

Feature Selection for Interval-Valued Data Based on D-S Evidence Theory

YICHUN PENG¹ AND QINLI ZHANG²

¹School of Computer Science and Engineering, Yulin Normal University, Yulin, Guangxi 537000, China

²School of Big Data and Artificial Intelligence, Chizhou University, Chizhou, Anhui 247000, China

Corresponding author: Qinli Zhang (zhangqinli2021@126.com)

ABSTRACT Feature selection is one basic and critical technology for data mining, especially in current “big data era”. Rough set theory (RST) is sensitive to noise in feature selection due to the strict condition of equivalence relation. However, D-S evidence theory is flexible to measure uncertainty of information. This paper introduces robust feature evaluation metrics “belief function” and “plausibility function” into feature selection algorithm to avoid the defect that classification effect is affected by noise. First of all, similarity between information values in an interval-valued information system (IVIS) is given and a variable parameter to control the similarity of samples is introduced. Then, θ -lower approximation and θ -upper approximation in IVIS are put forward. Next, belief function and plausibility function based on θ -lower approximation and θ -upper approximation are put forward. Finally, four feature selection algorithms in an IVIS based on D-S evidence theory are proposed. The experimental results on four real interval-valued datasets show that the proposed metrics are robust to noise, and the proposed algorithms are more effective than the existing algorithms.


INDEX TERMS Interval-valued data, IVIS, D-S evidence theory, belief function, plausibility function, feature selection.

I. INTRODUCTION

A. RESEARCH BACKGROUND AND RELATED WORKS

Rough set theory (RST) is put forward by Pawlak [24], [25]. It is a significant method to deal with imprecision, fuzziness and uncertainty. It also doesn't need any prior information beyond the data set that the problem needs to be processed [17], [37]. RST, based on classification mechanism, regards classification as equivalence relation in a specific space, and equivalence relation constitutes the division of space. The main idea of RST is to use the known knowledge base to describe the imprecise or uncertain knowledge. RST can effectively deal with the uncertainty of information system (IS). In recent years, RST has attracted many researchers' attention, and its application is mostly related to an IS [19]–[21], [44].

Another important method used to deal with uncertainty in an IS is D-S evidence theory. It is originated by Dempster's concepts of lower and upper probabilities [3], and extended into a theory by Shafer [33]. The theory is nowadays widely used for the objective and subjective uncertainty

The associate editor coordinating the review of this manuscript and approving it for publication was Francisco J. Garcia-Penalvo .

analysis [11], [12], [35]. The use of D-S evidence theory in risk analysis has many advantages over the conventional probabilistic approach. It provides convenient and comprehensive way to handle uncertain problems including imprecisely specified distributions, poorly known and unknown correlation between different variables, modeling uncertainty, small sample size, and measuring uncertainty. The basic representational structure in this theory is a belief structure. The primitive numeric measurements derived from the belief structure are a dual pair of belief and plausibility functions. Belief function and plausibility function can successfully distinguish the difference between “uncertainty” and “don't know”, and satisfy the relatively weaker condition than probability theory.

Feature selection is an effective way to eliminate negative effects caused by redundant features. In the framework of RST, feature selection is also called attribute reduction, which is to find a minimal feature subset that provides the same discriminating information as the whole set of features. Specifically, some features in an IS are redundant. We want to find a reduct that has the fewest features. Feature selection in an IS mean deleting redundant features under keeping the classification ability. The core step of feature selection

is to construct an feature evaluation function. This function can be used to select key representative features from high-dimensional data. Thus, feature selection can simplify data and reduce the computational complexity of machine learning. At present, this technology is widely applied in data mining, pattern recognition, and other real-life fields [36].

Belief and plausibility functions in D-S evidence theory can be used in feature selection. For example, Wu *et al.* [42] discussed feature selection in complete decision systems based on D-S evidence theory. Moreover, Wu [38] studied knowledge reduction in incomplete ISs without decision with evidence theory. Zhang *et al.* [49] proposed the concepts of belief reduction and plausibility reduction in complete ISs without decision.

An interval-valued information system (IVIS) means an IS where the information values are interval numbers. In order to study interval-valued data, some researchers have extended RST and established a generalized model of single IS. Dai *et al.* [9] introduced uncertainty measurement for an IVIS based on α -weak similarity. Zhang *et al.* [48] presented incremental updating of rough approximations in an IVIS under attribute generalization. Leung *et al.* [18] brought up the minimal attribute reduction method for an IVIS and obtained all classification rules hidden in an IS through a knowledge induction process. Xie *et al.* [44] considered new measurements of uncertainty and information structures for an IVIS. Yang *et al.* [46] investigated the dominance-based rough set in an IVIS, which contains both incomplete and imprecise evaluations of objects. Sakai *et al.* [29] developed a rule generation prototype system for incomplete information databases that can process IVISs.

In this paper, we introduce robust feature evaluation metrics “belief function” and “plausibility function” into feature selection algorithms to avoid the defect that classification effect is affected by noise in an IVIS. Similarity between information values in an IVIS is established and then θ -lower and θ -upper approximations in IVIS are presented. Second, feature selection algorithms in an IVIS based on D-S evidence theory are proposed. Finally, experiments are performed to evaluate the robustness of the proposed metric and the performance of the algorithms. Experimental results show that the proposed metrics are robust to noise and the proposed feature selection algorithms are effective.

This paper is organized as follows. In Section 2, we review binary relations, IVISs and D-S evidence theory. In Section 3, we study feature selection in an IVIS based on D-S evidence theory and give the corresponding algorithms. In Section 4, we perform numerical experiments and do effectiveness analysis of the proposed metrics. In Section 5, we evaluate the performance of the given algorithms. Section 6 concludes the paper.

The work flow of this paper is displayed in FIGURE 1.

II. PRELIMINARIES

The binary relation, IVIS and D-S evidence theory are briefly introduced in the section.

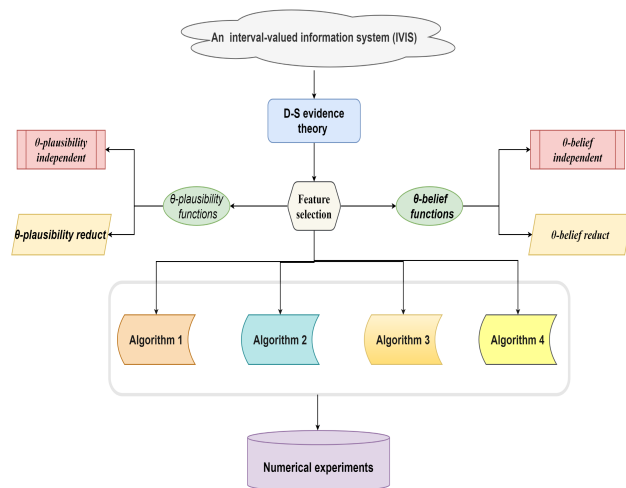


FIGURE 1. CV-Values of belief measurement and plausibility measurement on four interval-valued datasets.

In this paper, $U = \{u_1, u_2, \dots, u_n\}$ and $A = \{a_1, a_2, \dots, a_m\}$ denote two non-empty finite sets, 2^U denotes the family of all subsets on U , $|X|$ denotes the cardinality of $X \in 2^U$, $\delta = U \times U$ denotes the universal relation and $\Delta = \{(u, u) : u \in U\}$ denotes the identity relation.

$$\text{For any } X \in 2^U, P(X) = \frac{|X|}{|U|} = \frac{|X|}{n}.$$

A. BINARY RELATION

R is a binary relation on U if $R \subseteq U \times U$. If $(x, y) \in R$, we also denote it by xRy .

Let R be a binary relation on U . Then R is called

- (1) Reflexive, if xRx for any $x \in U$.
- (2) Symmetric, if xRy implies yRx for any $x, y \in U$.
- (3) Transitive, if xRy and yRz imply xRz for any $x, y, z \in U$.

R is called an equivalence relation on U , if R is reflexive, symmetric and transitive. R is called a tolerance relation on U , if R is reflexive and symmetric. Moreover, R is called a universal relation on U if $R = \delta$; R is said to be an identity relation on U if $R = \Delta$.

B. INTERVAL-VALUED NUMBER

$[R] = \{m = [m^-, m^+] : m^-, m^+ \in R, m^- \leq m^+\}$ is called an interval-valued number.

For any $r \in R$, express $r = [r, r]$.

