Applied Intelligence (2023) 53:9858–9877
https://doi.org/10.1007/s10489-022-03987-2

Feature selection using a weighted method in interval-valued decision information systems

Xiaoyan Zhang[1] · Zongying Jiang[1] · Weihua Xu[1]

Accepted: 10 July 2022 / Published online: 12 August 2022

Abstract

Recent developments in big data applications have heightened the need for understanding and processing high-dimensional data. It is necessary to extract some excellent features that effect the learning performance in high-dimensional data. Feature selection algorithm based on rough set theory as an important preprocessing method has been widely used in practical applications. Meanwhile, it should be noted that different attributes have different effects on model evaluation. Nevertheless, each feature or attribute has the same degree of importance in the interval-valued information system by using rough set models, ignoring the imbalance between features. Moreover, the monotonic classification effect of interval-valued data is easily affected by noise. For these two issues, we introduce different weights into neighborhood relations and propose a novel approach for feature selection-based weighted neighborhood rough sets for interval-valued information systems in this study. First, weighted neighborhood relations and some important properties are proposed by considering different attribute weights in the interval-valued information system. Then, we construct an interval-valued-based weighted neighborhood rough set (IVWNRS) model to solve the contradiction between the degree of dependency and the classification ability of the attribute subset. Furthermore, a heuristic algorithm is designed according to the degree of dependency to select an attribute subset that has both strong correlation and high dependency. Finally, we compare it with six other representative feature selection algorithms on fifteen public datasets to evaluate the performance of the proposed algorithm. Experimental results on different classifiers show that the IVWNRS algorithm has higher classification performance and is significantly effective.

Keywords Degree of dependency · Feature selection · Interval-valued · Information systems · Weighted neighborhood rough set

1 Introduction

Rough set theory, proposed by Pawlak [26] in 1982, has been widely used in machine learning, knowledge discovery, and approximate reasoning. Limited to strict equivalence relations, the classical Pawlak rough set can only deal with information systems with categorical attributes. Faced with complex surroundings, different types of data become more common. Naturally, intervals appear to be a method for describing uncertainty, such as temperature changes and blood pressure, and have received intense attention. To mine knowledge under the circumstance of uncertainty, such as interval values in information systems, some scholars have introduced neighborhood relations and fuzzy relations into Pawlak rough sets, which have formed interval-valued-based neighborhood rough sets [5, 21, 29, 46] and interval-valued-based fuzzy rough sets [14, 31], respectively. Similar to the Pawlak rough set, the aforementioned generalized rough set models are applied to feature selection [2, 11, 35, 39, 48, 49], rule extraction [1, 27, 28], dynamic learning [11, 24] and other fields.

Studies focus on fuzzy rough sets and neighborhood rough sets in interval-valued datasets. Different from interval-valued-based neighborhood rough sets, interval-valued-based fuzzy rough sets can quantify information. However, full use is rarely made of primitive information because the imbalance of data leads to partial information being employed rather than the complete information being employed. Even inappropriate binary relations are further introduced into the fuzzy decision information system. Questions such as these have hindered more logical and objective research. Contrary to fuzzy rough sets,

✉ Xiaoyan Zhang
zxy19790915@163.com

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

similarity relations in neighborhood rough sets measure any two objects without considering the effect of a third object, which has more extensive applications in real life. Therefore, the similarity relations in interval-valued based neighborhood rough sets are considered.

As an imperative part of rough set theory, feature selection always causes heated discussion. Generally, feature selection has the following framework: for a set of all attributes A, and B ⊆ A, B can be regarded as one selected feature subset if the constraints are satisfied

$(1) B\ satisfies\ the\ \rho - constraint;$

$(2) \forall\ C \subset B\ does\ not\ satisfy\ the\ \rho - constraint.$

Different feature selection evaluation approaches are determined by different $\rho - constraints$. Based on such basic construction, different scholars measure the feature subset with different methods [37, 38, 41, 45]. To make feature selection more suitable for neighborhood rough sets, issues such as the choice of parameters [51] controlling the size of information granules, running time [18, 32], incremental learning [30] and other fields have been explored. Unfortunately, the effectiveness of these approaches is weakened in interval-valued neighborhood rough sets. Through our further study [3, 7, 12, 13, 16, 23, 34], such situations can be ascribed to their overlooking attribute importance. Without assigning weights to attributes, attributes highly correlated with decisions have the same probability of being selected as common attributes. Such problems may not be evident in the training stage, but it generally shows poor generalization ability during the testing period. Moreover, in terms of time complexity, appropriate attribute weights can save running time of searching one feature subset [7]. Consequently, weighting attributes requires consideration.

After analyzing the process of obtaining one feature subset in interval-valued-based data from the perspective of binary relations and the attribute weights, we define a new model, namely, the interval-valued-based weighted neighborhood rough set (IVWNRS), to search one valuable feature subset by taking advantage of all data information. To attain one feature subset, we first calculate the correlation between attributes and decisions, and the significance of attributes highly relevant to decisions are improved; then, one minimal feature subset can be selected based on similarity relations by comprehensively using the data information. The flow of the methodology used is displayed in Fig. 1.

Inspired by above analysis, we try to design a feature selection algorithm for high-dimensional interval-valued data to obtain important features to improve the performance of learning model. Compared with the existing methods, the IVWNRS model could solve the imbalance in different features and the contradiction between the degree of dependency and the classification ability of feature subset. Thus, the proposed approach based on IVWNRS can comprehensively evaluate the features and improve the object classification performance in high-dimensional interval-valued data. The main contributions of our research can be summarized as follows.

- We proposed a feature selection algorithm for high-dimensional interval-valued based on interval-valued-based weighted neighborhood rough set model, which could depict the importance of features in more detail. Compared with some existing methods, this approach could select excellent features that have better object classification performance.
- The defined IVWNRS model is based on weighted neighborhood relation, which provides a way for dealing with interval-valued data. Meanwhile, the IVWNRS model focuses on solving the imbalance issue in different features and the contradiction between the degree of dependency and the classification ability, so as to improve its ability to describe uncertainty and better evaluation selection features.
- Moreover, a heuristic algorithm is designed according to the degree of dependency to select an attribute subset that has both strong correlation and high dependency according to the degree of dependency. All the experimental results demonstrate that our method has better classification performance than some state-of-the-art feature selection methods in interval-valued data.

The rest of this paper is organized as follows. In Section 2, related works are presented in more detail. In Section 3, we briefly review the basic concept of neighborhood rough sets and interval-valued-based decision information systems. To process interval-valued data, we propose a new model named interval-valued-based neighborhood rough sets without weighted attributes and find that the dependency of attribute subsets contradicts its performance. In Section 4, an interval-valued-based weighted neighborhood rough set is further presented to solve the defect, and a measure of evaluating attributes is proposed. In Section 5, we design a heuristic algorithm to find a reduct in an interval-value-based decision information system. In Section 6, we use fifteen datasets to compare the proposed model with the advanced feature selection methods from two different aspects. In Section 7, we summarize the paper and propose a vision for the future.

2 Related works

In the introduction, the information of different studies were omitted, they are described in detail in this section.

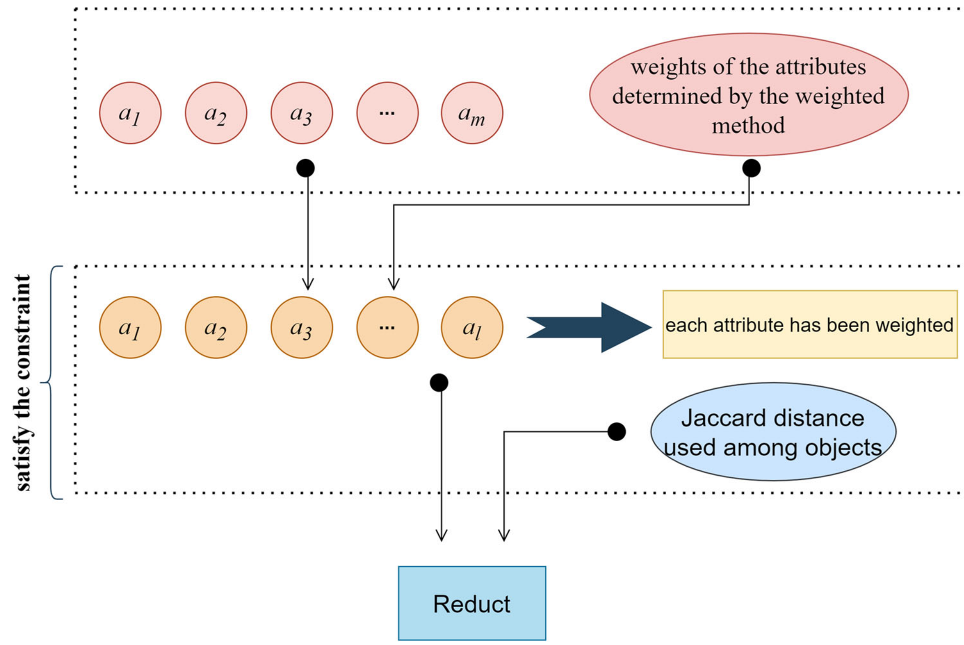

Fig. 1 The process of obtaining one feature subset

Based on neighborhood similarity relations, neighborhood rough sets can handle information systems with interval-valued attributes, and several researchers have done significant work. Sang et al. [29] studied incremental feature selection approaches based on a fuzzy dominance neighborhood rough set (FDNRS) for dynamic interval-valued ordered data. Yang et al. [46] introduced a novel fuzzy neighborhood rough set model and found that its matrix representation was suitable for quantifying the similarity relations in interval-valued information systems with fuzzy decisions (IvIS_FD). Considering that the existing rough set models neglect the test cost and misclassification cost in the data, Liu et al. [21] proposed a feature selection approach based on the cost sensitivity for interval-valued data after designing the cost-sensitive function. Inspired by the importance of the positive region, Chen et al. [5] used the similarity measure between interval numbers to establish the neighborhood and then constructed a rough set model based on the neighborhood. For information fusion, Huang et al. [14] used a novel data fusion method based on fuzzy information granulation, which translated multi-source interval-valued data into trapezoidal fuzzy granules. The intuitive comparison can be found in Table 1. Unfortunately, although the above neighborhood-based rough set models deal with interval values from different aspects, they all ignore the importance of attribute weights. However, in daily routines, the significance of different attributes is usually different. In this case, we have to distinguish one attribute from the left, that is, assigning different weights to different attributes. Furthermore, if we consider internal relevance between condition attributes and decisions ahead of time, we can place more emphasis on the attributes that are highly related to decisions, which paves the way for selecting attribute subsets with relatively high correlation and dependency.

The weight of an attribute stresses its importance. Since the weight of an attribute plays an important role in feature selection, some scholars have studied how to assign attribute weights in Pawlak rough sets, fuzzy rough sets, and decision-theoretic rough sets. To handle multiindividual participation and mutual compensation among risk factors, Luo et al. [23] proposed a methodology based on weighted multigranulation fuzzy rough sets (MGFRSs) over two universes to perform risk evaluation for PPP WTE incineration plant projects. Vluymans et al. [34] presented a strategy for selecting a suitable weighting scheme for ordered weighted average-based fuzzy rough sets in general. Facing the problem of high time consumption on large-scale datasets in classical reduction algorithms in fuzzy rough sets, Fan et al. [7] introduced weights into the concept of feature selection and built an optimization problem to find weights. In addition to the difference in the importance of different granulations in multisource systems, Guo et al. [12] provided a weighted generalized multigranulation interval-valued decision-theoretic rough set model (WGM-IVDTRS) for multisource decision fusion. Other methods for determining attribute weights can be found in [3, 13, 16].

