Information Sciences 563 (2021) 196–225

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

# Class-specific information measures and attribute reducts for hierarchy and systematicness

Xianyong Zhang [a,*], Hong Yao [a], Zhiying Lv [b], Duoqian Miao [c]

[a] School of Mathematical Sciences, Laurent Mathematics Center, Sichuan Normal University, Chengdu 610066, China
[b] School of Management, Chengdu University of Information Technology, Chengdu 610225, China
[c] Department of Computer Science and Technology, Tongji University, Shanghai 201804, China

ARTICLE INFO

*Article history:*
Received 24 February 2020
Received in revised form 22 January 2021
Accepted 31 January 2021
Available online 5 February 2021

*Keywords:*
Rough set theory
Class-specific attribute reducts
Classification-based attribute reducts
Information theory
Uncertainty measurement
Granular computing

ABSTRACT

Attribute reduction of rough set theory underlies knowledge acquisition and has two hierarchical types (classification-based and class-specific attribute reducts) and two perspectives from algebra and information theory; thus, there are four combined modes in total. Informational class-specific reducts are fundamental but lacking and are thus investigated by correspondingly constructing class-specific information measures. First, three types of information measures (i.e., information entropy, conditional entropy, and mutual information) are novelly established at the class level by hierarchical decomposition to acquire their hierarchical connection, systematical relationship, uncertainty semantics, and granulation monotonicity. Second, three types of informational class-specific reducts are correspondingly proposed to acquire their internal relationship, basic properties, and heuristic algorithm. Third, the informational class-specific reducts achieve their transverse connections, including the strength feature and consistency degeneration, with the algebraic class-specific reducts and their hierarchical connections, including the hierarchical strength and balance, with the informational classification-based reducts. Finally, relevant information measures and attribute reducts are effectively verified by decision tables and data experiments. Class-specific information measures deepen existing classification-based information measures by a hierarchical isomorphism, while the informational class-specific reducts systematically perfect attribute reduction by level and viewpoint isomorphisms; these results facilitate uncertainty measurement and information processing, especially at the class level.



## 1. Introduction

Rough set theory can effectively perform information processing for imprecise, inconsistent, and incomplete data [34]. Rough set theory has become a basic methodology of data analysis and is widely used in multiple fields of artificial intelligence and machine learning [2,15,26,44,50,56,58,66–68]. As a key topic of rough set theory, attribute reduction reduces dimensionality based on knowledge granulation to perform feature selection and thus reduces system complexity to promote knowledge acquisition. Therefore, attribute reduction is critical to rough sets and granular computing, and there are many relevant approaches in terms of decision tables. For example, Ref. [9] uses a set operational perspective to study

* Corresponding author.
*E-mail address:* xianyongzh@sina.com.cn (X. Zhang).

https://doi.org/10.1016/j.ins.2021.01.080

decision tables and their basic notions. Ref. [10] proposes new perspectives of granular computing in relation geometry induced by pairings, where a pairing is an abstract mathematical generalization of an information table. Ref. [16] discusses attribute reduction using the nullity-based matroid of rough sets. Ref. [37] proposes positive approximation to accelerate a heuristic process of attribute reduction. Ref. [41] investigates association reducts using data-based functional dependencies between sets of attributes. Ref. [42] uses generalized decision functions to define attribute dependencies and reducts in decision tables. Ref. [46] builds a matroidal structure of rough set theory by redefining rough approximation operators through matroidal approaches. Ref. [55] studies rough sets through matroids using graph and matrix approaches, and Ref. [65] proposes a reduct construction method based on discernibility matrix simplification. Regarding attribute reduction, there are two hierarchical types (classification-based and class-specific attribute reducts) [64] and two metric viewpoints from algebra and information theory [54]; thus, there are four combined modes in total based on these two dimensions (algebraic and informational classification-based and class-specific reducts), as shown in Fig. 1. In particular, attribute reduction adheres to granulation computing [8,21,35,39,52,57,60,61], and granulation monotonicity plays a vital role in reduct definition and algorithm construction [4,13,33,51,53].

A decision table serves as the formal context of attribute reduction and consists of a three-level structure: Macro-Top, Meso-Middle and Micro-Bottom [24,47,71]. The relevant three-level analysis regarding granular computing is related to tri-level thinking [62,63]. The classification-based and class-specific reducts are respectively located at Macro-Top and Meso-Middle, which are respectively called the classification and class levels. Classification-based reducts use classification optimization and dependency reasoning and have become mainstream [1,3,12,18,20,22,25,43]. Their initial mode resorts to positive regions and dependency degrees to become algebraic, while their developmental mode uses information measures to become informational. Information theory provides uncertainty measurements of information contents; thus, relevant information measures have been introduced into rough set theory to conduct uncertainty representation and information processing [6,7,11,38,48], and information reduction has been extensively studied [14,19,27,32,40,45,49]. At the classification level in particular, information entropy, conditional entropy, and mutual information are systematically developed and generate informational classification-based reducts and heuristic reduction algorithms [32,54]. Moreover, the possibility of considering multiple information measures rather than the classical positive region has been investigated in detail in [33], mainly by preserving given properties in terms of a (changed) discernibility matrix. Targeting pattern applications, class-specific reducts have recently been proposed to improve classification-based reducts and highlight the class optimization and corresponding rule extraction [64]. Furthermore, class-specific reducts are mainly researched from the algebraic viewpoint and thus have been developed by quantitative region preservation [73], three-way decisions [29], three-way probabilities [36], min–max attribute-object bireducts [28], cost sensitivity [30], and rule-based classifiers [23]. In contrast, information-driven reports on class-specific reducts are rare. Ref. [72] discusses three-way class-specific reducts from three-way weighted entropies, whose metric construction comes from bottom-middle integration [71], and Ref. [31] examines the computational formulations of class-specific reducts in three-way probabilistic rough set models based on fuzzy entropies.

Against the background of four reduct modes, existing results and desired developments are showed as follows, and both relevant background and development are described in Fig. 1.

(1) Three old reduct modes (informational classification-based reducts, algebraic classification-based reducts, and algebraic class-specific reducts) have been investigated in detail and applied in practice extensively. Regarding mutual relevance, the former two modes have already gained transverse connections [32,54], while the latter two have already yielded hierarchical connections [64].
(2) In contrast, informational class-specific reducts appear less frequently. Both their transverse connections with algebraic class-specific reducts and hierarchical connections with informational classification-based reducts are required but also lacking.

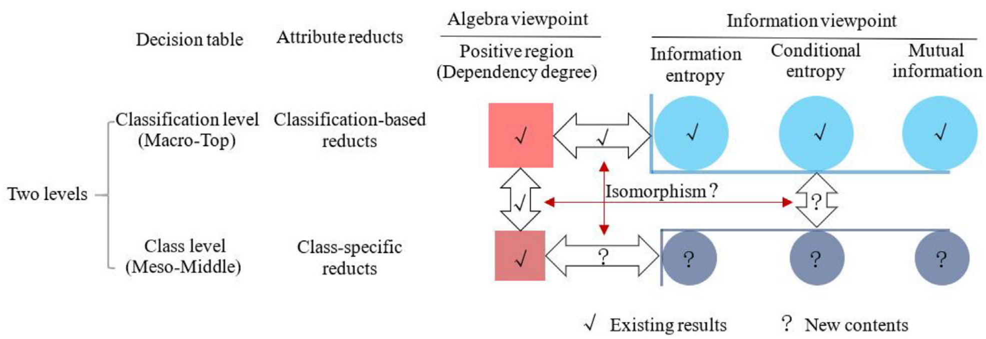



**Fig. 1.** Research background and development of four reduct modes.

Clearly, definitions and connections of informational class-specific reducts have become two important issues. In this paper, they are resolved by constructing new informational class-specific reducts; however, how to mine underlying class-specific information measures becomes a new question that is both critical and difficult to answer. Because classical information measures (information entropy, conditional entropy, and mutual information) at the classification level induce informational classification-based reducts, novel information measures, which include three similar information measures, at the class level are worthy of being hierarchically determined by the top-middle decomposition to accordingly induce informational class-specific reducts. This strategy of hierarchical decomposition and systematic simulation becomes natural, feasible, and scientific for measure mining and reduct construction, and also underlies the further connection analysis of relevant reduct modes. Class-specific information measures and informational class-specific reducts can be novelly constructed by referring to and simulating traditional notions at the classification level. Then, vertical and horizontal connections embracing the informational class-specific reducts can be described. As a result, concrete cases will be described in detail as follows, and three types of variant isomorphisms can be reasonably extracted or perfectly concluded.

(1) Regarding measures, the class-specific information entropy, conditional entropy, and mutual information are determined to achieve the systematical relationship, uncertainty semantics, and granulation monotonicity. This new system of class-specific information measures is hierarchically isomorphic to the existing system of classification-based information measures.
(2) Regarding reduct constructions, the informational class-specific reducts based on class-specific information entropy, conditional entropy, and mutual information are proposed to acquire the internal relationship, basic property, and heuristic algorithm.
(3) Regarding transverse reduct connections, the informational class-specific reducts have the strength feature and consistency degeneration, in contrast to the algebraic class-specific reducts. This new system of informational class-specific reducts is hierarchically isomorphic to the existing system of informational classification-based reducts.
(4) Regarding hierarchical reduct connections, the informational class-specific and classification-based reducts have the hierarchical strength and balance. The informational reducts with both levels are transversely isomorphic to the algebraic reducts with both levels, mainly by perspective systems and hierarchical reduct relations.
(5) Finally, the relevant information measures and attribute reducts are effectively verified by both decision tables and data experiments.

**Table 1**
A list of basic mathematical symbols.

| System | Symbol | Description |
|---|---|---|
| Decision table | $A, B, C, R; D$ | Condition attribute subsets; decision attribute set |
| | $E_A, E_B, E_C, E_R; E_D$ | Equivalent relations |
| | $[x]_A, [x]_B, [x]_C, [x]_R; D_j$ | Condition classes; decision class |
| | $\pi_A, \pi_B, \pi_C, \pi_R; \pi_D$ | Condition partitions; decision classification |
| | $B \xrightarrow{-(B-A)} A, C \xrightarrow{-(C-R)} R, R \xrightarrow{-\{r\}} R - \{r\}, C \xrightarrow{-\{c\}} C - \{c\}$ | Attribute deletion |
| | $\pi_B \xrightarrow{\preceq} \pi_A, \pi_C \xrightarrow{\preceq} \pi_R$ | Granulation coarsening |
| | $\bigcup_{t=1}^{k} [x]_C^t = [x]_R \ (k \geqslant 2)$ | Representative group of granular merging |
| | $* = A, B, C, R, \cdots$ (i.e., $* \in 2^C - \{\varnothing\}$) | General representation of condition attribute subsets |
| Algebra viewpoint | $POS(D_j \mid \pi_*), \gamma_*(D_j)$ | Class-specific positive region and dependency degree |
| | $RED^{AV}(D_j), CORE^{AV}(D_j)$ | Class-specific reduct set and attribute core |
| | $POS(\pi_D \mid \pi_*), \gamma_*(D)$ | Classification-based positive region and dependency degree |
| | $RED^{AV}(\pi_D), CORE^{AV}(\pi_D)$ | Classification-based reduct set and attribute core |
| Information viewpoint | $H_{D_j}(*), H(D_j); H(D_j \mid *), H(* \mid D_j); I(*; D_j), I(D_j; *)$ | Class-specific information entropy, conditional entropy, mutual information |
| | $H_W(D_j \mid *)$ ($H_W^{lin}(D_j \mid *)$), $H_W^{D_j}(*), H_W(* \mid D_j)$ | Likelihood (likelihood-linear), prior, posterior weighted entropies |
| | $RED^{IE}(D_j), RED^{CE}(D_j), RED^{MI}(D_j)$ | Class-specific reduct sets on three-way information measures |
| | $CORE^{IE}(D_j), CORE^{CE}(D_j), CORE^{MI}(D_j)$ | Class-specific attribute cores on three-way information measures |
| | $Sig^{IE}(*, c, D_j), Sig^{CE}(*, c, D_j), Sig^{MI}(*, c, D_j)$ | Class-specific attribute significance on three-way information measures |
| | $H(*), H(D); H(D \mid *), H(* \mid D); I(*; D), I(D; *)$ | Classification-based information entropy, conditional entropy, mutual information |
| | $RED^{IE}(\pi_D), RED^{CE}(\pi_D), RED^{MI}(\pi_D)$ | Classification-based reduct sets on three-way information measures |
| | $CORE^{IE}(\pi_D), CORE^{CE}(\pi_D), CORE^{MI}(\pi_D)$ | Classification-based attribute cores on three-way information measures |
| | $\sharp \in \{IE, CE, MI\}$ | General representation of three-way information types |
| | $RED^{\sharp}(D_j), CORE^{\sharp}(D_j); RED^{\sharp}(\pi_D), CORE^{\sharp}(\pi_D)$ | Informational class-specific and classification-based reduct sets and attribute cores |
| Generalized viewpoint | $RT, RT_{\mathrm{IE}}, RT_{\mathrm{strong}}, RT_{\mathrm{weak}}$ | General, entropic, strong, weak reduction targets |
| | $C \xrightarrow{-(C-R)} R \models RT$, $R \xrightarrow{-\{r\}} R - \{r\} \not\models RT, C \xrightarrow{-\{c\}} C - \{c\} \not\models RT$ | Reduction target satisfiability or unsatisfiability for attribute deletion |
| | $RED^{RT_{\mathrm{strong}}}, CORE^{RT_{\mathrm{strong}}}; RED^{RT_{\mathrm{weak}}}, CORE^{RT_{\mathrm{weak}}}$ | Strong and weak reduct sets and attribute cores |

The isomorphism construction of novel class-specific measures and reducts hierarchically deepen existing classification-based measures and reducts, respectively, while the correlation revelation of informational class-specific reducts systematically perfects the reduction framework with two-level-horizontal and two-viewpoint-longitudinal reducts.

The remainder of this paper is organized as follows. Section 2 reviews three existing reduct modes: algebraic and informational classification-based reducts and algebraic class-specific reducts. Section 3 systematically investigates three types of information measures at the class level: class-specific information entropy, conditional entropy, and mutual information. Section 4 proposes three types of informational class-specific reducts. Section 5 analyzes the transverse connections between the informational and algebraic class-specific reducts and the hierarchical connections between the informational class-specific and classification-based reducts. Section 6 provides examples and experimental verification. Finally, Section 7 concludes the paper. In preparation, Table 1 provides main mathematical notations in advance.

## 2. Existing attribute reducts from the algebraic and informational viewpoints

As shown in Fig. 1, there are two hierarchical types of attribute reducts (classification-based and class-specific reducts), which are reviewed below from two transverse viewpoints of algebra and information, as well as from a further generalized perspective.

### 2.1. Attribute reducts from the algebraic viewpoint

A decision table is a special data table $DT = (U, C \cup D, V, f)$ [5,17,33,49,59]. In this study, $U$ is a universe with a finite nonempty set of objects; $C$ and $D$ are finite nonempty sets of condition and decision attributes, respectively, where $C \cap D = \emptyset; V = \bigcup_{a \in C \cup D} V_a$ is the total value range, where $V_a$ is the value domain for $a \in C \cup D; f : U \times (C \cup D) \to V$ is an information function. An arbitrary subset of condition attributes $A \subseteq C$ can define equivalence relation

$$E_A = \{(x, y) \in U \times U | \forall a \in A[f(x, a) = f(y, a)]\},$$

and equivalence class $[x]_A = \{y \in U | yE_A x\}$, equivalence partition $\pi_A = U/E_A = \{[x]_A | x \in U\}$. Similarly, the total set of decision attributes $D$ induces decision relation $E_D$, decision class $D_j$, and decision classification $\pi_D = \{D_j | j = 1, 2, \cdots, m\}$. The decision table $DT = (U, C \cup D, V, f)$ can be simply noted as $(U, C \cup D)$ and has two forms regarding consistency and inconsistency. In terms of set inclusion relation $\subseteq$, if $E_C \subseteq E_D$, then the decision table is consistent; otherwise, it is inconsistent. As usual, let symbol $||$ denote the cardinality function of sets.

Within the framework of conditional $A$ and decisional $D$ and their granulation, decision table $(U, C \cup D)$ contains three-level granular structures [71]. The usual two levels are extracted into Table 2 to describe attribute reduction based on measures. According to [71], the third level is Micro-Bottom $([x]_A, D_j)$, and its category reducts can be transformed into the object-oriented attribute reducts, which are hierarchically connected with the middle class-specific reducts and top classification-based reducts. Targeting the top and middle levels, the region, measure, and reduct are discussed from the algebraic perspective below.

**Definition 1** [34]. The positive, negative, and boundary regions of a decision class $D_j$ given partition $\pi_A$ are respectively defined by:

$$POS(D_j|\pi_A) = \{x | [x]_A \subseteq D_j\}, NEG(D_j|\pi_A) = \{x | [x]_A \subseteq \tilde{~}D_j\}, BND(D_j|\pi_A) = U - POS(D_j|\pi_A) \cup NEG(D_j|\pi_A). \quad (1)$$

And the class-specific dependency degree can be determined by:

$$\gamma_A(D_j) = \frac{|POS(D_j|\pi_A)|}{|U|}. \quad (2)$$

Furthermore, the positive and boundary regions of decision classification $\pi_D$ given partition $\pi_A$ are respectively defined by:

$$POS(\pi_D|\pi_A) = \bigcup_{j=1}^{m} POS(D_j|\pi_A), \quad BND(\pi_D|\pi_A) = U - POS(\pi_D|\pi_A), \quad (3)$$

and the classification-based dependency degree becomes:

**Table 2**
Two levels of decision table and corresponding measures and reducts.

| Level | Structure composition | Algebraic dependency degree | Information measures | Attribute reducts |
|---|---|---|---|---|
| Classification level | Macro-Top $(\pi_A, \pi_D)$ | $\gamma_A(D)$ ✓ | $H(A), H(D\|A), I(A; D)$ ✓ | Classification-based reducts |
| Class level | Meso-Middle $(\pi_A, D_j)$ | $\gamma_A(D_j)$ ✓ | $H_{D_j}(A), H(D_j\|A), I(A; D_j)$ ? | Class-specific reducts |

Symbol ✓ means the existence in current references, while symbol ? reflects the innovative construction in this paper.

$$\gamma_A(D) = \frac{|POS(\pi_D|\pi_A)|}{|U|} = \sum_{j=1}^{m} \gamma_A(D_j). \tag{4}$$

**Proposition 1** [34]. If $A \subseteq B \subseteq C$, then

$$\begin{aligned} &\text{(i)} \quad POS(D_j|\pi_A) \subseteq POS(D_j|\pi_B), \quad \gamma_A(D_j) \leqslant \gamma_B(D_j); \\ &\text{(ii)} \quad POS(\pi_D|\pi_A) \subseteq POS(\pi_D|\pi_B), \quad \gamma_A(D) \leqslant \gamma_B(D). \end{aligned} \tag{5}$$

Regarding the two levels, the positive regions and their dependency degrees have basic derivation/dependency semantics and good granulation monotonicity; thus, they are naturally utilized to define attribute reducts. An attribute reduct $R$ is a minimum set of $C$ that produces the same positive region or dependency degree.

**Definition 2** [34]. $R \subseteq C$ is a classification-based reduct if it satisfies one set of dual conditions regarding sufficiency and necessity:

$$\begin{cases} (S)\ POS(\pi_D|\pi_R) = POS(\pi_D|\pi_C), \\ (N)\ \forall r \in R[POS(\pi_D|\pi_{R-\{r\}}) \subset POS(\pi_D|\pi_R)]; \end{cases} \quad \begin{cases} (S')\ \gamma_R(D) = \gamma_C(D), \\ (N')\ \forall r \in R[\gamma_{R-\{r\}}(D) < \gamma_R(D)]. \end{cases}$$

The set of all classification-based reducts is denoted by $RED^{AV}(\pi_D)$.

