeq}(U). \end{aligned} \tag{15}$$

The aim of attribute reduction is to figure out meaningful attributes without redundancy, and the two conditions are guarantees. The former ensures our selected feature subset owns the same discernibility as the initial one. The latter guarantees that every feature in the simplified subset is indispensable via getting rid of redundant elements. Hence, a reduct can be obtained when these two items are satisfied.

3.4 Feature selection in matrix form

When handling with high-dimensional data, it is intuitive and effective to transform the calculation to matrix. Therefore, there is a need to acquire WDNCE through relation matrices. Subsequently, we are going to give the conception of the operations on relation matrices.

Definition 14 Let an ODIS, $IS^{\succeq} = (U, AT \cup \{d\}, h, g)$, $\forall A \subseteq AT$, $\mathcal{W}_A^{\succeq}$ is a dominance relation under A, and its corresponding relation matrix with respect to A on U is defined as $\mathbb{M}_U^A = \left[m_{(i,j)}^A\right]_{n\times n}$, where

$$m_{(i,j)}^A = \begin{cases} 1, & y\,\mathcal{W}_A^{\succeq}\,x; \\ 0, & otherwise. \end{cases} \tag{16}$$

Property 3 *$\mathbb{M}_U^A = \left[m_{(i,j)}^A\right]_{n\times n}$ is a dominance relation matrix, it holds:*

(1) $m_{(i,j)}^A = 1$, *where* $i \in [1, n]$ *and* $i \in N^+$;
(2) $\sum_{j=1}^n m_{(i,j)}^A = \left|\mathcal{W}_A^+(x_i)\right|$ *and* $\sum_{i=1}^n m_{(i,j)}^A = \left|\mathcal{W}_A^-(x_i)\right|$, *where* $i, j \in [1, n]$ *and* $i, j \in N^+$.

Definition 15 Given $A_1, A_2 \subseteq AT \cup \{d\}$, $\mathbb{M}_U^{A_1}$ and $\mathbb{M}_U^{A_2}$ are relation matrices respectively under attribute subsets A_1 and A_2, operations $''\bigtriangledown''$ are defined as

$$\mathbb{M}_U^{A_1} \bigtriangledown \mathbb{M}_U^{A_2} = \left[m_{(i,j)}^{A_1} \times m_{(i,j)}^{A_2}\right]_{n\times n}. \tag{17}$$

From Eq. 17, we can know that $''\bigtriangledown''$ is used to produce a new relation matrix, and the new one considers both attribute subsets A_1 and A_2.

Definition 16 Given $A \subseteq AT \cup \{d\}$, $\mathbb{M}_U^A = \left[m_{(i,j)}^A\right]_{n\times n}$ be a relation matrix £and its corresponding diagonal matrix can be defined as $\widehat{\mathbb{M}} = \left[\widehat{m}_{(i,j)}^A\right]_{n\times n}$, where

$$\widehat{m}_{(i,j)}^{A\lambda} = \begin{cases} \sum_{k=1}^n m_{(i,k)}^A, & 1 \le i = j \le n; \\ 0, & 1 \le i, j \le n,\ i \neq j. \end{cases} \tag{18}$$

Additionally, the inverse matrix of $\widehat{\mathbb{M}}_U^A$ is represented as $\left(\widehat{\mathbb{M}}_U^A\right)^{-1} = \left[1/\widehat{m}_{(i,j)}^A\right]_{n\times n}$.

Definition 17 Let an ODIS , $IS^{\succeq} = (U, AT \cup \{d\}, h, g)$, $\forall A \subseteq AT$, grounded on two diagonal matrices, WDNCE of A to d is denoted as follows, where $\widehat{\mathbb{M}}_U^{A\cup\{d\}} = \widehat{\mathbb{M}}_U^A \bigtriangledown D_d^+$.

$$\mathcal{WE}_{d|A}^{\succeq} = -\frac{1}{|U|} log\left|\widehat{\mathbb{M}}_U^{A\cup\{d\}} \bigtriangledown \left(\widehat{\mathbb{M}}_U^A\right)^{-1}\right|. \tag{19}$$

Proof According to Definition 10. it can be found as follows. Through the proof below indicates that WDNCE generated by Eqs. 12 and 19 are the same.

$$\begin{aligned} &\mathcal{WE}_{d|A}^{\succeq}(U) = -\frac{1}{|U|}\sum_{i=1}^n log \prod_{i=j=1}^n \frac{\widehat{\mathbb{M}}_{(i,j)}^{A\cup\{d\}}}{\widehat{\mathbb{M}}_{(i,j)}^A} = \\ &-\frac{1}{|U|} log \frac{\prod_{i=j=1}^n \widehat{\mathbb{M}}_{(i,j)}^{A\cup\{d\}}}{\prod_{i=j=1}^n \widehat{\mathbb{M}}_{(i,j)}^A} = -\frac{1}{|U|} log \frac{\prod_{i=1}^n \left(\sum_{k=1}^n \mathcal{W}_{(i,k)}^{\succeq A\cup\{d\}}\right)}{\prod_{i=1}^n \left(\sum_{k=1}^n \widehat{\mathbb{M}}_{(i,k)}^A\right)} = \\ &-\frac{1}{|U|} log \frac{\prod_{i=1}^n \left|\mathcal{W}_{A\cup\{d\}}^+(x_i)\right|}{\prod_{i=1}^n \left|\mathcal{W}_A^+(x_i)\right|} = \\ &-\frac{1}{|U|} log \frac{\prod_{i=1}^n \left|\mathcal{W}_A^+(x_i) \cap D_d^+(x_i)\right|}{\prod_{i=1}^n \left|\mathcal{W}_A^+(x_i)\right|} = -\frac{1}{|U|}\sum_{i=1}^n log \frac{\left|\mathcal{W}_A^+(x_i) \cap D_d^+(x_i)\right|}{\left|\mathcal{W}_A^+(x_i)\right|} \end{aligned}$$

What's more, from Eq. 19 we can know that the center part of WDNCE is $\left|\widehat{\mathbb{M}}_U^{A\cup\{d\}} \bigtriangledown \left(\widehat{\mathbb{M}}_U^{A}\right)^{-1}\right|$, which directly shows the proportion of the diagonal matrices $\widehat{\mathbb{M}}_U^{A\cup\{d\}}$ to $\widehat{\mathbb{M}}_U^{A}$. And its practical meaning is in agreement with Eq. 12.