Feature selection aims to select one minimal feature subset with important information after eliminating redundant and inconsistent attributes. Based on Pawlak rough

Table 1 Summary of interval-valued-based generalized rough sets

	Fuzzy rough sets	Neighborhood rough sets
Interval-valued attributes	Huang et al. [14] Sun et al. [31]	Chen et al. [5] Liu et al. [21] Sang et al. [29] Yang et al. [46]

sets, state-of-the-art feature selection methods have been expanded to several fields, including fuzzy rough sets, neighborhood rough sets and decision-theoretic rough sets. Different from classical Pawlak rough sets, fuzzy rough sets can handle continuous attributes, and neighborhood rough sets can be used to deal with both continuous and categorical attributes. Thus, neighborhood rough sets have intrigued many authors for their wide application range. To date, there have been many feature selection approaches based on neighborhood knowledge. Wang et al. [38] proposed decision self-information for feature selection considering the uncertainty information in lower and upper approximations. Without any domain knowledge or specifying any parameters in advance, Zhou et al. [51] defined a new neighborhood rough set relation with adapted neighbors and proposed a new online streaming feature selection method based on this relation. The neighborhood discrimination index, characterizing the distinguishing information of a neighborhood relation, was defined by Wang et al. [37] to select one feature subset. To better describe the neighborhoods of category-mixed samples, Wang et al. [41] proposed a new model combining the advantages of the δ-neighborhood and k-nearest-neighbor for feature selection. Adapted for mutual information on continuous features and multiple labels, Gonzalez-Lopez et al. [10] proposed a distributed model. Since it is difficult to determine the neighborhood radius, Yang et al. [45] presented a pseudo label neighborhood relation to reduce the uncertainties in feature selection, where samples can be differentiated by not only distance but also the sample pseudo labels. In addition to the neighborhood radius defect, time complexity remains an issue, so Jiang et al. [18] designed an accelerator to speed up the process of obtaining features with better discriminating performance in supervised neighborhoods. Similarly, Sun et al. [32] proposed a novel feature selection method using the Fisher score and multilabel neighborhood rough sets (MNRS) in multilabel neighborhood decision systems to decrease running time while determining an appropriate neighborhood radius. In incremental learning, Sang et al. [30] introduced conditional entropy into neighborhood rough sets and proposed an incremental feature selection approach for dynamic ordered data based on this model. Obviously, far more feature selection methods exist than the abovementioned model; other feature selection methods can be found in [3, 4, 20, 22, 25, 36, 40, 42, 44]. For convenience, the studies based on different feature selection methods are displayed in Table 2.

3 Preliminaries

3.1 Classical neighborhood rough set theory

A five-tuple $IS = (U, AT \cup DT, F, G)$ is called a decision information system, where the universe $U = \{x_1, x_2, \cdots, x_n\}$ is a nonempty finite set of objects, $AT = \{a_1, a_2, \cdots, a_m\}$ is a nonempty finite set of conditional attributes, $DT = \{d_1, d_2, \cdots, d_r\}$ is a nonempty finite set of decision attributes, F is the relation from U to V_a satisfying $F = \{f : U \to V_a\}$, and G is the relation from U to V_d satisfying $G = \{f : U \to V_d\}$. What needs to be noted is that V_a and V_d denote the real domains of conditional attribute a and decision attribute d, respectively.

Given a decision information system $IS = (U, AT \cup DT, F, G)$, $U/DT = \{D_1, D_2, \cdots, D_s\}$ makes up a partition on U to DT.

In a given decision information system $IS = (U, AT \cup DT, F, G)$, for any $x \in U$, $B \subseteq AT$, the neighborhood similarity class of x under the conditional attribute subset B is defined as

$$NS_B^\delta(x) = \{y | d_B(x, y) \le \delta\}. \tag{1}$$

where d_B denotes the distance function under subset B and the neighborhood threshold δ ($\delta > 0$). The function $Dis : U \times U \to R^+$ can serve as a distance function if the function Dis satisfies the following properties:

(1) for any $x, y \in U$, $Dis(x, y) \ge 0$, $Dis(x, y) = 0$ if and only if $x = y$;
(2) for any $x, y \in U$, $Dis(x, y) = Dis(y, x)$;
(3) for any $x, y, z \in U$, $Dis(x, z) \le Dis(x, y) + Dis(y, z)$.

In a given decision information system $IS = (U, AT \cup DT, F, G)$, δ is set to a fixed value; for any $X \subseteq U$, $B \subseteq AT$, the lower and upper approximations of X with respect to B are, respectively defined as

$$\begin{aligned} \underline{RNS}_B^\delta(X) &= \{x \in U | NS_B^\delta(x) \subseteq X\}, \\ \overline{RNS}_B^\delta(X) &= \{x \in U | NS_B^\delta(x) \cap X \ne \varnothing\}. \end{aligned} \tag{2}$$

Analogously, according to formula (2), the lower and upper approximations of DT with respect to B are defined as

$$\begin{aligned} \underline{RNS}_B^\delta(DT) &= \cup_{i=1}^s \underline{RNS}_B^\delta(D_i), \\ \overline{RNS}_B^\delta(DT) &= \cup_{i=1}^s \overline{RNS}_B^\delta(D_i). \end{aligned} \tag{3}$$

Table 2 Summary of feature selection in generalized rough sets

	Fuzzy rough sets	Neighborhood rough sets	Decision-theoretic rough sets
Without weighted Method	Wan et al. [35] Wang et al. [39] Zhang et al. [48] Zhang et al. [49]	Chen et al. [2] Chen et al.[3] Chen et al. [4] Fu et al. [9] Jiang et al. [18] Liang et al. [20] Liu et al. [22] Mariello et al. [25] Sang et al. [30] Sun et al. [32] Wang et al. [36] Wang et al. [37] Wang et al. [38] Wang et al. [40] Wang et al. [41] Yang et al. [44] Yang et al. [45] Zhou et al. [51]	Guo et al. [11]
With weighted Method	Fan et al. [7] Hashemzadeh et al. [13] Luo et al. [23] Vluymans et al. [34]	Chen et al. [3] Huang et al. [16]	Guo et al. [12]

where $U/DT = \{D_1, D_2, \cdots, D_s\}$. The boundary and positive regions of DT with respect to B are defined as

$$\begin{aligned} BND_B^{\delta}(DT) &= \overline{RNS}_B^{\delta}(DT) - \underline{RNS}_B^{\delta}(DT), \\ POS_B^{\delta}(DT) &= \underline{RNS}_B^{\delta}(DT). \end{aligned} \tag{4}$$

Furthermore, the dependency degree of DT with respect to B in the decision information system $IS = (U, AT \cup DT, F, G)$ is described as

$$\gamma_B^{\delta}(DT) = \frac{|POS_B^{\delta}(DT)|}{|U|}. \tag{5}$$

where $|\cdot|$ represents the cardinality of set s. $\gamma_B^{\delta}(DT)$ is used to measure the ability of conditional attribute subset B to approximate DT. The larger the value of $\gamma_B^{\delta}(DT)$ has, the stronger the approximation ability of the attribute subset B is.

3.2 Interval-valued-based decision information system and neighborhood rough set

A five-tuple $IVIS = (U, AT \cup DT, F, G)$ is called an interval-valued decision information system when it is a special case of a decision information system with interval-valued attributes. Particularly, F is the relation from U to V_a satisfying $F = \{f : U \rightarrow V_a\}$, where V_a is interval number. That is, for any $a \in AT, x \in U$, the value of the object x under the conditional attribute a is denoted as $f(x, a) = [x_a^-, x_a^+](x_a^-, x_a^+ \in R \text{ and } x_a^- \leq x_a^+)$.

Considering that Euclidean distance in classical neighborhood rough sets cannot deal with interval numbers, Jaccard distance can be another choice. Given the two sets A and B, the difference between them can be described as

$$d_{J(A,B)} = 1 - \frac{|A \cap B|}{|A \cup B|}. \tag{6}$$

Now that the Jaccard distance can measure the difference between the two sets, it can naturally generalize to interval values. Assuming that $a = [a^-, a^+]$ and $b = [b^-, b^+]$ are two intervals, the Jaccard distances between them are defined as

$$d_{J(a,b)} = 1 - \frac{|a \cap b|}{|a \cup b|}. \tag{7}$$

where $|\cdot|$ represents the length of the interval, namely $|a \cap b|$ represents the length of the intersection of the interval a and b, and $|a \cup b|$ represents the length of the union of the interval a and b.

In a given interval-valued decision information system $IVIS = (U, AT \cup DT, F, G)$, for any $x \in U, B \subseteq AT$, the interval-valued neighborhood similarity class of x under the conditional attribute subset B is defined as

$$IVS_B^{\delta}(x) = \{y | d_J^B(x, y) \leq \delta\}. \tag{8}$$

Here, $d_J^B(x, y)$ is the Jaccard distance between objects x and y with respect to B in an interval-valued environment.

In a given interval-valued decision information system $IS = (U, AT \cup DT, F, G)$, δ is set to a fixed value; for any $X \subseteq U, B \subseteq AT$, the lower and upper approximations of X with respect to B are, respectively defined as

$$\begin{aligned} \underline{IVR}_B^{\delta}(X) &= \{x \in U | IVS_B^{\delta}(x) \subseteq X\}, \\ \overline{IVR}_B^{\delta}(X) &= \{x \in U | IVS_B^{\delta}(x) \cap X \neq \varnothing\}. \end{aligned} \tag{9}$$

Analogously, according to formula (9), the lower and upper approximations of DT with respect to B are, respectively defined as

$$\begin{aligned} \underline{IVR}_B^{\delta}(DT) &= \cup_{i=1}^{s} \underline{IVR}_B^{\delta}(D_i), \\ \overline{IVR}_B^{\delta}(DT) &= \cup_{i=1}^{s} \overline{IVR}_B^{\delta}(D_i). \end{aligned} \tag{10}$$

where $U/DT = \{D_1, D_2, \cdots, D_s\}$. The boundary and positive regions of DT with respect to B are defined as

$$\begin{aligned} BND_B^{\delta}(DT) &= \overline{IVR}_B^{\delta}(DT) - \underline{IVR}_B^{\delta}(DT), \\ POS_B^{\delta}(DT) &= \underline{IVR}_B^{\delta}(DT). \end{aligned} \tag{11}$$

Furthermore, the dependency degree of DT with respect to B in the interval-valued decision information system $IS = (U, AT \cup DT, F, G)$ is described as

$$\gamma_B^{\delta}(DT) = \frac{|POS_B^{\delta}(DT)|}{|U|}. \tag{12}$$

Next, we use an example to illustrate the proposed interval-valued-based neighborhood rough set theory.

Example 1 A given interval-valued decision information system $IS = (U, AT \cup DT, F, G)$ can be seen in Table 3, where the universe is $U = \{x_1, x_2, \cdots, x_{14}\}$, and the conditional and decision attribute sets are $AT = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7\}$ and $DT = \{d\}$. A partition on U to d is $\{\{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}, \{x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}\}\}$.

Given two conditional attribute subsets $B_1 = \{a_1, a_3\}$, $B_2 = \{a_1, a_5\}$ and a neighborhood threshold $\delta = 0.2$, the generated neighborhood similarity classes induced by B_1 and B_2 are shown in Table 4. According to the definition of dependency degree and the result in Table 4, we calculate that $\gamma_{B_1}^{\delta}(DT) = 1$ and $\gamma_{B_2}^{\delta}(DT) = 1$. Based on the above theory, the ability of B_1 to approximate DT is the same as that of B_2. However, there is a contradiction between the theory of dependency degree and the result of the $k-$nearest-neighbor (KNN) classifier with k ruled 3. The detailed results of two conditional attribute subsets with the KNN classifier are shown in Fig. 2. If not otherwise specified, in this paper, we transform $f(x, a)(\forall x \in U, \forall a \in AT)$ into real forms $G_{\eta}(f(x, a)) = (\lambda^-)^{1-\eta}(\lambda^+)^{\eta}$ by the geometry average interval sorting method [50] to obtain the classification result under different classifiers. In Fig. 2, we can see that under subset B_1, one sample is misclassified, while under subset B_2, none of the samples are classified incorrectly. A similar situation occurs to SVM as well. Therefore, it is limited to using dependency degree with attributes treated equally in interval-valued-based neighborhood rough sets. Therefore, in the following, a new model based on the interval-valued-based neighborhood rough set is proposed to solve this defect.

