Interval Dominance-Based Feature Selection for Interval-Valued Ordered Data

Wentao Li, Haoxiang Zhou, Weihua Xu, Xi-Zhao Wang, *Fellow, IEEE*, and Witold Pedrycz, *Life Fellow, IEEE*

***Abstract*— Dominance-based rough approximation discovers inconsistencies from ordered criteria and satisfies the requirement of the dominance principle between single-valued domains of condition attributes and decision classes. When the ordered decision system (ODS) is no longer single-valued, how to utilize the dominance principle to deal with multivalued ordered data is a promising research direction, and it is the most challenging step to design a feature selection algorithm in interval-valued ODS (IV-ODS). In this article, we first present novel thresholds of interval dominance degree (IDD) and interval overlap degree (IOD) between interval values to make the dominance principle applicable to an IV-ODS, and then, the interval-valued dominance relation in the IV-ODS is constructed by utilizing the above two developed parameters. Based on the proposed interval-valued dominance relation, the interval-valued dominance-based rough set approach (IV-DRSA) and their corresponding properties are investigated. Moreover, the interval dominance-based feature selection rules based on IV-DRSA are provided, and the relevant algorithms for deriving the interval-valued dominance relation and the feature selection methods are established in IV-ODS. To illustrate the effectiveness of the parameters variation on feature selection rules, experimental evaluation is performed using 12 datasets coming from the University of California-Irvine (UCI) repository.**

***Index Terms*— Dominance-based rough set, feature selection, interval value, ordered information system (OIS), rough approximation.**

Nomenclature

OIS	Ordered information system.
ODS	Ordered decision system.
IV-OIS	Interval-valued ordered information system.
IV-ODS	Interval-valued ordered decision system.
IDD	Interval dominance degree.
IOD	Interval overlap degree.
DRSA	Dominance-based rough set approach.
IV-DRSA	Interval-valued dominance-based rough set approach.
IV-UB	Upper boundary for interval value.
IV-MV	Mean value for interval value.
RSoIV	Comparison method in [29].
GDI	Comparison method in [5].

Manuscript received 28 October 2021; revised 3 May 2022; accepted 12 June 2022. Date of publication 23 June 2022; date of current version 6 October 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 61976245, in part by the Scientific and Technological Project of Construction of Double City Economic Circle in Chengdu-Chongqing Area under Grant KJCX2020009, and in part by the Science and Technology Research Program of the Chongqing Education Commission under Grant KJQN202100205 and Grant KJQN202100206. *(Corresponding author: Weihua Xu.)*

Wentao Li is with the College of Artificial Intelligence, Southwest University, Chongqing 400715, China, and also with the School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China (e-mail: drliwentao@gmail.com).

Haoxiang Zhou and Weihua Xu are with the College of Artificial Intelligence, Southwest University, Chongqing 400715, China (e-mail: rickzhoustr@gmail.com; chxuwh@gmail.com).

Xi-Zhao Wang is with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: xizhaowang@ieee.org).

Witold Pedrycz is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6R 2V4, Canada (e-mail: wpedrycz@ualberta.ca).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNNLS.2022.3184120.

Digital Object Identifier 10.1109/TNNLS.2022.3184120

I. Introduction

THE rough set theory, proposed by Pawlak [1], is an effective formal tool to handle imprecision, vagueness, and uncertainty in data analysis. This relatively new soft computing methodology has produced successful applications in many areas of science and engineering, such as pattern recognition [2], [3], data mining [4]–[6], image processing [7], and medical diagnosis [8]–[10]. The Pawlak rough set exhibits two characteristics in real-life applications: 1) the data type is classic, that is, each object in the information system has only one definite value with regard to each attribute and 2) the required binary relation is classic, that is, it is an equivalence relation. These two aspects lead to two important research directions of rough set theory. Existing extended rough set models can be roughly cast into two perspectives: 1) extending the data type, including fuzzy data [11]–[15], incomplete data [6], [16]–[18], set-valued data [19], interval-valued data [20]–[24], and so forth and 2) extending the binary relation, including similarity relation [19], tolerance relation [25], dominance relation [26]–[31], and others.

As a special data type different from the classic, interval value is a common data type in the fields of approximate reasoning, signal processing, and control. The traditional dataset in machine learning and data mining is a single real-valued, which is used to represent just one exact value, while each value in interval-valued data is expressed as an interval, which means a range. Interval-valued data have been applied to characterize the imprecise and uncertain situations in real applications. Compared with other types of data, interval-valued data have unique advantages and practical significance. For example, interval values can be used in cases where measurement inaccuracies are sometimes inevitable, and then, the result can be represented as an interval value. Besides, interval values are also useful when the values of a particular attribute are continuous, which cannot be met by other data types at the same time. Let us take the temperature as another example. The temperature in Chongqing on a certain day in April is 12 °C to 20 °C. Using real value cannot directly express that, so we always use an average temperature as a substitute. In that case, we can say that the average temperature is 16 °C. On the other hand, we can also use an interval value to precisely express the variation range, which is expressed as $[12\ ^\circ\mathrm{C}, 20\ ^\circ\mathrm{C}]$, and it can reflect the facts better than the average temperature 16 °C. Evidently, an interval value contains more information compared with a single value. This is one of the reasons why it is necessary to study interval-valued data.

Meanwhile, the binary relations used in a specific information system are no longer equivalence relations but preference relations, such as dominance relations. We call this kind of information system as OIS [32]. When the decision attribute also exhibits preference order in the OIS, it is defined as an ODS. It is necessary to propose an extension called the DRSA to take into account the ordering properties of the criteria. The innovation is mainly based on the substitution of the equivalence relation (indiscernibility relation) in an OIS by a dominance relation. Since Greco *et al.* [28] initially studied DRSA, many scholars have investigated a variety of rough set models based on dominance relation to solve different problems. Among these results, only a few involve research of IV-OIS [11], [24], [33]. The main reason for less research on IV-OIS is that the order relationship between interval values in IV-OIS is complicated, and there is some information loss between the interval-valued order relationship obtained after processing and the original interval value, that is, the interval value order relationship obtained after processing cannot fully reflect the order characteristics between interval values. As implied in [34], the error bound generated by the imprecise intervals and fuzzy data may affect the effectiveness of the classification for the IV-OIS. Therefore, it is vital to thoroughly explore the order relationship of interval values, and it is a promising research direction to put forward an interval dominance relation to deal with arbitrary two interval-valued data.

Feature selection is to select important features (or remove redundant features) from complex datasets to achieve the purpose of reducing dimensionality, so as to save the storage space of data and the time of data analysis. Depending on the operating principle, feature selection algorithms can broadly be divided into two categories, including the wrapper model and the filter model [35]. The wrapper model utilizes a learning algorithm and applies its performance to evaluate features. On the contrary, the filter approach applies significance criteria instead of a learning algorithm to select features. Significance criteria include information gain [7], consistency [36], dependence [37], [38], distance [39], and so on. The significance measures of features and selection criteria are important in feature selection, which influences the effectiveness and classification of the reduction sets. In the rough set theory, the importance of features depends on the approximate accuracy of the rough set, and the size of lower and upper approximations directly determines the accuracy of rough set models; thus, scholars have proposed various significance measures for features based on the approximate operators, such as maximal discernibility pairs, dependency, and inner and outer importance measures. Therefore, different feature selection algorithms have been derived from the above significance measure functions [40]–[46]. Feature selection methods based on DRSA have been extensively studied in the past decades, and they are used to deal with classical ordered datasets [47]–[50]. Although these methods can effectively remove redundant features from ordered data, they ignore the characteristics of interval-valued ordered data, which are more widely used and more reasonable existed than classical ordered data in real-life applications. Accordingly, an effective and efficient feature selection method is urgently requested to process interval-valued ordered data.

Based on the above analysis, in this article, we define the notions of IDD and IOD to conveniently describe the dominance relation between any interval values and then establish a general dominance-based rough set model to deal with IV-ODS and further explore the method of feature selection rules. The main contents and innovation of this article could be summarized in the following three aspects.

1) A novel dominance relation suitable for dealing with interval values is proposed based on the defined IDD and IOD. After discussing its important properties, the IV-DRSA models are constructed.
2) Rules of interval dominance-based feature selection for IV-ODS are presented, and the related algorithms for calculating interval-valued dominance relation and feature selection are derived in IV-ODS.
3) The experimental evaluation is performed using 12 public available datasets. The relevant experiment is designed to explore the optimal values of the two thresholds (IDD and IOD), and the superiority of interval dominance-based feature selection is shown by the analysis of experimental results.

This article is organized as follows. Some necessary and important concepts about DRSA are introduced in Section II, and the motivation for this article is described in detail. In Section III, we first present the notions of IDD and IOD, and the concept of interval-valued dominance relation is provided. Then, the concrete definitions of upper and lower approximations are further obtained, and the IV-DRSA is constructed accordingly. Moreover, we derived the reduction rules in IV-OIS by using the inner and outer significances. In Section IV, we mainly design two related algorithms to compute the proposed interval dominance relation and calculate the feature selection. In Section V, the corresponding experimental testing is conducted by 12 datasets from the University of California-Irvine (UCI) datasets to test the advantage of interval dominance-based feature selection methods for IV-ODS, and the preferable thresholds of IDD and IOD are further debated. Finally, Section VI covers some conclusions.

II. Related Work and Foundations

In this section, the basic notions of the DRSA model and IV-ODS are reviewed, and the motivation is depicted with a concrete and comprehensive analysis. Let us first explain the necessary symbolic notations listed in the Nomenclature.

Definition 1: An information system is a tuple $I = (U, \mathrm{AT}, V, f)$, where the following holds.

1) $U = \{x_1, x_2, x_3, \ldots, x_n\}$ is a nonempty and finite set of objects.
2) $\mathrm{AT} = \{a_1, a_2, \ldots, a_m\}$ is a nonempty and finite set of condition attributes, and $\mathrm{AT} \cap d = \emptyset$, where d is called the decision attribute set.
3) $V = V_{\mathrm{AT}} = \bigcup_{a_l \in \mathrm{AT}} V_{a_l}$ is the domain of attribute set AT, where V_{a_l} is the domain of a_l.
4) $f = \{f_{a_l} | U \to V_{a_l}, a_l \in \mathrm{AT}\}$, where f_{a_l} is the value of a_l on $x \in U$.

A decision information system is an information system $(U, \mathrm{AT} \cup d, V, f)$, where $\mathrm{AT} \cap d = \emptyset$, AT is the condition attribute set, while d is called the decision attribute set. The conditional attribute set AT contains important information about the characteristics of all aspects of the sample, and the decision attribute set d represents the classification label of the samples. V_d represents the domain of the decision attribute set d. Generally, samples with the same value of decision attributes are considered as the same category.

In a decision information system, if the domain of an attribute is ordered according to a decreasing or increasing preference, then the attribute is a criterion. A decision information system is called an ODS if all condition attributes are criteria.

Definition 2: Let $I = (U, \mathrm{AT} \cup d, V, f)$ be an ODS with decision attribute d, and A is a subset of AT, namely, $A \in \mathrm{AT}$. The dominance relation with regard to A can be defined as

$$D_{R_A} = \left\{(x, y) \in U^2 | f(x, a) \geq f(y, a) \quad \forall a \in A\right\}.$$

In the ODS, $\succeq_a$ is defined to denote the preference-ordered relation on condition attribute a. Therefore, $\forall a \in A$, if $x \succeq_a y$, that means x dominates y in the attribution set A. We can denote this as $x D_{R_A} y$. The set of all the objects that are dominating x on attributes $\forall a \in A$, namely, A-dominating set, is defined as

$$D^+_{R_A} = \{y \in U | y \succeq_A x\}.$$

Similarly, the set of all the objects that are dominated by x on attributes $\forall a \in A$, namely, A-dominated set, is defined as

$$D^-_{R_A} = \{y \in U | x \succeq_A y\}.$$

Based on the difference of the decision attributes' value, the universe of discourse U can be divided into a finite number of classes $\mathrm{Cl} = \{\mathrm{Cl}_t, t \in T\}$, $T = \{1, 2, 3, \ldots, n\}$. Each $x \in U$ belongs to one and only one decision class Cl_t. To each decision attribute value v_{d_t}, $\mathrm{Cl}_t = \{x \in U | f(x, d) = v_{d_t}\}$. Assume that these decision classes are ordered, that is, for $t, s \in T$, if $t \leq s$, then all the objects from Cl_s are preferred to any object from Cl_t. Then, the upward unions $\mathrm{Cl}^{\succeq}_t$ and downward unions $\mathrm{Cl}^{\preceq}_t$ of the decision classes are obtained as follows:

$$\mathrm{Cl}^{\succeq}_t = \bigcup_{s \geq t} \mathrm{Cl}_s$$
$$\mathrm{Cl}^{\preceq}_t = \bigcup_{s \leq t} \mathrm{Cl}_s, \quad t \in T.$$

The above upward unions and downward unions are the sets to be approximated in the DRSA.