**Definition 3** [64]. $R \subseteq C$ is a class-specific reduct if it satisfies one set of dual conditions regarding sufficiency and necessity:

$$\begin{cases} (s)\ POS(D_j|\pi_R) = POS(D_j|\pi_C), \\ (n)\ \forall r \in R[POS(D_j|\pi_{R-\{r\}}) \subset POS(D_j|\pi_R)]; \end{cases} \quad \begin{cases} (s')\ \gamma_R(D_j) = \gamma_C(D_j), \\ (n')\ \forall r \in R[\gamma_{R-\{r\}}(D_j) < \gamma_R(D_j)]. \end{cases}$$

The set of all class-specific reducts is denoted by $RED^{AV}(D_j)$.

Via joint sufficiency and individual necessity, classification-based and class-specific reducts are respectively determined in Definitions 2 and 3 from the algebraic viewpoint, and their hierarchical relationships have been described by [64]. Moreover, the attribute cores serve as an important attribute feature and are concretely defined at two levels by:

$$\begin{aligned} CORE^{AV}(\pi_D) &= \{c \in C | POS(\pi_D|\pi_{C-\{c\}}) \subset POS(\pi_D|\pi_C)\} = \{c \in C | \gamma_{C-\{c\}}(D) < \gamma_C(D)\}, \\ CORE^{AV}(D_j) &= \{c \in C | POS(D_j|\pi_{C-\{c\}}) \subset POS(D_j|\pi_C)\} = \{c \in C | \gamma_{C-\{c\}}(D_j) < \gamma_C(D_j)\}. \end{aligned} \tag{6}$$

According to the granulation monotonicity (Eq. (5)), they are respectively equivalent to

$$\begin{aligned} CORE^{AV}(\pi_D) &= \bigcap RED^{AV}(\pi_D), \\ CORE^{AV}(D_j) &= \bigcap RED^{AV}(D_j). \end{aligned} \tag{7}$$

Regarding the necessity condition of reducts, the concrete value/set relationship and single attribute expression are mainly used in this paper and are respectively equivalent to the inequality sign description and attribute subset style within the basic framework of granulation monotonicity.

### 2.2. Classification-based attribute reducts from the informational viewpoint

Next, information measures and classification-based reducts are recalled from the informational perspective.

At first, we define a mapping on $\sigma$-algebra $2^U$, i.e.,

$$p : 2^U \to \mathbf{Q}, \quad p(T) = \frac{|T|}{|U|}, \quad \forall T \subseteq U, \tag{8}$$

where $(U, 2^U, p)$ constitutes a probability space. This mathematical space establishes the usual probability framework of rough set theory, and informational measures can be systematically constructed by referring to information theory. For generalization, an arbitrary subset $A \subseteq C$ is chosen for representative descriptions. It is assumed that condition partition $\pi_A$ has number $n(A)$ of equivalence classes: $A_1, A_2, \cdots, A_{n(A)}$, i.e.,

$$\pi_A = \{[x]_A | x \in U\} = \{A_i | i = 1, 2, \cdots, n(A)\}.$$

We thus have decision classification $\pi_D = \{D_j | j = 1, 2, \cdots, m\}$. In preparation, we provide four related probabilities:

$$p(A_i) = \frac{|A_i|}{|U|}, \quad p(D_j) = \frac{|D_j|}{|U|}, \quad p(D_j|A_i) = \frac{|A_i \cap D_j|}{|A_i|}, \quad p(A_i|D_j) = \frac{|A_i \cap D_j|}{|D_j|}. \tag{9}$$

**Definition 4** ([32,54]). Regarding the classification, the information entropy, conditional entropy, and mutual information are respectively defined by:

$$\begin{aligned}
&H(A) = -\sum_{i=1}^{n(A)} p(A_i)\log_2 p(A_i), && H(D) = -\sum_{j=1}^{m} p(D_j)\log_2 p(D_j),\\
&H(D|A) = -\sum_{i=1}^{n(A)}\left(p(A_i)\sum_{j=1}^{m} p(D_j|A_i)\log_2 p(D_j|A_i)\right), && H(A|D) = -\sum_{j=1}^{m}\left(p(D_j)\sum_{i=1}^{n(A)} p(A_i|D_j)\log_2 p(A_i|D_j)\right),\\
&I(A;D) = H(D) - H(D|A), && I(D;A) = H(A) - H(A|D).
\end{aligned} \tag{10}$$

**Proposition 2** ([32,54]). Information measures have a basic relationship:

$$H(D) - H(D|A) = I(A;D) = I(D;A) = H(A) - H(A|D). \tag{11}$$

**Proposition 3** ([32,54]). If $A \subseteq B \subseteq C$, then

$$\begin{aligned}
&\text{(i)} \quad H(A) \leqslant H(B);\\
&\text{(ii)} \quad H(D|A) \geqslant H(D|B);\\
&\text{(iii)} \quad I(A;D) \leqslant I(B;D).
\end{aligned} \tag{12}$$

Via information theory, three types of information measures have the fundamental semantics and descriptive function of uncertainty measurements, and their systematicness is shown in Eq. (11). At the level of classification $\pi_D$, we are more concerned with only one set of information measures $(H(A), H(D|A), I(A;D))$, and the three main measures also have granulation monotonicity, as shown in Eq. (12). Therefore, they can be naturally utilized to define attribute reducts, and thus the informational classification-based reducts emerge to become minimum sets correspondingly preserving the initial and optimal uncertainty information.

**Definition 5** ([32,54]).

(1) $R \subseteq C$ is an IE-classification-based reduct (a classification-based reduct on information entropy) if it satisfies two conditions regarding sufficiency and necessity:

$$\begin{cases} (S1)\ H(R) = H(C),\\ (N1)\ \forall r \in R[H(R - \{r\}) < H(R)]. \end{cases}$$

The set of all these reducts is denoted by $RED^{IE}(\pi_D)$.

(2) $R \subseteq C$ is a CE-classification-based reduct (a classification-based reduct on conditional entropy) if it satisfies two conditions:

$$\begin{cases} (S2)\ H(D|R) = H(D|C),\\ (N2)\ \forall r \in R[H(D|(R - \{r\})) > H(D|R)]. \end{cases}$$

The set of all these reducts is denoted by $RED^{CE}(\pi_D)$.

(3) $R \subseteq C$ is an MI-classification-based reduct (a classification-based reduct on mutual information) if it satisfies two conditions:

$$\begin{cases} (S3)\ I(R;D) = I(C;D),\\ (N3)\ \forall r \in R[I(R - \{r\};D) < I(R;D)]. \end{cases}$$

The set of all these reducts is denoted by $RED^{MI}(\pi_D)$.

**Proposition 4** ([32,54]). If $R \subseteq C$, then

$$H(R) = H(C) \Longleftrightarrow \pi_R = \pi_C. \tag{13}$$

There are three types of informational classification-based reducts. By Proposition 4, the preservation of information entropy is equivalent to the preservation of knowledge granulation. Accordingly, an IE-classification-based reduct of decision table $(U, C \cup D)$ is equivalent to the attribute reduct of knowledge/granulation preservation of $\pi_C$, which is a reduct of an information sub-table $(U, C)$ without $D$. In contrast, the latter two types concern the dependency mechanism between conditional and decisional parts; thus, they more adhere to the classification task. Moreover, three types of attribute cores are determined from the informational viewpoint, i.e.,

$$\begin{aligned} & CORE^{IE}(\pi_D) = \{c \in C | H(C - \{c\}) < H(C)\}, \\ & CORE^{CE}(\pi_D) = \{c \in C | H(D|(C - \{c\})) > H(D|C)\}, \\ & CORE^{MI}(\pi_D) = \{c \in C | I(C - \{c\}; D) < I(C; D)\}. \end{aligned} \tag{14}$$

Via monotonicity in Eq. (12), they are equivalent to corresponding intersections of reduct sets, i.e.,

$$CORE^{IE}(\pi_D) = \bigcap RED^{IE}(\pi_D), \quad CORE^{CE}(\pi_D) = \bigcap RED^{CE}(\pi_D), \quad CORE^{MI}(\pi_D) = \bigcap RED^{MI}(\pi_D). \tag{15}$$

### 2.3. Attribute reducts from the generalized viewpoint

These algebraic and informational attribute reducts can be promoted from the generalized viewpoint, as introduced below.

Following [70], we can obtain a kind of generalized reduct from reduction targets, as well as the strong–weak relationships of attribute reducts. A reduction target ($RT$) is a state-specific preservation condition on attribute deletion, and its satisfiability (unsatisfiability) for the deletion process can be represented by the symbol $\models$ ($\not\models$). For example, the reduction target of information entropy preservation ($RT_{\mathrm{IE}}$) is defined by that deletion process $B \xrightarrow{-(B-A)} A$ ($\forall A \subseteq B \subseteq C$) submitting to $H(A) = H(B)$, and thus the relevant satisfiability is denoted by $B \xrightarrow{-(B-A)} A \models RT_{\mathrm{IE}}$. We can also obtain the satisfiability $C \xrightarrow{-(C-A)} A \models RT_{\mathrm{IE}}$ from the initial $C$ and the unsatisfiability $B \xrightarrow{-\{b\}} B - \{b\} \not\models RT_{\mathrm{IE}}, C \xrightarrow{-\{c\}} C - \{c\} \not\models RT_{\mathrm{IE}}$ regarding single-attribute deletion. According to the reduction target $RT$, we naturally define a corresponding reduct $R \subseteq C$ by two conditions:

$$\begin{cases} C \xrightarrow{-(C-R)} R \models RT, \\ \forall r \in R[R \xrightarrow{-\{r\}} R - \{r\} \not\models RT]. \end{cases}$$

Thus, we can achieve the reduct set $RED^{RT}$ and attribute core $CORE^{RT} = \{c \in C | C \xrightarrow{-\{c\}} C - \{c\} \not\models RT\}$. Now, we let $RT_{\mathrm{strong}}$ and $RT_{\mathrm{weak}}$ be two reduction targets. If $B \xrightarrow{-(B-A)} A \models RT_{\mathrm{strong}} \Rightarrow B \xrightarrow{-(B-A)} A \models RT_{\mathrm{weak}}$, then we consider that $RT_{\mathrm{strong}}$ is stronger than $RT_{\mathrm{weak}}$, which is weaker than $RT_{\mathrm{strong}}$. Furthermore, the strong and weak reduction targets respectively define the strong and weak reducts, and the latter two have a basic strength-weakness feature as follows.

**Proposition 5** [70]. It is assumed that the strong reduction target $RT_{\mathrm{strong}}$ induces $CORE^{RT_{\mathrm{strong}}}, RED^{RT_{\mathrm{strong}}}$, and the weak reduction target $RT_{\mathrm{weak}}$ generates $CORE^{RT_{\mathrm{weak}}}, RED^{RT_{\mathrm{weak}}}$. These attribute cores and reduct sets follow the strength relationship:

$$CORE^{RT_{\mathrm{strong}}} \supseteq CORE^{RT_{\mathrm{weak}}}, \tag{16}$$

$$\forall R^{RT_{\mathrm{strong}}} \in RED^{RT_{\mathrm{strong}}} \ \exists R^{RT_{\mathrm{weak}}} \in RED^{RT_{\mathrm{weak}}} \text{ such that } R^{RT_{\mathrm{weak}}} \subseteq R^{RT_{\mathrm{strong}}}. \tag{17}$$

By Proposition 5, parallel connections of reducts can usually be illustrated by the strength-weakness relation [70], which is reflexive and transitive. Thus, the systematic relationships of classification-based reducts [32,54] can be described as follows. The type based on information entropy is the strongest, which is supported by Proposition 4. The information types based on conditional entropy and mutual information are equivalent to offer the middle reduction strength, while the algebra type based on the positive region or dependency degree is the weakest. Moreover, the informational reducts based on conditional entropy (or mutual information) are necessarily the same as the algebraic reducts for consistent tables and may be different from the latter for inconsistent decision tables.

## 3. Systematic information measures at the class level

Based on the review in Section 2 (especially the framework shown in Fig. 1), two levels of attribute reducts (classification-based and class-specific reducts) and two construction viewpoints from the algebra and information together combine into four modes of reducts, and only the class-specific reducts from the informational viewpoint need in-depth explorations. For this topic, this section describes the underlying information measures from the classification level to the class level.

### 3.1. Hierarchical construction and systematic relationship

According to Definition 4, the classification level concerns six types of information measures (information entropy $H(A), H(D)$, conditional entropy $H(D|A), H(A|D)$, and mutual information $I(A;D), I(D;A)$ (Eq. (10)), where only a set of measures ($H(A), H(D|A), I(A;D)$) is used for the classification-based reducts according to Definition 5. By observing relevant analytic expressions in Eq. (10), the classification-based information measures exhibit three main cases of granular summation, and thus, we can correspondingly conduct the hierarchical decomposition on class information to extract class-specific measures.

(1) A measure, such as $H(D)$ and $H(A|D)$, may concern the class summation $\sum_{j=1}^{m}$ at the final process and can thus directly extract internal information on class $D_j$ by deleting external integration on summation $\sum_{j=1}^{m}$.
(2) A measure, such as $H(D|A)$, may concern both the class summation $\sum_{j=1}^{m}$ at the first process and the granular summation $\sum_{i=1}^{n(A)}$ at the second process (i.e., it has a form of double summations: $\sum_{i=1}^{n(A)}\sum_{j=1}^{m}$). Thus, the summation commutativity is first required, and then $\sum_{j=1}^{m}\sum_{i=1}^{n(A)}$ can induce internal $D_j$-class information based on $\sum_{i=1}^{n(A)}$.
(3) A measure, such as $H(A)$, may directly concern not the class summation $\sum_{j=1}^{m}$ but the granular summation $\sum_{i=1}^{n(A)}$ (i.e., it has a form of single summation: $\sum_{i=1}^{n(A)}$). Thus, the class summation $\sum_{j=1}^{m}$ must be introduced to achieve the form $\sum_{j=1}^{m}$ or style $\sum_{j=1}^{m}\sum_{i=1}^{n(A)}$; then, $D_j$-specific information can be extracted by deleting the class summation $\sum_{j=1}^{m}$.

According to these three strategies, we next perform corresponding transformations to extract six systematic class-specific measures that can be gradually divided into three groups: (1) $H(D_j)$ and $H(A|D_j)$, (2) $H(D_j|A)$ and $I(A;D_j)$, (3) $H_{D_j}(A)$ and $I(D_j;A)$.

**Definition 6.** Information entropy on class $D_j$ and conditional entropy on partition $\pi_A$ given class $D_j$ are respectively defined by:

$$\begin{aligned} &H(D_j) = -p(D_j)\log_2 p(D_j),\\ &H(A|D_j) = -p(D_j)\sum_{i=1}^{n(A)} p(A_i|D_j)\log_2 p(A_i|D_j) = -\sum_{i=1}^{n(A)} p(D_j)p(A_i|D_j)\log_2 p(A_i|D_j). \end{aligned} \tag{18}$$

**Proposition 6.** The information entropy and conditional entropy at the classification and class levels have hierarchical relationships:

$$H(D) = \sum_{j=1}^{m} H(D_j) \geqslant H(D_j),\ \ H(A|D) = \sum_{j=1}^{m} H(A|D_j) \geqslant H(A|D_j). \tag{19}$$

According to the $\sum_{j=1}^{m}$ formula (Eq. (10)), $H(D)$ and $H(A|D)$ have the direct class summation on class $D_j$; thus, they are easy for decomposition and extraction in terms of decision class. By Definition 6, information entropy $H(D_j)$ and conditional entropy $H(A|D_j)$ at the class level are proposed, and their hierarchical connections regarding decomposition/integration and size with $H(D)$ and $H(A|D)$ are shown in Proposition 6.

We now focus on $H(D|A)$ and $I(A;D)$. The two measures resort to double summations to more closely adhere to Macro-Top $(\pi_A, \pi_D)$, and relevant information extractions on decision class $D_j$ are significant for class-specific applications. Relevant mathematical transformations with summation commutativity are first given.

**Lemma 1.** *Conditional entropy $H(D|A)$ and mutual information $I(A;D)$ have the following summation forms (with commutativity or conversion):*

$$\begin{aligned} H(D|A) &= -\sum_{i=1}^{n(A)}\left(p(A_i)\sum_{j=1}^{m} p(D_j|A_i)\log_2 p(D_j|A_i)\right) = \sum_{j=1}^{m}\left(-\sum_{i=1}^{n(A)} p(A_i)p(D_j|A_i)\log_2 p(D_j|A_i)\right),\\ I(A;D) &= H(D) - H(D|A) = -\sum_{j=1}^{m} p(D_j)\log_2 p(D_j) - \sum_{j=1}^{m}\left(-\sum_{i=1}^{n(A)} p(A_i)p(D_j|A_i)\log_2 p(D_j|A_i)\right)\\ &= \sum_{j=1}^{m}\left(-p(D_j)\log_2 p(D_j) + \sum_{i=1}^{n(A)} p(A_i)p(D_j|A_i)\log_2 p(D_j|A_i)\right). \end{aligned} \tag{20}$$

**Definition 7.** Conditional entropy on class $D_j$ given partition $\pi_A$ and mutual information between partition $\pi_A$ and class $D_j$ are respectively defined by:

$$\begin{aligned} &H(D_j|A) = -\sum_{i=1}^{n(A)} p(A_i)p(D_j|A_i)\log_2 p(D_j|A_i),\\ &I(A;D_j) = -p(D_j)\log_2 p(D_j) + \sum_{i=1}^{n(A)} p(A_i)p(D_j|A_i)\log_2 p(D_j|A_i). \end{aligned} \tag{21}$$

**Proposition 7.** The conditional entropy and mutual information at the classification and class levels have hierarchical relationships:

$$H(D|A) = \sum_{j=1}^{m} H(D_j|A) \geqslant H(D_j|A),\ I(A;D) = \sum_{j=1}^{m} I(A;D_j) \geqslant I(A;D_j). \tag{22}$$

**Proposition 8.** $H(D_j), H(D_j|A), I(A;D_j)$ satisfy a relationship:

$$I(A;D_j) = H(D_j) - H(D_j|A). \tag{23}$$

By Lemma 1, Eq. (20) resorts to the summation commutativity to change the granular order of double summations, and thus $H(D|A)$ and $I(A;D)$ at the $\pi_D$-classification level effectively embrace and express the internal information of class $D_j$. By Definition 7, Eq. (21) naturally extracts and defines two corresponding measures at the $D_j$-class level (i.e., conditional entropy $H(D_j|A)$ and mutual information $I(A;D_j)$). Their hierarchical connections regarding decomposition/integration and size become natural in Proposition 7, while their systematic relationship with $H(D_j)$ is also clear in Proposition 8.