Table 4 Interval-valued neighborhood similarity class induced by $\{a_1, a_3\}$ and $\{a_1, a_5\}$

U	$IVS^{0.2}_{\{a_1,a_3\}}(x)$	$IVS^{0.2}_{\{a_1,a_5\}}(x)$
x_1	$\{x_1\}$	$\{x_1\}$
x_2	$\{x_2\}$	$\{x_2\}$
x_3	$\{x_3\}$	$\{x_3\}$
x_4	$\{x_4\}$	$\{x_4\}$
x_5	$\{x_5\}$	$\{x_5\}$
x_6	$\{x_6\}$	$\{x_6\}$
x_7	$\{x_7\}$	$\{x_7\}$
x_8	$\{x_8\}$	$\{x_8\}$
x_9	$\{x_9\}$	$\{x_9\}$
x_{10}	$\{x_{10}\}$	$\{x_{10}\}$
x_{11}	$\{x_{11}\}$	$\{x_{11}\}$
x_{12}	$\{x_{12}\}$	$\{x_{12}\}$
x_{13}	$\{x_{13}\}$	$\{x_{13}\}$
x_{14}	$\{x_{14}\}$	$\{x_{14}\}$

Table 3 An interval-valued decision information system

U	a_1	a_2	a_3	a_4	a_5	a_6	a_7	d
x_1	[27806,33596]	[1370,1910]	[185,211]	[254,254]	[406,406]	[171,171]	[143,143]	1
x_2	[27492,34092]	[1596,1753]	[185,193]	[262,262]	[415,415]	[170,170]	[143,143]	1
x_3	[19229,30885]	[1242,1910]	[155,170]	[246,246]	[380,384]	[166,166]	[148,148]	1
x_4	[19242,24742]	[1242,1753]	[167,167]	[245,245]	[383,383]	[163,163]	[132,132]	1
x_5	[19837,29034]	[1242,1242]	[158,174]	[238,238]	[372,372]	[169,169]	[144,144]	1
x_6	[16992,23492]	[1149,1149]	[151,168]	[235,235]	[343,343]	[163,163]	[142,142]	1
x_7	[24192,33042]	[1119,1994]	[160,185]	[251,251]	[399,399]	[169,169]	[142,142]	1
x_8	[41593,62291]	[1598,2492]	[200,227]	[260,260]	[443,443]	[175,175]	[142,142]	2
x_9	[68216,140265]	[1781,4172]	[216,250]	[276,276]	[480,480]	[181,181]	[145,145]	2
x_{10}	[45407,76392]	[1796,2979]	[201,247]	[273,273]	[447,447]	[174,174]	[142,142]	2
x_{11}	[50490,65399]	[1796,2497]	[195,210]	[275,275]	[475,475]	[178,178]	[143,143]	2
x_{12}	[27419,48679]	[1585,1896]	[190,191]	[251,251]	[452,452]	[173,173]	[143,143]	2
x_{13}	[36492,49092]	[1585,2171]	[193,207]	[264,264]	[450,450]	[171,171]	[143,143]	2
x_{14}	[39676,63455]	[1595,2496]	[192,220]	[270,270]	[470,470]	[175,175]	[146,146]	2

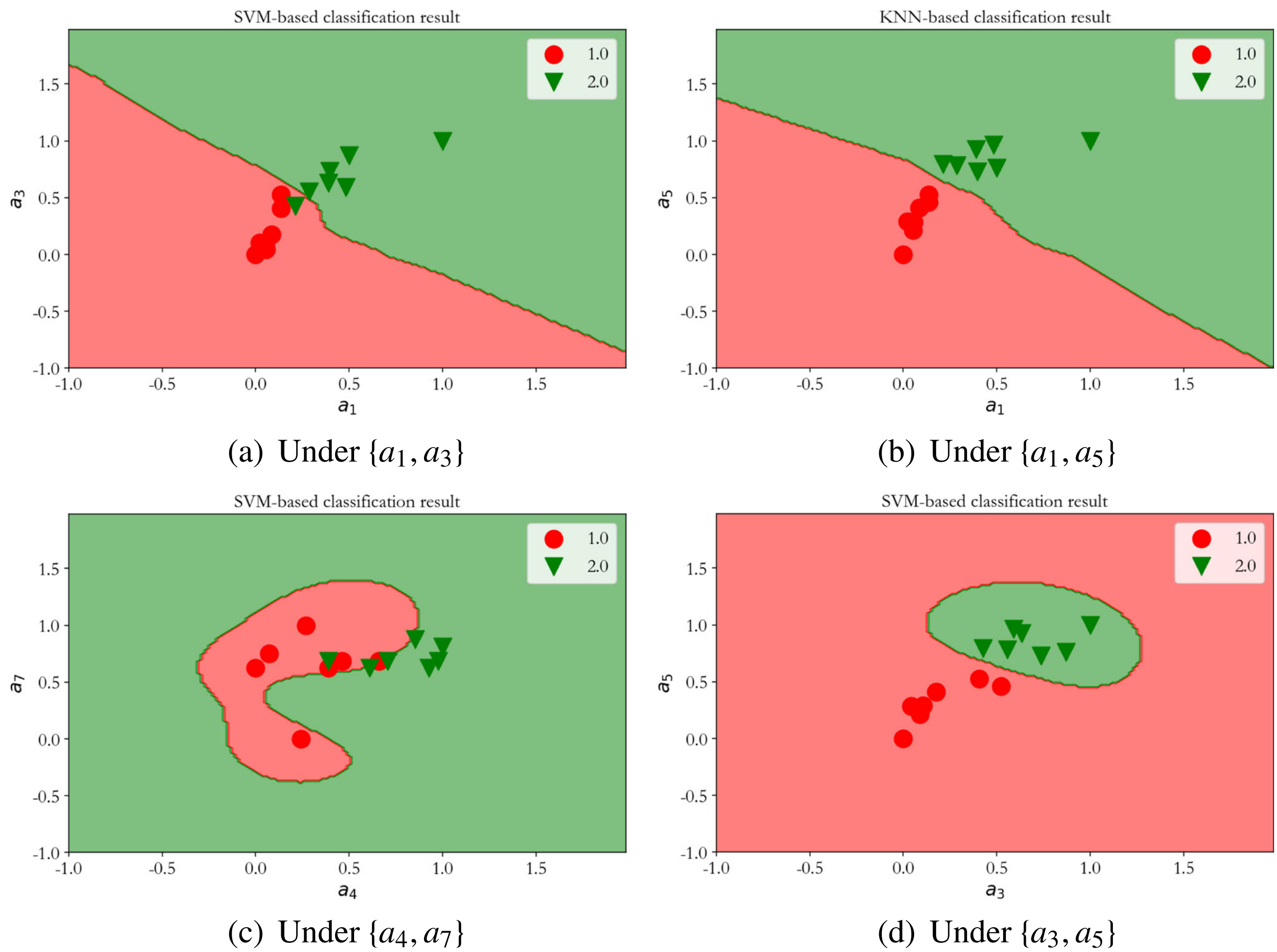

Fig. 2 Classification result under KNN and SVM

4 Interval-valued-based weighted neighborhood rough set

Through in-depth research, we find that different attributes have different significance for decision attributes. Ignoring such a difference will lead to attribute subsets with lower classification ability having a larger value of dependency degree, which is the problem with an interval-valued-based neighborhood rough set. In this part, a new rough set model named the interval-valued-based weighted neighborhood rough set is proposed to solve the aforementioned issue.

Given an interval-valued decision information system $IS = (U, AT \cup DT, F, G)$, $\forall x \in U, a \in AT$, $f(x, a) = [x_a^-, x_a^+]$ is the interval value of sample x with respect to attribute a. Considering that the correlation coefficient between the condition attribute and the decision attribute may be negative and multiplying the interval value by a negative number reverses its upper and lower bounds, e.g., $-5 \times [1, 2] = [-10, 5]$, converting the interval value to a real number becomes a necessity. Here, we still choose the geometry average interval sorting method. For any interval $\lambda = [\lambda^-, \lambda^+]$, its real number form , namely, $G_\eta(\lambda) = (\lambda^-)^{1-\eta}(\lambda^+)^\eta$, is definite when parameter η is given.

Let the coefficient matrix be

$$A = \begin{bmatrix} G_\eta(f(x_1, a_1)) & G_\eta(f(x_1, a_2)) & \cdots & G_\eta(f(x_1, a_m)) \\ G_\eta(f(x_2, a_1)) & G_\eta(f(x_2, a_2)) & \cdots & G_\eta(f(x_2, a_m)) \\ \vdots & \vdots & \ddots & \vdots \\ G_\eta(f(x_n, a_1)) & G_\eta(f(x_n, a_2)) & \cdots & G_\eta(f(x_n, a_m)) \end{bmatrix}$$

Because the real number forms are generally different when the parameter η is different, the matrix can be renamed the η—degree of preference coefficient matrix. The decision matrix is $Y = (f(x_1, d), f(x_2, d), \cdots, f(x_n, d))^T$, and the partition coefficients of attributes are $C = (c(a_1), c(a_2), \cdots, c(a_m))^T$. To determine the optimal partition coefficients of attributes, we transform the problem of seeking the optimal coefficients into an optimization problem as follows:

$$C^* = \underset{C}{\operatorname{argmin}} ||AC - Y|| \tag{13}$$

where $||\cdot||$ denotes the 2-norm of a vector. The optimal partition coefficients C can be attained when $AC = Y$. Specifically, two situations need to be considered for the solution. When A^TA is invertible, both sides of $AC = Y$ are multiplied by A^T to obtain $A^TAC = A^TY$. Finally, by solving $A^TAC = A^TY$, we can obtain

$$C = (A^TA)^{-1}A^TY. \tag{14}$$

When A^TA is not invertible, we attempt to add a penalty term based on objective function (13); then, formula (13) can be converted to $F(C) = ||AC - Y|| + ||C||^2$. Since $F(C)$ is a convex function, its minimum is obtained when $F'(C) = 0$. We can easily see that $F'(C) = 2A^T(AC - Y) + 2C$; therefore, $(A^TA + E)C = A^TY$, where E is an identity matrix. So

$$C = (A^TA + E)^{-1}A^TY. \tag{15}$$

$|C(a)|$ is the absolute value of $C(a)$, which reflects the relation between attribute a and decision D. The larger $|C(a)|$ is, the stronger the internal relevance of the attribute and the decision. Similar to such common practice, the attribute weighted method is designed in the following.

Definition 1 Given an interval-valued decision information system $IS = (U, AT \cup DT, F, G)$, $\forall a_i \in AT$, the weight of a_i is defined as

$$\omega(a_i) = \frac{|AT||C(a_i)|}{\sum_{a_i \in C}|C(a_i)|}. \tag{16}$$

Proposition 1 *Given an interval-valued decision information system $IS = (U, AT \cup DT, F, G)$, $\forall a_i \in AT$, the weight vector with attributes $\omega = (\omega(a_1), \omega(a_2), \cdots, \omega(a_m))^T$, we have*

$(1)\omega(a_i) \geq 0;$

$$(2) \sum_{a_i \in AT} \omega(a_i) = |AT|. \tag{17}$$

Proof (1) − (2) can be proved directly by Definition 1. □

From formula (17), each attribute weight is assigned by using the partition coefficients between the conditional attributes and decision attributes. The higher the correlation is, the higher the assigned weight of the corresponding attribute.

Given an interval-valued decision information system $IS = (U, AT \cup DT, F, G)$, $\omega = (\omega(a_1), \omega(a_2), \cdots, \omega(a_m))^T$ is a weight vector with attributes; for attribute subset $B(B \subseteq AT)$ and neighborhood threshold δ, the weighted interval-valued neighborhood similarity relation is defined as

$$\begin{aligned} & WIVS_B^\delta \\ &= \{(x,y)| \sqrt{\sum_{a_i \in B}(\omega(a_i)(1-\frac{|f(x,a_i)\cap f(y,a_i)|}{|f(x,a_i)\cup f(y,a_i)|}))^2} \leq \delta\} \\ &= \{(x,y)| \sqrt{\sum_{a_i \in B}\omega(a_i)^2(1-\frac{|f(x,a_i)\cap f(y,a_i)|}{|f(x,a_i)\cup f(y,a_i)|})^2} \leq \delta\}. \end{aligned} \tag{18}$$

where $\omega(a_i)$ is the weight of attribute a_i, $w(a_i) \geq 0$ and $\sum_{a_i \in AT}|\omega(a_i)| = |AT|$. Compared with the interval-valued decision information system treating the conditional attribute weights equally, the new model takes advantage of the significance of attributes. When $\omega(a_i) \geq 1$, the significance of attribute a will be increased; when $0 < \omega(a_i) < 1$, its significance will be decreased; when $\omega(a_i) = 1$, its significance will remain unchanged; and when $\omega(a_i) = 0$, the corresponding attribute a can be removed since it seems to be not important. Moreover, for $\forall a_i \in AT$, it can be seen that the interval-valued-based weighted neighborhood similarity relation will degenerate to an interval-valued-based neighborhood similarity relation when $\omega(a_i) = 1$. Therefore, the interval-valued-based weighted neighborhood rough set is a natural generalization of the interval-valued-based neighborhood rough set.