Definition 3: Let $I = (U, \mathrm{AT} \cup d, V, f)$ be an ODS and $A \subseteq \mathrm{AT}$. Then, the lower and upper approximations of $\mathrm{Cl}^{\succeq}_t (t \in T)$ are defined as follows:

$$\underline{R_A}\left(\mathrm{Cl}^{\succeq}_t\right) = \left\{x \in U | D^+_A(x) \subseteq \mathrm{Cl}^{\succeq}_t\right\}$$
$$\overline{R_A}\left(\mathrm{Cl}^{\succeq}_t\right) = \left\{x \in U | D^-_A(x) \cap \mathrm{Cl}^{\succeq}_t \neq \varnothing\right\}.$$

The lower and upper approximations of $\mathrm{Cl}^{\preceq}_t (t \in T)$ also can be defined as

$$\underline{R_A}\left(\mathrm{Cl}^{\preceq}_t\right) = \left\{x \in U | D^-_A(x) \subseteq \mathrm{Cl}^{\preceq}_t\right\}$$
$$\overline{R_A}\left(\mathrm{Cl}^{\preceq}_t\right) = \left\{x \in U | D^+_A(x) \cap \mathrm{Cl}^{\preceq}_t \neq \varnothing\right\}.$$

Once the lower and upper approximations are obtained, the positive region, the negative region, and the boundary region of $\mathrm{Cl}^{\succeq}_t$ and $\mathrm{Cl}^{\preceq}_t$ could be derived, which are

$$\mathrm{Pos}_A\left(\mathrm{Cl}^{\succeq}_t\right) = \underline{R_A}\left(\mathrm{Cl}^{\succeq}_t\right)$$
$$\mathrm{Neg}_A\left(\mathrm{Cl}^{\succeq}_t\right) = U - \overline{R_A}\left(\mathrm{Cl}^{\succeq}_t\right)$$
$$Bn_A\left(\mathrm{Cl}^{\succeq}_t\right) = \overline{R_A}\left(\mathrm{Cl}^{\succeq}_t\right) - \underline{R_A}\left(\mathrm{Cl}^{\succeq}_t\right);$$
$$\mathrm{Pos}_A\left(\mathrm{Cl}^{\preceq}_t\right) = \underline{R_A}\left(\mathrm{Cl}^{\preceq}_t\right)$$
$$\mathrm{Neg}_A\left(\mathrm{Cl}^{\preceq}_t\right) = U - \overline{R_A}\left(\mathrm{Cl}^{\preceq}_t\right)$$
$$Bn_A\left(\mathrm{Cl}^{\preceq}_t\right) = \overline{R_A}\left(\mathrm{Cl}^{\preceq}_t\right) - \underline{R_A}\left(\mathrm{Cl}^{\preceq}_t\right).$$

In the following, we introduce IV-ODS and the method of comparison of two interval values. An interval value is denoted as $u = [a, b]$, in which $a, b \in \mathbb{R}$ and $a < b$ always holds. As a matter of fact, the situation of $a > b$ can also be converted to $u' = [b, a]$, which satisfies the above definition. a is the lower boundary of interval value u, and b is the upper boundary of u. For $u = [a, b]$ and $v = [c, d]$, $u = v$ holds iff $a = c$ and $b = d$. An interval value is actually a family of continuous values and can be intuitively expressed as a region on the axis of reals. Furthermore, we define the set of all interval values as $\mathbb{IV} = \{u | u = [a, b], a, b \in \mathbb{R}\}$.

An IV-ODS is defined as $I = (U, \mathrm{AT} \cup d, V, f)$, in which $U = \{x_1, x_2, \ldots, x_n\}$ is nonempty finite set of objects; $\mathrm{AT} = \{a_1, a_2, \ldots, a_m\}$ is a nonempty finite sets of attributes or features, and $d = \{d_1, d_2, \ldots, d_k\}$ is the set of decision attributes; $V = V_{\mathrm{AT}} \cup V_d = (\bigcup_{a_i \in \mathrm{AT}} V_{a_i}) \cup V_d$, where V_{a_i} is the domain of conditional attributes a_i, V_d is domain of decision attributes d, and $V_{a_i} \in \mathbb{IV}$; and $f : U \times (\mathrm{AT} \cup d) \to V$ is a mapping from $U \times (\mathrm{AT} \cup d)$ to V and obviously $f(x, a_i) \in V_{a_i}$.

The difference between IV-ODS and classical ODS is that the former has all attributes with interval values, while the latter has all attributes with a single real value. Nowadays, there are many studies on the classical ODS in the field of rough sets; however, there are relatively few results on IV-ODS. The reason why less research on IV-ODS existed is that there is no reasonable approach to deal with the interval-valued order relationship. This also leads to the absence of appropriate references for studying if we want to deal with IV-ODSs, and there is no suitable rough decision model for IV-ODS.

In many real-life circumstances, data types are no longer single-valued, but multivalued, such as interval-valued. Compared with the single-valued data type, the first challenge of feature selection for the interval-valued data type is how to find an appropriate approach to deal with interval-valued data. The second challenge is that few studies can reasonably and comprehensively compare the size of interval values, that is, there are few reasonable studies on the ordered relationship between interval values. The existing research on how to deal with interval values could be roughly summarized into two aspects [11], [20]–[22], [24], [33], [42], [51], [52]. One of them, denoted as IV-UB, is using the upper boundary to replace it. The other one, denoted as IV-MV, is replacing the interval value with its mean value of two boundaries. For example, for an interval value $u = [2, 4]$, it was substituted by $a = 4$ and $b = 3$ in IV-UB and IV-MV, respectively.

Although these two methods can simply transform interval values into real values, there are disadvantages to both methods. Let us introduce the shortcomings as follows. In some cases, two different interval values finally turn out to be equal, which is unreasonable. For IV-UB, when two interval values have the same upper boundary, then they are considered to be the same. For instance, if $u = [3, 5]$ and $\upsilon = [2, 5]$, they both are replaced by the real value $a = 5$. As for IV-MV if two different interval values have the same mean value, then they are also regarded as the same. For example, if $u = [2, 6]$ and $u = [3, 5]$, they are replaced with $b = 4$, which is apparently improper.

Moreover, a single value cannot involve whole information contained in interval values. Sometimes, the loss of information makes a great difference, which means that both IV-UB and IV-MV are imprecise and cannot be properly used to deal with IV-ODS. We need to find a more effective method to compare the size of two interval values.

Meanwhile, two interval values may not have intersecting parts or partially intersecting parts. In the case of intersection, it may even occur that one interval value completely contains another interval value. To establish the partial ordered relation between two interval values, let us consider the relative position of two interval values on the number axis. Let $u = [a, b]$ and $\upsilon = [c, d]$ be two interval values in which $a < b$ and $c < d$ always hold. As shown in Fig. 1, A is the part of interval value u that is less than υ; the region denoted as B represents the common part of u and υ; and the region C is the part of interval value u that is bigger than υ. For different interval values, their relative positions on the number axis are different. First, let us consider a special case. When the length of B is zero, that means any value of υ is preferred to all values in u. In terms of the interval values themselves, we can say that υ is completely preferred to u. For a common situation where the length of B is greater than zero, we can use the proportion of the length of B in two interval values to describe the degree of one interval value that dominates the other one. The larger the proportion of B in the two interval values, the less obvious the dominance relationship between them. By comparing Fig. 1(a) and (b), we know that the smaller the region B, the clearer the partial order relationship between u and υ. From this point, we can introduce proper parameters to define the interval dominance relation.

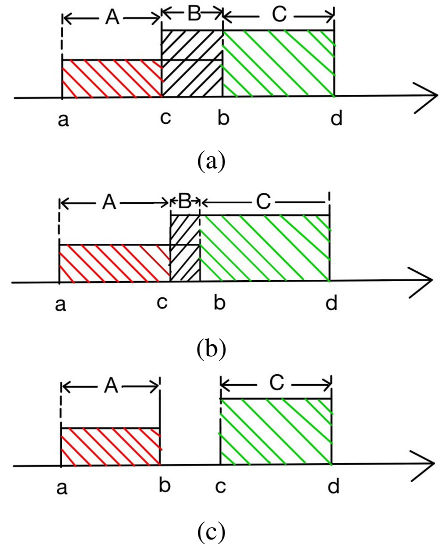

Fig. 1. Different cases on two interval values. (a) Large common part between u and υ. (b) Small common part between u and υ. (c) No common part between u and υ.

III. Interval-Valued Dominance-Based Rough Set Approach

The DRSA model can be successfully used to deal with a single real-valued information system. However, this model cannot be applied to IV-ODS directly. In this section, we first define the dominance relation between two interval values on basis of the definition of IDD and IOD. Then, we expand the DRSA model from the single real-valued situation to the interval-valued information system and determine the IV-DRSA model.

A. Interval Dominance Degree

The dominance relation for the single real-valued data is defined directly by comparing the size of the value. Nevertheless, the interval values cannot be compared directly. However, for special interval values, we can compare their sizes. For example, we have two interval values $u = [2, 3]$ and $\upsilon = [3, 4]$. It is obvious that an arbitrary value of υ is larger than any specific value of u. Here, we construct a quantitative index to describe the dominance relation between interval values, especially for the situation in which interval values have intersecting parts and cannot be compared.

Consider two interval values $u = [a, b]$ and $\upsilon = [c, d]$, in which $a < b$ and $c < d$ always hold.

Definition 4: The IDD, denoted as $g(u, \upsilon)$, is defined as

$$g(u, \upsilon) = \begin{cases} 1, & a < b \leq c < d \\ & or\ a = c < d = b \\ -1, & c < d \leq a < b \\ \dfrac{d - b}{\max\{b, d\} - \max\{a, c\}}, & \text{else.} \end{cases}$$

In order to construct the partial order in IV-OIS, antisymmetry is a basic property that must be satisfied. For two interval values, we discuss the important properties of IDD as follows.

Theorem 1: Given two interval values $u = [a, b]$ and $\upsilon = [c, d]$, then we have the following.

1) $-1 \leq g(u, v) \leq 1$.
2) $g(u, v) = -g(v, u)$ iff $a \neq c$ and $b \neq d$.

Proof:

1) From Definition 4, if $d \leq a$ or $a = c < d = b$, then $g(u, v) = -1$, and if $b \leq c$, then $g(u, v) = -1$. In other cases, $|d - b| < |\max\{b, d\} - \max\{a, c\}|$ always holds. Thus, $-1 \leq g(u, v) \leq 1$ holds.
2) It is obvious that, $g(u, v) = 1$ if $a = c$ and $b = d$. If $c < d \leq a < b$, then $g(u, v) = -1$ and $g(v, u) = 1$, that is, $g(u, v) = -g(v, u)$. In other cases, it is easy to figure out $g(u, v) = -g(v, u)$.

□

From Theorem 1, the IDD of an interval value on another interval value ranges from -1 to 1. Specifically, if the dominance degree of interval value u concerning v ranges from -1 to 0 and $g(u, v) \leq \alpha$ ($\alpha < 0$) holds, that means u is dominating v with the level α. On the contrary, if the dominance degree of u concerning v ranges from 0 to 1, and $g(u, v) \geq \alpha$ ($\alpha > 0$) holds, then we say that u is dominated by v with the level α. Generally, one interval value is not always completely dominating another one. By utilizing IDD, we can precisely describe the degree of one interval value that dominates another.

B. Interval Overlap Degree

The IDD describes the dominance relation on the one hand. From another perspective, we always expect as little overlap as possible between two interval values because the situation of two completely nonoverlapping interval values means absolute dominance relation. Meanwhile, sometimes, two different groups of interval values can have the same IDD. For example, the IDDs of [3, 5], [4, 6] and [2, 5], [4, 6] have the same IDD 0.5, while the overlap of latter interval values is obviously less than the former. Then, we give the definition of the IOD, which describes the level of overlap between two interval values.

Definition 5: Let $l_1 = b-a$ and $l_2 = d-c$. l_1 and l_2 are the lengths of interval values u and v, respectively. For possible cases, the IOD, denoted as $h(u, v)$, is defined as

$$h(u, v) = \begin{cases} 0, & d \leq a \text{ or } b \leq c \\ & \text{or } a = c < b = d \\ \dfrac{2(\min\{b, d\} - \max\{a, c\})}{l_1 + l_2}, & \text{else.} \end{cases}$$

Similarly, in order to construct the partial order in IV-OIS, we use the following theorem to prove the antisymmetry of interval-valued dominance relation.

