Information Sciences 412–413 (2017) 67–86

Contents lists available at ScienceDirect

Information Sciences

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

Three-layer granular structures and three-way informational measures of a decision table

Xianyong Zhang[a,b,*], Duoqian Miao[c]

[a] College of Mathematics and Software Science, Sichuan Normal University, Chengdu 610066, China
[b] Institute of Intelligent Information and Quantum Information, Sichuan Normal University, Chengdu 610066, China
[c] Department of Computer Science and Technology, Tongji University, Shanghai 201804, China

ARTICLE INFO

Article history:
Received 12 June 2016
Revised 18 April 2017
Accepted 21 May 2017
Available online 22 May 2017

Keywords:
Attribute reduction
Granular computing
Three-way decisions
Three-layer structures
Information theory
Bayes' theorem

ABSTRACT

Attribute reduction in rough set theory serves as a fundamental topic for information processing, and its basis is usually a decision table (D-Table). D-Table attribute reduction concerns three hierarchical types, and only classification-based reduction is related to information-theoretic representation. Aiming at inducing comprehensive D-Table attribute reduction with hierarchies and information, this paper concretely constructs a D-Table's three-layer granular structures and three-way informational measures via granular computing and Bayes' theorem. With regard to the D-Table, the micro-bottom, meso-middle, and macro-top are hierarchically organized according to the formal structure and systematic granularity. Then, different layers produce different three-way informational measures by developing Bayes' theorem. Thus, three-way weighted entropies originate from three-way probabilities at the micro-bottom and further evolve from the meso-middle to the macro-top, and their granulation monotonicity and evolution systematicness are acquired. Furthermore, three-way informational measures are analyzed by three-layer granular structures to achieve their hierarchical evolution, superiority, and algorithms. Finally, structural and informational results are effectively illustrated by a D-Table example. This study establishes D-Table's hierarchical structures to reveal constructional mechanisms and systematic relationships of informational measures. The obtained results underlie the D-Table's hierarchical, systematic, and informational attribute reduction, and they also enrich the three-way decisions theory.

1. Introduction

Rough set theory is fundamental for information processing. Its core reduction (especially attribute reduction) can effectively implement simplification and reasoning, and a decision table (D-Table) usually serves as the basis [4,23,35,36,44] but has no special reports on granular hierarchical structures. D-Table reduction includes "three steps": attribute reduction, category reduction, and rule reduction [25]. Category reduction can change into attribute reduction on a condition class [39], while rule reduction simply eliminates repeatability, and thus, the D-Table's attribute reduction is worthwhile to comprehensively research. Herein, D-Table concerns four basic elements, i.e., the decisional classification, decisional class, conditional

* Corresponding author at: College of Mathematics and Software Science, Sichuan Normal University, Chengdu 610066, China.
E-mail addresses: xianyongzh@sina.com.cn (X. Zhang), miaoduoqian@163.com.cn (D. Miao).

http://dx.doi.org/10.1016/j.ins.2017.05.032
0020-0255/© 2017 Published by Elsevier Inc.

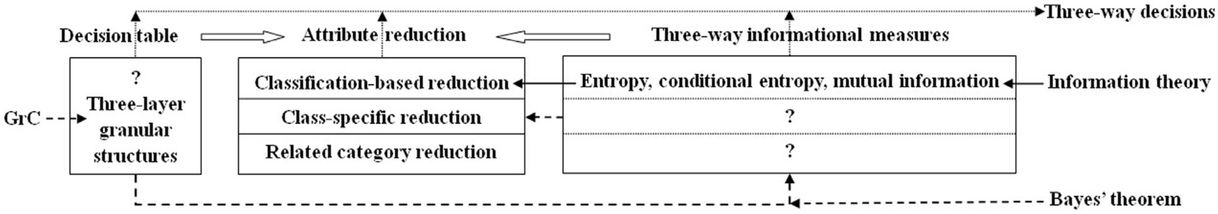

Fig. 1. Background, thought, and contribution of the new research.

classification, and conditional class, which are simply denoted by D-Classification, D-Class, C-Classification, and C-Class, respectively.

D-Table attribute reduction has "three types" of hierarchies. The classical and prevalent type [25] depends on D-Classification and C-Classification and, thus, can be called classification-based reduction, and it exhibits research pathways in the model region, discernibility matrix, decision rule, and information measure [13,20,31,38]. Classification-based reduction implements whole optimization to, on average, fit all of the D-Classes, but its compromise might not necessarily be suitable for specific optimization with regard to every D-Class. Thus, Yao and Zhang [43] utilized D-Class and C-Classification to establish a new attribute reduction called class-specific reduction to solve the relevant blind spots. Moreover, attribute reduction based on D-Class and C-Class can describe category reduction [39] and, thus, becomes the third type. With regard to the D-Table, the three types of attribute reduction are actually located at different levels, and thus, relevant hierarchical structures and informational evolution become significant; herein, granulation monotonicity plays an important role in attribute reduction with regard to C-Classification [13,23,33,48] and therefore becomes the main criterion for evaluating informational measures. Note that three-way decisions serve as a fundamental methodology with extensive applications. Hu [9] discussed three-way decisions based on semi-three-way decision spaces, and Li et al. [16] adopted the multi-granularity to study three-way cognitive concept learning. In particular, Yao [40] pointed out that the three-level analysis falls into the category of three-way decisions, so the three-layer attribute reduction and relevant three-level measure construction become a typical case and a good example of three-way decisions.

Information theory provides an effective approach for uncertainty processing, which also exists in rough set theory, and the relevant information measure has been introduced into rough set theory for uncertainty representation and measurement [2,3,7,10,17,34]. As a result, the attribute reduction depends on measures to be related to the informational description. Miao and Wang [21,22] offered the informational representation of knowledge reduction and decision reduction, where the entropy and mutual information are highlighted; Wang et al. [32] conducted a comparative study on attribute reduction from the algebra and information viewpoints, where the conditional entropy acts as a main tool; Ma et al. [20] constructed a sort of conditional information entropy to obtain the heuristic reduction algorithms; Jiang et al. [14] presented the relative decision entropy to propose a feature selection algorithm; moreover, Slezak [29] used the (conditional) entropy to define approximate reducts. On the whole, the entropy, conditional entropy, and mutual information become basic informational measures to describe attribute reduction. However, they mainly originate from the perspective of information theory so are applied to only the usual classification-based reduction. For these information-theoretic measures, their construction mechanisms and systematic relationships are worthwhile to thoroughly clarify from the viewpoint of rough sets themselves based on the D-Table and attribute reduction.

The above background is described in Fig. 1, and the current research mainly rests on the solid arrow "information theory→three existing measures→classification-based reduction" which concerns only one level. In contrast, our overall and long objective is to induce comprehensive D-Table attribute reduction from hierarchical and informal perspectives, and D-Table's hierarchical structures and informational measures naturally become a research basis. As a starting point, this paper concretely constructs D-Table's three-layer granular structures and three-way informational measures. This research thought is also clarified by Fig. 1 (mainly the virtual arrow as well as the question mark), and granular computing (GrC) and Bayes' theorem will be basically utilized.

(1) GrC within a trialistic framework [42] acts as a structural methodology to process hierarchical information. GrC extensively concerns information granulation and is applied in rough set theory [18,27,28,45]. Herein, D-Table and its informational measures concern D-Classification, D-Class, C-Classification, and C-Class to adhere to GrC, and thus, GrC-based multi-granule, multi-level, and multi-view can be fully utilized to implement structural construction and informational evolution, where granulation monotonicity must seek to underlie further attribute reduction.

(2) According to Bayes' theorem [1]:

$$p(W/D_t) = \frac{p(W) \times p(D_t/W)}{p(D_t)}, \tag{1}$$

where W and D_t mean the model parameter and observed data, respectively. The posterior probability is proportional to the produce of the prior probability and likelihood function, and thus, Bayesian inference highlights the posterior part [1]. Bayes' theorem and Bayesian inference are introduced into rough set theory to offer concrete results [8,26,30,41]. With

Table 1
Conditional and decisional classifications and classes.

Item	C-Classification	C-Class	D-Classification	D-Class
Mathematical symbol	$U/IND(\mathcal{A})$	$[x]^i_{\mathcal{A}},\ i = 1, .., n$	$U/IND(\mathcal{D})$	$X_j,\ j = 1, .., m$
Granular essence	Conditional granule set	Conditional granule	Decisional granule set	Decisional granule

regard to the D-Table, Bayes' theorem can induce hierarchical evolution and systematic relationships of informational measures, which thus enables it to become central.

The research developed in this paper is presented as follows. (1) With regard to the D-Table, three-layer structures (i.e., the macro-high, meso-middle, and micro-bottom) are hierarchically organized by the formal structure and systemic granularity. (2) Three-layer structures perform hierarchical bottom-middle-top evolution via both Bayes' theorem and integrated fusion to gain three-way informational measures, as well as their granulation monotonicity and system equations. Herein, the entropy $\Sigma p(.)\log p(.)$ is not applicable, and thus, we resort to the weighted entropy $\sum wp(.)\log p(.)$, which already exists in applications [15,24,47]. (3) Based on three-layer structures, the hierarchical evolution, superiority, and algorithms of three-way informational measures are discussed. (4) Finally, structural and informational results are effectively illustrated by a D-Table example. In terms of contributions, the D-Table's hierarchical structures are established to reveal constructional mechanisms and systematic relationships of informational measures, and thus, the obtained results deepen and interpret the existing information-theoretic measures and firmly underlie the D-Table's hierarchical, systematic, and informational reduction; in particular, our concrete research regarding the three-layer granular structures and three-way informational measures enriches the three-way decisions theory, especially from its wide view [19]. These contributions can be well reflected by Fig. 1 where our research covers the entire three levels.

The remainder of this paper is organized as follows. Section 2 reviews the D-Table and its information-theoretic measures on classifications. Section 3 presents the D-Table's three-layer granular structures. Section 4 hierarchically constructs the D-Table's three-way informational measures, including the three-way probabilities at the micro-bottom and the three-way weighted entropies at the meso-middle and macro-top. Section 5 utilizes the three-layer structures to make hierarchical analyses of three-way informational measures, including the hierarchical evolution, superiority, and algorithms. Section 6 provides a D-Table example for a relevant illustration. Finally, Section 7 concludes this paper and highlights our contributions on informational measures and attribute reduction, which enrich the three-way decisions.

2. A decision table and its information-theoretic measures on classifications

This section reviews the D-Table and its information-theoretic measures on classifications.

2.1. A decision table

This subsection introduces the D-Table by Pawlak [25].

Rough set theory focuses on the data that is represented in an information table:

$$(U, AT, \{V_a : a \in AT\}, \{I_a : a \in AT\}),$$

where U is the universe with finite objects, AT is the finite attribute set, V_a is the value domain for $a \in AT$, and I_a: $U \to V_a$ is an information function. Each object x takes a value $I_a(x)$ on attribute a. The D-Table is a special type of information table with $AT = \mathcal{C} \cup \mathcal{D}$ and $\mathcal{C} \cap \mathcal{D} = \emptyset$, where $\mathcal{C}$ and $\mathcal{D}$ denote the sets of condition attribute and decision attribute, respectively.

Next, D-Table is simply denoted by $(U, \mathcal{C} \cup \mathcal{D})$ for the sake of discussion. In view of attribute reduction, the condition part concerns the parameter $\mathcal{A} \subseteq \mathcal{C}$ (rather than only the constant $\mathcal{C}$), while the decision part concerns the constant $\mathcal{D}$.

$$IND(\mathcal{A}) = \{(x, y) \in U \times U : \forall a \in \mathcal{A}, I_a(x) = I_a(y)\}$$

serves as an equivalence relation to cause C-Class $[x]_{\mathcal{A}}$, which implies a type of basic granule. The classified structure $U/IND(\mathcal{A}) = \{[x]_{\mathcal{A}} : x \in U\}$ means knowledge or C-Classification, where $U/IND(\emptyset) = \{U\}$. Suppose that $U/IND(\mathcal{A}) = \{[x]^i_{\mathcal{A}} : i = 1, .., n\}$, and thus, $|U/IND(\mathcal{A})| = n$. Similarly, $\mathcal{D}$ can induce the equivalence relation $IND(\mathcal{D})$ and further D-Classification $U/IND(\mathcal{D}) = \{X_j :\ j = 1, .., m\}$, which consists of m D-Classes. Four basic notions of the D-Table are summarized in Table 1.

Attribute reduction requires a parameter change of $\mathcal{A} \subseteq \mathcal{C}$, and this change implies knowledge/C-Classification granulation. If $\mathcal{B} \subseteq \mathcal{A} \subseteq \mathcal{C}$, then $U/IND(\mathcal{A})$ and $U/IND(\mathcal{B})$ correspond to the finer and coarser granulation structures, respectively. As a result, a partial order of transformations emerges, and the relevant relation and granulation with regard to coarsening are denoted by

$$U/IND(\mathcal{A}) \xrightarrow{\preceq} U/IND(\mathcal{B}); \tag{2}$$

Accordingly,

$$\forall [x]_{\mathcal{B}} \in U/IND(\mathcal{B}),\ \exists k \in \mathbf{N},\ s.t.,\ \bigcup_{t=1}^{k} [x]^t_{\mathcal{A}} = [x]_{\mathcal{B}}.$$

According to [48], knowledge coarsening $U/IND(\mathcal{A}) \xrightarrow{\preceq} U/IND(\mathcal{B})$ implies some groups of granule merging, and a representative group is represented by

$$\bigcup_{t=1}^{k}[x]_{\mathcal{A}}^{t} \xrightarrow{=} [x]_{\mathcal{B}}. \tag{3}$$

For attribute reduction, knowledge granulation provides the GrC mechanism for presentational uncertainty, and granulation monotonicity becomes a fundamental criterion for evaluating uncertainty measures [13,23,33,48]. Based on joint integration, an uncertainty measure necessarily has granulation monotonicity, if it has monotonicity in every group of granular merging. Therefore, granulation monotonicity can be effectively probed by observing a representative group of granules merging.