Definition 2 Given an interval-valued decision information system $IS = (U, AT \cup DT, F, G)$ and a weighted neighborhood similarity relation $WIVS_B^\delta$, δ is set to a fixed value; for any $X \subseteq U, B \subseteq AT$, the lower and upper approximations of X with respect to B are respectively defined as

$$\begin{aligned} \underline{WIVR}_B^\delta(X) &= \{x \in U | WIVS_B^\delta(x) \subseteq X\}, \\ \overline{WIVR}_B^\delta(X) &= \{x \in U | WIVS_B^\delta(x) \cap X \neq \varnothing\}. \end{aligned} \tag{19}$$

Analogously, according to formula (19), the lower and upper approximations of DT with respect to the relation $WIVS_B^\delta$ are respectively defined as

$$\begin{aligned} \underline{WIVR}_B^\delta(DT) &= \cup_{i=1}^s \underline{WIVR}_B^\delta(D_i), \\ \overline{WIVR}_B^\delta(DT) &= \cup_{i=1}^s \overline{WIVR}_B^\delta(D_i). \end{aligned} \tag{20}$$

where $U/DT = \{D_1, D_2, \cdots, D_s\}$. The boundary and positive regions of DT with respect to the relation $WIVS_B^\delta$ are defined as

$$\begin{aligned} WBND_B^\delta(DT) &= \overline{WIVR}_B^\delta(DT) - \underline{WIVR}_B^\delta(DT), \\ WPOS_B^\delta(DT) &= \underline{WIVR}_B^\delta(DT). \end{aligned} \tag{21}$$

The size of the boundary and positive regions reflects the roughness of decision DT under the weighted neighborhood similarity relation $WIVS_B^\delta$ from different angles. On the one hand, from the perspective of the approximate set, $WPOS_B^\delta(DT)$ measures the roughness of $WIVS_B^\delta$ from the aspect of the lower approximation, and $WBND_B^\delta(DT)$ considers both the upper and lower approximations. On the other hand, considering classification, the samples of the lower approximation can be classified correctly by $WIVS_B^\delta$, while some samples of the upper approximation are correctly classified and others are misclassified. Therefore, the measurement ability of the lower approximation is better than that of the upper approximation. The relevant measurement ability of the attribute subset based on the lower approximation is as follows.

Definition 3 Given an interval-valued decision information system $IS = (U, AT \cup DT, F, G)$ and a weighted neighborhood similarity relation $WIVS_B^\delta$, the dependency degree of DT with respect to the relation $WIVS_B^\delta$ in the interval-valued decision information system $IS = (U, AT \cup DT, F, G)$ is described as

$$\gamma_B^\delta(DT) = \frac{|WPOS_B^\delta(DT)|}{|U|}. \tag{22}$$

where $0 \le \gamma_B^\delta(DT) \le 1$. $\gamma_B^\delta(DT)$ is used to evaluate the ability of conditional attribute subset B to approximate DT, where the attributes of AT have different weights. The definition shows that the larger $\gamma_B^\delta(DT)$ is, the stronger the approximation ability of Bs. Comprehensively considering the definition of the dependency degree, we find that there are three factors affecting the value of $\gamma_B^\delta(DT)$: the neighborhood threshold δ controlling the neighborhood granule size, the conditional attribute subset B characterizing the samples and attribute weights. When the attribute weights are given, γ increases with the decrease in δ or the increase in attributes.

To understand the calculation process of the interval-valued-based weighted neighborhood rough set and the difference between it and the interval-valued-based neighborhood rough set, we still use **Example 1** to calculate the dependency degree in the new rough set model. By formulas (13) and (14), we obtain partition coefficients of attributes $C = (0.000004816, 0.000278600, 0.008022000, 0.015960000, 0.014120000, 0.018570000, 0.011080000)$ and $\omega = (0.0004955, 0.0286592, 0.8253064, 1.6420429, 1.4527331, 1.9109832, 1.1397797)$. Now that $\omega(a_1)$, $\omega(a_2)$ and $\omega(a_3)$ are less than 1, a_1, a_2 and a_3 have little significance in decision-making; $\omega(a_4)$, $\omega(a_5)$, $\omega(a_6)$ and $\omega(a_7)$ are evidently more than 1, so they play an indispensable role in decision-making. Under such circumstances, we recalculate the dependency degree of conditional attribute subsets $B_1 = \{a_1, a_3\}$ and $B_2 = \{a_1, a_5\}$. The interval-valued-based weighted neighborhood similarity classes induced by subsets B_1 and B_2 are displayed in Table 5. According to Definitions (19), (20), (21) and (22), we obtain $WPOS_{B_1}^{0.2}(DT) = \{x_1, x_2, x_3, x_4, x_6, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{14}\}$ and $WPOS_{B_2}^{0.2}(DT) = \{x_1, x_2, x_3, x_4, x_6, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}\}$. Thus, we can obtain $\gamma_{B_1}^{0.2}(DT) = 0.93$ and $\gamma_{B_2}^{0.2}(DT) = 1$. That is, the ability of B_1 to approximate DT is less than that of B_2, which is obviously different from the interval-valued-based neighborhood rough set. However, the performance of the new rough set model is consistent with the result in Fig. 2, namely, the KNN classifier with k ruled 3. Therefore, we can conclude that the interval-valued-based weighted neighborhood rough set can address the issue caused by the interval-valued-based neighborhood rough set. As a result, it is worth measuring the significance of conditional attributes through dependency degree by applying the attribute weighting method to the interval-valued-based neighborhood rough set model.

In the next section, we give axiomatic proofs of the relations between γ and its determining factors.

Proposition 2 *Given an interval-valued decision information system $IS = (U, AT \cup DT, F, G)$, for $B_1 \subseteq B_2 \subseteq AT$ and a neighborhood threshold δ, we have*

$$(1) WIVS_{B_1}^\delta \supseteq WIVS_{B_2}^\delta;$$
$$(2) \forall X \subseteq U, \underline{WIVR}_{B_1}^\delta(X) \subseteq \underline{WIVR}_{B_2}^\delta(X), \overline{WIVR}_{B_1}^\delta(X) \supseteq \overline{WIVR}_{B_2}^\delta(X);$$
$$(3) WPOS_{B_1}^\delta(DT) \subseteq WPOS_{B_2}^\delta(DT), \gamma_{B_1}^\delta(DT) \le \gamma_{B_2}^\delta(DT). \tag{23}$$

Table 5 Interval-valued-based weighted neighborhood similarity classes induced by $\{a_1, a_3\}$ and $\{a_1, a_5\}$

U	$WIVS_{\{a_1,a_3\}}^{0.2}(x)$	$WIVS_{\{a_1,a_5\}}^{0.2}(x)$
x_1	$\{x_1\}$	$\{x_1\}$
x_2	$\{x_2\}$	$\{x_2\}$
x_3	$\{x_3\}$	$\{x_3\}$
x_4	$\{x_4\}$	$\{x_4\}$
x_5	$\{x_5\}$	$\{x_5\}$
x_6	$\{x_6\}$	$\{x_6\}$
x_7	$\{x_7\}$	$\{x_7\}$
x_8	$\{x_8\}$	$\{x_8\}$
x_9	$\{x_9\}$	$\{x_9\}$
x_{10}	$\{x_{10}\}$	$\{x_{10}\}$
x_{11}	$\{x_{11}\}$	$\{x_{11}\}$
x_{12}	$\{x_{12}\}$	$\{x_{12}\}$
x_{13}	$\{x_1, x_{13}\}$	$\{x_{13}\}$
x_{14}	$\{x_{14}\}$	$\{x_{14}\}$

Proof (1) $\forall x, y \in U, B_1 \subseteq B_2$, from Formula (19), we have $\sum_{a_i \in B_1}(\omega(a_i)(1-\frac{|f(x,a_i)\cap f(y,a_i)|}{|f(x,a_i)\cup f(y,a_i)|}))^2 \leq \sum_{a_i \in B_2}(\omega(a_i)(1-\frac{|f(x,a_i)\cap f(y,a_i)|}{|f(x,a_i)\cup f(y,a_i)|}))^2$, so $WIVS^{\delta}_{B_1} \supseteq WIVS^{\delta}_{B_2}$.

(2) for $B_1 \subseteq B_2, \forall x \in U$, according to (1), there is $WIVS^{\delta}_{B_1}(x) \supseteq WIVS^{\delta}_{B_2}(x)$, if $x \in \underline{WIVR}^{\delta}_{B_1}(X)$, we have $WIVS^{\delta}_{B_1}(x) \subseteq X$, because $WIVS^{\delta}_{B_1}(x) \supseteq WIVS^{\delta}_{B_2}(x)$, then $WIVS^{\delta}_{B_2}(x) \subseteq X$, from formula (19) we can obtain $x \in \underline{WIVR}^{\delta}_{B_2}(X)$, so $\underline{WIVR}^{\delta}_{B_1}(X) \subseteq \underline{WIVR}^{\delta}_{B_2}(X)$; similarly, we can obtain $\overline{WIVR}^{\delta}_{B_1}(X) \supseteq \overline{WIVR}^{\delta}_{B_2}(X)$.

(3) According to (2) and formulas(20)-(21), $WPOS^{\delta}_{B_1}(DT) = \cup_{i=1}^{s}\underline{WIVR}^{\delta}_{B_1}(D_i)$ and $WPOS^{\delta}_{B_2}(DT) = \cup_{i=1}^{s}\underline{WIVR}^{\delta}_{B_2}(D_i), \forall D_i \in U/DT$, there is $\underline{WIVR}^{\delta}_{B_1}(D_i) \subseteq \underline{WIVR}^{\delta}_{B_2}(D_i)$, so $WPOS^{\delta}_{B_1}(DT) \subseteq WPOS^{\delta}_{B_2}(DT)$; then we have $\gamma^{\delta}_{B_1}(DT) \leq \gamma^{\delta}_{B_2}(DT)$. □

Proposition 3 *Given an interval-valued decision information system $IS = (U, AT \cup DT, F, G)$, for $B \subseteq AT$ and two neighborhood thresholds δ_1 and δ_2 $(\delta_1 \leq \delta_2)$, we have*

$$\begin{aligned}&(1) WIVS^{\delta_1}_B \subseteq WIVS^{\delta_2}_B;\\&(2) \forall X \subseteq U, \underline{WIVR}^{\delta_1}_B(X) \supseteq \underline{WIVR}^{\delta_2}_B(X),\\&\overline{WIVR}^{\delta_1}_B(X) \subseteq \overline{WIVR}^{\delta_2}_B(X);\\&(3) WPOS^{\delta_1}_B(DT) \supseteq WPOS^{\delta_2}_B(DT),\\&\gamma^{\delta_1}_B(DT) \geq \gamma^{\delta_2}_B(DT).\end{aligned} \tag{24}$$

Proof (1) Because $\delta_1 \leq \delta_2$, $\forall x, y \in U$, if $\sqrt{\sum_{a_i \in B}(\omega(a_i)(1-\frac{|f(x,a_i)\cap f(y,a_i)|}{|f(x,a_i)\cup f(y,a_i)|}))^2} \leq \delta_1\}$, then $\sqrt{\sum_{a_i \in B}(\omega(a_i)(1-\frac{|f(x,a_i)\cap f(y,a_i)|}{|f(x,a_i)\cup f(y,a_i)|}))^2} \leq \delta_2\}$, so $WIVS^{\delta_1}_B \subseteq WIVS^{\delta_2}_B$.

(2) According to $\delta_1 \leq \delta_2$ and (1), there is $WIVS^{\delta_1}_B(x) \subseteq WIVS^{\delta_2}_B(x)$, if $x \in \underline{WIVR}^{\delta_2}_B(X)$, we obtain $WIVS^{\delta_2}_B(x) \subseteq X$, and then $WIVS^{\delta_1}_B(x) \subseteq X$; therefore, $x \in \underline{WIVR}^{\delta_1}_B(X)$, so $\underline{WIVR}^{\delta_1}_B(X) \supseteq \underline{WIVR}^{\delta_2}_B(X)$; similarly, we can obtain $\overline{WIVR}^{\delta_1}_B(X) \subseteq \overline{WIVR}^{\delta_2}_B(X)$.

