

Article

Double-Granule Conditional-Entropies Based on Three-Level Granular Structures

Taopin Mu ^{1,2}, Xianyong Zhang ^{1,2,*}  and Zhiwen Mo ^{1,2}

¹ School of Mathematical Sciences, Sichuan Normal University, Chengdu 610066, China

² Institute of Intelligent Information and Quantum Information, Sichuan Normal University, Chengdu 610066, China

* Correspondence: xianyongzh@sina.com.cn

Received: 30 April 2019; Accepted: 27 June 2019; Published: 3 July 2019



Abstract: Rough set theory is an important approach for data mining, and it refers to Shannon's information measures for uncertainty measurements. The existing local conditional-entropies have both the second-order feature and application limitation. By improvements of hierarchical granulation, this paper establishes double-granule conditional-entropies based on three-level granular structures (i.e., micro-bottom, meso-middle, macro-top), and then investigates the relevant properties. In terms of the decision table and its decision classification, double-granule conditional-entropies are proposed at micro-bottom by the dual condition-granule system. By virtue of successive granular summation integrations, they hierarchically evolve to meso-middle and macro-top, to respectively have part and complete condition-granulations. Then, the new measures acquire their number distribution, calculation algorithm, three bounds, and granulation non-monotonicity at three corresponding levels. Finally, the hierarchical constructions and achieved properties are effectively verified by decision table examples and data set experiments. Double-granule conditional-entropies carry the second-order characteristic and hierarchical granulation to deepen both the classical entropy system and local conditional-entropies, and thus they become novel uncertainty measures for information processing and knowledge reasoning.

Keywords: rough set theory; information theory; conditional entropy; uncertainty; granular computing; three-level granular structures

1. Introduction

Rough set theory can effectively implement data mining for the imprecise, inconsistent, and incomplete information [1], and it has been extensively applied in artificial intelligence and machine learning [2–8]. In rough set theory, attribute reduction based on decision tables is a main topic for approximate reasoning and knowledge discovery, and there are three main construction strategies: from the positive region, information measure, and a discernibility matrix [9–15]. By virtue of the discernibility matrix, Wei et al. [16] proposed an incremental reduction algorithm for dynamic data; Ma et al. [17] utilized the compressed binary discernibility matrix to construct an incremental reduction algorithm for group dynamic data; moreover, Nie and Zhou [18] proposed a new discernibility matrix defined by local conditional-entropies to compute the reduction core.

Information theory originated from Shannon's entropy system [19], and it provides an effective method for uncertainty measurement, such as in attribute reduction. Currently, information theory has been introduced into rough set theory for uncertainty analyses and information processing [20–25]. As far as attribute reduction is concerned, Miao [26] offered the informational representation of knowledge reduction and decision reduction, where entropy and mutual-information are highlighted; Wang et al. [27] conducted a comparative study on attribute reduction from the algebra and

information viewpoints, where the conditional-entropy acts as a main tool; Jiang et al. [28] presented the relative decision entropy to propose a feature selection algorithm; Slezak [29] used the conditional-entropy to define approximate reducts; moreover, Qian and Shu [30] provided the mutual information criterion to evaluate candidate features in incomplete data. In general, the entropy, conditional-entropy, and mutual-information together constitute the classical information system with integrality and comprehensiveness, and they can function on rough set applications (such as attribute reduction) but may exhibit different emphases in different application scenarios. In addition, information-theoretic measures have multiple variational forms [31–35]. As far as conditional-entropies are concerned, they are extensively applied in rough set theory from multiple pointcuts [26,27,29,31,34,36–39], while uncertainty measurement and reduction construction still serve as two basic issues. Aiming at probabilistic rough sets, Deng and Yao [40,41] used Shannon's entropy and conditional-entropy to interpret and determine probabilistic thresholds by an information-theoretic approach, and Ma et al. [42] considered variants of conditional-entropies to construct heuristic reduction algorithms for the probabilistic model. In particular, local conditional-entropies are put forward by adopting double condition-granules and their union locality [18], and they can distinctively determine a new discernibility matrix for reduction core computation; moreover, the information measures exhibit a novel feature of second-order expressions, especially when compared to the traditional entropy system with only single-granule descriptions [19,26,27].

Granular computing is a structural methodology of hierarchical computing and information processing [43,44], and its technology of multi-granularity and multiple levels is useful for uncertainty analyses and knowledge acquisition regarding data. In rough set theory, the information granulation is of extensive concern [45–49], and the granulation monotonicity plays an important role in attribute reduction [12,50–52]. In particular, a decision table acts as a formal background of data mining [12,53–55], and it involves condition/decision granules and classifications from granular structures. According to granular computing, Zhang and Miao [56] introduced three-layer granular structures of decision tables, and they further hierarchically constructed three-way informational measures based on weighted-entropies; moreover, Wang et al. [57] utilized three-layer granular structures to research three-way weighted combination-entropies. These studies adhere to three-level analyses, and the latter are directly related to granular computing [43] and three-way decisions [58], as well as their interplay. Recently, Yao [59] discussed three-way granular computing by making use of two particular types of three granules and three levels, where thinking in three levels results in an important model. Additionally, three-level analyses were extensively utilized in the location allocation and programming/optimization modeling [60–62].

According to [18], the new discernibility matrix is used for reduction core calculations, and its creative implementation mainly depends on local conditional-entropies. Therefore, local conditional-entropies focus on the granule-union locality rather than their underlying double-granule interaction, and the latter more essentially adheres to the second-order characteristic; moreover, they lack the condition granulation to restrict their uncertainty measurement function and information procession prospect based on knowledge. Motivated by the two issues, this paper utilizes the two-granular essence and three-hierarchical evolution to propose double-granule conditional-entropies based on three-level granular structures. Regarding the contribution, this novel type of information measures improves local conditional-entropies from both the granular interaction and hierarchical/conditional granulation, and they will achieve multiple important properties (including the integration hierarchy, number distribution, calculation algorithm, three bounds, and granulation non-monotonicity) to offer both robust measurement functions and knowledge-application prospects. Moreover, three-level granular structures here (including micro-bottom, meso-middle, macro-top) adopt only the condition part of decision table, and thus they differ from and push forward the previous ones, which include both the condition and decision parts [56].

The remainder of this paper is organized as follows. Section 2 reviews the decision table and local conditional-entropies; Section 3 proposes and studies double-granule conditional-entropies from

three-level granular structures; Section 4 provides a decision table example for mechanism illustration; Section 5 makes data experiments for effectiveness verification; finally, Section 6 concludes this paper.

2. Decision Table and Its Existing Entropy Measures

Rough set theory [1] focuses on the data that are represented in an information table

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

U is the universe with finite objects, AT is the finite attribute set, V_a is the value domain for $a \in AT$, and $I_a : U \rightarrow V_a$ is an information function to endow each object x with a value $I_a(x) = a(x)$ on attribute a . The decision table is a special type of information table with $AT = C \cup D$ and $C \cap D = \emptyset$, where C and D denote the sets of condition attribute and decision attribute, respectively, and it is simply denoted by $(U, C \cup D)$ in this paper. Furthermore, the granulation construction usually considers two parts.

- (1) The condition attribute subset $A \subseteq C$ induces an equivalence relation

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

and the latter provides the condition granulation or partition $U/IND(A) = \{A_i : i = 1, \dots, n\}$, where $A_i = [x]_A^i$ represents the equivalence granule to exhibit number $|U/IND(A)| = n$.