□

3.5 Heuristic feature selection mechanism

This subsection is going to introduce the non-incremental heuristic attribute reduction algorithm which is grounded on Definition 13. Concrete steps are presented in Algorithm 1.

Algorithm 1: WDNCE-HAR algorithm

Input: An ODIS, $IS^{\geq} = (U, AT \cup \{d\}, h, g)$, λ and δ
Output: Attribute subset C.
1 Initialize $C \longleftarrow \emptyset$;
2 Count the weight of each conditional attribute by Eq. 10;
3 Count WDNCE $\mathcal{WE}^{\geq}_{d|AT}$ through Eq. 19;
4 **for** $l = 1$ *to* $|AT|$ **do**
5   Count $sim^{\geq U}_{in}(a_l, AT, d)$ via Eq.13 ;
6   **if** $sim^{\geq U}_{in}(a_l, AT, d) > 0$ **then**
7     $C \longleftarrow C \cup \{a_l\}$
8   **end**
9 **end**
10 Let $A \longleftarrow C$;
11 **while** $\mathcal{WE}^{\geq}_{d|A}(U) \neq \mathcal{WE}^{\geq}_{d|AT}(U)$ **do**
12   **for** $m = 1$ *to* $|AT - A|$ **do**
13   Count $sim^{\geq U}_{out}(a_l, A, d)$ via Eq.14 ;
14   **end**
15   Choose $a_1 = argmax\left\{sim^{\geq U}_{out}(a_l, AT, d), a_1 \in (AT - A)\right\}$;
16   $A \longleftarrow A \cup \{a_1\}$;
17 **end**
18 **for** *each* $a \in A$ **do**
19   **if** $\mathcal{WE}^{\geq}_{d|A-\{a\}}(U) = \mathcal{WE}^{\geq}_{d|A}(U)$ **then**
20     $A \longleftarrow A - \{a\}$;
21   **end**
22 **end**
23 $C \longleftarrow A$;
24 Return C;

In algorithm WDNCE-HAR, Step 2 calculate the wights of the entire conditional feature sets, and the time complexity is $O(|U| \times |AT|)$. In Step 3, WDNCE based on the raw set AT is counted, corresponding to the time complexity of $O\left(|AT|^2|U|^2\right)$. Steps 4–10 apply inner significance measure to acquire indispensable features. Steps 11–17 add necessary features with the highest outer significance from the remains into the core feature set. In Steps 18–22, redundant attributes are deleted with the WDNCE unchanged. And all three segments' time complexity are $O\left(|AT|^2|U|^2\right)$.

4 Incremental approach for the weighted dominance-based neighborhood rough set with dynamic objects

In actual life, dynamic objects can be divided into two kinds: increasing samples and decreasing ones. Out of the similarity of the two situations, this study just focuses on the situation that samples are added into the origin dataset. Section 3 has already presented the way WDNCE selects necessary features. In spite of the use of matrix form, repeating calculation is time-wasting once objects vary, particularly faced with big data. On that ground, we introduce an incremental algorithm for attribute reduction.

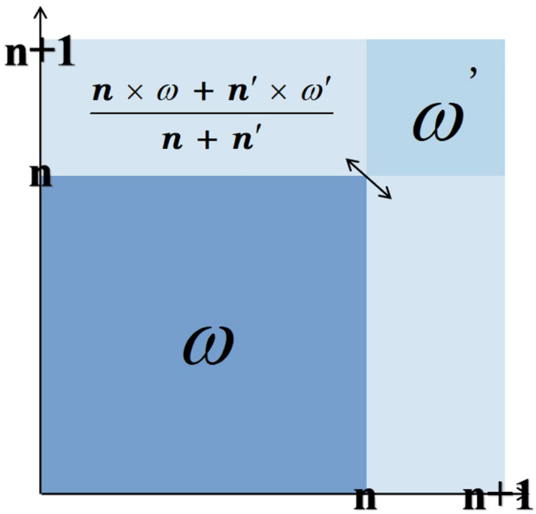

Fig. 3 Update principle of ω

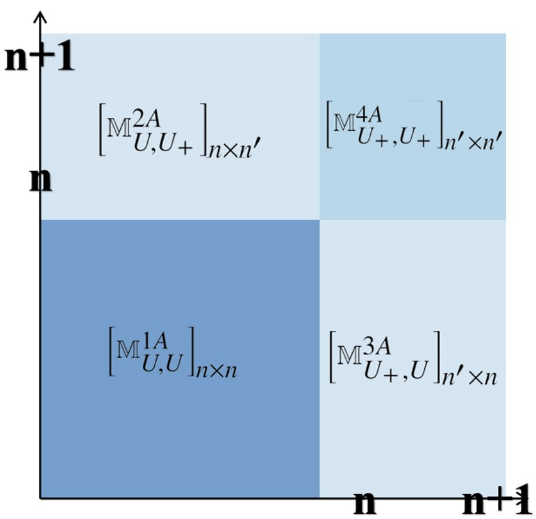

Fig. 4 Update principle of dominance relation matrix

4.1 Incremental mechanism of WDNCE when increasing objects

According to Eq. 12, it can be found that the update of WDNCE is the outcome of the variation of dominance relation matrix and its diagonal matrix. That is to say, before generating the new WDNCE, we need to introduce the changes of the weight, relation matrices and corresponding diagonal matrices and we start from the update of weight when there are objects added.

Definition 18 Let an ODIS, $IS^{\succeq} = (U, AT \cup \{d\}, h, g)$, where $U = \{x_1, x_2, \dots, x_n\}$, $\forall a \in AT$, we can get the weight from Eq. 11. Then new sample set come in, $U_+ = \{x_{n+1}, x_{n+2}, \dots, x_{n+n'}\}$. The update of ω is presented as

$$\widetilde{\omega} = \begin{cases} \omega, & 1 \le i,j \le n; \\ \omega', & n+1 \le i,j \le n+n'; \\ \frac{n\times\omega + n'\times\omega'}{n+n'} & otherwise, \end{cases} \tag{20}$$

where ω is the original weight generated under objects set U, and ω' is a new weight produced under objects set U_+. Moreover, when x_i and x_j are not in the above two situations, we treat the weights as the weighted average of ω and ω'. Figure 3 illustrates the update principle of ω. Next, we give the conception about the update of matrices.