For any $m, n \in [R]$, define

- (1) $m = n \Rightarrow m^- = n^-, m^+ = n^+$.
- (2) $m \leq n \Rightarrow m^- \geq n^-, m^+ \leq n^+$; $m < n \Rightarrow m \leq n, m \neq n$.

Definition 1 ([23]): Let $m, n \in [R]$. Then the possible degree of m relative to n is defined as follows:

$$p(m, n) = \min\{1, \max\{\frac{m^+ - n^-}{(m^+ - m^-) + (n^+ - n^-)}, 0\}\}.$$

Proposition 1 ([23]): The following properties hold:

- (1) $\forall m, n \in [R], 0 \leq p(m, n) \leq 1$;
- (2) $\forall m \in [R], p(m, m) = 0.5$;
- (3) $\forall m, n \in [R], p(m, n) + p(n, m) = 1$.

Definition 2 ([6]): Let $m, n \in [R]$. Then the similarity degree of m and n is defined as follows:

$$s(m, n) = 1 - |p(m, n) - p(n, m)|.$$

Proposition 2 ([6]): The following properties hold:

- (1) $\forall m, n \in [R], s(m, n) = s(n, m)$;
- (2) $\forall m, n \in [R], 0 \leq s(m, n) \leq 1$;
- (3) $\forall m, n \in [R], s(m, n) = 1 \Rightarrow m = n$.

Example 1: Pick $m = [4, 10]$ and $n = [6, 9]$. Then

$$p(m, n) = \min\{1, \max\{\frac{10 - 6}{(10 - 4) + (9 - 6)}, 0\}\} = \frac{4}{9},$$

$$p(n, m) = \min\{1, \max\{\frac{9 - 4}{(10 - 4) + (9 - 6)}, 0\}\} = \frac{5}{9},$$

$$s(m, n) = 1 - |p(m, n) - p(n, m)| = 1 - |\frac{4}{9} - \frac{5}{9}| = \frac{8}{9}.$$

C. AN IVIS

Definition 3 ([25]): Let U be a finite set of objects and A a finite set of attributes. Then the ordered pair (U, A) is referred to as an IS, if $a \in A$ is able to decide a function $a : U \rightarrow V_a$, where $V_a = \{a(u) : u \in U\}$.

Let (U, A) be an IS and $B \subseteq A$. Then we can define

$$ind(B) = \{(u, v) \in U \times U : \forall a \in A, a(u) = a(v)\}.$$

Obviously, $ind(B)$ is an equivalence relation on U .

Denote

$$[u]_B = \{v \in U : (u, v) \in ind(B)\}.$$

Then $[u]_B$ is known as the equivalence class of the object u under the equivalence relation $ind(B)$.

Definition 4 ([6]): Let (U, A) be an IS. Then (U, A) called an IVIS, for $\forall a \in A$ and $u \in U$, $a(u)$ is an interval-valued number. If $B \subseteq A$, then (U, B) is known as the subsystem of (U, A) .

Definition 5: Let (U, A) be an IVIS. Given $B \subseteq A$ and $\theta \in [0, 1]$. Define

$$R_B^\theta = \{(u, v) \in U \times U : \forall a \in B, s(a(u), a(v)) \geq \theta\}.$$

Clearly, R_B^θ is a tolerance relation on U and $R_B^\theta = \bigcap_{a \in B} R_a^\theta$

where $R_{\{a\}}^\theta = R_a^\theta$.

Definition 6: Let (U, A) be an IVIS. Given $\theta \in [0, 1]$ and $B \subseteq A$. Then $\forall u \in U$, the θ -tolerance class of u under R_B^θ is defined as

$$R_B^\theta(u) = \{v \in U : (u, v) \in R_B^\theta\}.$$

Clearly, $R_B^\theta(u) = \bigcap_{a \in B} R_a^\theta(u)$.

Definition 7: Let (U, A) be an IVIS. Given $X \in 2^U$, $B \subseteq A$ and $\theta \in [0, 1]$. Suppose that R_B^θ is the tolerance relation induced by the subsystem (U, B) with respect to θ . Based on the approximation space (U, R_B^θ) , we define a pair of operations \underline{R}_B^θ and \overline{R}_B^θ as follows:

$$\underline{R}_B^\theta(X) = \{u \in U : R_B^\theta(u) \subseteq X\},$$

$$\overline{R}_B^\theta(X) = \{u \in U : R_B^\theta(u) \cap X \neq \emptyset\}.$$

Then $\underline{R}_B^\theta(X)$ and $\overline{R}_B^\theta(X)$ are called θ -lower and θ -upper approximations of X , respectively. In general, \underline{R}_B^θ and \overline{R}_B^θ are named as θ -rough approximations of X .

Theorem 1: Let (U, A) be an IVIS. Then the following properties hold.

$$(1) \underline{R}_B^\theta(\emptyset) = \overline{R}_B^\theta(\emptyset) = \emptyset, \underline{R}_B^\theta(U) = \overline{R}_B^\theta(U) = U.$$

$$(2) \underline{R}_B^\theta(X) \subseteq X \subseteq \overline{R}_B^\theta(X).$$

$$(3) X \subseteq Y \Rightarrow \underline{R}_B^\theta(X) \subseteq \underline{R}_B^\theta(Y), \overline{R}_B^\theta(X) \subseteq \overline{R}_B^\theta(Y).$$

$$(4) \text{ If } B_1 \subseteq B_2 \subseteq A, \text{ then } \forall \theta \in [0, 1] \text{ and } X \in 2^U,$$

$$\underline{R}_{B_1}^\theta(X) \subseteq \underline{R}_{B_2}^\theta(X), \overline{R}_{B_2}^\theta(X) \subseteq \overline{R}_{B_1}^\theta(X).$$

$$(5) \text{ If } 0 \leq \theta_1 \leq \theta_2 \leq 1, \text{ then } \forall B \subseteq A \text{ and } X \in 2^U,$$

$$\underline{R}_B^{\theta_2}(X) \subseteq \underline{R}_B^{\theta_1}(X), \overline{R}_B^{\theta_1}(X) \subseteq \overline{R}_B^{\theta_2}(X).$$

$$(6) \underline{R}_B^\theta(X \cap Y) = \underline{R}_B^\theta(X) \cap \underline{R}_B^\theta(Y); \overline{R}_B^\theta(X \cup Y) = \overline{R}_B^\theta(X) \cup \overline{R}_B^\theta(Y).$$

$$(7) \underline{R}_B^\theta(U - X) = U - \overline{R}_B^\theta(X); \overline{R}_B^\theta(U - X) = U - \underline{R}_B^\theta(X).$$

Proof: Obviously. □

D. D-S EVIDENCE THEORY

D-S evidence theory, also called ‘‘Evidence theory’’ or ‘‘Belief function theory’’, is treated as a promising method of dealing with uncertainty in IS. Basic representational structure in D-S evidence theory is a belief structure [40].

Definition 8 ([40]): Let U be a non-empty finite set, a set function $m : 2^U \rightarrow I$ is referred to as a crisp basic probability assignment if it satisfies the following conditions:

$$m(\emptyset) = 0, \sum_{A \subseteq U} m(A) = 1.$$

A set $X \in 2^U$ with nonzero basic probability assignment is referred to as a focal element. We denote the family of all focal elements of m by \mathcal{M} . The pair (\mathcal{M}, m) is called a belief structure on U . A set function $Bel : 2^U \rightarrow I$ is referred to as a belief function if

$$Bel(X) = \sum_{A \subseteq X} m(A), \forall X \in 2^U.$$

A set function $Pl : 2^U \rightarrow I$ is referred to as a plausibility function if

$$Pl(X) = \sum_{A \cap X \neq \emptyset} m(A), \forall X \in 2^U.$$

A belief function $Bel : 2^U \rightarrow I$ can also be equivalently defined by axioms. That is, Bel is a belief function if it satisfies the following axioms:

- 1) $Bel(\emptyset) = 0$;
- 2) $Bel(U) = 1$;

3) For all $X_i \in 2^U, i = 1, 2, \dots, k,$

$$Bel(\bigcup_{i=1}^k X_i) \geq \sum_{\emptyset \neq J \subseteq \{1,2,\dots,k\}} (-1)^{|J|+1} Bel(\bigcap_{i \in J} X_i).$$

Belief and plausibility functions based on the same belief structure are connected by the dual property

$$Pl(X) = 1 - Bel(\sim X),$$

and furthermore, $Bel(X) \leq Pl(X),$ for all $X \in 2^U.$

A basic probability assignment m can also be represented by its belief function Bel by using the following Möbius transform:

$$m(X) = \sum_{Y \subseteq X} (-1)^{|X \setminus Y|} Bel(Y), \forall X \in 2^U$$

III. FEATURE SELECTION IN AN IVIS BASED ON D-S EVIDENCE THEORY

In this section, we study feature selection in an IVIS based on D-S evidence theory.