Commutation, decomposition, and extraction are rational because $H(D|A), I(A;D), H(A|D), H(D)$ concern the decision parts and relevant summation $\sum_{j=1}^{m}$. There are two surplus measures (information entropy $H(A)$ and mutual information $I(D;A)$), and $H(A)$ concerns only a single summation of the condition part, i.e.,

$$H(A) = -\sum_{i=1}^{n(A)} p(A_i)\log_2 p(A_i).$$

Thus, how to extract or define the information entropy for a decision class becomes difficult; furthermore, how to extract or define $I(D_j;A)$ by $I(D;A) = H(A) - H(A|D)$ also becomes a relevant question. For these construction issues, the systematic Eq. (11) regarding classification $\pi_D$ can be used, and we establish in advance a perfect system equation (with a variant isomorphism):

$$H(D_j) - H(D_j|A) = I(A;D_j) = I(D_j;A) = H_{D_j}(A) - H(A|D_j). \tag{24}$$

In other words, Eq. (24) can be rationally required by the hierarchical transmission of system relationships, and its transformation naturally inspires two natural formulae for definitions:

$$\begin{aligned} &I(D_j;A) = I(A;D_j), \\ &H_{D_j}(A) = I(D_j;A) + H(A|D_j) = I(A;D_j) + H(A|D_j). \end{aligned} \tag{25}$$

Next, we make the relevant construction and definition by system switching.

**Lemma 2.** *Entropy $H(D_j)$ has an expression on granulation summation:*

$$H(D_j) = -\sum_{i=1}^{n(A)} p(A_i)p(D_j|A_i)\log_2 p(D_j). \tag{26}$$

*Furthermore, $I(A;D_j)$ yields*

$$I(A;D_j) = H(D_j) - H(D_j|A) = -\sum_{i=1}^{n(A)} p(D_j \cap A_i)\log_2 \frac{|D_j| \times |A_i|}{|U| \times |D_j \cap A_i|}, \tag{27}$$

*and we have*

$$I(A;D_j) + H(A|D_j) = -\sum_{i=1}^{n(A)} p(D_j|A_i)p(A_i)\log_2 p(A_i). \tag{28}$$

**Definition 8.** Information entropy of partition $\pi_A$ on class $D_j$ and mutual information between partition $\pi_A$ and class $D_j$ are respectively defined by:

$$\begin{aligned} &H_{D_j}(A) = -\sum_{i=1}^{n(A)} p(D_j|A_i)p(A_i)\log_2 p(A_i), \\ &I(D_j;A) = -\sum_{i=1}^{n(A)} p(D_j \cap A_i)\log_2 \tfrac{|D_j| \times |A_i|}{|U| \times |D_j \cap A_i|}. \end{aligned} \tag{29}$$

**Proposition 9.** The information entropy and mutual information at the classification and class levels have hierarchical relationships:

$$H(A)=\sum_{j=1}^{m}H_{D_j}(A)\geqslant H_{D_j}(A),\ \ I(D;A)=\sum_{j=1}^{m}I(D_j;A)\geqslant I(D_j;A). \tag{30}$$

**Proof.** Only $\sum_{j=1}^{m}H_{D_j}(A)=H(A)$ is required to be proved, and we have

$$\begin{aligned}\sum_{j=1}^{m}H_{D_j}(A)&=-\sum_{j=1}^{m}\left(\sum_{i=1}^{n(A)}p(D_j|A_i)p(A_i)\log_2 p(A_i)\right)=-\sum_{j=1}^{m}\sum_{i=1}^{n(A)}p(D_j|A_i)p(A_i)\log_2 p(A_i)\\&=-\sum_{i=1}^{n(A)}\left(\sum_{j=1}^{m}p(D_j|A_i)p(A_i)\log_2 p(A_i)\right)=-\sum_{i=1}^{n(A)}\left(p(A_i)\log_2 p(A_i)\sum_{j=1}^{m}p(D_j|A_i)\right)=-\sum_{i=1}^{n(A)}p(A_i)\log_2 p(A_i)=H(A).\end{aligned} \tag{31}$$

□

**Proposition 10.** At the class level, six types of information measures have a basic relationship:

$$H(D_j)-H(D_j|A)=I(A;D_j)=I(D_j;A)=H_{D_j}(A)-H(A|D_j). \tag{32}$$

Lemma 2's proof is provided in Appendix A. According to the construction of Eqs. (24) (25), Eq. (26) of Lemma 2 transfers $H(D_j)$ into a form on $\sum_{i=1}^{n(A)}$ to match the granulation summation of $H(D_j|A)$, and then Eqs. (27) (28) of Lemma 2 respectively endow $I(A;D_j)$ and $I(A;D_j)+H(A|D_j)$ with styles on $\sum_{i=1}^{n(A)}$. Furthermore, Definition 8 naturally proposes entropy $H_{D_j}(A)$ and mutual information $I(D_j;A)$, Proposition 9 provides relevant hierarchical relationships, while Proposition 10 shows their systematicness with previous measures.

Herein, $H_{D_j}(A)$ is indirectly defined by a systematic transformation, i.e., Eq. (25), and its direct decomposition from $H(A)$ is indeed difficult based on the hierarchical connection proof (i.e., Eq. (31)).

(1) According to the inverse deduction of Eq. (31), the hierarchical decomposition of $H(A)$ actually concerns the decomposition of both granular summation $\sum_{i=1}^{n(A)}$ and class summation $\sum_{j=1}^{m}$ as well as relevant commutation; thus, the decomposition strategy for $H_{D_j}(A)$ is complex.

(2) In contrast, the conversion strategy for $H_{D_j}(A)$ becomes easy because we only need granular summation $\sum_{i=1}^{n(A)}$ of $H(D_j)$ as well as certain subsequent transformations. This conversion strategy is also useful when describing the systematic relationship.

(3) The direct decomposition and systematic switching respectively correspond to the reverse and forward deductions of Eq. (31); thus, their degrees regarding difficulty and heuristics are completely different. Based on the final result of Eq. (31), we can also say that $H_{D_j}(A)$ is hierarchically decomposed from $H(A)$.

The system function in Eq. (32) can be equivalently broken up into three groups of relationships:

$$I(D_j;A)=I(A;D_j),\ \ I(A;D_j)+H(D_j|A)=H(D_j),\ \ H_{D_j}(A)=I(A;D_j)+H(A|D_j). \tag{33}$$

The three connections can gain vivid diagrams, when we introduce five variables:

$$I(D_j;*),I(*;D_j),H(D_j|*),H(*|D_j),H_{D_j}(*)\quad(\forall *\in 2^C-\{\varnothing\}),$$

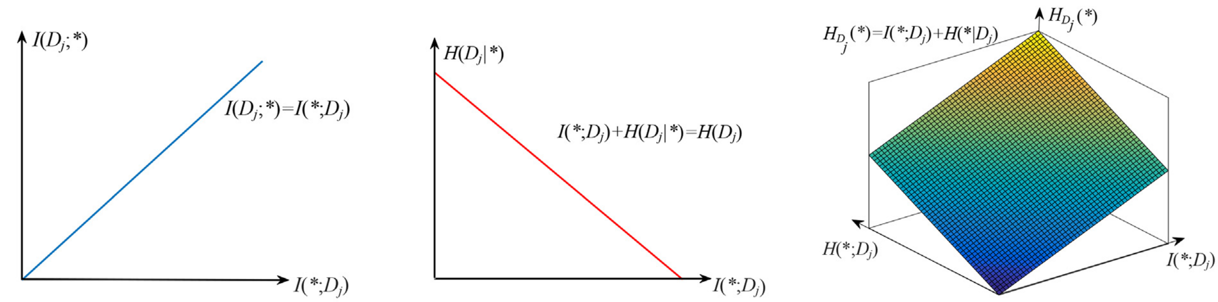
and the geometric characteristics are described in Fig. 2. Fig. 2 describes three groups of connections in Eq. (33) to deeply embody the systematicness in Eq. (32).

### 3.2. Comprehensive analysis

Thus far, information entropy $H_{D_j}(A),H(D_j)$, conditional entropy $H(D_j|A),H(A|D_j)$, and mutual information $I(A;D_j),I(D_j;A)$ at the class level have been systematically mined in the top-down direction, and they respectively correspond to $H(A),H(D),H(D|A),H(A|D)$, and $I(A;D),I(D;A)$ at the classification level, where the relevant hierarchical decomposition is concerned. As a result, they have a system equation at the class level, and Eq. (32) is completely similar to Eq. (11) at the classification level. This conclusion can be summarized to a variant isomorphism.

**Theorem 1. (Measure isomorphism)** *In terms of internal relations (i.e., Eqs. (11) (32)), the two information systems at the class and classification levels offer an isomorphism:*

$$\left(H_{D_j}(A),H(D_j);H(D_j|A),H(A|D_j);I(A;D_j),I(D_j;A)\right)\cong(H(A),H(D);H(D|A),H(A|D);I(A;D),I(D;A)). \tag{34}$$




(a) Two-dimensional line regarding $(I(A;D_j), I(D_j;A))$ (b) Two-dimensional line regarding $(I(A;D_j), H(D_j|A))$ (c) Three-dimensional plane regarding $(I(A;D_j), H(A|D_j), H_{D_j}(A))$

**Fig. 2.** Three geometric characteristics of systematic class-specific information measures.

**Proof.** This result is verified by Propositions 2 and 10. □

This metric isomorphism supports the proposed class-specific information measures in terms of the traditional classification-based information measures. As an example, the class-specific information measures carry corresponding uncertainty semantics from the top level to the middle level, and this translational strategy for semantics may outperform the direct and feasible approach. Concretely, a classification with multiple granules can be viewed as an information source; thus, the previous measures at Macro-Top $(\pi_A, \pi_D)$ concern two types of information sources, while the current measures at Meso-Middle $(\pi_A, D_j)$ concern only a type of information source and a fixed observation class.

(1) Information entropy $H_{D_j}(A)$ characterizes the average information content and uncertainty of classification $\pi_A$ regarding class $D_j$, while information entropy $H(D_j)$ directly represents the uncertainty of class $D_j$.
(2) Conditional entropy $H(D_j|A)$ quantifies the average information content and uncertainty of the premise classification $\pi_A$ and target class $D_j$, while conditional entropy $H(A|D_j)$ measures the average information content and uncertainty of the classification $\pi_A$ and premise class $D_j$.
(3) Mutual information $I(A;D_j)$ measures the information content from classification $\pi_A$ to class $D_j$, while mutual information $I(D_j;A)$ can be viewed to make a reverse description. $I(A;D_j)$ and $I(D_j;A)$ have the same connotation (i.e., $I(D_j;A)$ is mainly equal to $I(A;D_j)$) because they concern only a sole classification $\pi_A$. In contrast, $I(A;D)$ and $I(D;A)$ have different connotations but commutative equality because they concern two kinds of granulation, i.e., $\pi_A$ and $\pi_D$.

In summary, the class-specific information measures at the class level have been developed by three types of hierarchical decomposition and have the relevant analytical structure, uncertainty semantics, hierarchical relationship, and systematic equation in contrast to the previous classification-based information measures. The complete similarity of both information systems is related to the isomorphism. This study novelly transfers the information measure system from the classification level to the class level and thus highlights the class-specific uncertainty measurement and attribute reduction. Following the classical approach, only a set of class-specific measures (entropy $H_{D_j}(A)$, conditional entropy $H(D_j|A)$, and mutual information $I(A;D_j)$) is more focused on or mainly used, especially when constructing the class-specific attribute reducts.

In Ref. [71], three-way weighted entropies at Meso-Middle $(\pi_A, D_j)$ are proposed by the Bayes' system formula and bottom-middle integration evolution. These measures at the class level involve (1) likelihood weighted entropy $H_W(D_j|A)$ (or likelihood-linear weighted entropy $H_W^{\text{lin}}(D_j|A)$), (2) prior weighted entropy $H_W^{D_j}(A)$, and (3) posterior weighted entropy $H_W(A|D_j)$. Next, close connections between these new measures and previous weighted entropies are shown to reinforce the rationality and significance of the proposed new measures.

**Theorem 2.** *At the class level, two sets of information measures have the following equivalence:*

$$H_{D_j}(A) \equiv H_W^{D_j}(A),\ H(D_j|A) \equiv H_W(D_j|A),\ H(A|D_j) \equiv H_W(A|D_j),\ I(A;D_j) \equiv H_W^{\text{lin}}(D_j|A). \tag{35}$$

**Proof.** The equivalence is verified by comparing the decomposed definition of three-way information measures in Definitions 6–8 and the integrated definitions of three-way weighted entropies in [71]. □

Theorem 2 describes the equivalence between two systems of information measures. As main measures, information entropy $H_{D_j}(A)$, conditional entropies $H(D_j|A)$ and $H(A|D_j)$, and mutual information $I(A;D_j)$ are respectively equivalent to prior weighted entropy $H_W^{D_j}(A)$, likelihood weighted entropy $H_W(D_j|A)$, posterior weighted entropy $H_W(A|D_j)$, and likelihood-linear weighted entropy $H_W^{\text{lin}}(D_j|A)$. The two metric systems originate from two different strategies at three levels: the hierarchical decomposition in the top-middle direction and the hierarchical integration in the bottom-middle direction. However, they achieve the equivalence form at the same level regarding Meso-Middle; this conclusion describes both the concordance of three-level information construction and the rationality of two systems of middle measures. The new system at the class level both degrades the classical system of information measures in Definition 4 and promotes the existing system of weighted entropies in [71], thus retaining the fundamental connection significance between Macro-Top and Meso-Middle.

According to Theorem 2, class-specific information measures can directly achieve in-depth properties by virtue of relevant features of three-way weighted entropies. To construct attribute reducts, we mainly focus on the granulation monotonicity. By referring to corresponding results in [71,72], Corollaries 1 and 2 respectively provide the relevant granulation monotonicity and optimization condition for the primary measures: information entropy $H_{D_j}(A)$, conditional entropy $H(D_j|A)$, and mutual information $I(A;D_j)$. In this study, attribute relation $A \subseteq B \subseteq C$ naturally induces a process of granulation coarsening: $\pi_B \xrightarrow{\preceq} \pi_A$. This process usually contains multiple groups of granular merging [69]. Thus, attribute deletion $C \xrightarrow{-(C-R)} R$ and knowledge coarsening $\pi_C \xrightarrow{\preceq} \pi_R$ are presumed to be accompanied by a representative group of granular merging:

$$\bigcup_{t=1}^{k}[x]_C^t = [x]_R \ (k \geqslant 2). \tag{36}$$

This formula implies that $k$ original granules regarding $C$ (i.e., $[x]_C^1, \cdots, [x]_C^k$) are merged into an ultima granule regarding $R$ (i.e., $[x]_R$), where granular merging plays an important role in knowledge-based granulation mining [69]. Moreover, the inverse granulation refinement can gain the symbol $\xrightarrow{\succeq}$ and relevant descriptions.

**Corollary 1.** *Let $A \subseteq B \subseteq C$ and $\pi_D = \{D_1, \cdots, D_m\}$. For any $j = 1, \cdots, m$,*

$$\begin{aligned} &(i) \quad H_{D_j}(A) \leqslant H_{D_j}(B);\\ &(ii) \quad H(D_j|A) \geqslant H(D_j|B);\\ &(iii) \quad I(A;D_j) \leqslant I(B;D_j). \end{aligned} \tag{37}$$

**Corollary 2.** *Regarding $R \subseteq C$ and its coarsening $\pi_C \xrightarrow{\preceq} \pi_R$,*

$$\begin{aligned} &H_{D_j}(R) = H_{D_j}(C) \Longleftrightarrow \forall(\bigcup_{t=1}^{k}[x]_C^t = [x]_R), \quad \forall t \in \{1, \cdots, k\}\left[p(D_j \cap [x]_C^t) = p(D_j \cap [x]_R) = 0\right],\\ &H(D_j|R) = H(D_j|C) \wedge I(R;D_j) = I(C;D_j) \Longleftrightarrow \forall(\bigcup_{t=1}^{k}[x]_C^t = [x]_R), \quad \forall t \in \{1, \cdots, k\}\left[p(D_j|[x]_C^t) = p(D_j|[x]_R)\right], \end{aligned} \tag{38}$$

*where $\bigcup_{t=1}^{k}[x]_C^t = [x]_R$ ($k \geqslant 2$) denotes an arbitrary group of granular merging. Herein, the two granular probability descriptions on the right side of Eq. (38) are respectively called IP0-Condition and LPE-Condition [72].*

As shown in Corollary 1, the information entropy, conditional entropy, and mutual information at the class level exhibit relevant granulation monotonicity (Eq. (37)), and this important feature benefits from the integration of monotonicity in granular merging. The class-specific granulation monotonicity cannot directly come from the hierarchical decomposition of monotonicity at Macro-Top (Eq. (12)); conversely, the former monotonicity can jointly derive the latter. Moreover, the preservation condition regarding $A \subseteq B$ in Corollary 1 naturally yields the optimization condition regarding $R \subseteq C$ in Corollary 2. Regarding Corollary 2, each group of granular merging $\bigcup_{t=1}^{k}[x]_C^t = [x]_R$ is required to satisfy certain probability conditions. As a result, the two corollaries underlie the monotonic construction and systematic relationship of the following class-specific reducts.

## 4. Class-specific attribute reducts from the informational viewpoint

In the above section, three types of information measures (the information entropy, conditional entropy, and mutual information) have been systematically established at the class level and have fundamental uncertainty semantics and perfect granulation monotonicity. In this section, these measures are fully utilized to constitute a new system with three types of informational class-specific reducts, and internal relationships and basic properties are first researched after definitions are provided.

**Definition 9.**

(1) $R \subseteq C$ is an IE-class-specific reduct (a class-specific reduct on information entropy) if it satisfies two conditions regarding sufficiency and necessity:

$$\begin{cases} (s1)\ H_{D_j}(R) = H_{D_j}(C),\\ (n1)\ \forall r \in R[H_{D_j}(R-\{r\}) < H_{D_j}(R)]. \end{cases}$$

The set of all these reducts is denoted by $RED^{IE}(D_j)$.

(2) $R \subseteq C$ is a CE-class-specific reduct (a class-specific reduct on conditional entropy) if it satisfies two conditions:

$$\begin{cases} (s2)\ H(D_j|R) = H(D_j|C), \\ (n2)\ \forall r \in R[H(D_j|(R - \{r\})) > H(D_j|R)]. \end{cases}$$

The set of all these reducts is denoted by $RED^{CE}(D_j)$.

(3) $R \subseteq C$ is an MI-class-specific reduct (a class-specific reduct on mutual information) if it satisfies two conditions:

$$\begin{cases} (s3)\ I(R; D_j) = I(C; D_j), \\ (n3)\ \forall r \in R[I(R - \{r\}; D_j) < I(R; D_j)]. \end{cases}$$

The set of all these reducts is denoted by $RED^{MI}(D_j)$.

Three types of attribute cores are defined by:

$$\begin{aligned} &CORE^{IE}(D_j) = \{c \in C | H_{D_j}(C - \{c\}) < H_{D_j}(C)\}, \\ &CORE^{CE}(D_j) = \{c \in C | H(D_j|(C - \{c\})) > H(D_j|C)\}, \\ &CORE^{MI}(D_j) = \{c \in C | I(C - \{c\}; D_j) < I(C; D_j)\}. \end{aligned} \tag{39}$$

At the class level, Definition 9 utilizes the three information measures and two reduction conditions to typically define three types of class-specific reducts. $R \subseteq C$ implies the $C$-original coarsening $\pi_C \xrightarrow{\preceq} \pi_R$ and relevant granulation monotonicity, and thus corresponding reducts are minimum attribute subsets to maintain optimal information values. The three types of reducts are similar to those at the classification level in Definition 5.