Proposition 1 *Let an ODIS,* $IS^{\succeq} = (U, AT \cup \{d\}, h, g)$, *where* $U = \{x_1, x_2, \dots, x_n\}$, $\forall A \subseteq AT$, $\mathcal{W}^{\succeq}_A$ *is a dominance relation under A, the weighted dominance relation matrix with respect to A on U is defined as* $\mathbb{M}^A_U = \left[m^A_{(i,j)}\right]_{n\times n}$. *Then objects increase. New object set* $U_+ = \{x_{n+1}, x_{n+2}, \dots, x_{n+n'}\}$ *adds to U. Then the changed object set is* $\widetilde{U} = U \cup U_+$, *and the updated dominance relation matrix is presented as* $\mathbb{M}^A_{\widetilde{U}} = \left[\widetilde{m}^A_{(i,j)}\right]_{(n+n')\times(n+n')}$, *where*

$$\widetilde{m}^A_{(i,j)} = \begin{cases} m^A_{(i,j)}, & 1 \le i,j \le n; \\ 1, & x_j \mathcal{W}^{\succeq}_A x_i,\ (n+1 \le i \le n+n') \vee (n+1 \le j \le n+n'); \\ 0, & otherwise. \end{cases} \tag{21}$$

Proof Given that $U_+ = \{x_{n+1}, x_{n+2}, \dots, x_{n+n'}\}$ is joined to U, then $\widetilde{U} = U \cup U_+ = \{x_1, x_2, \dots, x_n, x_{n+1}, \dots, x_{n+n'}\}$. And we can divide the updated dominance relation into four parts, which is

$$\left[\begin{array}{c|c} \left[\mathbb{M}^{1A}_{U,U}\right]_{n\times n} & \left[\mathbb{M}^{2A}_{U,U_+}\right]_{n\times n'} \\ \hline \left[\mathbb{M}^{3A}_{U_+,U}\right]_{n'\times n} & \left[\mathbb{M}^{4A}_{U_+,U_+}\right]_{n'\times n'} \end{array}\right]$$

Relative presentation is in Fig. 4.

The first part of the matrix is on behalf of the dominance relation of $U \times U$ under A. The second section of the matrix represents the dominance relation of $U \times U_+$ under A. The third segment shows the the dominance relation of $U_+ \times U$ under A, while the last one displays the dominance relation of $U_+ \times U_+$ under A. When i and j are both between 1 and n, $\mathbb{M}^{1A}_{U,U} = \left[m^{1A}_{(i,j)}\right]_{n\times n}$, if $x_j \mathcal{W}^{\succeq}_A x_i$ satisfies, $m^{1A}_{(i,j)} = 1$, otherwise, 0. When i and j have at least one beyond n, $\widetilde{m}^A_{(i,j)} = 1$ only if $x_j \mathcal{W}^{\succeq}_A x_i$ holds. Similarly, $\widetilde{m}^{4A}_{(i,j)} = 1$ holds under the condition that $x_j \mathcal{W}^{\succeq}_A x_i$.

$\mathbb{M}^{1A}_{U,U}$ is the already existed part, and we do not need to recalculate it. Meanwhile, we can find that $\mathbb{M}^{2A}_{U,U_+}$, $\mathbb{M}^{3A}_{U_+,U}$, $\mathbb{M}^{4A}_{U_+,U_+}$ can be summarized a unified expression, and

$$\widetilde{m}^A_{(i,j)} = \begin{cases} 1, & x_j \mathcal{W}^{\succeq}_A x_i,\ (n+1 \le i \le n+n') \wedge (n+1 \le j \le n+n'); \\ 0, & otherwise. \end{cases}$$

Thus, the proposition can be proved.

Furthermore, we are going to introduce the way dominance diagonal matrix update.

□

Proposition 2 *Let an ODIS* , $IS^{\succeq} = (U, AT \cup \{d\}, h, g)$, *where* $U = \{x_1, x_2, \ldots, x_n\}$, $\forall A \subseteq AT$, *the existed diagonal matrix is* $\widehat{\mathbb{M}}_U^A = \left[\widehat{m}_{(i,j)}^A\right]_{n\times n}$. *After adding samples, the matrix is* $\widehat{\mathbb{M}}_{\widetilde{U}}^A = \left[\widehat{m}_{(i,j)}^A\right]_{(n+n')\times(n+n')}$, *where*

$$\widetilde{\widehat{m}}_{(i,j)}^A = \begin{cases} \widehat{m}_{(i,j)}^A + \sum_{k=n+1}^{n+n'} m_{(i,k)}^A, & 1 \le i = j \le n; \\ \sum_{k=1}^{n+n'} m_{(i,k)}^{\succeq A}, & n+1 \le i = j \le n+n'; \\ 0, & i,j \in \left[1, n+n'\right], i \ne j. \end{cases} \quad (22)$$

Proof In accordance with Definition 16, it can be known that if an element is not on the diagonal, then its corresponding value in dominance diagonal matrix is 0. That is to say, $\forall i,j \in \left[1, n+n'\right]$, if $i \ne j$, then $\widetilde{\widehat{m}}_{(i,j)}^A = 0$ always satisfies. And when $i = j$, there are two situations. First, for $\forall i = j \in [1, n]$, we have $\widetilde{\widehat{m}}_{(i,j)}^A = \sum_{k=1}^{n+n'} \widetilde{m}_{(i,k)}^A = \sum_{k=1}^{n} m_{(i,k)}^A + \sum_{k=n+1}^{n+n'} \widetilde{m}_{(i,k)}^A$. For $\forall i,k \in [1, n]$, corresponding weights are unchanged, that means the existed weighted dominance-based neighborhood relation for objects in U always holds. So there is $\widetilde{\widehat{m}}_{(i,j)}^A = \sum_{k=1}^{n} m_{(i,k)}^A + \sum_{k=n+1}^{n+n'} \widetilde{m}_{(i,k)}^A = \widehat{m}_{(i,j)}^A + \sum_{k=n+1}^{n+n'} \widetilde{\widehat{m}}_{(i,k)}^A$. Besides, for any $n+1 \le i = j \le n+n'$, $\widetilde{\widehat{m}}_{(i,j)}^A = \sum_{k=1}^{n+n'} \widetilde{m}_{(i,k)}^A$. Then we can get the the update principle of diagonal matrices.