2.2. Information-theoretic measures on classifications

Regarding the D-Table, this subsection reviews information-theoretic measures on classifications by Miao [21].
First, we can define a mapping on σ-algebra 2^U, i.e.,

$$p : 2^U \to \mathbf{Q},\ \ p(T) = \frac{|T|}{|U|},\ \ \forall T \subseteq U, \tag{4}$$

where |.| means the set cardinality. $(U, 2^U, p)$ constitutes a probability space, where conditional probability $p(T/T_0) = \frac{|T|}{|T_0|}$ (suppose $T_0 \subseteq U$ and $|T_0| \neq 0$). This mathematical space establishes the usual probability framework of rough set theory, and thus, informational measures on classifications can be directly constructed by referring to information theory.

Definition 1 ([21])**.** The entropy on C-Classification $U/IND(\mathcal{A})$ is

$$H(\mathcal{A}) = -\sum_{i=1}^{n} p([x]_{\mathcal{A}}^{i})\log\, p([x]_{\mathcal{A}}^{i}), \tag{5}$$

where the function *log* has the base number of 2; similarly, we can obtain

$$H(\mathcal{D}) = -\sum_{j=1}^{m} p(X_j)\log\, p(X_j). \tag{6}$$

The conditional entropy on D-Classification $U/IND(\mathcal{D})$ given C-Classification $U/IND(\mathcal{A})$ is

$$H(\mathcal{D}/\mathcal{A}) = -\sum_{i=1}^{n}\left[p([x]_{\mathcal{A}}^{i})\sum_{j=1}^{m} p(X_j/[x]_{\mathcal{A}}^{i})\log\, p(X_j/[x]_{\mathcal{A}}^{i})\right]; \tag{7}$$

Similarly, we can obtain

$$H(\mathcal{A}/\mathcal{D}) = -\sum_{j=1}^{m}\left[p(X_j)\sum_{i=1}^{n} p([x]_{\mathcal{A}}^{i}/X_j)\log\, p([x]_{\mathcal{A}}^{i}/X_j)\right]. \tag{8}$$

Furthermore, the mutual information is the difference between the informational and conditional entropies, i.e.,

$$I(\mathcal{A};\mathcal{D}) = H(\mathcal{D}) - H(\mathcal{D}/\mathcal{A}) = H(\mathcal{A}) - H(\mathcal{A}/\mathcal{D}) = I(\mathcal{D};\mathcal{A}). \tag{9}$$

Three information-theoretic measures (on classifications) are determined in rough set theory, and they hold the basic semantics or function. A classification with multiple granules can be viewed as an information source, and its entropy measures its average information content to represent its uncertainty. C-Classification and D-Classification act as two information sources. Their conditional entropy measures the average information content and uncertainty for the main classification after achieving the premise classification, while their mutual information measures the information content from one classification to the other to represent the reduced uncertainty of one classification when knowing the other. In short, the information-theoretic measures become fundamental for uncertainty representation in rough set theory.

Theorem 1 ([21])**.** *The entropy, conditional entropy, and mutual information have granulation monotonicity. Concretely,*

$$U/IND(\mathcal{A}) \xrightarrow{\preceq} U/IND(\mathcal{B}) \implies H(\mathcal{B}) \le H(\mathcal{A}),\ \ H(\mathcal{D}/\mathcal{B}) \ge H(\mathcal{D}/\mathcal{A}),\ \ I(\mathcal{B};\mathcal{D}) \le I(\mathcal{A};\mathcal{D}). \tag{10}$$

For knowledge granulation, information-theoretic measures have monotonicity, and thus, they can be utilized to construct informational reduction via the informational preservation criterion.

Definition 2 ([21,32])**.** $\mathcal{B}$ is an entropy-based reduct of $\mathcal{C}$ if it satisfies the following two conditions:

(1) $H(\mathcal{B}) = H(\mathcal{C})$;

(2) $H(\mathcal{B} - \{b\}) < H(\mathcal{B}), \forall b \in \mathcal{B}$.

Table 2
Basic descriptions of a D-Table's three-layer granular structures.

Structure	Composition	Granular scale	Granular level	Simple name	Relevant reduction
I	$U/IND(\mathcal{A})$, $U/IND(\mathcal{D})$	Macro	Top	Macro-Top	Classification-based attribute reduction
II	$U/IND(\mathcal{A})$, X_j	Meso	Middle	Meso-Middle	Class-specific attribute reduction
III	$[x]^i_{\mathcal{A}}$, X_j	Micro	Bottom	Micro-Bottom	Category reduction or transformational attribute reduction

Structure IV contains $[x]^i_{\mathcal{A}}$ and $U/IND(\mathcal{D})$ to concern the granular meso scale and middle level, but it never has reduction connections.

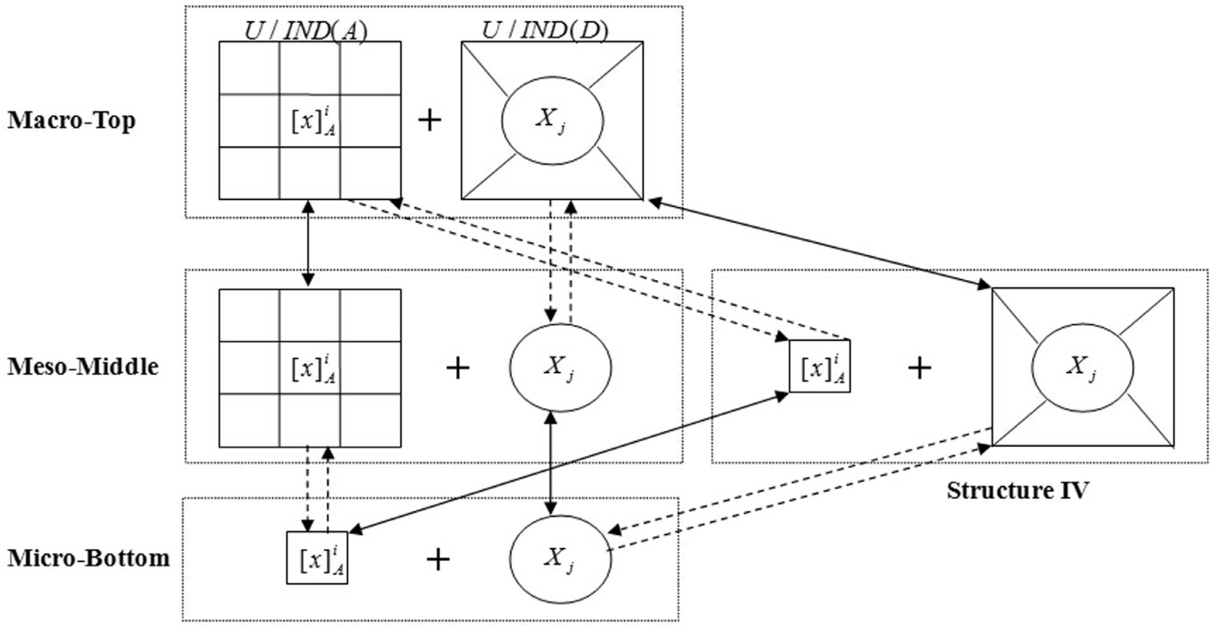

Fig. 2. Hierarchical/Granular relationships of the D-Table's three-layer granular structures.

$\mathcal{B}$ is a mutual information-based reduct of $\mathcal{C}$ if it satisfies the following two conditions:

(1) $I(\mathcal{B}; \mathcal{D}) = I(\mathcal{C}; \mathcal{D})$;

(2) $I(\mathcal{B} - \{b\}; \mathcal{D}) < I(\mathcal{B}; \mathcal{D}), \forall b \in \mathcal{B}$.

Two types of informational reducts are determined, and they correspond to the informational viewpoint and representation [21,32]. The former is applied to the information table $(U, \mathcal{C})$ and is related to the knowledge-based reduct, while the latter is applied to the D-Table $(U, \mathcal{C} \cup \mathcal{D})$ and can also be represented by the conditional entropy.

3. Three-layer granular structures of a decision table

Aiming at the D-Table $(U, \mathcal{C} \cup \mathcal{D})$, this section utilizes GrC technology to establish three-layer granular structures, which underlie later information construction.

According to the four granular notions presented in Table 1, D-Table contains two types of classifications to contain multiple granules. C-Classification $U/IND(\mathcal{A})$ has n C-Classes with regard to $[x]^i_{\mathcal{A}}$, while D-Classification $U/IND(\mathcal{D})$ has m D-Classes with regard to X_j. The relevant classification and class lead to three-layer granular structures, as shown in Table 2.

According to Table 2, the D-Table carries three-layer granular structures, which are accompanied by four concrete structures. The macro scale and top level produce Structure I, which is the bearing basis of classification-based attribute reduction [25]. In particular, the meso scale and middle level produce two types of symmetrical structures: Structures II and IV; the former underlies class-specific attribute reduction [43], while the latter has no significance for (attribute) reduction. The micro scale and bottom level produce Structure III, which is related to category reduction or its transformation [25,39]. Therefore, the main Structures I, II, III constitute three-layer granular structures and thus are named the vivid Macro-Top, Meso-Middle, and Micro-Bottom, respectively, while Structure IV implements some supplementary material.

Three-layer granular structures and their hierarchical/granular relationships are also elaborated in Fig. 2. There, two composition elements are individually set up to become clear, although they simultaneously exist in the universe U. The arrow shows a relevant change process between the classification and class. From the GrC perspective, Macro-Top→Meso-

Middle→Micro-Bottom implies concretization for a specific class in the top-bottom direction, while the opposite means generalization for the classes family in the bottom-to-top direction.

Next, three-layer granular structures are mainly considered from a systematic viewpoint, and the numeric result and decomposition/merging relationships are described.

Macro-Top, Meso-Middle, Structure IV have 1, m, n parallel patterns, respectively, while the Micro-Bottom has $m \times n$ parallel patterns. Thus, Macro-Top becomes sole for the given $\mathcal{A}$ and the constant $\mathcal{D}$, m D-Classes imply parallel m Meso-Middles for the given C-Classification $U/IND(\mathcal{A})$, while the n C-Classes imply parallel n patterns of Structure IV for the given D-Classification $U/IND(\mathcal{D})$.

The following conclusion presents the systematic decomposition (from a higher level to a lower level), which matches the downward arrow direction in Fig. 2.

(1) One Macro-Top can be decomposed into m Meso-Middles, while one Meso-Middle can be decomposed into n Micro-Bottoms.
(2) One Macro-Top can be decomposed into n patterns of Structure IV, while one pattern of Structure IV can be decomposed into m Micro-Bottoms.
(3) One Macro-Top can be decomposed into $m \times n$ Micro-Bottoms. The top-to-bottom decomposition has two equivalent approaches through having two middle structures with regard to the Meso-Middle and Structure IV.
In contrast, the following conclusion presents systematic merging (from a lower level to a higher level), which matches the upward arrow direction.

(1) The related n Micro-Bottoms can be merged into one Meso-Middle, while the related m Meso-Middles can be merged into one Macro-Top.
(2) The related m Micro-Bottoms can be merged into one pattern of Structure IV, while related n patterns of Structure IV can be merged into one Macro-Top.
(3) All $m \times n$ Micro-Bottoms can be merged into one Macro-Top. The bottom-top merging has two equivalent approaches based on the Meso-Middle and Structure IV, if the middle structures are considered.
The relevant merging in the direction Micro-Bottom→Meso-Middle→Macro-Top provides a powerful GrC mechanism for hierarchical construction and informational fusion.

Thus far, the D-Table has gained its three-layer granular structures as well as their hierarchical and systematic relationships. The relevant construction and result are owed to the GrC technology with regard to the multi-granule, multi-level and multi-view, and they are actually determined by only the D-Table's formal structure. Therefore, they exhibit general significance for the D-Table's measure construction and attribute reduction.

4. Three-way informational measures of a decision table

Based on the above three-layer granular structures, this section investigates three-way informational measures and their hierarchies and systematicness. Herein, the bottom-middle-top merging direction is utilized to implement the hierarchical evolution and integrated fusion of the information measures. In Section 4.1, three-way probabilities are analyzed at the Micro-Bottom; in Section 4.2, they are hierarchically integrated to three weighted entropies at the Meso-Middle; in Section 4.3, three-way weighted entropies integratedly evolve from the Meso-Middle to Macro-Top. In particular, Bayes' theorem at the Micro-Bottom acts as a basic point for informational evolution; this theorem and its development present information systematicness at three levels. Moreover, granulation monotonicity is achieved, especially at the Meso-Middle and Macro-Top levels.

4.1. Three-way probabilities at the micro-bottom

At the Micro-Bottom, C-Class $[x]^i_{\mathcal{A}}$ and D-Class X_j are of concern. They exist in approximate space $(U, \mathcal{A})$ and can produce some fundamental measures, including probabilities. This subsection analyzes three-way probabilities, mainly by connecting the Meso-Middle and its reasoning mechanism. Three-way probabilities become bottomed measures that underlie informational construction at higher levels.

Within the probability framework (related to Eq. (4)), we can produce the following product formula:

$$p(X_j) \times p([x]^i_{\mathcal{A}}/X_j) = p([x]^i_{\mathcal{A}} \cap X_j) = p([x]^i_{\mathcal{A}}) \times p(X_j/[x]^i_{\mathcal{A}}). \tag{11}$$

Based on the mathematical transformation, this probabilistic formula induces two types of Bayes' theorem, which are related to two middle structures. To clarify the related mechanism, contrasting the Meso-Middle and Structure IV is first provided in Fig. 3.