**Lemma 3.** *Regarding $R \subseteq C$ and its coarsening $\pi_C \xrightarrow{\preceq} \pi_R$,*

$$H_{D_j}(R) = H_{D_j}(C) \Rightarrow H(D_j|R) = H(D_j|C) \Longleftrightarrow I(R; D_j) = I(C; D_j). \tag{40}$$

**Proof.** At first, $H(D_j|R) = H(D_j|C) \Longleftrightarrow I(R; D_j) = I(C; D_j)$ naturally holds according to basic relation $I(A; D_j) = H(D_j) - H(D_j|A)$, where $H(D_j)$ is a constant at the fixed level of class $D_j$. According to Eq. (38),

$$\forall (\bigcup_{t=1}^{k} [x]_C^t = [x]_R), \forall t \in \{1, \cdots, k\} \left[p(D_j \cap [x]_C^t) = p(D_j \cap [x]_R) = 0 \Rightarrow p(D_j|[x]_C^t) = p(D_j|[x]_R)\right]. \tag{41}$$

IP0-Condition thus necessarily leads to LPE-Condition regarding granular merging $\bigcup_{t=1}^{k} [x]_C^t = [x]_R$ in Corollary 2. Therefore, $H_{D_j}(R) = H_{D_j}(C) \Rightarrow H(D_j|R) = H(D_j|C) \wedge I(R; D_j) = I(C; D_j)$. □

**Theorem 3.** *IE-class-specific reducts are stronger than CE-class-specific reducts and MI-class-specific reducts, while the latter two are equivalent. Thus, we have basic relationships:*

$$\begin{aligned} &(1) \quad \forall R^{IE} \in RED^{IE}(D_j), \exists R^{CE} \in RED^{CE}(D_j), \exists R^{MI} \in RED^{MI}(D_j), \quad s.t., \quad R^{CE} \subseteq R^{IE}, \quad R^{MI} \subseteq R^{IE}; \\ &(2) \quad RED^{CE}(D_j) \equiv RED^{MI}(D_j); \\ &(3) \quad CORE^{IE}(D_j) \supseteq CORE^{CE}(D_j) \equiv CORE^{MI}(D_j). \end{aligned}$$

**Proof.** In terms of the strong–weak reduction theory [70], this conclusion can be proven by Lemma 3 regarding reduction targets and Proposition 5 regarding reduction manifestations. □

Lemma 3 provides the strong–weak relationships of information reduction targets. The preservations of conditional entropy and multiple information become equivalent, while the preservation of information entropy becomes stronger to derive the former two. Accordingly, Theorem 3 provides the strong–weak relationships of information reducts as well as relevant descriptions, and these results are similar to those of informational classification-based reducts. According to this equivalence, the three types of information measures essentially provide only two types of informational class-specific reducts (IE-class-specific reducts and MI-class-specific reducts (or CE-class-specific reducts)). Because mutual information has better uncertainty semantics between classification $\pi_A$ and class $D_j$, MI-class-specific reducts are more focused on than IE-class-specific reducts, and this new type of class-specific reducts is utilized to completely omit its equivalent CE-class-specific reducts.

Basic properties of informational class-specific reducts are now discussed via granulation monotonicity. Representative MI-class-specific reducts are mainly considered, while the two surplus types have similar results.

**Lemma 4.** *Regarding mutual information, the individual necessity has two equivalent conditions:*

$(n3) \quad \forall r \in R[I(R - \{r\}; D_j) < I(R; D_j)],$

$(n3') \quad \forall R' \subset R[I(R'; D_j) < I(R; D_j)].$

**Proof.** (1) First consider $(n3) \Rightarrow (n3')$. $\forall R' \subset R, \exists r \in R - R' \subset R$, s.t., $R' \subseteq R - \{r\}$. According to granulation monotonicity (Eq. (37)), $I(R'; D_j) \leqslant I(R - \{r\}; D_j)$. According to condition $(n3), I(R - \{r\}; D_j) < I(R; D_j)$. Hence, $I(R'; D_j) < I(R; D_j)$, i.e., condition $(n3')$ holds. (2) Then prove $(n3') \Rightarrow (n3)$. $\forall r \in R, R' = R - \{r\} \subset R$. According to condition $(n3'), I(R - \{r\}; D_j) = I(R'; D_j) < I(R; D_j)$, so condition $(n3)$ holds. □

**Proposition 11.** $R \in RED^{MI}(D_j)$ if and only if $R$ satisfies two conditions of $(s3)$ and $(n3')$.

**Proposition 12.** Regarding MI-class-specific reducts, the core becomes the intersection of all reducts:

$$CORE^{MI}(D_j) = \bigcap_{R \in RED^{MI}(D_j)} R.$$

**Proof.** (1) If $c \notin \bigcap_{R \in RED^{MI}(D_j)} R$, so $\exists R \in RED^{MI}(D_j)$ $(c \notin R)$. Thus, $I(R; D_j) = I(C; D_j)$ and $R \subseteq C - \{c\} \subset C$. According to granulation monotonicity (Eq. (37)), $I(C - \{c\}; D_j) = I(C; D_j)$, so $c \notin CORE^{MI}(D_j)$. Hence, $\bigcap_{R \in RED^{MI}(D_j)} R \supseteq CORE^{MI}(D_j)$. (2) If $c \notin CORE^{MI}(D_j)$, so $I(C - \{c\}; D_j) = I(C; D_j)$. Suppose there exists an MI-class-specific reduct $R \subseteq C - \{c\}$ for $C - \{c\}$. Thus, $I(R; D_j) = I(C - \{c\}; D_j) = I(C; D_j)$, and $R$ satisfies the individual necessity condition, i.e., $(n3)$, so $R \in RED^{MI}(D_j)$. However, $c \notin R$ and $c \notin \bigcap_{R \in RED^{MI}(D_j)} R$. Therefore, $CORE^{MI}(D_j) \supseteq \bigcap_{R \in RED^{MI}(D_j)} R$. According to the above two items with double inclusions, we prove $CORE^{MI}(D_j) = \bigcap_{R \in RED^{MI}(D_j)} R$. □

**Definition 10.** Regarding mutual information, the significance of attribute $c \in C - R$ on $R$ is defined by:

$$Sig^{MI}(R, c, D_j) = I(R \cup \{c\}; D_j) - I(R; D_j). \tag{42}$$

In terms of mutual information, $Sig^{MI}(R, c, D_j)$ represents the importance degree of attribute $c$ regarding class $D_j$ when it is added to basic attribute subset $R$; thus, it can be utilized to develop a heuristic algorithm for relevant reduction.

**Algorithm 1.** (MI-CSR) A heuristic algorithm for an MI-class-specific reduct

**Input:** A decision table $(U, C \cup D)$ and a class index $j \in \{1, 2, \cdots, m\}$.
**Output:** An MI-class-specific reduct $R \in RED^{MI}(D_j)$.

```
1: Calculate CORE^{MI}(D_j).
2: Let R = CORE^{MI}(D_j).
3: while I(R; D_j) < I(C; D_j) do
4:    ∀c ∈ C − R, calculate Sig^{MI}(R, c, D_j); choose c_0 = arg max_{c∈C−R} Sig^{MI}(R, c, D_j), and let R ← R ∪ {c_0}.
5: end while
6: R* = R.
7: for each r ∈ R* do
8:    if I(R − {r}; D_j) = I(R; D_j) then
9:       R = R − {r}.
10:   end if
11: end for
12: return R.
```

Algorithm 1 concerns two stages.

(1) The first stage makes the addition construction based on $CORE^{MI}(D_j)$ and the heuristic search based on $Sig^{MI}(R, c, D_j)$, and includes Steps 1–5. Step 1 provides $CORE^{MI}(D_j)$, and Step 2 chooses the core as the starting point for construction. In Steps 3–5 with a "while" loop, the addition attribute $c_0$ is chosen by the highest heuristic information of $Sig^{MI}(R, c, D_j)$ to

perform quick research, and thus a superset of $CORE^{MI}(D_j)$ is constructed to satisfy the reduction condition of joint sufficiency (condition $(s3)$).

(2) The second stage conducts the backward deletion for attribute redundancy. Step 6 stores $R$ in $R^*$ for loop search. Steps 7–11 use a "for" loop to sequentially delete unnecessary attributes; thus, the remainder subset $R$ satisfies the reduction condition of individual necessity (condition $(n3)$).

As a result, Algorithm 1 effectively yields an MI-class-specific reduct that satisfies both conditions $(s3)$ and $(n3)$ as output in Step 12. In this study, associated actions based on mutual information are viewed as basic operations to estimate the computational complexity regarding condition attributes and their number $|C|$. For calculation times, Step 1 requires $|C|$ comparisons; in the worst case, Step 4 includes $|C|$ subtractions, $|C|-1$ comparisons, and 1 renewal. Furthermore, the "while" loop in Steps 4–5 concerns $(1+2|C|)|C|$ operations; the "for" loop in Steps 7–11 has an upper bound of $2|C|$ operations. Thus, the time complexity follows $T(|C|) = 2|C|^2 + 3|C|$ in the worst case to offer an asymptotic analysis of polynomial complexity (i.e., $T(|C|) = O(|C|^2)$), while the space complexity similarly exhibits $S(|C|) = O(|C|^2)$. This algorithm is also feasible in terms of calculation complexity.

Regarding MI-class-specific reducts, Lemma 4 provides an equivalent condition of necessity, and thus Proposition 11 provides an equivalent form of reducts. Proposition 12 provides the relationship between the core and reduct, and thus the computable core in each reduct serves as an important basis for reduct construction. Accordingly, the core in Eq. (39) and the attribute significance in Eq. (42) are used to design a heuristic algorithm (i.e., Algorithm 1), which can efficiently seek an MI-class-specific reduct. Note that these results are conventional and that they can be similarly obtained for IE-class-specific reducts as well as CE-class-specific reducts. Moreover, relevant descriptions are neglected. In particular, the idea and framework of Algorithm 1 can be used to develop heuristic algorithms of other reducts based on monotonic measures. Accordingly, the heuristic algorithms and relevant labels are uniformly used for both the classification-basic and class-specific reducts based on algebraic dependency degree or information measures, mainly in later examples and experiments.

## 5. Systematic connections embracing informational class-specific attribute reducts

The informational class-specific reducts have been systematically developed by three types of information measures at the class level. In this section, we show their relevant mode connections, mainly their transverse connections with the algebraic class-specific reducts and their hierarchical connections with the informational classification-based reducts. The two types of reduct connections have been previously described and are required in Fig. 1 and constitute two subsections.

### 5.1. Transverse connections between informational and algebraic class-specific attribute reducts

Herein, internal relationships of entire class-specific reducts are discussed from the strength-weakness perspective, which serves as a basic technology of reduct comparison [70]. For this purpose, multiple reduction targets and their strength revelation become both key and difficult. The attribute reduction concerns knowledge coarsening. According to [69], knowledge coarsening consists of two types of granular actions: granular preservation and granular merging. However, the former has no impact on knowledge coarsening; thus, only the latter causes knowledge coarsening and its measurement. In terms of knowledge coarsening, both the granulation monotonicity and preservation strength can be shown by focusing on only a representative of granular merging. As previously stated in Corollary 2 and its explanation, the measure preservation of reduction targets requires knowledge coarsening $\pi_C \xrightarrow{\preceq} \pi_R$ and its representative granular merging $\bigcup_{t=1}^{k} [x]_C^t = [x]_R$ $(k \geqslant 2)$.

**Lemma 5.** *Regarding $R \subseteq C$ and its coarsening $\pi_C \xrightarrow{\preceq} \pi_R$,*

$$\begin{array}{l} H(D_j|R) = H(D_j|C) \Longleftrightarrow I(R;D_j) = I(C;D_j) \Rightarrow POS(D_j|R) = POS(D_j|C) \Longleftrightarrow \gamma_R(D_j) = \gamma_C(D_j), \\ POS(D_j|R) = POS(D_j|C) \Longleftrightarrow \gamma_R(D_j) = \gamma_C(D_j) \not\Rightarrow H(D_j|R) = H(D_j|C) \Longleftrightarrow I(R;D_j) = I(C;D_j). \end{array} \quad (43)$$

**Corollary 3.** *Regarding $R \subseteq C$ and its coarsening $\pi_C \xrightarrow{\preceq} \pi_R$,*

(1) *If $D_j$ is a consistent class [73], then the preservations of mutual information (conditional entropy) and dependency degree (positive region) become equivalent, i.e.,*

$$I(R;D_j) = I(C;D_j) \Longleftrightarrow H(D_j|R) = H(D_j|C) \Longleftrightarrow \gamma_R(D_j) = \gamma_C(D_j) \Longleftrightarrow POS(D_j|R) = POS(D_j|C).$$

(2) *Otherwise, $D_j$ is an inconsistent class [73], and then the preservations of mutual information (conditional entropy) and dependency degree (positive region) may be different, where the former is stronger than the latter.*

**Proof.** Regarding the consistent class $D_j$, which is equivalent to $BND(D_j|C) = \varnothing$ [73], the latter two cases in Lemma 5's proof in Appendix B never emerge, and thus the strength-weakness relation degenerates into the equivalent case. Regarding the inconsistent class $D_j$, which is equivalent to $BND(D_j|C) \neq \varnothing$ [73], the latter two cases may exist and thus the strength-weakness property still holds to imply the difference. □

Lemma 5's proof is offered in Appendix B, while the consistent and inconsistent classes are proposed in [73] to respectively represent the restrictive features of consistent and inconsistent decision tables on a specific decision class. Thus, Lemma 5 and Corollary 3 show the relationships between the preservations of mutual information (conditional entropy) and positive region (dependency degree). The informational preservation is usually stronger than the algebraic preservation; in the special case of a consistent class, the former degenerates into the latter to achieve the equivalence. On this basis, Theorem 4 naturally provides relationships between MI-class-specific/CE-class-specific reducts and algebraic class-specific reducts (the usual strength-weakness difference and the degenerated equivalence). Furthermore, Corollary 4 derives the strength-weakness relationship between IE-class-specific reducts and algebraic class-specific reducts because Theorem 3 shows that IE-class-specific reducts are stronger than MI-class-specific/CE-class-specific reducts.

**Theorem 4.** *MI-class-specific/CE-class-specific reducts are stronger than the algebraic class-specific reducts, and thus:*

$$(1)\quad \forall R \in RED^{MI}(D_j) \equiv RED^{CE}(D_j), \quad \exists R^{AV} \in RED^{AV}(D_j), \quad s.t., \quad R^{AV} \subseteq R;$$
$$(2)\quad CORE^{MI}(D_j) \equiv CORE^{CE}(D_j) \supseteq CORE^{AV}(D_j).$$

*Particularly, if the decision class is consistent [73], then the three types of reducts are equivalent, and thus:*

$$RED^{MI}(D_j) \equiv RED^{CE}(D_j) \equiv RED^{AV}(D_j),$$
$$CORE^{MI}(D_j) \equiv CORE^{CE}(D_j) \equiv CORE^{AV}(D_j).$$

**Proof.** The conclusion holds by using Lemma 5, Corollary 3, and Proposition 5. □

**Corollary 4.** *IE-class-specific reducts are stronger than algebraic class-specific reducts, and thus:*

$$(1)\quad \forall R^{IE} \in RED^{IE}(D_j), \exists R^{AV} \in RED^{AV}(D_j), \quad s.t., \quad R^{AV} \subseteq R^{IE};$$
$$(2)\quad CORE^{IE}(D_j) \supseteq CORE^{AV}(D_j).$$

We now summarize the systematic relationships of class-specific reducts by relevant results concluded in Fig. 3. In total, there are four types of class-specific reducts, which come from the algebraic region/measure, information entropy, conditional entropy, and mutual information, while the latter three constitute the informational class-specific reducts. Theorems 3 and 4 and Corollary 4 are summarized in Fig. 3, wherein arrows denote the reduct relationships from the strong to weak. As a result, IE-class-specific reducts are strongest, and the algebraic class-specific reducts are weakest, while the middle two (i.e., MI-class-specific and CE-class-specific reducts) are equivalent and thus have the same reduction strength. The strength theory of reducts can generate a structural algorithm from a strong reduct to a weak one. As an example, Algorithm 2 yields an algebraic class-specific reduct in a given MI-class-specific reduct; when considering the basic operations related to condition attributes, there exists only a "for" loop in Steps 2–6 to have an upper bound $2|R^{MI}|$ that is less than or equal to $2|C|$; thus, the temporal and spatial complexities are respectively $T(|C|) = O(|C|)$ and $S(|C|) = O(|C|)$ in terms of asymptotic analysis.

**Algorithm 2.** (MI-CSR→AV-CSR) A constructional algorithm from an MI-class-specific reduct to an algebraic class-specific reduct

**Input:** A decision table $(U, C \cup D)$, a class index $j \in \{1, 2, \cdots, m\}$, and an MI-class-specific reduct $R^{MI} \in RED^{MI}(D_j)$.
**Output:** An algebraic class-specific reduct $R \in RED^{AV}(D_j)$ satisfying $R \subseteq R^{MI}$.

1: Let $R = R^{MI}$.
2: **for** each $r \in R^{MI}$ **do**
3: **if** $\gamma_{R-\{r\}}(D_j) = \gamma_R(D_j)$ **then**
4: $R = R - \{r\}$.
5: **end if**
6: **end for**
7: **return** $R$.

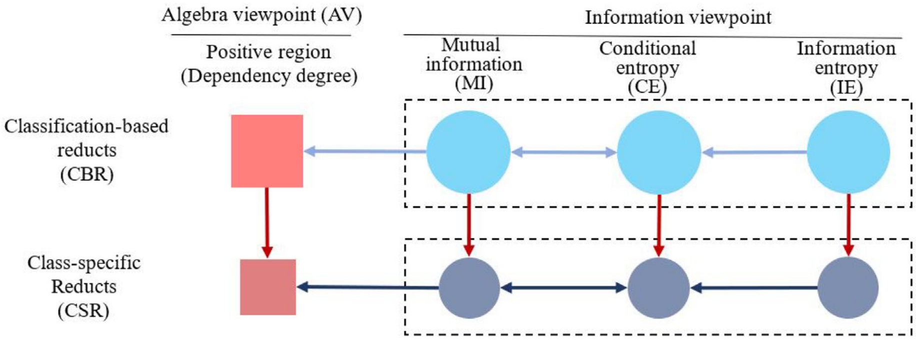



**Fig. 3.** Strength-weakness relationships of class-specific and classification-based reducts from informational and algebraic viewpoints.

**Theorem 5. (Reduct-level isomorphism)** *In terms of the reduction strength order, the two hierarchical systems of class-specific and classification-based reducts offer an isomorphism:*

$$\left(RED^{AV}(D_j); RED^{MI}(D_j), RED^{CE}(D_j), RED^{IE}(D_j)\right) \cong \left(RED^{AV}(\pi_D); RED^{MI}(\pi_D), RED^{CE}(\pi_D), RED^{IE}(\pi_D)\right). \tag{44}$$

**Proof.** The conclusion holds by observing these relevant results or their summary Fig. 3. □

These systematic relationships of class-specific reducts are in complete accordance with those of classification-based reducts, which have been reviewed by Refs. [32,54], as discussed in Section 2. The two hierarchical systems of reducts are described by two horizontal levels in Fig. 3. Thus, the relevant hierarchical isomorphism is concluded in Theorem 5. Furthermore, regarding the corresponding hierarchical relationships between the two hierarchical systems, another isomorphism between two reduct perspectives must be investigated. As part of this discussion, we would like to provide a related conclusion about hierarchical reduction strength: the upper classification-based reducts are stronger than the lower class-specific reducts in terms of each hierarchical correspondence. This result is also labeled in the longitudinal direction of Fig. 3. Both the horizontal and longitudinal relationships of reduct strength-weakness are offered in Fig. 3, where the reduction strength relation actually has transitivity and reflexivity.