□

4.2 Incremental algorithm for feature selection

In this subsection, we propose an incremental algorithm facing increasing samples based on WDNCE. In Algorithm 2, Step 1 increases dataset's objects. Step 2 is used to update the weights and it is worth mentioning that the origin weights are reserved for the initial data. And Step 2 also updates all kinds of initial diagonal matrices by Proposition 1. The time complexity of Step 2 is $|AT|\left|\widetilde{U}\right|\left|U_+\right|$. In the following, we use Step 3 to generate the new WDNCE via using Eq. 19. Steps 4–8 differentiates the relationship between WDNCE under the previous reduct and under the raw attribute set for the new dataset. If the value of the former is not higher than the latter, we consider the previous reduct as the reduct under new dataset. If not, Steps 9–14 depict the way the necessary attributes are added into the previous reduct and their relative steps are $|AT - C|\left|\widetilde{U}\right|^2$. After finding out the reduct, Steps 16–21 aims at getting rid of redundant features and get the final reduct in Steps 22 and 23. And corresponding time complexity of Steps 16–21 are $|C|^2\left|\widetilde{U}\right|^2$.

5 Experimental analysis

In this part, we evaluate the proposed method via twelve datasets downloaded from UCI, which are shown in Table 4. All experiments are carried out on a private computer and its configuration is below. Its CPU is i5-11300H with the memory of 16GB and the system of Windows 11. In addition, the algorithms are coded by Python.

Before starting numerical experiments, we need to process the data. First, we transform all the symbols of categorical features into numerical value based on their semantic meaning. Additionally, values of the whole dataset need to be normalized via $v_{ai} = \frac{v_{ai} - min(V_{ai})}{max(V_{ai}) - min(V_{ai})}$.

In order to demonstrate the advantages of the proposed method WDNCE-IAR in effectiveness and efficiency, the model is evaluated from two aspects, effectiveness and

Table 4 The description of data sets

No.	Data sets	Abbreviation	Samples	Attributes	Classes
1	Wine	Wine	178	13	3
2	Sonar,mines vs rocks	Sonar	208	60	2
3	Seeds	Seeds	210	7	3
4	Heart Failure Clinical Records	Heart	299	13	2
5	Leaf	Leaf	340	15	30
6	Ionosphere	Iono	351	34	2
7	Forest Fires	Forest	517	13	2
8	Wisconsin Diagnostic Breast Cancer	Wdbc	569	31	2
9	Breast Cancer	Breast	699	10	2
10	Maternal Health Risk	Health	1014	7	3
11	Cardiotocography	Card	2126	21	3
12	Statlog (Image Segmentation)	Image	2310	19	7

efficiency. For effectiveness, we focus on the reduct size and the precision. WDNCE-IAR is compared with five other algorithms, WDNCE-HAR, DCE-HAR [28], DCE-IAR [28], NCMI [23], HKCMI [42]. WDNCE-HAR is a heuristic feature selection mechanism based on the weighted dominance-based neighborhood conditional entropy shown in Algorithm 1. DCE-HAR and DCE-IAR apply dominance conditional entropy into attribute reduction, and the latter is a dynamic approach on the basis of the former. NCMI, as an interaction attribute reduction, is grounded on neighborhood conditional mutual information. HKCMI combines fuzzy complementary mutual information on clustering issues. In addition, two classifiers BayesNet and RandomTree are utilized to verify the classification performance of the reduct derived from those methods. We carry out experiments adopting 10-fold cross-validation. And for efficiency, speed-up ratio is our evaluation criteria.

Algorithm 2: WDNCE-IAR algorithm

Input: An original ODIS, $IS^{\succeq} = (U, AT \cup \{d\}, h, g)$ and its reduct C, two parameters λ and σ, four initial diagonal matrices $\widehat{\mathbb{M}}_U^{AT}$, $\widehat{\mathbb{M}}_U^{AT\cup\{d\}}$, $\widehat{\mathbb{M}}_U^{C}$, $\widehat{\mathbb{M}}_U^{C\cup\{d\}}$ and $U_+ = \{x_{n+1}, x_{n+2}, \ldots, x_{n+n'}\}$.

Output: A new reduct $R_{\widetilde{U}}$ on $\widetilde{U} = U \cup U_+$.

1 Add new samples to the set $\widetilde{U} \longleftarrow U \cup U_+$;

2 Update the weight via $\widetilde{w} \leftarrow w$ by Definition 18; Update the four diagonal matrices $\widehat{\mathbb{M}}_{\widetilde{U}}^{AT} \longleftarrow \widehat{\mathbb{M}}_U^{AT}$, $\widehat{\mathbb{M}}_{\widetilde{U}}^{AT\cup\{d\}} \longleftarrow \widehat{\mathbb{M}}_U^{AT\cup\{d\}}$, $\widehat{\mathbb{M}}_{\widetilde{U}}^{C} \longleftarrow \widehat{\mathbb{M}}_U^{C}$, $\widehat{\mathbb{M}}_{\widetilde{U}}^{C\cup\{d\}} \longleftarrow \widehat{\mathbb{M}}_U^{C\cup\{d\}}$ by Proposition 1 ;

3 Calculate WDNCE $\mathcal{WE}_{d|AT}^{\succeq}\left(\widetilde{U}\right)$ and $\mathcal{WE}_{d|C}^{\succeq}\left(\widetilde{U}\right)$ via using Eq. 19 ;

4 **if** $\mathcal{WE}_{d|C}^{\succeq}\left(\widetilde{U}\right) \leq \mathcal{WE}_{d|AT}^{\succeq}\left(\widetilde{U}\right)$ **then**

5   turn to Step 16 ;

6 **else**

7   turn to Step 10 ;

8 **end**

9 For each $a \in (AT - C)$, count $sim_{out}^{\widetilde{U}}(a, C, d)$ through Eq. 14. Next, based on outer significance measure, rank all these features by descending sort, and record the outcome as $\left\{\widetilde{a_1}, \widetilde{a_2}, \ldots, \widetilde{a_{|AT-C|}}\right\}$;

