ELSEVIER

Attribute reduction based on intuitionistic fuzzy dominance mutual information in intuitionistic fuzzy information systems

Xiaofeng Liu [a], Hong Mo [b], Jianhua Dai [c,*]

[a] *School of Mathematics Sciences, Zhejiang Normal University, Jinhua, Zhejiang 321004, China*
[b] *College of Electric and Information Engineering, Changsha University of Science and Technology, Changsha, Hunan 410114, China*
[c] *Hunan Provincial Key Laboratory of Intelligent Computing and Language Information Processing, Hunan Normal University, Changsha, Hunan 410081, China*

ARTICLE INFO

ABSTRACT

Intuitionistic fuzzy information system (IFIS) is an extension of fuzzy information system that can represent more uncertain information and more accurately describe the essence of fuzziness. Attribute reduction is an important problem in processing and analyzing IFISs. The article tries to propose an attribute reduction method in view of intuitionistic fuzzy dominance mutual information in IFISs, whose information values are intuitionistic fuzzy numbers. First, an intuitionistic fuzzy dominance relation is established in IFISs according to intuitionistic fuzzy dominance degrees, and the intuitionistic fuzzy information structure generated by the intuitionistic fuzzy dominance relation is constructed. Then, the intuitionistic fuzzy dominance entropy and its variations are researched, and some of their properties are discussed. Subsequently, an attribute reduction method and its algorithm based on intuitionistic fuzzy dominance mutual information are given. Furthermore, numerical studies and statistical tests are presented to evaluate the performance of the proposed method. Theoretical research and experiments show that the raised attribute reduction method is applicable to IFISs.

1. Introduction

Since Zadeh proposed the concept of information granule [1], information granularity has attracted great interest among researchers. Zadeh points out that there are three basic concepts that underline human cognition: causation, organization, and information granulation [2]. As a new interdisciplinary research field, granular computing has generated research topics such as granular thinking, granular logic, granular reasoning, granular analysis, granular processing, granular problem solving, and so on. Since the emergence of granular computing, people have studied the theory of granular computing from different perspectives, and some methods have been used to research granular computing, such as rough sets [3], fuzzy sets [4], concept lattice [5], and quotient space [6].

Rough set theory, as a newer technique for researching uncertainty, was introduced by Pawlak [3]. In rough sets, the so-called information system is a knowledge representation system. Many applications of rough sets, such as attribute reduction, pattern recognition, uncertainty reasoning, and decision analysis, are often associated with information systems. Attribute reduction (i.e., feature selection) is a core content in the analysis and processing of information systems, and the idea of rough sets can be well

---

reflected in the process of attribute reduction. Attributes in an information system are not equally important, reduction is to delete redundant and irrelevant attributes, so as to eliminate their impact on the calculation process and the final results, and improve the efficiency and accuracy on the processing and analysis of systems. For an information system, the construction of the evaluation standard of the importance of attributes in the system is the key to attribute reduction, so as to retain the attribute subsets with the same or similar classification ability as the raw attribute set. Notably, the research on attribute reduction in information systems has yielded remarkable achievements [7–9].

Shannon [10] expanded the fundamental concept of entropy to information theory. Entropy can express the substance of knowledge and information in various forms, so it has wide applications, especially in uncertainty measurement [11–13]. In attribute reduction, the entropy of measuring uncertainty can also be used as a standard to evaluate the importance of attributes. Xie et al. [14] gave a novel fuzzy-rough attribute reduction approach via local information entropy. Dai et al. [15] proposed a novel feature selection method based on fuzzy combination entropy considering global and local feature correlation. Chen et al. [16] researched two-dimensional improved attribute reductions based on distance granulation and condition entropy in incomplete interval-valued decision systems. In addition, more and more variations of entropy are being used as criteria to reduce attributes, especially mutual information. Mutual information considers both attribute relevance and redundancy, selecting optimal attributes by maximizing relevance and minimizing redundancy. Mutual information was first used for reducing attributes by Battiti [17], while Hu et al. [18] gave a greedy forward search strategy by extending mutual information to fuzzy mutual information. During and after this period, many reduction methods using mutual information as a criterion have emerged, such as the minimal-Redundancy-Maximal-Relevance (mRMR) method [19], the conditional mutual information-based feature selection (CMIFS) [20], the independent classification information (ICI) function [21], the granularity fuzzy max-relevance and independence (GFMRI) criterion [22], and so on.

Intuitionistic fuzzy (IF) sets, a generalized concept of fuzzy sets, were proposed by Atanassov [23]. In an IF set, each element corresponds to a membership degree and a non-membership degree, and the total of them is not more than 1. By setting membership degree and non-membership degree, IF set can solve three kinds of state information: positive degree, negative degree and hesitation degree, so it is able to handle more complex and uncertain problems and describe the essence of fuzziness more precisely. With the popularization of IF sets, some scholars began to devote themselves to the study of this theory [24–26]. At present, it has gradually penetrated into many theoretical and applied fields, such as pattern recognition, medical diagnosis, logical programming, approximate reasoning, and decision-making.

In intuitionistic fuzzy information systems (IFISs) [27], each attribute determines an IF set on the interval [0, 1], i.e., the information values of an IFIS are IF numbers (including memberships and non-memberships). It is worth mentioning that the IFIS given by Hesameddini in [28] is the only one with an IF decision class (i.e., the decision class corresponds to an IF set). Furthermore, the IFIS used by Feng et al. [29] or Tan et al. [30] is a pair with a family of IF relations and a universe, it is actually an IF knowledge base or IF relation information system. To the best of our knowledge, there are few studies on developing attribute reduction in IFISs. Tan et al. [31] gave an attribute reduction algorithm for IF information by using IF conditional entropy. However, they only considered the relevance (significance) of an attribute to the decision one by one, ignoring the redundancies produced by the selected attributes. From this point, we intend to construct a new IF dominance entropy and its variations based on IF dominance relation, and then reduce attributes based on IF dominance mutual information for IFISs.

The article aims to present an attribute reduction method based on IF dominance mutual information in IFISs. There are four main contributions to the article: (1) The IF dominance relation is raised in IFISs, and the IF information structure generated by the IF dominance relation is constructed. (2) The IF dominance entropy and its variations are studied, and some of their properties are discussed. (3) Based on the IF dominance mutual information, an attribute reduction method and its algorithm are given. (4) Some experiments are presented to evaluate the performance of the proposed reduction method.

The remainder of the article is structured as follows. Section 2 reviews some fundamental knowledge, such as information entropy and IF sets. Section 3 presents an IF dominance relation induced by IFIS, constructs an IF information structure based on the IF dominance relation, introduces IF dominance entropy and its variations, and discusses some of their properties. Section 4 introduces an attribute reduction method in view of IF dominance mutual information. Section 5 gives some experiments to evaluate the performance of the proposed attribute reduction method. Section 6 summarizes the article.

2. Preliminaries

This section briefly covers some basic content on decision systems, information entropy, intuitionistic fuzzy (IF) sets, and IF relations [3,10,32].

In the article, denote $I = [0, 1]$.

2.1. Decision systems and information entropy

An information system is a pair (U, A), where U is an object set of universe and A is an attribute set. Each attribute $a \in A$ corresponds to an information function $a : U \to V_a$, V_a is the set of function values of a.

Additionally, (U, A) is referred to as a decision system if $A = C \cup D$, where D is the set of decision attributes, C is the set of condition attributes.

For a given information system (U, A), $B \subseteq A$, define

$$ind(B) = \{(x, y) | \forall a \in B, a(x) = a(y)\}. \tag{2.1}$$

Obviously, $ind(B)$ is an equivalence relation, and $ind(B) = \bigcap_{a \in B} ind(\{a\})$.

Furthermore, $U/B = \{X_1, X_2, \cdots, X_s\}$ is the partition generated by $ind(B)$.

For a decision system $(U, C \cup D)$, $Y_k \in U/D = \{Y_1, Y_2, \cdots, Y_r\}$, if $B \subseteq C$, $X_l \in U/B = \{X_1, X_2, \cdots, X_s\}$, then

1) $H(B) = -\sum_{l=1}^{s} p(X_l) \log_2(p(X_l))$ is the information entropy about B, where $p(X_l) = \frac{|X_l|}{|U|}$ is the probability of equivalence class X_l, and $|X_l|$ is the cardinality of X_l;

2) $H(D|B) = -\sum_{l=1}^{s} p(X_l) \sum_{k=1}^{r} p(Y_k|X_l) \log_2(p(Y_k|X_l))$ is the conditional entropy about D under B, where $p(Y_k|X_l) = \frac{|Y_k \cap X_l|}{|X_l|}$;

3) $H(B \cup D) = -\sum_{l=1}^{s} \sum_{k=1}^{r} p(X_l \cap Y_k) \log_2(p(X_l \cap Y_k))$ is the joint entropy about B with D, where $p(X_l \cap Y_k) = \frac{|Y_k \cap X_l|}{|U|}$;

4) $H(B; D) = H(D) - H(D|B)$ is the mutual information about B with D.

Obviously, $H(D|B) + H(B) = H(B \cup D)$.

For a decision system $(U, C \cup D)$, if the relation generated by $B \subseteq C$ is fuzzy, then $H(B) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|R_B(x_i)|}{|U|}$ means the fuzzy information entropy, where $R_B(x_i) = \sum_{x_j \in U} R_B(x_i, x_j)/x_j$, $R_B(x_i, x_j)$ means the similarity degree (or relational degree) between x_i and x_j relative to B.

Similarly, $H(D|B) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|R_B(x_i) \cap R_D(x_i)|}{|R_B(x_i)|}$ is the fuzzy conditional entropy about D under B.

$H(B \cup D) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|R_B(x_i) \cap R_D(x_i)|}{|U|}$ is the fuzzy joint entropy about B with D.

$H(B; D) = H(D) - H(D|B)$ is the fuzzy mutual information about B with D.

2.2. Intuitionistic fuzzy sets and intuitionistic fuzzy relations

For an universe set U, define

$$P = \{<x, \mu_P(x), \nu_P(x)> | x \in U\}. \tag{2.2}$$

Then P is an IF set on U, where $\mu_P : U \to I$, $\nu_P : U \to I$ and $\mu_P(x) + \nu_P(x) \in I$.

The degrees $\mu_P(x)$ and $\nu_P(x)$ mean the membership degree and non-membership degree of x to P, respectively. The hesitancy degree is defined as $\tau_P(x) = 1 - \mu_P(x) - \nu_P(x)$. Moreover, an IF number α is denoted by $\alpha = (\mu_P, \nu_P)$.

In particular, $\tilde{0} = \{<x, 0, 1> | x \in U\}$ and $\tilde{1} = \{<x, 1, 0> | x \in U\}$ are IF empty set and IF universal set, respectively.

While introducing the concept of IF sets, Atanassov also gave some operators to compare them. For two IF sets P_1 and P_2 on U,

(1) $P_2 \supseteq P_1 \Leftrightarrow \forall\ x \in U$, $\mu_{P_2}(x) \geq \mu_{P_1}(x)$ and $\nu_{P_2}(x) \leq \nu_{P_1}(x)$;
(2) $P_2 = P_1 \Leftrightarrow P_2 \subseteq P_1$ and $P_2 \supseteq P_1$;
(3) $P_1 \cap P_2 = \{<x, \mu_{P_1}(x) \wedge \mu_{P_2}(x), \nu_{P_1}(x) \vee \nu_{P_2}(x)> | x \in U\}$;
(4) $P_1 \cup P_2 = \{<x, \mu_{P_1}(x) \vee \mu_{P_2}(x), \nu_{P_1}(x) \wedge \nu_{P_2}(x)> | x \in U\}$;
(5) $P_1^c = \{<x, \nu_{P_1}(x), \mu_{P_1}(x)> | x \in U\}$.

The cardinality of IF set P is expressed as [33,34]:

$$|P| = \sum_{x \in U} \frac{\mu_P(x) + 1 - \nu_P(x)}{2}. \tag{2.3}$$

The IF relation, raised by Bustince and Burillo [35], is expressed as

$$R = \{<(x_1, x_2), \mu_R(x_1, x_2), \nu_R(x_1, x_2)> | (x_1, x_2) \in U \times U\}, \tag{2.4}$$

where $\mu_R(x_1, x_2) + \nu_R(x_1, x_2) \in I$, $\mu_R : U \times U \to I$, $\nu_R : U \times U \to I$.

The degrees $\mu_R(x_1, x_2)$ and $\nu_R(x_1, x_2)$ mean the similarity degree and diversity degree of x_1 to x_2, respectively.

An IF relation can be expressed by matrix $R = (\mu_R(x_1, x_2), \nu_R(x_1, x_2))_{|U| \times |U|}$.

There are some main properties of IF relations [35,36]. For each $x_1, x_2, x_3 \in U$, an IF relation on U is said to be

(1) serial, if $\bigwedge_{x_2 \in U} \mu_R(x_1, x_2) = 1$ and $\bigwedge_{x_2 \in U} \nu_R(x_1, x_2) = 0$;
(2) reflexive, if $\mu_R(x_1, x_1) = 1$, $\nu_R(x_1, x_1) = 0$;
(3) antireflexive, if $\mu_R(x_1, x_1) = 0$, $\nu_R(x_1, x_1) = 1$;
(4) symmetric, if $\mu_R(x_1, x_2) = \mu_R(x_2, x_1)$, $\nu_R(x_1, x_2) = \nu_R(x_2, x_1)$;
(5) antisymmetrical intuitionistic, if $x_1 \neq x_2$ and $\mu_R(x_1, x_2) \neq \mu_R(x_2, x_1)$, $\nu_R(x_1, x_2) \neq \nu_R(x_2, x_1)$, $\tau_R(x_1, x_2) = \tau_R(x_2, x_1)$;
(6) perfect antisymmetrical intuitionistic, if $x_1 \neq x_2$ and $\mu_R(x_1, x_2) > 0$ (or $\mu_R(x_1, x_2) = 0$ and $\nu_R(x_1, x_2) < 1$);
(7) transitive, if $\mu_R(x_1, x_2) \geq \bigvee_{x_3 \in U} (\mu_R(x_1, x_3) \wedge \mu_R(x_3, x_2))$ and $\nu_R(x_1, x_2) \leq \bigwedge_{x_3 \in U} (\nu_R(x_1, x_3) \vee \nu_R(x_3, x_2))$;
(8) C-transitive, if $\mu_R(x_1, x_2) \leq \bigwedge_{x_3 \in U} (\mu_R(x_1, x_3) \vee \mu_R(x_3, x_2))$ and $\nu_R(x_1, x_2) \geq \bigvee_{x_3 \in U} (\nu_R(x_1, x_3) \wedge \nu_R(x_3, x_2))$.