- (2) Similarly, the decision attribute set D induces the equivalence relation $IND(D)$ and further decision classification $U/IND(D) = \{D_j : j = 1, \dots, m\}$, which consists of $|U/IND(D)| = m$ decision classes.

The decision table $(U, C \cup D)$ and its granulation from $A \subseteq C$ and D constitute the basic background for information measure construction. The probability space $(U, 2^U, P)$ establishes the usual probability framework, where

$$P : 2^U \rightarrow \mathbf{Q}, P(X) = \frac{|X|}{|U|}, \forall X \subseteq U, \tag{1}$$

and thus two usual probabilities are

$$P(A_i) = \frac{|A_i|}{|U|}, P(D_j/A_i) = \frac{|A_i \cap D_j|}{|A_i|}. \tag{2}$$

Definition 1 ([26,27,56]). *The entropy on condition A , conditional-entropy on D given A , and mutual-information between A and D are respectively defined by*

$$\begin{aligned} H(A) &= - \sum_{i=1}^n P(A_i) \log_2 P(A_i), \\ H(D/A) &= - \sum_{i=1}^n \left(P(A_i) \sum_{j=1}^m P(D_j/A_i) \log_2 P(D_j/A_i) \right), \\ I(A;D) &= H(D) - H(D/A), \end{aligned} \tag{3}$$

where

$$H(D) = - \sum_{j=1}^m P(D_j) \log_2 P(D_j).$$

Theorem 1 ([26,27,56]). *The entropy, conditional-entropy, and mutual-information have granulation monotonicity. Concretely,*

$$U/IND(A) \succeq U/IND(B) \implies H(B) \geq H(A), H(D/B) \leq H(D/A), I(B;D) \geq I(A;D). \tag{4}$$

In terms of the decision table $(U, C \cup D)$, the classical system of Shannon entropies has been introduced into rough set theory, as shown by Definition 1 and Theorem 1. As three basic information measures, the entropy, conditional-entropy, and mutual-information have uncertainty semantics and granulation monotonicity, so they are extensively used in attribute reduction and heuristic algorithms [26,27,42]. The granulation relation $U/IND(A) \succeq U/IND(B)$ is equivalent to $IND(A) \supseteq IND(B)$, that is,

$$\forall B_{i^*} \in U/IND(B), \exists A_i \in U/IND(A), \text{ s.t., } B_{i^*} \subseteq A_i,$$

and it is usually induced by $A \subseteq B \subseteq C$; furthermore, relevant granulation monotonicity/non-monotonicity becomes an important index to assess and apply uncertainty measures.

According to the decision table and its formal structure, Zhang and Miao [56] recently introduced three-level granular structures, i.e.,

micro-bottom (A_i, D_j) , meso-middle $(U/IND(A), D_j)$, macro-top $(U/IND(A), U/IND(D))$,

and further investigated weighted-entropy constructions. As a result, the previous entropy system (Equation (3)) is actually located at macro-top and has an equivalent construction from the weighted-entropy system; at meso-middle, Zhang et al. [10] established three-way informational class-specific reducts to be compared with the algebraic class-specific reducts [9].

In particular, Nie and Zhou [18] proposed a new discernibility matrix for computing the reduction core, and they tactfully utilized a kind of novel information of so-called local conditional-entropy. As our preliminary, the relevant entropy and matrix are reviewed as follows, where let $U/IND(C) = \{C_k : k = 1, \dots, r\}$ and the cardinality form is mainly adopted.

Definition 2 ([18]). The local conditional-entropy on decision table $(U, C \cup D)$ is defined by:

$$\begin{aligned} &\forall C_p, C_q \in U/IND(C) \ (1 \leq p, q \leq r), \\ &H_{C_p \cup C_q}(D/C) = - \frac{|C_p|}{|C_p \cup C_q|} \sum_{j=1}^m \frac{|C_p \cap D_j|}{|C_p|} \log_2 \frac{|C_p \cap D_j|}{|C_p|} \\ &\quad - \frac{|C_q|}{|C_p \cup C_q|} \sum_{j=1}^m \frac{|C_q \cap D_j|}{|C_q|} \log_2 \frac{|C_q \cap D_j|}{|C_q|}. \end{aligned} \tag{5}$$

Definition 3 ([18]). The discernibility matrix $DM = (r_{i'j'})_{|U| \times |U|}$ on decision table $(U, C \cup D)$ is defined by:

$$r_{i'j'} = \begin{cases} c \in C, \text{ if } \min(|dx_{i'}|, |dx_{j'}|) = 1, c(x_{i'}) \neq c(x_{j'}), D(x_{i'}) \neq D(x_{j'}), \\ c \in C, \text{ if } \min(|dx_{i'}|, |dx_{j'}|) > 1, c(x_{i'}) \neq c(x_{j'}), D(x_{i'}) \neq D(x_{j'}), \\ \quad H_{[x_{i'}]_C \cup [x_{j'}]_C}(D/(C - \{c\})) > H_{[x_{i'}]_C \cup [x_{j'}]_C}(D/C), \\ \emptyset, \text{ otherwise} \end{cases} \tag{6}$$

where $dx = \{D(y) : y \in [x]_C\}$ ($\forall x, y \in U$) represents the set of decision values induced by conditional class $[x]_C$ while $|dx|$ means the corresponding cardinality [63]. In Equation (6), let $[x_{i'}]_C = C_p$, $[x_{j'}]_C = C_q$, and then

$$H_{[x_{i'}]_C \cup [x_{j'}]_C}(D/(C - \{c\})) = H_{C_p \cup C_q}(D/(C - \{c\})) = - \sum_{j=1}^m \frac{|(C_p \cup C_q) \cap D_j|}{|C_p \cup C_q|} \log_2 \frac{|(C_p \cup C_q) \cap D_j|}{|C_p \cup C_q|} \tag{7}$$

is determined to represent the conditional-entropy of local decision table when accompanied by new universe $C_p \cup C_q$ after deleting attribute c ; moreover, $H_{[x_{i'}]_C \cup [x_{j'}]_C}(D/C) = H_{C_p \cup C_q}(D/C)$ is clear according to Equation (5).

3. Double-Granule Conditional-Entropies Based on Three-Level Granular Structures

The local conditional-entropy in Equation (5) implements effective uncertainty descriptions to guide the in-depth discernibility matrix and core calculation [18], thus exhibiting fundamental significance. However, this basic measure has three flawed aspects, and corresponding improvements for general applications.

- (1) According to Equation (5), the locality mainly refers to less range $C_p \cup C_q$ in universe U . More essentially, we can stand on the dual granules C_p and C_q to propose a novel notion of double-granule conditional-entropies, and it differs from the usual entropy system with only the single-granule representation which implies a kind of first-order style. Moreover, the measure properties are lacking in [18], and we will provide in-depth properties such as restriction bounds and granulation non-monotonicity.
- (2) Regarding granular structures, all decision classes D_j ($j = 1, \dots, m$) (or decision classification $U/IND(D)$) are considered, but condition granules involve only two factors C_p and C_q . A condition partition $U/IND(C)$ needs considering in practice to provide a system description of knowledge granulation, so we also focus on granulation $U/IND(C)$ to introduce three-level granular structures for hierarchical constructions of double-granule conditional-entropies.
- (3) Finally, the initial concept is limited to only C for expressing the discernibility matrix and reduction core, and a general subset $A \subseteq C$ has better theoretical and practical prospects, especially for the knowledge-based applications (such as attribute reduction or feature selection).