(3) According to (2) and formula (20)-(21), $WPOS^{\delta_1}_B(DT) = \cup_{i=1}^{s}\underline{WIVR}^{\delta_1}_B(D_i)$ and $WPOS^{\delta_2}_B(DT) = \cup_{i=1}^{s}\underline{WIVR}^{\delta_2}_B(D_i)$, $\forall D_i \in U/DT$, there is $\underline{WIVR}^{\delta_1}_B(D_i) \supseteq \underline{WIVR}^{\delta_2}_B(D_i)$, so $WPOS^{\delta_1}_B(DT) \supseteq WPOS^{\delta_2}_B(DT)$; then, we have $\gamma^{\delta_1}_B(DT) \geq \gamma^{\delta_2}_B(DT)$. □

5 Feature selection in the interval-valued decision information systems based weighted neighborhood rough set

In Section 4, we are committed to investigating the mechanics of the interval-valued-based weighted neighborhood rough set. Furthermore, in this section, we settle corresponding feature selection issues based on this new rough set model.

Definition 4 Given an interval-valued decision information system $IS = (U, AT \cup DT, F, G)$, a neighborhood threshold δ, an attribute subset $B \subseteq AT$ and an attribute $a_i \in B$, a_i is the redundant attribute of B with respect to DT if $\gamma^{\delta}_{B-\{a_i\}}(DT) = \gamma^{\delta}_B(DT)$, and a_i is the necessary attribute of B with respect to DT if $\gamma^{\delta}_{B-\{a_i\}}(DT) < \gamma^{\delta}_B(DT)$. Under this circumstance, the conditional subset B is a reduct of AT relative to DT when the following constraints are established:

$$\begin{aligned}&(1)\gamma^{\delta}_B(DT) = \gamma^{\delta}_{AT}(DT);\\&(2)\forall a_i \in B, \gamma^{\delta}_{B-\{a_i\}}(DT) < \gamma^{\delta}_B(DT).\end{aligned} \tag{25}$$

From the above discussion, the dependency degree of the interval-valued-based weighted neighborhood similarity relation can serve as an evaluation metric of the significance of conditional attribute subsets. Its ability to distinguish samples from different decisions increases with the increasing dependency degree. Following Definition 4, for any interval-valued decision information system with m conditional attributes, there are a total of $2^m - 1$ suspected attribute subsets needing verification. Evidently, it is unrealistic to calculate the dependency degree of each candidate attribute subset one by one. Hence, several strategies for finding a reduct, such as genetic algorithm, branch and bound, and greedy search, have been proposed to save time. In this paper, a greedy search algorithm is chosen as the strategy to find one optimal attribute subset. Before that, two measures to evaluate the significance of an attribute relative to an attribute subset are defined to construct the greedy search algorithm.

Definition 5 Given an interval-valued decision information system $IS = (U, AT \cup DT, F, G)$, a neighborhood threshold δ, an attribute subset $B \subseteq AT$ and an attribute $a_i \in B$, the internal significance of a_i relative to B under DT is defined as

$$INS(a_i, B, DT) = \gamma^{\delta}_B(DT) - \gamma^{\delta}_{B-\{a_i\}}(DT). \tag{26}$$

Obviously, according to formulas (22) and (23), we have $0 \leq INS(a_i, B, DT) \leq 1$. a_i is not an internal necessary attribute relative to B if $INS(a_i, B, DT) = 0$; then, a_i can be removed from B. In addition, a_i is an internal necessary attribute relative to B if $INS(a_i, B, DT) > 0$,

Definition 6 Given an interval-valued decision information system $IS = (U, AT \cup DT, F, G)$, a neighborhood threshold δ, an attribute subset $B \subseteq AT$ and an attribute $a_i \in AT - B$, the external significance of a_i relative to B under DT is defined as

$$EXS(a_i, B, DT) = \gamma^{\delta}_{B \cup a_i}(DT) - \gamma^{\delta}_{B}(DT). \tag{27}$$

Similarly, according to formulas (22) and (23), we have $0 \leq EXS(a_i, B, DT) \leq 1$. a_i is not an external necessary attribute relative to B if $EXS(a_i, B, DT) = 0$; then, a_i cannot be added to B. In addition, a_i is an external necessary attribute relative to B if $EXS(a_i, B, DT) > 0$.

To obtain one attribute subset characterizing the samples while simultaneously controlling the attribute size, two greedy searches strategies are proposed: forward search and backward search. In the forward search stage, attributes whose corresponding external significance are greater than zero are added to the candidate attribute subset. During backward searching, attributes that serve as internal unnecessary attributes are removed from the candidate attribute subset. Moreover, to continue the forward search smoothly, $\gamma^{\delta}_{B}(DT)$ is ruled 0 if B is an empty set. Details about the feature selection algorithm found on the interval-valued-based weighted neighborhood rough set (IVWNRS) are shown in Algorithm 1. In Algorithm 1, there is an unknown parameter affecting the size of the information granules, which needs to be set in advance. Ordinarily, δ is set by the prior knowledge of experts or the result of an isometric search. In the experimental section, we show how to search the appropriate value for the threshold δ.

It is not hard to find that the interval-valued-based weighted neighborhood similarity relation plays an important role in interval-valued-based weighted neighborhood rough set theory. As the core of the interval-valued-based weighted neighborhood similarity relation, the Jaccard distance between x and y ($\forall x, y \in U$) is viewed as a basic operation, and then the time complexity of the feature selection algorithm in the worst case can be further calculated. In steps 1-2, the initial state of RED is an empty set, and the weights of all conditional attributes are calculated with time complexity ignored. Steps 3-15 describe the sequential forward search of the feature selection algorithm by adding attributes to RED. For each cycle, we need to calculate $|U/DT||U|^2(1 + |AT - RED|)$ times, and there are $|AT|$ cycles at most. In this situation, the time complexity from 3-15 is $\mathcal{O}(|U/DT||U|^2|AT|^2)$ in all. After forward searching, there may be redundant attributes in RED, so steps 16-23 are devoted to removing these attributes. Through our calculation, the time complexity in this part is $\mathcal{O}(|RED||U/DT||U|^2)$. In summary, the time complexity of this feature selection algorithm is $\mathcal{O}(|U/DT||U|^2|AT|^2)$.

Algorithm 1 Feature selection founded on interval-valued-based weighted neighborhood rough set.

Input: An interval-valued decision information system $IS = (U, AT \cup DT, F, G)$, and a threshold δ

Output: One feature subset RED

1 Initialize RED $\leftarrow \varnothing$;// RED is initialized to an empty set;
2 Compute the weight of each conditional attribute by formula (15);
3 **while** *AT-RED* $\neq \varnothing$ **do**
4   Compute the dependency degree $\gamma^{\delta}_{RED}(DT)$;
5   **for** *each* $a_i \in AT - RED$ **do**
6   Compute the dependency degree $\gamma^{\delta}_{RED \cup a_i}(DT)$;
7         $EXS(a_i, RED, DT) = \gamma^{\delta}_{RED \cup \{a_i\}}(DT) - \gamma^{\delta}_{RED}(DT)$
8   **end**
9   Find a_k with $EXS(a_k, RED, DT) = \max_{a_i \in AT - RED}(EXS(a_i, RED, DT))$;
10   **if** $EXS(a_k, RED, DT) = 0$ **then**
11   break;
12   **else**
13         $RED \leftarrow RED \cup \{a_i\}$; // Add the appropriate attribute to RED;
14   **end**
15 **end**
16 **for** $a_i \in RED$ **do**
17   Compute the dependency degree $\gamma^{\delta}_{RED}(DT)$ and $\gamma^{\delta}_{RED-\{a_i\}}(DT)$;
18     $INS(a_i, RED, DT) = \gamma^{\delta}_{RED}(DT) - \gamma^{\delta}_{RED-\{a_i\}}(DT)$;
19   **if** $INS(a_i, RED, DT) = 0$ **then**
20         $RED \leftarrow RED - \{a_i\}$; // Eliminate the attribute satisfying the constraint ;
21   **end**
22 **end**
23 return RED;

Example 2 Further considering **Example 1** in terms of feature selection after using the weighted method, we set $\eta = 0.5$ and neighborhood threshold $\delta = 0.2$. According to Definition 4, it can be calculated that $\gamma^{0.2}_{\{a_1,a_2,a_3,a_4,a_5,a_6,a_7\}}(DT) = 1$ and $\gamma^{0.2}_{\{a_5\}}(DT) = 1$. Obviously, there is no redundant attribute in attribute subset $\{a_5\}$. Therefore, $\{a_5\}$

is one reduct in **Example 1**. The result is consistent with that of the feature selection algorithm, but it has low efficiency. In the feature selection algorithm, we directly obtain the result $\{a_5\}$ without considering each suspected attribute subset.

6 Numerical experiment

In this section, we design a series of experiments to verify the effectiveness and robustness of the proposed IVWNRS algorithm. Emphasizing the importance of the attribute weights, the interval-valued based weighted neighborhood rough set(IVWNRS) is compared with the interval-valued based neighborhood rough set (IVNRS), local neighborhood rough set [40], distance measure based fuzzy rough set(AVDP) [39] (the parameter γ is set 0.005), multigranularity attribute selector(MGAS) [22], bucket and attribute group based neighborhood rough set(BAGR) [4] and multilevel neighborhood-based sequential three-way decision(MNS3WD) [44] (we only consider the horizontal granularity based on γ and set the decision thresholds (α, β) in (0.15,0.3) and (0,0.15)). Furthermore, to show the importance of employing interval-valued information comprehensively and assigning weights to attributes, the compared algorithms are carried out in datasets with real forms using the geometry average interval sorting method. Influenced by magnitude, these datasets are normalized into [0,1].

To make the result more persuasive, we compare the feature selection methods from two aspects: 1) the classification accuracies under different classifiers and 2) the number of selected attributes. All algorithms are executed in Jupyter Notebook 5.0.0 and MATLAB R2018a and run in a hardware environment with an Inter(R) Core(TM) i5-9300H CPU @2.40 GHz with 8.00GB RAM.

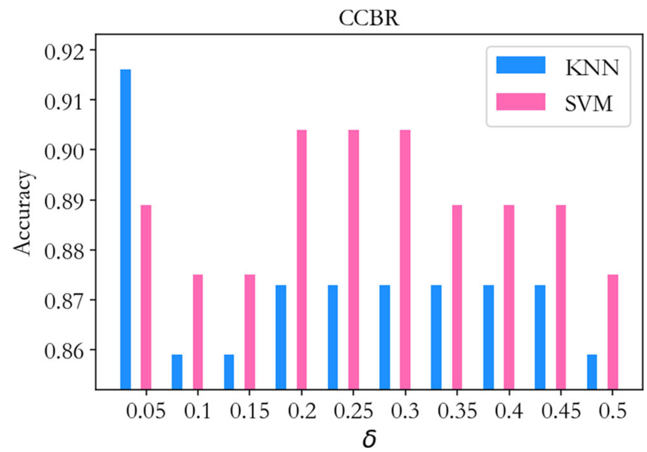

Fig. 3 CCBR

In classification ability, we use KNN (ruled k=3) and SVM (with default parameters) classifiers to evaluate the performance of these algorithms. Fifteen datasets downloaded from the UCI machine learning repository are tested, and their details are described in Table 6. Since the dataset values are all real numbers, we construct interval-valued datasets by multiplying error precision α. Then, $f(x,a)(\forall x \in U, \forall a \in AT)$, the value of object x under attribute a can be expressed as $[(1-\alpha)\times f(x,a), (1+\alpha)\times f(x,a)]$. Specifically, when $f(x,a) \leq 0$, we have $[(1+\alpha)\times f(x,a), (1-\alpha)\times f(x,a)]$. In this paper, we set the error precision $\alpha = 0.05$. To eliminate as many accidental errors as possible, 10-fold cross validation is employed to