Theorem 2: For two interval values $u = [a, b]$ and $v = [c, d]$, we have the following properties.

1) $0 \leq h(u, v) \leq 1$.
2) $h(u, v) = h(v, u)$.

Proof:

1) From Definition 5, if $d \leq a$ or $b \leq c$, $h(u, v) = 0$, while, in other cases, $0 \leq 2(\min\{b, d\} - \max\{a, c\}) \leq l_1 + l_2$ always holds; hence, $0 \leq h(u, v) \leq 1$.
2) It can be easily derived from Definition 5.

□

From Theorem 2, the IOD of two interval values ranges from 0 to 1. The smaller the IOD is, the more obvious the difference between two interval values. With the IOD, we can precisely describe the degree of one interval value that overlaps another, which is also an important indicator to measure the dominance relation between two interval values.

C. Interval-Valued Dominance Relation

Now, any two interval values are comparable by using two proposed parameters, namely, IDD and IOD. In an IV-ODS, by giving the constraint conditions of two objects on a specific attribute set, we can define the interval-valued dominance relation.

Definition 6: Let $I = (U, \text{AT} \cup d, V, f)$ be an IV-ODS. $x, y \in U$, and A is a subset of conditional attribute AT. The interval-valued dominance relation on attribute set A is defined as

$$D_A^{\alpha,\beta} = \left\{(x, y) \in U^2 | g(u, v) \geq \alpha \wedge h(u, v) \leq \beta\right\}$$

where $u = f(x, a)$ and $v = f(y, a)$.

For the object y, $u = f(x, a)$, and $v = f(y, a)$, if $g(u, v) \geq \alpha \wedge h(u, v) \leq \beta$ holds for any $a \in A$, then objects x and y satisfy the dominance relation at the level α and β. We can also denote it as $y D_A^{\alpha,\beta} x$.

In the following, we discuss some properties of interval dominance relation, which are important for the feature selection algorithm and save a lot of computational time.

Theorem 3: Let $I = (U, \text{AT} \cup d, V, f)$ be an IV-ODS, and $D_A^{\alpha,\beta}$ is an interval-based dominance relation with thresholds α, β $(0 < \alpha < 1)$; $\forall x, y \in U$, we have the following.

1) *Reflexivity:* $(x, x) \in D_A^{\alpha,\beta}$.
2) *Antisymmetry:* $(x, y) \in D_A^{\alpha,\beta} \Rightarrow (y, x) \notin D_A^{\alpha,\beta}$.
3) *Transitivity:* $(x, y) \in D_A^{\alpha,\beta}, (y, z) \in D_A^{\alpha,\beta} \Rightarrow (x, z) \in D_A^{\alpha,\beta}$.

Proof: Let $u = f(x, a)$, $v = f(y, a)$, and $w = f(z, a)$ be different interval values and $a \in A$. From Definitions 4 and 5, we can derive that $g(u, u) = 1 > \alpha$ and $h(u, u) = 0 < \beta$ hold; then, the reflexivity is proven. If $(x, y) \in D_A^{\alpha,\beta}$, then we have $g(u, v) \geq \alpha$ and $h(u, v) \leq \beta$. Notice that $g(u, v) = -g(v, u)$ and $h(u, v) = h(v, u)$, so $g(v, u) \geq \alpha$ and $h(v, u) \leq \beta$ do not hold. Then, antisymmetry is proven. As for the proof of transitivity, from $(x, y) \in D_A^{\alpha,\beta}, (y, z) \in D_A^{\alpha,\beta}$, we can obtain that $g(u, v) \geq \alpha$ and $h(u, v) \leq \beta$, and $g(v, w) \geq \alpha$ and $h(v, w) \leq \beta$. Apparently, we can derive $g(u, w) \geq \alpha$ and $h(u, w) \leq \beta$, that is, $(x, z) \in D_A^{\alpha,\beta}$. □

It is not difficult to observe that the complete dominance relation, namely, $\alpha = 1$ and $\beta = 0$, may be a little strict. Thus, we can specify the thresholds of IDD and IOD as α and β in which $0 < \alpha < 1$ and $0 < \beta < 1$, respectively. On the basis of this, we can define the set of objects that dominate x on the attribute set A called the A-dominating set.

Definition 7: Let $I = (U, \text{AT} \cup d, V, f)$ be an IV-ODS. $x, y \in U$, and $A \subseteq \text{AT}$. The A-dominating set is defined as

$$D_A^{\alpha,\beta,+}(x) = \left\{y \in U | y D_A^{\alpha,\beta} x\right\}.$$

TABLE I
INTERVAL-VALUED INFORMATION SYSTEM

U	a_1	a_2	a_3
x_1	[0.2,1.5]	[2.1,4.7]	[1.2,3.3]
x_2	[1.2,2.5]	[0.8,1.6]	[2.1,3.7]
x_3	[0.5,1.1]	[0.6,2.7]	[0.7,4.2]
x_4	[4.0,4.5]	[2.7,3.6]	[1.3,1.8]
x_5	[2.0,2.6]	[1.1,2.8]	[2.9,4.1]

Similarly, the set of objects, which is dominated by x on the attribute set A called the A-dominated set, is given as

$$D_A^{\alpha,\beta,-}(y) = \left\{y \in U | x D_A^{\alpha,\beta} y\right\}.$$

From Definitions 4 and 5, we always set $0 < \alpha < 1$ in the A-dominating set $D_A^{\alpha,\beta,+}(x)$, while $-1 < \alpha < 0$ in the A-dominated set $D_A^{\alpha,\beta,-}(x)$.

For any two different objects x and y, it is easy to find that the IDDs of two objects are opposite numbers on a specific attribute. On the other hand, the IOD of two objects on a given attribute is just the same. Therefore, we use two matrices to denote IDD and IOD of all the objects.

Theorem 4: Given an IV-ODS $I = (U, \mathrm{AT} \cup d, V, f)$ and two thresholds α and β, let $a_k \in \mathrm{AT}$. The matrices of IDD and IOD are derived as

$$M_{Da_k} = [g(u, v)]_{i\times j}$$
$$M_{Oa_k} = [h(u, v)]_{i\times j}$$

where $g(u, v)$ and $h(u, v)$ are the IDD and IOD in which $u = f(x, a_k)$ and $v = f(y, a_k)$.

For the conditional attribute $a \in \mathrm{AT}$, we specify that the element in row i and column j represents the IDD (IOD) of the object x_j with respect to object x_i in the matrix. From Theorems 1 and 2, we obtain that the matrix of IDD is an antisymmetric matrix, and IOD is a symmetric matrix.

Example 1: Table I shows an interval-valued information system $I = (U, A, V, f)$, in which $U = \{x_1, x_2, \ldots, x_5\}$ is the set of objects and $A = \{a_1, a_2, a_3\}$ is the attributes set. In this part, we specify two thresholds as $\alpha = 0.5$ and $\beta = 0.25$. For any $a_t \in A$, we can, respectively, obtain a matrix of IDD denoted as M_{Da_t} and a matrix of IOD denoted as M_{Oa_t}. For example, in terms of information system in Table I, for the attribute a_1 we have

$$M_{Da_1} = \begin{bmatrix} 1 & 0.77 & -0.4 & 1 & 1 \\ -0.77 & 1 & -1 & 1 & 0.17 \\ 0.4 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & 1 & -1 \\ -1 & -0.17 & -1 & 1 & 1 \end{bmatrix}$$
$$M_{Oa_1} = \begin{bmatrix} 0 & 0.23 & 0.63 & 0 & 0 \\ 0.23 & 0 & 0 & 0 & 0.53 \\ 0.63 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0.53 & 0 & 0 & 0 \end{bmatrix}.$$

For any $a_t \in A$, we can figure out M_{Da_t} and M_{Oa_t}. As shown in above, all elements on the main diagonal of M_{Da_t} are 1, and all elements on the main diagonal of M_{Oa_t} are 0.

The thresholds of IDD and IOD make a great influence on interval dominance relation, which is associated with classification, so it is important to explore the influence of the parameter size on the IV-DRSA model. Thus, in Theorems 5 and 6, we will discuss the difference between the two thresholds of IDD and IOD.

Theorem 5: Given an IV-ODS $I = (U, \mathrm{AT} \cup d, V, f)$ and two sets of thresholds α_1, α_2 and β_1, β_2, which satisfy $\alpha_2 \le \alpha_1$, $\beta_1 \le \beta_2$, and $A \subseteq \mathrm{AT}$, then the following properties holds.

1) $0 < \alpha_2 \le \alpha_1 < 1 : D_A^{\alpha_1,\beta_1,+}(x) \subseteq D_A^{\alpha_2,\beta_2,+}(x)$.
2) $-1 < \alpha_2 \le \alpha_1 < 0 : D_A^{\alpha_1,\beta_1,-}(x) \subseteq D_A^{\alpha_2\beta_2,-}(x)$.

Proof:

1) For any $y \in D_A^{\alpha_1,\beta_1,+}(x)$, $g(u, v) \ge \alpha_1 \ge \alpha_2$ and $h(u, v) \le \beta_1 \le \beta_2$ hold on any $a \in A$ in which $u = f(x, a)$, $v = f(x, a)$. From Definition 6, we can easily know that $y \in D_A^{\alpha_2,\beta_2,+}(x)$; thus, $D_A^{\alpha_1,\beta_1,+}(x) \subseteq D_A^{\alpha_2,\beta_2,+}(x)$.
2) For any $y \in D_A^{\alpha_1,\beta_1,-}(x)$, $g(u, v) \le \alpha_1 \le \alpha_2$ and $h(u, v) \le \beta_1 \le \beta_2$ hold on $\forall a \in A$ in which $u = f(x, a)$, $v = f(t, a)$. From Definition 6, we can derive that $y \in D_A^{\alpha_2,\beta_2,-}(x)$; thus, $D_A^{\alpha_1,\beta_1,-}(x) \subseteq D_A^{\alpha_2\beta_2,-}(x)$.

□

With different ranges of threshold, the relationships between A-dominating sets and A-dominated sets are revealed. When $-1 < \alpha < 0$ is satisfied, a bigger threshold divides the set of the object into smaller A-dominated sets. When $0 < \alpha < 1$, smaller A-dominating sets are obtained by a bigger threshold. The value of the threshold directly decides the size of A-dominating sets and A-dominated sets if the other conditions are totally the same, which means that we can judge the belonging relationship between two A-dominating sets (or A-dominated sets) simply by comparing the size of thresholds. Thus, from Theorem 5, we can derive that the sizes of IDD and IOD actually limit the size of A-dominating sets and A-dominated sets. In the case of the A-dominating set, larger values of thresholds α and β mean stricter limitations. Different values of two thresholds cause different partitions on the same set of objects and then influence classification. The appropriate values of α and β will be debated later.

If, $\forall a \in A$, the IDD of y on x is no less than threshold α $(0 < \alpha < 1)$, and IOD of y on x is no more than threshold β, then we say that object y dominates x at the levels α and β.

In the process of feature selection, the condition attribute set may change over time. Theorem 6 is about the influence of attribute set variation on the IV-DRSA model when α and β are the same.