Fig. 3 extracts the middle structure of Fig. 2 but integrates two composition elements with conditional and decisional parts. This figure provides two structural mechanisms and the relevant systems. (1) Regarding the Meso-Middle, C-Classes $[x]^i_{\mathcal{A}}$ $(i = 1, \ldots, n)$ form a partition of U. One D-Class X_j exhibits the following total probability:

$$p(X_j) = \sum_{i=1}^{n} p([x]^i_{\mathcal{A}})p(X_j/[x]^i_{\mathcal{A}}),$$

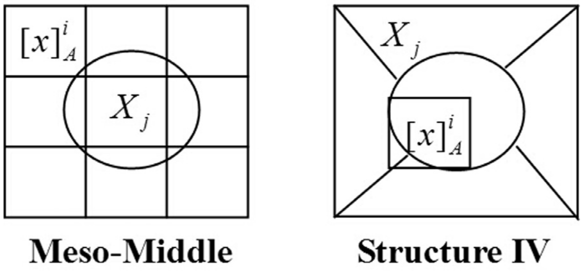

Fig. 3. Meso-Middle and Structure IV underlying Bayes' theorem.

and Bayes' formula becomes

$$p([x]^i_A/X_j) = \frac{p([x]^i_A) \times p(X_j/[x]^i_A)}{\sum\limits_{i=1}^{n} p([x]^i_A)p(X_j/[x]^i_A)}. \tag{12}$$

On this basis, the Micro-Bottom induces Bayes' theorem:

$$p([x]^i_A/X_j) = \frac{p([x]^i_A) \times p(X_j/[x]^i_A)}{p(X_j)}. \tag{13}$$

Four probabilities emerge, but $p(X_j)$ can be viewed as a constant (in view of the Meso-Middle). The three surplus probabilities are worthwhile to discuss and are named the likelihood, prior, and posterior probabilities in [47]. (2) In contrast, Structure IV and its partition with regard to X_j $(j = 1, \ldots, m)$ induce Bayes' theorem at the Micro-Bottom:

$$p(X_j/[x]^i_A) = \frac{p(X_j) \times p([x]^i_A/X_j)}{p([x]^i_A)}. \tag{14}$$

$p(X_j)$ and $p(X_j/[x]^i_A)$ are called the prior and posterior probabilities, respectively, in [30].

Both Bayes' theorems (i.e., Eqs. (13) and (14)) correspond to the product formula (i.e., Eq. (11)), which becomes equivalent and symmetrical in terms of the mathematics, but they have different emphases. The former is related to the Meso-Middle, which underlies class-specific attribute reduction [43], and thus, it becomes important for three-layer granular structures and their hierarchical reduction. The latter is related to Structure IV, and thus, it adheres to the practical determination mechanism (that X_j and $[x]^i_A$ act as the internal cause and external manifestation, respectively), but it goes against the hierarchical structure and reduction related to specific D-Class. Correspondingly, the probability naming based on Eq. (13) in [47] originates from the formal structure with regard to the Meso-Middle, while the prior and posterior probabilities based on Eq. (14) in [30] follow the practical pattern related to Eq. (1) and thus become more general. To link appropriately to the main Meso-Middle, the first Bayes' theorem (i.e., Eq. (13)) is adopted, and thus, three-way probabilities emerge but their naming is neglected.

Definition 3. At the Micro-Bottom, three-way probabilities are defined by

$$p(X_j/[x]^i_A) = \frac{|[x]^i_A \cap X_j|}{|[x]^i_A|},\ \ p([x]^i_A) = \frac{|[x]^i_A|}{|U|},\ \ p([x]^i_A/X_j) = \frac{|[x]^i_A \cap X_j|}{|X_j|}. \tag{15}$$

Theorem 2. *Three-way probabilities hold systematicness with regard to Bayes' theorem (i.e., Eq. (13)).*

In view of the Meso-Middle, the three-way probabilities are determined to exhibit systematicness. According to Fig. 3, they could have some explanation from rule reasoning, where $[x]^i_A$ and X_j become the predecessor and successor for a decision rule $[x]^i_A \Rightarrow X_j$, respectively [25]. Moreover, their quantitative features can be revealed by their forms in Eq. (15). (1) Conditional probability $p(X_j/[x]^i_A)$ depends on predecessor $[x]^i_A$ to describe successor X_j. It relatively measures the interaction information between $[x]^i_A$ and X_j on the basis of the former. Its definition exhibits the informational feature with regard to the relativity, concentration, and locality. (2) C-Class probability $p([x]^i_A)$ measures only the predecessor and never concerns the successor and further interaction. It embodies absoluteness, directness, and globality with regard to U. (3) Conditional probability $p([x]^i_A/X_j)$ describes the predecessor on the condition of the successor. It absolutely measures the interaction information on the premise of the successor, and it exhibits absoluteness, directness, and globality with regard to X_j.

$p([x]^i_A)$ never concerns rule reasoning. $p(X_j/[x]^i_A)$ and $p([x]^i_A/X_j)$ utilize core interaction information $|[x]^i_A \cap X_j|$ to directly reflect reasoning, but they have different preference premises for the successor and predecessor, which concern the predecessor-successor and successor-predecessor directions, respectively. From the double-quantitative perspective [5,6,37,46,48], they depend on the information concentration and data directness to express a type of relative and absolute reasoning possibilities, respectively.

Theorem 3. *For three-way probabilities, $p([x]^i_{\mathcal{A}})$ and $p([x]^i_{\mathcal{A}}/X_j)$ have granulation monotonicity:*

$$U/IND(\mathcal{A}) \xrightarrow{\preceq} U/IND(\mathcal{B}) \implies p([x]^i_{\mathcal{B}}) \le p([x]^i_{\mathcal{A}}),\ p([x]^i_{\mathcal{B}}/X_j) \le p([x]^i_{\mathcal{A}}/X_j), \tag{16}$$

but $p(X_j/[x]^i_{\mathcal{A}})$ does not necessarily have granulation monotonicity.

Three-way probabilities draw basic conclusions of granulation monotonicity/non-monotonicity, where

$$U/IND(\mathcal{A}) \xrightarrow{\preceq} U/IND(\mathcal{B}) \implies [x]^i_{\mathcal{B}} \subseteq [x]^i_{\mathcal{A}} \tag{17}$$

acts as a granular root. Hence, only $p([x]^i_{\mathcal{A}})$ and $p([x]^i_{\mathcal{A}}/X_j)$ can be directly utilized for category reduction (or its transformational attribute reduction), which is located at the Micro-Bottom.

Three-way probabilities gain systematicness and monotonicity/non-monotonicity, and their reasoning connotations and quantification features underlie further combined construction for benign measures. For example, $p([x]^i_{\mathcal{A}})$ and $p(X_j/[x]^i_{\mathcal{A}})$ are fused by the weighed product to mine a monotonic uncertainty measure, which is used to implement hierarchical construction and attribute reduction [48]. Next, three-way probabilities are fused by the weighed product of information function $p(.)\log p(.)$, and the further sum integration with regard to the C-Classes will lead to three-way weighted entropies at the Meso-Middle.

4.2. Three-way weighted entropies at the meso-middle

Based on three-way probabilities at the Micro-Bottom, this subsection constructs three-way weighted entropies at the Meso-Middle and reveals their granulation monotonicity and evolution systematicness. Relevant results take a link function to underlie the latter informational construction at the Macro-Top.

A promotional measure at the Meso-Middle requires probability fusion when integrating C-Classes into C-Classification. First, the entropy is inspected in view of its fundamentality. With regard to C-Classification $U/IND(\mathcal{A})$, $p([x]^i_{\mathcal{A}})$ and $p([x]^i_{\mathcal{A}}/X_j)$ form their probabilistic distributions, but $p(X_j/[x]^i_{\mathcal{A}})$ cannot because usually $\sum_{i=1}^n p(X_j/[x]^i_{\mathcal{A}}) \neq 1$. By using the informational function $\Sigma p(.)\log p(.)$, we generally derive three-way entropies at the Meso-Middle:

$$H(X_j/\mathcal{A}) = -\sum_{i=1}^n p(X_j/[x]^i_{\mathcal{A}})\log p(X_j/[x]^i_{\mathcal{A}}), \tag{18}$$

$$H^{X_j}(\mathcal{A}) = -\sum_{i=1}^n p([x]^i_{\mathcal{A}})\log p([x]^i_{\mathcal{A}}), \tag{19}$$

$$H(\mathcal{A}/X_j) = -\sum_{i=1}^n p([x]^i_{\mathcal{A}}/X_j)\log p([x]^i_{\mathcal{A}}/X_j). \tag{20}$$

According to entropy properties, $H^{X_j}(\mathcal{A})$ and $H(\mathcal{A}/X_j)$ naturally have granulation monotonicity:

$$U/IND(\mathcal{A}) \xrightarrow{\preceq} U/IND(\mathcal{B}) \implies H^{X_j}(\mathcal{B}) \le H^{X_j}(\mathcal{A}),\ H(\mathcal{B}/X_j) \le H(\mathcal{A}/X_j);$$

However, $H(X_j/\mathcal{A})$ does not necessarily have granulation monotonicity. The granulation monotonicity/non-monotonicity is verified by an example in Appendix A.

Three-way entropies go against the monotonicity demand and lack clear systematicness. We need to go beyond function $\Sigma p(.)\log p(.)$ to seek benign measures that have both monotonicity and systematicness. This requirement triggers the following evolution of weighted entropies.

Bayes' theorem provides systematicness of three-way probabilities and, thus, becomes the starting point. Herein, we first make a key transformation for Bayes' theorem. According to Eq. (13) with stable X_j,

$$p([x]^i_{\mathcal{A}}/X_j) = \frac{p([x]^i_{\mathcal{A}}) \cdot p(X_j/[x]^i_{\mathcal{A}})}{p(X_j)},\ \forall i \in \{1, \ldots, n\}.$$

By calculating the information item $-p(.)\log p(.)$ with regard to $p([x]^i_{\mathcal{A}}/X_j)$,

$$-p([x]^i_{\mathcal{A}}/X_j)\log p([x]^i_{\mathcal{A}}/X_j) = -\frac{p([x]^i_{\mathcal{A}}) \cdot p(X_j/[x]^i_{\mathcal{A}})}{p(X_j)}[\log p([x]^i_{\mathcal{A}}) + \log p(X_j/[x]^i_{\mathcal{A}}) - \log p(X_j)].$$

Based on the factor $p(X_j)$ multiplication and further consolidation,

$$\begin{aligned} &-p(X_j)p([x]^i_{\mathcal{A}}/X_j)\log p([x]^i_{\mathcal{A}}/X_j) \\ &= -p(X_j/[x]^i_{\mathcal{A}})p([x]^i_{\mathcal{A}})\log p([x]^i_{\mathcal{A}}) - p([x]^i_{\mathcal{A}})p(X_j/[x]^i_{\mathcal{A}})\log p(X_j/[x]^i_{\mathcal{A}}) + p([x]^i_{\mathcal{A}})p(X_j/[x]^i_{\mathcal{A}})\log p(X_j). \end{aligned}$$

According to the i-based summation,

$$-\sum_{i=1}^{n} p(X_j)p([x]^i_{\mathcal{A}}/X_j)\log\, p([x]^i_{\mathcal{A}}/X_j)$$
$$= -\sum_{i=1}^{n} p(X_j/[x]^i_{\mathcal{A}})p([x]^i_{\mathcal{A}})\log\, p([x]^i_{\mathcal{A}}) - \sum_{i=1}^{n} p([x]^i_{\mathcal{A}})p(X_j/[x]^i_{\mathcal{A}})\log\, p(X_j/[x]^i_{\mathcal{A}}) + \sum_{i=1}^{n} p([x]^i_{\mathcal{A}})p(X_j/[x]^i_{\mathcal{A}})\log\, p(X_j). \tag{21}$$

The final item in Eq. (21) becomes

$$\sum_{i=1}^{n} p([x]^i_{\mathcal{A}})p(X_j/[x]^i_{\mathcal{A}})\log\, p(X_j) = \sum_{i=1}^{n} p([x]^i_{\mathcal{A}} \cap X_j)\log\, p(X_j) = \left[\sum_{i=1}^{n} p([x]^i_{\mathcal{A}} \cap X_j)\right] \times \log\, p(X_j) = p(X_j)\log\, p(X_j). \tag{22}$$

The above step-by-step deduction implies the hierarchical evolution of Bayes' theorem. Bayes' theorem and its three-way probabilities at the Micro-Bottom evolve in the entropy direction, and thus, weight-based entropies and their relationships emerge at the Mecro-Middle. Concretely, Eq. (22) provides a constant that is based on X_j, and thus, systematic Eq. (21) concerns three weighted and informational items. Next, we introduce the weighted entropy, which already has basic applications [15,24,47]. Suppose that (ξ, p_i) denotes a probability distribution and $w_i \geq 0$ means the weight; then, the weighted entropy is defined by

$$H_W(\xi) = -\sum_{i=1}^{n} w_i p_i \log\, p_i.$$

The weighted entropy introduces weights into the entropy, where the weights reflect the importance degrees for information receivers or attention degrees of information receivers, and the weighted entropy develops the entropy while degenerates into the latter via $w_i = 1$.

According to the probability distribution, $p([x]^i_{\mathcal{A}})$ and $p([x]^i_{\mathcal{A}}/X_j)$ naturally produce weighted entropies, while $p(X_j/[x]^i_{\mathcal{A}})$ tolerably induces a weighted entropy in a general sense. Thus, Eq. (21) produces three-way weighted entropies and their systematicness, and the symbol $H_W(.)$ is used to differ from the entropy $H(.)$.

Definition 4. At the Meso-Middle, three-way weighted entropies are defined by

$$\begin{aligned} H_W(X_j/\mathcal{A}) &= -\sum_{i=1}^{n} p([x]^i_{\mathcal{A}})p(X_j/[x]^i_{\mathcal{A}})\log\, p(X_j/[x]^i_{\mathcal{A}}), \\ H_W^{X_j}(\mathcal{A}) &= -\sum_{i=1}^{n} p(X_j/[x]^i_{\mathcal{A}})p([x]^i_{\mathcal{A}})\log\, p([x]^i_{\mathcal{A}}), \\ H_W(\mathcal{A}/X_j) &= -\sum_{i=1}^{n} p(X_j)p([x]^i_{\mathcal{A}}/X_j)\log\, p([x]^i_{\mathcal{A}}/X_j). \end{aligned} \tag{23}$$

Proposition 1. *The weighted entropy $H_W(\mathcal{A}/X_j)$ is the product of the constant $p(X_j)$ and the entropy $H(\mathcal{A}/X_j)$. In other words,*

$$H_W(\mathcal{A}/X_j) = p(X_j) \times \left[-\sum_{i=1}^{n} p([x]^i_{\mathcal{A}}/X_j)\log\, p([x]^i_{\mathcal{A}}/X_j)\right] = p(X_j)H(\mathcal{A}/X_j).$$

Three-way weighted entropies originate from three-way entropies by introducing weight coefficients of specific probabilities, and thus, they implement double-quantitative fusion [5,6,37,46,48] to acquire better informational features. $H_W^{X_j}(\mathcal{A})$ improves absolute $H^{X_j}(\mathcal{A})$ by introducing relative $p(X_j/[x]^i_{\mathcal{A}})$ to the importance weights, while $H_W(\mathcal{A}/X_j)$ and $H_W(X_j/\mathcal{A})$ improve the relative $H(\mathcal{A}/X_j)$ and $H(X_j/\mathcal{A})$ by introducing absolute $p(X_j)$ and $p([x]^i_{\mathcal{A}})$, respectively. Herein, $H_W(\mathcal{A}/X_j)$ has a simpler form according to Proposition 1. In other words, three-way weighted entropies inherit the essential uncertainty semantics of three-way properties by using different probability weights, and thus, they become robust for uncertainty measurement. Their superiority is next clarified by their perfect monotonicity and systematicness.