Along the above thoughts, this section mainly establishes double-granule conditional-entropies based on a universal attribute-subset $A \subseteq C$ and investigates relevant algorithms and properties, and we particularly use a kind of three-level granular structures.

From a viewpoint of only condition granulation, basic descriptions of three-level granular structures are provided in Table 1, and relevant concepts are usually intuitionistic and descriptive according to a supporting figure with granular structures: Figure 1. Micro-bottom (A_p, A_q) focuses on only two granules, meso-middle

$$(A_p, U/IND(A) = \{A_q : q = 1, \dots, n\})$$

consists of one granule and a partition, while macro-top

$$(U/IND(A) = \{A_p : p = 1, \dots, n\}, U/IND(A) = \{A_q : q = 1, \dots, n\})$$

considers the same partition with different construction origins. The three-level granular structures carry a kind of hierarchical integration (or decomposition) relationship, and they provide $n \times n$, n , and one parallel patterns, respectively; they will be presented in a table form with the $n \times n$ mainbody data as well as the edge statistics. Moreover, they differ from the existing three-level granular structures for decision tables, which consider not only the condition granulation (with A_i and $U/IND(A)$) but also decision granulation (with D_j and $U/IND(D)$) [56].

Table 1. Three-level granular structures based on condition granulation of the decision table.

Structure Naming	Composition System	Granular Scale	Granular Level	Number of Parallel Patterns
Micro-Bottom	(A_p, A_q)	Micro	Bottom	$n \times n$
Meso-Middle	$(A_p, U/IND(A))$ $= (A_p, \{A_q : q = 1, \dots, n\})$	Meso	Middle	n
Macro-Top	$(U/IND(A), U/IND(A))$ $= (\{A_p : p = 1, \dots, n\}, \{A_q : q = 1, \dots, n\})$	Macro	Top	1

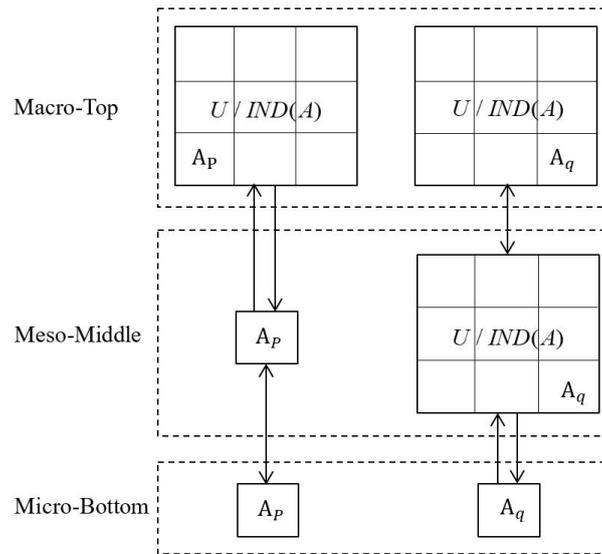


Figure 1. Schematic diagram of three-level granular structures.

3.1. Double-Granule Conditional-Entropy at Micro-Bottom

The local conditional-entropies are actually at only micro-bottom, i.e., (C_p, C_q) regarding C . As a basis of hierarchical development, this subsection improves local conditional-entropies to construct double-granule conditional-entropies at micro-bottom (A_p, A_q) $(p, q \in \{1, \dots, n\})$, which comes from an arbitrary condition-attribute subset $A \subseteq C$. We first suppose weight coefficients

$$\omega_p = \frac{|A_p|}{|A_p| + |A_q|}, \omega_q = \frac{|A_q|}{|A_p| + |A_q|}, \tag{8}$$

where

$$\omega_p + \omega_q = 1.$$

Definition 4. At micro-bottom (A_p, A_q) , the double-granule conditional-entropy is defined by

$$\begin{aligned} H_{(A_p, A_q)}(D/A) &= -\omega_p \sum_{j=1}^m P(D_j/A_p) \log_2 P(D_j/A_p) - \omega_q \sum_{j=1}^m P(D_j/A_q) \log_2 P(D_j/A_q) \\ &= -\frac{|A_p|}{|A_p| + |A_q|} \sum_{j=1}^m \frac{|A_p \cap D_j|}{|A_p|} \log_2 \frac{|A_p \cap D_j|}{|A_p|} - \frac{|A_q|}{|A_p| + |A_q|} \sum_{j=1}^m \frac{|A_q \cap D_j|}{|A_q|} \log_2 \frac{|A_q \cap D_j|}{|A_q|}. \end{aligned} \tag{9}$$

Proposition 1. The double-granule conditional-entropy based on A_p becomes

$$\begin{aligned} H_{(A_p, A_p)}(D/A) &= -\sum_{j=1}^m P(D_j/A_p) \log_2 P(D_j/A_p) \\ &= -\sum_{j=1}^m \frac{|A_p \cap D_j|}{|A_p|} \log_2 \frac{|A_p \cap D_j|}{|A_p|}. \end{aligned} \tag{10}$$

By using probabilistic and cardinal forms, Definition 4 proposes the double-granule conditional-entropy at micro-bottom. In contrast to the local conditional-entropy in [18], our measure generally adopts the same essence but a different viewpoint. In other words, Equation (9) with forms (A_p, A_p) and $|A_q| + |A_p|$ is equivalent to Equation (5) with styles $A_p \cup A_p$ and $|A_q \cup A_p|$ when

$$A_q \neq A_p \implies |A_q| + |A_p| = |A_q \cup A_p|,$$

but the former becomes different and coherent when

$$A_q = A_p \implies |A_q| + |A_p| = 2|A_q \cup A_p| > |A_q \cup A_p|;$$

moreover, it more tends to the double-granule description rather than the granule-union locality. In Equation (9), conditional-information measures

$$-\sum_{j=1}^m P(D_j/A_p) \log_2 P(D_j/A_p), -\sum_{j=1}^m P(D_j/A_q) \log_2 P(D_j/A_q)$$

represent the uncertainty of decision classification $U/IND(D)$ regarding condition granules A_p and A_q , respectively, and they are integrated into $H_{(A_p,A_q)}(D/A)$ by two complementary weight coefficients ω_p and ω_q . As a result, $H_{(A_p,A_q)}(D/A)$ embodies a kind of information fusion of double-granule A_p, A_q to describe decision classification $U/IND(D)$ and its uncertainty, from the perspective of conditional information. Therefore, $H_{(A_p,A_q)}(D/A)$ is naturally called the double-granule conditional-entropy, and it is actually located at micro-bottom (A_p, A_q) . In particular, the double-granule measures utilize the double-granule fusion to capture a new feature of second-order, because main entropy systems (such as those in Equation (3)) utilize only the single-granule description which correspondingly refers to the so-called first-order information. Proposition 1 focuses on a specific case of $A_q = A_p$, and the concrete result $H_{(A_p,A_p)}(D/A)$ degenerates into a one-order measure regarding conditional-entropy.