3. Intuitionistic fuzzy dominance relations and intuitionistic fuzzy dominance entropy

3.1. Intuitionistic fuzzy dominance relations

Definition 3.1 *([27])*. Let (U, A) be an information system. (U, A) is called an intuitionistic fuzzy information system (IFIS) if $\forall\, a \in A$, V_a is an IF set. Let $(U, C \cup D)$ be a decision system, $(U, C \cup D)$ is called an intuitionistic fuzzy decision system (IFDS) if $\forall\, a \in C$, V_a is an IF set.

It is remarkable that decision attribute set D divides the universe U into a family of decision classes $U / D = \{D_1, D_2, \cdots, D_l\}$ based on decision labels, each decision class can be represented by an IF set:

$$D_i = \{<x, \mu_{D_i}(x), \nu_{D_i}(x)> | x \in U\}, \text{ where } (\mu_{D_i}(x), \nu_{D_i}(x)) = \begin{cases} (1, 0), & x \in D_i, \\ (0, 1), & x \notin D_i. \end{cases}$$

Below, we present an IF dominance degree between two objects whose attribute values are IF number.

Definition 3.2. For a given IFIS (U, A) and $B \subseteq A$, define

$$R_B^{\mu, \geq}(x_i, x_j) = \frac{1}{\lambda} |\{a \in B | \mu_a(x_i) \geq \mu_a(x_j) \wedge \nu_a(x_i) \leq \nu_a(x_j)\}|, \tag{3.1}$$

$$R_B^{\nu, \preceq}(x_i, x_j) = \frac{1}{\lambda} |\{a \in B | \mu_a(x_i) \geq \mu_a(x_j) \vee \nu_a(x_i) \leq \nu_a(x_j)\}|, \tag{3.2}$$

where λ is a parameter that adjusts $R_B^{\mu, \geq}(x_i, x_j)$ and $R_B^{\nu, \preceq}(x_i, x_j)$ to be in I for any object pair. Here, $\lambda = |A|$.

Let $(R_B^{\mu, \geq}(x_i, x_j), 1 - R_B^{\nu, \preceq}(x_i, x_j))$ be the degree that object x_i dominates x_j under B, where $R_B^{\mu, \geq}(x_i, x_j)$ can be seen as the degree to which the membership degree of x_i dominates that of x_j, $1 - R_B^{\nu, \preceq}(x_i, x_j)$ can be seen as the degree to which the non-membership degree of x_i dominates that of x_j.

Actually, $R_B^{\nu, \preceq}(x_i, x_j)$ means the degree to which the non-membership degree of x_i does not dominate the non-membership degree of x_j. It can be seen that $1 - R_B^{\nu, \preceq}(x_i, x_j) = \frac{|A|}{|A|} - \frac{1}{|A|} |\{a \in B | \mu_a(x_i) \geq \mu_a(x_j) \vee \nu_a(x_i) \leq \nu_a(x_j)\}|$. Since $\{a \in B | \mu_a(x_i) \geq \mu_a(x_j) \vee \nu_a(x_i) \leq \nu_a(x_j)\} \subseteq A$, so $1 - R_B^{\nu, \preceq}(x_i, x_j) = \frac{1}{|A|} |A - \{a \in B | \mu_a(x_i) \geq \mu_a(x_j) \vee \nu_a(x_i) \leq \nu_a(x_j)\}| = \frac{1}{|A|} |\{a \in B | \mu_a(x_i) < \mu_a(x_j) \wedge \nu_a(x_i) > \nu_a(x_j)\}|$, it is clear that $R_B^{\mu, \geq}(x_i, x_j) + (1 - R_B^{\nu, \preceq}(x_i, x_j)) \leq 1$.

Definition 3.3. For an IFIS (U, A), $B \subseteq A$, $U = \{x_1, x_2, \cdots, x_n\}$, an IF dominance relation $R_B^{\geq}$ is a relation matrix:

$$R_B^{\geq} = \begin{pmatrix} (R_B^{\mu, \geq}(x_1, x_1), 1 - R_B^{\nu, \preceq}(x_1, x_1)) & (R_B^{\mu, \geq}(x_1, x_2), 1 - R_B^{\nu, \preceq}(x_1, x_2)) & \cdots & (R_B^{\mu, \geq}(x_1, x_n), 1 - R_B^{\nu, \preceq}(x_1, x_n)) \\ (R_B^{\mu, \geq}(x_2, x_1), 1 - R_B^{\nu, \preceq}(x_2, x_1)) & (R_B^{\mu, \geq}(x_2, x_2), 1 - R_B^{\nu, \preceq}(x_2, x_2)) & \cdots & (R_B^{\mu, \geq}(x_2, x_n), 1 - R_B^{\nu, \preceq}(x_2, x_n)) \\ \cdots & \cdots & \cdots & \cdots \\ (R_B^{\mu, \geq}(x_n, x_1), 1 - R_B^{\nu, \preceq}(x_n, x_1)) & (R_B^{\mu, \geq}(x_n, x_2), 1 - R_B^{\nu, \preceq}(x_n, x_2)) & \cdots & (R_B^{\mu, \geq}(x_n, x_n), 1 - R_B^{\nu, \preceq}(x_n, x_n)) \end{pmatrix}. \tag{3.3}$$

It is easy to see that the memberships and non-memberships of $R_B^{\geq}$ are affected by B. The membership degrees of $R_B^{\geq}$ become larger and the non-membership degrees of $R_B^{\geq}$ become smaller as B becomes larger. In other words, $R_B^{\geq}$ has the following property.

Proposition 3.4. *Suppose $B_2 \subseteq B_1$, then $R_{B_2}^{\geq} \subseteq R_{B_1}^{\geq}$.*

Proposition 3.5. *$R_A^{\geq}$ is reflexive.*

Proof. According to equations (3.1) and (3.2), for any $x_i \in U$, $R_A^{\mu, \geq}(x_i, x_i) = \frac{1}{\lambda} |\{a \in A | \mu_a(x_i) \geq \mu_a(x_i) \wedge \nu_a(x_i) \leq \nu_a(x_i)\}| = \frac{1}{\lambda} |A| = 1$, $1 - R_A^{\nu, \preceq}(x_i, x_i) = 1 - \frac{1}{\lambda} |\{a \in A | \mu_a(x_i) \geq \mu_a(x_i) \vee \nu_a(x_i) \leq \nu_a(x_i)\}| = 1 - \frac{1}{\lambda} |A| = 0$.

So $R_A^{\geq}$ is reflexive. $\square$

Below, we give an example to illustrate the relation matrices induced by attribute subsets.

Example 3.6. For the IFIS (U, A) shown in Table 1, pick $\lambda = |A|$, the relation matrices induced by $B_1 = \{a_1\}$, $B_2 = \{a_1, a_2, a_3\}$ and A can be computed as follows:

Table 1
An IFIS (U, A).

	a_1	a_2	a_3	a_4	a_5
x_1	(0.7, 0.2)	(0.4, 0.3)	(0.5, 0.4)	(0.3, 0.3)	(0.4, 0.4)
x_2	(0.4, 0.4)	(0.6, 0.2)	(0.2, 0.7)	(0.4, 0.6)	(0.7, 0.3)
x_3	(0.6, 0.1)	(0.8, 0.1)	(0.1, 0.7)	(0.2, 0.8)	(0.1, 0.6)
x_4	(0.3, 0.6)	(0.3, 0.7)	(0.6, 0.3)	(0.5, 0.3)	(0.2, 0.6)
x_5	(0.5, 0.3)	(0.7, 0.3)	(0.4, 0.5)	(0.7, 0.1)	(0.9, 0.1)
x_6	(0.4, 0.2)	(0.6, 0.3)	(0.1, 0.1)	(0.7, 0.3)	(0.5, 0.1)
x_7	(0.6, 0.4)	(0.1, 0.2)	(0.8, 0.1)	(0.3, 0.1)	(0.4, 0.5)
x_8	(0.6, 0.3)	(0.1, 0.5)	(0.3, 0.7)	(0.1, 0.6)	(0.3, 0.6)
x_9	(0.2, 0.8)	(0.2, 0.4)	(0.5, 0.3)	(0.2, 0.5)	(0.4, 0.3)

$$
R^{\geq}_{B_1} = \begin{pmatrix}
(0.2,0.8) & (0.2,0.8) & (0,0.8) & (0.2,0.8) & (0.2,0.8) & (0.2,0.8) & (0.2,0.8) & (0.2,0.8) & (0.2,0.8) \\
(0,1) & (0.2,0.8) & (0,1) & (0.2,0.8) & (0,1) & (0,0.8) & (0,0.8) & (0,1) & (0.2,0.8) \\
(0,0.8) & (0.2,0.8) & (0.2,0.8) & (0.2,0.8) & (0.2,0.8) & (0.2,0.8) & (0.2,0.8) & (0.2,0.8) & (0.2,0.8) \\
(0,1) & (0,1) & (0,1) & (0.2,0.8) & (0,1) & (0,1) & (0,1) & (0,1) & (0.2,0.8) \\
(0,1) & (0.2,0.8) & (0,1) & (0.2,0.8) & (0.2,0.8) & (0,0.8) & (0,0.8) & (0,0.8) & (0.2,0.8) \\
(0,0.8) & (0.2,0.8) & (0,1) & (0.2,0.8) & (0,0.8) & (0.2,0.8) & (0,0.8) & (0,0.8) & (0.2,0.8) \\
(0,1) & (0.2,0.8) & (0,0.8) & (0.2,0.8) & (0,0.8) & (0,0.8) & (0.2,0.8) & (0,0.8) & (0.2,0.8) \\
(0,1) & (0.2,0.8) & (0,0.8) & (0.2,0.8) & (0.2,0.8) & (0,0.8) & (0.2,0.8) & (0.2,0.8) & (0.2,0.8) \\
(0,1) & (0,1) & (0,1) & (0,1) & (0,1) & (0,1) & (0,1) & (0,1) & (0.2,0.8)
\end{pmatrix},
$$

$$
R^{\geq}_{B_2} = \begin{pmatrix}
(0.6,0.4) & (0.4,0.6) & (0.2,0.6) & (0.4,0.6) & (0.4,0.4) & (0.2,0.4) & (0.2,0.6) & (0.6,0.4) & (0.4,0.4) \\
(0.2,0.8) & (0.6,0.4) & (0.2,0.8) & (0.4,0.6) & (0,0.8) & (0.2,0.4) & (0.2,0.6) & (0.2,0.6) & (0.4,0.6) \\
(0.2,0.6) & (0.4,0.4) & (0.6,0.4) & (0.4,0.6) & (0.4,0.6) & (0.4,0.4) & (0.4,0.6) & (0.4,0.4) & (0.4,0.6) \\
(0.2,0.8) & (0.2,0.8) & (0.2,0.8) & (0.6,0.4) & (0.2,0.8) & (0,0.8) & (0,0.8) & (0.2,0.6) & (0.4,0.4) \\
(0.2,0.8) & (0.4,0.4) & (0.2,0.8) & (0.4,0.6) & (0.6,0.4) & (0.2,0.4) & (0,0.6) & (0.4,0.4) & (0.4,0.6) \\
(0.2,0.4) & (0.2,0.4) & (0.2,0.8) & (0.4,0.4) & (0,0.4) & (0.6,0.4) & (0,0.4) & (0.2,0.4) & (0.4,0.4) \\
(0.2,0.6) & (0.4,0.4) & (0.2,0.6) & (0.4,0.4) & (0.2,0.4) & (0.2,0.4) & (0.6,0.4) & (0.4,0.4) & (0.4,0.4) \\
(0,1) & (0.4,0.6) & (0.2,0.6) & (0.2,0.6) & (0.2,0.8) & (0,0.6) & (0.2,0.6) & (0.6,0.4) & (0.2,0.8) \\
(0.2,0.8) & (0.2,0.8) & (0.2,0.8) & (0,0.6) & (0.2,0.8) & (0,0.8) & (0,0.8) & (0.4,0.6) & (0.6,0.4)
\end{pmatrix},
$$

$$
R^{\geq}_{A} = \begin{pmatrix}
(1,0) & (0.4,0.4) & (0.6,0.2) & (0.6,0.2) & (0.4,0.4) & (0.2,0.2) & (0.4,0.2) & (1,0) & (0.6,0) \\
(0.4,0.4) & (1,0) & (0.6,0.4) & (0.6,0.4) & (0,0.8) & (0.2,0.2) & (0.4,0.2) & (0.6,0.2) & (0.6,0.2) \\
(0.2,0.6) & (0.4,0.4) & (1,0) & (0.4,0.4) & (0.4,0.6) & (0.4,0.4) & (0.4,0.6) & (0.4,0) & (0.4,0.4) \\
(0.4,0.6) & (0.4,0.6) & (0.6,0.4) & (1,0) & (0.2,0.8) & (0,0.6) & (0,0.6) & (0.4,0.2) & (0.6,0.2) \\
(0.6,0.4) & (0.8,0) & (0.6,0.4) & (0.8,0.2) & (1,0) & (0.6,0) & (0.4,0.2) & (0.8,0) & (0.8,0.2) \\
(0.6,0) & (0.4,0) & (0.6,0.4) & (0.8,0) & (0,0) & (1,0) & (0.2,0) & (0.6,0) & (0.8,0) \\
(0.4,0.2) & (0.4,0.2) & (0.6,0.2) & (0.6,0) & (0.2,0.2) & (0.2,0.2) & (1,0) & (0.8,0) & (0.6,0) \\
(0,0) & (0.4,0.4) & (0.4,0.2) & (0.4,0.4) & (0.2,0.8) & (0,0.6) & (0.2,0.6) & (1,0) & (0.2,0.8) \\
(0.4,0.6) & (0.2,0.4) & (0.6,0.4) & (0.2,0.4) & (0.2,0.8) & (0,0.8) & (0.2,0.6) & (0.8,0.2) & (1,0)
\end{pmatrix}.
$$

A set of fuzzy information granules can be correspondingly induced by a fuzzy binary relation for a given universe [37], and it can be seen as a fuzzy binary granular structure. Similarly, a set of IF information granules can be correspondingly induced by an IF dominance relation, and it can be seen as an IF information structure.