10 **while** $\mathcal{WE}_{d|C}^{\succeq}\left(\widetilde{U}\right) > \mathcal{WE}_{d|AT}^{\succeq}\left(\widetilde{U}\right)$ **do**

11   **for** $k = 1$ *to* $|AT - C|$ **do**

12   Choose $C \leftarrow C \cup \{\widetilde{a_k}\}$ and count $\mathcal{WE}_{d|C}^{\succeq}\left(\widetilde{U}\right)$;

13   **end**

14 **end**

15 **for** *each* $a \in C$ **do**

16   Count WDNCE $\mathcal{WE}_{d|(C-\{a\})}^{\succeq}\left(\widetilde{U}\right)$ through Eq. 19;

17   **if** $\mathcal{WE}_{d|(C-\{a\})}^{\succeq}\left(\widetilde{U}\right) \leq \mathcal{WE}_{d|C}^{\succeq}\left(\widetilde{U}\right)$ **then**

18         $C \longleftarrow C - \{a\}$

19   **end**

20 **end**

21 $R_{\widetilde{U}} \longleftarrow C$;

22 Return $R_{\widetilde{U}}$;

Table 5 The comparative classification results of different algorithms on BayesNet classifier (%)

Data set	RAW	*NCMI*	*HKCMI*	*DCE − HAR*	*DCE − IAR*	*WDNCE − HAR*	*WDNCE − IAR*
Wine	0.9624	0.9494(9)	0.9629(9)	0.9685(9)	0.9705(10)	0.9595(4)	**0.9891**(5)
Sonar	0.7085	0.7298(46)	0.6984(8)	0.6929(6)	0.6945(5)	0.7342(4)	**0.7868**(6)
Seeds	0.8974	0.9057(7)	0.9286(6)	0.9418(6)	0.9386(6)	0.9054(6)	**0.9442**(6)
Heart	0.7636	0.7866(8)	0.7911(9)	0.7445(11)	0.7211(11)	0.7993(6)	**0.8316**(6)
Leaf	0.6426	0.7015(7)	0.5882(5)	0.6153(9)	0.6084(8)	0.6367(10)	**0.7081**(12)
Iono	0.8896	**0.9269**(12)	0.8491(12)	0.7523(18)	0.7587(18)	0.8852(46)	0.9186(5)
Forest	0.9461	0.9527(8)	0.9288(8)	0.8997(8)	0.9022(9)	0.9541(6)	**0.9656**(7)
Wdbc	0.9265	0.9483(7)	0.9239(12)	0.9294(13)	0.9266(7)	0.9324(13)	**0.9502**(15)
Breast	0.9619	0.9526(4)	0.9649(8)	0.9667(8)	0.9643(9)	0.9648(4)	**0.9681**(4)
Health	0.6059	**0.6565**(6)	0.6203(5)	0.6109(6)	0.6109(6)	0.6254(6)	0.6354(6)
Card	0.7416	**0.8605**(5)	0.7602(10)	0.7545(10)	**0.7574**(10)	0.7462(11)	0.8164(10)
Iamge	0.6722	**0.8323**(4)	0.5335(7)	0.6817(10)	0.6923(10)	0.6852(14)	0.7148(14)
Average	0.8099	0.8502	0.7975	0.7965	0.7955	0.8191	**0.8524**

5.1 Effectiveness analysis

In this subsection, we are going to compare WDNCE-IAR with the other five mechanisms in terms of their classification precision. Randomly choosing 50% of the objects from one dataset, the remaining half samples are viewed as subsequent-added objects. When generating a new reduct, the rest is added to the origin. Respectively assessing the classification accuracy under the above algorithms, the results are presented below in Tables 5 and 6 where the bold parts refer to the highest accuracy. Table 5 is about classification outcome under BayesNet, likewise, Table 6, RandomTree. In the table, ”RAW” means the classification accuracy of the raw data. Extra, the number in bracket following the accuracy represents the feature number of the ultimate reduct.

As illustrated in Tables 5 and 6 no matter in what classifier, the classification precision of WDNCE-IAR is better than that of other methods for most of the datasets. Moreover, the proposed method select less features without significantly worsening classification performances. Conversely, sometimes there is a decent classification accuracy instead.

Considering WDNCE-IAR involves two parameter λ and σ, then we are going to further explore the classification performance under the combination of different parameters' values. Figure 5 shows the classification precision under BayesNet classifier for six datasets with the variation of both λ and σ. The abscissa represents the parameter σ and the ordinate is on behalf of the parameter λ. Classification results are displayed on z-axis. The values of λ are set from 0.1 to 0.2 with a step of 0.025. Meanwhile σ ranges from 0 to 0.2 in steps of 0.05.

Through Fig. 5, it is evident that classification accuracy is actually influenced by parameters λ and σ, which, to some extent, demonstrates the necessity of our research. For dataset Seed, Iono and Forest, their figures are fluctuating with the change of two parameters. However, there are some instances that the value of σ has an inapparent impact on

Table 6 The comparative classification results of different algorithms on RandomForest classifier (%)

Data set	RAW	*NCMI*	*HKCMI*	*DCE − HAR*	*DCE − IAR*	*WDNCE − HAR*	*WDNCE − IAR*
Wine	0.9718	0.9315(9)	0.9631(9)	0.9536(9)	**0.9833**(10)	0.9581(4)	0.9723(5)
Sonar	0.7959	0.7255(46)	0.7397(8)	0.7535(6)	0.7395(5)	0.7881(4)	**0.8023**(6)
Seeds	0.8955	0.9233(7)	0.9312(6)	0.9536(6)	0.9215(6)	0.9055(6)	**0.9331**(6)
Heart	0.8229	0.8214(8)	0.7978(9)	0.8083(11)	0.7423(11)	0.8185(6)	**0.8231**(6)
Leaf	0.6601	0.6559(7)	0.5817(5)	0.6348(9)	0.6514(8)	0.6629(10)	**0.6885**(12)
Iono	0.9246	0.9155(12)	0.9182(12)	0.9008(18)	0.9242(18)	0.9243(46)	**0.9329**(5)
Forest	0.9683	**0.9835**(8)	0.9534(8)	0.9638(8)	0.9609(9)	0.9766(6)	0.9764(7)
Wdbc	0.9502	0.9443(7)	0.9474(12)	0.9412(24)	0.9441(24)	0.9618(13)	**0.9618**(15)
Breast	0.9638	0.9597(4)	0.9659(8)	0.9677(8)	0.9677(9)	0.9672(4)	**0.96721**(4)
Health	0.7926	0.7935(6)	0.7974(5)	0.7971(6)	0.7994(6)	0.7887(6)	**0.8066**(6)
Card	0.9216	0.8951(5)	0.9232(10)	0.9195(10)	0.9186(10)	0.9227(11)	**0.9232**(10)
Iamge	0.8477	0.8281(4)	0.5383(7)	0.8219(10)	0.8326(10)	0.8527(14)	**0.8534**(14)
Average	0.8762	0.8646	0.8385	0.8679	0.8655	0.8773	**0.8871**