Theorem 6: Given an IV-ODS $I = (U, \mathrm{AT} \cup d, V, f)$, $A_1 \subseteq A_2 \subseteq \mathrm{AT}$. $\mathrm{Cl}_t^{\succeq}$ and $\mathrm{Cl}_t^{\preceq}$ are the upward union and the downward union of the decision classes, respectively. The following properties hold:

$$D_{A_1}^{\alpha,\beta,+} \supseteq D_{A_2}^{\alpha,\beta,+}, D_{A_1}^{\alpha,\beta,-} \supseteq D_{A_2}^{\alpha,\beta,-}.$$

Proof: From Definition 6 and Theorem 5, for any $(x, y) \in D_{A_2}^{\alpha,\beta}$, we can know that $g(u, v) \geq \alpha \wedge h(u, v) \leq \beta$ holds with regard to $\forall a \in A_2$, where $u = f(x, a)$ and $v = f(y, a)$. From $A_1 \in A_2$, then, $\forall a \in A_1$, there is $a \in A_2$. Therefore, $g(u, v) \geq \alpha \wedge h(u, v) \leq \beta$ holds with regard to $\forall a \in A_1$, that is, $(x, y) \in D_{A_1}^{\alpha,\beta,+}$. In other words, for any $(x, y) \in D_{A_2}^{\alpha,\beta,+}$, $(x, y) \in D_{A_1}^{\alpha,\beta,+}$ holds, and then, $D_{A_1}^{\alpha,\beta,+} \supseteq D_{A_2}^{\alpha,\beta,+}$ is proven. The proof of $D_{A_1}^{\alpha,\beta,-} \supseteq D_{A_2}^{\alpha,\beta,-}$ is similar to the proof process above. □

Definition 8: The interval-valued dominance-based rough approximation of the upward union $\mathrm{Cl}_t^{\succeq}$ and the downward union $\mathrm{Cl}_t^{\preceq}$ is defined as

$$\underline{R_A^{\alpha,\beta}}(\mathrm{Cl}_t^{\succeq}) = \left\{x \in U | D_A^{\alpha,\beta,+}(x) \subseteq \mathrm{Cl}_t^{\succeq}\right\}$$
$$\overline{R_A^{\alpha,\beta}}(\mathrm{Cl}_t^{\succeq}) = \left\{x \in U | D_A^{\alpha,\beta,-}(x) \cap \mathrm{Cl}_t^{\succeq} \neq \varnothing\right\}.$$

The interval-valued dominance-based rough approximations of downward union $\mathrm{Cl}_t^{\preceq}$ are defined as

$$\underline{R_A^{\alpha,\beta}}(\mathrm{Cl}_t^{\preceq}) = \left\{x \in U | D_A^{\alpha,\beta,-}(x) \subseteq \mathrm{Cl}_t^{\preceq}\right\}$$
$$\overline{R_A^{\alpha,\beta}}(\mathrm{Cl}_t^{\preceq}) = \left\{x \in U | D_A^{\alpha,\beta,+}(x) \cap \mathrm{Cl}_t^{\preceq} \neq \varnothing\right\}.$$

One of the most important properties of utilizing a heuristic algorithm in feature selection is monotonicity. In Theorems 7 and 8, we provide the important monotonicity of the IV-DRSA model.

Theorem 7: Let $I = (U, \mathrm{AT} \cup d, V, f)$ be an IV-ODS; $\mathrm{Cl}_t^{\succeq}$ and $\mathrm{Cl}_t^{\preceq}$ are the upward union and the downward union of the decision classes, respectively, and A is a subset of AT. α and β are IDD and IOD; then, we have the following.

1) $\underline{R_A^{\alpha,\beta}}(\mathrm{Cl}_t^{\succeq}) \subseteq \mathrm{Cl}_t^{\succeq} \subseteq \overline{R_A^{\alpha,\beta}}(\mathrm{Cl}_t^{\succeq})$.
2) $\underline{R_A^{\alpha,\beta}}(\mathrm{Cl}_t^{\preceq}) \subseteq \mathrm{Cl}_t^{\preceq} \subseteq \overline{R_A^{\alpha,\beta}}(\mathrm{Cl}_t^{\preceq})$.

Proof:

1) From Definition 8, for any $x \in \underline{R_A^{\alpha,\beta}}(\mathrm{Cl}_t^{\succeq})$, we have $x \in \mathrm{Cl}_t^{\succeq}$. Hence, $\underline{R_A^{\alpha,\beta}}(\mathrm{Cl}_t^{\succeq}) \subseteq \mathrm{Cl}_t^{\succeq}$. $\forall x \in \mathrm{Cl}_t^{\succeq}$, it is obvious that $x \in D_A^{\alpha,\beta,-}(x)$; hence, $\mathrm{Cl}_t^{\succeq} \in \overline{R_A^{\alpha,\beta}}(\mathrm{Cl}_t^{\succeq})$.
2) The proof of (2) is similar to the proof of (1). □

The positive region, the negative region, and the boundary region of $\mathrm{Cl}_t^{\succeq}$ are shown as

$$\begin{aligned}\mathrm{Pos}_A^{\alpha,\beta}(\mathrm{Cl}_t^{\succeq}) &= \underline{R_A^{\alpha,\beta}}(\mathrm{Cl}_t^{\succeq}) \\ &= \left\{x \in U | D_A^{\alpha,\beta,+}(x) \subseteq \mathrm{Cl}_t^{\succeq}\right\} \\ \mathrm{Neg}_A^{\alpha,\beta}(\mathrm{Cl}_t^{\succeq}) &= U - \overline{R_A^{\alpha,\beta}}(\mathrm{Cl}_t^{\succeq}) \\ &= \left\{x \in U | D_A^{\alpha,\beta,-}(x) \cap \mathrm{Cl}_t^{\succeq} = \varnothing\right\} \\ \mathrm{Bnd}_A^{\alpha,\beta}(\mathrm{Cl}_t^{\succeq}) &= \overline{R_A^{\alpha,\beta}}(\mathrm{Cl}_t^{\succeq}) - \underline{R_A^{\alpha,\beta}}(\mathrm{Cl}_t^{\succeq}).\end{aligned}$$

The three-way decision rules of the upward union $\mathrm{Cl}_t^{\succeq}$ are given as follows.

P: If $x \in \mathrm{Cl}_t^{\succeq}$ and $D_A^{\alpha,\beta}(x) \subseteq \mathrm{Cl}_t^{\succeq}$, decide $\mathrm{Pos}_A^{\alpha,\beta}(\mathrm{Cl}_t^{\succeq})$.

N: If $x \in \mathrm{Cl}_t^{\succeq}$ and $D_A^{\alpha,\beta}(x) \cap \mathrm{Cl}_t^{\succeq} \neq \varnothing$, then decide $\mathrm{Neg}_A^{\alpha,\beta}(\mathrm{Cl}_t^{\succeq})$.

B: Else, decide $\mathrm{Bnd}_A^{\alpha,\beta}(\mathrm{Cl}_t^{\succeq})$.

The positive region, the negative region, and the boundary region of $\mathrm{Cl}_t^{\preceq}$ are

$$\begin{aligned}\mathrm{Pos}_A^{\alpha,\beta}(\mathrm{Cl}_t^{\preceq}) &= \underline{R_A^{\alpha,\beta}}(\mathrm{Cl}_t^{\preceq}) \\ &= \left\{x \in U | D_A^{\alpha,\beta,-}(x) \subseteq \mathrm{Cl}_t^{\preceq}\right\} \\ \mathrm{Neg}_A^{\alpha,\beta}(\mathrm{Cl}_t^{\preceq}) &= U - \overline{R_A^{\alpha,\beta}}(\mathrm{Cl}_t^{\preceq}) \\ &= \left\{x \in U | D_A^{\alpha,\beta,+}(x) \cap \mathrm{Cl}_t^{\preceq} = \varnothing\right\} \\ \mathrm{Bnd}_A^{\alpha,\beta}(\mathrm{Cl}_t^{\preceq}) &= \overline{R_A^{\alpha,\beta}}(\mathrm{Cl}_t^{\preceq}) - \underline{R_A^{\alpha,\beta}}(\mathrm{Cl}_t^{\preceq}).\end{aligned}$$

Theorem 8: Given an IV-ODS $I = (U, \mathrm{AT} \cup d, V, f)$, $A_1 \subseteq A_2 \subseteq \mathrm{AT}$. $\mathrm{Cl}_t^{\succeq}$ and $\mathrm{Cl}_t^{\preceq}$ are the upward union and the downward union of the decision class, respectively. The following properties hold.

1) $\overline{R_{A_1}^{\alpha,\beta}}(\mathrm{Cl}_t^{\succeq}) \supseteq \overline{R_{A_2}^{\alpha,\beta}}(\mathrm{Cl}_t^{\succeq})$, $\underline{R_{A_1}^{\alpha,\beta}}(\mathrm{Cl}_t^{\succeq}) \supseteq \underline{R_{A_2}^{\alpha,\beta}}(\mathrm{Cl}_t^{\succeq})$, $\overline{R_{A_1}^{\alpha,\beta}}(\mathrm{Cl}_t^{\preceq}) \supseteq \overline{R_{A_2}^{\alpha,\beta}}(\mathrm{Cl}_t^{\preceq})$, and $\underline{R_{A_1}^{\alpha,\beta}}(\mathrm{Cl}_t^{\preceq}) \supseteq \underline{R_{A_2}^{\alpha,\beta}}(\mathrm{Cl}_t^{\preceq})$.
2) $\mathrm{Pos}_{A_1}^{\alpha,\beta}(\mathrm{Cl}_t^{\succeq}) \subseteq \mathrm{Pos}_{A_2}^{\alpha,\beta}(\mathrm{Cl}_t^{\succeq})$ and $\mathrm{Pos}_{A_1}^{\alpha,\beta}(\mathrm{Cl}_t^{\preceq}) \subseteq \mathrm{Pos}_{A_2}^{\alpha,\beta}(\mathrm{Cl}_t^{\preceq})$.

Proof:

1) Based on Theorem 6 and $A_1 \subseteq A_2$, we can directly get $D_{A_1}^{\alpha,\beta,+} \supseteq D_{A_2}^{\alpha,\beta,+}$. If $x \in \overline{R_{A_2}^{\alpha,\beta}}(\mathrm{Cl}_t^{\succeq})$, then $D_{A_2}^{\alpha,\beta,+}(x) \cap \mathrm{Cl}_t^{\succeq} \neq \varnothing$. Thus, $D_{A_1}^{\alpha,\beta,+}(x) \cap \mathrm{Cl}_t^{\succeq} \neq \varnothing$ also holds, and $\overline{R_{A_1}^{\alpha,\beta}}(\mathrm{Cl}_t^{\succeq}) \supseteq \overline{R_{A_2}^{\alpha,\beta}}(\mathrm{Cl}_t^{\succeq})$ is proven.
Similarly, $\underline{R_{A_1}^{\alpha,\beta}}(\mathrm{Cl}_t^{\succeq}) \supseteq \underline{R_{A_2}^{\alpha,\beta}}(\mathrm{Cl}_t^{\succeq})$, $\overline{R_{A_1}^{\alpha,\beta}}(\mathrm{Cl}_t^{\preceq}) \supseteq \overline{R_{A_2}^{\alpha,\beta}}(\mathrm{Cl}_t^{\preceq})$, and $\underline{R_{A_1}^{\alpha,\beta}}(\mathrm{Cl}_t^{\preceq}) \supseteq \underline{R_{A_2}^{\alpha,\beta}}(\mathrm{Cl}_t^{\preceq})$ can also be proved.
2) According to item 1), we have $\underline{R_{A_1}^{\alpha,\beta}}(\mathrm{Cl}_t^{\succeq}) \supseteq \underline{R_{A_2}^{\alpha,\beta}}(\mathrm{Cl}_t^{\succeq})$. It is obvious that $\mathrm{Pos}_{A_1}^{\alpha,\beta}(\mathrm{Cl}_t^{\succeq}) \subseteq \mathrm{Pos}_{A_2}^{\alpha,\beta}(\mathrm{Cl}_t^{\succeq})$ and $\mathrm{Pos}_{A_1}^{\alpha,\beta}(\mathrm{Cl}_t^{\preceq}) \subseteq \mathrm{Pos}_{A_2}^{\alpha,\beta}(\mathrm{Cl}_t^{\preceq})$. Then, the proof process of the above property is completed.

□

The three-way decision rules of the upward union $\mathrm{Cl}_t^{\succeq}$ are given as follows.

P: If $x \in \mathrm{Cl}_t^{\preceq}$ and $D_A^{\alpha,\beta,+}(x) \subseteq \mathrm{Cl}_t^{\succeq}$, decide $\mathrm{Pos}_A^{\alpha,\beta}(\mathrm{Cl}_t^{\preceq})$.

N: If $x \in \mathrm{Cl}_t^{\preceq}$ and $D_A^{\alpha,\beta,-}(x) \cap \mathrm{Cl}_t^{\preceq} \neq \varnothing$, then decide $\mathrm{Neg}_A^{\alpha,\beta}(\mathrm{Cl}_t^{\preceq})$.

B: Else, decide $Bn_A^{\alpha,\beta}(\mathrm{Cl}_t^{\preceq})$.

Example 2: Table II is about an IV-ODS, in which $U = \{x_1, x_2, \ldots, x_{10}\}$ and $\mathrm{AT} = \{a_1, a_2, \ldots, a_5\}$.