Definition 3.7. For an IFIS (U, A), $B \subseteq A$, the IF information structure generated by $R^{\geq}_B$ is

$$
S(R^{\geq}_B) = (S_{R^{\geq}_B}(x_1), S_{R^{\geq}_B}(x_2), \cdots, S_{R^{\geq}_B}(x_{|U|})), \tag{3.4}
$$

where $S_{R^{\geq}_B}(x_i)$ is also called the IF information granule of x_i relative to $R^{\geq}_B$.

If $R^{\mu,\geq}_B(x_i, x_j) + (1 - R^{\nu,\preceq}_B(x_i, x_j)) = 1$ ($\forall x_i, x_j \in U$), i.e., the IF dominance relation degenerates to the classical fuzzy relation, then the IF information structure can be degenerated to fuzzy information structure proposed by Qian et al. [37].

According to equation (2.3), the cardinality of IF information granule $S_{R^{\geq}_B}(x_i)$ is

$$
|S_{R^{\geq}_B}(x_i)| = \sum_{j=1}^{|U|} \frac{R^{\mu,\geq}_B(x_i, x_j) + 1 - (1 - R^{\nu,\preceq}_B(x_i, x_j))}{2} = \sum_{j=1}^{|U|} \frac{R^{\mu,\geq}_B(x_i, x_j) + R^{\nu,\preceq}_B(x_i, x_j)}{2}. \tag{3.5}
$$

3.2. Intuitionistic fuzzy dominance entropy

We now present several information measures for the discernibility power of an IF information granule. When the induced IF dominance relation $R_B^\geq$ degrades to a fuzzy relation, i.e., $S_{R_B^\geq}(x)$ degrades to a fuzzy information granule, the defined IF dominance entropy degrades to fuzzy information entropy.

Definition 3.8. For a given IFIS (U, A), the IF dominance entropy about $B \subseteq A$ can be expressed as

$$H(B) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|S_{R_B^\geq}(x_i)|}{|U|}. \tag{3.6}$$

When $U/D = \{D_1, D_2, \cdots, D_k, \cdots, D_r\}$, $S_{R_D}(x_i) = \frac{(R_D^\mu(x_i, x_1), R_D^\nu(x_i, x_1))}{x_1} + \cdots + \frac{(R_D^\mu(x_i, x_j), R_D^\nu(x_i, x_j))}{x_j} + \cdots + \frac{(R_D^\mu(x_i, x_{|U|}), R_D^\nu(x_i, x_{|U|}))}{x_{|U|}}$ and

$(R_D^\mu(x_i, x_j), R_D^\nu(x_i, x_j)) = \begin{cases} (1, 0), & x_i, x_j \in D_k \\ (0, 1), & x_i, x_j \notin D_k \end{cases}$, $H(D) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|S_{R_D}(x_i)|}{|U|} = -\sum_{k=1}^{r} \frac{|D_k|}{|U|} \log_2 \frac{|D_k|}{|U|}$.

Theorem 3.9. *For an IFIS (U, A):*
(1) $H(B) \geq 0$ *if $B \subseteq A$;*
(2) $H(B_1) \leq H(B_2)$ *if $B_2 \subseteq B_1 \subseteq A$.*

Proof. (1) Since $\forall x_i \in U$, $|R_B^\geq(x_i)| \leq |U|$, by equation (3.6), we have $H(B) \geq 0$.

(2) Since $B_2 \subseteq B_1$, by Proposition 3.4, we have $R_{B_2}^\geq \subseteq R_{B_1}^\geq$. So $\forall x_i \in U$, $S_{R_{B_2}^\geq}(x_i) \subseteq S_{R_{B_1}^\geq}(x_i)$. Then $|S_{R_{B_2}^\geq}(x_i)| \leq |S_{R_{B_1}^\geq}(x_i)|$. By equation (3.6), we have $H(B_1) \leq H(B_2)$. $\square$

Moreover, if the induced IF dominance relation $R_B^\geq$ is an IF universal relation, that is, $\forall x_i \in U$, $S_{R_B^\geq}(x_i)$ is an IF universal set, then $H(B) = 0$.

We can see from the above theorem that (1) it is non-negative for the IF dominance entropy of an attribute subset; (2) the value of the IF dominance entropy decreases as the attribute subset becomes larger.

Next, the measure in the fuzzy context will be generalized to measure the importance of attributes in IFDSs.

Definition 3.10. For a given IFDS $(U, C \cup D)$ and $B \subseteq C$, define

$$H(D|B) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|S_{R_B^\geq}(x_i) \cap S_{R_D}(x_i)|}{|S_{R_B^\geq}(x_i)|}. \tag{3.7}$$

$H(D|B)$ is called the IF dominance conditional entropy about D under B.

In this article, $H(D|B) = 0$ if $B = \emptyset$.

Theorem 3.11. *For an IFDS $(U, C \cup D)$, $B \subseteq C$, one has*
(1) $0 \leq H(D|B)$.
(2) $R_D \subseteq R_B^\geq \Leftrightarrow H(D) - H(B) = H(D|B)$.

Proof. (1) Due to $\forall x_i \in U$, $S_{R_B^\geq}(x_i) \cap S_{R_D}(x_i) \subseteq S_{R_B^\geq}(x_i)$. So $|S_{R_B^\geq}(x_i) \cap S_{R_D}(x_i)| \leq |S_{R_B^\geq}(x_i)|$.

By equation (3.7), $H(D|B) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|S_{R_B^\geq}(x_i) \cap S_{R_D}(x_i)|}{|S_{R_B^\geq}(x_i)|} \geq -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 1 = 0$.

(2) "$\Rightarrow$" Since $R_D \subseteq R_B^\geq$, then $\forall x_i \in U$, $S_{R_B^\geq}(x_i) \cap S_{R_D}(x_i) = S_{R_D}(x_i)$. By equation (3.7), $H(D|B) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|S_{R_B^\geq}(x_i) \cap S_{R_D}(x_i)|}{|S_{R_B^\geq}(x_i)|} =$

$-\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|S_{R_D}(x_i)| \cdot |U|}{|S_{R_B^\geq}(x_i)| \cdot |U|} = -\frac{1}{|U|} \sum_{i=1}^{|U|} (\log_2 \frac{|S_{R_D}(x_i)|}{|U|} - \log_2 \frac{|S_{R_B^\geq}(x_i)|}{|U|}) = H(D) - H(B)$.

"$\Leftarrow$" By $H(D) - H(B) = H(D|B)$, one has $\forall x_i \in U$, $|S_{R_B^\geq}(x_i) \cap S_{R_D}(x_i)| = |S_{R_D}(x_i)|$. Since $\forall x_i \in U$, $S_{R_B^\geq}(x_i) \cap S_{R_D}(x_i) \subseteq S_{R_D}(x_i)$, then $\forall x_i \in U$, $S_{R_B^\geq}(x_i) \cap S_{R_D}(x_i) = S_{R_D}(x_i)$. So $\forall x_i \in U$, $S_{R_D}(x_i) \subseteq S_{R_B^\geq}(x_i)$. Thus $R_D \subseteq R_B^\geq$. $\square$

It can be seen from the above theorem that (1) the IF dominance conditional entropy is non-negative; (2) the IF dominance conditional entropy and the IF dominance entropy are linearly related if and only if the relation induced by D is finer than the IF dominance relation induced by B.

At the same time, we present the definitions of IF dominance joint entropy and IF dominance mutual information.

Definition 3.12. For a given IFDS $(U, C \cup D)$ and $B \subseteq C$, define

$$H(D \cup B) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|S_{R_B^\geq}(x_i) \cap S_{R_D}(x_i)|}{|U|}. \tag{3.8}$$

$H(D \cup B)$ is called the IF dominance joint entropy about D with B.

Obviously, $H(D \cup B) = H(B \cup D)$.

Theorem 3.13. *For an IFDS $(U, C \cup D)$:*
(1) $H(D \cup B) \geq \max\{H(B), H(D)\}$ *if $B \subseteq C$;*
(2) $H(D \cup B_2) \leq H(D \cup B_1)$ *if $B_1 \subseteq B_2 \subseteq C$.*

Proof. (1) Since $\forall\, x_i \in U$, $S_{R_B^\geq}(x_i) \cap S_{R_D}(x_i) \subseteq S_{R_B^\geq}(x_i)$, we have $|S_{R_B^\geq}(x_i) \cap S_{R_D}(x_i)| \leq |S_{R_B^\geq}(x_i)|$.

By equation (3.8), $H(D \cup B) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|S_{R_B^\geq}(x_i) \cap S_{R_D}(x_i)|}{|U|} \geq -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|S_{R_B^\geq}(x_i)|}{|U|} = H(B)$.

Similarly, $H(D) \leq H(D \cup B)$.

(2) If $B_1 \subseteq B_2$, by Proposition 3.4, one has $R_{B_1}^\geq \subseteq R_{B_2}^\geq$. Then $\forall\, x_i \in U$, $S_{R_{B_1}^\geq}(x_i) \subseteq S_{R_{B_2}^\geq}(x_i)$. So $S_{R_{B_1}^\geq}(x_i) \cap S_{R_D}(x_i) \subseteq S_{R_{B_2}^\geq}(x_i) \cap S_{R_D}(x_i)$, i.e., $|S_{R_{B_1}^\geq}(x_i) \cap S_{R_D}(x_i)| \leq |S_{R_{B_2}^\geq}(x_i) \cap S_{R_D}(x_i)|$. Thus $H(D \cup B_2) \leq H(D \cup B_1)$. $\square$

Theorem 3.13 shows that (1) the IF dominance joint entropy about D with B is not less than any IF dominance entropy about D and B; (2) the IF dominance joint entropy becomes smaller as the conditional attribute subset increases.

Definition 3.14. For an IFDS $(U, C \cup D)$, $B \subseteq C$, define

$$H(D; B) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|S_{R_B^\geq}(x_i)| \cdot |S_{R_D}(x_i)|}{|S_{R_B^\geq}(x_i) \cap S_{R_D}(x_i)| \cdot |U|}. \tag{3.9}$$

$H(D; B)$ is called the IF dominance mutual information about D with B.

Clearly, $H(D; B) = H(B; D)$.
Next, we give a theorem to describe the relationships among IF dominance entropy and its variations.

Theorem 3.15. *For an IFDS $(U, C \cup D)$, $B \subseteq C$, one has*
(1) $H(B \cup D) = H(B) + H(D|B) = H(D) + H(B|D)$;
(2) $H(B; D) = H(B) + H(D) - H(B \cup D)$;
(3) $H(B; D) = H(B) - H(B|D) = H(D) - H(D|B)$.

Proof. (1) According to equations (3.6) and (3.7),

$$H(B) + H(D|B) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|S_{R_B^\geq}(x_i)|}{|U|} - \frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|S_{R_B^\geq}(x_i) \cap S_{R_D}(x_i)|}{|S_{R_B^\geq}(x_i)|}$$

$$= -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \left(\frac{|S_{R_B^\geq}(x_i)|}{|U|} \cdot \frac{|S_{R_B^\geq}(x_i) \cap S_{R_D}(x_i)|}{|S_{R_B^\geq}(x_i)|} \right) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|S_{R_B^\geq}(x_i) \cap S_{R_D}(x_i)|}{|U|} = H(B \cup D).$$

Similarly, $H(D) + H(B|D) = H(B \cup D)$.
Thus $H(B \cup D) = H(B) + H(D|B) = H(D) + H(B|D)$.

(2) According to equations (3.6) and (3.8),

$$H(B) + H(D) - H(B \cup D) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|S_{R_B^\geq}(x_i)|}{|U|} - \frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|S_{R_D}(x_i)|}{|U|} - \left(-\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|S_{R_B^\geq}(x_i) \cap S_{R_D}(x_i)|}{|U|} \right)$$

$$= -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \left(\frac{|S_{R_B^\geq}(x_i)|}{|U|} \cdot \frac{|S_{R_D}(x_i)|}{|U|} \Big/ \frac{|S_{R_B^\geq}(x_i) \cap S_{R_D}(x_i)|}{|U|} \right)$$

$$= -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|S_{R_B^\geq}(x_i)| \cdot |S_{R_D}(x_i)|}{|S_{R_B^\geq}(x_i) \cap S_{R_D}(x_i)| \cdot |U|} = H(B; D).$$

(3) $H(B) - H(B|D) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|S_{R_B^\geq}(x_i)|}{|U|} - \left(-\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|S_{R_D}(x_i) \cap S_{R_B^\geq}(x_i)|}{|S_{R_D}(x_i)|} \right)$

$$= -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \left(\frac{|S_{R_B^\geq}(x_i)|}{|U|} \Big/ \frac{|S_{R_D}(x_i) \cap S_{R_B^\geq}(x_i)|}{|S_{R_D}(x_i)|} \right) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|S_{R_B^\geq}(x_i)| \cdot |S_{R_D}(x_i)|}{|S_{R_B^\geq}(x_i) \cap S_{R_D}(x_i)| \cdot |U|} = H(B; D).$$

Similarly, $H(D) - H(D|B) = H(B; D)$.