Theorem 4. *At the Meso-Middle, three-way weighted entropies have granulation monotonicity. Concretely,*

$$U/IND(\mathcal{A}) \xrightarrow{\preceq} U/IND(\mathcal{B}) \implies H_W(X_j/\mathcal{B}) \geq H_W(X_j/\mathcal{A}),\ H_W^{X_j}(\mathcal{B}) \leq H_W^{X_j}(\mathcal{A}),\ H_W(\mathcal{B}/X_j) \leq H_W(\mathcal{A}/X_j). \tag{24}$$

Theorem 4 is proved in Appendix B. Although the entropy $H(X_j/\mathcal{A})$ is initially non-monotonic, the weighted entropy $H_W(X_j/\mathcal{A})$ becomes monotonic. For the $H_W(X_j/\mathcal{A})$ monotonicity, the relevant proof becomes difficult and requires a mathematical trick, and thus, a concave feature of the function $-u\log u$ is effectively utilized in Appendix B.

Theorem 5. *Three-way weighted entropies have systematicness:*

$$H_W(\mathcal{A}/X_j) = H_W^{X_j}(\mathcal{A}) + H_W(X_j/\mathcal{A}) + p(X_j)\log\ p(X_j). \tag{25}$$

In other words, $H_W(\mathcal{A}/X_j)$ is a linear translation of the sum of $H_W^{X_j}(\mathcal{A})$ and $H_W(X_j/\mathcal{A})$, where $p(X_j)\log p(X_j)$ is a constant at the Meso-Middle.

Eq. (25) originates from Eqs. (21)–(23), and it develops Bayes' theorem at the Micro-Bottom to establish a systematic equation of three-way weighted entropies. Furthermore, it can change into

$$H_W(\mathcal{A}/X_j) = H_W^{X_j}(\mathcal{A}) - [-p(X_j)\log\ p(X_j) - H_W(X_j/\mathcal{A})]; \tag{26}$$

the linear transformation item with regard to the above $H_W(X_j/\mathcal{A})$, which exhibits non-negativity by using $U/IND(\mathcal{A}) \xrightarrow{\preceq} U/IND(\emptyset)$, i.e.,

$$\begin{aligned} &-p(X_j)\log\ p(X_j) - H_W(X_j/\mathcal{A}) \\ &\geq -p(X_j)\log\ p(X_j) - H_W(X_j/\emptyset)) = -p(X_j)\log\ p(X_j) + p(U)p(X_j/U)\log\ p(X_j/U) = 0. \end{aligned}$$

According to Eqs. (25) and (26), eliminating the conversion distance can produce a new measure to simplify the systematic equation.

Definition 5. At the Meso-Middle, the linear weighted entropy with regard to the weighted entropy $H_W(X_j/\mathcal{A})$ is defined by

$$H_W^{\text{lin}}(X_j/\mathcal{A}) = -p(X_j)\log\ p(X_j) - H_W(X_j/\mathcal{A}). \tag{27}$$

Corollary 1. *The linear weighted entropy $H_W^{\text{lin}}(X_j/\mathcal{A})$ has granulation monotonicity:*

$$U/IND(\mathcal{A}) \xrightarrow{\preceq} U/IND(\mathcal{B}) \implies H_W(X_j/\mathcal{B}) \leq H_W^{\text{lin}}(X_j/\mathcal{A}). \tag{28}$$

Corollary 2. *Three-way weighted entropies have the equivalent systematicness:*

$$H_W(\mathcal{A}/X_j) = H_W^{X_j}(\mathcal{A}) - H_W^{\text{lin}}(X_j/\mathcal{A}). \tag{29}$$

In other words, $H_W(\mathcal{A}/X_j)$ is the difference of $H_W^{X_j}(\mathcal{A})$ and $H_W^{\text{lin}}(X_j/\mathcal{A})$.

The linear weighted entropy $H_W^{\text{lin}}(X_j/\mathcal{A})$ corresponds to $H_W(X_j/\mathcal{A})$ by virtue of a specific linear transformation. The former uses the superscript *lin* (which means *linear*) to differ from the latter, but both are viewed as only one item for three-way weighted entropies. In contrast to $H_W(X_j/\mathcal{A})$, $H_W^{\text{lin}}(X_j/\mathcal{A})$ exhibits opposite granulation monotonicity, and it simplifies the systematicness of three-way weighted entropies.

In summary, this section at the Meso-Middle becomes important to link the Micro-Bottom and Macro-Top. Bayes' theorem (i.e., Eq. (13)) provides three-way probabilities systematicness, and it further plays a fundamental role in the informational evolution of weighted entropies. It induces essential measures and systematic equations of three-way weighted entropies. Next, three-way weighted entropies are promoted from the Meso-Middle to the Macro-Top.

4.3. Three-way weighted entropies at the macro-top

For three-way weighted entropies at the Meso-Middle, their monotonicity and systematicness are established. They can hierarchically evolve to Macro-Top by using the natural sum integration with regard to multiple D-Classes. This subsection constructs three-way weighted entropies at the Macro-Top and offers their monotonicity and systematicness. Their equivalent relationships with the previous information-theoretic measures are finally clarified.

Definition 6. At Macro-Top, three-way weighted entropies are defined by

$$\begin{aligned} H_W(\mathcal{D}/\mathcal{A}) &= \sum_{j=1}^{m} H_W(X_j/\mathcal{A}) \ \left(or\ H_W^{\text{lin}}(\mathcal{D}/\mathcal{A}) = \sum_{j=1}^{m} H_W^{\text{lin}}(X_j/\mathcal{A})\right), \\ H_W^{\mathcal{D}}(\mathcal{A}) &= \sum_{j=1}^{m} H_W^{X_j}(\mathcal{A}), \\ H_W(\mathcal{A}/\mathcal{D}) &= \sum_{j=1}^{m} H_W(\mathcal{A}/X_j). \end{aligned} \tag{30}$$

Corollary 3. *$H_W^{\text{lin}}(\mathcal{D}/\mathcal{A})$ is a linear transformation of $H_W(\mathcal{D}/\mathcal{A})$. According to Eqs. (27) and (30),*

$$H_W^{\text{lin}}(\mathcal{D}/\mathcal{A}) = -\sum_{j=1}^{m} p(X_j)\log\ p(X_j) - \sum_{j=1}^{m} H_W(X_j/\mathcal{A}) = H(\mathcal{D}) - H_W(\mathcal{D}/\mathcal{A}), \tag{31}$$

where entropy $H(\mathcal{D}) = -\sum_{j=1}^{m} p(X_j)\log p(X_j)$ *is in contrast.*

With regard to the Meso-Middle, the Macro-Top exhibits the hierarchical promotion and systematic integration from D-Classes to D-Classification. Accordingly, three-way weighted entropies at Macro-Top are integratedly fused by three-way weighted entropies at the Meso-Middle, and they exhibit a type of informational sum. $H_W^{\text{lin}}(\mathcal{D}/\mathcal{A})$ and $H_W(\mathcal{D}/\mathcal{A})$ exhibit a linear transformation to be viewed as only one item. Three-way weighted entropies at Macro-Top depend on the sum integration to naturally inherit monotonicity and systematicness at the Meso-Middle, and the relevant features are presented as follows.

Theorem 6. *At Macro-Top, three-way weighted entropies have granulation monotonicity. Concretely,*

$$\begin{aligned}&U/IND(\mathcal{A}) \xrightarrow{\preceq} U/IND(\mathcal{B}) \implies \\ &H_W(\mathcal{D}/\mathcal{B}) \geq H_W(\mathcal{D}/\mathcal{A}),\ H_W^{\text{lin}}(\mathcal{D}/\mathcal{B}) \leq H_W^{\text{lin}}(\mathcal{D}/\mathcal{A}),\ H_W^{\mathcal{D}}(\mathcal{B}) \leq H_W^{\mathcal{D}}(\mathcal{A}),\ H_W(\mathcal{B}/\mathcal{D}) \leq H_W(\mathcal{A}/\mathcal{D}).\end{aligned} \tag{32}$$

Theorem 7. *Three-way weighted entropies have systematicness:*

$$\begin{aligned}H_W(\mathcal{A}/\mathcal{D}) &= H_W^{\mathcal{D}}(\mathcal{A}) + H_W(\mathcal{D}/\mathcal{A}) - H(\mathcal{D}) \\ &= H_W^{\mathcal{D}}(\mathcal{A}) - [H(\mathcal{D}) - H_W(\mathcal{D}/\mathcal{A})] = H_W^{\mathcal{D}}(\mathcal{A}) - H_W^{\text{lin}}(\mathcal{D}/\mathcal{A}).\end{aligned} \tag{33}$$

In other words, $H_W(\mathcal{A}/\mathcal{D})$ *is a linear translation of the sum of* $H_W^{\mathcal{D}}(\mathcal{A})$ *and* $H_W(\mathcal{D}/\mathcal{A})$ *or the difference between* $H_W^{\mathcal{D}}(\mathcal{A})$ *and* $H_W^{\text{lin}}(\mathcal{D}/\mathcal{A})$.

At Macro-Top, an information-theoretic system already exists to contain the entropy, conditional entropy, and mutual information [21], which are reviewed in Definition 1. Herein, three-way weighted entropies establish a novel informational system. Both of the systems' relationships are analyzed next.

Theorem 8. *The weight-entropy system and information-theoretic system are equivalent. Concretely, the weighted entropies* $H_W(\mathcal{D}/\mathcal{A})$, $H_W^{\text{lin}}(\mathcal{D}/\mathcal{A})$, $H_W^{\mathcal{D}}(\mathcal{A})$, *and* $H_W(\mathcal{A}/\mathcal{D})$ *are equivalent to the conditional entropy* $H(\mathcal{D}/\mathcal{A})$, *mutual information* $I(\mathcal{A};\mathcal{D})$, *entropy* $H(\mathcal{A})$, *and conditional entropy* $H(\mathcal{A}/\mathcal{D})$, *respectively. In other words,*

$$H_W(\mathcal{D}/\mathcal{A}) = H(\mathcal{D}/\mathcal{A}),\ H_W^{\text{lin}}(\mathcal{D}/\mathcal{A}) = I(\mathcal{A};\mathcal{D}),\ H_W^{\mathcal{D}}(\mathcal{A}) = H(\mathcal{A}),\ H_W(\mathcal{A}/\mathcal{D}) = H(\mathcal{A}/\mathcal{D}). \tag{34}$$

Proof.

$$\begin{aligned}(1)\ H_W(\mathcal{D}/\mathcal{A}) &= H_W(X_1/\mathcal{A}) + \ldots + H_W(X_m/\mathcal{A}) \\ &= +[-p([x]_{\mathcal{A}}^1)p(X_1/[x]_{\mathcal{A}}^1)\log p(X_1/[x]_{\mathcal{A}}^1) - \ldots - p([x]_{\mathcal{A}}^n)p(X_1/[x]_{\mathcal{A}}^n)\log p(X_1/[x]_{\mathcal{A}}^n)] + \ldots \\ &\quad + [-p([x]_{\mathcal{A}}^1)p(X_m/[x]_{\mathcal{A}}^1)\log p(X_m/[x]_{\mathcal{A}}^1) - \ldots - p([x]_{\mathcal{A}}^n)p(X_m/[x]_{\mathcal{A}}^n)\log p(X_m/[x]_{\mathcal{A}}^n)] \\ &= -p([x]_{\mathcal{A}}^1)[p(X_1/[x]_{\mathcal{A}}^1)\log p(X_1/[x]_{\mathcal{A}}^1) + \ldots + p(X_m/[x]_{\mathcal{A}}^1)\log p(X_m/[x]_{\mathcal{A}}^1)] - \ldots \\ &\quad - p([x]_{\mathcal{A}}^n)[p(X_1/[x]_{\mathcal{A}}^1)\log p(X_1/[x]_{\mathcal{A}}^n) + \ldots + p(X_m/[x]_{\mathcal{A}}^n)\log p(X_m/[x]_{\mathcal{A}}^1)] \\ &= -\sum_{i=1}^{n}\left[p([x]_{\mathcal{A}}^i)\sum_{j=1}^{m} p(X_j/[x]_{\mathcal{A}}^i)\log p(X_j/[x]_{\mathcal{A}}^i)\right] \\ &= H(\mathcal{D}/\mathcal{A}).\end{aligned} \tag{35}$$

$$(2)\ H_W^{\text{lin}}(\mathcal{D}/\mathcal{A}) = H(\mathcal{D}) - H_W(\mathcal{D}/\mathcal{A}) = H(\mathcal{D}) - H(\mathcal{D}/\mathcal{A}) = I(\mathcal{A};\mathcal{D}). \tag{36}$$

$$\begin{aligned}(3)\ H_W^{\mathcal{D}}(\mathcal{A}) &= \sum_{j=1}^{m} H_W^{X_j}(\mathcal{A}) = -\sum_{j=1}^{m}\left[\sum_{i=1}^{n} p(X_j/[x]_{\mathcal{A}}^i)p([x]_{\mathcal{A}}^i)\log p([x]_{\mathcal{A}}^i)\right] \\ &= -\sum_{j=1}^{m}[p(X_j/[x]_{\mathcal{A}}^1)p([x]_{\mathcal{A}}^1)\log p([x]_{\mathcal{A}}^1) + \ldots + p(X_j/[x]_{\mathcal{A}}^n)p([x]_{\mathcal{A}}^n)\log p([x]_{\mathcal{A}}^n)] \\ &= -\left[\sum_{j=1}^{m} p(X_j/[x]_{\mathcal{A}}^1)\right]p([x]_{\mathcal{A}}^1)\log p([x]_{\mathcal{A}}^1) - \ldots - \left[\sum_{j=1}^{m} p(X_j/[x]_{\mathcal{A}}^n)\right]p([x]_{\mathcal{A}}^n)\log p([x]_{\mathcal{A}}^n) \\ &= -p([x]_{\mathcal{A}}^1)\log p([x]_{\mathcal{A}}^1) - \ldots - p([x]_{\mathcal{A}}^n)\log p([x]_{\mathcal{A}}^n) \\ &= H(\mathcal{A}).\end{aligned} \tag{37}$$