Table 6 Data description

Datasets	Abbreviation	Samples	Attributes	Labels
Cervical Cancer Behavior Risk	CCBR	72	19	2
Data for Software Engineering Teamwork Assessment in Education Setting	DSETAES	74	109	2
Tae	Tae	151	5	3
Ultrasonic Flowmeter Diagnostics	UFD	179	44	4
Haberman	Haberman	306	3	2
Ionosphere	Ionosphere	351	34	2
Whole Scale Customers	WSC	440	7	3
Wdbc	Wdbc	569	31	2
Hill-Valley	HV	606	100	2
Blood Transfusion Service Center	BTSC	748	4	2
Contraceptive Method Choice	CMC	1473	8	2
Wireless Indoor Localization	WIL	2000	7	4
Estimation of Obesity Levels based on Eating Habits and Physical Condition	EOLEHPC	2111	16	7
Electrical Grid Stability Simulated	EGSS	10000	13	2
A141 2020 Predictive Maintenance	APM	10000	14	3

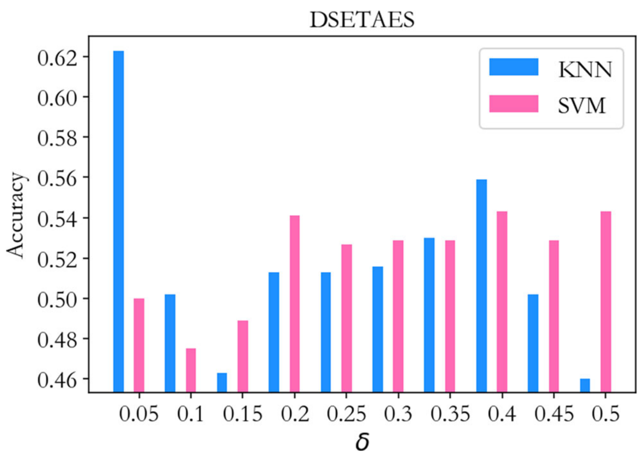

Fig. 4 DSETAES

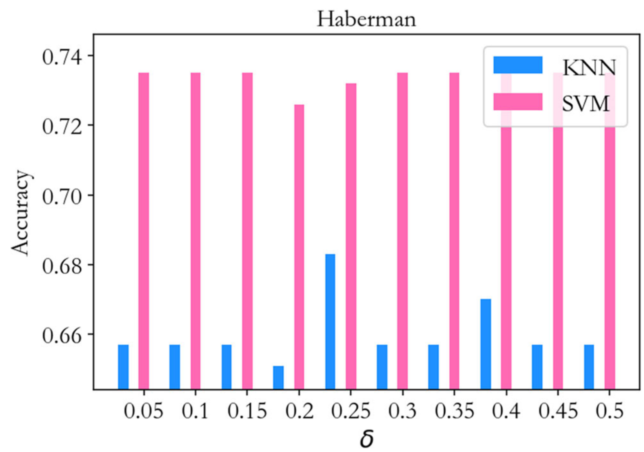

Fig. 7 Haberman

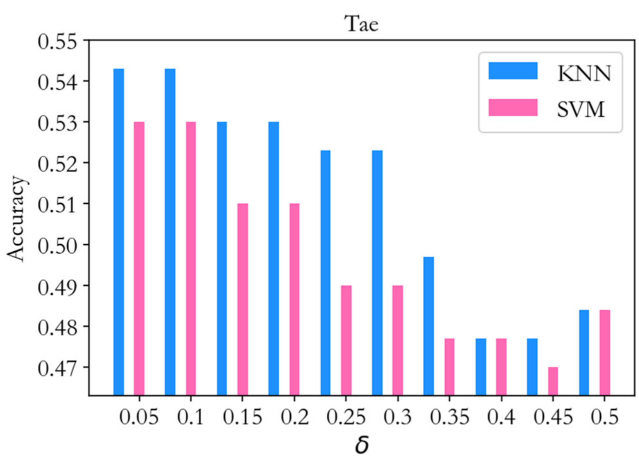

Fig. 5 Tae

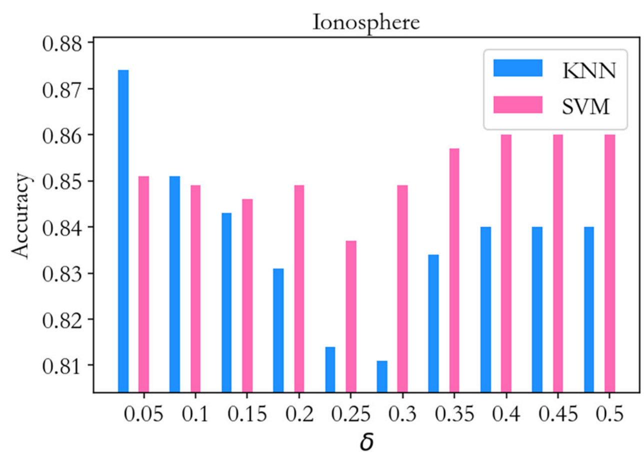

Fig. 8 Ionosphere

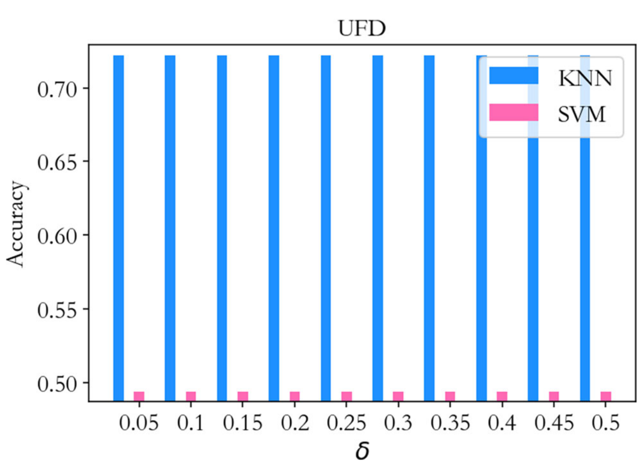

Fig. 6 UFD

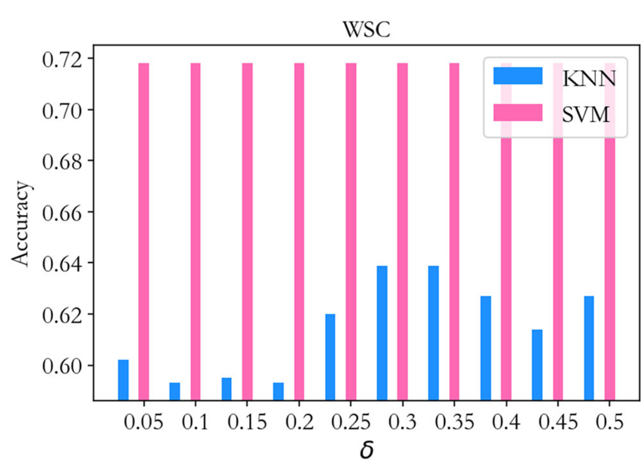

Fig. 9 WSC

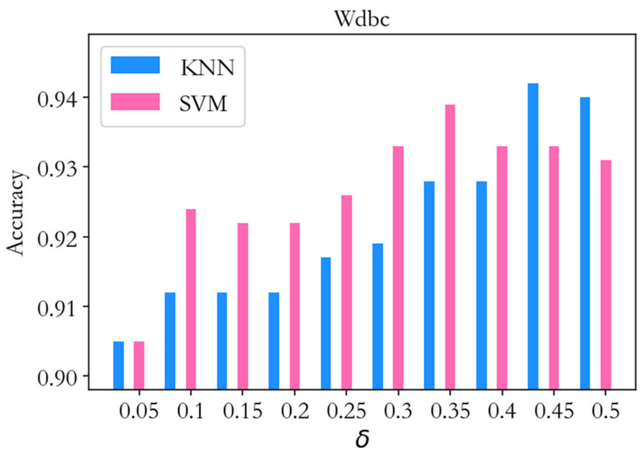

Fig. 10 Wdbc

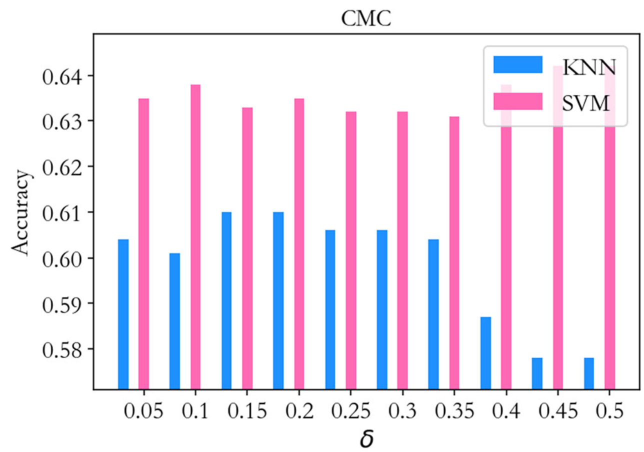

Fig. 13 CMC

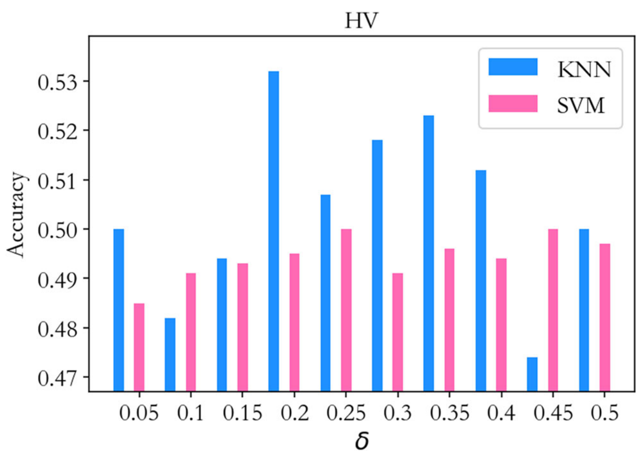

Fig. 11 HV

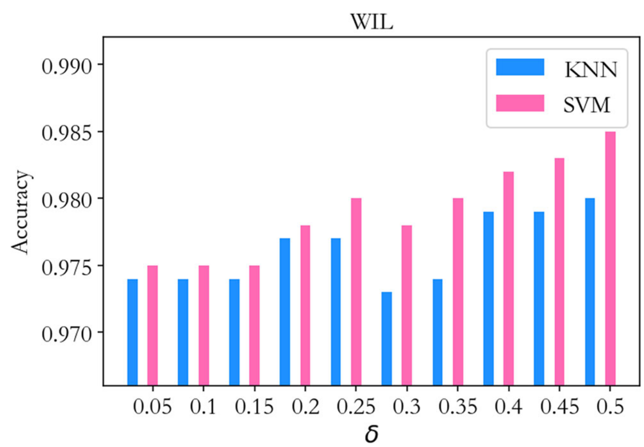

Fig. 14 WIL

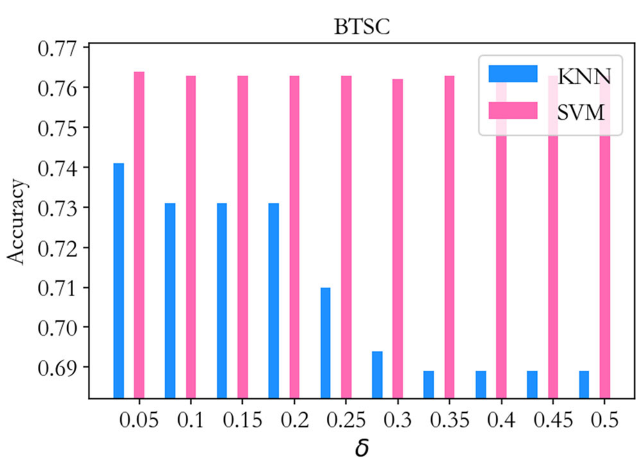

Fig. 12 BTSC

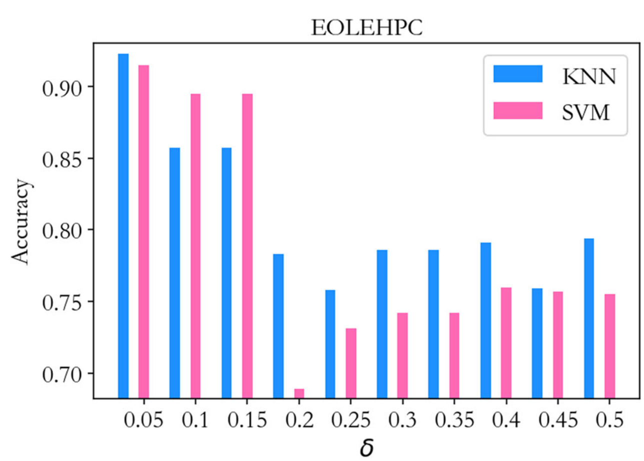

Fig. 15 EOLEHPC

Table 7 Classification accuracies of different algorithms with KNN

Datasets	Row Data	IVWNRS	IVNRS	LCER	AVDP	MGAS	BAGR	MNS3WD
CCBR	0.904	**0.916**	0.891	0.766	0.620	0.823	0.821	0.773
DSETAES	0.575	**0.623**	0.566	0.568	0.595	0.595	0.489	0.514
Tae	0.517	**0.543**	0.444	0.470	0.411	0.510	0.470	0.411
UFD	0.848	0.833	0.760	0.850	0.594	0.844	**0.872**	0.522
Haberman	0.687	0.657	0.687	**0.690**	0.684	**0.690**	**0.690**	0.480
Ionosphere	0.860	**0.874**	0.494	0.754	0.520	0.520	0.843	0.706
WSC	0.600	0.602	0.598	0.600	**0.655**	0.600	0.600	**0.655**
Wdbc	0.974	0.928	0.903	0.945	0.847	**0.995**	0.951	0.613
HV	0.548	0.500	0.492	0.548	0.490	0.548	**0.550**	0.474
BTSC	0.739	0.740	**0.749**	0.742	0.689	0.742	0.742	0.677
CMC	0.593	0.604	**0.605**	0.597	0.425	0.597	0.597	0.539
WIL	**0.988**	0.976	0.976	0.979	0.551	0.987	0.984	0.750
EOLEHPC	0.795	**0.924**	**0.924**	0.783	0.136	0.905	0.820	0.172
EGSS	0.901	**1.000**	**1.000**	0.952	0.549	0.901	0.951	0.606
APM	**0.502**	0.485	0.485	**0.502**	0.455	**0.502**	**0.502**	0.464
Average	0.735	**0.747**	0.705	0.716	0.548	0.717	0.725	0.557

evaluate the performance of these algorithms. The dataset is divided into 10 pieces: one piece is chosen as the testing set, and the remaining pieces are used as the training set. During the training stage, we use a feature selection algorithm to select attributes; then, in the testing stage, the classification accuracies of KNN and SVM can be calculated under the selected attributes. After ten cycles, the final classification accuracies of KNN and SVM can be acquired by calculating the corresponding average classification accuracy. To make the comparison among different models more objective, we set the parameters in advance. η in the geometry average interval sorting method is set to 0.5 with a moderate attitude.

We search neighborhood threshold δ from 0.05 to 0.5 with steps 0.05 and find the optimal δ when it performs relatively well on all datasets under both KNN and SVM by IVWNRS. The search result under thirteen datasets shows that $\delta = 0.05$ can satisfy our requirements in most cases, as shown in Figs. 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 and 15. It must be stated that isometric research is used in thirteen datasets instead of fifteen because the remaining datasets require such a long time to explore that we are unable to afford the time. Thus, the neighborhood radius of the two datasets naturally depends on the thirteen datasets without searching step by step.

Table 8 Classification accuracies of different algorithms with SVM

Datasets	Row Data	IVWNRS	IVNRS	LCER	AVDP	MGAS	BAGR	MNS3WD
CCBR	**0.904**	0.889	0.877	0.695	0.707	0.807	0.877	0.761
DSETAES	0.555	0.500	0.457	0.498	0.457	0.457	**0.575**	0.543
Tae	0.516	**0.530**	0.436	0.476	0.358	0.516	0.476	0.312
UFD	0.653	**0.776**	0.497	0.617	0.472	0.617	0.606	0.528
Haberman	0.719	**0.735**	0.719	0.719	0.733	0.719	0.719	**0.735**
Ionosphere	**0.934**	0.851	0.640	0.802	0.640	0.640	0.903	0.749
WSC	**0.718**	**0.718**	**0.718**	**0.718**	**0.718**	**0.718**	**0.718**	**0.718**
Wdbc	0.979	0.923	0.910	0.960	0.866	**0.995**	0.960	0.624
HV	0.498	0.485	0.493	0.498	**0.500**	0.498	0.498	0.497
BTSC	0.763	**0.765**	0.760	0.763	0.763	0.763	0.763	0.762
CMC	0.633	0.635	**0.639**	0.633	0.599	0.633	0.633	0.622
WIL	0.982	0.975	0.975	0.980	0.640	0.983	**0.984**	0.780
EOLEHPC	0.816	**0.913**	**0.913**	0.844	0.169	0.911	0.863	0.307
EGSS	0.982	**0.998**	**0.998**	0.992	0.638	0.982	0.990	0.685
APM	**0.600**	**0.600**	**0.600**	**0.600**	**0.600**	**0.600**	**0.600**	**0.600**
Average	0.750	**0.753**	0.709	0.720	0.591	0.723	0.744	0.615