Thus $H(B; D) = H(B) - H(B|D) = H(D) - H(D|B)$. $\square$

Example 3.16. For an IFDS $(U, C \cup D)$, where $U/D = \{\{x_1, x_6, x_7\}, \{x_2, x_5, x_8, x_9\}, \{x_3, x_4\}\}$, and (U, C) is shown in Table 1, $C = \{a_1, a_2, a_3, a_4, a_5\}$. Pick $\lambda = |C|$, $B_1 = \{a_1\}$, $B_2 = \{a_1, a_2, a_3\}$, then the defined fuzzy entropies are computed as follows:

$$H(D) = -\sum_{k=1}^{3} \frac{|X_k|}{9} \log_2 \frac{|X_k|}{9} = 1.5305.$$

By equation (3.6), $H(B_1) = -\frac{1}{9} \sum_{i=1}^{9} \log_2 \frac{|S_{R_{B_1}^{\geq}}(x_i)|}{9} = 3.3128$, $H(B_2) = 1.5175$, $H(C) = 0.7747$;

By equation (3.7), $H(D|B_1) = -\frac{1}{9} \sum_{i=1}^{9} \log_2 \frac{|S_{R_{B_1}^{\geq}}(x_i) \cap S_{R_D}(x_i)|}{|S_{R_{B_1}^{\geq}}(x_i)|} = 1.1816$, $H(D|B_2) = 1.2891$, $H(D|C) = 1.2813$;

By equation (3.8), $H(B_1 \cup D) = -\frac{1}{9} \sum_{i=1}^{9} \log_2 \frac{|S_{R_{B_1}^{\geq}}(x_i) \cap S_{R_D}(x_i)|}{9} = 4.4944$, $H(B_2 \cup D) = 2.8066$, $H(C \cup D) = 2.0560$;

By equation (3.9), $H(B_1; D) = -\frac{1}{9} \sum_{i=1}^{9} \log_2 \frac{|S_{R_{B_1}^{\geq}}(x_i)| \cdot |S_{R_D}(x_i)|}{|S_{R_{B_1}^{\geq}}(x_i) \cap S_{R_D}(x_i)| \cdot 9} = 0.3489$, $H(B_2; D) = 0.2414$, $H(C; D) = 0.2492$.

Clearly, $H(B \cup D) = H(B) + H(D|B)$, $H(B; D) = H(B) + H(D) - H(B \cup D)$.

3.3. The influence of parameter λ

It can be seen that the IF dominance relation introduced in Definition 3.2 and Definition 3.3 is crucial for defining the IF dominance entropy. However, there is a parameter λ in the IF dominance relation. In a dataset, if the number of attributes is larger, the membership degrees of objects to the relation are smaller, the non-membership degrees are larger, and the values of IF dominance entropy are larger. Next, we consider the influence of parameter λ on the IF dominance entropy.

We apply the strategy of an increasing natural sequence as the value of parameter λ in Definition 3.2 and Definition 3.3 to deduce the computation of IF dominance entropy. The relationships between IF dominance entropies under different parameters are given as follows.

Theorem 3.17. *For an IFDS $(U, C \cup D)$, $B \subseteq C$, an increasing natural sequence of numbers $\{\lambda_l | l = 1, 2, \cdots, s\}$ is used as the value domain of parameter λ in Definition 3.2. When $\lambda = \lambda_l$ or $\lambda = \lambda_{l+1}$, the IF dominance entropies about B are recorded as $H(B)_l$ and $H(B)_{l+1}$, respectively; the IF dominance conditional entropies about D under B are recorded as $H(D|B)_l$ and $H(D|B)_{l+1}$, respectively; the IF dominance joint entropies about D with B are recorded as $H(B \cup D)_l$ and $H(B \cup D)_{l+1}$, respectively; the IF dominance mutual information about D with B are recorded as $H(B; D)_l$ and $H(B; D)_{l+1}$, respectively. Then*

(1) $H(B)_{l+1} = \log_2 \frac{\lambda_{l+1}}{\lambda_l} + H(B)_l$.

(2) $H(D|B)_{l+1} = H(D|B)_l$.

(3) $H(B \cup D)_{l+1} = \log_2 \frac{\lambda_{l+1}}{\lambda_l} + H(B \cup D)_l$.

(4) $H(B; D)_{l+1} = H(B; D)_l$.

Proof. Let $p_{ij} = |\{a \in B | \mu_a(x_i) \geq \mu_a(x_j) \wedge \nu_a(x_i) \leq \nu_a(x_j)\}|$, $q_{ij} = |\{a \in B | \mu_a(x_i) \geq \mu_a(x_j) \vee \nu_a(x_i) \leq \nu_a(x_j)\}|$.

(1) According to equations (3.1) and (3.2), $\forall x_i, x_j \in U$, when $\lambda = \lambda_l$, $(R_B^{\mu, \geq})_l(x_i, x_j) = \frac{1}{\lambda_l} p_{ij}$, $(R_B^{\nu, \leq})_l(x_i, x_j) = \frac{1}{\lambda_l} q_{ij}$;

when $\lambda = \lambda_{l+1}$, $(R_B^{\mu, \geq})_{l+1}(x_i, x_j) = \frac{1}{\lambda_{l+1}} p_{ij}$, $(R_B^{\nu, \leq})_{l+1}(x_i, x_j) = \frac{1}{\lambda_{l+1}} q_{ij}$. Then

$$(R_B^{\geq})_l(x_i) = \frac{(\frac{1}{\lambda_l} p_{i1}, 1 - \frac{1}{\lambda_l} q_{i1})}{x_1} + \frac{(\frac{1}{\lambda_l} p_{i2}, 1 - \frac{1}{\lambda_l} q_{i2})}{x_2} + \cdots + \frac{(\frac{1}{\lambda_l} p_{i|U|}, 1 - \frac{1}{\lambda_l} q_{i|U|})}{x_{|U|}},$$

$$(R_B^{\geq})_{l+1}(x_i) = \frac{(\frac{1}{\lambda_{l+1}} p_{i1}, 1 - \frac{1}{\lambda_{l+1}} q_{i1})}{x_1} + \frac{(\frac{1}{\lambda_{l+1}} p_{i2}, 1 - \frac{1}{\lambda_{l+1}} q_{i2})}{x_2} + \cdots + \frac{(\frac{1}{\lambda_{l+1}} p_{i|U|}, 1 - \frac{1}{\lambda_{l+1}} q_{i|U|})}{x_{|U|}}.$$

By equation (3.5), $|(R_B^{\geq})_l(x_i)| = \sum_{j=1}^{|U|} \frac{\frac{1}{\lambda_l} p_{ij} + \frac{1}{\lambda_l} q_{ij}}{2}$, $|(R_B^{\geq})_{l+1}(x_i)| = \sum_{j=1}^{|U|} \frac{\frac{1}{\lambda_{l+1}} p_{ij} + \frac{1}{\lambda_{l+1}} q_{ij}}{2}$, i.e.,

$$|S_{(R_B^{\geq})_l}(x_i)| = \sum_{j=1}^{|U|} \frac{\frac{1}{\lambda_l} p_{ij} + \frac{1}{\lambda_l} q_{ij}}{2} = \frac{1}{\lambda_l} \sum_{j=1}^{|U|} \frac{p_{ij} + q_{ij}}{2}, \quad |S_{(R_B^{\geq})_{l+1}}(x_i)| = \sum_{j=1}^{|U|} \frac{\frac{1}{\lambda_{l+1}} p_{ij} + \frac{1}{\lambda_{l+1}} q_{ij}}{2} = \frac{1}{\lambda_{l+1}} \sum_{j=1}^{|U|} \frac{p_{ij} + q_{ij}}{2}.$$

Thus $\lambda_l |S_{(R_B^{\geq})_l}(x_i)| = \lambda_{l+1} |S_{(R_B^{\geq})_{l+1}}(x_i)|$, i.e., $|S_{(R_B^{\geq})_{l+1}}(x_i)| = (\lambda_l / \lambda_{l+1}) |S_{(R_B^{\geq})_l}(x_i)|$.

Based on equation (3.6), $H(B)_{l+1} = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|S_{(R_B^{\geq})_{l+1}}(x_i)|}{|U|} = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{(\lambda_l / \lambda_{l+1}) |S_{(R_B^{\geq})_l}(x_i)|}{|U|}$

$= -\frac{1}{|U|} \sum_{i=1}^{|U|} (\log_2 \frac{\lambda_l}{\lambda_{l+1}} + \log_2 \frac{|S_{(R_B^{\geq})_l}(x_i)|}{|U|}) = \log_2 \frac{\lambda_{l+1}}{\lambda_l} + H(B)_l$.

(2) Suppose that $U/D = \{D_1, \cdots, D_k, \cdots, D_r\}$, $\forall x_i \in U$,

$$S_{R_D}(x_i) = \frac{(R_D^{\mu}(x_i, x_1), R_D^{\nu}(x_i, x_1))}{x_1} + \cdots + \frac{(R_D^{\mu}(x_i, x_j), R_D^{\nu}(x_i, x_j))}{x_j} + \cdots + \frac{(R_D^{\mu}(x_i, x_{|U|}), R_D^{\nu}(x_i, x_{|U|}))}{x_{|U|}}.$$

$\forall x_j \in U$, when $x_i, x_j \in D_k$, $(R_D^\mu(x_i, x_j), R_D^\nu(x_i, x_j)) = (1, 0)$;

when $x_i, x_j \notin D_k$, $(R_D^\mu(x_i, x_j), R_D^\nu(x_i, x_j)) = (0, 1)$.

$\forall x_i, x_j \in U$, when $\lambda = \lambda_l$, $(R_B^{\mu,\geq})_l(x_i, x_j) = \frac{1}{\lambda_l} p_{ij}$, $(R_B^{\nu,\leq})_l(x_i, x_j) = \frac{1}{\lambda_l} q_{ij}$;

when $\lambda = \lambda_{l+1}$, $(R_B^{\mu,\geq})_{l+1}(x_i, x_j) = \frac{1}{\lambda_{l+1}} p_{ij}$, $(R_B^{\nu,\leq})_{l+1}(x_i, x_j) = \frac{1}{\lambda_{l+1}} q_{ij}$. Then

$$S_{(R_B^\geq)_l}(x_i)(x_j) = (\frac{1}{\lambda_l} p_{ij}, 1 - \frac{1}{\lambda_l} q_{ij}), \ S_{(R_B^\geq)_{l+1}}(x_i)(x_j) = (\frac{1}{\lambda_{l+1}} p_{ij}, 1 - \frac{1}{\lambda_{l+1}} q_{ij}).$$

So, when $x_i, x_j \in D_k$, $(S_{(R_B^\geq)_l}(x_i) \cap S_{R_D}(x_i))(x_j) = (\frac{1}{\lambda_l} p_{ij}, 1 - \frac{1}{\lambda_l} q_{ij})$, $(S_{(R_B^\geq)_{l+1}}(x_i) \cap S_{R_D}(x_i))(x_j) = (\frac{1}{\lambda_{l+1}} p_{ij}, 1 - \frac{1}{\lambda_{l+1}} q_{ij})$;

when $x_i, x_j \notin D_k$, $(S_{(R_B^\geq)_l}(x_i) \cap S_{R_D}(x_i))(x_j) = (0, 1)$, $(S_{(R_B^\geq)_{l+1}}(x_i) \cap S_{R_D}(x_i))(x_j) = (0, 1)$.

Thus $\lambda_l |S_{(R_B^\geq)_l}(x_i) \cap S_{R_D}(x_i)| = \lambda_{l+1} |S_{(R_B^\geq)_{l+1}}(x_i) \cap S_{R_D}(x_i)|$, i.e., $|S_{(R_B^\geq)_{l+1}}(x_i) \cap S_{R_D}(x_i)| = (\lambda_l / \lambda_{l+1}) |S_{(R_B^\geq)_l}(x_i) \cap S_{R_D}(x_i)|$.

According to (1), we have

$$|S_{(R_B^\geq)_{l+1}}(x_i)| = (\lambda_l / \lambda_{l+1}) |S_{(R_B^\geq)_l}(x_i)|.$$

By equation (3.7), $H(D|B)_{l+1} = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|S_{(R_B^\geq)_{l+1}}(x_i) \cap S_{R_D}(x_i)|}{|S_{(R_B^\geq)_{l+1}}(x_i)|} = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{(\lambda_l / \lambda_{l+1}) |S_{(R_B^\geq)_l}(x_i) \cap S_{R_D}(x_i)|}{(\lambda_l / \lambda_{l+1}) |S_{(R_B^\geq)_l}(x_i)|}$

$= -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|S_{(R_B^\geq)_l}(x_i) \cap S_{R_D}(x_i)|}{|S_{(R_B^\geq)_l}(x_i)|} = H(D|B)_l$.

(3) According to Theorem 3.15, one has

$H(B \cup D)_{l+1} = H(D|B)_{l+1} + H(B)_{l+1} = H(D|B)_l + \log_2 \frac{\lambda_{l+1}}{\lambda_l} + H(B)_l = \log_2 \frac{\lambda_{l+1}}{\lambda_l} + H(B \cup D)_l$.