Proposition 2. At micro-bottom, double-granule conditional-entropies offer $n \times n$ values, i.e.,

$$H_{(A_p,A_q)}(D/A) \quad (p, q = 1, \dots, n).$$

Since both A_p and A_q have n granules based on $p = 1, \dots, n$ and $q = 1, \dots, n$, $H_{(A_p,A_q)}(D/A)$ offers number $n \times n$ (Proposition 2) to correspond to $n \times n$ micro-bottoms. The $n \times n$ kinds of double-granule conditional-entropies are arranged in Table 2, and the mainbody refers to an $n \times n$ square symmetric matrix where

$$H_{(A_p,A_q)}(D/A) = H_{(A_q,A_p)}(D/A).$$

Based on Equation (9), Algorithm 1 resorts to a “for” loop to effectively offer a double-granule conditional-entropy $H_{(A_p,A_q)}(D/A)$ for two arbitrary granules $A_p, A_q \in U/IND(A)$. Furthermore, we can achieve all $n \times n$ entropies values by adding two “for” loops regarding $p = 1, \dots, n$ and $q = 1, \dots, n$.

Table 2. Matrix distribution of double-granule conditional-entropies at micro-bottom.

$U/IND(A)$	A_1	\dots	A_q	\dots	A_n
A_1	$H_{(A_1,A_1)}(D/A)$	\dots	$H_{(A_1,A_q)}(D/A)$	\dots	$H_{(A_1,A_n)}(D/A)$
\vdots	\vdots	\ddots	\vdots	\ddots	\vdots
A_p	$H_{(A_p,A_1)}(D/A)$	\dots	$H_{(A_p,A_q)}(D/A)$	\dots	$H_{(A_p,A_n)}(D/A)$
\vdots	\vdots	\ddots	\vdots	\ddots	\vdots
A_n	$H_{(A_n,A_1)}(D/A)$	\dots	$H_{(A_n,A_q)}(D/A)$	\dots	$H_{(A_n,A_n)}(D/A)$

Algorithm 1: Calculation of double-granule conditional-entropy at micro-bottom

Input: Decision table $(U, C \cup D)$, target subset $A \subseteq C$, and two special indexes

$$p, q \in \{1, \dots, n\};$$

Output: Double-granule conditional-entropy $H_{(A_p, A_q)}(D/A)$ at micro-bottom (A_p, A_q) .

- 1: Compute $U/IND(A)$ to obtain two concrete granules $A_p, A_q \in U/IND(A)$, and determine ω_p, ω_q .
 - 2: Compute $U/IND(D)$ to obtain all decision classes $D_j (j = 1, \dots, m)$.
 - 3: Let $H_p = 0, H_q = 0$.
 - 4: **for** $j \in \{1, \dots, m\}$ **do**
 - 5: $H_p \leftarrow H_p - P(D_j/A_p)\log_2 P(D_j/A_p)$,
 $H_q \leftarrow H_q - P(D_j/A_q)\log_2 P(D_j/A_q)$.
 - 6: **end for**
 - 7: Obtain $H_{(A_p, A_q)}(D/A) = \omega_p H_p + \omega_q H_q$.
 - 8: **return** $H_{(A_p, A_q)}(D/A)$.
-

Theorem 2. At micro-bottom, the double-granule conditional-entropy has lower and upper bounds. Concretely,

$$\underline{H}_{(A_p, A_q)}(D/A) \leq H_{(A_p, A_q)}(D/A) \leq \bar{H}_{(A_p, A_q)}(D/A),$$

where

$$\begin{aligned} \underline{H}_{(A_p, A_q)}(D/A) &= -\frac{|A_p|}{2|U|} \sum_{j=1}^m P(D_j/A_p)\log_2 P(D_j/A_p) - \frac{|A_q|}{2|U|} \sum_{j=1}^m P(D_j/A_q)\log_2 P(D_j/A_q), \\ \bar{H}_{(A_p, A_q)}(D/A) &= -\sum_{j=1}^m P(D_j/A_p)\log_2 P(D_j/A_p) - \sum_{j=1}^m P(D_j/A_q)\log_2 P(D_j/A_q). \end{aligned} \tag{11}$$

Proof. $|A_p|, |A_q| \in [1, |U|)$ implies

$$\begin{aligned} \omega_p &= \frac{|A_p|}{|A_p| + |A_q|} \in \left[\frac{|A_p|}{2|U|}, 1 \right), \\ \omega_q &= \frac{|A_q|}{|A_p| + |A_q|} \in \left[\frac{|A_q|}{2|U|}, 1 \right), \end{aligned} \tag{12}$$

so $H_{(A_p, A_q)}(D/A) \in [\underline{H}_{(A_p, A_q)}(D/A), \bar{H}_{(A_p, A_q)}(D/A)]$. \square

In Theorem 2, the double bounds of $H_{(A_p, A_q)}(D/A)$ are acquired by the enlarging and reducing of weight coefficients. Regarding Equation (12),

$$(A_q \neq A_p) \vee \left(A_q = A_p \wedge |A_q| = |A_p| \leq \frac{|U|}{2} \right) \implies \omega_p \geq \frac{|A_p|}{|U|} > \frac{|A_p|}{2|U|}, \omega_q \geq \frac{|A_q|}{|U|} > \frac{|A_q|}{2|U|}; \tag{13}$$

on the other hand,

$$(A_q = A_p) \wedge \left(|A_q| = |A_p| > \frac{|U|}{2} \right) \implies \omega_p = \frac{1}{2} \in \left[\frac{|A_p|}{2|U|}, \frac{|A_p|}{|U|} \right), \omega_q = \frac{1}{2} \in \left[\frac{|A_q|}{2|U|}, \frac{|A_q|}{|U|} \right). \tag{14}$$

In other words, ω_p and ω_q have theoretical lower bounds $\frac{|A_p|}{2|U|}$ and $\frac{|A_q|}{2|U|}$, respectively, but they usually have closer lower bounds $\frac{|A_p|}{|U|}$ and $\frac{|A_q|}{|U|}$, respectively. Therefore, $H_{(A_p,A_q)}(D/A)$ can theoretically achieve $\underline{H}_{(A_p,A_q)}(D/A)$, such as in the case

$$A_p = A_q = U \implies \left(\omega_p = \frac{1}{2} = \frac{|A_p|}{2|U|} \right) \wedge \left(\omega_q = \frac{1}{2} = \frac{|A_q|}{2|U|} \right);$$

usually, it may be practically restricted by a better measure:

$$\underline{H}'_{(A_p,A_q)}(D/A) = -\frac{|A_p|}{|U|} \sum_{j=1}^m P(D_j/A_p) \log_2 P(D_j/A_p) - \frac{|A_q|}{|U|} \sum_{j=1}^m P(D_j/A_q) \log_2 P(D_j/A_q), \quad (15)$$

which offers

$$\underline{H}'_{(A_p,A_q)}(D/A) \geq \underline{H}_{(A_p,A_q)}(D/A). \quad (16)$$

We below provide another upper bound of $H_{(A_p,A_q)}(D/A)$, which may be better than $\overline{H}_{(A_p,A_q)}(D/A)$ in some cases.