A. BELIEF REDUCTION AND PLAUSIBILITY REDUCTION

Theorem 2: Let (U, A) be an IVIS. Given $B \subseteq A$ and $\theta \in [0, 1], \forall X \in 2^U,$ define

$$Bel_B^\theta(X) = P(\underline{R}_B^\theta(X)), Pl_B^\theta(X) = P(\overline{R}_B^\theta(X)).$$

Then Bel_B^θ and Pl_B^θ are θ -belief and θ -plausibility functions on $U,$ respectively and the corresponding basic probability assignment is

$$m_B^\theta(Y) = P(j_B^\theta(Y)), \forall Y \in 2^U,$$

where $j_B^\theta(Y) = \{u \in U : R_B^\theta(u) = Y\}.$

Proof:

(1) Since $j_B^\theta(\emptyset) = \emptyset,$ we know $m_B^\theta(\emptyset) = \frac{|\emptyset|}{n} = 0.$

$\forall u \in U,$ pick $Y^* = R_B^\theta(u),$ we have $u \in j_B^\theta(Y^*).$

Then $u \in \bigcup_{Y \in 2^U} j_B^\theta(Y).$ Hence $U \subseteq \bigcup_{Y \in 2^U} j_B^\theta(Y).$ Since $\bigcup_{Y \in 2^U} j_B^\theta(Y) \subseteq U,$ we have

$$\bigcup_{Y \in 2^U} j_B^\theta(Y) = U.$$

Note that $Y_1 \neq Y_2, j_B^\theta(Y_1) \cap j_B^\theta(Y_2) = \emptyset.$ This follows that

$$\sum_{Y \in 2^U} |j_B^\theta(Y)| = n.$$

Then

$$\sum_{Y \in 2^U} m_B^\theta(Y) = \sum_{Y \in 2^U} \frac{|j_B^\theta(Y)|}{n} = 1.$$

So, m_B^θ is a crisp basic probability assignment according to Definition 8.

(2) $\forall X \in 2^U, u \in \underline{R}_B^\theta(X),$ by Definition 7, we have $R_B^\theta(u) \subseteq X.$ Denote $R_B^\theta(u) = Y^*.$ Then we have $u \in j_B^\theta(Y^*)$ and $Y^* \subseteq X.$ So $u \in \bigcup_{Y \subseteq X} j_B^\theta(Y).$

Hence

$$\underline{R}_B^\theta(X) \subseteq \bigcup_{Y \subseteq X} j_B^\theta(Y).$$

$\forall X \in 2^U, u \in \bigcup_{Y \subseteq X} j_B^\theta(Y),$ we have $\exists Y^* \subseteq X, u \in j_B^\theta(Y^*).$ Then $R_B^\theta(u) = Y^*$ and $Y^* \subseteq X.$ This follows that $R_B^\theta(u) \subseteq X.$ So $u \in \underline{R}_B^\theta(X).$

Hence

$$\bigcup_{Y \subseteq X} j_B^\theta(Y) \subseteq \underline{R}_B^\theta(X).$$

Thus

$$\underline{R}_B^\theta(X) = \bigcup_{Y \subseteq X} j_B^\theta(Y).$$

Note that $Y_1 \neq Y_2, j_B^\theta(Y_1) \cap j_B^\theta(Y_2) = \emptyset.$ This follows that $\forall X \in 2^U,$

$$|\underline{R}_B^\theta(X)| = \sum_{Y \subseteq X} |j_B^\theta(Y)|.$$

Then we obtain $\forall X \in 2^U,$

$$\frac{|\underline{R}_B^\theta(X)|}{n} = \sum_{Y \subseteq X} \frac{|j_B^\theta(Y)|}{n}.$$

Hence

$$Bel_B^\theta(X) = \sum_{Y \subseteq X} m_B^\theta(Y).$$

By Definition 8, Bel_B^θ is a belief function.

(3) $\forall X \in 2^U, u \in \overline{R}_B^\theta(X),$ by Definition 7, we have $R_B^\theta(u) \cap X \neq \emptyset.$ Denote $R_B^\theta(u) = Y^*.$ Then we have $u \in j_B^\theta(Y^*)$ and $Y^* \cap X \neq \emptyset.$ So $u \in \bigcup_{Y \cap X \neq \emptyset} j_B^\theta(Y).$

Hence

$$\overline{R}_B^\theta(X) \subseteq \bigcup_{Y \cap X \neq \emptyset} j_B^\theta(Y).$$

$\forall X \in 2^U, u \in \bigcup_{Y \cap X \neq \emptyset} j_B^\theta(Y),$ we have $\exists Y^* \cap X \neq \emptyset, u \in j_B^\theta(Y^*).$ Then $R_B^\theta(u) = Y^*$ and $Y^* \cap X \neq \emptyset.$ This follows that $R_B^\theta(u) \cap X \neq \emptyset.$ So $u \in \overline{R}_B^\theta(X).$

Hence

$$\bigcup_{Y \cap X \neq \emptyset} j_B^\theta(Y) \subseteq \overline{R}_B^\theta(X).$$

Thus

$$\overline{R}_B^\theta(X) = \bigcup_{Y \cap X \neq \emptyset} j_B^\theta(Y).$$

Note that $Y_1 \neq Y_2, j_B^\theta(Y_1) \cap j_B^\theta(Y_2) = \emptyset.$ This follows that $\forall X \in 2^U,$

$$|\overline{R}_B^\theta(X)| = \sum_{Y \cap X \neq \emptyset} |j_B^\theta(Y)|.$$

Then we obtain $\forall X \in 2^U$,

$$\frac{|\overline{R}_B^\theta(X)|}{n} = \sum_{Y \cap X \neq \emptyset} \frac{|J_B^\theta(Y)|}{n}.$$

Hence

$$Pl_B^\theta(X) = \sum_{Y \cap X \neq \emptyset} m_B^\theta(Y).$$

By Definition 8, Pl_B^θ is a plausibility function. \square

Proposition 3: Let (U, A) be an IVIS. If $C \subseteq B \subseteq A$, then $\forall X \in 2^U$ and $\theta \in [0, 1]$, the following inequalities hold,

$$Bel_C^\theta(X) \leq Bel_B^\theta(X) \leq P(X) \leq Pl_B^\theta(X) \leq Pl_C^\theta(X).$$

Proof: $\forall X \in 2^U$, by Theorem 1(2), we have

$$\underline{R}_B^\theta(X) \subseteq X \subseteq \overline{R}_B^\theta(X)$$

Then

$$|\underline{R}_B^\theta(X)| \leq |X| \leq |\overline{R}_B^\theta(X)|$$

So

$$Bel_B^\theta(X) \leq P(X) \leq Pl_B^\theta(X).$$

Since $C \subseteq B$, by Theorem 1(4), we have

$$\underline{R}_C^\theta(X) \subseteq \underline{R}_B^\theta(X), \overline{R}_B^\theta(X) \subseteq \overline{R}_C^\theta(X).$$

Then

$$|\underline{R}_C^\theta(X)| \leq |\underline{R}_B^\theta(X)|, |\overline{R}_B^\theta(X)| \leq |\overline{R}_C^\theta(X)|.$$

So

$$Bel_C^\theta(X) \leq Bel_B^\theta(X), Pl_B^\theta(X) \leq Pl_C^\theta(X).$$

Thus $\forall X \in 2^U$,

$$Bel_C^\theta(X) \leq Bel_B^\theta(X) \leq P(X) \leq Pl_B^\theta(X) \leq Pl_C^\theta(X).$$

\square

In this paper, denote

$$U/R_A^\theta = \{R_A^\theta(u) : u \in U\} = \{D_1, D_2, \dots, D_t\}.$$

For any $B \subseteq A$, put

$$Bel^\theta(B) = \sum_{i=1}^t Bel_B^\theta(D_i),$$

$$Pl^\theta(B) = \sum_{i=1}^t Pl_B^\theta(D_i).$$

Definition 9: Let (U, A) be an IVIS. Given $B \subseteq A$ and $\theta \in [0, 1]$.