Table 9 Number of selected features

Datasets	IVWNRS	IVNRS	LCER	AVDP	MGAS	BAGR	MNS3WD
CCBR	2.90	2.80	2.70	3.90	4.80	3.30	1.00
DSETAES	1.00	1.00	1.00	1.00	1.00	3.90	1.00
Tae	3.90	3.30	4.00	1.00	5.00	4.00	1.00
UFD	3.00	1.00	34.10	1.00	42.10	13.10	1.00
Haberman	2.00	3.00	3.00	1.00	3.00	3.00	1.00
Ionosphere	2.00	1.00	1.00	1.00	1.00	5.00	1.00
WSC	2.00	2.00	7.00	1.00	7.00	7.00	1.00
Wdbc	2.00	2.00	5.00	1.00	30.10	5.10	1.00
HV	2.00	2.00	100.00	1.00	100.00	95.60	1.00
BTSC	1.60	3.00	4.00	1.00	4.00	4.00	1.00
CMC	7.00	6.90	8.00	1.00	8.00	8.00	1.00
WIL	4.00	4.00	6.00	1.00	7.00	6.00	1.00
EOLEHPC	4.90	4.90	7.80	1.00	16.00	7.40	1.00
EGSS	1.50	1.50	5.00	1.00	12.00	5.20	1.00
APM	2.00	2.00	7.80	1.00	11.00	7.80	1.00
Average	2.79	2.69	13.09	1.19	16.79	11.89	1.00

The classification accuracies of the row data and the seven feature selection algorithms under KNN and SVM are displayed in Tables 7 and 8, where the best performance is highlighted in bold and underlined over different feature selection algorithms. Compared with the average accuracy of the raw data, IVWNRS improved by 1.2% and 0.3% with KNN and SVM, other algorithms were slightly inferior to the raw data, AVDP declined by 18.7% and 15.9% and MNS3WD declined by 17.8% and 13.5%. As the improved IVNRS model, the performance of IVWNRS with KNN and SVM improved by 4.2% and 4.4%, respectively, after solving the defect of IVNRS. Evidently, for classification accuracies, IVWNRS performs best in most cases. Therefore, we can conclude that IVWNRS improves the classification ability by compensating for the IVNRS shortcomings based on utilizing the pivotal information of the raw data (Table 9).

Average accuracy alone may not evaluate the performances of the algorithms completely mentioned above, so the hypothesis test is employed to describe their significant

Table 10 Rank of the seven algorithms with KNN

Datasets	IVWNRS	IVNRS	LCER	AVDP	MGAS	BAGR	MNS3WD
CCBR	1	2	6	7	3	4	5
DSETAES	1	5	4	2.5	2.5	7	6
Tae	1	5	3.5	6.5	2	3.5	6.5
UFD	4	5	2	6	3	1	7
Haberman	6	4	2	5	2	2	7
Ionosphere	1	7	3	5.5	5.5	2	4
WSC	3	7	5	1.5	5	5	1.5
Wdbc	4	5	3	6	1	2	7
HV	4	5	2.5	6	2.5	1	7
BTSC	5	1	3	6	3	3	7
CMC	2	1	4	7	4	4	6
WIL	4.5	4.5	3	7	1	2	6
EOLEHPC	1.5	1.5	5	7	3	4	6
EGSS	1.5	1.5	3	7	5	4	6
APM	4.5	4.5	2	7	2	2	6
Average	3.89	2.82	3.50	5.71	3.04	3.18	5.86

Table 11 Rank of the seven algorithms with SVM

Datasets	IVWNRS	IVNRS	LCER	AVDP	MGAS	BAGR	MNS3WD
CCBR	1	2.5	7	6	4	2.5	5
DSETAES	3	6	4	6	6	1	2
Tae	1	5	3.5	6	2	3.5	7
UFD	1	6	2.5	7	2.5	4	5
Haberman	1.5	5.5	5.5	3	5.5	5.5	1.5
Ionosphere	2	6	3	6	6	1	4
WSC	4	4	4	4	4	4	4
Wdbc	4	5	2.5	6	1	2.5	7
HV	7	6	2	1	2	2	5
BTSC	1	7	2.5	2.5	2.5	2.5	6
CMC	2	1	4	7	4	4	6
WIL	4.5	4.5	3	7	2	1	6
EOLEHPC	1.5	1.5	5	7	3	4	6
EGSS	1.5	1.5	3	7	4	5	6
APM	4	4	4	4	4	4	4
Average	2.60	4.37	3.70	5.30	3.50	3.10	4.97

differences. First, we examine whether the seven algorithms are significantly different from each other. Here we used the Friedman test statistic, namely

$$\chi^2 = \frac{12n}{k(k+1)}\left(\sum_{i=1}^{k} r_i^2 - \frac{k(k+1)^2}{4}\right) \tag{28}$$

$$F = \frac{(N-1)\chi^2}{N(k-1)-\chi^2} \tag{29}$$

where r_i is the average ordering of the $i-th$ algorithm, N is the number of datasets and k is the number of algorithms. The variable F represents the F distribution with $k-1$ and $(k-1)(N-1)$ freedom. Generally, we conduct the Friedman test at a 90% significance level and set the null hypothesis that these algorithms have no significant differences. If the value of the Friedman statistic is more than the critical value $F(6, 84) = 1.8455$, we can conclude that there are significant differences among these algorithms; otherwise, they are indistinguishable.

Tables 10 and 11 show the orderings of the seven algorithms under KNN and SVM. Following the Friedman statistic, we obtain $F = 7.9049$ under KNN and $F = 1.1559$ under SVM. Obviously, there are significant differences under KNN, while the seven algorithms are recognized as equivalent under SVM.

Now that the algorithms with KNN have significant differences, we have to proceed with the Nemenyi test under KNN as a post hoc test. In this case, any pair of algorithms can be recognized as significantly different if the distance between the average ordering values exceeds the critical value, where the critical value can be calculated as

$$CD_\alpha = q_\alpha\sqrt{\frac{k(k+1)}{6N}} \tag{30}$$

From [6], we obtain that $q_{0.1} = 2.693$ when $k = 7$ and $\alpha = 0.1$. Therefore, we use $CD_{0.1} = 2.1243$ in this paper.

Supported by Table 10, we can calculate the distance between the average orderings of any pair of algorithms. Through calculation, we find that the distances between the

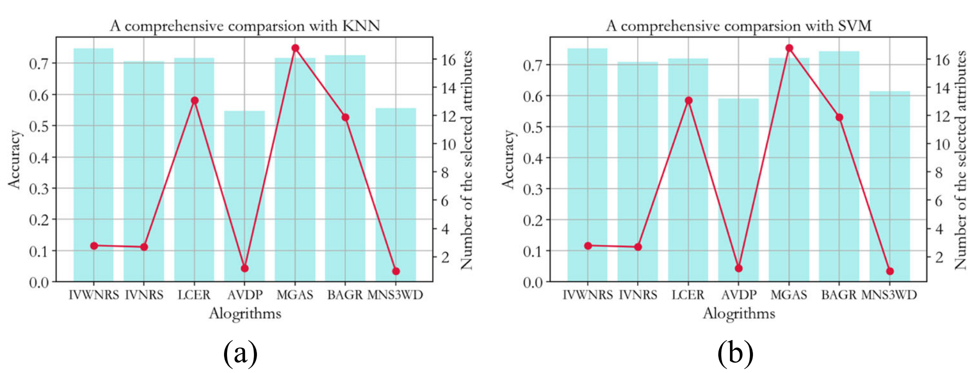

Fig. 16 A comprehensive comparison among all algorithms

average orderings of IVWNRS to AVDP and MNS3WD are 2.8667 and 2.9333 respectively. Consequently, IVWNRS is significantly better than AVDP and MNS3WD under KNN. However, the distance between IVWNRS and other algorithms is less than 2.1243, which means that they have no significant differences.