Theorem 3. *At micro-bottom, the double-granule conditional-entropy has an upper bound. Concretely,*

$$\begin{aligned} H_{(A_p,A_q)}(D/A) &\leq H^*_{(A_p,A_q)}(D/A) \\ &= -\sum_{j=1}^m \frac{|A_p \cap D_j| + |A_q \cap D_j|}{|A_p| + |A_q|} \log_2 \frac{|A_p \cap D_j| + |A_q \cap D_j|}{|A_p| + |A_q|}. \end{aligned} \quad (17)$$

Proof. As shown in Figure 2, function $f(P) = -P \log_2 P$ ($P \in [0,1]$) is convex, where $f''(P) = -\frac{1}{P \ln 2} < 0$. Thus, let

$$P_p = P(D_j/A_p) = \frac{|A_p \cap D_j|}{|A_p|}, P_q = P(D_j/A_q) = \frac{|A_q \cap D_j|}{|A_q|},$$

and then the famous ‘‘Jensen’s inequality’’ in mathematics could induce

$$\omega_p + \omega_q = 1 \implies -\omega_p P_p \log_2 P_p - \omega_q P_q \log_2 P_q \leq -(\omega_p P_q + \omega_p P_p) \log_2 (\omega_p P_q + \omega_p P_p),$$

where

$$\omega_p P_q + \omega_p P_p = \frac{|A_p|}{|A_p| + |A_q|} \frac{|A_p \cap D_j|}{|A_p|} + \frac{|A_q|}{|A_p| + |A_q|} \frac{|A_q \cap D_j|}{|A_q|} = \frac{|A_p \cap D_j| + |A_q \cap D_j|}{|A_p| + |A_q|}.$$

In other words, we can get

$$\begin{aligned} H_{(A_p,A_q)}(D/A) &= \sum_{j=1}^m [-\omega_p P_p \log_2 P_p - \omega_q P_q \log_2 P_q] \\ &\leq \sum_{j=1}^m -(\omega_p P_q + \omega_p P_p) \log_2 (\omega_p P_q + \omega_p P_p) \\ &= -\sum_{j=1}^m \frac{|A_p \cap D_j| + |A_q \cap D_j|}{|A_p| + |A_q|} \log_2 \frac{|A_p \cap D_j| + |A_q \cap D_j|}{|A_p| + |A_q|} \\ &= H^*_{(A_p,A_q)}(D/A). \end{aligned}$$

□

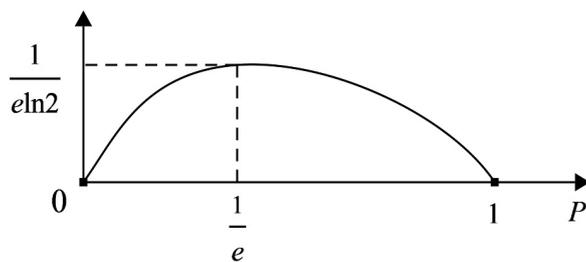


Figure 2. Convex figure of information function $f(P) = -P\log_2 P$.

In Theorem 3, the convex property of information function $f(P) = -P\log_2 P$ is utilized to provide a new upper bound $H_{(A_p,A_q)}^*(D/A)$ of central measure $H_{(A_p,A_q)}(D/A)$. When comparing Equations (7) and (17), we can surprisingly discover that $H_{(A_p,A_q)}^*(D/A)$ highly adheres to

$$H_{A_p \cup A_q}(D/(A - \{a\})) = - \sum_{j=1}^m \frac{|(A_p \cup A_q) \cap D_j|}{|A_p \cup A_q|} \log_2 \frac{|(A_p \cup A_q) \cap D_j|}{|A_p \cup A_q|}, \tag{18}$$

which naturally comes from $H_{C_p \cup C_q}(D/(C - \{c\}))$ (Equation (7)). In fact,

$$H_{(A_p,A_q)}^*(D/A) = H_{A_p \cup A_q}(D/(A - \{a\})) \tag{19}$$

when $A_p \neq A_q$; when

$$A_q = A_p \implies \frac{|A_p \cap D_j| + |A_q \cap D_j|}{|A_p| + |A_q|} = \frac{|A_p \cap D_j|}{|A_p|} \geq \frac{|A_p \cap D_j|}{2|A_p|} = \frac{|(A_p \cup A_q) \cap D_j|}{|A_p| + |A_q|},$$

where $A_p \cup A_q = A_p$, there is a difference between two measures, and we obtain

$$H_{(A_p,A_p)}^*(D/A) = - \sum_{j=1}^m \frac{|A_p \cap D_j|}{|A_p|} \log_2 \frac{|A_p \cap D_j|}{|A_p|} \neq - \sum_{j=1}^m \frac{|A_p \cap D_j|}{2|A_p|} \log_2 \frac{|A_p \cap D_j|}{2|A_p|} = H_{A_p \cup A_q}(D/(A - \{a\})). \tag{20}$$

Thus far, $H_{(A_p,A_q)}(D/A)$ has one lower bound $\underline{H}_{(A_p,A_q)}(D/A)$ and two upper bounds $\overline{H}_{(A_p,A_q)}(D/A)$, $H_{(A_p,A_q)}^*(D/A)$. An interesting question naturally emerges, i.e., can we necessarily determine the size relationship between $\overline{H}_{(A_p,A_q)}(D/A)$ and $H_{(A_p,A_q)}^*(D/A)$ to provide an exact bound? Unfortunately, the answer is negative, and the later example and experiment will reveal the size uncertainty. We simply provide a mechanism analysis. Let

$$P_{pq} = \frac{|A_p \cap D_j| + |A_q \cap D_j|}{|A_p| + |A_q|},$$

and its numerator/denominator be the corresponding sum of numerators/denominators of P_p and P_q . According to [64], we can obtain

$$P_{pq} \in [\min(P_p, P_q), \max(P_p, P_q)]$$

but P_{pq} produces an uncertainty location between P_p and P_q . In view of the information function $f(P) = -P\log_2 P$ and its maximum point $(\frac{1}{e}, \frac{1}{e \ln 2})$ (Figure 2),

$$f(P_p) + f(P_q), f(P_{pq})$$

never having the necessary size relationships, so

$$\overline{H}_{(A_p, A_q)}(D/A) = \sum_{j=1}^m (f(P_p) + f(P_q)), H_{(A_p, A_q)}^*(D/A) = \sum_{j=1}^m f(P_{pq})$$

also never have the necessary size relationships. In summary, $\overline{H}_{(A_p, A_q)}(D/A)$ and $H_{(A_p, A_q)}^*(D/A)$ adopt different views to become irrelevant and interactive, and they together restrict $H_{(A_p, A_q)}(D/A)$. With the addition of lower bound of $\underline{H}_{(A_p, A_q)}(D/A)$, there are in total three bounds to systematically emerge. Similar to $H_{(A_p, A_q)}(D/A)$ and its distributional Table 2, they can also be arranged in a table with an $n \times n$ square symmetric matrix, i.e., Table 3, and thus Table 3 correspondingly restricts Table 2.

Finally, consider relevant granulation monotonicity/non-monotonicity. In fact, micro-bottom and its double-granule conditional-entropies focus on only two condition granules and thus never consider the condition granulation and further monotonicity/non-monotonicity. Moreover, $U/IND(A) \succeq U/IND(B)$ implies the granulation refining and granule decomposition from A to B ; thus $A_p, A_q \in U/IND(A)$ and $B_{p^*}, B_{q^*} \in U/IND(B)$ exhibit complex correspondence and uncertainty change, so we cannot mine fine relationships between $H_{(A_p, A_q)}(D/A)$ and $H_{(B_{p^*}, B_{q^*})}(D/B)$.

Table 3. Three bounds of double-granule conditional-entropies at micro-bottom.