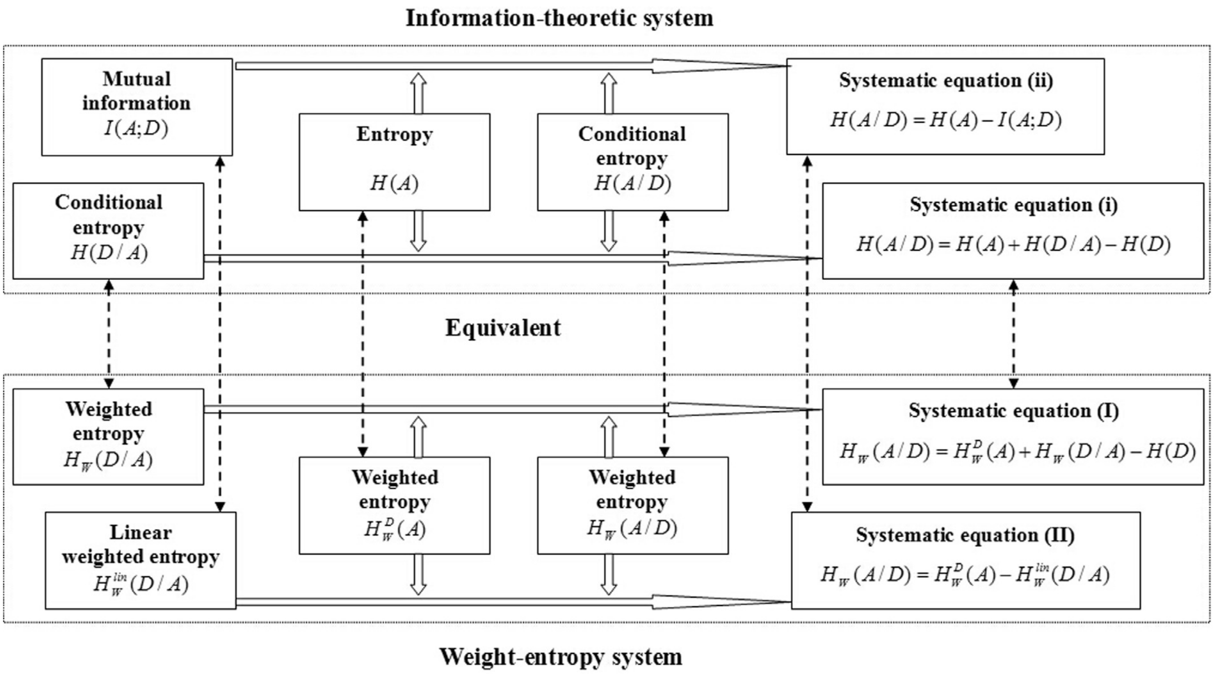

Fig. 4. Equivalence of the weight-entropy system and the information-theoretic system.

(4)
$$
\begin{aligned}
H_W(\mathcal{A}/\mathcal{D}) &= \sum_{j=1}^{m} H_W(\mathcal{A}/X_j) = \sum_{j=1}^{m} p(X_j)H(\mathcal{A}/X_j) \\
&= -\sum_{j=1}^{m}\left[p(X_j)\sum_{i=1}^{n} p([x]_{\mathcal{A}}^i/X_j)\log\, p([x]_{\mathcal{A}}^i/X_j)\right] \\
&= H(\mathcal{A}/\mathcal{D}).
\end{aligned} \tag{38}
$$

□

Corollary 4. *For the weight-entropy system and information-theoretic system, the corresponding informational measures have the same granulation monotonicity.*

Corollary 5. *The weight-entropy system and information-theoretic system have equivalent systematic equations. In other words,*

$$
\begin{aligned}
H_W(\mathcal{A}/\mathcal{D}) = H_W^{\mathcal{D}}(\mathcal{A}) + H_W(\mathcal{D}/\mathcal{A}) - H(\mathcal{D}) &\iff H(\mathcal{A}/\mathcal{D}) = H(\mathcal{A}) + H(\mathcal{D}/\mathcal{A}) - H(\mathcal{D}), \\
H_W(\mathcal{A}/\mathcal{D}) = H_W^{\mathcal{D}}(\mathcal{A}) - H_W^{\mathrm{lin}}(\mathcal{D}/\mathcal{A}) &\iff H(\mathcal{A}/\mathcal{D}) = H(\mathcal{A}) - I(\mathcal{A};\mathcal{D}).
\end{aligned} \tag{39}
$$

Theorem 8 and its corollaries reveal the equivalence and correspondence between the two informational systems. The relevant results are well clarified in a relationship figure: Fig. 4, where vertical virtual arrows show the corresponding equivalence for informational measures and systematicness.

What is the root cause for the theoretical equivalence of both systems? Based on the proof in Eqs. (35)–(38), the weight-entropy system mainly adopts an integration order on the C-Class-first and D-Class-second principle, i.e., n C-Classes are first fused by the weighted entropy function while m D-Classes are, second, integrated by the sum operation. In contrast, the information-theoretic system adopts the opposite order, i.e., m D-Classes are first fused by the entropy function, while n C-Classes are, second, integrated by the weight-sum operation. The above results and proofs reflect some operational commutativity, which can be fully reflected by the sum operation orders with regard to i and j. This commutativity ensures the final systematic equivalence, which is mainly from a mathematical viewpoint. In the next section, both systems will be compared by the D-Table's three-layer structures to manifest the reduction advancement of three-way informational measures.

Table 3
Three-layer granular structures and three-way informational measures.

Hierarchical structure	Components	Three-way informational measures	Granulation monotonicity	Systematic equations
Micro-Bottom	$[x]^i_{\mathcal{A}}$, X_j	$p(X_j/[x]^i_{\mathcal{A}})$, $p([x]^i_{\mathcal{A}})$, $p([x]^i_{\mathcal{A}}/X_j)$	Has in part	$p([x]^i_{\mathcal{A}}/X_j)$ $= p([x]^i_{\mathcal{A}}) \times p(X_j/[x]^i_{\mathcal{A}}) \div p(X_j)$
Meso-Middle	$U/IND(\mathcal{A})$, X_j	$H_W(X_j/\mathcal{A})$ (or $H_W^{\mathrm{lin}}(X_j/\mathcal{A})$), $H_W^{X_j}(\mathcal{A})$, $H_W(\mathcal{A}/X_j)$	Has	$H_W(\mathcal{A}/X_j)$, $= H_W^{X_j}(\mathcal{A}) + H_W(X_j/\mathcal{A}) + p(X_j)\log\, p(X_j)$, $= H_W^{X_j}(\mathcal{A}) - H_W^{\mathrm{lin}}(X_j/\mathcal{A})$
Macro-Top	$U/IND(\mathcal{A})$, $U/IND(\mathcal{D})$	$H_W(\mathcal{D}/\mathcal{A})$ (or $H_W^{\mathrm{lin}}(\mathcal{D}/\mathcal{A})$), $H_W^{\mathcal{D}}(\mathcal{A})$, $H_W(\mathcal{A}/\mathcal{D})$	Has	$H_W(\mathcal{A}/\mathcal{D})$ $= H_W^{\mathcal{D}}(\mathcal{A}) + H_W(\mathcal{D}/\mathcal{A}) - H(\mathcal{D})$ $= H_W^{\mathcal{D}}(\mathcal{A}) - H_W^{\mathrm{lin}}(\mathcal{D}/\mathcal{A})$

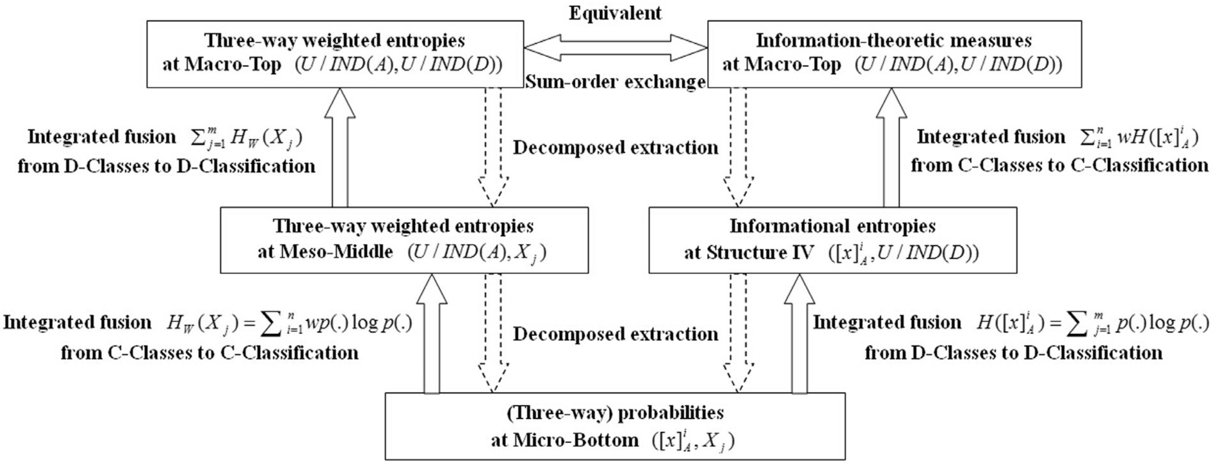

Fig. 5. Hierarchical evolution of informational measures.

5. Hierarchical analyses of three-way informational measures based on three-layer granular structures

Three-way information measures are hierarchically established, and their final weight-entropy system is equivalent to the existing information-theoretic system. By virtue of the three-layer structures, this section makes some hierarchical analyses of three-way informational measures, including the evolution summary, comparison superiority, and invocation algorithms.

5.1. Hierarchical evolution of three-way informational measures

Based on the three-layer structures, this subsection summarizes the hierarchical evolution of three-way informational measures.

Let us first clarify the overall research concept. According to the systemic classification and class, D-Table $(U, \mathcal{C}\bigcup\mathcal{D})$ contains three-layer structures to implement the hierarchical evolution of information measures. In probability space, three-way probabilities at the Micro-Bottom are determined by Bayes' theorem, which is related to the Meso-Middle. At the Meso-Middle, the three-way entropies lack complete monotonicity and clear systematicness; thus, Bayes' theorem evolves via the weighted information function $wp(.)\log\, p(.)$ and integrated sum operation Σ, and thus, three-way weighted entropies emerge to gain benign monotonicity and perfect systematicness. Finally, three-way weighted entropies of all D-Classes are integratedly fused by the sum operation, and thus, three-way weighted entropies at Macro-Top naturally emerge to acquire their monotonicity and systematicness. Herein, three-layer granular structures and three-way informational measures are concluded in Table 3.

Table 3 is utilized to review some of the basic results. (1) The Micro-Bottom concerns C-Class and D-Class to produce the three-way probabilities. They partly have granulation monotonicity and follow Bayes' theorem, where $p(X_j)$ is a constant. (2) The Meso-Middle concerns C-Classification and D-Class to construct three-way weighted entropies. They have granulation monotonicity and evolution systematicness. The latter hierarchically develops Bayes' theorem, and the constant $-p(X_j)\log\, p(X_j)$ adjusts $H_W(X_j/\mathcal{A})$ to $H_W^{\mathrm{lin}}(X_j/\mathcal{A})$. (3) Macro-Top includes C-Classification and D-Classification to establish further three-way weighted entropies. They also exhibit granulation monotonicity and evolution systematicness. The latter deeply develops Bayes' theorem, and the constant $H(\mathcal{D})$ transforms $H_W(\mathcal{D}/\mathcal{A})$ to $H_W^{\mathrm{lin}}(\mathcal{D}/\mathcal{A})$.

Based on the above reviews, we next emphasize the hierarchical evolution of three-way informational measures. From the GrC viewpoint, the development process depends on D-Table's three-layer structures, and Fig. 5, which is relevant, is first provided.

According to the left half of Fig. 5, three-way informational measures go through the Meso-Middle to exhibit two steps of integrated promotion among the three-layer structures. (1) The Micro-Bottom evolves to the Meso-Middle by integrating C-Classes to C-Classification. Accordingly, three-way probabilities are integratedly fused into three-way weighted entropies by the weighted entropy function $\sum_{i=1}^{n} wp(.)\log\ p(.)$, and the relevant monotonicity and systematicness are determined by developing Bayes' theorem. (2) The Meso-Middle evolves to the Macro-Top by integrating D-Classes to D-Classification. Accordingly, three-way weighted entropies at the Meso-Middle are integratedly fused into three-way weighted entropies at Macro-Top by the sum operation $\sum_{j=1}^{m}$, and the relevant monotonicity and systematicness are naturally gained by the internal integration. The above hierarchical evolution adopts the bottom-middle-top construction technology, which is a basic GrC strategy, to implement the integrated fusion of three-way informational measures. In contrast, the top-middle-bottom strategy can conversely implement decomposed extraction, which is reflected by the downward virtual arrows in Fig. 5. On the whole, the hierarchical evolution of the three-way informational measures corresponds to a type of bidirectional information construction.

5.2. Hierarchical superiority of three-way informational measures

According to Fig. 5, information measures adhere to three-layer structures to have two types of hierarchical evolution. Both types go through different middle structures with regard to the Meso-Middle and Structure IV, and they are called the weight-entropy evolution and information-theoretic evolution, respectively, according to their top measures. This subsection mainly makes their parallel comparison to reveal hierarchical superiority of the weight-entropy evolution, which is related to three-way informational measures.