(4) According to Theorem 3.15, one has

$H(B; D)_{l+1} = H(D) - H(D|B)_{l+1} = H(D) - H(D|B)_l = H(B; D)_l$. $\quad \square$

Theorem 3.17 shows that the parameter λ has no effect on IF dominance mutual information and IF dominance conditional entropy, the IF dominance entropy under parameter λ_{l+1} and the IF dominance entropy under parameter λ_l display a linearity relevance, the IF dominance joint entropy under parameter λ_{l+1} and the IF dominance joint entropy under parameter λ_l display a linearity relevance.

Example 3.18. (Continuation of Example 3.16) Pick $\lambda = |C|$, $\lambda_1 = |C| + 1$, we have

$H(B_1)_\lambda = 3.3128$, $H(B_2)_\lambda = 1.5175$, $H(C)_\lambda = 0.7747$;

$H(D|B_1)_\lambda = 1.1816$, $H(D|B_2)_\lambda = 1.2891$, $H(D|C)_\lambda = 1.2813$;

$H(D \cup B_1)_\lambda = 4.4944$, $H(D \cup B_2)_\lambda = 2.8066$, $H(D \cup C)_\lambda = 2.0560$;

$H(D; B_1)_\lambda = 0.3489$, $H(D; B_2)_\lambda = 0.2414$, $H(D; C)_\lambda = 0.2492$.

$H(B_1)_{\lambda+1} = 3.5758$, $H(B_2)_{\lambda+1} = 1.7805$, $H(C)_{\lambda+1} = 1.0377$;

$H(D|B_1)_{\lambda+1} = 1.1816$, $H(D|B_2)_{\lambda+1} = 1.2891$, $H(D|C)_{\lambda+1} = 1.2813$;

$H(D \cup B_1)_{\lambda+1} = 4.7574$, $H(D \cup B_2)_{\lambda+1} = 3.0696$, $H(D \cup C)_{\lambda+1} = 2.3190$;

$H(D; B_1)_{\lambda+1} = 0.3489$, $H(D; B_2)_{\lambda+1} = 0.2414$, $H(D; C)_{\lambda+1} = 0.2492$.

Since $\log_2 \frac{\lambda+1}{\lambda} = 0.2630$, one can see that

$H(B)_{\lambda+1} = \log_2 \frac{\lambda+1}{\lambda} + H(B)_\lambda$, $H(D|B)_{\lambda+1} = H(D|B)_\lambda$,

$H(B \cup D)_{\lambda+1} = \log_2 \frac{\lambda+1}{\lambda} + H(B \cup D)_\lambda$, $H(B; D)_{\lambda+1} = H(B; D)_\lambda$.

According to Theorem 3.17, the following corollary is easily obtained.

Corollary 3.19. *For an IFDS $(U, C \cup D)$ and $\{\lambda_l | l = 1, 2, \cdots, s\}$, $a \in C - B$, where $B \subseteq C$, we have*

(1) $H(B)_{l+1} - H(\{a\} \cup B)_{l+1} = H(B)_l - H(\{a\} \cup B)_l$.

(2) $H(D|B)_{l+1} - H(D|\{a\} \cup B)_{l+1} = H(D|B)_l - H(D|\{a\} \cup B)_l$.

(3) $H(B \cup D)_{l+1} - H((\{a\} \cup B) \cup D)_{l+1} = H(B \cup D)_l - H((\{a\} \cup B) \cup D)_l$.

(4) $H(B; D)_{l+1} - H(\{a\} \cup B; D)_{l+1} = H(B; D)_l - H(\{a\} \cup B; D)_l$.

4. Attribute reduction framework in intuitionistic fuzzy information systems

4.1. Attribute reduction framework

In attribute reduction, an attribute subset in IFISs may have the same observation of discernibility power as that of the raw attribute set, which indicates that adding an attribute subset (or a new attribute) will not increase the discernibility power. From this point of view, the relative reduction of IFDSs is given by Tan et al. [31].

Definition 4.1 *([31]).* For an IFDS $(U, C \cup D)$, $B \subseteq C$ is a relative reduct of $(U, C \cup D)$ iff

1. $H(D|C) = H(D|B)$;

2. $\forall b \in B$, $H(D|B) \neq H(D|(B - \{b\}))$.

It can be seen from the above definition that each attribute in the relative reduct is jointly sufficient and independently necessary. That is to say, the relative reduct is the minimal subset with the original entropy of the IFDS.

However, this attribute reduction method only considered the relevance (significance) of an attribute to the decision one by one, ignoring the redundancies produced by selected attributes. Heretofore, Wang et al. [21] have proposed the maximizing independent classification information method to reduce attributes. Based on the idea of maximizing independent classification information, we construct an attribute reduction method by using IF dominance mutual information for IFDSs.

Before that, we give Algorithm 1 for calculating IF dominance mutual information.

Algorithm 1 Algorithm for calculating IF dominance mutual information.

Input: IFDS $(U, C \cup D)$, $B \subseteq C$.
Output: IF dominance mutual information $H(B; D)$.
1: Initialize $H(D|B) = 0$, $H(D) = 0$, $R_B^{\mu,\geq} = zeros(|U|)$, $R_B^{\nu,\leq} = zeros(|U|)$;
2: Compute $U/D = \{D_1, D_2, \cdots, D_r\}$;
3: **for** each $a \in B$ **do**
4:     **for** each $x_i \in U$ **do**
5:         **for** each $x_j \in U$ **do**
6:             Let $p = 0$, $q = 0$;
7:             **if** $\mu_a(x_i) \geq \mu_a(x_j)$ and $\nu_a(x_i) \leq \nu_a(x_j)$ **then**
8:                 $p = p + 1$;
9:                 $R_a^{\mu,\geq}(x_i, x_j) = p/|C|$;
10:            **end if**
11:            **if** $\mu_a(x_i) \geq \mu_a(x_j)$ or $\nu_a(x_i) \leq \nu_a(x_j)$ **then**
12:                $q = q + 1$;
13:                $R_a^{\nu,\leq}(x_i, x_j) = q/|C|$;
14:            **end if**
15:        **end for**
16:    **end for**
17:    $R_B^{\mu,\geq}(x_i, x_j) = R_B^{\mu,\geq}(x_i, x_j) + R_a^{\mu,\geq}(x_i, x_j)$, $R_B^{\nu,\leq}(x_i, x_j) = R_B^{\nu,\leq}(x_i, x_j) + R_a^{\nu,\leq}(x_i, x_j)$;
18: **end for**
19: **for** each $D_k \in U/D$ **do**
20:    **for** each $x_i \in U$ **do**
21:        Let $s = 0$, $t = 0$;
22:        **for** each $x_j \in U$ **do**
23:            $s = s + R_B^{\mu,\geq}(x_i, x_j) + R_B^{\nu,\leq}(x_i, x_j)$;
24:            **if** $x_i, x_j \in D_k$ **then**
25:                $t = t + R_B^{\mu,\geq}(x_i, x_j) + R_B^{\nu,\leq}(x_i, x_j)$;
26:            **end if**
27:        **end for**
28:    **end for**
29:    $H(D|B) = H(D|B) + (-(1/|U|)\log_2(t/s))$; $H(D) = H(D) + (-(|D_k|/|U|)\log_2(|D_k|/|U|))$;
30: **end for**
31: $H(B; D) = H(D) - H(D|B)$;
32: **return** $H(B; D)$.

Next, we introduce two kinds of the extended IF dominance mutual information.

Definition 4.2. For an IFDS $(U, C \cup D)$, $B_1, B_2 \subseteq C$, $B_1 \cap B_2 = \emptyset$, the joint IF dominance mutual information about D with B_1 and B_2 can be defined as

$$H(B_1, B_2; D) = H(B_1; D) + H(B_2; D) - H(B_1 \cup B_2; D). \tag{4.1}$$

Obviously, $H(B_1, B_2; D) = H(B_2, B_1; D)$.

Definition 4.3. For an IFDS $(U, C \cup D)$, $B_1, B_2 \subseteq C$, $B_1 \cap B_2 = \emptyset$, the conditional IF dominance mutual information about D with B_1 under B_2 can be defined as

$$H(B_1 | B_2; D) = H(B_1 \cup B_2; D) - H(B_2; D). \tag{4.2}$$

Since the parameter λ has no effect on IF dominance conditional entropy and IF dominance mutual information, the parameter λ has no effect on joint IF dominance mutual information and conditional IF dominance mutual information.

From Definition 4.2 and Definition 4.3, the following theorem is easily obtained.

Theorem 4.4. *For an IFDS* $(U, C \cup D)$*, if* $B_1 \cap B_2 = \emptyset$*, where* $B_1, B_2 \subseteq C$*, then*

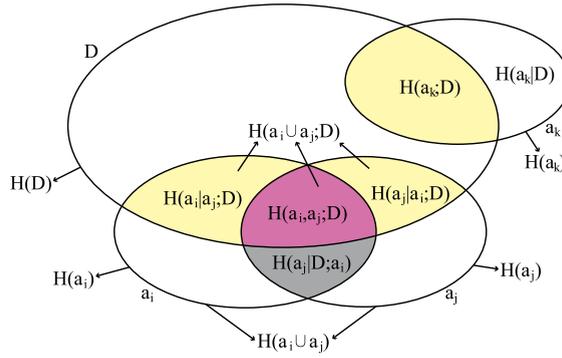$$H(B_1; D) = H(B_1 | B_2; D) + H(B_1, B_2; D). \tag{4.3}$$

Fig. 1. The relationship among IF dominance entropy and its variations.

Fig. 1 shows the relationship among IF dominance entropy and its variations. If there is a reduced (selected) attribute a_i, the relevance of attribute a_i to decision D can be expressed by the IF dominance mutual information about a_i with D, i.e., $H(a_i; D)$. For a candidate attribute (attribute to be reduced) a_j, the relevance of attribute a_j to decision D is $H(a_j; D)$. However, it can be seen from Fig. 1 that there exists redundant information between a_i and a_j, i.e., $H(a_i, a_j; D)$. Based on this, the independent classification information (ICI) of a_i and a_j relative to D can be represented by the conditional IF dominance mutual information, i.e., $ICI(a_i; a_j; D) = H(a_i|a_j; D) + H(a_j|a_i; D)$. Assume that the candidate attribute is a_k, and there is no redundant information between a_i and a_k, then ICI of a_i and a_k relative to D is $ICI(a_i; a_k; D) = H(a_i|a_k; D) + H(a_k|a_i; D) = H(a_i; D) + H(a_k; D)$.

For an IFDS $(U, C \cup D)$, $a_j \in C - B$ is a candidate attribute, where $B = \{a_1, a_2, \cdots, a_{j-1}\}$ is the reduct set. Then ICI of a_j and B relative to D can be defined as

$$\sum_{a_i \in B} ICI(a_i; a_j; D) = \sum_{a_i \in B} (H(a_i|a_j; D) + H(a_j|a_i; D)). \tag{4.4}$$

According to the idea of max-relevance and max-independent [21], the attribute reduction criterion can be designed as below.

Definition 4.5. For an IFDS $(U, C \cup D)$, $a_j \in C - B$ is a candidate attribute, where $B = \{a_1, a_2, \cdots, a_{j-1}\}$ is the reduct set. Define

$$Sig(a_j, D, B) = H(a_j; D) + \sum_{a_i \in B} ICI(a_i; a_j; D). \tag{4.5}$$

$Sig(a_j, D, B)$ is called the significance of a_j for D relative to B. If $B = \emptyset$, then $Sig(a_j, D, B) = H(a_j; D)$.

The criterion for reducing attributes is to find the attribute a_j that maximizes significance $Sig(a_j, B, D)$, when $Sig(a_j, B, D)$ is bigger, it means that a_j is more important for D relative to the reduct set B.

Obviously,

$$
\begin{aligned}
Sig(a_j, D, B) &= H(a_j; D) + \sum_{a_i \in B} ICI(a_i; a_j; D) = H(a_j; D) + \sum_{a_i \in B} (H(a_i|a_j; D) + H(a_j|a_i; D)) \\
&= H(a_j; D) + \sum_{a_i \in B} (H(a_i \cup a_j; D) - H(a_j; D) + H(a_j \cup a_i; D) - H(a_i; D)) \\
&= (1 - |B|)H(a_j; D) + \sum_{a_i \in B} (2H(a_i \cup a_j; D) - H(a_i; D)).
\end{aligned}
\tag{4.6}
$$

Next, we give a heuristic attribute reduction algorithm for IFDSs in Algorithm 2.

Algorithm 2 Heuristic attribute reduction algorithm for IFDSs.

Input: IFDS $(U, C \cup D)$.
Output: An attribute reduction subset B.
1: Initialize $B = \emptyset$, $H(B; D) = 0$;
2: Compute $H(C; D)$ on the basis of Algorithm 1;
3: **while** $|H(B; D) - H(C; D)| > \alpha$ **do**
4:     **for** each $b \in C - B$ **do**
5:         Compute $Sig(b, D, B)$ by equation (4.6);
6:     **end for**
7:     Find b' maximizing $Sig(b, D, B)$ and the front attribute is selected if the maximum value corresponds to multiple attributes;
8:     $B = B \cup \{b'\}$;
9:     Compute $H(B; D)$ on the basis of Algorithm 1;
10: **end while**
11: **return** B.

Table 2
Medical evaluation system.