(1) B is referred to as a classical θ -consistent set of A , if $R_B^\theta = R_A^\theta$.

(2) B is referred to as a θ -belief consistent set of A , if

$$\forall i, Bel_B^\theta(D_i) = Bel_A^\theta(D_i).$$

(3) B is referred to as a θ -plausibility consistent set of A , if

$$\forall i, Pl_B^\theta(D_i) = Pl_A^\theta(D_i).$$

Definition 10: Suppose that (U, A) is an IVIS. Given $a \in B \subseteq A$ and $\theta \in [0, 1]$.

(1) a is called θ -independent in B , if $R_B^\theta \neq R_{B-\{a\}}^\theta$.

(2) a is called θ -belief independent in B , if $\exists i, Bel_B^\theta(D_i) \neq Bel_{B-\{a\}}^\theta(D_i)$.

(3) a is called θ -plausibility independent in B , if $\exists i, Pl_B^\theta(D_i) \neq Pl_{B-\{a\}}^\theta(D_i)$.

Definition 11: Assume that (U, A) is an IVIS. Given $B \subseteq A$ and $\theta \in [0, 1]$.

(1) B is called θ -independent, if $\forall a \in B$, a is θ -independent in B .

(2) B is called θ -belief independent, if $\forall a \in B$, a is θ -belief independent in B .

(3) B is called θ -plausibility independent, if $\forall a \in B$, a is θ -plausibility independent in B .

Definition 12: Let (U, A) be an IVIS. Given $B \subseteq A$ and $\theta \in [0, 1]$.

(1) B is called a θ -reduct of A , if B is both θ -consistent and θ -independent.

(2) B is called a θ -belief reduct of A , if B is both θ -belief consistent and θ -belief independent.

(3) B is called a θ -plausibility reduct of A , if B is both θ -plausibility consistent and θ -plausibility independent.

In this paper, the family of all θ -consistent (resp., θ -belief, θ -plausibility) subsets of A is denoted by $co^\theta(A)$ (resp., $co_b^\theta(A)$, $co_p^\theta(A)$), and the family of all θ -reducts (resp., θ -belief reducts, θ -plausibility reducts) A is denoted by $red^\theta(A)$ (resp., $red_b^\theta(A)$, $red_p^\theta(A)$).

Obviously,

$$B \in red^\theta(A) \iff B \in co^\theta(A) \text{ and } \forall C \subset B, C \notin co^\theta(A);$$

$$B \in red_b^\theta(A) \iff B \in co_b^\theta(A) \text{ and } \forall C \subset B, C \notin co_b^\theta(A);$$

$$B \in red_p^\theta(A) \iff B \in co_p^\theta(A) \text{ and } \forall C \subset B, C \notin co_p^\theta(A).$$

Theorem 3: Let (U, A) be an IVIS. Given $B \subseteq A$ and $\theta \in [0, 1]$, then the following conditions are equivalent:

(1) $B \in co^\theta(A)$;

(2) $B \in co_b^\theta(A)$;

(3) $Bel^\theta(B) = Bel^\theta(A)$.

Proof: (1) \Rightarrow (2).

Suppose $B \in co^\theta(A)$, we have $R_B^\theta = R_A^\theta$.

Then $\forall u \in U, R_B^\theta(u) = R_A^\theta(u)$.

This implies that $\forall i$,

$$R_B^\theta(u) \subseteq D_i \iff R_A^\theta(u) \subseteq D_i.$$

Then $\forall u \in U$,

$$u \in \underline{R}_B^\theta(D_i) \iff u \in \underline{R}_A^\theta(D_i).$$

So $\forall i$,

$$|\underline{R}_B^\theta(D_i)| = |\underline{R}_A^\theta(D_i)|.$$

Hence $\forall i$,

$$Bel_B^\theta(D_i) = Bel_A^\theta(D_i).$$

By Definition 9, $B \in co_b^\theta(A)$.

(2) \Rightarrow (3).

Suppose $B \in co_b^\theta(A)$. By Definition 9, $\forall i$, we have

$$Bel_B^\theta(D_i) = Bel_A^\theta(D_i).$$

Then

$$\sum_{i=1}^t Bel_B^\theta(D_i) = \sum_{i=1}^t Bel_A^\theta(D_i).$$

Thus

$$Bel^\theta(B) = Bel^\theta(A).$$

(3) \Rightarrow (1).

Suppose $Bel^\theta(B) = Bel^\theta(A)$. Then

$$\sum_{i=1}^t Bel_B^\theta(D_i) = \sum_{i=1}^t Bel_A^\theta(D_i).$$

This implies that

$$\sum_{i=1}^t (Bel_A^\theta(D_i) - Bel_B^\theta(D_i)) = 0.$$

By Proposition 3, $\forall i, Bel_A^\theta(D_i) - Bel_B^\theta(D_i) \geq 0$.

Then $\forall i$,

$$Bel_A^\theta(D_i) = Bel_B^\theta(D_i).$$

So $\forall i$,

$$|R_B^\theta(D_i)| = |R_A^\theta(D_i)|.$$

Since $B \subseteq A$, by Theorem 1(4), we have $\forall i$,

$$R_B^\theta(D_i) \subseteq R_A^\theta(D_i).$$

Hence $\forall i$,

$$R_B^\theta(D_i) = R_A^\theta(D_i).$$

This implies that $\forall u \in U, \forall i$,

$$R_B^\theta(u) \subseteq D_i \Leftrightarrow R_A^\theta(u) \subseteq D_i.$$

$\forall u \in U$, we have

$$R_A^\theta(u) \subseteq R_B^\theta(u).$$

Put $R_A^\theta(u) = D_{i^*}$. Then $R_A^\theta(u) \subseteq D_{i^*}$. This implies that $R_B^\theta(u) \subseteq D_{i^*}$.

So, $\forall v \in R_B^\theta(u)$, we obtain $v \in D_{i^*}$. Then $v \in R_A^\theta(u)$.

Hence

$$R_B^\theta(u) \subseteq R_A^\theta(u).$$

Thus

$$R_B^\theta(u) = R_A^\theta(u).$$

By Definition 9, $B \in co^\theta(A)$. □

Corollary 1: Let (U, A) be an IVIS. Given $B \subseteq A$ and $\theta \in [0, 1]$, then the following conditions are equivalent:

- (1) $B \in red^\theta(A)$;
- (2) $B \in red_b^\theta(A)$;
- (3) $Bel^\theta(B) = Bel^\theta(A)$ and $\forall C \subset B, Bel^\theta(C) \neq Bel^\theta(B)$.

Proof: It immediately follows from Theorem 3 and Definition 12. □

Theorem 4: Let (U, A) be an IVIS. Given $B \subseteq A$ and $\theta \in [0, 1]$, then $co^\theta(A) \subseteq co_p^\theta(A)$.

Proof:

Suppose $B \in co^\theta(A)$, we have $R_B^\theta = R_A^\theta$.

Then $\forall u \in U, R_B^\theta(u) = R_A^\theta(u)$.

This implies that $\forall i \in \{1, 2, \dots, t\}$,

$$R_B^\theta(u) \cap D_i \neq \emptyset \Leftrightarrow R_A^\theta(u) \cap D_i \neq \emptyset.$$

That is $\forall u \in U$,

$$u \in \overline{R_B^\theta(D_i)} \Leftrightarrow u \in \overline{R_A^\theta(D_i)}.$$

Then $\forall i, \overline{R_B^\theta(D_i)} = \overline{R_A^\theta(D_i)}$.

Hence $\forall i$,

$$Pl_B^\theta(D_i) = Pl_A^\theta(D_i).$$

By Definition 9, $B \in co_p^\theta(A)$.

Hence $co^\theta(A) \subseteq co_p^\theta(A)$. □

Definition 13: Let (U, A) be an IVIS. Given $B \subseteq A, \theta \in [0, 1]$ and $a \in A - B$.

(1) θ -belief significance of the feature a relative to B is defined as

$$sig_b^\theta(a, B) = Bel^\theta(B) - Bel^\theta(B - \{a\}).$$

(2) θ -plausibility significance of the feature a relative to B is defined as

$$sig_p^\theta(a, B) = Pl^\theta(B - \{a\}) - Pl^\theta(B).$$

We stipulate

$$sig_b^\theta(a, \emptyset) = 0, sig_p^\theta(a, \emptyset) = 0.$$

Theorem 5: Let (U, A) be an IVIS. Suppose $\theta \in [0, 1]$ and $B \subseteq A$, then the following conditions are equivalent:

- (1) $B \in red_b^\theta(A)$;
- (2) $Bel^\theta(B) = Bel^\theta(A)$ and $\forall a \in B, sig_b^\theta(a, B) > 0$.