The weight-entropy evolution is analyzed above according to the left half of Fig. 5. According to the right half, the information-theoretic evolution also has two steps of integrated promotion, which are realized by Structure IV. (1) The Micro-Bottom evolves to Structure IV by integrating D-Classes to D-Classification. Accordingly, the probabilities are integratedly fused into entropies by the entropy function $H = \sum_{j=1}^{m} p(.)\log\ p(.)$, but granulation monotonicity is never involved. (2) Structure IV evolves to Macro-Top by integrating C-Classes to C-Classification. Accordingly, the entropies at Structure IV are integratedly fused into three-way weighted entropies at Macro-Top by the weighted sum operation $\sum_{i=1}^{n} wH$, and the relevant monotonicity and systematicness are finally established. The above analyses mainly depend on the basic information-theoretic forms in Definition 1. In fact, the proof of Theorem 8 implies some transformations, such as

$$H(\mathcal{D}/\mathcal{A}) = -\sum_{i=1}^{n}\left[p([x]_{\mathcal{A}}^{i})\sum_{j=1}^{m} p(X_j/[x]_{\mathcal{A}}^{i})\log\ p(X_j/[x]_{\mathcal{A}}^{i})\right] = -\sum_{i=1}^{n}\left[\sum_{j=1}^{m} p([x]_{\mathcal{A}}^{i})p(X_j/[x]_{\mathcal{A}}^{i})\log\ p(X_j/[x]_{\mathcal{A}}^{i})\right],$$

and thus, a type of transformative weighted entropies can be properly mined at Structure IV to change the above operation. Except for the bottom-middle-top integrated fusion, the top-middle-bottom decomposed extraction also exists to supplement the bidirectional information construction.

Next, the weight-entropy evolution and information-theoretic evolution are compared, mainly in the integrated fusion direction. (1) For the granular integration, the former takes effect first from C-Classes to C-Classification and then from D-Classes to D-Classification, while the latter adopts the inverse integration order. (2) For information fusion, the former uses the weighted entropy first and the sum second, while the latter uses the entropy first and the weighted sum second. (3) For granulation monotonicity, the former has monotonicity at the Meso-Middle to further simplify the monotonicity at the Macro-Top, while the latter never involves monotonicity at Structure IV and, thus, the monotonicity at Macro-Top becomes complex. (4) For systematicness, both have the basic Bayes' theorem, and thus, they can establish systemic equations at the three granular levels. (5) For the results at Macro-Top, both produce two equivalent systems: the weight-entropy system and information-theoretic system, which have equivalent measures and systematicness (Fig. 4). The equivalence transformation relies on the changing order of the sum. As a result, the sum-order exchange and decomposition extraction can be utilized for the information-theoretic measures to produce three-way weighted entropies at the Meso-Middle. Moreover, three-way entropies can be calculated from the Meso-Middle to the Macro-Top, and thus, they have a closer hierarchical mechanism, because the information-theoretic measures are usually directly calculated to become more complex.

The weight-entropy evolution and information-theoretic evolution depend on the Meso-Middle and Structure IV, respectively, and thus, they become symmetrical and parallel but have different emphases. For the former, Bayes' theorem at the Micro-Bottom (Eq. (13)) focuses on rule reasoning, and three-way entropies exhibit granulation monotonicity at both the Meso-Middle and Macro-Top. For the latter, Bayes' theorem at the Micro-Bottom (Eq. (14)) focuses on practical determination, and its granulation monotonicity is applied to only Macro-Top. Both have similar granular and hierarchical mechanisms, but the former has more superiority in view of the attribute reduction. Only the weight-entropy evolution is useful for class-specific attribute reduction, which exists at the Meso-Middle [43]. Both can be equivalently used for classification-based attribute reduction, which exists at Macro-Top, but the former becomes more thorough and simple for granulation monotonicity. In other words, the former (with three-way informational measures) effectively underlies hierarchical attribute reduction based on D-Table's three-layer structures.

Algorithm 1 An algorithm of three-way probabilities.

Input: C-Class $[x]^i_{\mathcal{A}}$ and D-Class X_j, as well as $|U|$;
Output: Three-way probabilities $p(X_j/[x]^i_{\mathcal{A}})$, $p([x]^i_{\mathcal{A}})$, $p([x]^i_{\mathcal{A}}/X_j)$.
Compute $|[x]^i_{\mathcal{A}}|$, $|X_j|$, $|[x]^i_{\mathcal{A}} \cap X_j|$.
According to Eq. (15), compute
$p(X_j/[x]^i_{\mathcal{A}}) = \frac{|[x]^i_{\mathcal{A}} \cap X_j|}{|[x]^i_{\mathcal{A}}|}$, $p([x]^i_{\mathcal{A}}) = \frac{|[x]^i_{\mathcal{A}}|}{|U|}$, $p([x]^i_{\mathcal{A}}/X_j) = \frac{|[x]^i_{\mathcal{A}} \cap X_j|}{|X_j|}$.
return $p(X_j/[x]^i_{\mathcal{A}})$, $p([x]^i_{\mathcal{A}})$, $p([x]^i_{\mathcal{A}}/X_j)$.

Algorithm 2 A probability-based algorithm of the three-way weighted entropies at the Meso-Middle.

Input: C-Classification $U/IND(\mathcal{A}) = \{[x]^i_{\mathcal{A}} : i = 1, .., n\}$ and D-Class X_j, as well as $|U|$;
Output: Three-way weighted entropies at the Meso-Middle: $H_W(X_j/\mathcal{A})$ and $H^{\text{lin}}_W(X_j/\mathcal{A})$, $H^{X_j}_W(\mathcal{A})$, $H_W(\mathcal{A}/X_j)$.
Compute $|X_j|$ and $p(X_j)$.
$H_W(X_j/\mathcal{A}) = 0$, $H^{X_j}_W(\mathcal{A}) = 0$, $H_W(\mathcal{A}/X_j) = 0$.
for $i \in \{1, .., n\}$ **do**
  Compute $p(X_j/[x]^i_{\mathcal{A}})$, $p([x]^i_{\mathcal{A}})$, $p([x]^i_{\mathcal{A}}/X_j)$ by Algorithm 1.
  According to Eq. (23), let
  $H_W(X_j/\mathcal{A}) \leftarrow H_W(X_j/\mathcal{A}) - p([x]^i_{\mathcal{A}})p(X_j/[x]^i_{\mathcal{A}})\log p(X_j/[x]^i_{\mathcal{A}})$,
  $H^{X_j}_W(\mathcal{A}) \leftarrow H^{X_j}_W(\mathcal{A}) - p(X_j/[x]^i_{\mathcal{A}})p([x]^i_{\mathcal{A}})\log p([x]^i_{\mathcal{A}})$,
  $H_W(\mathcal{A}/X_j) \leftarrow H_W(\mathcal{A}/X_j) - p(X_j)p([x]^i_{\mathcal{A}}/X_j)\log p([x]^i_{\mathcal{A}}/X_j)$.
end for
According to Eq. (27), compute $H^{\text{lin}}_W(X_j/\mathcal{A}) = -p(X_j)\log p(X_j) - H_W(X_j/\mathcal{A})$.
return $H_W(X_j/\mathcal{A})$ and $H^{\text{lin}}_W(X_j/\mathcal{A})$, $H^{X_j}_W(\mathcal{A})$, $H_W(\mathcal{A}/X_j)$.

Algorithm 3 An algorithm of three-way weighted entropies from the Meso-Middle to the Macro-Top.

Input: C-Classification $U/IND(\mathcal{A})$ and D-Classification $U/IND(\mathcal{D}) = \{X_j : j = 1, .., m\}$;
Output: Three-way weighted entropies at Macro-Top: $H_W(\mathcal{D}/\mathcal{A})$ and $H^{\text{lin}}_W(\mathcal{D}/\mathcal{A})$, $H^{\mathcal{D}}_W(\mathcal{A})$, $H_W(\mathcal{A}/\mathcal{D})$.
$H_W(\mathcal{D}/\mathcal{A}) = 0$ and $H^{\text{lin}}_W(\mathcal{D}/\mathcal{A}) = 0$, $H^{\mathcal{D}}_W(\mathcal{A}) = 0$, $H_W(\mathcal{A}/\mathcal{D}) = 0$.
for $j \in \{1, .., m\}$ **do**
  Compute $H_W(X_j/\mathcal{A})$ and $H^{\text{lin}}_W(X_j/\mathcal{A})$, $H^{X_j}_W(\mathcal{A})$, $H_W(\mathcal{A}/X_j)$ by Algorithm 2.
  According to Eq. (30), let
  $H_W(\mathcal{D}/\mathcal{A}) \leftarrow H_W(\mathcal{D}/\mathcal{A}) + H_W(X_j/\mathcal{A})$,
  $H^{\text{lin}}_W(\mathcal{D}/\mathcal{A}) \leftarrow H^{\text{lin}}_W(\mathcal{D}/\mathcal{A}) + H^{\text{lin}}_W(X_j/\mathcal{A})$,
  $H^{\mathcal{D}}_W(\mathcal{A}) \leftarrow H^{\mathcal{D}}_W(\mathcal{A}) + H^{X_j}_W(\mathcal{A})$,
  $H_W(\mathcal{A}/\mathcal{D}) \leftarrow H_W(\mathcal{A}/\mathcal{D}) + H_W(\mathcal{A}/X_j)$.
end for
return $H_W(\mathcal{D}/\mathcal{A})$ and $H^{\text{lin}}_W(\mathcal{D}/\mathcal{A})$, $H^{\mathcal{D}}_W(\mathcal{A})$, $H_W(\mathcal{A}/\mathcal{D})$.

5.3. Hierarchical algorithms of three-way informational measures

In view of the above hierarchical evolution and superiority, this subsection develops hierarchical algorithms of three-way informational measures.

Algorithm 1 utilizes the division operation in Definition 3 to yield three-way probabilities at the Micro-Bottom, where the cardinality determination becomes a basis. Algorithm 2 utilizes the weight-entropy function in Definition 4 to yield three-way weighted entropies at the Meso-Middle. There, the constant probability $p(X_j)$ at the Micro-Bottom is directly calculated. In the "for" loop with regard to the C-Classes, three-way probabilities are computed by invoking Algorithm 1, and three-way entropies implement the summation for weighted information item $wp(.)\log\ p(.)$. Finally, $H^{\text{lin}}_W(X_j/\mathcal{A})$ is achieved by its linear transformation definition with regard to $H_W(X_j/\mathcal{A})$, i.e., Definition 5. Algorithm 3 utilizes the sum operation in Definition 6 to yield three-way weighted entropies at Macro-Top. In the "for" loop with regard to the D-Classes, three-way entropies at the Meso-Middle are computed by invoking Algorithm 2, and they are further integrated into three-way entropies at Macro-Top.

Algorithms 1–3 closely follow the bottom-middle-top evolution of three-way informational measures to exhibit a strong hierarchical relation, and the algorithm at a lower level underlies the algorithm at a higher level. Simply speaking, the

Table 4
A decision table.

U	a	b	c	d
x_1	3	3	3	1
x_2	1	1	1	1
x_3	1	1	1	1
x_4	2	2	2	1
x_5	4	4	3	2
x_6	1	1	1	2
x_7	1	3	1	2
x_8	2	2	2	2
x_9	5	4	3	3
x_{10}	1	3	1	3
x_{11}	1	1	1	3
x_{12}	2	2	2	3

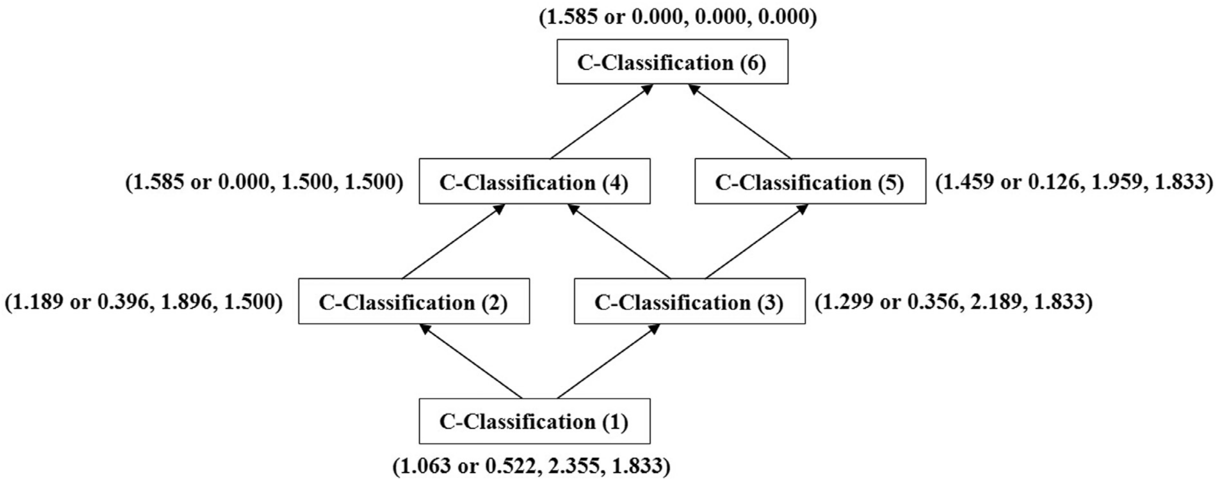

Fig. 6. Granulation hierarchies and three-way weighted entropies of C-Classifications.

three algorithms are applied to three different granular levels, and the former underlies the later by virtue of the invocation. Finally, the computational complexity is simply given according to the Micro-Bottom, which is located within the loop framework. Algorithm 1 concerns only one Micro-Bottom, and thus, its complexity can be viewed as $O(1)$. Algorithm 2 concerns one Meso-Middle or n Micro-Bottoms, and thus, its complexity becomes $O(n)$. Algorithm 3 concerns one Macro-Top or m Meso-Middles or $m \times n$ Micro-Bottoms, and thus, its complexity becomes $O(mn)$.

6. Example illustration based on a decision table

Aiming at the D-Table, this section illustrates three-layer granular structures and three-way informational measures by an example from [47]. The basic information of D-Table $(U, \mathcal{C} \cup \mathcal{D})$ is provided in Table 4, where $U = \{x_1, \ldots, x_{12}\}$, $\mathcal{C} = \{a, b, c\}$, $\mathcal{D} = \{d\}$; $U/IND(\mathcal{D}) = \{X_1, X_2, X_3\}$, $X_1 = \{x_1, \ldots, x_4\}$, $X_2 = \{x_5, \ldots, x_8\}$, and $X_3 = \{x_9, \ldots, x_{12}\}$.