U	a_1	a_2	a_3	a_4	a_5	a_6	D
x_1	(0.9,0.1)	(0.5,0.0)	(0.9,0.0)	(0.6,0.1)	(0.4,0.5)	(0.5,0.4)	A
x_2	(0.4,0.5)	(0.2,0.7)	(0.1,0.6)	(0.3,0.6)	(0.5,0.0)	(0.4,0.1)	C
x_3	(0.5,0.3)	(0.6,0.3)	(0.6,0.2)	(0.5,0.4)	(0.3,0.5)	(0.5,0.3)	B
x_4	(0.5,0.4)	(0.5,0.4)	(0.6,0.3)	(0.6,0.4)	(0.4,0.4)	(0.7,0.1)	B
x_5	(0.7,0.2)	(0.6,0.1)	(0.7,0.1)	(0.5,0.1)	(0.8,0.1)	(0.7,0.2)	A
x_6	(0.4,0.1)	(0.3,0.5)	(0.3,0.3)	(0.2,0.6)	(0.2,0.3)	(0.3,0.0)	C
x_7	(0.6,0.2)	(0.6,0.3)	(0.6,0.4)	(0.8,0.2)	(0.3,0.4)	(0.5,0.3)	B
x_8	(0.5,0.1)	(0.5,0.3)	(0.6,0.3)	(0.4,0.3)	(0.5,0.3)	(0.4,0.4)	B
x_9	(0.2,0.5)	(0.2,0.6)	(0.4,0.5)	(0.1,0.8)	(0.1,0.2)	(0.1,0.3)	C
x_{10}	(0.9,0.0)	(0.7,0.1)	(0.8,0.1)	(0.9,0.0)	(0.1,0.8)	(0.8,0.1)	A

Strictly speaking, the termination condition in Algorithm 2 should be $|H(B; D) - H(C; D)| > 0$, but in most cases, $H(B; D)$ and $H(C; D)$ are difficult to equal because of the existences of imprecision and inaccuracy. Moreover, a large number of experiments show that there is a slight difference between the value $H(B; D)$ and the value $H(C; D)$ when a currently better attribute subset B is obtained in attribute set C. In view of this reason (problem), we relax the condition of attribute reduction and consider the soft attribute subset reduction: $|H(B; D) - H(C; D)| > \alpha$, where α is a threshold to allow the possibility of a near miss for the search strategy.

In Algorithm 1, the time complexity is not more than $O(|U|^2|C| + |U/D||U|^2)$. In Algorithm 2, the complexity of step 5 for computing the significance can be done in $O(|U|^2|C|^2)$. Hence, steps 3-10 can be done within $O(|B|(|U|^2|C| + |U/D||U|^2 + |U|^2|C|^2))$. To summarize, the time complexity of Algorithm 2 is not more than $O(|U|^2|C|^2|B|)$.

4.2. Illustrative example

The attribute reduction method in IFISs, which can be used to directly reduce attributes of IF data, has been presented above. It is worth noting that an increasing number of real-world problems are being described in terms of IF numbers. IF set is introduced into data environment analysis to form IF data environment analysis [38]. IF neural network is generated by combining IF set and neural network [39]. IF set is used to describe IF time series generated by time series data [40]. An IF data-driven product ranking model is established based on emotion analysis and IF set [41]. Additionally, multi-attribute decision-making methods are constructed in various fields by using IF information [42].

Furthermore, incomplete data is prevalent in reality due to measurement failure, omission, and storage loss. Processing incomplete data is a crucial task. IF sets offer an alternative approach for handling incomplete data. Membership and non-membership can be assigned to missing values to realize the filling of missing values, which can enhance the similarity of samples of the same category and the difference of samples of different categories, so that the problem of incomplete data can be dealt with more effectively and accurately.

An example is given below to illustrate the method proposed in the article.

Example 4.6. Consider a real case of medical evaluation system from [43], represented in Table 2, where $U = \{x_1, x_2, \cdots, x_{10}\}$ represents ten hospitals from different regions. $C = \{a_1, a_2, a_3, a_4, a_5, a_6\}$ is composed of six evaluation indexes, where $a_i (i = 1, 2, 3, 4, 5, 6)$ represents equipment, service level, technical level, management level, treatment direction, and hospital scale, respectively. The value of the decision attribute A, B, C represents the final level in this evaluation. We take the value $\alpha = 0.17$ in Algorithm 2 for illustration.

First, initialize the set of attribute reduction to the empty set, i.e., $B = \emptyset$.

According to Algorithm 1, $H(C; D) = 0.3278$.

For each attribute a_i in C, calculate the significance $Sig(a_i, D, \emptyset)$:

$Sig(a_1, D, \emptyset) = H(a_1; D) = 0.4077$, $Sig(a_2, D, \emptyset) = 0.4761$, $Sig(a_3, D, \emptyset) = 0.5029$,

$Sig(a_4, D, \emptyset) = 0.4403$, $Sig(a_5, D, \emptyset) = 0.2588$, $Sig(a_6, D, \emptyset) = 0.0988$.

Obviously, $Sig(a_3, D, \emptyset)$ is the maximum, so attribute a_3 is selected to add to set B, i.e., $B = \{a_3\}$. And $H(B; D) = 0.5029$.

Then, for each attribute a_i in $C - \{a_3\}$, calculate the significance $Sig(a_i, D, \{a_3\})$:

$Sig(a_1, D, \{a_3\}) = 2H(a_1 \cup a_3; D) - H(a_3; D) = 0.4251$, $Sig(a_2, D, \{a_3\}) = 0.5059$,

$Sig(a_4, D, \{a_3\}) = 0.4699$, $Sig(a_5, D, \{a_3\}) = 0.0879$, $Sig(a_6, D, \{a_3\}) = 0.0974$.

$Sig(a_2, D, \{a_3\})$ is the maximum, so attribute a_2 is added to set B, i.e., $B = \{a_2, a_3\}$. And $H(B; D) = 0.5044$.

Next, for each attribute a_i in $C - \{a_2, a_3\}$, calculate the significance $Sig(a_i, D, \{a_2, a_3\})$:

$Sig(a_1, D, \{a_2, a_3\}) = 2H(a_1 \cup a_2; D) - H(a_2; D) + 2H(a_1 \cup a_3; D) - H(a_3; D) - H(a_1; D) = 0.4264$,

$Sig(a_4, D, \{a_2, a_3\}) = 0.4949$, $Sig(a_5, D, \{a_2, a_3\}) = -0.0702$, $Sig(a_6, D, \{a_2, a_3\}) = 0.0774$.

$Sig(a_4, D, \{a_2, a_3\})$ is the maximum, so attribute a_4 is selected to add to set B, i.e., $B = \{a_2, a_3, a_4\}$. And $H(B; D) = 0.4912$.

At this moment, $|H(B; D) - H(C; D)| = 0.4912 - 0.3278 = 0.1634 < 0.17 = \alpha$, the algorithm terminates.

So, attribute reduction subset $B = \{a_2, a_3, a_4\}$.

Table 3
Description of data.

No.	Datasets	Objects	Attributes	Class
1	wine	178	13+1	3
2	heart	270	13+1	2
3	australian	690	14+1	2
4	segment	2310	18+1	7
5	waveform	5000	21+1	3
6	wdbc	569	30+1	2
7	ionosphere	351	33+1	2
8	soner	208	60+1	2
9	hillvalley	1212	100+1	2
10	musk2	707	166+1	2
11	featMIAS	322	280+1	4
12	Yale	165	1024+1	15
13	ORL	400	1024+1	40
14	colon	62	2000+1	2
15	srbct	83	2308+1	4
16	web	149	2556+1	5
17	arcene	200	10000+1	2
18	orlraws10P	100	10304+1	10
19	CLL_SUB_111	111	11340+1	3
20	MLL	72	12582+1	3
21	GLI_85	85	22283+1	2

5. Experimental analysis

This part presents some experiments to evaluate the performance of the proposed method.

5.1. Experiment preparation

We compare the raised method (IFDMI) based on IF dominance mutual information with several existing reduction methods: the fuzzy boundary region-based method (FBR) [44], the fuzzy positive region-based method (FPR) [45], the IF positive region-based method (IFPR) [30] and the IF conditional entropy-based method (IFCE) [31]. The datasets, which come from the UCI repository of machine learning databases, the ASU feature selection datasets and the Keng Ridge Bio-medical (KRBM) Data Set Repository, employed in this part, are summarized in Table 3.

Before the experiments, we need to pre-process the above datasets to form the IF datasets. For a decision system $(U, C \cup \{d\})$ formed by a dataset, $\forall\, x \in U$ and $a \in C$, fuzzify the information values:

$$(a(x))' = \frac{a(x) - \min_i a(x_i)}{\max_i a(x_i) - \min_i a(x_i)}.$$

It is known that each attribute of an information system can generate a fuzzy relation. In a generated fuzzy information system, the fuzzy similarity relation between objects x_i and x_j under attribute a can be calculated by:

$$R_a(x_i, x_j) = 1 - |(a(x_i))' - (a(x_j))'|.$$

Each attribute in a fuzzy information system corresponds to a fuzzy set on the object set. By combining the fuzzy set formed by attribute a and the fuzzy relation induced by this attribute, we can construct a fuzzy rough set:

$$\underline{R_a}(a)(x_i) = \inf_{x_j \in U} ((1 - R_a(x_i, x_j)) \vee a(x_j)), \quad \overline{R_a}(a)(x_i) = \sup_{x_j \in U} (R_a(x_i, x_j) \wedge a(x_j)).$$

In rough sets, the positive region is the set of elements in U that definitely belong to subset X of U based on knowledge R, while the negative region is the set of elements in U that definitely do not belong to subset X based on knowledge R. In fuzzy rough sets, the positive region refers to the degree to which an element x_i must belong to fuzzy set a, and the negative region refers to the degree to which an element x_i must not belong to fuzzy set a. Thus, we can construct an IF set a on the object set U:

$$\mu_a(x) = \underline{R_a}(a)(x), \quad \nu_a(x) = 1 - \overline{R_a}(a)(x),$$

where $\nu_a(x)$ and $\mu_a(x)$ mean the non-membership degree and membership degree of x to a, respectively, and $\mu_a(x) + \nu_a(x) \in I$, clearly.

In this way, each conditional information value $a(x)$ in original decision system $(U, C \cup \{d\})$ is transformed into IF information value $(\mu_a(x), \nu_a(x))$.

The parameter α is used to terminate the main loop in Algorithm 2. For a given dataset, generally speaking, the number of the selected attributes increases if the value of the parameter α decreases. As suggested by Wang et al. [46], the parameter α is usually fixed at 0.01 for high-dimensional datasets and fixed at 0.001 for low-dimensional datasets.

The attribute reduction subsets of all datasets can be obtained using different attribute reduction methods. The corresponding classification accuracies of these subsets are then calculated by classifiers. In this part, the k-nearest-neighbor rule (KNN, K = 3) and the tree-based J48 (C4.5) classifiers are used to check the classification performances of these attribute reduction subsets, and the 10-fold cross-validation technique is adopted.

Table 4
The attribute numbers of reduct sets.

| Datasets	FBR	FPR	IFPR	IFCE	IFDMI
wine	9	11	6	**5**	6
heart	11	11	8	9	**6**
australian	10	10	**8**	9	**8**
segment	9	12	8	**6**	11
waveform	14	15	**11**	12	14
wdbc	24	21	17	19	**13**
ionosphere	11	27	28	**3**	10
soner	20	17	**13**	19	25
hillvalley	24	**14**	16	21	16
musk2	20	26	21	20	**13**
featMIAS	28	31	23	20	**17**
Yale	219	203	**67**	146	83
ORL	154	172	71	135	**58**
colon	8	7	6	8	**5**
srbct	216	181	67	40	**31**
web	**29**	43	132	34	38
arcene	22	19	16	15	**12**
orlraws10P	12	16	14	11	**8**
CLL_SUB_111	**8**	12	24	19	15
MLL	**6**	8	7	**6**	6
GLI_85	7	**5**	12	10	7
average	41.00	41.00	27.38	27.00	**19.14**
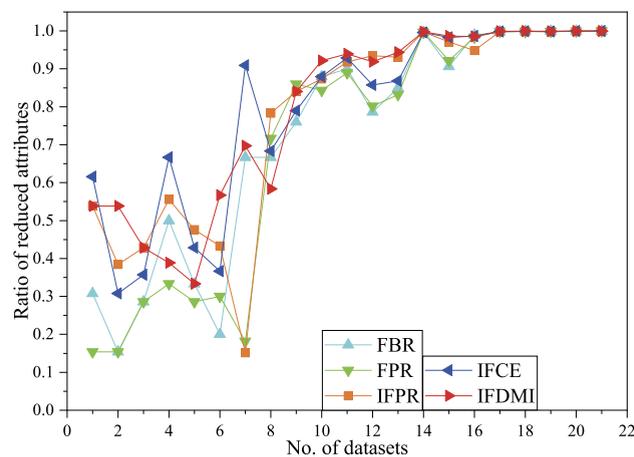
Fig. 2. The ratios of reduced attributes for each dataset.

5.2. Results and analysis

Table 4 displays the attribute numbers of reduct sets using the proposed method and the four compared methods for each dataset. The table shows that all five methods can decrease the size of attribute subsets, indicating their effectiveness in reducing redundant attributes. This effect is particularly noticeable on high-dimensional datasets. Such as MLL and GLI_85 datasets, the reduction obtained by all methods is at least 99.94% lower in dimensionality and can even reach 99.98%. As to orlraws10P dataset, method IFDMI reduces attribute set to 8 attributes while the other four methods reduce to at least 16 attributes. In the reduction of low dimensional datasets, it is more obvious that method IFDMI is better than the other four methods, such as wdbc and musk2 datasets. The attribute number of reduct sets on average by method IFDMI is 19.14, which is smaller than those by the other four methods. Overall, the proposed method is the most effective of the five methods. Moreover, the reduction rates of attribute sets of the five methods are drawn in Fig. 2, where the ordinate is the ratio of attribute numbers of reduced attributes and original attribute set on each dataset and the abscissa means the sequence number for 21 datasets. Although method IFCE can get the smallest reduction on wine, segment and ionosphere datasets, the curve depicted by method IFDMI is in most cases higher than those of the other four methods. Therefore, when it comes to reducing the size of data, the proposed method outperforms the four compared methods.