The decision attribute has three possible values: 1, 2, and 3. Considering the preference order, we can divide the subject set divided into three parts as

$$\mathrm{Cl}_1 = \{x_1, x_8, x_{10}\}$$
$$\mathrm{Cl}_2 = \{x_2, x_5, x_7, x_9\}$$
$$\mathrm{Cl}_3 = \{x_3, x_4, x_6\}.$$

Then, the upward union and the downward union can be obtained as

$$\mathrm{Cl}_2^{\succeq} = \mathrm{Cl}_2 \cup \mathrm{Cl}_3 = \{x_2, x_3, x_4, x_5, x_6, x_7, x_9\}$$
$$\mathrm{Cl}_3^{\succeq} = \mathrm{Cl}_3 = \{x_3, x_4, x_6\}$$
$$\mathrm{Cl}_1^{\preceq} = \mathrm{Cl}_1 = \{x_1, x_8, x_9\}$$
$$\mathrm{Cl}_2^{\preceq} = \mathrm{Cl}_1 \cup \mathrm{Cl}_2 = \{x_1, x_2, x_5, x_7, x_8, x_9, x_{10}\}.$$

TABLE II
INTERVAL-VALUED DECISION TABLE

U	a_1	a_2	a_3	a_4	a_5	d
x_1	[1,3]	[1.3,2.1]	[2,3]	[2.3,4]	[2,4.1]	1
x_2	[3,4]	[5.3,6.2]	[3,5]	[4,6]	[6,8]	2
x_3	[5,7]	[5.8,6.5]	[5,6]	[7.1,7.5]	[9,10]	3
x_4	[7.2,10.2]	[6.2,8]	[6.1,9]	[6,8.2]	[7.6,9.3]	3
x_5	[2,5]	[3.1,3.7]	[2,4]	[1,2]	[3,5.2]	2
x_6	[5.2,7.1]	[7,9]	[4.9,7.6]	[6,9]	[7,9]	3
x_7	[2,4]	[3,4]	[1.5,2.6]	[2.3,3.4]	[1,3.2]	2
x_8	[1.2,1.8]	[1.5,2.3]	[2.3,4.1]	[2.1,3.2]	[0.8,4.5]	1
x_9	[2,3]	[3,3.5]	[2.5,3.1]	[3,4]	[2.4,4.1]	2
x_{10}	[2,5]	[1,4]	[2,5]	[1,2.1]	[1,3]	1

TABLE III
A-DOMINATING SETS IN EXAMPLE 2

U	$D_A^{\alpha,\beta,+}$	U	$D_A^{\alpha,\beta,+}$
x_1	$\{x_1,x_2,x_3,x_4,x_6\}$	x_6	$\{x_6\}$
x_2	$\{x_2,x_4\}$	x_7	$\{x_3,x_4,x_6,x_7\}$
x_3	$\{x_3\}$	x_8	$\{x_3,x_4,x_6,x_8\}$
x_4	$\{x_4\}$	x_9	$\{x_2,x_3,x_4,x_6,x_9\}$
x_5	$\{x_3,x_4,x_5,x_6\}$	x_{10}	$\{x_3,x_4,x_6,x_{10}\}$

We set the level as $\alpha = 0.5$ and $\beta = 0.25$. The A-dominating sets of each $x \in U$ are shown in Table III. Then, we calculate the lower and upper approximations of $\mathrm{Cl}_2^{\succeq}$ as

$$\underline{R_A^{\alpha,\beta}}(\mathrm{Cl}_2^{\succeq}) = \{x_2,x_3,x_4,x_5,x_6,x_7,x_9\}$$
$$\overline{R_A^{\alpha,\beta}}(\mathrm{Cl}_2^{\succeq}) = \{x_2,x_3,x_4,x_5,x_6,x_7,x_9\}.$$

The lower and upper approximations of $\mathrm{Cl}_3^{\succeq}$ are

$$\underline{R_A^{\alpha,\beta}}(\mathrm{Cl}_3^{\succeq}) = \{x_3,x_4,x_6\}$$
$$\overline{R_A^{\alpha,\beta}}(\mathrm{Cl}_3^{\succeq}) = \{x_3,x_4,x_6\}.$$

The lower and upper approximations of $\mathrm{Cl}_1^{\preceq}$ are

$$\underline{R_A^{\alpha,\beta}}(\mathrm{Cl}_1^{\preceq}) = \{x_1,x_8,x_9\}$$
$$\overline{R_A^{\alpha,\beta}}(\mathrm{Cl}_1^{\preceq}) = \{x_1,x_8,x_9\}.$$

The lower and upper approximations of $\mathrm{Cl}_2^{\preceq}$ are

$$\underline{R_A^{\alpha,\beta}}(\mathrm{Cl}_2^{\preceq}) = \{x_1,x_2,x_7,x_8,x_9,x_{10}\},$$
$$\overline{R_A^{\alpha,\beta}}(\mathrm{Cl}_2^{\preceq}) = \{x_1,x_2,x_5,x_7,x_8,x_9,x_{10}\}.$$

D. Significance Based on IV-ODS and Reduction Rules

In what follows, the dependency degree in IV-ODS is defined, and then, the inner and outer significances are obtained. By using inner and outer significances, the reduction rules are further constructed.

Definition 9: For an IV-ODS $I = (U, \mathrm{AT} \cup d, V, f)$, $\mathrm{Pos}_A^{\alpha,\beta}(\mathrm{Cl}_t^{\succeq})$ is the positive region of $\mathrm{Cl}_t^{\succeq}$. Let A be a subset of AT and $a \in \mathrm{AT}$. The dependency degrees of $\mathrm{Cl}_t^{\succeq}$ and $\mathrm{Cl}_t^{\preceq}$ are defined as

$$\sigma_A^{\alpha,\beta}(\mathrm{Cl}_t^{\succeq}) = \frac{|\mathrm{Pos}_A^{\alpha,\beta}(\mathrm{Cl}_t^{\succeq})|}{|U|}$$
$$\sigma_A^{\alpha,\beta}(\mathrm{Cl}_t^{\preceq}) = \frac{|\mathrm{Pos}_A^{\alpha,\beta}(\mathrm{Cl}_t^{\preceq})|}{|U|}.$$

The inner and outer significances of the attribute set A with regard to decision attribution d are defined as follows:

$$\mathrm{Sig}_{\mathrm{in}}^{\succeq}(a, A, \mathrm{Cl}_t^{\succeq}) = \sigma_A^{\alpha,\beta}(\mathrm{Cl}_t^{\succeq}) - \sigma_{A-\{a\}}^{\alpha,\beta}(\mathrm{Cl}_t^{\succeq})$$
$$\mathrm{Sig}_{\mathrm{in}}^{\preceq}(a, A, \mathrm{Cl}_t^{\succeq}) = \sigma_A^{\alpha,\beta}(\mathrm{Cl}_t^{\preceq}) - \sigma_{A-\{a\}}^{\alpha,\beta}(\mathrm{Cl}_t^{\preceq})$$
$$\mathrm{Sig}_{\mathrm{out}}^{\succeq}(a, A, \mathrm{Cl}_t^{\succeq}) = \sigma_{A\cup\{a\}}^{\alpha,\beta}(\mathrm{Cl}_t^{\succeq}) - \sigma_A^{\alpha,\beta}(\mathrm{Cl}_t^{\succeq})$$
$$\mathrm{Sig}_{\mathrm{out}}^{\preceq}(a, A, \mathrm{Cl}_t^{\succeq}) = \sigma_{A\cup\{a\}}^{\alpha,\beta}(\mathrm{Cl}_t^{\preceq}) - \sigma_A^{\alpha,\beta}(\mathrm{Cl}_t^{\preceq}).$$

It is obvious that both the inner and outer significances of an attribute set range from 0 to 1.

Definition 10: Given an IV-ODS $I = (U, \mathrm{AT} \cup d, V, f)$, let A be a subset of AT and $a \in A$. If $\sigma_{A-a}^{\alpha,\beta}(\mathrm{Cl}_t^{\succeq}) = \sigma_A^{\alpha,\beta}(\mathrm{Cl}_t^{\succeq})$, then we call a the redundant attribute of A with regard to $\mathrm{Cl}_t^{\succeq}$; on the contrary, a is called the necessary attribute of A with respect to $\mathrm{Cl}_t^{\succeq}$. In terms of the upward union $\mathrm{Cl}_t^{\succeq}$, we can obtain a reduct A of AT if the following conditions are satisfied.

1) $\sigma_{A-\{a\}}^{\alpha,\beta}(\mathrm{Cl}_t^{\succeq}) = \sigma_A^{\alpha,\beta}(\mathrm{Cl}_t^{\succeq})$.
2) $\forall a \in A$, $\sigma_{A-\{a\}}^{\alpha,\beta}(\mathrm{Cl}_t^{\succeq}) < \sigma_A^{\alpha,\beta}(\mathrm{Cl}_t^{\succeq})$.

Definition 10 actually provides a rule to gain one reduct by constantly deleting attributes from the conditional attribute set AT when item 1) is satisfied. Once an attribute is deleted from A, then recalculate the dependency degree, and compare it with the original dependency degree. When item 2) is satisfied, it means that A has been a reduct.

IV. ALGORITHM FOR INTERVAL DOMINANCE-BASED FEATURE SELECTION

In this section, two algorithms are designed to calculate interval-valued dominance relation and feature selection, as shown in Algorithms 1 and 2, respectively. Meanwhile, the detailed descriptions of Algorithms 1 and 2, and complexity analysis are addressed as follows.

For Algorithm 1, the IV-ODS and two thresholds need to be given in advance. The calculation result D is a set that stores the interval dominance sets of each object in universe U. Step 3 initializes a temporary list ls_1 to store calculation results. Step 5 initializes two temporary lists to record calculations in the second loop. In step 7, we use Definitions 4 and 5 to compute the IDD and IOD. Step 14 adds the results to $D_{\mathrm{AT}}^{\alpha,\beta,+}$. We can obtain the time complexity of Algorithm 1 to be $O(|U|^2 \times |\mathrm{AT}|)$.

As for Algorithm 2, the parameter red is initialized as an empty set to store the reduct. Steps 2 and 3 compute the interval-valued dominance relation $D_a^{\alpha,\beta}$ on each $a \in \mathrm{AT}$. The process of calculating reduction can be divided into two parts. First, we choose the attributes in which the inner significance degree $\mathrm{Sig}_{in}^{\succeq}(a, \mathrm{red}, \mathrm{Cl}_t^{\succeq})$ is greater than 0 and add these attributes into red. Then, $\forall a \in$ red, compute the outer

Algorithm 1 Calculate Interval-Valued Dominance Relation

Require: IV-ODS $I = (U, AT \cup d, V, f)$; thresholds α, β.
Ensure: The interval dominance relation $D_{AT}^{\alpha,\beta,+}$.

Set $D_{AT}^{\alpha,\beta,+} \leftarrow \varnothing$; //Initialize $D_{AT}^{\alpha,\beta,+}$ to an empty set;
for each $x \in U$ **do**
  Set $ls\_1 \leftarrow \varnothing$; //Initialize the temporary list ls_1;
  for each $y \in U$ **do**
    Set $ls\_2 \leftarrow \varnothing, ls\_3 \leftarrow \varnothing$; /*Initialize two temporary lists to record calculation results.*/
    for each $a \in AT$ **do**
      Compute $g(v, u)$ and $h(v, u)$ which $u = f(x, a)$ and $v = f(y, a)$ respectively by Definition 4 and Definition 5;
      Add $g(v, u)$ to ls_2; Add $h(v, u)$ to ls_3; //Record calculation results.
    end for
    if all elements in $ls\_2[i]$ and $ls\_3[i]$ satisfy that $ls\_2[i] \geq \alpha$ and $ls\_3[i] \leq \beta$ **then**
      Add y to ls_1;
    end if
  end for
  Add ls_1 to $D_{AT}^{\alpha,\beta,+}$; //Store calculation results.
end for
return $D_{AT}^{\alpha,\beta,+}$;

Algorithm 2 Feature Selection Based on IV-DRSA

Require: IV-ODS $I = (U, AT \cup d, V, f)$; thresholds α, β.
Ensure: One reduct red.

1: Set $red \leftarrow \varnothing$; //Initialize red to an empty set;
2: Compute the interval dominance relation $D_{AT}^{\alpha,\beta,+}$;
3: Compute upward union $Cl_t^{\succeq}$;
4: **while** $AT - red \neq \varnothing$ **do**
5:   **for** each $a \in AT - red$ **do**
6:     Compute $Sig_{out}^{\succeq}(a, red, Cl_t^{\succeq})$;
7:   **end for**
8:   Find a_k with maximum value of $Sig_{out}^{\succeq}(a_k, red, Cl_t^{\succeq})$;
9:   **if** $Sig_{out}^{\succeq}(a_k, red, Cl_t^{\succeq}) = 0$ **then**
10:     Break;
11:   **else**
12:     $red \leftarrow red \cup \{a_k\}$;
13:   **end if**
14: **end while**
15: **for** each $a \in red$ **do**
16:   Compute $Sig_{in}^{\succeq}(a, red, Cl_t^{\succeq})$;
17:   **if** $Sig_{in}^{\succeq}(a, red, Cl_t^{\succeq}) = 0$ **then**
18:     $red \leftarrow red - \{a_k\}$;
19:   **end if**
20: **end for**
21: **return** red;

significance degree $\mathrm{Sig}_{\mathrm{out}}^{\succeq}(a, \mathrm{red}, \mathrm{Cl}_t^{\succeq})$, and check whether a can be added into red. When $\sigma_{\mathrm{red}}^{\alpha,\beta}(\mathrm{Cl}_t^{\succeq}) = \sigma_{\mathrm{AT}}^{\alpha,\beta}(\mathrm{Cl}_t^{\succeq})$, the attribute set red is iterated into a reduct, and then, stop the algorithm and output results.