### *5.2. Hierarchical connections between informational class-specific and classification-based attribute reducts*

In terms of information measures, we now show the hierarchical connections between the class-specific and classification-based reducts, mainly the basic strength-weakness connections, the family-based balance, and the final perspective-transverse isomorphism. At each level of class or classification, there are four types of reducts: the algebraic mode concerns the dependency degree, while the information mode uses the mutual information, conditional entropy, and information entropy, as summarized in Fig. 3. For the latter three informational types of hierarchical correspondence, we use a generic expression. Let $\sharp \in \{MI, CE, IE\}$ broadly denote one information label of $MI, CE, IE$ (i.e., $\sharp$ implies *MI*, or *CE*, or *IE*). Thus, a uniform description is developed regarding the three-way informational measures and reducts.

**Theorem 6.** *Regarding the three-way information measures,*

$$\begin{gathered}
H(R) = H(C) \Longleftrightarrow \forall 1 \leqslant j \leqslant m[H_{D_j}(R) = H_{D_j}(C)],\\
\forall r \in R[H(R-\{r\}) \neq H(R)] \Longleftrightarrow \forall r, \exists 1 \leqslant j \leqslant m[H_{D_j}(R-\{r\}) < H_{D_j}(R)];\\
H(D|R) = H(D|C) \Longleftrightarrow \forall 1 \leqslant j \leqslant m[H(D_j|R) = H(D_j|C)],\\
\forall r \in R[H(D|(R-\{r\})) \neq H(D|R)] \Longleftrightarrow \forall r, \exists 1 \leqslant j \leqslant m[H(D_j|(R-\{r\})) > H(D_j|R)];\\
I(R;D) = I(C;D) \Longleftrightarrow \forall 1 \leqslant j \leqslant m[I(R;D_j) = I(C;D_j)],\\
\forall r \in R[I(R-\{r\};D) \neq I(R;D)] \Longleftrightarrow \forall r, \exists 1 \leqslant j \leqslant m[I(R-\{r\};D_j) < I(R;D_j)].
\end{gathered}$$

**Proof.** This conclusion can be reached by both the hierarchical integration/decomposition in Eqs. (22) and (30) and the granulation monotonicity in Eq. (37). □

**Corollary 5.** *Regarding $\sharp$-classification-based and $\sharp$-class-specific reducts,*
*(1) Suppose $R \in RED^{IE}(\pi_D)$ and $R_j \in RED^{IE}(D_j)$, and then*

$$R \in RED^{IE}(D_j) \Longleftrightarrow \forall r \in R[H_{D_j}(R - \{r\}) < H_{D_j}(R)],$$
$$R_j \in RED^{IE}(\pi_D) \Longleftrightarrow H(R_j) = H(C).$$

*(2) Suppose $R \in RED^{CE}(\pi_D)$ and $R_j \in RED^{CE}(D_j)$, and then*

$$R \in RED^{CE}(D_j) \Longleftrightarrow \forall r \in R[H(D_j|(R - \{r\})) > H(D_j|R)],$$
$$R_j \in RED^{CE}(\pi_D) \Longleftrightarrow H(D|R_j) = H(D|C).$$

*(3) Suppose $R \in RED^{MI}(\pi_D)$ and $R_j \in RED^{MI}(D_j)$, and then*

$$R \in RED^{MI}(D_j) \Longleftrightarrow \forall r \in R[I(R - \{r\}; D_j) < I(R; D_j)],$$
$$R_j \in RED^{MI}(\pi_D) \Longleftrightarrow I(R_j; D) = I(C; D).$$

Regarding the three-way informational measures and reducts, Theorem 6 focuses on the reduction targets to clarify the corresponding hierarchical relationships, mainly regarding the two reduction conditions of joint sufficiency and individual necessity, while Corollary 5 shows the corresponding transition of hierarchical reducts regarding all decision classes. For each measure related to $\sharp$, the information preservation at the classification level is equivalent to the information preservation at the class level. From the hierarchical perspective, the classification-based reducts become stronger, while the class-specific reducts can resort to the family of all decision classes to balance and bound the classification-based reducts. Relevant theories and algorithms are provided next.

**Theorem 7.** *For each type of informational reducts, the classification-based reducts are accordingly stronger than the class-specific reducts, and thus:*

$$(1) \quad \forall R \in RED^{\sharp}(\pi_D), \exists R_j \in RED^{\sharp}(D_j), s.t., R_j \subseteq R;$$
$$(2) \quad CORE^{\sharp}(\pi_D) \supseteq CORE^{\sharp}(D_j).$$

**Proof.** According to Theorem 6, when considering a fixed type $\sharp$ of informational reducts, the classification-based reduction target naturally derives the class-specific reduction target (i.e., the former is stronger than the latter). Hence, this theorem holds by Proposition 5, which comes from the strong–weak reduction theory [70]. □

**Algorithm 3.** (MI-CBR→MI-CSR) A construction algorithm from an MI-classification-based reduct to a family of MI-class-specific reducts

**Input:** A decision table $(U, C \cup D)$ and an MI-classification-based reduct $R \in RED^{MI}(\pi_D)$.
**Output:** A family of MI-class-specific reducts $(R_1, R_2, \cdots, R_m)$ satisfying $\forall j \in \{1, 2, \cdots, m\}[R_j \subseteq R]$.
1: **for** j = 1 to m **do**
2: $R_j = R$.
3: **for** each $r \in R$ **do**
4: **if** $I(R_j - \{r\}; D_j) = I(R_j; D_j)$ **then**
5: $R_j = R_j - \{r\}$.
6: **end if**
7: **end for**
8: **end for**
9: **return** $(R_1, R_2, \cdots, R_m)$.

Theorem 7 yields the strength-weakness conclusion, and all of these hierarchical strength-weakness relationships have been labeled in Fig. 3. Algorithm 3 utilizes a given MI-classification-based reduct to generate a family of MI-class-specific reducts; regarding the basic operations related to condition attributes, the inner "for" loop in Steps 3–7 requires $2|R| \in (0, 2|C|]$ in the worst case, and the dual loops have an upper bound of $(1 + 2|R|)m \in (0, (1 + 2|C|)m] \subseteq (0, (1 + 2|C|)|U|]$ where $|R| \leqslant |C|$ and $m \leqslant |U|$; thus, both the temporal and the spatial complexities are $O(|C||U|)$ in terms of asymptotic analysis. Moreover, similar algorithms can be produced for the other two kinds of reducts based on information entropy and conditional entropy.

**Theorem 8.** *For each type of informational reducts, the family union of class-specific reducts regarding all decision classes necessarily includes a classification-based reduct. Thus:*

(1) $\forall R_j \in RED^{\sharp}(D_j)(j \in \{1,2,\cdots,m\}), \exists R \in RED^{\sharp}(\pi_D), s.t., R \subseteq \bigcup_{j=1}^{m} R_j$;

(2) $CORE^{\sharp}(\pi_D) \subseteq \bigcup_{j=1}^{m} CORE^{\sharp}(D_j)$.

**Proof.** This result comes from the reduct condition connections in Theorem 6 as well as relevant reduct definitions. □

**Algorithm 4.** (MI-CSR→MI-CBR) A construction algorithm from a family of MI-class-specific reducts to an MI-classification-based reduct

**Input:** A decision table $(U, C \cup D)$ and a family of MI-class-specific reducts $(R_1, \cdots, R_m)$ $(\forall j \in \{1,2,\cdots,m\}, R_j \in RED^{MI}(D_j))$.

**Output:** An MI-classification-based reduct $R \in RED^{MI}(\pi_D)$ satisfying $R \subseteq \bigcup_{j=1}^{m} R_j$.

1: $R = \bigcup_{j=1}^{m} R_j$.
2: **for** each attribute $r \in \bigcup_{j=1}^{m} R_j$ **do**
3: **if** $I(R - \{r\}; D) = I(R; D)$ **then**
4: $R = R - \{r\}$;
5: **end if**
6: **end for**
7: **return** $R$.

Theorem 8 and Algorithm 4 describe the other direction, which is opposite to these strength-weakness from classification-based reducts to class-specific reducts (family). From the family of class-specific reducts and their union set, we can seek a classification-based reduct, and Algorithm 4 provides the algorithm regarding mutual information, which can induce the surplus two algorithms. In terms of the basic operations embracing the sole "for" loop, this algorithm concerns $2|\bigcup_{j=1}^{m} R_j| \in (0, 2|C|]$ in the worst case; thus, the temporal and spatial complexities are respectively $T(|C|) = O(|C|)$ and $S(|C|) = O(|C|)$. By summarizing these two sides, we can directly provide the further balance conclusion.

**Theorem 9.** *For each type of informational reducts, there exists a kind of balance between the classification-based reducts and a family of class-specific reducts, and then:*

(1) $\forall R \in RED^{\sharp}(\pi_D), \exists R_j \in RED^{\sharp}(D_j)(j \in \{1,2,\cdots,m\}), s.t., R = \bigcup_{j=1}^{m} R_j$;

(2) $CORE^{\sharp}(\pi_D) = \bigcup_{j=1}^{m} CORE^{\sharp}(D_j)$.

**Proof.** According to Theorem 7, $\forall R \in RED^{\sharp}(\pi_D), \exists R_j \in RED^{\sharp}(D_j)$, s.t., $R_j \subseteq R$, and thus $\bigcup_{j=1}^{m} R_j \subseteq R$. For this union set and according to Theorem 8, $\exists R' \in RED^{\sharp}(\pi_D)$, s.t., $R' \subseteq \bigcup_{j=1}^{m} R_j$. Hence, we achieve $R' \subseteq \bigcup_{j=1}^{m} R_j \subseteq R$. However, $R \in RED^{\sharp}(\pi_D), R' \in RED^{\sharp}(\pi_D)$, and $R' \subseteq R$ together deduce $R' = R$. Therefore, $R' = \bigcup_{j=1}^{m} R_j = R$.

Moreover, Theorem 7 can derive $CORE^{\sharp}(\pi_D) \supseteq \bigcup_{j=1}^{m} CORE^{\sharp}(D_j)$; with the addition of $CORE^{\sharp}(\pi_D) \subseteq \bigcup_{j=1}^{m} CORE^{\sharp}(D_j)$ in Theorem 8, we can naturally achieve $CORE^{\sharp}(\pi_D) = \bigcup_{j=1}^{m} CORE^{\sharp}(D_j)$. □

Theorem 9 resorts to the reduction strength-weakness and its reverse to provide an in-depth balance conclusion between two types of hierarchical reducts, where a family of all decision classes is fully used. According to Theorem 9 and its proof, the $\sharp$-classification-based reduct $R$ is input into Algorithm $\sharp$-CBR→$\sharp$-CSR to obtain $\sharp$-class-specific reducts' union $\bigcup_{j=1}^{m} R_j$. This union is then input into Algorithm $\sharp$-CSR→$\sharp$-CBR to recover the initial $R$; thus, the combination of Algorithms $\sharp$-CBR→$\sharp$-CSR and $\sharp$-CSR→$\sharp$-CBR generates an identity mapping in $RED^{\sharp}(\pi_D)$. In contrast, the other combination of Algorithms $\sharp$-CSR→$\sharp$-CBR and $\sharp$-CBR→$\sharp$-CSR cannot yield an identity mapping.

Based on the family of reducts in $RED^{\sharp}(D_j)$ $(j = 1, 2, \cdots, m)$, $RED^{\sharp}(\pi_D)$ can gain an upper bound by the union form, which comes from the balance property of Theorem 9, while $RED^{\sharp}(\pi_D)$ can supplement a lower bound by the intersection strategy from Theorem 7. The relevant conclusion is stated as follows.

**Theorem 10.** *$\sharp$-classification-based and $\sharp$-class-specific reducts have the following description of double bounds,*

$$\bigcap_{j=1}^{m} RED^{\sharp}(D_j) \subseteq RED^{\sharp}(\pi_D) \subseteq \left\{ \bigcup_{j=1}^{m} R_j | R_j \in RED^{\sharp}(D_j), \quad j = 1, 2, \cdots, m \right\}. \tag{45}$$

**Proof.** In terms of the direct proof, the two subset symbols (i.e., $\subseteq$) respectively come from Theorems 7 and 9. Moreover, the double-bound conclusion of the algebra-hierarchical reducts has been proven in Ref. [64], and we can mainly refer to those deductions on Eqs. (8)–(10) in that paper; thus, the proof can be utilized to similarly and effectively verify the double-bound result of the information–hierarchical reducts in this study. □

Conversely, all of these hierarchical results of reduct strength, balance, and bound regarding the information reducts are similar to those regarding the algebraic reducts in [64]. This conclusion describes the similarity of hierarchical development between the informational and algebraic approaches, as shown in Figs. 1 and 3, and its rationality benefits from the hierarchical relevance and integrated construction between the two levels of Macro-Top and Meso-Middle. As a result, we can directly extract a varietal isomorphism conclusion between the transverse viewpoints.

**Theorem 11. (Reduct-viewpoint isomorphism)** *In terms of the reduct-hierarchical relations, the two transverse systems of informational and algebraic reducts offer an isomorphism:*

$$\left(RED^{\sharp}(\pi_D), RED^{\sharp}(D_j)\right) \cong \left(RED^{AV}(\pi_D), RED^{AV}(D_j)\right). \tag{46}$$

**Proof.** This assertion can be completely summarized and verified by comparing these information–hierarchical reducts results (i.e., Theorems 6–10 and Corollary 5) and the matching algebra-hierarchical reducts conclusions in Ref. [64]. □

## 6. Example illustration and experiment verification

### *6.1. Example illustration*

These informational measures and reducts at the class level are first shown in decision tables.

**Example 1.** This example uses a consistent decision table $(U, C \cup D)$, provided in Table 3. Herein, we have $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\}, C = \{c_1, c_2, c_3, c_4, c_5\}$, while $D = \{d\}$ generates three decision classes in classification $\pi_D$:

$$D_1 = \{x_1, x_2, x_3\}, \quad D_2 = \{x_4, x_5, x_6\}, \quad D_3 = \{x_7, x_8, x_9\}.$$

The class-specific information measures are first investigated, and their systematicness, hierarchy, and monotonicity require verification based on attribute subsets and metric calculation. For this purpose, we resort to an attribute-enlargement chain:

$$C_1 = \{c_1\} \subset C_2 = \{c_1, c_2\} \subset \cdots \subset C_{|C|} = \{c_1, c_2, \cdots, c_{|C|}\}, \tag{47}$$

which can deeply and effectively probe the hierarchical feature and structural information by relevant granulation refining sequence:

$$U/\pi_{C_1} \succeq U/\pi_{C_2} \succeq \cdots \succeq U/\pi_{C_{|C|}}. \tag{48}$$

**Table 3**
Consistent decision table of Example 1.

| $U$ | $c_1$ | $c_2$ | $c_3$ | $c_4$ | $c_5$ | $d$ |
|---|---|---|---|---|---|---|
| $x_1$ | 0 | 1 | 3 | 0 | 0 | 1 |
| $x_2$ | 0 | 1 | 3 | 0 | 0 | 1 |
| $x_3$ | 0 | 1 | 3 | 2 | 2 | 1 |
| $x_4$ | 1 | 0 | 1 | 0 | 1 | 2 |
| $x_5$ | 0 | 0 | 1 | 0 | 1 | 2 |
| $x_6$ | 1 | 0 | 2 | 2 | 0 | 2 |
| $x_7$ | 1 | 0 | 1 | 0 | 2 | 3 |
| $x_8$ | 1 | 0 | 1 | 0 | 2 | 3 |
| $x_9$ | 1 | 1 | 2 | 2 | 1 | 3 |

**Table 4**
Measured values based on the attribute-addition chain of Example 1.

| Measure | $\{c_1\}$ | $\{c_1, c_2\}$ | $\{c_1, c_2, c_3\}$ | $\{c_1, c_2, c_3, c_4\}$ | $\{c_1, c_2, c_3, c_4, c_5\}$ |
|---|---|---|---|---|---|
| $\gamma_{C_k}(D)$ | 0 | 0.5556 | 0.6667 | 0.6667 | 1 |
| $\gamma_{C_k}(D_1)$ | 0 | 0.3333 | 0.3333 | 0.3333 | 0.3333 |
| $\gamma_{C_k}(D_2)$ | 0 | 0.1111 | 0.2222 | 0.2222 | 0.3333 |
| $\gamma_{C_k}(D_3)$ | 0 | 0.1111 | 0.1111 | 0.1111 | 0.3333 |
| $H(C_k)$ | 0.9911 | 1.7527 | 2.1133 | 2.4194 | 2.7255 |
| $H_{D_1}(C_k)$ | 0.3900 | 0.5283 | 0.5283 | 0.8344 | 0.8344 |
| $H_{D_2}(C_k)$ | 0.3184 | 0.6122 | 0.8805 | 0.8805 | 1.0566 |
| $H_{D_3}(C_k)$ | 0.2827 | 0.6122 | 0.7044 | 0.7044 | 0.8344 |
| $H(D)$ | 1.5850 | 1.5850 | 1.5850 | 1.5850 | 1.5850 |
| $H(D_1)$ | 0.5283 | 0.5283 | 0.5283 | 0.5283 | 0.5283 |
| $H(D_2)$ | 0.5283 | 0.5283 | 0.5283 | 0.5283 | 0.5283 |
| $H(D_3)$ | 0.5283 | 0.5283 | 0.5283 | 0.5283 | 0.5283 |
| $H(D\|C_k)$ | 0.9000 | 0.4444 | 0.3061 | 0.3061 | 0 |
| $H(D_1\|C_k)$ | 0.1383 | 0 | 0 | 0 | 0 |
| $H(D_2\|C_k)$ | 0.5160 | 0.2222 | 0.1761 | 0.1761 | 0 |
| $H(D_3\|C_k)$ | 0.2457 | 0.2222 | 0.1300 | 0.1300 | 0 |
| $H(C_k\|D)$ | 0.3061 | 0.6122 | 0.8344 | 1.1405 | 1.1405 |
| $H(C_k\|D_1)$ | 0 | 0 | 0 | 0.3061 | 0.3061 |
| $H(C_k\|D_2)$ | 0.3061 | 0.3061 | 0.5283 | 0.5283 | 0.5283 |
| $H(C_k\|D_3)$ | 0 | 0.3061 | 0.3061 | 0.3061 | 0.3061 |
| $I(C_k;D), I(D;C_k)$ | 0.6850 | 1.1405 | 1.2789 | 1.2789 | 1.5850 |
| $I(C_k;D_1), I(D_1;C_k)$ | 0.3900 | 0.5283 | 0.5283 | 0.5283 | 0.5283 |
| $I(C_k;D_2), I(D_2;C_k)$ | 0.0123 | 0.3061 | 0.3522 | 0.3522 | 0.5283 |
| $I(C_k;D_3), I(D_3;C_k)$ | 0.2827 | 0.3061 | 0.3983 | 0.3983 | 0.5283 |

By definition calculation, the relevant algebraic and informational measures at the classification and class levels are provided in Table 4.

Table 4 can be utilized to observe three main conclusions of information measures.

(1) For each attribute subset $C_k$ $(k = 1, 2, 3, 4, 5)$, we can achieve:

$$I(D_j;C_k) = I(C_k;D_j), I(C_k;D_j) + H(D_j|C_k) = H(D_j), \quad H_{D_j}(C_k) = I(C_k;D_j) + H(C_k|D_j). \tag{49}$$

Thus, all six types of class-specific information measures indeed satisfy the system function Eq. (33) or Eq. (32).