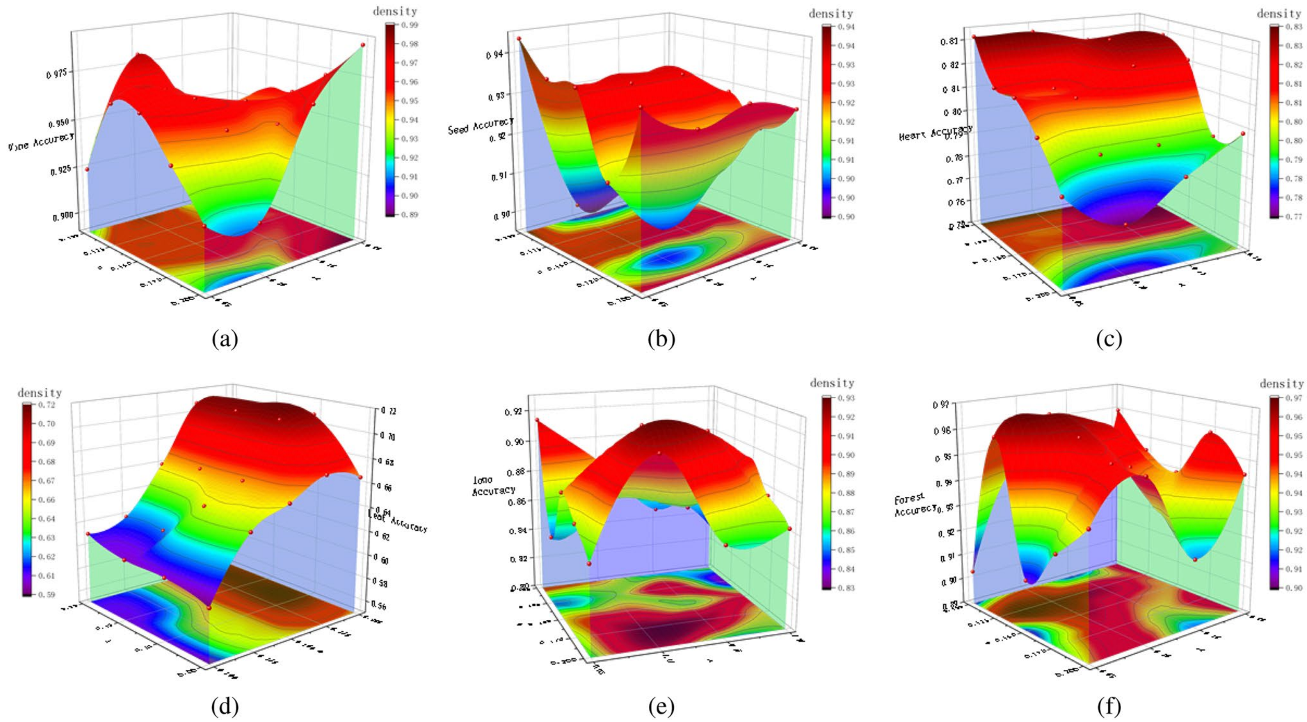

Fig. 5 Classification accuracy under BayesNet classifier for six datasets based on various parameters λ and σ

classification performances. For datasets Leaf, the influence of σ is slight, since when λ is fixed, the variation of σ poorly influences the classification precision. It is nearly the same with Dataset Wine and Dataset Heart except for the situation that σ is closer to 0.2 and λ is next to 0. And from Fig. 5, we can find combinations of parameters corresponds to superior performance are not fixed. For example, in the dataset Wine, the best accuracy is acquired with $\sigma = 0.2$, $\lambda = 0.2$. While the precision reaches the maximum in dataset Forest when σ and λ are set 0.125 and 0.1 respectively.

In order to further verify the effectiveness of WDNCE-IAR and WDNCE-HAR, Wilcoxon signed-ranked test is applied to evaluate whether the the proposed methods WDNCE-IAR and WDNCE-HAR statically outperform comparative approaches. Then corresponding P-values are generated in Table 7 where the bold parts are smaller than 0.1. When setting a significance level of 10 %, most of the P-values are smaller than 0.1, particularly for WDNCE-IAR, which indicates the classification accuracy of the two methods are statically higher than others. And that demonstrates the effectiveness of the two designed methods.

5.2 Efficiency comparison

This subsection aims at demonstrating the efficiency of WDNCE-IAR in accordance with speed-up ratio, which is calculated as $speed-up\ ratio = \frac{T_{HAR}}{T_{IAR}}$. We simulate the dynamic process in the following way. Firstly, randomly choose 50% samples as the initial dataset.Then, from the remaining set, objects wait to be added. We record the time, compute the ratio and compare it with the numerical value 1. If the result is greater than one, then static mechanism spends more time on selecting features, which further indicates the necessity of our study.