Eight attribute subsets $\mathcal{A} \subseteq \mathcal{C}$ exist to produce only six types of C-Classifications:

$(1) U/IND(\{a, b, c\}) = U/IND(\{a, b\}) = \{\{x_2, x_3, x_6, x_{11}\}, \{x_4, x_8, x_{12}\}, \{x_1\}, \{x_5\}, \{x_7, x_{10}\}, \{x_9\}\}$,
$(2) U/IND(\{a, c\}) = U/IND(\{a\}) = \{\{x_2, x_3, x_6, x_7, x_{10}, x_{11}\}, \{x_4, x_8, x_{12}\}, \{x_1\}, \{x_5\}, \{x_9\}\}$,
$(3) U/IND(\{b, c\}) = \{\{x_2, x_3, x_6, x_{11}\}, \{x_4, x_8, x_{12}\}, \{x_1\}, \{x_5, x_9\}, \{x_7, x_{10}\}\}$,
$(4) U/IND(\{c\}) = \{\{x_2, x_3, x_6, x_7, x_{10}, x_{11}\}, \{x_4, x_8, x_{12}\}, \{x_1, x_5, x_9\}\}$,
$(5) U/IND(\{b\}) = \{\{x_2, x_3, x_6, x_{11}\}, \{x_4, x_8, x_{12}\}, \{x_1, x_7, x_{10}\}, \{x_5, x_9\}\}$,
$(6) U/IND(\emptyset) = \{U\}$.

The relevant granulation hierarchies are described in a Hasse diagram: Fig. 6, where the arrow shows knowledge roughening $U/IND(.) \xrightarrow{\preceq} U/IND(..)$.

Herein, C-Classification $U/IND(\mathcal{C})$ and its six C-Classes are utilized to illustrate three-layer structures and three-way measures, where $n = 6$, $m = 3$ and X_1 acts as a representative D-Class.

Table 5
Three-way weighted entropies for hierarchical C-Classification.

C-Classification: $U/IND(\mathcal{A})$	X_1-based three-way weighted entropies: $(H_W(X_1/\mathcal{A})$ or $H_W^{\text{lin}}(X_1/\mathcal{A}), H_W^{X_1}(\mathcal{A}), H_W(\mathcal{A}/X_1))$	$U/IND(\mathcal{D})$-based three-way weighted entropies: $(H_W(\mathcal{D}/\mathcal{A})$ or $H_W^{\text{lin}}(\mathcal{D}/\mathcal{A}), H_W^{\mathcal{D}}(\mathcal{A}), H_W(\mathcal{A}/\mathcal{D}))$
(1)	(0.298 or 0.230, 0.730, 0.500)	(1.063 or 0.522, 2.355, 1.833)
(2)	(0.396 or 0.132, 0.632, 0.500)	(1.189 or 0.396, 1.896, 1.500)
(3)	(0.298 or 0.230, 0.730, 0.500)	(1.229 or 0.356, 2.189, 1.833)
(4)	(0.528 or 0.000, 0.500, 0.500)	(1.585 or 0.000, 1.500, 1.500)
(5)	(0.431 or 0.097, 0.597, 0.500)	(1.459 or 0.126, 1.959, 1.833)
(6)	(0.528 or 0.000, 0.000, 0.000)	(1.585 or 0.000, 0.000, 0.000)

(1) There are $6 \times 3 = 18$ Micro-Bottoms, and $([x]_{\mathcal{C}}^2 = \{x_4, x_8, x_{12}\}, X_1)$ is considered. The three-way probabilities are

$$p(X_1/[x]_{\mathcal{C}}^2) = 1/3, \ \ p([x]_{\mathcal{C}}^2) = 3/12, \ \ p([x]_{\mathcal{C}}^2/X_1) = 1/4,$$

and Bayes' theorem (i.e., Eq. (13)) becomes

$$p([x]_{\mathcal{C}}^2/X_1) = \frac{p([x]_{\mathcal{C}}^2) \times p(X_1/[x]_{\mathcal{C}}^2)}{p(X_1)} = \frac{3/12 \times 1/3}{4/12} = 1/4,$$

where constant $p(X_1) = 4/12$.

(2) There are $m = 3$ Meso-Middles, and $(U/IND(\mathcal{C}), X_1)$ is considered. According to the probability integration of $n = 6$ C-Classes, the three-way weighted entropies at the Meso-Middle are

$$H_W(X_1/\mathcal{C}) = 0.298 \text{ or } H_W^{\text{lin}}(X_1/\mathcal{C}) = 0.230, \ \ H_W^{X_1}(\mathcal{C}) = 0.730, \ \ H_W(\mathcal{C}/X_1) = 0.500,$$

where the constant $-p(X_1)\log\, p(X_1) = 0.528$ is used to gain $H_W^{\text{lin}}(X_1/\mathcal{C})$ from $H_W(X_1/\mathcal{C})$, and the relevant systematic equation (i.e., Eq. (29)) becomes

$$H_W(\mathcal{C}/X_1) = H_W^{X_1}(\mathcal{C}) - H_W^{\text{lin}}(X_1/\mathcal{C}) = 0.730 - 0.230 = 0.500.$$

(3) There is only one Macro-Top $(U/IND(\mathcal{C}), U/IND(\mathcal{D}))$. According to the weight-entropy integration of $m = 3$ D-Classes, the three-way weighted entropies at Macro-Top are

$$H_W(\mathcal{D}/\mathcal{C}) = 1.063 \text{ or } H_W^{\text{lin}}(\mathcal{D}/\mathcal{C}) = 0.522, \ \ H_W^{\mathcal{D}}(\mathcal{C}) = 2.355, \ \ H_W(\mathcal{C}/\mathcal{D}) = 1.833,$$

where the constant $H(\mathcal{D}) = 1.585$ is used to gain $H_W^{\text{lin}}(\mathcal{D}/\mathcal{C})$ from $H_W(\mathcal{D}/\mathcal{C})$, and the relevant systematic equation (i.e., Eq. (33)) becomes

$$H_W(\mathcal{C}/\mathcal{D}) = H_W^{\mathcal{D}}(\mathcal{C}) - H_W^{\text{lin}}(\mathcal{D}/\mathcal{C}) = 2.355 - 0.522 = 1.833.$$

The above analyses partly illustrate three-layer structures and three-way informational measures (as well as relevant systematicness from Bayes' theorem), which are mainly summarized in Table 3. In particular, three-way information measures are calculated by basic definitions or relevant Algorithms 1–3, where the weight-entropy evolution (in Fig. 5) plays an important role.

For the six types of C-Classification, their three-way weighted entropies of X_1 and $U/IND(\mathcal{D})$ are presented by a three-dimensional vector in Table 5, and the final weighted entropies at Macro-Top are also marked in Fig. 6. Thus, we have utilized Table 5 and Fig. 6 to verify the granulation monotonicity of the three-way informational measures.

Fig. 6 and Table 5 are used to provide further illustration, mainly at Macro-Top. For the three-way weight-entropy vector (in Fig. 6 or in the third column of Table 5), the first element contains $H_W(\mathcal{D}/\mathcal{A})$ and $H_W^{\text{lin}}(\mathcal{D}/\mathcal{A})$, and their sum is equal to the constant $H(\mathcal{D}) = 1.585$. The third element $H_W(\mathcal{A}/\mathcal{D})$ is the difference between the second element $H_W^{\mathcal{D}}(\mathcal{A})$ and the first element $H_W^{\text{lin}}(\mathcal{D}/\mathcal{A})$, and this result reflects the systematic equation: Eq. (33). The measure change in the arrow direction in Fig. 6 effectively verifies the granulation monotonicity. In contrast to three-way weighted entropies, Definition 1 (which exhibits order exchange) is utilized to calculate the entropy, conditional entropy, and mutual information, and these informational measures yield the same results. Thus, the equivalence of the weight-entropy system and information-theoretic system (in Fig. 4) is verified. However, the hierarchical calculation of the three-way weighted entropies becomes ingenious in contrast to the direct but complex calculation of the information-theoretic measures.

7. Conclusions

D-Table is a basis for implementing attribute reduction. Aiming at D-Table and its attribute reduction, this paper utilizes the systematic granularities of the classification and class to establish three-layer granular structures (Fig. 2), and these structures underlie extensive hierarchical applications, including measure mining and reduction construction. Thus, this paper further constructs three-way informational measures at three-layer structures, including three-way probabilities at the Micro-Bottom and three-way weighted entropies at the Meso-Middle and Macro-Top. Three-way informational measures originate from Bayes' theorem (Fig. 3) and perform a thorough hierarchical evolution (Fig. 5) with monotonicity and systematicness (Table 3), and notably, they have hierarchical invocation algorithms (Algorithms 1–3).

Table A.6
Probability calculation and granulation monotonicity/non-monotonicity of three-way entropies.

D-Class	Type symbols	$[x]^1_{\mathcal{A}}$ probability	$[x]^2_{\mathcal{A}}$ probability	$[x]_{\mathcal{B}}$ probability	Entropy on $\mathcal{A}$	Entropy on $\mathcal{B}$	Entropy relationship
X_1	$p(X_1/[x]_.)$ or $H(X_1/.)$	0.025	0.700	0.3625	0.4932	0.5307	$H(X_1/\mathcal{B}) > H(X_1/\mathcal{A})$
X_1	$p([x]_.)$ or $H^{X_1}(.)$	0.500	0.500	1.0000	1.0000	0.0000	$H^{X_1}(\mathcal{B}) < H^{X_1}(\mathcal{A})$
X_1	$p([x]_./X_1)$ or $H(./X_1)$	1/29	28/29	1.0000	0.2164	0.0000	$H(\mathcal{B}/X_1) < H(\mathcal{A}/X_1)$
X_2	$p(X_2/[x]_.)$ or $H(X_2/.)$	0.975	0.300	0.6375	0.5567	0.4141	$H(X_2/\mathcal{B}) < H(X_2/\mathcal{A})$
X_2	$p([x]_.)$ or $H^{X_2}(.)$	0.500	0.500	1.0000	1.0000	0.0000	$H^{X_2}(\mathcal{B}) < H^{X_2}(\mathcal{A})$
X_2	$p([x]_./X_2)$ or $H(./X_2)$	39/51	12/51	1.0000	0.7871	0.0000	$H(\mathcal{B}/X_2) < H(\mathcal{A}/X_2)$

This study basically concerns GrC-based multi-granule, multi-level, and multi-view. The relevant work holds two fundamental contributions for informational measures, and the latter is especially related to attribute reduction.

(1) D-Table's three-layer granular structures – a novel viewpoint – is adopted to construct informational measures, which are related to attribute reduction. Two types of hierarchical evolution, which are based on the Meso-Middle and Structure IV (Fig. 5), thoroughly establish the structural mechanisms and systemic relationships for the novel weight-entropy system and the existing information-theoretic system. These two systems exhibit theoretical equivalence (Fig. 4), and both have practical emphases (Fig. 5), and thus, the former's results obtained in this paper deepen and explain the latter, which directly refers to information theory. In particular, the monotonicity evolution based on the Meso-Middle can simplify the final monotonicity at Macro-Top, which was previously attached to information-theoretic measures (Theorem 1).

(2) Based on the basic comparison, three-way informational measures focus more on C-Classification monotonicity and D-Class specificity to have the hierarchical evolution be superior, especially from the attribute reduction perspective. As a result, this study emphasizes granulation monotonicity and systematic equations at each level to underlie further attribute reduction, including informational, systematic, and hierarchical reduction. For example, the three-way attribute reduction at Macro-Top in [47] can be deepened, and the informational attribute reduction at the Meso-Middle can be defined as in Definition 2 to link class-specific attribute reduction, and the three-layer attribute reduction in Table 2 is worthwhile to explore by introducing the category reduction transformation.
Thus, our whole research regarding the three-layer granular structures and three-way informational measures enrich the three-way decisions theory, especially in its generalized sense [19].

In summary, this paper focuses on D-Table, reduction, and information to conduct an appropriate study of rough set theory. Three-layer granular structures and three-way informational measures are worthwhile to investigate further for data analysis, including the extensive construction of uncertainty measures and the comprehensive promotion of attribute reduction. Our research team is devoting its focus to follow-up work on reduction from the informational, systematic, and hierarchical perspectives. Moreover, the concerned structure and measure mainly aim at the classical rough sets with equivalence relations, and the relationships between informational measures and extended rough sets (including neighborhood rough sets, fuzzy rough sets, dominance rough sets, set-valued rough sets) are extensively discussed [2,3,7,10–12]. According to these extended models, the relevant work of this study is worthwhile to generalize deeply for future studies and practical applications.

Acknowledgments

The authors thank all of the editors and reviewers for their valuable suggestions, which have substantially improved this paper.

This work was supported by National Natural Science Foundation of China (61673285, 61673301, 61203285, 61273304), Sichuan Youth Science and Technology Foundation of China (2017JQ0046), and Scientific Research Project of Sichuan Provincial Education Department of China (15ZB0028).

Appendix A. Example illustration for granulation monotonicity/non-monotonicity of three-way entropies

Suppose that C-Classification $U/IND(\mathcal{A})$ has two C-Classes $[x]^1_{\mathcal{A}}$ and $[x]^2_{\mathcal{A}}$, while D-Classification $U/IND(\mathcal{D})$ has two D-Classes X_1 and X_2. The relevant cardinality information is given as follows: (1) $|[x]^1_{\mathcal{A}}| = 40 = |[x]^2_{\mathcal{A}}|$; (2) $|X_1| = 29,\ |X_2| = 51$; and (3) $|[x]^1_{\mathcal{A}} \cap X_1| = 1,\ |[x]^1_{\mathcal{A}} \cap X_2| = 39,\ |[x]^2_{\mathcal{A}} \cap X_1| = 28,\ |[x]^2_{\mathcal{A}} \cap X_2| = 12$.