The complexity of Algorithm 2 is analyzed as follows. In steps 2 and 3, we compute the approximation to further obtain the positive region, and the time complexity is $O(|U|^2 \times |\mathrm{AT}|)$. Steps 4 and 5 are used to compute the upward union, and the time complexity of this part is $O(|U|)$. The time complexity of steps 6–11, 12–14, 15–20, and 21–25 are, respectively, $O(|U|^2 \times |\mathrm{AT}|^2)$, $O(|U|^2 \times |\mathrm{AT}|^2 + |\mathrm{AT}|)$, $O(1)$, and $O(1)$. In summary, the time complexity of Algorithm 2 is

$$O\left(|U|^2 \times |\mathrm{AT}|^2 + |\mathrm{AT}|\right) = O\left(|U|^2 \times |\mathrm{AT}|^2\right).$$

On the other hand, in Algorithm 2, the space required to store the information table is $O(|U| \times (|\mathrm{AT}| + |d|))$. In order to store the upward union, upper and lower approximation, boundary region, inner and outer significance degrees, and the feature selection, the spatial complexities are $O(|d| \times |U|)$, $O(|d| \times |U|^2)$, $O(|d| \times |U|^2)$, $O(2 \times |\mathrm{AT}|^2)$, and $O(|\mathrm{AT}|)$, respectively.

V. Experimental Analysis

In this section, the corresponding experiment evaluation is designed to verify the advantages of the proposed algorithms. Detailed information about the tested datasets is given in Table IV. All the 12 datasets are originally from UCI, and we further preprocess these datasets to obtain interval-valued data inspired by the statistical method provided in [21] and [53]. First, we compute one of the reductions on each dataset. We adopt tenfold cross-validation to train the datasets and attempt to find out the optimal value of thresholds α and β. There are several studies on classifiers directly for interval-valued datasets in the literature of clustering [21], [53]. From the clustering results reported, it is not hard to discover that all of these methods have a common limitation, namely, the clustering effect of the direct classification for interval-valued data is not very prominent. Thus, two classical classifiers, K-nearest neighbor (KNN, in this article $K = 3$) and support vector machine (SVM), are used to evaluate classification accuracy with the selected dataset, which is computed by the feature selection algorithm. The classification accuracies of KNN and SVM and the average classification accuracies are computed for the reduced testing data. The interval value plays an important part mainly in the process of feature selection because it decides the IV-DRSA model, but, when we apply two classifiers, the dataset is restored to the real value. By comparing the results of raw data, IV-UB, IV-MV, and two methods from the literature [11], [33], we analyze the effectiveness and superiority of the proposed IV-DRSA model. All algorithms are executed in Python 3.8 and run in a hardware environment with Inter[1] Core[2] i5-7200 CPU @ 2.50 and 2.70 GHz with 4-GB RAM.

A. Robustness of Preprocessed Data

The main idea of data preprocessing is presented as follows: taking the initial real value of the datasets as the center point and then randomly generating two proper biases as

[1]Registered trademark.
[2]Trademarked.

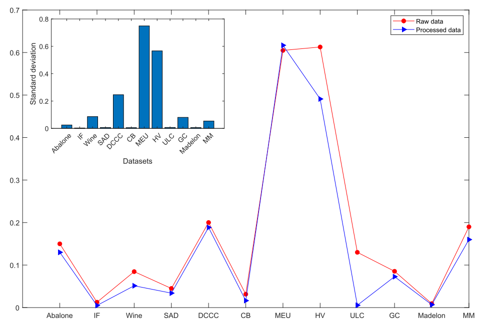

Fig. 2. Robustness of the processed data on each dataset.

δ_+ and δ_-. Denoting the original value in datasets as x^*, we take δ_+ and δ_- in the range of $0 < \delta_+(\delta_-) < 0.1 \cdot x^*$. Then, an interval value can be obtained as $[x^* - \delta_-, x^* + \delta_+]$. In this way, we successfully simulate the situation that, in reality, the attribute value cannot be given accurately. By adding a deviation to the original data, the change range of the data is expressed as an interval value, which has little impact on the distribution of the original data. Besides, the robustness of the preprocessed data is further evaluated by the definition of roughness. The roughness measure $\rho(D_A^{\alpha,\beta}, X) = 1 - (|\underline{R_A^{\alpha,\beta}}(X)|/|\overline{R_A^{\alpha,\beta}}(X)|)$ is utilized to verify the difference between original data and preprocessed data. We randomly obtained the interval-valued data by the previous method and further compute the roughness of the processed data. Then, we repeat the process ten times and compute the average roughness of processed data, which are used to compare with the original. Besides, the standard deviation of roughness represented by a histogram is used to reflect the change of roughness. From Fig. 2, we can intuitively observe that the preprocessed data exhibit better robustness than the original data, and the interval-valued preprocessing does not change the distribution of the original data.

B. Effectiveness Evaluations of Proposed Algorithm

In this section, the effectiveness of the feature selection algorithm based on IV-DRSA is validated. First, we find a relatively optimal threshold of α and β. Second, we apply the feature selection algorithm with optimal thresholds on interval-valued data to obtain a reduct. Finally, we use the selected feature to carry out the classification and calculate the classification on two classifiers and draw a certain conclusion about the experiment.

In fact, thresholds α and β decide the IV-DRSA model. In particular, the classification accuracies are expected to be currently lower when α is too small or too large. We use three interval values to illustrate how the different values of α and β affect the classification. As shown in Fig. 3, three interval values are represented as lines with different colors, and dotted lines indicate different classifications with different thresholds. For example, in Fig. 3(a)–(c), u_1, u_2, and u_3 are discernible with $\alpha = 0.3$ and $\beta = 0.4$. u_1 is dominated by u_2 and u_3 when $\alpha = 0.5$ and $\beta = 0.2$; however, between u_2 and u_3, there is no dominance relationship, which means that u_2 and u_3 are indiscernible, that is, because smaller value

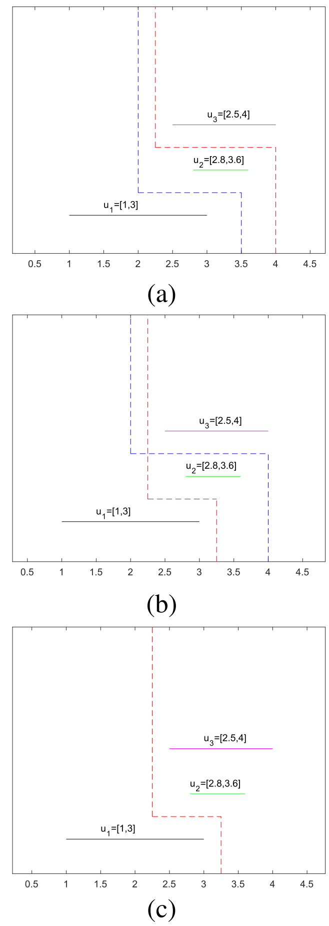

Fig. 3. Comparison between classifications derived by different thresholds. (a) Classification with thresholds $\alpha = 0.3$ and $\beta = 0.4$. (b) Classification with thresholds $\alpha = 0.4$ and $\beta = 0.3$. (c) Classification with thresholds $\alpha = 0.5$ and $\beta = 0.2$.

of α means more stricter limitation, while too large value of α may cause the difference between classes being not obvious. In terms of β, it also has an optimal value with the maximum classification accuracy. To find the optimal values, we iterate the values of α and β from 0.1 to 0.6 with a step 0.05 and then calculate one reduct of datasets and the classification accuracies of reduced data. The classification accuracies are shown in Fig. 4. By analyzing the change of classification accuracy with α and β, we can obtain the following conclusions.

1) Different datasets may have different optimal parameter values, and each dataset may also have multiple optimal values, which can be obtained by training on particular datasets. For example, the Wine dataset achieves the maximum classification accuracy with the thresholds of $\alpha = 0.5$ and $\beta = 0.2$, while, in the HV dataset, the optimal values of two thresholds are $\alpha = 0.4$ and $\beta = 0.55$. Meanwhile, $\alpha = 0.5$, $\beta = 0.25$ and $\alpha = 0.5$, $\beta = 0.2$ all are the optimal values of dataset GC. Overall, the optimal parameter values of 12 datasets in this article are close to $\alpha = 0.5$, $\beta = 0.25$.
2) In general, with the increase in threshold, the classification accuracy increases at the beginning and then

TABLE IV
DATA DESCRIPTION

No	Data set	Sample	Feature	Class
1	Abalone	4177	8	20
2	Internet Firewall (IF)	65532	12	4
3	Wine	178	13	3
4	Spoken Arabic Digit (SAD)	8800	13	10
5	Default of Credit Card Clients (DCCC)	30000	24	2
6	Connection Bench (CB)	208	60	2
7	MEU-Mobile KSD (MEU)	2856	71	56
8	Hill-Valley	606	101	2
9	Urban Land Cover (ULC)	168	148	9
10	Gait Classification (GC)	48	321	16
11	Madelon	4400	500	32
12	MicroMass (MM)	931	1300	4

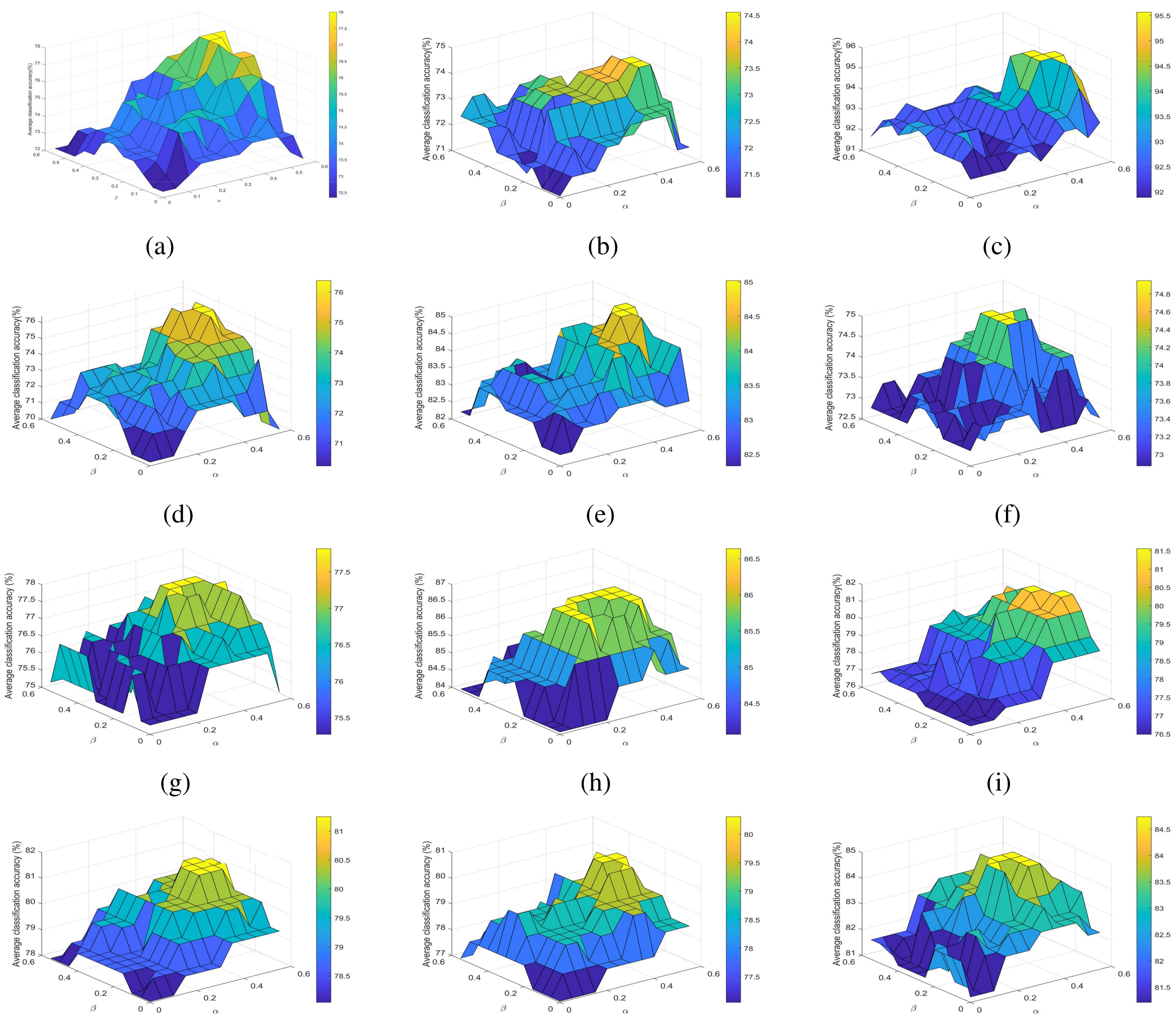

Fig. 4. Classification accuracies with threshold β on 12 datasets. (a) Abalone. (b) IF. (c) Wine. (d) SAD. (e) DCCC. (f) CB. (g) MEU. (h) HV. (i) ULC. (j) GC. (k) Madelon. (l) MM.