Table 7 P-values of the comparison results on two classifiers

	BayesNet		RandomForest	
	WDNCE-IAR	WDNCE-HAR	WDNCE-IAR	WDNCE-HAR
RAW	**0.0013**	**0.0114**	**0.0013**	0.3051
NCMI	0.2525	0.9369	**0.0021**	**0.0273**
HKCMI	**0.0013**	**0.0388**	**0.0019**	**0.0539**
DCE-HAR	**0.0013**	**0.0539**	**0.0067**	**0.0458**
DCE-IAR	**0.0013**	**0.0734**	**0.0043**	0.1120

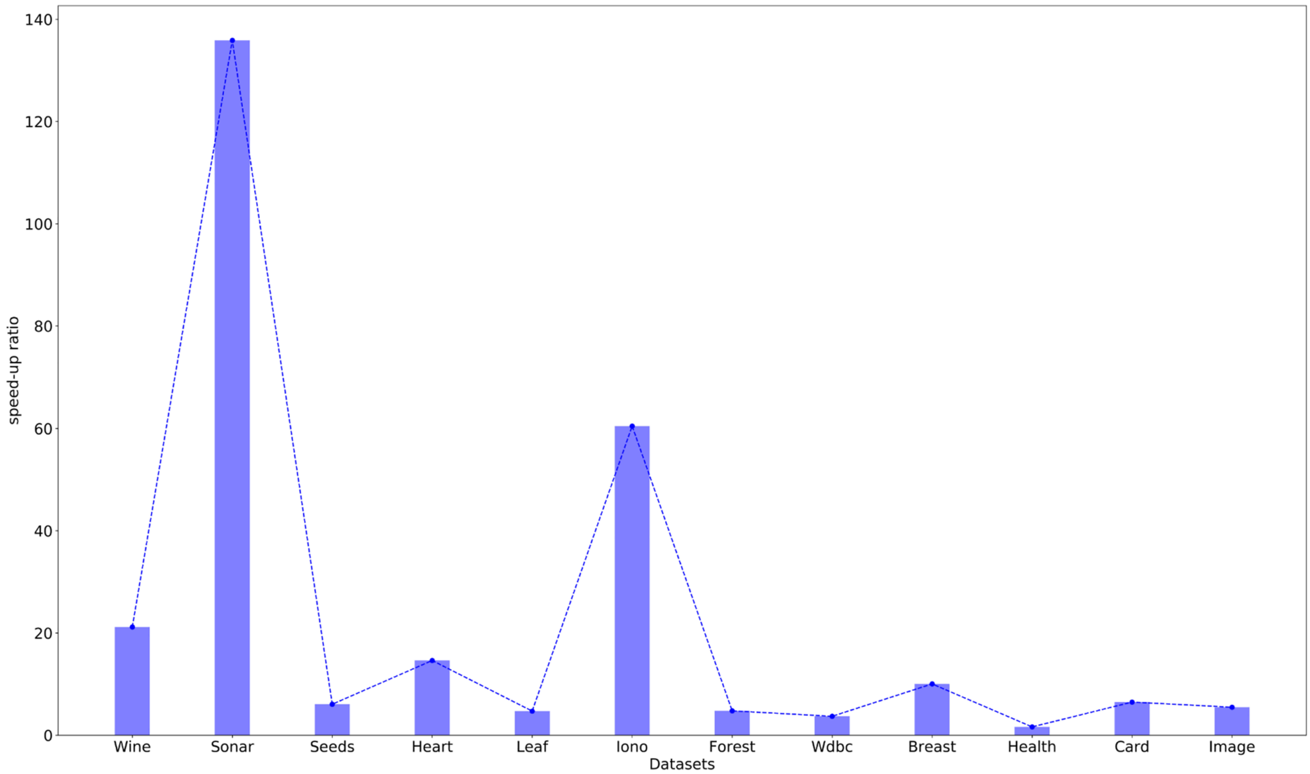

Fig. 6 The speed-up ratio of datasets

The results of numerical experiments are shown in Fig. 6, where the x-axis denotes the type of datasets and the y-axis shows the values of speed-up ratio. From the picture, almost all the ratios are beyond 4. For datasets Breast, Heart and Wine, WDNCE-IAR is at least ten times faster than WDNCE-HAR. And it is noteworthy that the speed-up ratio for datasets Iono and Sonar even exceed 60. That is to say WDNCE-IAR is efficient in attribute reduction and saves lots of time in making decisions.

6 Conclusion and future work

When applying neighborhood rough sets to feature selection, traditional approaches always directly assign the same weights to all conditional features. These methods ignore the inner relationship between conditional attributes and decisions. And if we can mine them in advance, we are able to highlight those features having high correlations with the decision. Hence, it is easier and more effective to figure out essential conditional features. Likewise, it is unavoidable that in the age of big data, the demand of feature selection has surged since oceans of data spring up everyday. Static methods are inappropriate to access updated massive data while incremental learning is conducive to efficiently cope with. Inspired by these two limitations, the weighted dominance-based neighborhood rough set(WDNRS) is came up with and a relative entropy is also introduced. Considering that matrix form is instrumental in reducing the complexity of calculation, a related incremental algorithm in matrix form is also proposed for dynamic ordered data with updating samples. Experimental results show that the accuracy rises through assigning different weights and incremental algorithm contributes to shorten the run time. In general, the proposed method is capable of effectively and efficiently selecting necessary and non-redundant features from update ordered data.

There are two respects for our future research. For one thing, in the information age, the variation of collected data is various. And in this study, we just keep an eye on the change of objects. Later, we intend to extend incremental feature-selection mechanisms which pay attention to the single variation of features or the simultaneous varying of both objects and attributes. For another thing, no matter what variable changes, features or samples, different datasets corresponds to different wights. However, it is a waste of time if we count them from scratch. And there is a need to ensure the precision of the generated weights. Hence we will investigate the updated weights in the future.

Acknowledgements This paper is supported by the National Natural Science Foundation of China (NO. 61976245).