Table 5 and Table 6 display the classification performances of all five methods tested by KNN (K = 3) and C4.5, respectively. The highest classification accuracies in all reduct sets are shown in bold. It can be observed that, in most cases, the classification accuracies of all five methods can be improved or maintained when compared with the raw datasets. The improvement is even significant in many cases. For example, on CLL_SUB_111 dataset with respect to KNN, the best accuracy obtained by method IFDMI

Table 5
The classification accuracies of reduct sets on KNN.

Datasets	Original data	FBR	FPR	IFPR	IFCE	IFDMI
wine	94.94 ± 4.12	**97.75 ± 2.76**	96.07 ± 3.75	94.38 ± 4.87	96.07 ± 4.37	94.94 ± 5.24
heart	80.00 ± 23.65	80.74 ± 22.54	80.37 ± 23.04	**84.44 ± 21.56**	80.00 ± 23.22	83.70 ± 19.98
australian	85.22 ± 19.07	82.46 ± 21.53	84.49 ± 19.09	82.75 ± 21.72	85.22 ± 19.42	**87.39 ± 20.64**
segment	96.06 ± 1.38	96.10 ± 1.43	96.10 ± 1.38	95.93 ± 1.53	95.37 ± 1.63	**96.67 ± 1.24**
waveform	80.24 ± 15.43	76.88 ± 18.69	79.28 ± 16.62	78.10 ± 16.78	**81.22 ± 14.64**	77.53 ± 18.63
wdbc	96.84 ± 4.75	96.66 ± 4.98	**97.19 ± 4.92**	96.13 ± 5.51	97.01 ± 4.80	95.96 ± 5.33
ionosphere	84.90 ± 15.27	87.18 ± 12.69	87.18 ± 13.85	83.76 ± 15.84	87.46 ± 14.06	**88.89 ± 11.67**
soner	86.06 ± 15.99	**87.98 ± 15.83**	85.58 ± 17.26	86.06 ± 18.54	85.07 ± 19.34	85.10 ± 17.58
hillvalley	54.70 ± 47.53	52.64 ± 48.87	54.46 ± 46.59	54.21 ± 47.72	53.47 ± 48.19	**55.86 ± 45.85**
musk2	92.65 ± 11.39	91.65 ± 8.47	90.66 ± 11.54	91.94 ± 10.90	88.40 ± 13.16	**92.93 ± 10.36**
featMIAS	71.43 ± 17.53	63.98 ± 19.34	63.04 ± 20.16	57.76 ± 24.13	62.42 ± 20.37	**71.74 ± 17.84**
Yale	63.03 ± 6.14	52.73 ± 7.23	52.12 ± 7.42	40.61 ± 8.80	**55.15 ± 6.97**	43.64 ± 8.14
ORL	90.00 ± 0.95	76.75 ± 1.69	79.50 ± 1.47	80.75 ± 1.44	82.50 ± 1.23	**83.25 ± 1.32**
colon	75.81 ± 27.15	82.26 ± 19.72	83.87 ± 17.59	79.03 ± 21.31	**87.10 ± 17.06**	83.87 ± 25.56
srbct	80.72 ± 12.69	79.52 ± 12.69	84.34 ± 11.51	78.31 ± 13.09	**100.00 ± 0.00**	86.75 ± 8.75
web	38.26 ± 25.41	53.69 ± 21.69	55.03 ± 20.55	34.23 ± 26.98	49.66 ± 20.42	**57.05 ± 18.41**
arcene	87.00 ± 18.95	74.00 ± 31.07	72.00 ± 32.07	76.50 ± 25.59	80.50 ± 24.26	**82.50 ± 23.93**
orlraws10P	95.00 ± 2.12	90.00 ± 3.02	**94.00 ± 2.76**	86.00 ± 3.66	93.00 ± 2.57	93.00 ± 2.44
CLL_SUB_111	56.76 ± 30.17	67.57 ± 26.21	71.17 ± 20.86	66.67 ± 25.22	66.67 ± 25.81	**72.07 ± 23.43**
MLL	81.94 ± 13.44	**95.83 ± 5.54**	86.11 ± 12.23	88.89 ± 9.79	89.17 ± 8.30	90.28 ± 11.01
GLI_85	85.88 ± 17.54	90.59 ± 12.48	**92.94 ± 8.60**	91.76 ± 10.54	83.53 ± 19.48	91.76 ± 10.54
average	79.88 ± 15.75	79.86 ± 15.17	80.26 ± 14.92	77.53 ± 15.98	80.90 ± 14.73	**81.66 ± 14.66**

Table 6
The classification accuracies of redut sets on C4.5.

Datasets	Original data	FBR	FPR	IFPR	IFCE	IFDMI
wine	93.82 ± 4.86	94.38 ± 4.49	94.38 ± 4.53	91.57 ± 6.58	93.82 ± 5.45	**96.63 ± 2.93**
heart	76.30 ± 28.06	78.52 ± 25.67	75.93 ± 29.39	78.15 ± 26.06	**78.89 ± 27.87**	78.00 ± 25.66
australian	86.09 ± 19.38	85.65 ± 19.43	**85.80 ± 20.00**	84.20 ± 20.74	85.51 ± 20.29	85.36 ± 20.49
segment	96.62 ± 1.10	96.10 ± 1.27	96.23 ± 1.20	94.72 ± 1.72	95.97 ± 1.29	**96.80 ± 1.04**
waveform	75.92 ± 16.92	74.24 ± 19.40	76.68 ± 16.86	76.30 ± 17.73	76.42 ± 17.20	**76.96 ± 17.34**
wdbc	94.02 ± 6.58	94.20 ± 6.55	94.90 ± 5.91	93.67 ± 7.36	94.73 ± 6.12	**95.43 ± 5.47**
ionosphere	90.60 ± 9.57	92.02 ± 9.87	92.31 ± 8.75	90.88 ± 10.57	90.03 ± 15.94	**92.59 ± 9.41**
soner	71.15 ± 28.63	73.56 ± 27.82	**77.40 ± 23.32**	72.60 ± 27.07	72.60 ± 27.54	74.04 ± 27.94
hillvalley	49.65 ± 50.00	**49.67 ± 50.00**	**49.67 ± 50.00**	**49.67 ± 50.00**	**49.67 ± 50.00**	**49.67 ± 50.00**
musk2	88.68 ± 11.75	90.52 ± 11.21	89.53 ± 11.44	**90.81 ± 10.87**	87.84 ± 13.95	89.82 ± 12.68
featMIAS	69.25 ± 15.89	60.25 ± 19.65	62.11 ± 20.17	57.45 ± 25.06	67.08 ± 17.40	**67.70 ± 17.62**
Yale	48.48 ± 7.13	44.85 ± 7.41	46.06 ± 7.38	39.39 ± 8.47	43.64 ± 7.55	**47.27 ± 7.59**
ORL	58.75 ± 2.12	**59.00 ± 2.16**	56.25 ± 2.28	54.50 ± 2.36	55.75 ± 2.29	58.75 ± 2.14
colon	82.26 ± 18.27	83.87 ± 17.09	77.42 ± 26.51	83.87 ± 17.09	75.81 ± 26.85	**87.10 ± 19.58**
srbct	84.34 ± 8.30	85.54 ± 7.42	**91.57 ± 4.76**	72.29 ± 14.06	87.95 ± 6.73	75.90 ± 12.74
web	51.68 ± 19.67	50.34 ± 21.10	53.69 ± 19.09	37.58 ± 25.85	51.68 ± 20.06	**61.75 ± 15.36**
arcene	81.00 ± 19.29	69.00 ± 35.55	69.00 ± 35.36	**72.00 ± 31.98**	71.50 ± 31.54	70.50 ± 31.01
orlraws10P	75.00 ± 5.00	72.00 ± 5.68	74.00 ± 5.88	75.00 ± 5.51	**80.00 ± 4.22**	78.00 ± 4.70
CLL_SUB_111	59.46 ± 26.79	62.16 ± 25.12	63.97 ± 26.51	60.36 ± 26.30	58.56 ± 29.09	**65.77 ± 27.24**
MLL	84.72 ± 10.61	88.89 ± 10.15	87.50 ± 11.55	86.11 ± 10.53	**91.67 ± 6.75**	87.50 ± 9.69
GLI_85	74.12 ± 25.63	**89.41 ± 12.14**	81.18 ± 21.49	88.24 ± 13.70	78.82 ± 21.21	84.71 ± 17.46
average	75.81 ± 15.98	75.91 ± 16.15	75.98 ± 16.78	73.78 ± 17.12	75.62 ± 17.11	**77.15 ± 16.10**

exceeds the raw data with a margin of 15.31%. On web dataset with respect to C4.5, the best accuracy obtained by method IFDMI exceeds the raw data with a margin of 10.07%. It is important to note that in some cases, reducing the number of attributes using reduction methods may result in a decrease in accuracy compared to the raw data. On orlraws10P dataset with KNN, the accuracy of the raw data is 95.00%, whereas that of the reduced data by method IFDMI is 93.00% with a margin of decrease of 2.00%. However, it is more satisfactory that there is a dimension decrease of 99.92% by the reduction method. Moreover, it is observed that there are 10 classification accuracies of the reduct sets obtained by method IFDMI are higher than or equal to those obtained by other four methods on KNN, and method IFDMI outperforms the other four methods 11 times on C4.5. In addition, the average performance of method IFDMI is higher than that of other four methods on KNN and C4.5. Fig. 3 also displays the classification accuracies on KNN and C4.5. It is obvious that method IFDMI has obtained superior classification results on most datasets. As a consequence, the performance of the presented method is relatively good in comparison to the other four methods.

The reasons can be explained by the following aspects. Firstly, the proposed IF dominance relation is based on IFISs, and the attribute values of the sample are IF numbers, which ensures that our method can deal with IF data directly and is conducive to processing IF data. Secondly, the proposed relation is an IF relation, it can maintain both the maximal degrees of samples' membership
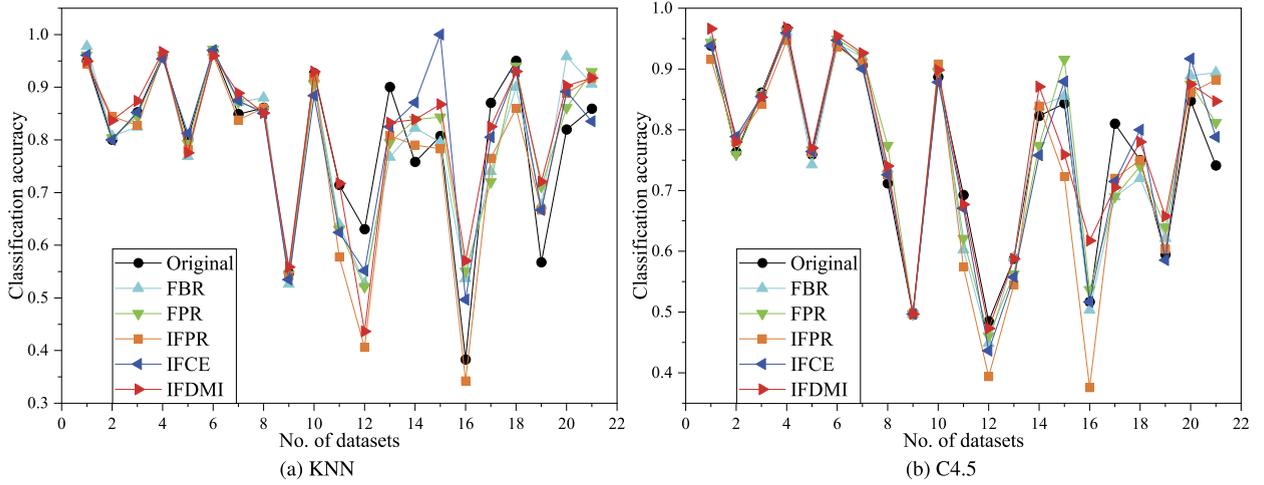
Fig. 3. The classification accuracies on KNN and C4.5.

Table 7
Statistical test of five methods under KNN and C4.5.

Classifiers	Mean rankings					χ_F^2	F_F
	FBR	FPR	IFPR	IFCE	IFDMI		
KNN	3.2381	2.8095	3.8095	3.0238	2.1190	12.8095	3.5987
C4.5	2.9762	2.7381	3.9048	3.3095	2.0714	15.5048	4.5273

to their own classes and non-membership to other classes, and thus can guarantee the maximum rate of correct classification and the minimum rate of error classification simultaneously. Thirdly, we consider the dominance degrees and the non-dominance degrees between the pairs of samples under attributes, which can better distinguish samples. For each attribute in the attribute subset, its proportion of all attributes is considered under the dominance degrees, so that the role of each attribute can be reflected to the greatest extent.

5.3. Statistical tests

In order to further study whether there are striking differences in the classification accuracies of each classifier under all compared methods, Friedman test was performed from a statistical point of view to test the validity of the presented method. After that, the corresponding Bonferroni-Dunn test was also performed.

Friedman test is shown [47] as

$$\chi_F^2 = \frac{12N}{k(k+1)}\left(\sum_{j=1}^{k} R_j - \frac{k(k+1)^2}{4}\right), \; F_F = \frac{(N-1)\chi_F^2}{N(k-1)-\chi_F^2},$$

where N and k mean experimental datasets and the number of methods, respectively; F_F means a F-distribution under $(k-1)$ and $(k-1)(N-1)$ freedom degrees, $R_j = \sum_{i=1}^{N} r_{ji}/N$ is the average ranking of the jth method in all datasets, and r_{ji} is the ranking of the jth method under the ith dataset.

If the baseline of Bonferroni-Dunn test is called critical difference (CD), then the expression of CD is [48]

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

where α means the significance level, q_α means a critical value [48].