In addition to the classification accuracy, the size of the selected attributes makes up an important part of feature selection. The purpose of feature selection is to find a set of relatively few and informative condition attributes. The average size of the selected attributes with 10-fold cross validation is shown in Table 9. In Table 9, we can see that IVWNRS is only slightly inferior to IVNRS, AVDP and MNS3WD with 2.79 attributes selected.

To make the comparison independent of any single isolated aspect, a comprehensive comparison seems inevitable combined with the performance and size of the selected attributes. As shown in Fig. 16, in the number of selected attributes, AVDP and MNS3WD select relatively fewer attributes, while in performance, they all have low accuracies. At its root, such a situation is ascribed to their inappropriate choices of attributes. Although the size of selected attributes of IVWNRS is in the middle among all algorithms, it is capable of classification tasks. In general, unlike any other algorithm, IVWNRS can select relatively few conditional attributes with discernment information to obtain satisfactory classification performance.

7 Conclusion and future work

Through our study and experiments, we found that removing redundant attributes in datasets can improve classification ability with less storage space. However, problems exist in interval-valued-based datasets by using other state-of-the-art feature selection methods. On the one hand, these algorithms overlooking the importance of the attribute weights makes them miss the indispensable attributes. On the other hand, their direct use of real-valued datasets converted from interval-valued causes a lack of information. Aimed at such defects, we introduced the weighted method by calculating the $\eta-$degree of preference coefficient of attributes into neighborhood relations to measure the interval-valued-based weighted neighborhood rough set (IVWNRS). The neighborhood relations were calculated by interval processing methods such as Jaccard distance. To evaluate the significance of attribute subsets, the dependency degree was used. Finally, a greedy search based on feature selection in IVWNRS was designed to find one minimal attribute subset with high relevance and dependence on decisions. Experimental results show that IVWNRS can accomplish classification tasks compared to other state-of-the-art methods.

This paper focuses on the correlation coefficient between attributes and decisions, but the correlation coefficient is determined by decision preference to a large extent, and the correlation coefficient between attributes is not fully mined. In the future, the correlation coefficient between interval-valued attributes and real-valued decisions needs further study. Simultaneously, attributes will be assigned to different weights after analyzing the correlation coefficient between attributes. Other measures of uncertainty addressing interval values can also be considered to improve classification accuracy.

Acknowledgements This paper is supported by the National Natural Science Foundation of China (Nos. 61976245, 61772002)

References

1. Brtka V, Stokic E, Srdic B (2008) Automated extraction of decision rules for leptin dynamics-A rough sets approach. J Biomed Inform 41(4):667–674
2. Chen L, Chen D, Wang H (2019) Fuzzy kernel alignment with application to attribute reduction of heterogeneous data. IEEE Trans Fuzzy Syst 27(7):1469–1478
3. 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
4. Chen Y, Liu KY, Song JJ et al (2020) Attribute Group for Attribute Reduction. Inf Sci 535:64–80
5. Chen HF, Long JW, Qu XP (2019) A positive Region-Based attribute reduction approach in interval valued decision information system. J Chongqing Univ Tech(Natural Sci) 33(11):130–136
6. Demsar J (2006) Statistical comparisons of classifiers over multiple data sets. J Mach Learn Res 7:1-30
7. Fan XQ, Li XF, Zhao SY, Chen H, Li CP (2018) Weighted attribute reduction based on fuzzy rough sets. Comput Sci 45(01):133–139
8. Fu YQ (2019) Design of attribute subset selection and fusion classification method via dominant rough sets. Value Eng 38(28):226–229
9. Fu WQ, Khalil AM (2021) Graded rough sets based on neighborhood operator over two different universes and their applications in Decision-making problems. J Intell Fuzzy Syst 41(2):2639–2664
10. Gonzalez-Lopez J, Ventura S, Cano A (2020) Distributed selection of continuous features in multilabel classification using mutual information. IEEE Trans Neural Netw Learn Syst 31(7):2280–2293
11. Guo YT, Tsang ECC, Hu M et al (2020) Incremental updating approximations for double-quantitative decision-theoretic rough sets with the lariation of Objects. Knowl-Based Syst:189
12. Guo YT, Tsang ECC, Xu WH et al (2020) Adaptive weighted generalized multi-granulation interval-valued decision-theoretic rough sets. Knowl-Based Syst:187
13. Hashemzadeh M, Oskouei AG, Farajzadeh N (2019) New fuzzy c-means clustering method based on feature-weight and cluster-weight learning. Appl Soft Comput 78:324-345
14. Huang YY, Li TR, Luo C, Fujita H, Horng SJ (2018) Dynamic fusion of multisource interval-valued data by fuzzy granulation. IEEE Trans Fuzzy Syst 26(6):1–1

15. Huang QQ, Li TR et al (2020) Dynamic dominance rough set approach for processing composite ordered data. Knowl-Based Syst:187
16. Huang J, Wei Y, Yi J, Liu M (2018) An improved KNN Based on class contribution and feature weighting, 2018 10th international conference on measuring technology and mechatronics automation (ICMTMA). IEEE:313–316
17. Jiang HB, Hu BQ (2021) A decision-theoretic fuzzy rough set in hesitant fuzzy information systems and its application in multi-attribute decision-making. Inf Sci 579:103–127
18. Jiang ZH, Liu KY et al (2020) Accelerator for supervised neighborhood based attribute reduction. Int J Approx Reason 119:122–150
19. Kong QZ, Zhang XW, Xu WH, Xie ST (2020) Attribute reducts of multi-granulation information system. Artif Intell Rev 53(2):1353–1371
20. Liang S, Yang X, Chen X et al (2018) Stable attribute reduction for neighborhood rough set. Filomat 32(5):1809–1815
21. Liu Q, Dai JH, Chen JL (2021) Cost-sensitive feature selection for interval-valued data. J Nanjing Univ(Nature Sci) 57(1):121–129
22. Liu KY, Yang XB, Fujita H et al (2019) An efficient selector for multi-granularity attribute reduction. Inf Sci 505:457–472
23. Luo C, Ju YB, Dong PW et al (2021) Risk assessment for ppp waste-to-energy incineration plant projects in china based on hybrid weight methods and weighted multigranulation fuzzy rough sets. Sustainable cities and society:74
24. Luo C, Li T, Huang Y, Fujita H (2019) Updating Three-Way decisions in incomplete multi-scale information systems. Inf Sci 476:274–289
25. Mariello A, Battiti R (2018) Feature selection based on the neighborhood entropy. IEEE Trans Neural Netw Learn Syst 29(12):6313–6322
26. Pawlak L (1982) Sets, rough. Int J Comput Inform Sci 11:341–356
27. Qian YH, Liang JY, Dang CY (2008) Converse approximation and rule extraction from decision tables in rough set theory. Comput Math Appl 55(8):1754–1765
28. Ren YG, Zhang YP, Zhang ZP (2020) Collaborative filtering recommendation algorithm based on rough set rule extraction. J Commun 41(1):76–83
29. Sang BB, Chen H, Yang L et al (2021) Feature selection for dynamic interval-valued ordered data based on fuzzy dominance neighborhood rough set. Knowl-Based Syst 227:107–223
30. Sang BB, Chen HM, Yang L et al (2021) Incremental feature selection using a conditional entropy based on fuzzy dominance neighborhood rough sets. IEEE Trans Fuzzy Syst 99:1–1
31. Sun BZ, Gong ZT, Chen DG (2008) Fuzzy rough set theory for the interval-valued fuzzy information systems. Inf Sci 178(13):2794–2815
32. Sun L, Wang TX, Ding WP et al (2021) Feature selection using fisher score and multilabel neighborhood rough sets for multilabel classification. Inf Sci 578:887–912
33. Tsang ECC, Hu Q, Chen D (2016) Feature and instance reduction for PNN classifiers based on fuzzy rough sets. Int J Mach Learn Cybernetics 7:1–11
34. Vluymans S, Parthalain NM, Cornelis C et al (2019) Weight selection strategies for ordered weighted average based fuzzy rough sets. Inf Sci 501:155–171
35. Wan ZC, Song J, Shen YL (2018) Variable intuitionistic fuzzy multi-granulation rough set model and its approximate distribution reduction algorithms. J Comput Appl 38(2):390–398
36. Wang YB, Chen XJ, Dong K (2019) Attribute reduction via local conditional entropy. Int J Mach Learn Cyber 10:3619–3634
37. Wang C, Hu Q, Wang X et al (2018) Feature selection based on neighborhood discrimination index. IEEE Trans Neural Netw Learn Syst 29(7):2986–2999
38. Wang CZ, Huang Y, Shao MW, Hu QH, Chen DG (2020) Feature selection based on neighborhood self-information. IEEE Trans Cybern 50(9):4031–4042
39. Wang C, Huang Y, Shao M et al (2019) Fuzzy rough Set-Based attribute reduction using distance measures. Knowl-Based Syst 164:205–212
40. Wang Q, Qian Y, Liang X et al (2018) Local neighborhood rough set. Knowl-Based Syst 153:53–64
41. Wang CZ, Shi YP, Fan XD, Shao MW (2019) Attribute reduction based on K-Nearest neighborhood rough sets. Int J Approx Reason 106:18–31
42. Xu WH, Li WL (2016) Granular computing approach to two-way learning based on formal concept analysis in fuzzy datasets. IEEE Trans Cybern 46(2):366–379
43. Xu WH, Yu JH (2020) A novel approach to information fusion in multi-source datasets: a granular computing viewpoint. Inf Sci 378:410–423
44. Yang X, Li TR, Liu D et al (2020) A multilevel neighborhood sequential decision approach of three-way granular computing. Inf Sci 538:119–141
45. Yang XB, Liang SC, Yu HL, Gao S, Qian YH (2019) Pseudo-label neighborhood rough set: measures and attribute reductions. Int J Approx Reason 105:112–129
46. Yang L, Qin KY, Sang BB, Xu WH (2021) Dynamic fuzzy neighborhood rough set approach for interval-valued information systems with fuzzy decision. Appl Soft Comput:111
47. Yang L, Xu WH, Zhang XY, Sang BB (2020) Multi-granulation method for information fusion in multi-source decision information system. Int J Approx Reason 122:47–65
48. Zhang X, Mei C, Chen D, Yang Y, Li J (2019) Active incremental feature selection using a Fuzzy-Rough-Set-Based information entropy. IEEE Trans. Fuzzy Syst. 28(5):901–915
49. Zhang X, Mei C, Chen D et al (2016) Feature selection in mixed data: a method using a novel fuzzy rough set-based information entropy. Pattern Recognit 56:1–15
50. Zhao HD (2012) The geometric average sorting method of interval numbers and its application inner Mongolia university for nationalities
51. Zhou P, Hu XG, Li PP, Wu XD (2019) Online streaming feature selection using adapted neighborhood rough set. Inf Sci 481:258–279

Xiaoyan Zhang received the B.Sc. degree in mathematics from Yanbei Normal University, Datong, China, the M.Sc. degree in mathematics from Guangxi University, Nanning, China, and the PhD degree in mathematics from Northwest University, Xi'an China, in 2001, 2005 and 2017, respectively. She is currently an Associate Professor of College of Artificial Intelligence, Southwest University, Chongqing, China. She has published over 40 articles in international journals and book chapters. Her current research interests include mathematical foundations of artificial intelligence, fuzzy sets, granular computing, knowledge discovery, etc.

Zongying Jiang is currently pursuing the B.Sc. degree from the College of Artificial Intelligence, Southwest University, Chongqing, China. His current research interests include fuzzy sets, granular computing.

Weihua Xu received the Ph.D. degree from School of Sciences, Xi'an Jiaotong University, Xi'an, China, in 2007, and the M.Sc. degree from School of Mathematics and Information Sciences, Guangxi University, Nanning, China, in 2004. He is currently a professor with the College of Artificial Intelligence, Southwest University, Chongqing, China. He serves in the editorial boards of several international journals, and has published 4 monographs and over 140 articles in international journals. His current research interests include granular computing, approximate reasoning, fuzzy sets, data mining, knowledge discovery, and cognitive computing.