References

1. Zhong J, Wang J, Peng W, Zhang Z, Li M (2015) A feature selection method for prediction essential protein. Tsinghua Sci Technol 20(5):491–499
2. Bang S, Kang J, Jhun M, Kim E (2017) Hierarchically penalized support vector machine with grouped variables. Int J Mach Learn Cyb 8(4):1211–1221
3. Abedini M, Kirley M (2013) An enhanced XCS rule discovery module using feature ranking. Int J Mach Learn Cyb 4(3):173–187
4. Gu S, Cheng R, Jin Y (2018) Feature selection for high-dimensional classification using a competitive swarm optimizer. Soft Comput 22(3):811–822
5. Pilyugina N, Tsukahara A, Tanaka K (2021) Comparing methods of feature extraction of brain activities for octave illusion classification using machine learning. Sensors 21(19):6407–6407
6. Ali W (2017) Phishing website detection based on supervised machine learning with wrapper features selection. Int J Adv Comput 8(9):72
7. Tuo Q, Zhao H, Hu Q (2018) Hierarchical feature selection with subtree based graph regularization. Knowl-Based Syst 163:996–1008
8. Roffo S, Melzi S, Castellani U, Vinciarelli A, Cristani M (2020) Infinite feature selection: a raph-based feature filtering approach. IEEE Trans Pattern Anal 43(12):4396–4410
9. Pawlak Z (1982) Rough sets. Int J Comput Inform Sci 11(5):341–365
10. Yao Y (1998) Relational interpretations of neighborhood operators and rough set approximation operators. Inform Sci 111(1–4):239–259
11. Ziarko W (1993) Variable precision rough set model. J Comput Syst Sci 46(1):39–59
12. Greco S, Matarazzo B, Slowinski R (1999) Rough approximation of a preference relation by dominance relations. Eur J Oper Res 117(1):63–83
13. Wang P, Wu Q, He J, Shang X (2018) Approximation operator based on neighborhood systems. Symmetry-basel 10(11):539–539
14. Hu Q, Yu D, Me Z (2008) Neighborhood classifiers. Expert Syst Appl 34(2):866–876
15. Sun L, Zhang X, Qian Y, Xu J, Zhang S (2019) Feature selection using neighborhood entropy-based uncertainty measures for gene expression data classification. Inform Sci 502:18–41
16. Sun L, Yin T, Ding W, Qian Y, Xu J (2022) Feature selection with missing labels using multilabel fuzzy neighborhood rough sets and maximum relevance minimum redundancy. IEEE Trans Fuzzy Syst 30(5):1197–1211
17. Sun L, Li M, Ding W, Zhang E, Mu X, Xu J (2022) AFNFS: Adaptive fuzzy neighborhood-based feature selection with adaptive synthetic over-sampling for imbalanced data. Inform Sci 612:724–744
18. Greco S, Matarazzo B, Slowinski R (2001) Rough sets theory for multicriteria decision analysis. Eur J Oper Res 129(1):1–47
19. Zhang X, Chen D, Tsang E (2016) Generalized dominance rough set models for the dominance intuitionistic fuzzy information systems. Inform Sci 378:1–25
20. Ali A, Ali MI, Rehman N (2019) Soft dominance based rough sets with applications in information systems. Int J Approx Reason 113:171–195
21. Chen H, Li T, Luo C, Hu J (2015) Dominance-based neighborhood rough sets and its attribute reduction. Rough sets and knowledge technology, lecture notes in computer science. Springer, Berlin, Heidelberg, pp 89–99
22. Chen H, Li T, Cai Y, Luo C, Fujita H (2016) Parallel attribute reduction in dominance-based neighborhood rough set. Inform Sci 373:351–368
23. Wan J, Chen H, Yuan Z, Li T, Yang X, Sang B (2017) A novel hybrid feature selection method considering feature interaction in neighborhood rough set. Knowl-Based Syst 227:107167
24. Wang S, Li X, Xia J, Xia J, Zhang X (2010) Weighted neighborhood classifier for the classification of imbalanced tumor dataset. J Circuit Syst Copm 19(1):259–273
25. Tsang E, Hu Q, Chen D (2016) Feature and instance reduction for PNN classifiers based on fuzzy rough sets. Int J Mach Learn Cyb 7(1):1–11
26. Hu M, Tsang E, Guo Y, Chen D, Xu W (2021) A novel approach to attribute reduction based on weighted neighborhood rough sets. Knowl-Based Syst 220:106908
27. Liang J, Wang F, Dang C, Qian Y (2014) A group incremental approach to feature selection applying rough set technique. IEEE Trans Knowl Data En 26(2):294–308
28. Sang B, Chen H, Yang L, Zhou D, Li T, Xu W (2021) Incremental attribute reduction approaches for ordered data with time-evolving objects. Knowl-Based Syst 212:106583
29. Sang B, Chen H, Yang L, Li T, Xu W (2022) Incremental feature selection using a conditional entropy based on fuzzy dominance neighborhood rough sets. IEEE Trans Fuzzy Syst 30(6):1683–1697
30. Yuan K, Xu W, Li W, Ding W (2021) An incremental learning mechanism for object classification based on progressive fuzzy three-way concept. Inform Sci 584:127–147
31. Jing Y, Li T, Huang J, Zhang Y (2016) An incremental attribute reduction approach based on knowledge granularity under the attribute generalization. Int J Approx Reason 76:80–95
32. Chen D, Dong L, Mi J (2019) Incremental mechanism of attribute reduction based on discernible relations for dynamically increasing attribute. Soft Comput 24(1):321–332
33. Dong L, Chen D (2020) Incremental attribute reduction with rough set for dynamic datasets with simultaneously increasing samples and attributes. Int J Mach Learn Cyb 11(6):1339–1355
34. Zhang X, Chen X, Xu W, Ding W (2022) Dynamic information fusion in multi-source incomplete interval-valued information system with variation of information sources and attributes. Inform Sci 608:1–27
35. Xu W, Yuan K, Li W (2022) Dynamic updating approximations of local generalized multi-granulation neighborhood rough set. Appl Intell 52(8):9148–9173
36. Huang Y, Li T, Luo C, Fujita H, Horng S (2017) Matrix-based dynamic updating rough fuzzy approximations for data mining. Knowl-Based Syst 119:273–283
37. Wang S, Li T, Luo C, Fujita H (2016) Efficient updating rough approximations with multi-dimensional variation of ordered data. Inform Sci 372:690–708
38. Wang S, Li T, Luo C, Chen H, Fujita H (2019) Domain-wise approaches for updating approximations with multi-dimensional variation of ordered information systems. Inform Sci 478:100–124
39. Shannon C, Weaver W (1948) The mathematical theory of communication. Bell Syst Tech J 27(3/4):373–423
40. Hu Q, Yu D, Xie Z, Liu J (2006) Fuzzy probabilistic approximation spaces and their information measures. IEEE Trans Fuzzy Syst 14(2):191–201
41. Hu Q, Che X, Zhang L, Zhang D, Guo M, Yu D (2012) Rank entropy-based decision trees for monotonic classification. IEEE Trans Knowl Data En 24(11):2052–2064

42. Yuan Z, Chen H, Yang X, Li T, Liu K (2021) Fuzzy complementary entropy using hybrid-kernel function and its unsupervised attribute reduction. Knowl-Based Syst 231:107398