(2) For each attribute subset $C_k$ $(k = 1, 2, 3, 4, 5)$, we always have:

$$H(C_k) = \sum_{j=1}^{3} H_{D_j}(C_k) \geqslant H_{D_j}(C_k), \ H(D) = \sum_{j=1}^{3} H(D_j) \geqslant H(D_j),$$

$$H(D|C_k) = \sum_{j=1}^{3} H(D_j|C_k) \geqslant H(D_j|C_k), \ H(C_k|D) = \sum_{j=1}^{3} H(C_k|D_j) \geqslant H(C_k|D_j),$$

$$I(C_k;D) = \sum_{j=1}^{3} I(C_k;D_j) \geqslant I(C_k;D_j), \ I(D;C_k) = \sum_{j=1}^{3} I(D_j;C_k) \geqslant I(D_j;C_k).$$

This fact shows the hierarchical relationships of integration/decomposition and size.

(3) For each decision class $D_j$ $(j = 1, 2, 3)$, $H_{D_j}(C_k)$ and $I(C_k;D_j)$ never decrease while $H(D_j|C_k)$ never increases in the attribute chain from $C_1$ to $C_5$. These results verify the granulation monotonicity in Eq. (37).

Next, we present the informational class-specific reducts and their systematic and hierarchical relationships. According to the notions and calculations, relevant results of reducts, cores, and algorithms are provided in Table 5, and these results provide several observations as follows.

(1) Regarding the basic properties of a reduct type, such as the descriptions in Proposition 12 and Algorithm 1, the core is the intersection of all reducts, while the heuristic algorithm provides one reduct.

(2) The horizontal strength-weakness in Fig. 3, such as those in Theorems 3, 4, Corollary 4, and Algorithm 2, can be naturally verified. IE-class-specific reducts are stronger than MI-class-specific/CE-class-specific and algebraic class-specific reducts, and the strict strength is reflected by both cases $D_1$ and $D_2$. MI-class-specific/CE-class-specific reducts are equivalent to algebraic class-specific reducts; this result describes the conclusion of a consistent class in Theorem 4 because the consistency of a decision table implies the consistency of all decision classes. Algorithm MI-CSR→AV-CSR (i.e., Algorithm 2) is effective, and the similar Algorithms IE-CSR→MI-CSR and IE-CSR→AV-CSR are effective and can be powerfully obtained by cases $D_1$ and $D_2$.

**Table 5**
Reduct results of Example 1.

| Level | $RED^{AV}$ | $CORE^{AV}$ | Algorithm AV-CBR/CSR | $RED^{MI}$ | $CORE^{MI}$ | Algorithm MI-CBR/CSR | $RED^{IE}$ | $CORE^{IE}$ | Algorithm IE-CBR/CSR |
|---|---|---|---|---|---|---|---|---|---|
| $D$ | $\{c_2, c_5\}$, $\{c_3, c_5\}$, $\{c_4, c_5\}$. | $\{c_5\}$ | $\{c_2, c_5\}$ | $\{c_2, c_5\}$, $\{c_3, c_5\}$, $\{c_4, c_5\}$. | $\{c_5\}$ | $\{c_2, c_5\}$, | $\{c_1, c_2, c_5\}$, $\{c_1, c_3, c_5\}$, $\{c_1, c_4, c_5\}$. | $\{c_1, c_5\}$ | $\{c_1, c_2, c_5\}$ |
| $D_1$ | $\{c_1, c_2\}$, $\{c_1, c_5\}$, $\{c_2, c_5\}$, $\{c_3\}$, $\{c_4, c_5\}$. | $\varnothing$ | $\{c_3\}$ | $\{c_1, c_2\}$, $\{c_1, c_5\}$, $\{c_2, c_5\}$, $\{c_3\}$, $\{c_4, c_5\}$. | $\varnothing$ | $\{c_3\}$ | $\{c_1, c_2, c_4\}$, $\{c_1, c_5\}$, $\{c_2, c_5\}$, $\{c_3, c_5\}$, $\{c_4, c_5\}$. | $\varnothing$ | $\{c_3, c_5\}$ |
| $D_2$ | $\{c_2, c_5\}$, $\{c_3, c_5\}$, $\{c_4, c_5\}$. | $\{c_5\}$ | $\{c_2, c_5\}$ | $\{c_2, c_5\}$, $\{c_3, c_5\}$, $\{c_4, c_5\}$. | $\{c_5\}$ | $\{c_2, c_5\}$, | $\{c_1, c_2, c_5\}$, $\{c_1, c_3, c_5\}$, $\{c_1, c_4, c_5\}$. | $\{c_1, c_5\}$ | $\{c_1, c_2, c_5\}$ |
| $D_3$ | $\{c_2, c_5\}$, $\{c_3, c_5\}$, $\{c_4, c_5\}$. | $\{c_5\}$ | $\{c_2, c_5\}$ | $\{c_2, c_5\}$, $\{c_3, c_5\}$, $\{c_4, c_5\}$. | $\{c_5\}$ | $\{c_2, c_5\}$, | $\{c_2, c_5\}$, $\{c_3, c_5\}$, $\{c_4, c_5\}$. | $\{c_5\}$ | $\{c_2, c_5\}$ |

(3) The hierarchical strength-weakness in Fig. 3, such as those in Theorem 7 and Algorithm 3, can be directly verified. In particular, Algorithm MI-CBR→MI-CSR (i.e., Algorithm 3) is effective, and the similar Algorithm IE-CBR→IE-CSR is also effective. Via the family function, the inverse relationship and hierarchical balance in Theorems 8, 9 Algorithm 4, as well as the similar Algorithm IE-CSR→IE-CBR, can be verified. □

The decision table used in Example 1 is consistent; thus, all decision classes consistently show the equivalency between MI-class-specific/CE-class-specific and algebraic class-specific reducts in Theorem 4. Furthermore, an inconsistent decision table is provided to verify the difference and strength-weakness of the two types of class-specific reducts.

**Example 2.** An inconsistent decision table is given in Table 6, where object pairs $(x_2, x_4)$ and $(x_3, x_5)$ imply the inconsistency, and the relevant reduct results are offered in Table 7. According to inconsistent class $D_1$, MI-class-specific/CE-class-specific and algebraic class-specific reducts are different, while the former are stronger. Thus, Theorem 4 and Algorithm 2 are true. Regarding $D_1$, Algorithm 2 effectively yields an algebraic class-specific reduct $\{c_2, c_3\}$ when inputting MI-class-specific/CE-class-specific and class-specific reduct $\{c_2, c_3\}$. □

**Table 6**
Inconsistent decision table of Example 2.

| $U$ | $c_1$ | $c_2$ | $c_3$ | $d$ |
|---|---|---|---|---|
| $x_1$ | 1 | 0 | 0 | 1 |
| $x_2$ | 0 | 1 | 1 | 1 |
| $x_3$ | 0 | 0 | 1 | 1 |
| $x_4$ | 0 | 1 | 1 | 2 |
| $x_5$ | 0 | 0 | 1 | 2 |
| $x_6$ | 0 | 2 | 1 | 2 |
| $x_7$ | 0 | 1 | 0 | 2 |

**Table 7**
Reduct results of Example 2.

| Level | $RED^{AV}$ | $CORE^{AV}$ | Algorithm AV-CBR/CSR | $RED^{MI}$ | $CORE^{MI}$ | Algorithm MI-CBR/CSR | $RED^{IE}$ | $CORE^{IE}$ | Algorithm IE-CBR/CSR |
|---|---|---|---|---|---|---|---|---|---|
| $D$ | $\{c_2, c_3\}$ | $\{c_2, c_3\}$ | $\{c_2, c_3\}$ | $\{c_2, c_3\}$ | $\{c_2, c_3\}$ | $\{c_2, c_3\}$ | $\{c_2, c_3\}$ | $\{c_2, c_3\}$ | $\{c_2, c_3\}$ |
| $D_1$ | $\{c_1\}, \{c_2, c_3\}$ | $\varnothing$ | $\{c_1\}$ | $\{c_2, c_3\}$ | $\{c_2, c_3\}$ | $\{c_2, c_3\}$ | $\{c_2, c_3\}$ | $\{c_2, c_3\}$ | $\{c_2, c_3\}$ |
| $D_2$ | $\{c_2, c_3\}$ | $\{c_2, c_3\}$ | $\{c_2, c_3\}$ | $\{c_2, c_3\}$ | $\{c_2, c_3\}$ | $\{c_2, c_3\}$ | $\{c_2, c_3\}$ | $\{c_2, c_3\}$ | $\{c_2, c_3\}$ |

**Table 8**
Basic descriptions of three UCI datasets in terms of decision table $(OB, C \cup D)$.

| Label | Name | Object number $\lvert OB\rvert$ | Condition attribute number $\lvert C\rvert$ | Decision class number $\lvert\pi_D\rvert$ | Consistent |
|---|---|---|---|---|---|
| (1) | Monk-3 | 432 | 6 | 2 | Yes |
| (2) | Tic-Tac-Toe | 958 | 9 | 2 | Yes |
| (3) | CMC | 1473 | 9 | 3 | No |

**Table 9**
Measured values based on the attribute-addition chain of three UCI datasets.

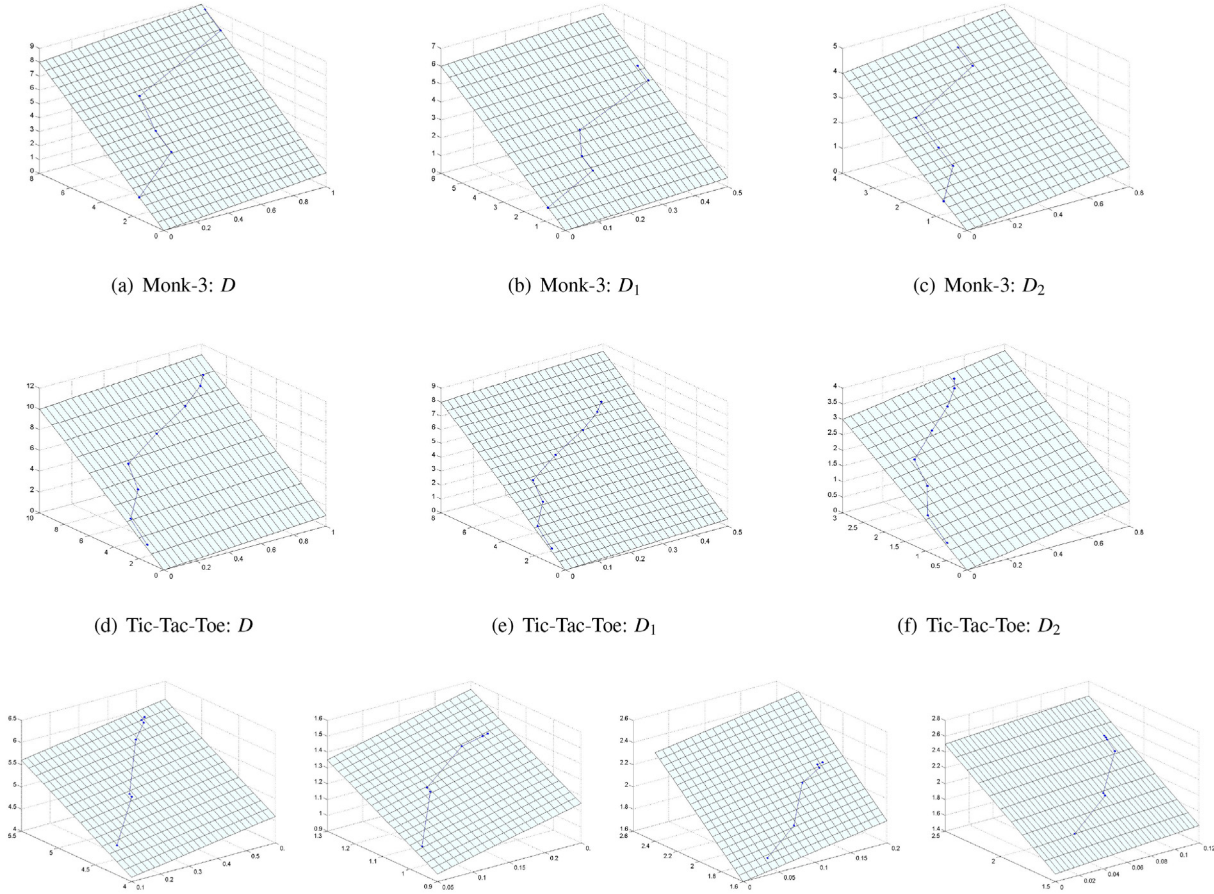
| Measure | $\{c_1\}$ | $\{c_1, c_2\}$ | $\{c_1, c_2, c_3\}$ | $\{c_1, \ldots, c_4\}$ | $\{c_1, \ldots, c_5\}$ | $\{c_1, \ldots, c_6\}$ | $\{c_1, \ldots, c_7\}$ | $\{c_1, \ldots, c_8\}$ | $\{c_1, \ldots, c_9\}$ |
|---|---|---|---|---|---|---|---|---|---|
| (1) $\gamma_{C_k}(D)$ | 0 | 0 | 0 | 0.2222 | 1 | 1 | – | – | – |
| $\gamma_{C_k}(D_1)$ | 0 | 0 | 0 | 0 | 0.5278 | 0.5278 | – | – | – |
| $\gamma_{C_k}(D_2)$ | 0 | 0 | 0 | 0.2222 | 0.4722 | 0.4722 | – | – | – |
| $H(C_k)$ | 1.5820 | 3.1699 | 4.1699 | 5.7549 | 7.7549 | 8.7549 | – | – | – |
| $H_{D_1}(C_k)$ | 0.8365 | 1.6730 | 2.2008 | 3.0373 | 4.0929 | 4.6206 | – | – | – |
| $H_{D_2}(C_k)$ | 0.7485 | 1.4969 | 1.9691 | 2.7176 | 3.6620 | 4.1343 | – | – | – |
| $H(D)$ | 0.9918 | 0.9918 | 0.9918 | 0.9918 | 0.9918 | 0.9918 | – | – | – |
| $H(D_1)$ | 0.4866 | 0.4866 | 0.4866 | 0.4866 | 0.4866 | 0.4866 | – | – | – |
| $H(D_2)$ | 0.5112 | 0.5112 | 0.5112 | 0.5112 | 0.5112 | 0.5112 | – | – | – |
| $H(D\|C_k)$ | 0.9978 | 0.6788 | 0.6788 | 0.6310 | 0 | 0 | – | – | – |
| $H(D_1\|C_k)$ | 0.4866 | 0.3071 | 0.3071 | 0.2631 | 0 | 0 | – | – | – |
| $H(D_2\|C_k)$ | 0.5112 | 0.3717 | 0.3717 | 0.3679 | 0 | 0 | – | – | – |
| $H(C_k\|D)$ | 1.5850 | 2.8590 | 3.8509 | 5.3881 | 6.7571 | 7.7571 | – | – | – |
| $H(C_k\|D_1)$ | 0.8365 | 1.4935 | 2.0213 | 2.8138 | 3.6062 | 4.1340 | – | – | – |
| $H(C_k\|D_2)$ | 0.7485 | 1.3574 | 1.8297 | 2.5743 | 3.1509 | 3.6231 | – | – | – |
| $I(C_k; D), I(D; C_k)$ | 0.0000 | 0.3190 | 0.3190 | 0.3668 | 0.9978 | 0.9978 | – | – | – |
| $I(C_k; D_1), I(D_1; C_k)$ | 0.0000 | 0.1795 | 0.1795 | 0.2235 | 0.4866 | 0.4866 | – | – | – |
| $I(C_k; D_2), I(D_2; C_k)$ | 0.0000 | 0.1395 | 0.1395 | 0.1432 | 0.5112 | 0.5112 | – | – | – |
| (2) $\gamma_{C_k}(D)$ | 0 | 0 | 0.1253 | 0.1628 | 0.4188 | 0.7766 | 0.9436 | 1 | 1 |
| $\gamma_{C_k}(D_1)$ | 0 | 0 | 0.0877 | 0.1221 | 0.3173 | 0.5393 | 0.6253 | 0.6534 | 0.6534 |
| $\gamma_{C_k}(D_2)$ | 0 | 0 | 0.0376 | 0.0407 | 0.1013 | 0.2463 | 0.3184 | 0.3466 | 0.3466 |
| $H(C_k)$ | 1.5281 | 3.0912 | 4.5760 | 6.0351 | 7.3153 | 8.4617 | 9.3530 | 9.9039 | 9.9039 |
| $H_{D_1}(C_k)$ | 0.9972 | 2.0242 | 2.9824 | 3.9399 | 4.7527 | 5.5015 | 6.0967 | 6.4716 | 6.4716 |
| $H_{D_2}(C_k)$ | 0.5309 | 1.0669 | 1.5936 | 2.0951 | 2.5616 | 2.9602 | 3.2563 | 3.4322 | 3.4322 |
| $H(D)$ | 0.9310 | 0.9310 | 0.9310 | 0.9310 | 0.9310 | 0.9310 | 0.9310 | 0.9310 | 0.9310 |
| $H(D_1)$ | 0.4011 | 0.4011 | 0.4011 | 0.4011 | 0.4011 | 0.4011 | 0.4011 | 0.4011 | 0.4011 |
| $H(D_2)$ | 0.5298 | 0.5298 | 0.5298 | 0.5298 | 0.5298 | 0.5298 | 0.5298 | 0.5298 | 0.5298 |
| $H(D\|C_k)$ | 0.9174 | 0.9005 | 0.7532 | 0.7057 | 0.4508 | 0.2059 | 0.0564 | 0 | 0 |
| $H(D_1\|C_k)$ | 0.3963 | 0.3904 | 0.3317 | 0.3161 | 0.2130 | 0.0987 | 0.0282 | 0 | 0 |
| $H(D_2\|C_k)$ | 0.5211 | 0.5101 | 0.4214 | 0.3896 | 0.2378 | 0.1072 | 0.0282 | 0 | 0 |
| $H(C_k\|D)$ | 1.5146 | 3.0606 | 4.3982 | 5.8098 | 6.8351 | 7.7366 | 8.4785 | 8.9729 | 8.9729 |
| $H(C_k\|D_1)$ | 0.9924 | 2.0135 | 2.9130 | 3.8549 | 4.5655 | 5.1991 | 5.7338 | 6.0705 | 6.0705 |
| $H(C_k\|D_2)$ | 0.5222 | 1.0471 | 1.4852 | 1.9549 | 2.2695 | 2.5375 | 2.7547 | 2.9024 | 2.9024 |
| $I(C_k; D), I(D; C_k)$ | 0.0136 | 0.0305 | 0.1778 | 0.2252 | 0.4802 | 0.7251 | 0.8746 | 0.9310 | 0.9310 |
| $I(C_k; D_1), I(D_1; C_k)$ | 0.0049 | 0.0107 | 0.0694 | 0.0850 | 0.1882 | 0.3024 | 0.3729 | 0.4011 | 0.4011 |
| $I(C_k; D_2), I(D_2; C_k)$ | 0.0087 | 0.0198 | 0.1084 | 0.1402 | 0.2920 | 0.4226 | 0.5016 | 0.5298 | 0.5298 |
| (3) $\gamma_{C_k}(D)$ | 0.0576 | 0.3453 | 0.5971 | 0.9353 | 0.9424 | 0.9712 | 0.9856 | 0.9856 | 0.9856 |
| $\gamma_{C_k}(D_1)$ | 0.0360 | 0.2086 | 0.2806 | 0.4173 | 0.4173 | 0.4317 | 0.4388 | 0.4388 | 0.4388 |
| $\gamma_{C_k}(D_2)$ | 0.0072 | 0.0288 | 0.1007 | 0.1871 | 0.1942 | 0.1942 | 0.2014 | 0.2014 | 0.2014 |
| $\gamma_{C_k}(D_3)$ | 0.0144 | 0.1079 | 0.2158 | 0.3309 | 0.3309 | 0.3453 | 0.3453 | 0.3453 | 0.3453 |
| $H(C_k)$ | 4.4861 | 5.0265 | 5.0612 | 5.6763 | 5.8585 | 5.8931 | 5.9217 | 5.9217 | 5.9217 |
| $H_{D_1}(C_k)$ | 1.0441 | 1.2215 | 1.2362 | 1.3643 | 1.3929 | 1.3929 | 1.4000 | 1.4000 | 1.4000 |
| $H_{D_2}(C_k)$ | 1.7081 | 1.8623 | 1.8623 | 2.0915 | 2.1582 | 2.1782 | 2.1853 | 2.1853 | 2.1853 |
| $H_{D_3}(C_k)$ | 1.7339 | 1.9428 | 1.9627 | 2.2204 | 2.3074 | 2.3221 | 2.3364 | 2.3364 | 2.3364 |
| $H(D)$ | 1.5517 | 1.5517 | 1.5517 | 1.5517 | 1.5517 | 1.5517 | 1.5517 | 1.5517 | 1.5517 |
| $H(D_1)$ | 0.4914 | 0.4914 | 0.4914 | 0.4914 | 0.4914 | 0.4914 | 0.4914 | 0.4914 | 0.4914 |
| $H(D_2)$ | 0.5307 | 0.5307 | 0.5307 | 0.5307 | 0.5307 | 0.5307 | 0.5307 | 0.5307 | 0.5307 |
| $H(D_3)$ | 0.5295 | 0.5295 | 0.5295 | 0.5295 | 0.5295 | 0.5295 | 0.5295 | 0.5295 | 0.5295 |
| $H(D\|C_k)$ | 1.4133 | 1.2624 | 1.2617 | 1.1191 | 1.0600 | 1.0593 | 1.0450 | 1.0450 | 1.0450 |
| $H(D_1\|C_k)$ | 0.4310 | 0.3733 | 0.3729 | 0.3019 | 0.2733 | 0.2733 | 0.2662 | 0.2662 | 0.2662 |
| $H(D_2\|C_k)$ | 0.4872 | 0.4348 | 0.4348 | 0.3940 | 0.3647 | 0.3644 | 0.3573 | 0.3573 | 0.3572 |
| $H(D_3\|C_k)$ | 0.4951 | 0.4543 | 0.4540 | 0.4232 | 0.4220 | 0.4216 | 0.4216 | 0.4216 | 0.4216 |
| $H(C_k\|D)$ | 4.3477 | 4.7372 | 4.7712 | 5.2436 | 5.3668 | 5.4008 | 5.4151 | 5.4151 | 5.4151 |
| $H(C_k\|D_1)$ | 0.9836 | 1.1033 | 1.1176 | 1.1748 | 1.1748 | 1.1748 | 1.1748 | 1.1748 | 1.1748 |
| $H(C_k\|D_2)$ | 1.6646 | 1.7665 | 1.7665 | 1.9548 | 1.9922 | 2.0119 | 2.0119 | 2.0119 | 2.0119 |
| $H(C_k\|D_3)$ | 1.6995 | 1.8675 | 1.8871 | 2.1141 | 2.1998 | 2.2141 | 2.2284 | 2.2284 | 2.2284 |
| $I(C_k; D), I(D; C_k)$ | 0.1384 | 0.2893 | 0.2900 | 0.4326 | 0.4917 | 0.4924 | 0.5067 | 0.5067 | 0.5067 |
| $I(C_k; D_1), I(D_1; C_k)$ | 0.0605 | 0.1181 | 0.1186 | 0.1895 | 0.2181 | 0.2181 | 0.2253 | 0.2253 | 0.2253 |
| $I(C_k; D_2), I(D_2; C_k)$ | 0.0435 | 0.0959 | 0.0959 | 0.1367 | 0.1660 | 0.1663 | 0.1734 | 0.1734 | 0.1734 |
| $I(C_k; D_3), I(D_3; C_k)$ | 0.0344 | 0.0753 | 0.0756 | 0.1063 | 0.1076 | 0.1080 | 0.1080 | 0.1080 | 0.1080 |