Proof: It can be proved by Theorem 3. □

Theorem 6: Let (U, A) be an IVIS. Suppose $\theta \in [0, 1]$ and $B \subseteq A$, then the following conditions are equivalent:

- (1) $B \in red_p^\theta(A)$;
- (2) $Pl^\theta(B) = Pl^\theta(A)$ and $\forall a \in B, sig_p^\theta(a, B) > 0$.

Proof: It can be proved by Theorem 3. □

B. FEATURE SELECTION ALGORITHMS IN AN IVIS

Feature selection algorithms based on θ -belief and θ -plausibility functions are given as follows.

Algorithm 1 uses belief function to obtain the feature which is added to the current selected coordinated set in each loop. This algorithm terminates when the addition of any remaining feature does not increase the evaluating function.

Algorithm 1 θ -Belief Feature Selection Algorithm in an IVIS

```

input : An IVIS  $(U, A)$  and  $\theta \in [0, 1]$ .
output : One  $\theta$ -belief reduct  $B$ .
begin
    Let  $B \leftarrow \emptyset$ ,
     $start = 1$ .
    Compute  $Bel^\theta(A)$ .
    while  $start$  do
        for each  $a \in A - B$  do
            if  $Bel^\theta(B \cup \{a\}) \neq Bel^\theta(A)$  then
                 $B \leftarrow B \cup \{a\}$ 
            else
                 $start = 0$ 
            end if
        end for
    end while
    return  $B$ 
end
    
```

Algorithm 2 θ -Plausibility Feature Selection Algorithm in an IVIS

```

input : An IVIS  $(U, A)$  and  $\theta \in [0, 1]$ .
output : One  $\theta$ -plausibility reduct  $B$ .
begin
    Let  $B \leftarrow A$ ,
     $start = 1$ .
    Compute  $Pl^\theta(A)$ .
    while  $start$  do
        for each feature  $a \in B$  do
            if  $Pl^\theta(B - \{a\}) = Pl^\theta(B)$  then
                 $B \leftarrow B - \{a\}$ 
            else
                 $start = 0$ 
            end if
        end for
    end while
    return  $B$ 
end
    
```

For attribute set A , the worst search time for a reduct will result in $|A|(|A| + 1)/2$ evaluations using the evaluation function. So, the overall time complexity of Algorithm 1 is $O(|A|^2)$.

Algorithm 2 applies plausibility function to determining the feature which is thrown away from the current selected coordinated set in each loop. This algorithm terminates when reducing any remaining features increase the evaluation function. For attribute set A , the worst search time for a reduct will also result in $|A|(|A| + 1)/2$ evaluations using the evaluation function. So, the overall time complexity of Algorithm 2 is also $O(|A|^2)$.

We can give another two feature selection algorithms based on θ -belief significance and θ -plausibility significance in an IVIS (see Algorithms 3 and 4).

For Algorithms 3 and 4, the worst search time for a reduct will result in $|A|(|A| + 1)(|A| + 2)/6$ evaluations using the evaluation function. So the time complexity of Algorithms 3-4 is $O(|A|^3)$.

Algorithm 3 An Algorithm of θ -Reduction Based on θ -Belief Significance in an IVIS

```

input : An IVIS  $(U, A)$  and  $\theta \in [0, 1]$ .
output : One  $\theta$ -belief reduct  $B$ .
begin
    Initialize the reduct to the empty set, i.e.,  $B \leftarrow \emptyset$ ,
    while  $Bel^\theta(B) \neq Bel^\theta(A)$  do
         $Bel^\theta(B \cup \{\hat{a}\}) = \max_{a \in A-B} Bel^\theta(B \cup \{a\})$ , where  $\hat{a} \in A - B$ .
         $B \leftarrow B \cup \{\hat{a}\}$ 
    end while
    return  $B$ 
end
    
```

Algorithm 4 An Algorithm of θ -Reduction Based on θ -Plausibility Significance in an IVIS

```

input : An IVIS  $(U, A)$  and  $\theta \in [0, 1]$ .
output : One  $\theta$ -plausibility reduct  $B$ .
begin
    Initialize the reduct to the universal set, i.e.,  $B \leftarrow A$ ,
     $Pl^\theta(B - \{\hat{a}\}) = \min_{a \in B} Pl^\theta(B - \{a\})$ , where  $\hat{a} \in B$ .
    while  $Pl^\theta(B - \{\hat{a}\}) = Pl^\theta(A)$  do
         $B \leftarrow B - \{\hat{a}\}$ 
         $Pl^\theta(B - \{\hat{a}\}) = \min_{a \in B} Pl^\theta(B - \{a\})$ , where  $\hat{a} \in B$ .
    end while
    return  $B$ 
end
    
```

IV. NUMERICAL EXPERIMENTS AND EFFECTIVENESS ANALYSIS OF THE PROPOSED METRICS

A. MONOTONICITY ANALYSIS

In order to test the proposed belief and plausibility functions, we apply them to four real-life interval-valued datasets which are shown in TABLE 1.

TABLE 1. Four interval-valued datasets from the real world.

NO.	Date sets	Objects	Features	Classes
1	Car	33	7	4
2	Face	27	6	9
3	Fish	12	13	4
4	Water	316	48	2

For the dataset Car [2], let $C_i = \{a_1, \dots, a_i\}$ ($i = 1, \dots, 7$). And for each i , $Bel^\theta(C_i)$ and $Pl^\theta(C_i)$ denote the belief measurement and plausibility measurement of subsystem (U, C_i) , respectively. Then the two measurement sets on the dataset Car are defined as follows:

$$\begin{aligned}
 X_{Bel^\theta}(Car) &= \{Bel^\theta(C_1), \dots, Bel^\theta(C_7)\}, \\
 X_{Pl^\theta}(Car) &= \{Pl^\theta(C_1), \dots, Pl^\theta(C_7)\},
 \end{aligned}$$

For the dataset Face [1], let $F_i = \{a_1, \dots, a_i\}$ ($i = 1, \dots, 6$). And for each i , $Bel^\theta(F_i)$ and $Pl^\theta(F_i)$ denote the belief measurement and plausibility measurement of subsystem (U, F_i) , respectively. Then the two measurement sets on the dataset Face are defined as follows:

$$\begin{aligned}
 X_{Bel^\theta}(Face) &= \{Bel^\theta(F_1), \dots, Bel^\theta(F_6)\}, \\
 X_{Pl^\theta}(Face) &= \{Pl^\theta(F_1), \dots, Pl^\theta(F_6)\},
 \end{aligned}$$

For the dataset Fish [16], let $Y_i = \{a_1, \dots, a_i\}$ ($i = 1, \dots, 13$). And for each i , $Bel^\theta(Y_i)$ and $Pl^\theta(Y_i)$ denote the belief measurement and plausibility measurement of subsystem (U, Y_i), respectively. Then the two measurement sets on the dataset Fish are defined as follows:

$$X_{Bel^\theta}(Fish) = \{Bel^\theta(Y_1), \dots, Bel^\theta(Y_{13})\},$$

$$X_{Pl^\theta}(Fish) = \{Pl^\theta(Y_1), \dots, Pl^\theta(Y_{13})\},$$

For the dataset Water [26], let $W_i = \{a_1, \dots, a_i\}$ ($i = 1, \dots, 48$). And for each i , $Bel^\theta(W_i)$ and $Pl^\theta(W_i)$ denote the belief measurement and plausibility measurement of subsystem (U, W_i), respectively. Then the two measurement sets on the dataset Water are defined as follows:

$$X_{Bel^\theta}(Water) = \{Bel^\theta(W_1), \dots, Bel^\theta(W_{48})\},$$

$$X_{Pl^\theta}(Water) = \{Pl^\theta(W_1), \dots, Pl^\theta(W_{48})\},$$

Measurements of belief and plausibility on four interval-valued datasets are shown in FIGURES 2-5.