$U/IND(A)$	A_1	...	A_q	...	A_n
A_1	$[\underline{H}_{(A_1,A_1)}(D/A), \overline{H}_{(A_1,A_1)}(D/A)]$ $H_{(A_1,A_1)}^*(D/A)$...	$[\underline{H}_{(A_1,A_q)}(D/A), \overline{H}_{(A_1,A_q)}(D/A)]$ $H_{(A_1,A_q)}^*(D/A)$...	$[\underline{H}_{(A_1,A_n)}(D/A), \overline{H}_{(A_1,A_n)}(D/A)]$ $H_{(A_1,A_n)}^*(D/A)$
\vdots	\vdots	\ddots	\vdots	\ddots	\vdots
A_p	$[\underline{H}_{(A_p,A_1)}(D/A), \overline{H}_{(A_p,A_1)}(D/A)]$ $H_{(A_p,A_1)}^*(D/A)$...	$[\underline{H}_{(A_p,A_q)}(D/A), \overline{H}_{(A_p,A_q)}(D/A)]$ $H_{(A_p,A_q)}^*(D/A)$...	$[\underline{H}_{(A_p,A_n)}(D/A), \overline{H}_{(A_p,A_n)}(D/A)]$ $H_{(A_p,A_n)}^*(D/A)$
\vdots	\vdots	\ddots	\vdots	\ddots	\vdots
A_n	$[\underline{H}_{(A_n,A_1)}(D/A), \overline{H}_{(A_n,A_1)}(D/A)]$ $H_{(A_n,A_1)}^*(D/A)$...	$[\underline{H}_{(A_n,A_q)}(D/A), \overline{H}_{(A_n,A_q)}(D/A)]$ $H_{(A_n,A_q)}^*(D/A)$...	$[\underline{H}_{(A_n,A_n)}(D/A), \overline{H}_{(A_n,A_n)}(D/A)]$ $H_{(A_n,A_n)}^*(D/A)$

3.2. Double-Granule Conditional-Entropy at Meso-Middle

As analyzed above, double-granule conditional-entropies at micro-bottom never consider the condition granulation to lack robust functions of uncertainty descriptions. In terms of fixed decision granulation $U/IND(D)$, $H_{(A_p, A_q)}(D/A)$ at micro-bottom (A_p, A_q) involves only two condition granules A_p, A_q and their interactive uncertainty information. For the function promotion, the condition granulation $U/IND(A)$ with systematic granules is worth introducing based on double-granule conditional-entropy $H_{(A_p, A_q)}(D/A)$. Thus, we will gradually strengthen the knowledge granulation $U/IND(A)$ to establish better double-granule conditional-entropies, by virtue of three-level granular structures (Table 1). This subsection discusses double-granule conditional-entropies at meso-middle

$$(A_p, U/IND(A) = \{A_1, \dots, A_n\}) \quad (p \in \{1, \dots, n\}).$$

Definition 5. At meso-middle $(A_p, U/IND(A))$, the double-granule conditional-entropy is defined by

$$\begin{aligned} H_{(A_p)}(D/A) &= - \sum_{q=1}^n \left(\omega_p \sum_{j=1}^m P(D_j/A_p) \log_2 P(D_j/A_p) \right) - \sum_{q=1}^n \left(\omega_q \sum_{j=1}^m P(D_j/A_q) \log_2 P(D_j/A_q) \right) \\ &= - \sum_{q=1}^n \left(\frac{|A_p|}{|A_p| + |A_q|} \sum_{j=1}^m \frac{|A_p \cap D_j|}{|A_p|} \log_2 \frac{|A_p \cap D_j|}{|A_p|} \right) - \sum_{q=1}^n \left(\frac{|A_q|}{|A_p| + |A_q|} \sum_{j=1}^m \frac{|A_q \cap D_j|}{|A_q|} \log_2 \frac{|A_q \cap D_j|}{|A_q|} \right). \end{aligned} \tag{21}$$

Corollary 1. At meso-middle, the double-granule conditional-entropy has an analytic expression:

$$H_{(A_p)}(D/A) = - \left(\sum_{q=1}^n \frac{|A_p|}{|A_p| + |A_q|} \right) \left(\sum_{j=1}^m P(D_j/A_p) \log_2 P(D_j/A_p) \right) - \sum_{q=1}^n \left(\frac{|A_q|}{|A_p| + |A_q|} \sum_{j=1}^m P(D_j/A_q) \log_2 P(D_j/A_q) \right). \tag{22}$$

Theorem 4. Double-granule conditional-entropies have a hierarchical integration from micro-bottom to meso-middle, i.e.,

$$H_{(A_p)}(D/A) = \sum_{q=1}^n H_{(A_p, A_q)}(D/A) = H_{(A_p, A_1)}(D/A) + \dots + H_{(A_p, A_q)}(D/A) + \dots + H_{(A_p, A_n)}(D/A). \tag{23}$$

By Definition 5 (Corollary 1) and Theorem 4, meso-middle’s measure $H_{(A_p)}(D/A)$ (which can also be noted by $H_{(A_p, U/IND(A))}(D/A)$) hierarchically integrates double-granule conditional-entropies $H_{(A_p, A_q)}(D/A)$ by condition-granular summation on $q = 1, \dots, n$. Thus, $H_{(A_p)}(D/A)$ inherits the features of double-granule and conditional-entropy, it considers a granule A_p and condition granulation $U/IND(A)$ to be at meso-middle $(A_p, U/IND(A))$, so it is called the double-granule conditional-entropy at meso-middle. As a transition, $H_{(A_p)}(D/A)$ combines granule A_p and partition $U/IND(A)$ to describe decision classification $U/IND(D)$ and its uncertainty, from the perspective of conditional information.

Similar to and based on previous discussions on $H_{(A_p, A_q)}(D/A)$ (Section 3.1), we will provide corresponding properties of $H_{(A_p)}(D/A)$, including the number distribution, calculation algorithm, three bounds, and granulation monotonicity/non-monotonicity.

Proposition 3. At meso-middle, double-granule conditional-entropies offer n values, i.e.,

$$H_{(A_p)}(D/A) \quad (p = 1, \dots, n).$$

In Proposition 3, double-granule conditional-entropies naturally exhibit number n to correspond to n meso-middles. The n values can be stored in an n -dimension vector to be related to the previous distributional Table 2. By enlarging Table 2, they are represented by the marginal vector of the bottom or right in Table 4, and they exactly correspond to the relevant row/column sum of micro-bottom’s information values. According to Equations (21) and (23), Algorithm 2 resorts

to two “for” loops to effectively offer a double-granule conditional-entropy $H_{(A_p)}(D/A)$ for an arbitrary granule $A_p \in U/IND(A)$. In fact, the inner loop invokes Algorithm 1 to calculate an arbitrary double-granule conditional-entropy at micro-bottom, while the outer loop integrates n related bottomed measures to produce $H_{(A_p)}(D/A)$. Furthermore, we can achieve all n middle entropies values by adding a “for” loop regarding $p = 1, \dots, n$.

Table 4. Marginal distribution of double-granule conditional-entropies at meso-middle and macro-top.