decreases. When two thresholds achieve a specific value, the classification then begins to decrease. This is in good agreement with the previous theoretical analysis. For example, in the dataset CB, when the value of α is kept unchanged, the classification accuracy first increases and then decreases when β changes from 0.1 to 0.6, and the maximum value is obtained near $\beta = 0.3$. For α, the change of classification accuracy conforms to

TABLE V
CLASSIFICATION ACCURACY OF RAW DATA AND REDUCED DATA WITH KNN (%)

Data set	Raw data	IV-UB	IV-MV	RSoIV	GDI	IV-DRSA
Abalone	75.33±3.68	74.00±5.62	76.00±4.72	71.62±4.39	76.28±4.82	**78.00±2.56**
IF	67.25±3.87	71.67±5.26	69.93±4.31	72.47±2.95	71.92±7.38	**74.56±5.21**
Wine	94.90±4.21	94.93±6.19	95.46±4.51	93.28±5.19	92.95±4.48	**95.57±4.98**
SAD	72.35±4.69	71.43±2.48	75.30±1.25	75.34±2.75	74.74±4.66	**76.38±6.32**
DCCC	80.71±6.25	79.54±4.26	81.45±7.21	80.35±4.26	82.33±3.59	**85.02±5.76**
CB	69.28±4.98	72.23±5.49	71.49±6.37	73.21±2.07	70.59±4.55	**74.75±7.10**
MEU	72.31±3.07	71.94±3.21	72.06±3.56	74.68±4.37	75.77±6.71	**77.73±4.61**
HV	80.49±7.01	81.49±9.69	80.70±6.03	84.58±3.58	85.35±5.77	**86.49±8.29**
ULC	81.44±4.49	82.79±4.74	84.85±6.43	82.39±1.82	84.78±3.27	**85.71±3.91**
GC	75.32±7.99	74.69±8.08	72.48±8.20	76.31±3.91	78.24±3.77	**80.39±5.21**
Madelon	77.64±3.49	79.57±5.61	81.45±2.97	80.35±3.74	80.04±6.32	**82.33±4.06**
MM	78.76±6.56	79.54±1.77	81.45±5.39	80.35±3.85	82.33±4.60	**83.68±5.76**
Average	77.15±5.02	77.82±5.20	78.55±5.08	78.74±3.57	79.61±4.99	**81.72±5.31**

TABLE VI
CLASSIFICATION ACCURACY OF RAW DATA AND REDUCED DATA WITH SVM (%)

Data set	Raw data	IV-UB	IV-MV	RSoIV	GDI	IV-DRSA
Abalone	74.59±2.14	72.46±1.08	77.33±2.10	74.42±6.27	75.38±3.72	**79.36±3.61**
IF	70.95±5.54	68.49±8.32	72.44±8.37	73.21±4.48	70.52±2.86	**73.49±4.04**
Wine	95.43±2.68	96.11±4.57	96.67±3.88	95.94±5.49	96.54±3.34	**96.87±5.81**
SAD	73.42±6.44	72.25±5.42	76.47±3.21	76.03±5.43	76.47±3.08	**76.79±4.26**
DCCC	79.46±5.47	80.10±3.66	81.86±2.74	81.09±3.59	83.47±2.88	**85.77±6.15**
CB	70.58±3.49	73.34±5.77	71.27±3.14	70.76±4.56	72.35±3.87	**75.15±7.56**
MEU	71.49±2.52	75.42±7.56	72.88±4.38	71.37±5.41	74.81±3.60	**78.91±3.43**
HV	79.06±3.48	76.67±8.43	78.05±4.24	86.38±3.09	84.21±4.16	**87.64±5.84**
ULC	80.44±2.31	81.43±4.16	82.86±5.23	84.37±5.41	81.48±4.33	**84.82±2.04**
GC	74.18±6.57	73.76±2.64	74.05±7.48	77.56±4.19	75.04±5.06	**81.25±4.22**
Madelon	75.79±5.47	79.54±3.22	81.45±5.41	80.35±3.56	81.47±4.15	**82.31±3.32**
MM	76.27±2.34	78.68±1.09	80.08±5.43	81.49±4.35	82.18±3.64	**84.74±2.77**
Average	76.81±4.04	77.35±4.67	78.78±4.63	79.16±4.65	75.49±3.72	**82.22±4.70**

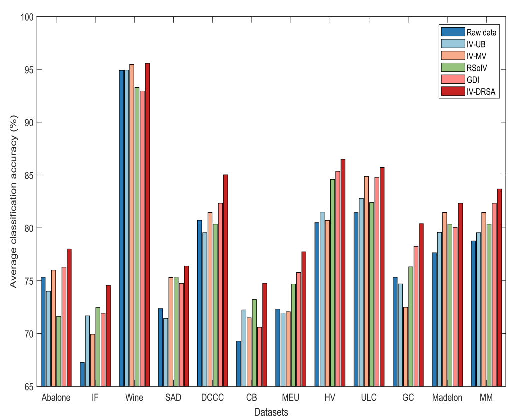

Fig. 5. Classification accuracy with KNN.

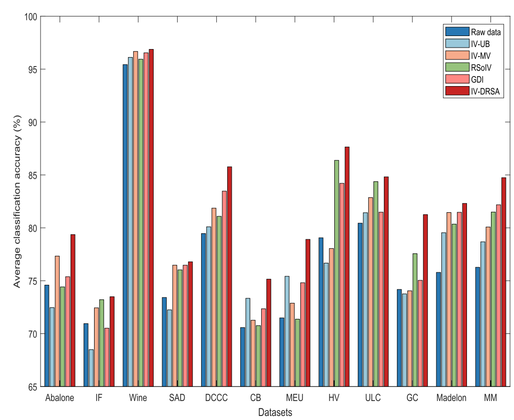

Fig. 6. Classification accuracy with SVM.

the same pattern. That is because too loose restrictions might increase the misclassification rate, while too strict restrictions may cause overfitting.

In order to demonstrate the advantage of Algorithm 2, we first obtain a reduct of datasets with the optimal thresholds $\alpha = 0.5$ and $\beta = 0.25$, which are trained in former experiments. Then, we apply KNN and SVM to classify the raw data, reduced data by IV-UB, IV-MV, the rough set model on the interval-valued information system (RSoIV) [33], the graded dominance interval (GDI)-based rough set approach [11], and IV-DRSA, respectively. Then, we compare the classification accuracies of the above methods. The results are shown in Tables V and VI and Figs. 5 and 6. In Tables V and VI, the highest classification accuracies between different methods are underlined and bold.

From a comparison of the results in Tables V and VI, we can obtain that, in most cases, the IV-DRSA model has the maximum classification accuracies compared with the other methods. Meanwhile, the average classification accuracy of IV-DRSA is also the highest. On the other hand, the IV-DRSA model has a better performance with the SVM classifier than KNN on datasets that we discussed. In the case of 12 datasets in this article, the average classification accuracy with KNN improves 5.68%, 4.77%, 3.84%, 3.54%, and 2.17% compared with raw data, IV-UB, IV-MV, RSoIV, and GDI, respectively. On the other hand, the average classification accuracy with SVM is improved by 7.04%, 6.30%, 4.37%, 3.49%, and 8.91% compared with results when using raw data, IV-UB, IV-MV, RSoIV, and GDI, respectively. In Figs. 5 and 6, the histogram is used to compare the classification accuracy of several methods more intuitively.

VI. Conclusion

As a more widely existed continuous data type, the interval value exhibits more extensive description and characterization ability compared with general data types, such as discrete dataset-valued data and fuzzy data. It is of great significance in real life. However, how to deal with interval-valued data reasonably has become a well-known challenge. To solve certain limitations of the classical rough set model with regard to ordered properties, DRSA was first presented by Greco *et al.* [28]. It is stated that DRSA could eliminate the inconsistencies between the decision attribute set and the condition attribute set in an OIS, which ensures that there is no inconsistency in the three-way decision regions obtained from the upper and lower approximations. This significant nature makes DRSA more widely applicable than indiscernibility relation-based rough set models. The main objective of this article is to build a general and novel IV-DRSA for an interval-valued environment by considering IDD and IOD between arbitrary interval values in IV-ODSs. The relative properties of the proposed interval-valued dominance relation are considered and verified through important theorems. Based on IV-DRSA, we discuss the feature selection methods for the IV-ODS and develop an effective algorithm to compute the reduction. In the experimental evaluation, we further discover the optimal values of two thresholds ensuring the relative higher classification accuracies. The generalization ability of the feature selection algorithm proposed in this article is mainly reflected by testing different datasets and studying the variance in the test results. This article mainly focuses on the basic construction of the IV-DRSA model and considers that thresholds of the IDD and IOD are the same values on all the attributes. However, the tolerance for different attributes could be different, which is more general and reasonable. In the future work, we will take into account the difference between attributes, focus on the fuzzy interval-valued classifier, investigate the classification suitable for the interval-valued data, and study the feature selection methods for heterogeneous interval-valued data with the ordered relationship.

Acknowledgment

The authors would like to thank the Associate Editor and the reviewers for their insightful comments and suggestions.