### 6.2. Experiment verification

The previous theoretical results of informational measures and reducts have been shown in detail and verified by two decision tables: the consistent Table 3 and inconsistent Table 6. For more sufficient manifestations, data experiments are finally supplemented to perform similar and further verification.

**Table 10**
Reduct results of three UCI datasets.

| Level | $RED^{AV}$ | $CORE^{AV}$ | A lgorithm AV-CBR/ CSR | $RED^{MI}$ | $CORE^{MI}$ | A lgorithm MI-CBR/ CSR | $RED^{IE}$ | $CORE^{IE}$ | A lgorithm IE-CBR/ CSR |
|---|---|---|---|---|---|---|---|---|---|
| (1) $D$, $D_1, D_2$ | $\{c_2, c_4, c_5\}$ | $\{c_2, c_4, c_5\}$ | $\{c_2, c_4, c_5\}$ | $\{c_2, c_4, c_5\}$ | $\{c_2, c_4, c_5\}$ | $\{c_2, c_4, c_5\}$ | $\{c_1, c_2, \ldots, c_6\}$ | $\{c_1, c_2, \ldots, c_6\}$ | $\{c_1, c_2, \ldots, c_6\}$ |
| (2) $D$, $D_1, D_2$ | $C - \{c_1\}$<br>$C - \{c_2\}$<br>$\ldots$<br>$C - \{c_9\}$ | $\varnothing$ | $C - \{c_9\}$ | $C - \{c_1\}$<br>$C - \{c_2\}$<br>$\ldots$<br>$C - \{c_9\}$ | $\varnothing$ | $C - \{c_9\}$ | $C - \{c_1\}$<br>$C - \{c_2\}$<br>$\ldots$<br>$C - \{c_9\}$ | $\varnothing$ | $C - \{c_9\}$ |
| (3) $D$ | $\{c_1, c_2, c_4, c_6, c_8\}$,<br>$\{c_1, c_3, c_4, c_6, c_7\}$,<br>$\{c_1, c_3, c_4, c_6, c_8\}$. | $\{c_1, c_4, c_6\}$ | $\{c_1, c_3, c_4, c_6, c_8\}$ | $\{c_1, c_2, c_3, c_4, c_6, c_8\}$ | $\{c_1, c_2, c_3, c_4, c_6, c_8\}$ | $\{c_1, c_2, c_3, c_4, c_6, c_8\}$ | $\{c_1, c_2, c_3, c_4, c_6, c_8\}$ | $\{c_1, c_2, c_3, c_4, c_6, c_8\}$ | $\{c_1, c_2, c_3, c_4, c_6, c_8\}$ |
| (3) $D_1$ | $\{c_1, c_2, c_4, c_6, c_7\}$,<br>$\{c_1, c_3, c_4, c_6, c_7\}$,<br>$\{c_1, c_3, c_4, c_6, c_8\}$. | $\{c_1, c_4, c_6\}$ | $\{c_1, c_3, c_4, c_6, c_8\}$ | $\{c_1, c_2, c_4, c_6, c_7\}$,<br>$\{c_1, c_3, c_4, c_6, c_7\}$,<br>$\{c_1, c_3, c_4, c_6, c_8\}$. | $\{c_1, c_4, c_6\}$ | $\{c_1, c_3, c_4, c_6, c_8\}$ | $\{c_1, c_2, c_3, c_4, c_6, c_8\}$,<br>$\{c_1, c_2, c_4, c_6, c_7\}$. | $\{c_1, c_2, c_4, c_6\}$ | $\{c_1, c_2, c_4, c_6, c_7\}$ |
| (3) $D_2$ | $\{c_1, c_2, c_3, c_4, c_8\}$,<br>$\{c_1, c_2, c_4, c_7, c_8\}$,<br>$\{c_1, c_3, c_4, c_6, c_8\}$,<br>$\{c_1, c_4, c_6, c_7, c_8\}$. | $\{c_1, c_4\}$ | $\{c_1, c_3, c_4, c_8\}$ | $\{c_1, c_2, c_3, c_4, c_8\}$,<br>$\{c_1, c_2, c_4, c_7, c_8\}$,<br>$\{c_1, c_3, c_4, c_6, c_8\}$,<br>$\{c_1, c_4, c_6, c_7, c_8\}$. | $\{c_1, c_4\}$ | $\{c_1, c_2, c_3, c_4, c_8\}$ | $\{c_1, c_2, c_3, c_4, c_5, c_8\}$,<br>$\{c_1, c_2, c_4, c_5, c_7\}$,<br>$\{c_1, c_3, c_4, c_6, c_8\}$,<br>$\{c_1, c_4, c_6, c_7, c_8\}$. | $\{c_1, c_4\}$ | $\{c_1, c_3, c_4, c_6, c_7\}$ |
| (3) $D_3$ | $\{c_1, c_2, c_4, c_6, c_7, c_8\}$,<br>$\{c_1, c_3, c_4, c_6\}$. | $\{c_1, c_4, c_6\}$ | $\{c_1, c_3, c_4, c_6\}$ | $\{c_1, c_2, c_4, c_6, c_7, c_8\}$,<br>$\{c_1, c_3, c_4, c_6\}$. | $\{c_1, c_4, c_6\}$ | $\{c_1, c_3, c_4, c_6\}$ | $\{c_1, c_3, c_4, c_6, c_8\}$ | $\{c_1, c_3, c_4, c_6, c_8\}$ | $\{c_1, c_3, c_4, c_6, c_8\}$ |



(a) Monk-3: $D$ (b) Monk-3: $D_1$ (c) Monk-3: $D_2$

(d) Tic-Tac-Toe: $D$ (e) Tic-Tac-Toe: $D_1$ (f) Tic-Tac-Toe: $D_2$

(g) CMC: $D$ (h) CMC: $D_1$. (i) CMC: $D_2$. (j) CMC: $D_3$.

**Fig. 4.** Three-dimensional plane of three-way information measures of three UCI datasets.

According to the three datasets described in Table 8, Tables 9 and 10, respectively exhibit the measure values and reduct results. Both tables can be utilized to analyze the theoretical properties. We similarly verify both the three points of measures and the three aspects of reducts concerned in Example 1, and the related details are omitted. Fig. 4 shows the three-dimensional points $(I(A;D), H(A|D), H(A))$ and $\left(I(A;D_j), H(A|D_j), H_{D_j}(A)\right)$ and their existent planes, thus presenting the three-way summations

$$H(A) = I(D;A) + H(A|D), \;\; H_{D_j}(A) = I(A;D_j) + H(A|D_j) \tag{50}$$

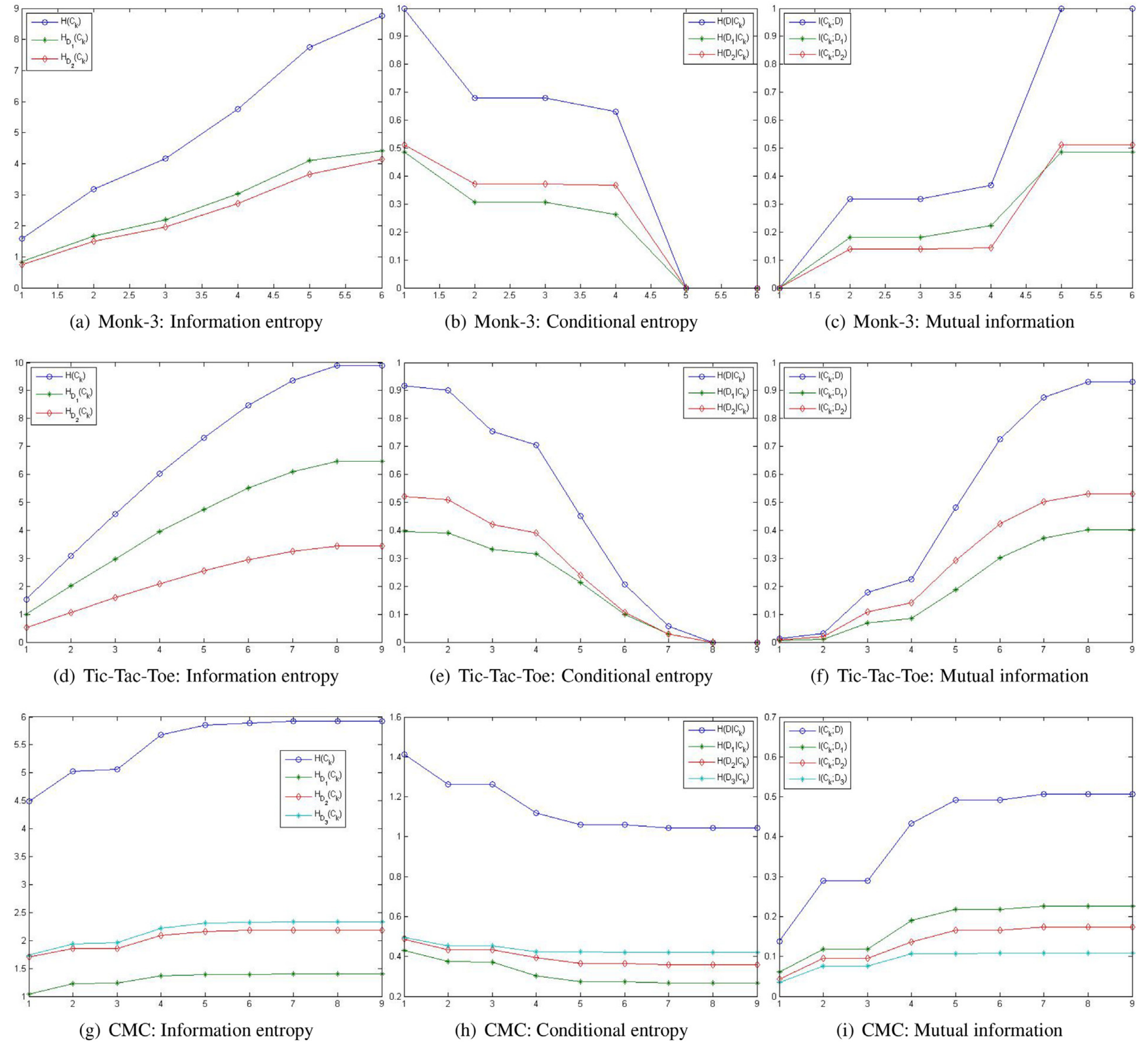
in Eqs. (11) and (33), respectively; moreover, Fig. 5 describes the three-way main information measures based on the attribute-addition chain, thus verifying the granulation monotonicity of two metric groups:

$$H(A), H(D|A), I(A;D); \;\; H(A)_{D_j}, H(D_j|A), I(A;D_j) \tag{51}$$

in Proposition 3 and Corollary 1, respectively.

## 7. Conclusions

Class-specific attribute reducts are useful for pattern recognition and have become a basic type of attribute reduction. Targeting their scarce information discussions and required system connections, this paper studies the informational class-specific reducts and their connections with the algebraic class-specific and classification-based reducts, while the decomposition mining of information measures from the classification level to the class level becomes a basis and a key. The class-specific information entropy, conditional entropy, and mutual information come from three strategies of hierarchical decomposition. Opposing the classification-based information measures, these measures obtain similar features, including the systematical relationship, uncertainty semantics, and granulation monotonicity, and they embody the hierarchical



(a) Monk-3: Information entropy (b) Monk-3: Conditional entropy (c) Monk-3: Mutual information

(d) Tic-Tac-Toe: Information entropy (e) Tic-Tac-Toe: Conditional entropy (f) Tic-Tac-Toe: Mutual information

(g) CMC: Information entropy (h) CMC: Conditional entropy (i) CMC: Mutual information

**Fig. 5.** Information measure monotonicity based on the attribute-addition chain of three UCI datasets.

isomorphism. Accordingly, we propose three types of informational class-specific reducts (IE-class-specific, CE-class-specific, and MI-class-specific reducts), and their systematic connections can be uncovered by the hierarchical and transverse isomorphisms.

(1) Transverse connections are mainly embodied by the strong–weak relation. IE-class-specific reducts are stronger than equivalent CE-class-specific and MI-class-specific reducts; furthermore, the latter informational class-specific reducts are stronger than AV-class-specific reducts, and this strong-weak connection degenerates for a consistent class but holds for an inconsistent class. Therefore, the informational class-specific reducts are isomorphic to the informational classification-based reducts because the latter have similar descriptions. This hierarchical isomorphism describes two types of transverse connections, and the same strong-weak relations among the informational class-specific and classification-based reducts are presented in the horizontal direction of Fig. 3.

(2) Alternately, hierarchical connections are mainly embodied by the strong-weak relation and family-based balance. The informational class-specific reducts are respectively weaker than the informational classification-based reducts, while the former resort to their family of all decision classes to balance the latter. The hierarchical strength-weakness and balance become true from both the informational and algebraic perspectives, and the hierarchical strength-weakness of both perspectives are described in the vertical direction of Fig. 3. The two-level reduct systems between the informational and algebraic viewpoints provide a transverse isomorphism regarding the hierarchical strength-weakness and balance.

Thus, we novelly establish three types of variant isomorphisms: the measure-hierarchical, reduct-level-longitudinal, and reduct-viewpoint-transverse isomorphisms. The former focuses on the information measures, while the latter two are symmetrical to embrace the vertical-horizontal reduction system, and thus all become interesting and deep. As a result, the class-specific information measures hierarchically deepen the existing classification-based information measures, while the informational class-specific reducts systematically perfect the attribute reduction framework with two-level and two-viewpoint modes. The obtained results facilitate uncertainty measurement and information processing, especially at the class level. The class-specific information measures and attribute reducts are worthy of practical application for pattern recognition in a future study, especially by combing the hierarchical guidance from Macro-Top; moreover, the measures and reducts at Micro-Bottom require more detailed research to describe three-level measures and reducts.

## CRediT author statement

**Xianyong Zhang:** Conceptualization, Methodology, Formal analysis. **Hong Yao:** Software, Visualization. **Zhiying Lv:** Validation, Investigation. **Duoqian Miao:** Conceptualization, Supervision.

## 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.


## Acknowledgments

The authors would like to thank both the editors and reviewers for their valuable suggestions, especially anonymous Reviewer 2 and his or her review, which all substantially improved this paper.

This work was supported by National Natural Science Foundation of China (61673285, 61673301, 61976158), Sichuan Science and Technology Program of China (2021YJ0085, 2019YJ0529), and A Joint Research Project of Laurent Mathematics Center of Sichuan Normal University and National-Local Joint Engineering Laboratory of System Credibility Automatic Verification.