$U/IND(A)$	A_1	\dots	A_q	\dots	A_n	Meso-Middle
A_1	$H_{(A_1,A_1)}(D/A)$	\dots	$H_{(A_1,A_q)}(D/A)$	\dots	$H_{(A_1,A_n)}(D/A)$	$H_{(A_1)}(D/A)$
\vdots	\vdots	\ddots	\vdots	\ddots	\vdots	\vdots
A_p	$H_{(A_p,A_1)}(D/A)$	\dots	$H_{(A_p,A_q)}(D/A)$	\dots	$H_{(A_p,A_n)}(D/A)$	$H_{(A_p)}(D/A)$
\vdots	\vdots	\ddots	\vdots	\ddots	\vdots	\vdots
A_n	$H_{(A_n,A_1)}(D/A)$	\dots	$H_{(A_n,A_q)}(D/A)$	\dots	$H_{(A_n,A_n)}(D/A)$	$H_{(A_n)}(D/A)$
Meso-Middle	$H_{(A_1)}(D/A)$	\dots	$H_{(A_q)}(D/A)$	\dots	$H_{(A_n)}(D/A)$	Macro-Top: $H(D/A)$

Algorithm 2: Calculation of double-granule conditional-entropy at meso-middle

- Input:** Decision table $(U, C \cup D)$, target subset $A \subseteq C$, and a special index $p \in \{1, \dots, n\}$;
Output: Double-granule conditional-entropy $H_{(A_p)}(D/A)$ at meso-middle $(A_p, U/IND(A))$.
- 1: Compute $U/IND(A)$ to obtain all condition classes $A_i (i = 1, \dots, n)$ and a fixed granule $A_p \in U/IND(A)$.
 - 2: Compute $U/IND(D)$ to obtain all decision classes $D_j (j = 1, \dots, m)$.
 - 3: Let $H_{(A_p)}(D/A) = 0$.
 - 4: **for** $q \in \{1, \dots, n\}$ **do**
 - 5: Compute ω_p, ω_q .
 - 6: Let $H_p = 0, H_q = 0$.
 - 7: **for** $j \in \{1, \dots, m\}$ **do**
 - 8: $H_p \leftarrow H_p - P(D_j/A_p)\log_2 P(D_j/A_p),$
 $H_q \leftarrow H_q - P(D_j/A_q)\log_2 P(D_j/A_q).$
 - 9: **end for**
 - 10: Obtain $H_{(A_p,A_q)}(D/A) = \omega_p H_p + \omega_q H_q$.
 - 11: $H_{(A_p)}(D/A) \leftarrow H_{(A_p)}(D/A) + H_{(A_p,A_q)}(D/A)$.
 - 12: **end for**
 - 13: **return** $H_{(A_p)}(D/A)$.

Theorem 5. At meso-middle, the double-granule conditional-entropy has a lower bound and two upper bounds. Concretely,

$$\begin{aligned}
 H_{(A_p)}(D/A) &\in [\underline{H}_{(A_p)}(D/A), \overline{H}_{(A_p)}(D/A)], \\
 H_{(A_p)}(D/A) &\leq H_{(A_p)}^*(D/A),
 \end{aligned}
 \tag{24}$$

where

$$\begin{aligned}
 \underline{H}_{(A_p)}(D/A) &= \sum_{q=1}^n \underline{H}_{(A_p, A_q)}(D/A) = -\frac{n|A_p|}{2|U|} \sum_{j=1}^m P(D_j/A_p) \log_2 P(D_j/A_p) - \sum_{q=1}^n \left(\frac{|A_q|}{2|U|} \sum_{j=1}^m P(D_j/A_q) \log_2 P(D_j/A_q) \right), \\
 \overline{H}_{(A_p)}(D/A) &= \sum_{q=1}^n \overline{H}_{(A_p, A_q)}(D/A) = -n \sum_{j=1}^m P(D_j/A_p) \log_2 P(D_j/A_p) - \sum_{q=1}^n \sum_{j=1}^m P(D_j/A_q) \log_2 P(D_j/A_q), \\
 H_{(A_p)}^*(D/A) &= \sum_{q=1}^n H_{(A_p, A_q)}^*(D/A) = -\sum_{q=1}^n \sum_{j=1}^m \frac{|A_p \cap D_j| + |A_q \cap D_j|}{|A_p| + |A_q|} \log_2 \frac{|A_p \cap D_j| + |A_q \cap D_j|}{|A_p| + |A_q|}.
 \end{aligned}
 \tag{25}$$

Theorem 5 naturally comes from Theorems 2–4. The three bounds in Equation (25) hierarchically integrate previous three bounds at micro-bottom (Equations (11) and (17)) to correspondingly restrict $H_{(A_p)}(D/A)$. They can be supplemented into distributional Table 4, and following Table 5 provides the relevant part.

Table 5. Three bounds of double-granule conditional-entropies at meso-middle and macro-top.

$U/IND(A)$	$H_{(A_p)}(D/A)$	$\underline{H}_{(A_p)}(D/A)$	$\overline{H}_{(A_p)}(D/A)$	$H_{(A_p)}^*(D/A)$
A_1	$H_{(A_1)}(D/A)$	$\underline{H}_{(A_1)}(D/A)$	$\overline{H}_{(A_1)}(D/A)$	$H_{(A_1)}^*(D/A)$
\vdots	\vdots	\vdots	\vdots	\vdots
A_p	$H_{(A_p)}(D/A)$	$\underline{H}_{(A_p)}(D/A)$	$\overline{H}_{(A_p)}(D/A)$	$H_{(A_p)}^*(D/A)$
\vdots	\vdots	\vdots	\vdots	\vdots
A_n	$H_{(A_n)}(D/A)$	$\underline{H}_{(A_n)}(D/A)$	$\overline{H}_{(A_n)}(D/A)$	$H_{(A_n)}^*(D/A)$
Macro-Top	$H(D/A)$	$\underline{H}(D/A)$	$\overline{H}(D/A)$	$H^*(D/A)$

At meso-middle, $H_{(A_p)}(D/A)$ introduces the condition granulation $U/IND(A)$, but it still needs condition granule A_p . Thus, we cannot make a positive assertion regarding granulation monotonicity/non-monotonicity. In fact, $U/IND(A) \succeq U/IND(B)$ also implies chaos between $H_{(A_p)}(D/A)$ and $H_{(B_{p^*})}(D/B)$.

3.3. Double-Granule Conditional-Entropy at Macro-Top

As analyzed above, double-granule conditional-entropies at meso-middle consider the condition granulation, but in an insufficient way, and $H_{(A_p)}(D/A)$ also depends on a single condition granule A_p . For the thorough granulation and robust description, systematic measures $H_{(A_p)}(D/A)$ ($p = 1, \dots, n$) can be further integrated to generate double-granule conditional-entropies at macro-top. Based on the previous thought and result in Sections 3.1 and 3.2, this subsection further discusses double-granule conditional-entropies at macro-top

$$(U/IND(A) = \{A_p : p = 1, \dots, n\}, U/IND(A) = \{A_q : q = 1, \dots, n\}),$$

which is given in Table 1. We will directly provide the relevant integration definition, number distribution, calculation algorithm, three bounds, and we finally uncover an important conclusion of granulation non-monotonicity.