Granular merging $[x]^1_{\mathcal{A}} \cup [x]^2_{\mathcal{A}} \xrightarrow{=} [x]_{\mathcal{B}}$ produces knowledge coarsening $U/IND(\mathcal{A}) \xrightarrow{\preceq} U/IND(\mathcal{B})$, where $U/IND(\mathcal{A}) = \{[x]^1_{\mathcal{A}}, [x]^2_{\mathcal{A}}\}$ and $U/IND(\mathcal{B}) = \{[x]_{\mathcal{B}}\}$. According to Eqs. (18)–(20), three-way entropies based on X_1 and X_2 are calculated in Table A.6.

For three-way entropies, the final entropy relationship in Table A.6 verifies their granulation monotonicity/non-monotonicity, especially the non-monotonicity that is based on $H(X_1/\mathcal{B}) > H(X_1/\mathcal{A})$ and $H(X_2/\mathcal{B}) < H(X_2/\mathcal{A})$.

Appendix B. Proof of Theorem 4: granulation monotonicity proof of three-way weighted entropies at the Meso-Middle

Proof. Because knowledge granulation consists of granular merging, the informational monotonicity of granular merging can be integrated to informational monotonicity in knowledge granulation. Therefore, we are only required to prove the informational monotonicity for a representative group of granular merging, i.e.,

$$\bigcup_{t=1}^{k}[x]_{\mathcal{A}}^{t} \xrightarrow{=} [x]_{\mathcal{B}}.$$

(1) $f(u) = -u\log u$ ($u \in [0, 1]$) is a concave function, and thus,

$$\sum_{t=1}^{k}\lambda_t = 1 \Longrightarrow -\sum_{t=1}^{k}\lambda_t p_t \log\ p_t \leq -\left[\sum_{t=1}^{k}\lambda_t p_t\right]\log\left[\sum_{t=1}^{k}\lambda_t p_t\right]. \tag{B.1}$$

This concave property can effectively deduce the key inequality relation in the following transformation:

$$\begin{aligned}
&-\sum_{t=1}^{k}p([x]_{\mathcal{A}}^{t})p(X_j/[x]_{\mathcal{A}}^{t})\log\ p(X_j/[x]_{\mathcal{A}}^{t})\\
&=-\sum_{t=1}^{k}p([x]_{\mathcal{B}})\frac{|[x]_{\mathcal{A}}^{t}|}{|[x]_{\mathcal{B}}|}p(X_j/[x]_{\mathcal{A}}^{t})\log\ p(X_j/[x]_{\mathcal{A}}^{t})\\
&=p([x]_{\mathcal{B}})\left[-\sum_{t=1}^{k}\frac{|[x]_{\mathcal{A}}^{t}|}{|[x]_{\mathcal{B}}|}p(X_j/[x]_{\mathcal{A}}^{t})\log\ p(X_j/[x]_{\mathcal{A}}^{t})\right]\\
&\leq -p([x]_{\mathcal{B}})\left[\sum_{t=1}^{k}\frac{|[x]_{\mathcal{A}}^{t}|}{|[x]_{\mathcal{B}}|}p(X_j/[x]_{\mathcal{A}}^{t})\right]\log\ \frac{\sum_{t=1}^{k}|[x]_{\mathcal{A}}^{t}\cap X_j|}{|[x]_{\mathcal{B}}|}\\
&=-p([x]_{\mathcal{B}})\frac{|[x]_{\mathcal{B}}\cap X_j|}{|[x]_{\mathcal{B}}|}\log\ \frac{|[x]_{\mathcal{B}}\cap X_j|}{|[x]_{\mathcal{B}}|}\\
&=-p([x]_{\mathcal{B}})p(X_j/[x]_{\mathcal{B}})\log\ p(X_j/[x]_{\mathcal{B}}).
\end{aligned} \tag{B.2}$$

(2) Note that

$$\begin{aligned}
H_W^{X_j}(\mathcal{A}) &= -\sum_{i=1}^{n}p(X_j/[x]_{\mathcal{A}}^{i})p([x]_{\mathcal{A}}^{i})\log\ p([x]_{\mathcal{A}}^{i})\\
&= -\sum_{i=1}^{n}p([x]_{\mathcal{A}}^{i}\cap X_j)\log\ p([x]_{\mathcal{A}}^{i}).
\end{aligned} \tag{B.3}$$

As a result,

$$\begin{aligned}
&-\sum_{t=1}^{k}p([x]_{\mathcal{A}}^{t}\cap X_j)\log\ p([x]_{\mathcal{A}}^{t})\\
&=-p([x]_{\mathcal{A}}^{1}\cap X_j)\log\ p([x]_{\mathcal{A}}^{1})-\ldots-p([x]_{\mathcal{A}}^{k}\cap X_j)\log\ p([x]_{\mathcal{A}}^{k})\\
&\geq -p([x]_{\mathcal{A}}^{1}\cap X_j)\log\ p([x]_{\mathcal{B}})-\ldots-p([x]_{\mathcal{A}}^{k}\cap X_j)\log\ p([x]_{\mathcal{B}})\\
&=-\sum_{t=1}^{k}p([x]_{\mathcal{A}}^{t}\cap X_j)\log\ p([x]_{\mathcal{B}})\\
&=-\left[\sum_{t=1}^{k}p([x]_{\mathcal{A}}^{t}\cap X_j)\right]\log\ p([x]_{\mathcal{B}})\\
&=-p([x]_{\mathcal{B}}\cap X_j)\log\ p([x]_{\mathcal{B}}).
\end{aligned} \tag{B.4}$$

(3) $H_W(\mathcal{A}/X_j) = -\sum_{i=1}^{n}p([x]_{\mathcal{A}}^{i}\cap X_j)\log\ p([x]_{\mathcal{A}}^{i}/X_j)$, so $H_W(\mathcal{A}/X_j) \geq H_W(\mathcal{B}/X_j)$ can be proved by a similar process with regard to Eq. (B.4). □

References

[1] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.
[2] Y.M. Chen, Y. Xue, Y. Ma, F.F. Xu, Measures of uncertainty for neighborhood rough sets, Knowl. Based Syst. 120 (2017) 226–235.
[3] J.H. Dai, H.W. Tian, Entropy measures and granularity measures for set-valued information systems, Inf. Sci. 240 (2013) 72–82.
[4] W.S. Du, B.Q. Hu, Attribute reduction in ordered decision tables via evidence theory, Inf. Sci. (364–365) (2016) 91–110.
[5] B.J. Fan, E.C.C. Tsang, W.H. Xu, J.H. Yu, Double-quantitative rough fuzzy set based decisions: a logical operations method, Inf. Sci. 378 (2017) 264–281.
[6] B.W. Fang, B.Q. Hu, Probabilistic graded rough set and double relative quantitative decision-theoretic rough set, Int. J. Approximate Reasoning 74 (2016) 1–12.
[7] A. Ganivada, S.S. Ray, S.K. Pal, Fuzzy rough granular self-organizing map and fuzzy rough entropy, Theor. Comput. Sci. 466 (2012) 37–63.
[8] S. Greco, Z. Pawlak, R. Slowinski, Can Bayesian confirmation measures be useful for rough set decision rules? Eng. Appl. Artif. Intell. 17 (4) (2004) 345–361.
[9] B.Q. Hu, Three-way decisions based on semi-three-way decision spaces, Inf. Sci. 382 (2017) 415–440.
[10] Q.H. Hu, D.R. Yu, Z.X. Xie, J.F. Liu, Fuzzy probabilistic approximation spaces and their information measures, IEEE Trans. Fuzzy Syst. 14 (2) (2006) 191–201.
[11] Q.H. Hu, L. Zhang, D.G. Chen, W. Pedrycz, D.R. Yu, Gaussian kernel based fuzzy rough sets: model, uncertainty measures and applications, Int. J. Approximate Reasoning 51 (4) (2010) 453–471.
[12] Q.H. Hu, L. Zhang, D. Zhang, W. Pan, S. An, W. Pedrycz, Measuring relevance between discrete and continuous features based on neighborhood mutual information, Expert Syst. Appl. 38 (9) (2011) 10737–10750.
[13] X.Y. Jia, L. Shang, B. Zhou, Y.Y. Yao, Generalized attribute reduct in rough set theory, Knowl. Based Syst. 91 (2016) 204–218.
[14] F. Jiang, Y.F. Sui, L. Zhou, A relative decision entropy-based feature selection approach, Pattern Recognit. 48 (7) (2015) 2151–2163.
[15] F.F. Jin, L.D. Pei, H.Y. Chen, L.G. Zhou, Interval-valued intuitionistic fuzzy continuous weighted entropy and its application to multi-criteria fuzzy group decision making, Knowl. Based Syst. 59 (2014) 132–141.
[16] J.H. Li, C.C. Huang, J.J. Qi, Y.H. Qian, W.Q. Liu, Three-way cognitive concept learning via multi-granularity, Inf. Sci. 378 (2017) 244–263.
[17] J.Y. Liang, Z.Z. Shi, D.Y. Li, M.J. Wierman, Information entropy, rough entropy and knowledge granularity in incomplete information systems, Int. J. Gen. Syst. 35 (6) (2006) 641–654.
[18] J.Y. Liang, J.H. Wang, Y.H. Qian, A new measure of uncertainty based on knowledge granulation for rough sets, Inf. Sci. 179 (4) (2009) 458–470.
[19] D. Liu, D.C. Liang, Generalized Three-way Decisions and Special Three-way decisions, Sciencepaper Online, Beijing, 2016. [http://www.paper.edu.cn/releasepaper/content/201606-581], (in Chinese).
[20] 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.
[21] D.Q. Miao, Rough Set Theory and Its Application in Machine Learing, Institute of Automation, The Chinese Academy of Sciences, Beijing, 1997 Ph. d. thesis. (in Chinese)
[22] D.Q. Miao, J. Wang, An information representation of the concepts and operations in rough set theory, J. Software 10 (2) (1999) 113–116. (in Chinese)
[23] 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.
[24] O. Parkash, P.K. Sharma, R. Mahajan, New measures of weighted fuzzy entropy and their applications for the study of maximum weighted fuzzy entropy principle, Inf. Sci. 178 (11) (2008) 2389–2395.
[25] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, 9, Kluwer Academic Publishers, Dordrecht, 1991.
[26] Z. Pawlak, Rough sets, decision algorithms and Bayes' theorem, Eur. J. Oper. Res. 136 (1) (2002) 181–189.
[27] Y.H. Qian, H. Zhang, Y.L. Sang, J.Y. Liang, Multigranulation decision-theoretic rough sets, Int. J. Approximate Reasoning 55 (1) (2014) 225–237.
[28] A. Skowron, J. Stepaniuk, R. Swiniarski, Modeling rough granular computing based on approximation spaces, Inf. Sci. 184 (1) (2012) 20–43.
[29] D. Slezak, Approximate entropy reducts, Fundam. Inform. 53 (2002) 365–390.
[30] D. Slezak, W. Ziarko, The investigation of the Bayesian rough set model, Int. J. Approximate Reasoning 40 (1–2) (2005) 81–91.
[31] S.H. Teng, M. Lu, A.F. Yang, J. Zhang, Y.J. Nian, M. He, Efficient attribute reduction from the viewpoint of discernibility, Inf. Sci. 326 (2016) 297–314.
[32] G.Y. Wang, J. Zhao, J.J. An, Y. Wu, A comparative study of algebra viewpoint and information viewpoint in attribute reduction, Fundam. Inform. 68 (3) (2005) 289–301.
[33] G.Y. Wang, X.A. Ma, H. Yu, Monotonic uncertainty measures for attribute reduction in probabilistic rough set model, Int. J. Approximate Reasoning 59 (2015) 41–67.
[34] W. Wei, J.Y. Liang, Y.H. Qian, C.Y. Dang, Can fuzzy entropies be effective measures for evaluating the roughness of a rough set? Inf. Sci. 232 (2013) 143–166.
[35] W. Wei, J.H. Wang, J.Y. Liang, X. Mi, C.Y. Dang, Compacted decision tables based attribute reduction, Knowl. Based Syst. 86 (2015) 261–277.
[36] W.Z. Wu, Y.H. Qian, T.J. Li, S.M. Gu, On rule acquisition in incomplete multi-scale decision tables, Inf. Sci. 378 (2017) 282–302.
[37] W.H. Xu, Y.T. Guo, Generalized multigranulation double-quantitative decision-theoretic rough set, Knowl. Based Syst. 105 (2016) 190–205.
[38] Y.Y. Yao, Y. Zhao, Discernibility matrix simplification for constructing attribute reducts, Inf. Sci. 179 (7) (2009) 867–882.
[39] Y.Y. Yao, R. Fu, The concept of reducts in Pawlak three-step rough set analysis, LNCS Trans. Rough Sets XVI, LNCS 7736 (2013) 53–72.
[40] Y.Y. Yao, Three-way decisions and cognitive computing, Cognit. Comput. 8 (4) (2016) 543–554.
[41] Y.Y. Yao, B. Zhou, Two Bayesian approaches to rough sets, Eur. J. Oper. Res. 251 (3) (2016) 904–917.
[42] Y.Y. Yao, A triarchic theory of granular computing, Granular Comput. 1 (2) (2016) 145–157.
[43] Y.Y. Yao, X.Y. Zhang, Class-specific attribute reducts in rough set theory, Inf. Sci. (2017). (in Press).
[44] H.Y. Zhang, S.Y. Yang, Feature selection and approximate reasoning of large-scale set-valued decision tables based on α-dominance-based quantitative rough sets, Inf. Sci. 378 (2017) 328–347.
[45] X.H. Zhang, D.Q. Miao, C.H. Liu, M.L. Le, Constructive methods of rough approximation operators and multigranulation rough sets, Knowl. Based Syst. 91 (2016) 114–125.
[46] X.Y. Zhang, D.Q. Miao, Quantitative information architecture, granular computing and rough set models in the double-quantitative approximation space on precision and grade, Inf. Sci. 268 (2014) 147–168.
[47] X.Y. Zhang, D.Q. Miao, Three-way weighted entropies and three-way attribute reduction, in: D.Q. Miao, W. Pedrycz, D. Slezak, G. Peters, Q.H. Hu, R.Z. Wang (Eds.), Proceedings of the 9th International RSKT Conference, LNCS (LNAI), 8818, 2014b, pp. 707–719.
[48] X.Y. Zhang, D.Q. Miao, Double-quantitative fusion of accuracy and importance: systematic measure mining, benign integration construction, hierarchical attribute reduction, Knowl. Based Syst. 91 (2016) 219–240.