## Appendix A. Proof of Lemma 2

**Proof.** According to Eq. (18),

$$H(D_j) = -p(D_j)\log_2 p(D_j) = -\left(\sum_{i=1}^{n(A)} p(A_i \cap D_j)\right)\log_2 p(D_j) = -\sum_{i=1}^{n(A)} p(A_i \cap D_j)\log_2 p(D_j) = -\sum_{i=1}^{n(A)} p(A_i)p(D_j|A_i)\log_2 p(D_j). \quad \text{(A.1)}$$

Note that the inverse deduction of above equation becomes clearer. On this basis of Eq. (A.1), we further have

$$\begin{aligned} I(A;D_j) = H(D_j) - H(D_j|A) &= -\sum_{i=1}^{n(A)} [p(A_i)p(D_j|A_i)\log_2 p(D_j) - p(A_i)p(D_j|A_i)\log_2 p(D_j|A_i)] \\ &= -\sum_{i=1}^{n(A)} p(D_j \cap A_i)\log_2 \frac{|D_j| \times |A_i|}{|U| \times |D_j \cap A_i|}, \end{aligned} \quad \text{(A.2)}$$

$$\begin{aligned} I(A;D_j) + H(A|D_j) &= -\sum_{i=1}^{n(A)} p(D_j \cap A_i)\log_2 \frac{|D_j| \times |A_i|}{|U| \times |D_j \cap A_i|} - \sum_{i=1}^{n(A)} p(D_j)p(A_i|D_j)\log_2 p(A_i|D_j) \\ &= -\sum_{i=1}^{n(A)} p(D_j \cap A_i)\log_2 \frac{|D_j| \times |A_i|}{|U| \times |D_j \cap A_i|} - \sum_{i=1}^{n(A)} p(D_j \cap A_i)\log_2 \frac{|D_j \cap A_i|}{|D_j|} = -\sum_{i=1}^{n(A)} p(D_j \cap A_i)\log_2 \frac{|A_i|}{|U|} \\ &= -\sum_{i=1}^{n(A)} p(D_j|A_i)p(A_i)\log_2 p(A_i). \end{aligned} \quad \text{(A.3)}$$

## Appendix B. Proof of Lemma 5

**Proof.** We emphatically consider the equality conversions on granular merging $\bigcup_{t=1}^{k} [x]_C^t = [x]_R$ $(k \geqslant 2)$, and this merging formula represents each group in $\pi_C \xrightarrow{\preceq} \pi_R$. Only $I(R;D_j) = I(C;D_j)$ and $POS(D_j|R) = POS(D_j|C)$ are considered for the unidirectional derivation because they respectively correspond to $H(D_j|R) = H(D_j|C)$ and $\gamma_R(D_j) = \gamma_C(D_j)$.

According to Eq. (38) in Corollary 2, the IPE-Condition of $I(R;D_j) = I(C;D_j)$ must be analyzed from the three-way regions (the positive, negative, and boundary regions). There are three and only three cases, and relevant regional derivations are discussed as follows.

(1) If $p(D_j|[x]_C^1) = p(D_j|[x]_C^2) = \cdots = p(D_j|[x]_C^k) = 0$ and $p(D_j|[x]_R) = 0$, then we have $[x]_C^1, [x]_C^2, \ldots, [x]_C^k \subseteq NEG(D_j|C)$ and $[x]_R \subseteq NEG(D_j|R)$. In this case, the granular merging concerns only the negative regions (i.e., it never impacts the positive regions).
(2) If $p(D_j|[x]_C^1) = p(D_j|[x]_C^2) = \cdots = p(D_j|[x]_C^k) = 1$ and $p(D_j|[x]_R) = 1$, then we have $[x]_C^1, [x]_C^2, \cdots, [x]_C^k \subseteq POS(D_j|C)$ and $[x]_R \subseteq POS(D_j|R)$. In this case, the granular merging concerns only the positive regions but never changes the positive regions.
(3) Otherwise, $p(D_j|[x]_C^1) = p(D_j|[x]_C^2) = \cdots = p(D_j|[x]_C^k) \in (0,1)$ and $p(D_j|[x]_R) = (0,1)$, and then we have $[x]_C^1, [x]_C^2, \ldots, [x]_C^k \subseteq BND(D_j|C)$ and $[x]_R \subseteq BND(D_j|R)$. In this case, the granular merging concerns only the boundary regions.

The three cases never cause the change in positive regions; thus, the positive regions are never impacted by granular merging $\bigcup_{t=1}^{k} [x]_C^t = [x]_R$. Furthermore, the positive regions are preserved in $\pi_C \xrightarrow{\preceq} \pi_R$, i.e., $POS(D_j|C) = POS(D_j|R)$. Therefore, $I(R;D_j) = I(C;D_j) \Rightarrow POS(D_j|R) = POS(D_j|C)$.

Conversely, it is assumed that $POS(D_j|R) = POS(D_j|C)$. By observation, $\bigcup_{t=1}^{k} [x]_C^t = [x]_R$ has multiple distributions on three-way regions; however, there are two and only two special cases of granular merging which destroy the IPE-Condition but not the positive regions.

(1) If $p(D_j|[x]_C^1), p(D_j|[x]_C^2), \cdots, p(D_j|[x]_C^k) \in (0,1)$ and $p(D_j|[x]_R) \in (0,1)$, and if $p(D_j|[x]_C^1) = p(D_j|[x]_C^2) = \cdots = p(D_j|[x]_C^k)$ does not hold, then we have $[x]_C^1, [x]_C^2, \cdots, [x]_C^k \subseteq BND(D_j|C)$ and $[x]_R \subseteq BND(D_j|R)$. In this case, the granular merging concerns only the boundary regions but never satisfies the IPE-Condition in Eq. (38).
(2) If $k$ probabilities $p(D_j|[x]_C^1), p(D_j|[x]_C^2), \cdots, p(D_j|[x]_C^k)$ exactly/completely reach two types of interval $(0,1)$ and equal 0, then granular merging destroys the boundary/negative regions but not the positive regions; this process also affects the IPE-Condition.

Based on each case, granular merging $\bigcup_{t=1}^{k} [x]_C^t = [x]_R$ impacts the IPE-Condition but not the positive regions. Furthermore, knowledge coarsening $\pi_C \xrightarrow{\preceq} \pi_R$ can have the unchanged positive regions and the changed information values (i.e., $POS(D_j|C) = POS(D_j|R)$ and $I(R;D_j) \neq I(C;D_j)$). We can thus achieve $POS(D_j|R) = POS(D_j|C) \nRightarrow I(R;D_j) = I(C;D_j)$. □

## References


[1] M. Abdolrazzagh-Nezhad, Enhanced cultural algorithm to solve multi-objective attribute reduction based on rough set theory, Math. Comput. Simul. 170 (2020) 332–350.
[2] M. Aggarwal, Probabilistic variable precision fuzzy rough sets, IEEE Trans. Fuzzy Syst. 24 (1) (2016) 29–39.
[3] M.J. Benitez-Caballero, J. Medina, E. Ramirez-Poussa, D. Slezak, Rough-set-driven approach for attribute reduction in fuzzy formal concept analysis, Fuzzy Sets Syst. 391 (2020) 117–138.
[4] J. Blaszczynski, S. Greco, R. Slowinski, M. Szelg, Monotonic variable consistency rough set approaches, Int. J. Approx. Reason. 50 (7) (2009) 979–999.
[5] D. Calvanese, M. Dumas, U. Laurson, F.M. Maggi, M. Montali, I. Teinemaa, Semantics, analysis and simplification of DMN decision tables, Inf. Syst. 78 (2018) 112–125.
[6] A. Campagner, D. Ciucci, Orthopartitions and soft clustering: soft mutual information measures for clustering validation, Knowl.-Based Syst. 180 (2019) 51–61.
[7] Y.M. Chen, Z.J. Zhang, J.Z. Zheng, Y. Ma, Y. Xue, Gene selection for tumor classification using neighborhood rough sets and entropy measures, J. Biomed. Inform. 67 (2017) 59–68.
[8] G. Chiaselotti, T. Gentile, F. Infusino, Granular computing on information tables: families of subsets and operators, Inf. Sci. 442–443 (2018) 72–102.
[9] G. Chiaselotti, T. Gentile, F. Infusino, Decision systems in rough set theory: a set operatorial perspective, J. Algebra Appl. 18 (01) (2019) 1950004.
[10] G. Chiaselotti, T. Gentile, F. Infusino, New perspectives of granular computing in relation geometry induced by pairings, Annali dell'Universitá di Ferrara 65 (1) (2019) 57–94.
[11] J.H. Dai, B.J. Wei, X.H. Zhang, Q.L. Zhang, Uncertainty measurement for incomplete interval-valued information systems based on α)weak similarity, Knowl.-Based Syst. 136 (2017) 159–171.
[12] L. D'eer, C. Cornelis, Decision reducts and bireducts in a covering approximation space and their relationship to set definability, Int. J. Approx. Reason. 109 (2019) 42–54..
[13] C. Gao, Z.H. Lai, J. Zhou, J.J. Wen, W.K. Wong, Granular maximum decision entropy-based monotonic uncertainty measure for attribute reduction, Int. J. Approx. Reason. 104 (2019) 9–24.
[14] C. Gao, Z.H. Lai, J. Zhou, C.R. Zhao, D.Q. Miao, Maximum decision entropy-based attribute reduction in decision-theoretic rough set model, Knowl.-Based Syst. 143 (2018) 179–191.
[15] Y.T. Guo, E.C.C. Tsang, W.H. Xu, D.G. Chen, Local logical disjunction double-quantitative rough sets, Inf. Sci. 500 (2019) 87–112.
[16] A.P. Huang, H. Zhao, W. Zhu, Nullity-based matroid of rough sets and its application to attribute reduction, Inf. Sci. 263 (2014) 153–165.
[17] B. Huang, H.X. Li, G.F. Feng, X.Z. Zhou, Dominance-based rough sets in multi-scale intuitionistic fuzzy decision tables, Appl. Math. Comput. 348 (2019) 487–512.
[18] X.Y. Jia, L. Shang, B. Zhou, Y.Y. Yao, Generalized attribute reduct in rough set theory, Knowl.-Based Syst. 91 (2016) 204–218.

[19] F. Jiang, Y.F. Sui, L. Zhou, A relative decision entropy-based feature selection approach, Pattern Recogn. 48 (7) (2015) 2151–2163.
[20] J. Konecny, P. Krajca, On attribute reduction in concept lattices: the polynomial time discernibility matrix-based method becomes the CR-method, Inf. Sci. 491 (2019) 48–62.
[21] Q.Z. Kong, X.W. Zhang, W.H. Xu, S.T. Xie, Attribute reducts of multi-granulation information system, Artif. Intell. Rev. 53 (2) (2020) 1353–1371.
[22] G.M. Lang, M.J. Cai, H. Fujita, Q.M. Xiao, Related families-based attribute reduction of dynamic covering decision information systems, Knowl.-Based Syst. 162 (2018) 161–173.
[23] M.S. Lazo-Cortes, J.F. Martinez-Trinidad, J.A. Carrasco-Ochoa, Class-specific reducts vs. classic reducts in a rule-based classifier: a case study, Lect. Notes Comput. Sci. 10880 (2018) 23–30.
[24] S.J. Liao, X.Y. Zhang, Z.W. Mo, Three-level and three-way uncertainty measurements for interval-valued decision systems, Int. J. Mach. Learn. Cybern. (2020), https://doi.org/10.1007/s13042-020-01247-8.
[25] G.L. Liu, Y.B. Feng, J.T. Yang, A common attribute reduction form for information systems, Knowl.-Based Syst. 193 (2020) 105466.
[26] B.H. Long, W.H. Xu, X.Y. Zhang, L. Yang, The dynamic update method of attribute-induced three-way granular concept in formal contexts, Int. J. Approx. Reason. 126 (2020) 228–248.
[27] X.A. Ma, G.Y. Wang, H. Yu, T.R. Li, Decision region distribution preservation reduction in decision-theoretic rough set model, Inf. Sci. 278 (2014) 614–640.
[28] X.A. Ma, Y.Y. Yao, Min-max attribute-object bireducts: on unifying models of reducts in rough set theory, Inf. Sci. 501 (2019) 68–83.
[29] X.A. Ma, Y.Y. Yao, Three-way decision perspectives on class-specific attribute reducts, Inf. Sci. 450 (2018) 227–245.
[30] X.A. Ma, X.R. Zhao, Cost-sensitive three-way class-specific attribute reduction, Int. J. Approx. Reason. 105 (2019) 153–174.
[31] X.A. Ma, Fuzzy entropies for class-specific and classification-based attribute reducts in three-way probabilistic rough set models, Int. J. Mach. Learn. Cybern. 12 (2021) 433–457.
[32] D.Q. Miao, Rough set theory and its application in machine learning (Ph. D. Thesis), Institute of Automation, The Chinese Academy of Sciences, Beijing, 1997 (in Chinese).
[33] D.Q. Miao, Y. Zhao, Y.Y. Yao, H.X. Li, F.F. Xu, Relative reducts in consistent and inconsistent decision tables of the Pawlak rough set model, Inf. Sci. 179 (24) (2009) 4140–4150.
[34] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, vol. 9, Kluwer Academic Publishers, Dordrecht, 1991.
[35] W. Pedrycz, Information granules and their use in schemes of knowledge management, Sci. Iran. 18 (3) (2011) 602–610.
[36] L.S. Peng, W.B. Qian, Y.L. Wang, Algorithms of three-way probabilistic attribute reducts for class-specific, J. Chin. Comput. Syst. 40 (9) (2019) 1851–1857 (in Chinese).
[37] Y.H. Qian, J.Y. Liang, W. Pedrycz, C.Y. Dang, Positive approximation: an accelerator for attribute reduction in rough set theory, Artif. Intell. 174 (2010) 597–618.
[38] R.K. Shiraz, H. Fukuyama, M. Tavana, D.D. Caprio, An integrated data envelopment analysis and free disposal hull framework for cost-efficiency measurement using rough sets, Appl. Soft Comput. 46 (2016) 204–219.
[39] A. Skowron, J. Stepaniuk, R. Swiniarski, Modeling rough granular computing based on approximation spaces, Inf. Sci. 184 (1) (2012) 20–43.
[40] D. Slezak, Approximate entropy reducts, Fundam. Inf. 53 (2002) 365–390.
[41] D. Slezak, Rough sets and functional dependencies in data: Foundations of association reducts, in: M.L. Gavrilova, C.J.K. Tan, Y. Wang, K.C.C. Chan (Eds.), Transactions on Computational Science V, Lecture Notes in Computer Science, vol. 5540, 2009, pp. 182–205. .
[42] D. Slezak, On generalized decision functions: reducts, networks and ensembles, in: Y. Yao, Q. Hu, H. Yu, J. Grzymala-Busse (Eds.) Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, Lecture Notes in Computer Science, vol. 9437, 2015, pp. 13–23. .
[43] S. Stawicki, D. Slezak, A. Janusz, S. Widz, Decision bireducts and decision reducts – a comparison, Int. J. Approx. Reason. 84 (2017) 75–109.
[44] B.Z. Sun, W.M. Ma, X.T. Chen, X. Zhang, Multigranulation vague rough set over two universes and its application to group decision making, Soft. Comput. 23 (18) (2019) 8927–8956.
[45] L. Sun, L.Y. Wang, W.P. Ding, Y.H. Qian, J.C. Xu, Neighborhood multi-granulation rough sets-based attribute reduction using Lebesgue and entropy measures in incomplete neighborhood decision systems, Knowl.-Based Syst. 192 (2019) 105373.
[46] J.G. Tang, K. She, F. Min, W. Zhu, A matroidal approach to rough set theory, Theoret. Comput. Sci. 471 (2013) 1–11.
[47] L.Y. Tang, X.Y. Zhang, Z.W. Mo, A weighted complement-entropy system based on tri-level granular structures, Int. J. Gen. Syst. 49 (8) (2020) 872–905.
[48] M. Tang, H.C. Liao, Managing information measures for hesitant fuzzy linguistic term sets and their applications in designing clustering algorithms, Inf. Fusion 50 (2019) 30–42.
[49] N.N. Thuy, S. Wongthanavasu, On reduction of attributes in inconsistent decision tables based on information entropies and stripped quotient sets, Expert Syst. Appl. 137 (2019) 308–323.
[50] S. Vluymans, N.M. Parthalain, C. Cornelis, Y. Saeys, Weight selection strategies for ordered weighted average based fuzzy rough sets, Inf. Sci. 501 (2019) 155–171.
[51] C.Z. Wang, Y. Wang, M.W. Shao, Y.H. Qian, D.G. Chen, Fuzzy rough attribute reduction for categorical data, IEEE Trans. Fuzzy Syst. 28 (5) (2020) 818–830.
[52] C.Z. Wang, Y. Huang, M.W. Shao, X.D. Fan, Fuzzy rough set-based attribute reduction using distance measures, Knowl.-Based Syst. 164 (2019) 205–212.
[53] G.Y. Wang, X.A. Ma, H. Yu, Monotonic uncertainty measures for attribute reduction in probabilistic rough set model, Int. J. Approx. Reason. 59 (2015) 41–67.
[54] G.Y. Wang, J. Zhao, J.J. An, Y. Wu, A comparative study of algebra viewpoint and information viewpoint in attribute reduction, Fundam. Inf. 68 (3) (2005) 289–301.
[55] S.P. Wang, Q.X. Zhu, W. Zhu, F. Min, Graph and matrix approaches to rough sets through matroids, Inf. Sci. 288 (2014) 1–11.
[56] W.H. Xu, W.T. Li, Granular computing approach to two-way learning based on formal concept analysis in fuzzy datasets, IEEE Trans. Cybern. 46 (2) (2016) 366–379.
[57] W.H. Xu, J.H. Yu, A novel approach to information fusion in multi-source datasets: a granular computing viewpoint, Inf. Sci. 378 (2017) 410–423.
[58] J.L. Yang, Y.Y. Yao, Semantics of soft sets and three-way decision with soft sets, Knowl.-Based Syst. 194 (2020) 105538.
[59] L. Yang, W.H. Xu, X.Y. Zhang, B.B. Sang, Multi-granulation method for information fusion in multi-source decision information system, Int. J. Approx. Reason. 122 (2020) 47–65.
[60] J.T. Yao, A.V. Vasilakos, W. Pedrycz, Granular computing: perspectives and challenges, IEEE Trans. Cybern. 43 (6) (2013) 1977–1989.
[61] Y.Y. Yao, Three-way decision and granular computing, Int. J. Approx. Reason. 103 (2018) 107–123.
[62] Y.Y. Yao, Three-way granular computing, rough sets, and formal concept analysis, Int. J. Approx. Reason. 116 (2020) 106–125.
[63] Y.Y. Yao, Tri-level thinking: models of three-way decision, Int. J. Mach. Learn. Cybern. 11 (2020) 947–959.
[64] Y.Y. Yao, X.Y. Zhang, Class-specific attribute reducts in rough set theory, Inf. Sci. 418–419 (2017) 601–618.
[65] Y.Y. Yao, Y. Zhao, Discernibility matrix simplification for constructing attribute reducts, Inf. Sci. 179 (2009) 867–882.
[66] Z. Yuan, X.Y. Zhang, S. Feng, Hybrid data-driven outlier detection based on neighborhood information entropy and its developmental measures, Expert Syst. Appl. 112 (2018) 243–257.
[67] Q.H. Zhang, S.H. Yang, G.Y. Wang, Measuring uncertainty of probabilistic rough set model from its three regions, IEEE Trans. Syst., Man Cybern. Syst. 47 (12) (2017) 3299–3309.
[68] X.Y. Zhang, H.Y. Gou, Z.Y. Lv, D.Q. Miao, Double-quantitative distance measurement and classification learning based on the tri-level granular structure of neighborhood system, Knowl.-Based Syst. 217 (2021) 106799.
[69] X.Y. Zhang, D.Q. Miao, Quantitative/qualitative region-change uncertainty/certainty in attribute reduction: comparative region-change analyses based on granular computing, Inf. Sci. 334–335 (2016) 174–204.

[70] X.Y. Zhang, D.Q. Miao, Reduction target structure-based hierarchical attribute reduction for two-category decision-theoretic rough sets, Inf. Sci. 277 (2014) 755–776.
[71] X.Y. Zhang, D.Q. Miao, Three-layer granular structures and three-way informational measures of a decision table, Inf. Sci. 412–413 (2017) 67–86.
[72] X.Y. Zhang, J.L. Yang, L.Y. Tang, Three-way class-specific attribute reducts from the information viewpoint, Inf. Sci. 507 (2020) 840–872.
[73] X.Y. Zhang, X. Tang, J.L. Yang, Z.Y. Lv, Quantitative three-way class-specific attribute reducts based on region preservations, Int. J. Approx. Reason. 117 (2020) 96–121.