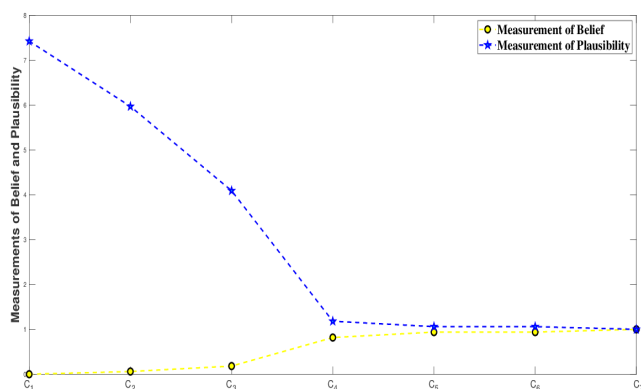


FIGURE 2. Measurements of Belief and Plausibility on dataset Car.

From FIGURES 2-5, the following conclusions can be obtained:

- (1) Belief measurement increases monotonically with the increase of the number of features.
- (2) Plausibility measurement decreases monotonically with the increase of the number of features.
- (3) The uncertainty of an IVIS decreases with the increase of the number of features.
- (4) Belief function and plausibility function can be used to measure the uncertainty of an IVIS.

B. DISPERSION ANALYSIS

Standard deviation is mainly used to measure the dispersion of numerical data. The larger the standard deviation, the higher the dispersion of the data. On the contrary, it means that the dispersion of data is lower. In this paper, the standard deviation coefficient is applied for effectiveness analysis of the proposed measurements.

Let $X = \{x_1, \dots, x_n\}$ be a data set. Then arithmetic average value, standard deviation and standard deviation coefficient

of X are denoted by $\bar{x}, \sigma(X)$ and $CV(X)$, respectively. They are defined as follows:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \sigma(X) = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}, CV(X) = \frac{\sigma(X)}{\bar{x}}.$$

According to the above formula, we calculate the CV-values of belief measurement and plausibility measurement on four interval-valued datasets. The results are shown in FIGURE 6.

FIGURE 6 shows that the CV-values of $Pl^\theta(B)$ are much larger than the CV-values of $Bel^\theta(B)$ on datasets Fish and Water, but there is little difference between them on datasets Car and Face. This is to say, the dispersion of $Bel^\theta(B)$ is small, so $Bel^\theta(B)$ has the better performance in measuring the uncertainty of an IVIS than $Pl^\theta(B)$.

C. CORRELATION ANALYSIS

In statistics, Pearson correlation coefficient is used to measure the linear correlation between two data sets.

Suppose that $X = \{x_1, \dots, x_n\}$ and $Y = \{y_1, \dots, y_n\}$ are two data sets. Pearson correlation coefficient between X and Y , denoted by $r(X, Y)$, is defined as

$$r(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}},$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$.

Obviously,

$$-1 \leq r(X, Y) \leq 1.$$

The correlation degree between X and Y can be obtained according to TABLE 2.

TABLE 2. Different degrees of correlation between X and Y.

$r(X, Y)$	Correlation between X and Y	Abbreviation
$r(X, Y) = 1$	Completely positive correlation	CPC
$0.7 \leq r(X, Y) < 1$	Height positive correlation	HPC
$0.4 \leq r(X, Y) < 0.7$	Moderate positive correlation	MPC
$0 < r(X, Y) < 0.4$	Low positive correlation	LPC
$r(X, Y) = 0$	No correlation	NC
$-0.4 < r(X, Y) < 0$	Low negative correlation	LNC
$-0.7 < r(X, Y) \leq -0.4$	Moderate negative correlation	MNC
$-1 < r(X, Y) \leq -0.7$	Height negative correlation	HNC
$r(X, Y) = -1$	Completely negative correlation	CNC

r-values between belief measurement and plausibility measurement are given on each of four interval-valued datasets according to the above equation. The results are shown in TABLES 3-6.

TABLES 3-6 show that the correlation between belief measurement and plausibility measurement is MNC on the dataset Water, while the correlation between belief measurement and plausibility measurement is HNC on the datasets Car, Face and Fish.

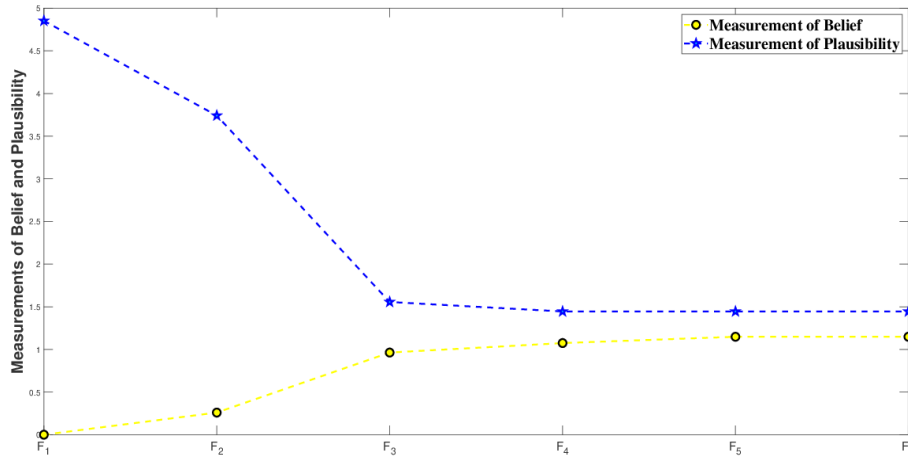


FIGURE 3. Measurements of Belief and Plausibility on dataset Face.

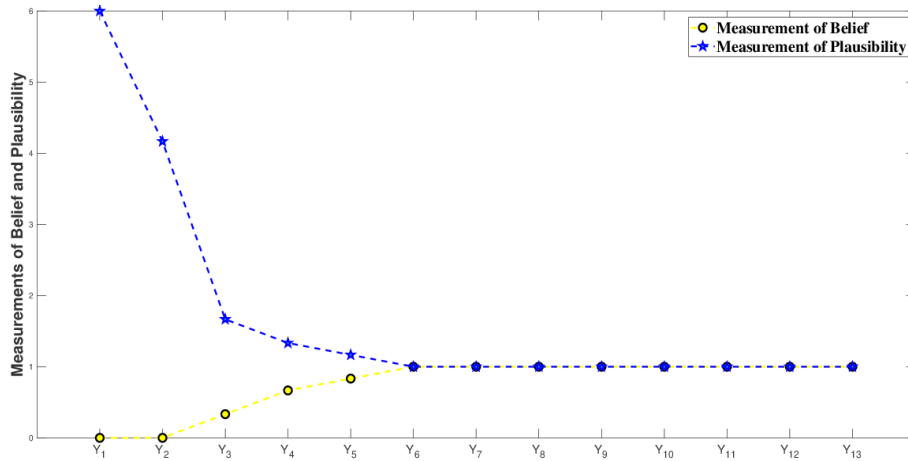


FIGURE 4. Measurements of Belief and Plausibility on dataset Fish.

TABLE 3. *r*-values between belief measurement and plausibility measurement on dataset Car.

<i>r</i>	$X_{Bel\theta}(Car)$	$X_{Pl\theta}(Car)$
$X_{Bel\theta}(Car)$	1	
$X_{Pl\theta}(Car)$	-0.9652	1

TABLE 4. *r*-values between belief measurement and plausibility measurement on dataset Face.

<i>r</i>	$X_{Bel\theta}(Face)$	$X_{Pl\theta}(Face)$
$X_{Bel\theta}(Face)$	1	
$X_{Pl\theta}(Face)$	-0.9915	1

TABLE 5. *r*-values between belief measurement and plausibility measurement on dataset Fish.

<i>r</i>	$X_{Bel\theta}(Fish)$	$X_{Pl\theta}(Fish)$
$X_{Bel\theta}(Fish)$	1	
$X_{Pl\theta}(Fish)$	-0.8948	1

V. PERFORMANCE ANALYSIS OF FEATURE SELECTION ALGORITHMS

In order to verify the performance of the proposed feature selection algorithms, they are applied to four real

TABLE 6. *r*-values between belief measurement and plausibility measurement on dataset Water.

<i>r</i>	$X_{Bel\theta}(Water)$	$X_{Pl\theta}(Water)$
$X_{Bel\theta}(Water)$	1	
$X_{Pl\theta}(Water)$	-0.6944	1

interval-valued datasets described in TABLE 1. Interval-valued data can be regarded as a kind of data with noise. Since there are very few feature selection algorithms for interval-valued data, we only compare our algorithm with the algorithm in reference [7]. We compare the classification accuracy of all feature selection results. The classification accuracy of all experiments is the result of 10-fold cross validation.