Let $\alpha = 0.1$. If the performances of all methods are tantamount, then the critical value of $F(4, 4*20)$ is 2.0160, the critical value of $q_{0.1}$ is 2.4590, and $CD = 1.1999$.

For those five attribute reduction methods, the null hypothesis of Friedman test can be built if the classification accuracies of all methods are equal. Based on the classification accuracies of twenty-one datasets shown in Table 5 and Table 6, one can easily calculate the rankings of five methods, and obtain their mean rankings under KNN and C4.5. The results, those calculated for the values of χ_F^2 and F_F, are represented in Table 7.

It is evident from Table 7 that two Friedman statistical values both exceed 2.0160, indicating that the original hypothesis is not accepted. That is to say, all methods have striking differences in the performance.
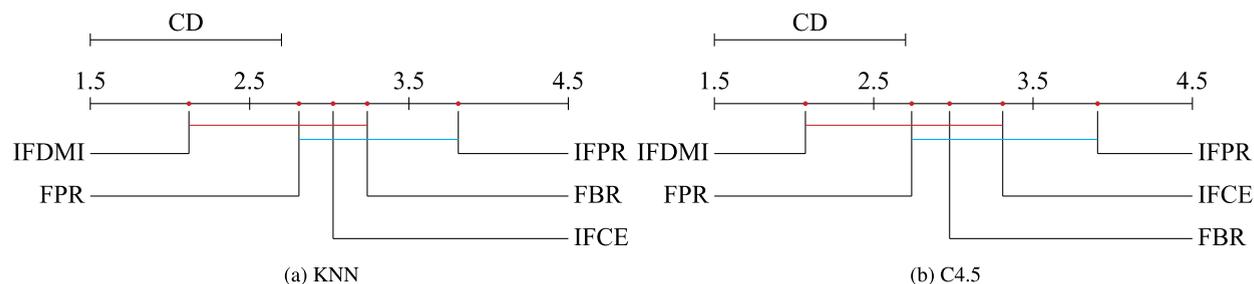
Fig. 4. Comparisons of accuracies with five methods under KNN and C4.5.

To make a more objective and intuitive comparison of the differences among several methods, the methods with no obvious difference are connected by the graph proposed by Demsar and Schuurmans [48]. These graphs can clearly illustrate the critical values among all methods. Fig. 4 shows the comparisons of accuracies with five methods under KNN and C4.5, where the critical value is shown as the top line, the mean ranking of each method is represented by coordinate axis, and the closer to the right-hand side, the higher the average ranking. Besides, the methods with no striking difference in the performance are connected by horizontal lines.

We can see from Fig. 4 that the striking differences of five methods are obvious. In Fig. 4, IFMDI performs obviously better than IFPR under KNN and C4.5. IFDMI is at the top in the rankings although there is no striking difference among IFDMI, FPR, IFCE and FBR. In summary, the introduced method is overall superior to the other four methods.

6. Conclusions

Intuitionistic fuzzy information system (IFIS) is an extension of fuzzy information system that can represent more uncertain information and more accurately describe the essence of fuzziness. When the entropy measurement of uncertainty is used to study attribute reduction in IFISs, intuitionistic fuzzy (IF) conditional entropy only considers the relevance (significance) of an attribute to the decision one by one, ignoring the produced redundancies of selected attributes. With that in mind, the article proposes an attribute reduction method based on IF dominance mutual information for IFISs. The IF dominance relation is defined based on IF dominance degree. The IF dominance entropy and its variations are constructed in view of IF dominance relation, and some of their properties are discussed. Based on IF dominance mutual information, an attribute reduction method and its algorithm are proposed. Furthermore, some experiments to evaluate the performance of the proposed method are presented. The proposed method is efficacious and applicable for IFISs, which is clearly shown by theoretical research and experimental analysis. These results will contribute to the analysis and processing of IF data.

The proposed method aims to reduce attributes in IFISs. However, its scalability is limited as it only targets IF data. It cannot directly process general data unless it is preprocessed to be in IF format. The IF dominance relation defined in the article is relatively complex, which leads to the complexity of the computation of IF dominance mutual information, and thus affects the scalability of the algorithm for calculating IF dominance mutual information. Besides, the proposed attribute reduction method only considers the redundancy among the attributes to be selected and the selected attributes, without taking into account the redundancy among the selected attributes. In the future, we will try to simplify the IF dominance relation or construct a simple IF relation, while also considering the redundancy among the selected attributes.

CRediT authorship contribution statement

Xiaofeng Liu: Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Hong Mo:** Conceptualization, Methodology, Writing – review & editing. **Jianhua Dai:** Conceptualization, Formal analysis, Investigation, Methodology, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

References

[1] L.A. Zadeh, Fuzzy logic equals computing with words, IEEE Trans. Fuzzy Syst. 4 (2) (1996) 103–111.

[2] L.A. Zadeh, Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, Fuzzy Sets Syst. 90 (2) (1997) 111–127.

[3] Z. Pawlak, Rough sets, Int. J. Comput. Inf. Sci. 11 (5) (1982) 341–356.

[4] L.A. Zadeh, Fuzzy sets, Inf. Control 8 (3) (1965) 338–353.

[5] W.Z. Wu, Y. Leung, J.S. Mi, Granular computing and knowledge reduction in formal contexts, IEEE Trans. Knowl. Data Eng. 21 (10) (2009) 1461–1474.

[6] L. Zhang, B. Zhang, Theory and Application of Problem Solving-Theory and Application of Granular Computing in Quotient Spaces, Tsinghua University Publishers, 2007.

[7] N.N. Thuy, S. Wongthanavasu, A novel feature selection method for high-dimensional mixed decision tables, IEEE Trans. Neural Netw. Learn. Syst. 33 (7) (2022) 3024–3037.

[8] Y. Kang, J.H. Dai, Attribute reduction in inconsistent grey decision systems based on variable precision grey multigranulation rough set model, Appl. Soft Comput. 133 (2023) 109928.

[9] X.Y. Zhang, J.R. Li, Incremental feature selection approach to interval-valued fuzzy decision information systems based on λ-fuzzy similarity self-information, Inf. Sci. 625 (2023) 593–619.

[10] C.E. Shanon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (3) (1948) 379–423.

[11] C. Luo, T.R. Li, H.M. Chen, J.C. Lv, Z. Yi, Fusing entropy measures for dynamic feature selection in incomplete approximation spaces, Knowl.-Based Syst. 252 (2022) 109329.

[12] H.X. Huang, Z.W. Li, F. Liu, C.F. Wen, New uncertainty measurement for hybrid data and its application in attribute reduction, Inf. Sci. 664 (2024) 120334.

[13] W.H. Xu, K. Cai, D.D. Wang, A novel information fusion method using improved entropy measure in multi-source incomplete interval-valued datasets, Int. J. Approx. Reason. 164 (2024) 109081.

[14] L.L. Xie, G.P. Lin, J.J. Li, Y.D. Lin, A novel fuzzy-rough attribute reduction approach via local information entropy, Fuzzy Sets Syst. 473 (2023) 108733.

[15] J.H. Dai, Q. Liu, X.T. Zou, C.C. Zhang, Feature selection based on fuzzy combination entropy considering global and local feature correlation, Inf. Sci. 652 (2024) 119753.

[16] B.W. Chen, X.Y. Zhang, Z. Yuan, Two-dimensional improved attribute reductions based on distance granulation and condition entropy in incomplete interval-valued decision systems, Inf. Sci. 657 (2024) 119910.

[17] R. Battiti, Using mutual information for selecting features in supervised neural net learning, IEEE Trans. Neural Netw. 5 (4) (1994) 537–550.

[18] Q.H. Hu, Z.X. Xie, D.R. Yu, Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation, Pattern Recognit. 40 (12) (2007) 3509–3521.

[19] H.C. Peng, F.H. Long, C. Ding, Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) (2005) 1226–1238.

[20] H.R. Cheng, Z.G. Qin, C.S. Feng, Y. Wang, F.G. Li, Conditional mutual information-based feature selection analyzing for synergy and redundancy, ETRI J. 33 (2) (2011) 210–218.

[21] J. Wang, J.M. Wei, Z.L. Yang, S.Q. Wang, Feature selection by maximizing independent classification information, IEEE Trans. Knowl. Data Eng. 29 (4) (2017) 828–841.

[22] J.H. Dai, J.L. Chen, Feature selection via normative fuzzy information weight with application into tumor classification, Appl. Soft Comput. 92 (2020) 106299.

[23] K.T. Atanassov, Intuitionistic fuzzy sets, Fuzzy Sets Syst. 20 (1) (1986) 87–96.

[24] D.J. Yu, L.B. Sheng, Z.S. Xu, Analysis of evolutionary process in intuitionistic fuzzy set theory: a dynamic perspective, Inf. Sci. 601 (2022) 175–188.

[25] X.D. Liu, Y.K. Sun, H. Garg, S.T. Zhang, Analysis of distance measures in intuitionistic fuzzy set theory: a line integral perspective, Expert Syst. Appl. 226 (2023) 120221.

[26] S.C. Ngan, An extension framework for creating operators and functions for intuitionistic fuzzy sets, Inf. Sci. 666 (2024) 120336.

[27] B. Huang, H.X. Li, D.K. Wei, Dominance-based rough set model in intuitionistic fuzzy information systems, Knowl.-Based Syst. 28 (2012) 115–123.

[28] E. Hesameddini, M. Jafarzadeh, H. Latifizadeh, Rough set theory for the intuitionistic fuzzy information systems, Int. J. Mod. Math. Sci. 6 (3) (2013) 132–143.

[29] T. Feng, J.S. Mi, S.P. Zhang, Belief functions on general intuitionistic fuzzy information systems, Inf. Sci. 271 (1) (2014) 143–158.

[30] A.H. Tan, W.Z. Wu, Y.H. Qian, J.Y. Liang, J.K. Chen, J.J. Li, Intuitionistic fuzzy rough set-based granular structures and attribute subset selection, IEEE Trans. Fuzzy Syst. 27 (3) (2019) 527–539.

[31] A.H. Tan, S.W. Shi, W.Z. Wu, J.J. Li, W. Pedrycz, Granularity and entropy of intuitionistic fuzzy information and their applications, IEEE Trans. Cybern. 52 (1) (2022) 192–204.

[32] Q.H. Hu, D.R. Yu, J.F. Liu, C.X. Wu, Neighborhood rough set based heterogeneous feature subset selection, Inf. Sci. 178 (18) (2008) 3577–3594.

[33] E. Szmidt, J. Kacprzyk, Entropy for intuitionistic fuzzy sets, Fuzzy Sets Syst. 118 (3) (2001) 467–477.

[34] I.K. Vlachos, G.D. Sergiadis, Subsethood, entropy, and cardinality for interval-valued fuzzy sets-an algebraic derivation, Fuzzy Sets Syst. 158 (12) (2007) 1384–1396.

[35] H. Bustince, P. Burillo, Structures on intuitionistic fuzzy relations, Fuzzy Sets Syst. 78 (3) (1996) 293–303.

[36] W.Z. Wu, L. Zhou, On intuitionistic fuzzy topologies based on intuitionistic fuzzy reflexive and transitive relations, Soft Comput. 15 (2011) 1183–1194.

[37] Y.H. Qian, J.Y. Liang, W.Z. Wu, C.Y. Dang, Information granularity in fuzzy binary GrC model, IEEE Trans. Fuzzy Syst. 19 (2) (2011) 253–264.

[38] J. Puri, S.P. Yadav, Intuitionistic fuzzy data envelopment analysis: an application to the banking sector in India, Expert Syst. Appl. 42 (11) (2015) 4982–4998.

[39] R.J. Kuo, F.E. Zulvia, The application of gradient evolution algorithm to an intuitionistic fuzzy neural network for forecasting medical cost of acute hepatitis treatment in Taiwan, Appl. Soft Comput. 111 (2021) 107711.

[40] M. Pant, K. Bisht, S. Negi, Computational-based partitioning and Strong (α, β)-cut based novel method for intuitionistic fuzzy time series forecasting, Appl. Soft Comput. 142 (2023) 110336.

[41] J.H. Dahooie, R. Raafat, A.R. Qorbani, T. Daim, An intuitionistic fuzzy data-driven product ranking model using sentiment analysis and multi-criteria decision-making, Technol. Forecast. Soc. Change 173 (2021) 121158.

[42] W.J. Wang, J.M. Zhan, J.S. Mi, A three-way decision approach with probabilistic dominance relations under intuitionistic fuzzy information, Inf. Sci. 582 (2022) 114–145.

[43] X.Y. Zhang, J.L. Hou, A novel rough set method based on adjustable-perspective dominance relations in intuitionistic fuzzy ordered decision tables, Int. J. Approx. Reason. 154 (2023) 218–241.

[44] R. Jensen, Q. Shen, New approaches to fuzzy-rough feature selection, IEEE Trans. Fuzzy Syst. 17 (4) (2009) 824–838.

[45] Y.H. Qian, Q. Wang, H.H. Cheng, J.Y. Liang, C.Y. Dang, Fuzzy-rough feature selection accelerator, Fuzzy Sets Syst. 258 (2015) 61–78.

[46] C.Z. Wang, Q.H. Hu, X.Z. Wang, D.G. Chen, Y.H. Qian, Z. Dong, Feature selection based on neighborhood discrimination index, IEEE Trans. Neural Netw. Learn. Syst. 29 (7) (2018) 2986–2999.

[47] M. Friedman, A comparison of alternative tests of significance for the problem of m rankings, Ann. Math. Stat. 11 (1) (1940) 86–92.

[48] J. Demsar, D. Schuurmans, Statistical comparison of classifiers over multiple data sets, J. Mach. Learn. Res. 7 (2006) 1–30.