References

[1] Z. Pawlak, "Rough sets," *Int. J. Comput. Inf. Sci.*, vol. 11, no. 5, pp. 341–356, Oct. 1982.
[2] Z. Fang, J. Lu, F. Liu, J. Xuan, and G. Zhang, "Open set domain adaptation: Theoretical bound and algorithm," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 10, pp. 4309–4322, Oct. 2021.
[3] F. Liu, J. Lu, and G. Zhang, "Unsupervised heterogeneous domain adaptation via shared fuzzy equivalence relations," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 6, pp. 3555–3568, Dec. 2018.
[4] F. Liu, G. Zhang, and J. Lu, "Multisource heterogeneous unsupervised domain adaptation via fuzzy relation neural networks," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 11, pp. 3308–3322, Nov. 2021.
[5] F. Liu, G. Zhang, and J. Lu, "Heterogeneous domain adaptation: An unsupervised approach," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 12, pp. 5588–5602, Dec. 2020.
[6] C. L. P. Chen and C.-Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on big data," *Inf. Sci.*, vol. 275, no. 11, pp. 314–347, Aug. 2014.
[7] C. Lee and G. G. Lee, "Information gain and divergence-based feature selection for machine learning-based text categorization," *Inf. Process. Manage.*, vol. 42, no. 1, pp. 155–165, 2006.
[8] A. T. Azar, H. H. Inbarani, and K. R. Devi, "Improved dominance rough set-based classification system," *Neural Comput. Appl.*, vol. 28, no. 8, pp. 2231–2246, Aug. 2017.
[9] W. Kotłowski, K. Dembczyński, S. Greco, and R. Słowiński, "Stochastic dominance-based rough set model for ordinal classification," *Inf. Sci.*, vol. 178, no. 21, pp. 4019–4037, Nov. 2008.
[10] M. Kryszkiewicz, "Rough set approach to incomplete information systems," *Inf. Sci.*, vol. 112, nos. 1–4, pp. 39–49, 1998.
[11] B. Huang, "Graded dominance interval-based fuzzy objective information systems," *Knowl.-Based Syst.*, vol. 24, no. 7, pp. 1004–1012, Oct. 2011.
[12] A. Petrosino and G. Salvi, "Rough fuzzy set based scale space transforms and their use in image analysis," *Int. J. Approx. Reasoning*, vol. 41, no. 2, pp. 212–228, Feb. 2006.
[13] C. Wang *et al.*, "A fitting model for feature selection with fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 4, pp. 741–753, Aug. 2017.
[14] W.-H. Xu and W.-T. Li, "Granular computing approach to two-way learning based on formal concept analysis in fuzzy datasets," *IEEE Trans. Cybern.*, vol. 46, no. 2, pp. 366–379, Feb. 2016.
[15] J. Zhou, W. Pedrycz, C. Gao, Z. Lai, and X. Yue, "Principles for constructing three-way approximations of fuzzy sets: A comparative evaluation based on unsupervised learning," *Fuzzy Sets Syst.*, vol. 413, pp. 74–98, Jun. 2021.
[16] W. S. Du and B. Q. Hu, "Dominance-based rough set approach to incomplete ordered information systems," *Inf. Sci.*, vols. 346–347, pp. 106–129, Jun. 2016.
[17] M. Kryszkiewicz, "Rules in incomplete information systems," *Inf. Sci.*, vol. 113, nos. 3–4, pp. 271–292, Feb. 1999.
[18] X. Yang, J. Yang, C. Wu, and D. Yu, "Dominance-based rough set approach and knowledge reductions in incomplete ordered information system," *Inf. Sci.*, vol. 178, no. 4, pp. 1219–1234, Feb. 2008.
[19] S. Singh, S. Shreevastava, T. Som, and G. Somani, "A fuzzy similarity-based rough set approach for attribute selection in set-valued information systems," *Soft Comput.*, vol. 24, no. 6, pp. 4675–4691, Mar. 2020.
[20] M. B. Gorzałczany, "Interval-valued fuzzy controller based on verbal model of object," *Fuzzy Sets Syst.*, vol. 28, no. 1, pp. 45–53, Oct. 1988.
[21] Y. Leung, M. M. Fischer, W.-Z. Wu, and J.-S. Mi, "A rough set approach for the discovery of classification rules in interval-valued information systems," *Int. J. Approx. Reasoning*, vol. 47, no. 2, pp. 233–246, Feb. 2008.
[22] Y. Qian, J. Liang, and C. Dang, "Interval ordered information systems," *Comput. Math. Appl.*, vol. 56, no. 8, pp. 1994–2009, Oct. 2008.
[23] I. B. Turksen, "Interval valued fuzzy sets based on normal forms," *Fuzzy Sets Syst.*, vol. 20, no. 2, pp. 191–210, Oct. 1986.
[24] X. Yang and M. Zhang, "Dominance-based fuzzy rough approach to an interval-valued decision system," *Frontiers Comput. Sci. China*, vol. 5, no. 2, pp. 195–204, Jun. 2011.
[25] Y. Y. Yao, "Relational interpretations of neighborhood operators and rough set approximation operators," *Inf. Sci.*, vol. 111, nos. 1–4, pp. 239–259, 1998.

[26] H. Chen, T. Li, and D. Ruan, "Maintenance of approximations in incomplete ordered decision systems while attribute values coarsening or refining," *Knowl.-Based Syst.*, vol. 31, pp. 140–161, Jul. 2012.

[27] H. Chen, T. Li, Y. Cai, C. Luo, and H. Fujita, "Parallel attribute reduction in dominance-based neighborhood rough set," *Inf. Sci.*, vol. 373, pp. 351–368, Dec. 2016.

[28] S. Greco, B. Matarazzo, and R. Slowinski, "Rough approximation by dominance relations," *Int. J. Intell. Syst.*, vol. 17, no. 2, pp. 153–171, Feb. 2002.

[29] S. Greco, R. Slowinski, and Y. Yao, "Bayesian decision theory for dominance-based rough set approach," in *Proc. Int. Conf. Rough Sets Knowl. Technol.*, in Lecture Notes in Computer Science, 2007, pp. 134–141.

[30] M. Inuiguchi, Y. Yoshioka, and Y. Kusunoki, "Variable-precision dominance-based rough set approach and attribute reduction," *Int. J. Approx. Reasoning*, vol. 50, no. 8, pp. 1199–1214, Sep. 2009.

[31] W. Li, X. Xue, W. Xu, T. Zhan, and B. Fan, "Double-quantitative variable consistency dominance-based rough set approach," *Int. J. Approx. Reasoning*, vol. 124, pp. 1–26, Sep. 2020.

[32] W. Li, W. Xu, X. Zhang, and J. Zhang, "Updating approximations with dynamic objects based on local multigranulation rough sets in ordered information systems," *Artif. Intell. Rev.*, vol. 55, no. 3, pp. 1821–1855, Mar. 2022.

[33] B. Sun, W. Ma, and Z. Gong, "Dominance-based rough set theory over interval-valued information systems," *Expert Syst.*, vol. 31, no. 2, pp. 185–197, May 2014.

[34] G. Ma, F. Liu, G. Zhang, and J. Lu, "Learning from imprecise observations: An estimation error bound based on fuzzy random variables," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, Jul. 2021, pp. 1–8.

[35] W. Shu, W. Qian, and Y. Xie, "Incremental feature selection for dynamic hybrid data using neighborhood rough set," *Knowl.-Based Syst.*, vol. 194, Apr. 2020, Art. no. 105516.

[36] H. Uğuz, "A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm," *Knowl.-Based Syst.*, vol. 24, no. 7, pp. 1024–1032, Oct. 2011.

[37] M. S. Raza and U. Qamar, "An incremental dependency calculation technique for feature selection using rough sets," *Inf. Sci.*, vols. 343–344, pp. 41–65, May 2016.

[38] Y. Liu, F. Tang, and Z. Zeng, "Feature selection based on dependency margin," *IEEE Trans. Cybern.*, vol. 45, no. 6, pp. 1209–1221, Jun. 2015.

[39] K. Kira and L. A. Rendell, "The feature selection problem: Traditional methods and a new algorithm," in *Proc. AAAI*, vol. 92, 1992, pp. 129–134.

[40] J. Dai, H. Hu, W.-Z. Wu, Y. Qian, and D. Huang, "Maximal-discernibility-pair-based approach to attribute reduction in fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 4, pp. 2174–2187, Aug. 2018.

[41] L. Guan, "A heuristic algorithm of attribute reduction in incomplete ordered decision systems," *J. Intell. Fuzzy Syst.*, vol. 36, no. 4, pp. 3891–3901, Apr. 2019.

[42] B. Huang, D.-K. Wei, H.-X. Li, and Y.-L. Zhuang, "Using a rough set model to extract rules in dominance-based interval-valued intuitionistic fuzzy information systems," *Inf. Sci.*, vol. 221, pp. 215–229, Feb. 2013.

[43] E. C. C. Tsang, D. Chen, D. S. Yeung, X. Z. Wang, and J. W. T. Lee, "Attributes reduction using fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 5, pp. 1130–1141, Oct. 2008.

[44] C. Wang, Y. Huang, M. Shao, and X. Fan, "Fuzzy rough set-based attribute reduction using distance measures," *Knowl.-Based Syst.*, vol. 164, pp. 205–212, Jan. 2019.

[45] W. Xu, Y. Li, and X. Liao, "Approaches to attribute reductions based on rough set and matrix computation in inconsistent ordered information systems," *Knowl.-Based Syst.*, vol. 27, pp. 78–91, Mar. 2012.

[46] Y. Yang, D. Chen, H. Wang, and X. Wang, "Incremental perspective for feature selection based on fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 3, pp. 1257–1273, Jun. 2018.

[47] P. Zhou, P. Li, S. Zhao, and X. Wu, "Feature interaction for streaming feature selection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 10, pp. 4691–4702, Oct. 2021.

[48] R. Susmaga, "Reducts and constructs in classic and dominance-based rough sets approach," *Inf. Sci.*, vol. 271, pp. 45–64, Jul. 2014.

[49] W. Qian and W. Shu, "Attribute reduction in incomplete ordered information systems with fuzzy decision," *Appl. Soft Comput.*, vol. 73, pp. 242–253, Dec. 2018.

[50] W. S. Du and B. Q. Hu, "A fast heuristic attribute reduction approach to ordered decision systems," *Eur. J. Oper. Res.*, vol. 264, no. 2, pp. 440–452, Jan. 2018.

[51] N. Xie, M. Liu, Z. Li, and G. Zhang, "New measures of uncertainty for an interval-valued information system," *Inf. Sci.*, vol. 470, pp. 156–174, Jan. 2019.

[52] J. Dai, W. Wang, and J.-S. Mi, "Uncertainty measurement for interval-valued information systems," *Inf. Sci.*, vol. 251, pp. 63–78, Dec. 2013.

[53] L. V. Utkin, A. I. Chekh, and Y. A. Zhuk, "Binary classification SVM-based algorithms with interval-valued training data using triangular and Epanechnikov kernels," *Neural Netw.*, vol. 80, pp. 53–66, Aug. 2016.

Wentao Li received the M.Sc. degree from the School of Mathematics and Statistics, Chongqing University of Technology, Chongqing, China, in 2015, and the Ph.D. degree from the Department of Mathematics, Harbin Institute of Technology, Harbin, China, in 2019.

From 2016 to 2018, he was a Joint Ph.D. Student with the University of Alberta, Edmonton, AB, Canada. He is currently a Lecturer with the College of Artificial Intelligence, Southwest University, Chongqing. He has published over 30 articles in international journals. His current research interests include fuzzy sets, rough sets, and granular computing.

Haoxiang Zhou is currently pursuing the bachelor's degree with the College of Artificial Intelligence, Southwest University, Chongqing, China.

His main research interests include rough sets, feature selection, granular computing, and their applications.

Weihua Xu received the B.Sc. degree in mathematics from Yanbei Normal University, Datong, China, in 2001, the M.Sc. degree in mathematics from Guangxi University, Nanning, China, in 2004, and the Ph.D. degree in applied mathematics from Xi'an Jiaotong University, Xi'an China, in 2007.

He is currently a Professor with the College of Artificial Intelligence, Southwest University, Chongqing, China. He has published four monographs and more than 120 articles in international journals. His current research interests include approximate reasoning, fuzzy sets, rough sets, concept lattices, and cognitive concept learning.

Dr. Xu also serves as a PC Member of Granular Computing and Knowledge Discovering in China, National Conference on Intelligent Information Processing. He also serves on the editorial boards of *International Journal of Machine Learning and Cybernetics* and *Journal of Intelligent and Fuzzy Systems*.

Xi-Zhao Wang (Fellow, IEEE) received the B.Sc. degree in mathematics from Hebei University, Shijiazhuang, China, in 1983, the M.Sc. degree in computer science from Shanghai Jiao Tong University, Shanghai, China, in 1987, and the Ph.D. degree in computer science from the Harbin Institute of Technology, Harbin, China, in 1998.

He is currently a Professor of computer science and software engineering with the School of Computer and Software, Shenzhen University, Shenzhen, China. His main research interests include machine learning and uncertainty information processing, including inductive learning with fuzzy representation, approximate reasoning and expert systems, neural networks and their sensitivity analysis, statistical learning theory, fuzzy measures and fuzzy integrals, random weight network, and the recent topic: machine learning theories and methodologies in the big-data environment.

Dr. Wang is also a CAAI Fellow. He also serves on the Advisory Board of the IEEE TRANSACTIONS ON FUZZY SYSTEMS and is a member of a number of editorial boards of other international journals. He is also the Editor-in-Chief of *International Journal of Machine Learning and Cybernetics*, the Deputy Chair of the CAAI Machine Learning Committee and the Knowledge Engineering Committee, the Overseas High-Level (peacock B class) Talent of Shenzhen City, the Prizewinner of the First-Class Award of the Natural Science Advances of Hebei Province, and a Model Teacher of China.

Witold Pedrycz (Life Fellow, IEEE) received the M.Sc., Ph.D., and D.Sci. degrees in automatic control and computer science from the Silesian University of Technology, Gliwice, Poland, in 1977, 1980, and 1984, respectively.

He is currently a Professor and the Canada Research Chair in computational intelligence with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. He is also with the Systems Research Institute of the Polish Academy of Sciences, Warsaw, Poland. His main research interests include computational intelligence, fuzzy modeling and granular computing, knowledge discovery and data mining, fuzzy control, pattern recognition, knowledge-based neural networks, relational computing, and software engineering.

Dr. Pedrycz was elected as a Foreign Member of the Polish Academy of Sciences in 2009. In 2012, he was elected as a fellow of the Royal Society of Canada. He was a recipient of the prestigious Norbert Wiener Award from the IEEE Systems, Man, and Cybernetics Society in 2007, the IEEE Canada Computer Engineering Medal, the Cajastur Prize for Soft Computing from the European Centre for Soft Computing, the Killam Prize, and the Fuzzy Pioneer Award from the IEEE Computational Intelligence Society. He is also the Editor-in-Chief of *Information Sciences*, *WIREs Data Mining and Knowledge Discovery*, and *Granular Computing*.