Classical classifiers cannot classify interval-valued data. In this paper, we transform KNN (*K*-Nearest Neighbors) to classify interval-valued datasets by redefining the distance between two samples.

Definition 14: Let $X = (x_1, x_2, \dots, x_t)$ and $Y = (y_1, y_2, \dots, y_t)$ be two samples of interval-valued dataset, where both x_i and y_i are interval numbers at the *i*-th feature.

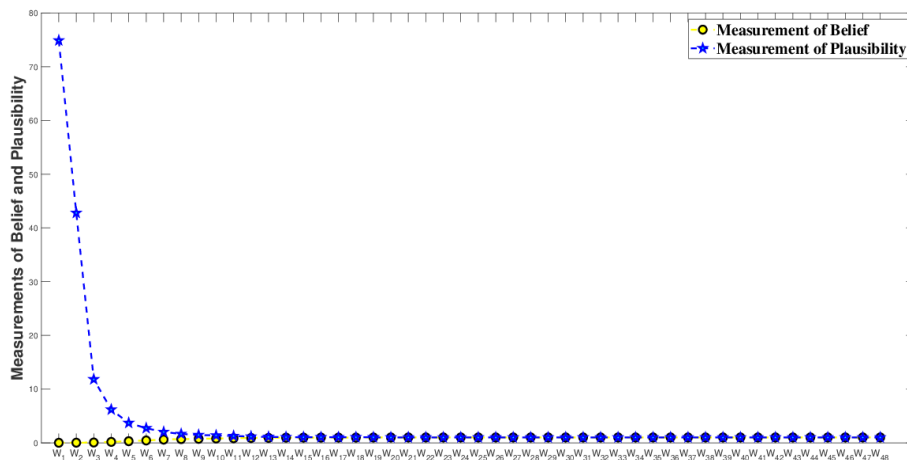


FIGURE 5. Measurements of Belief and Plausibility on dataset Water.

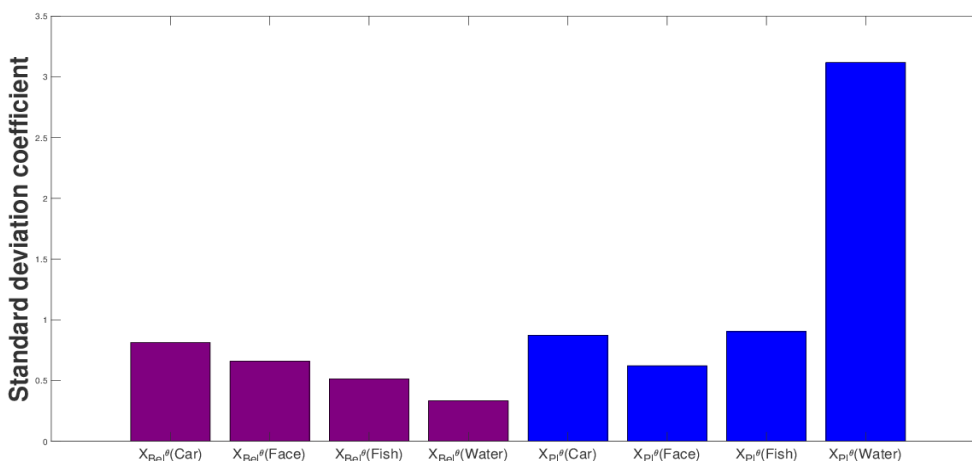


FIGURE 6. CV-Values of belief measurement and plausibility measurement on four interval-valued datasets.

TABLE 7. The results of feature selection and classification accuracy of our algorithm on four interval-valued datasets.

c Data sets	Selected features	Number of selected features	Classification accuracy
Car	{3, 4, 7}	3	0.7305
Face	{2, 3, 4}	3	1
Fish	{1, 6, 7}	3	0.9228
Water	{10, 19, 27, 31, 36, 43}	6	0.7899

TABLE 8. The results of feature selection and classification accuracy of the algorithm in reference [7] on four interval-valued datasets.

c Data sets	Selected features	Number of selected features	Classification accuracy
Car	{2, 3, 4, 5, 6, 7}	6	0.7058
Face	{1, 3, 4, 5, 6}	5	1
Fish	{2, 6, 7, 9, 10, 11}	6	0.9153
Water	{2, 5, 7, 14, 19, 20, 29, 31, 43}	9	0.7533

Then the distance between X and Y is defined as follows:

$$d(X, Y) = \sqrt{\sum_{i=1}^t (p(x_i, y_i) - p(y_i, x_i))^2}$$

If the features of dataset are disordered in each search, the algorithms may produce different reduction results. This paper only shows the results of a random search of all four datasets using Algorithm 1. If our algorithm can show good performance on four datasets, it can further prove the

superiority of our algorithm because of the results of a search in which the order of features is randomly arranged.

We use our algorithm and the algorithm in reference [7] to select features from four interval-valued datasets in TABLE 1. The value of parameter θ affects the result of feature selection. The value of parameter θ is set to 0.2 in this paper. The results of feature selection and the corresponding classification accuracy are listed in TABLEs 7 and 8.

The proposed algorithm not only has higher classification accuracy, but also selects fewer features than the algorithm in reference [7], as shown in TABLEs 7 and 8. This fully confirms that the proposed algorithm is robust to noise.

VI. CONCLUSION AND FUTURE WORK

Interval-valued data is difficult to deal with and can be considered as a kind of data with noise. Based on D-S evidence theory, this paper has proposed two new metrics to measure the uncertainty in IVIS, and applied them to feature selection for interval-valued data to solve the problem of noise sensitivity of RST. The experimental results on four real interval-valued datasets have showed that the proposed metrics are robust to noise, and the proposed feature selection algorithms are efficiency and accuracy than other algorithms based on RST. Our findings provide a new idea for feature selection for interval-valued data. In the future, we will study the optimization of parameter θ in the proposed feature selection algorithms.

REFERENCES

- [1] L. Billard, A. Douzal-Chouakria, and E. Diday, "Symbolic principal component for interval-valued observations," Univ. Georgia, Athens, GA, USA, Tech. Rep., Nov. 2008.
- [2] F. D. A. T. de Carvalho, R. M. C. R. de Souza, M. Chavent, and Y. Lechevallier, "Adaptive Hausdorff distances and dynamic clustering of symbolic interval data," *Pattern Recognit. Lett.*, vol. 27, no. 3, pp. 167–179, Feb. 2006.
- [3] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Ann. Math. Statist.*, vol. 38, no. 2, pp. 325–339, Apr. 1967.
- [4] I. Düntsch and G. Gediga, "Uncertainty measures of rough set prediction," *Artif. Intell.*, vol. 106, no. 1, pp. 109–137, Nov. 1998.
- [5] J. Dai, Q. Xu, W. Wang, and H. Tian, "Conditional entropy for incomplete decision systems and its application in data mining," *Int. J. Gen. Syst.*, vol. 41, no. 7, pp. 713–728, Oct. 2012.
- [6] J. Dai, W. Wang, Q. Xu, and H. Tian, "Uncertainty measurement for interval-valued decision systems based on extended conditional entropy," *Knowl.-Based Syst.*, vol. 27, pp. 443–450, Mar. 2012.
- [7] J. Dai, W. Wang, and J.-S. Mi, "Uncertainty measurement for interval-valued information systems," *Inf. Sci.*, vol. 251, pp. 63–78, Dec. 2013.
- [8] J.-H. Dai, H. Hu, G.-J. Zheng, Q.-H. Hu, H.-F. Han, and H. Shi, "Attribute reduction in interval-valued information systems based on information entropies," *Frontiers Inf. Technol. Electron. Eng.*, vol. 17, no. 9, pp. 919–928, Sep. 2016.
- [9] J. Dai, B. Wei, X. Zhang, and Q. Zhang, "Uncertainty measurement for incomplete interval-valued information systems based on α -weak similarity," *Knowl.-Based Syst.*, vol. 136, pp. 159–171, Nov. 2017.
- [10] J. Dai, Y. Yan, Z. Li, and B. Liao, "Dominance-based fuzzy rough set approach for incomplete interval-valued data," *J. Intell. Fuzzy Syst.*, vol. 34, no. 1, pp. 423–436, Jan. 2018.
- [11] D. Dubois and H. Prade, "Representation and combination of uncertainty with belief functions and possibility measures," *Comput. Intell.*, vol. 4, no. 3, pp. 244–264, Sep. 1988.
- [12] D. Yong, S. WenKang, Z. ZhenFu, and L. Qi, "Combining belief functions based on distance of evidence," *Decis. Support Syst.*, vol. 38, no. 3, pp. 489–493, Dec. 2004.
- [13] G. Facchinetti, R. G. Ricci, and S. Muzzioli, "Note on ranking fuzzy triangular numbers," *Int. J. Intell. Syst.*, vol. 13, no. 7, pp. 613–622, Jul. 1998.
- [14] A. Frank and A. Asuncion. *UCI Machine Learning Repository*. [Online]. Available: <http://archive.ics.uci.edu/ml>
- [15] M. Holčápek, "A graded approach to cardinal theory of finite fuzzy sets, Part I: Graded equipollence," *Fuzzy Sets Syst.*, vol. 298, pp. 158–193, Sep. 2016.
- [16] L. Hedjazi, J. Aguilar-Martin, and M.-V. Le Lann, "Similarity-margin based feature selection for symbolic interval data," *Pattern Recognit. Lett.*, vol. 32, no. 4, pp. 578–585, Mar. 2011.
- [17] Y. Huang, T. Li, C. Luo, H. Fujita, and S.-J. Horng, "Dynamic variable precision rough set approach for probabilistic set-valued information systems," *Knowl.-Based Syst.*, vol. 122, pp. 131–147, Apr. 2017.
- [18] 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.
- [19] Z. Li, D. Huang, X. Liu, N. Xie, and G. Zhang, "Information structures in a covering information system," *Inf. Sci.*, vol. 507, pp. 449–471, Jan. 2020.
- [20] Z. Li, X. Liu, J. Dai, J. Chen, and H. Fujita, "Measures of uncertainty based on Gaussian kernel for a fully fuzzy information system," *Knowl.-Based Syst.*, vol. 196, May 2020, Art. no. 105791.
- [21] Z. Li, P. Zhang, X. Ge, N. Xie, G. Zhang, and C. Wen, "Uncertainty measurement for a fuzzy relation information system," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 12, pp. 2338–2352, Dec. 2019.
- [22] J. Navarrete, D. Viejo, and M. Cazorla, "Color smoothing for RGB-D data using entropy information," *Appl. Soft Comput.*, vol. 46, pp. 361–380, Sep. 2016.
- [23] Y. Nakahara, M. Sasaki, and M. Gen, "On the linear programming problems with set coefficients," *Comput. Ind. Eng.*, vol. 23, pp. 301–304, Nov. 1992.
- [24] Z. Pawlak, "Rough sets," *Int. J. Comput. Inf. Sci.*, vol. 11, pp. 341–356, Oct. 1982.
- [25] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning About Data*. Dordrecht, The Netherlands: Kluwer, 1991.
- [26] J. Quevedo, V. Puig, G. Cembrano, J. Blanch, J. Aguilar, D. Saporta, G. Benito, M. Hedo, and A. Molina, "Validation and reconstruction of flow meter data in the Barcelona water distribution network," *Control Eng. Pract.*, vol. 18, no. 6, pp. 640–651, Jun. 2010.
- [27] C. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, Jul. 1948.
- [28] R. W. Swiniarski and A. Skowron, "Rough set methods in feature selection and recognition," *Pattern Recognit. Lett.*, vol. 24, no. 6, pp. 833–849, Mar. 2003.
- [29] H. Sakai, M. Nakata, and D. Slezak, "A prototype system for rule generation in Lipski's incomplete information databases," in *Proc. 13th Rough Sets, Fuzzy Sets, Data Mining Granular Comput.*, 2011, pp. 175–182.
- [30] R. Slowinski and D. Vanderpooten, "A generalized definition of rough approximations based on similarity," *IEEE Trans. Knowl. Data Eng.*, vol. 12, no. 2, pp. 331–336, Mar. 2000.
- [31] A. Skowron, "The relationship between rough set theory and evidence theory," *Bull. Polish Acad. Sci., Math.*, vol. 37, no. 1, pp. 87–90, 1989.
- [32] A. Skowron, "The rough sets theory and evidence theory," *Fundamenta Informaticae*, vol. 13, no. 3, pp. 245–262, Jul. 1990.
- [33] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ, USA: Princeton Univ. Press, 1976.
- [34] E. Shekarian, N. Kazemi, S. H. Abdul-Rashid, and E. U. Olugu, "Fuzzy inventory models: A comprehensive review," *Appl. Soft Comput.*, vol. 55, pp. 588–621, Jun. 2017.
- [35] Q. Sun, X. Q. Ye, and W. K. Gu, "A new combination rules of evidence theory," *Acta Electronica Sinica*, vol. 28, pp. 117–119, Oct. 2000.
- [36] L. Sun, X. Zhang, Y. Qian, J. Xu, and S. Zhang, "Feature selection using neighborhood entropy-based uncertainty measures for gene expression data classification," *Inf. Sci.*, vol. 502, pp. 18–41, Oct. 2019.
- [37] A. Tan, W.-Z. Wu, and Y. Tao, "A unified framework for characterizing rough sets with evidence theory in various approximation spaces," *Inf. Sci.*, vols. 454–455, pp. 144–160, Jul. 2018.
- [38] W.-Z. Wu, "Attribute reduction based on evidence theory in incomplete decision systems," *Inf. Sci.*, vol. 178, no. 5, pp. 1355–1371, Mar. 2008.
- [39] W.-Z. Wu, Y. Leung, and J.-S. Mi, "On generalized fuzzy belief functions in infinite spaces," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 2, pp. 385–397, Apr. 2009.
- [40] W.-Z. Wu, Y. Leung, and W.-X. Zhang, "Connections between rough set theory and Dempster-Shafer theory of evidence," *Int. J. Gen. Syst.*, vol. 31, no. 4, pp. 405–430, Jul. 2002.

- [41] W. Z. Wu and J. S. Mi, "Knowledge reduction in incomplete information systems based on Dempster-Shafer theory of evidence," in *Proc. 1st Int. Conf. Rough Sets Knowl. Technol.*, in Lecture Notes in Artificial Intelligence, J. F. Peters, A. Skowron, Y. Y. Yao, and G. Wang, Eds., 2006, pp. 254–261.
- [42] W.-Z. Wu, M. Zhang, H.-Z. Li, and J.-S. Mi, "Knowledge reduction in random information systems via Dempster-Shafer theory of evidence," *Inf. Sci.*, vol. 174, nos. 3–4, pp. 143–164, Aug. 2005.
- [43] Y. Wang, X. Chen, and K. Dong, "Attribute reduction via local conditional entropy," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 12, pp. 3619–3634, Dec. 2019.
- [44] 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.
- [45] Y. Y. Yao, "Relational interpretations of neighborhood operators and rough set approximation operators," *Inf. Sci.*, vol. 111, nos. 1–4, pp. 239–259, Nov. 1998.
- [46] X. Yang, D. Yu, J. Yang, and L. Wei, "Dominance-based rough set approach to incomplete interval-valued information system," *Data Knowl. Eng.*, vol. 68, no. 11, pp. 1331–1347, Nov. 2009.
- [47] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, pp. 338–356, Dec. 1965.
- [48] Y. Zhang, T. Li, C. Luo, J. Zhang, and H. Chen, "Incremental updating of rough approximations in interval-valued information systems under attribute generalization," *Inf. Sci.*, vol. 373, pp. 461–475, Dec. 2016.
- [49] M. Zhang, L. Da Xu, W.-X. Zhang, and H.-Z. Li, "A rough set approach to knowledge reduction based on inclusion degree and evidence reasoning theory," *Expert Syst.*, vol. 20, no. 5, pp. 298–304, Nov. 2003.



YICHUN PENG received the Ph.D. degree in environmental science from the University of Chinese Academy of Sciences, Beijing, China, in 2015. He is currently an Associate Professor with the School of Computer Science and Engineering, Yulin Normal University, Yulin, China. His current research interests include geographic information systems and remote sensing, neural networks, and rough set theory.



QINLI ZHANG received the Ph.D. degree from the School of Information Engineering, Jiangnan University, Wuxi, China, in 2009. He is currently a Professor with the School of Big Data and Artificial Intelligence, Chizhou University, Chizhou, China. His main research interests include fuzzy set theory, rough set theory, and bioinformatics.

• • •