Definition 6. At macro-top $(U/IND(A), U/IND(A))$, the double-granule conditional-entropy is defined by

$$\begin{aligned}
 H(D/A) &= -\sum_{p=1}^n \sum_{q=1}^n \left(\omega_p \sum_{j=1}^m P(D_j/A_p) \log_2 P(D_j/A_p) \right) - \sum_{p=1}^n \sum_{q=1}^n \left(\omega_q \sum_{j=1}^m P(D_j/A_q) \log_2 P(D_j/A_q) \right) \\
 &= -\sum_{p=1}^n \sum_{q=1}^n \left(\frac{|A_p|}{|A_p| + |A_q|} \sum_{j=1}^m \frac{|A_p \cap D_j|}{|A_p|} \log_2 \frac{|A_p \cap D_j|}{|A_p|} \right) - \sum_{p=1}^n \sum_{q=1}^n \left(\frac{|A_q|}{|A_p| + |A_q|} \sum_{j=1}^m \frac{|A_q \cap D_j|}{|A_q|} \log_2 \frac{|A_q \cap D_j|}{|A_q|} \right).
 \end{aligned}
 \tag{26}$$

Corollary 2. *At macro-top, the double-granule conditional-entropy has an analytic expression:*

$$H(D/A) = - \sum_{p=1}^n \left(\sum_{q=1}^n \frac{|A_p|}{|A_p| + |A_q|} \right) \left(\sum_{j=1}^m P(D_j/A_p) \log_2 P(D_j/A_p) \right) - \sum_{p=1}^n \sum_{q=1}^n \left(\frac{|A_q|}{|A_p| + |A_q|} \sum_{j=1}^m P(D_j/A_q) \log_2 P(D_j/A_q) \right). \quad (27)$$

Theorem 6. *Double-granule conditional-entropies have a hierarchical integration from micro-bottom and meso-middle to macro-top, i.e.,*

$$\begin{aligned} H(D/A) &= \sum_{p=1}^n H_{(A_p)}(D/A) = H_{(A_1)}(D/A) + \cdots + H_{(A_n)}(D/A) \\ &= \sum_{p=1}^n \sum_{q=1}^n H_{(A_p, A_q)}(D/A) = H_{(A_1, A_1)}(D/A) + \cdots + H_{(A_n, A_n)}(D/A). \end{aligned} \quad (28)$$

By Definition 6 (Corollary 2) and Theorem 6, macro-top's measure $H(D/A)$ hierarchically integrates meso-middle's entropies $H_{(A_p)}(D/A)$ by a single summation on $p = 1, \dots, n$, and thus it further hierarchically integrates micro-bottom's entropies $H_{(A_p, A_q)}(D/A)$ by double summations on $p, q = 1, \dots, n$. As a result, $H(D/A)$ inherits the features of double-granule and conditional-entropy. It considers only conditional granulation $U/IND(A)$ to be at macro-top ($U/IND(A), U/IND(A)$), so it is called the double-granule conditional-entropy at macro-top. As an ultimate measure, $H(D/A)$ completely utilizes the $U/IND(A)$ granulation information to effectively describe decision classification $U/IND(D)$ and its uncertainty, thus holding robust measurement functions for knowledge granulation. Moreover, $H(D/A)$ can be noted by $H_{(U/IND(A), U/IND(A))}(D/A)$.

Proposition 4. *At macro-top, the double-granule conditional-entropy offers only one value, i.e.,*

$$H(D/A) \text{ at macro-top } (U/IND(A), U/IND(A)).$$

In Proposition 4, the double-granule conditional-entropy naturally exhibits number 1 to correspond to the sole macro-top. In fact, the first top entropy comes from the fusion of either n middle entropies or $n \times n$ bottom entropies; thus, three-level entropies accord with three-level granular structures (Table 1) from the quantitative and structural perspective, and they embody the hierarchical integration. In particular, the sole conditional-entropy $H(D/A)$ is put into the lower-right corner of Table 4, thus corresponding to the summations of central $n \times n$ micro values and marginal n meso values. According to Equations (26) and (28), Algorithm 3 resorts to three "for" loops to effectively offer the double-granule conditional-entropy $H(D/A)$. The two inner loops invoke Algorithm 2 to calculate an arbitrary double-granule conditional-entropy at meso-middle (where the central loop invokes Algorithm 1 to construct micro-bottom's entropies), while the outer loop integrates n related meso-middle's information values to produce $H(D/A)$. In other words, Algorithms 1–3 exhibit a kind of hierarchical evolution based on circulation development, and thus they constitute a novel kind of three-level algorithms.

Algorithm 3: Calculation of double-granule conditional-entropy at macro-top

Input: Decision table $(U, C \cup D)$, target subset $A \subseteq C$;
Output: Double-granule conditional-entropy $H(D/A)$ at Macro-Top $(U/IND(A), U/IND(D))$.

- 1: Compute $U/IND(A)$ to obtain all condition classes $A_i (i = 1, \dots, n)$.
- 2: Compute $U/IND(D)$ to obtain all decision classes $D_j (j = 1, \dots, m)$.
- 3: Let $H(D/A) = 0$.
- 4: **for** $p \in \{1, \dots, n\}$ **do**
- 5: Let $H_{(A_p)}(D/A) = 0$.
- 6: **for** $q \in \{1, \dots, n\}$ **do**
- 7: Compute ω_p, ω_q .
- 8: Let $H_p = 0, H_q = 0$.
- 9: **for** $j \in \{1, \dots, m\}$ **do**
- 10: $H_p \leftarrow H_p - P(D_j/A_p)\log_2 P(D_j/A_p),$
 $H_q \leftarrow H_q - P(D_j/A_q)\log_2 P(D_j/A_q).$
- 11: **end for**
- 12: Obtain $H_{(A_p, A_q)}(D/A) = \omega_p H_p + \omega_q H_q$.
- 13: $H_{(A_p)}(D/A) \leftarrow H_{(A_p)}(D/A) + H_{(A_p, A_q)}(D/A)$.
- 14: **end for**
- 15: $H(D/A) \leftarrow H(D/A) + H_{(A_p)}(D/A)$.
- 16: **end for**
- 17: **return** $H(D/A)$.

Theorem 7. At macro-top, the double-granule conditional-entropy has a lower bound and two upper bounds. Concretely,

$$\begin{aligned}
 H(D/A) &\in [\underline{H}(D/A), \overline{H}(D/A)], \\
 H(D/A) &\leq H^*(D/A),
 \end{aligned}
 \tag{29}$$

where

$$\begin{aligned}
 \underline{H}(D/A) &= \sum_{p=1}^n \underline{H}_{(A_p)}(D/A) = \sum_{p=1}^n \sum_{q=1}^n \underline{H}_{(A_p, A_q)}(D/A) \\
 &= -n \sum_{p=1}^n \left(\frac{|A_p|}{2|U|} \sum_{j=1}^m P(D_j/A_p) \log_2 P(D_j/A_p) \right) - n \sum_{q=1}^n \left(\frac{|A_q|}{2|U|} \sum_{j=1}^m P(D_j/A_q) \log_2 P(D_j/A_q) \right) \\
 &= -n \sum_{p=1}^n \left(\frac{|A_p|}{|U|} \sum_{j=1}^m P(D_j/A_p) \log_2 P(D_j/A_p) \right),
 \end{aligned}
 \tag{30}$$

$$\begin{aligned}
 \overline{H}(D/A) &= \sum_{p=1}^n \overline{H}_{(A_p)}(D/A) = \sum_{p=1}^n \sum_{q=1}^n \overline{H}_{(A_p, A_q)}(D/A) \\
 &= -n \sum_{p=1}^n \sum_{j=1}^m P(D_j/A_p) \log_2 P(D_j/A_p) - n \sum_{q=1}^n \sum_{j=1}^m P(D_j/A_q) \log_2 P(D_j/A_q) \\
 &= -2n \sum_{p=1}^n \sum_{j=1}^m P(D_j/A_p) \log_2 P(D_j/A_p),
 \end{aligned}
 \tag{31}$$

$$\begin{aligned}
 H^*(D/A) &= \sum_{p=1}^n H_{(A_p)}^*(D/A) = \sum_{p=1}^n \sum_{q=1}^n H_{(A_p, A_q)}^*(D/A) \\
 &= - \sum_{p=1}^n \sum_{q=1}^n \sum_{j=1}^m \frac{|A_p \cap D_j| + |A_q \cap D_j|}{|A_p| + |A_q|} \log_2 \frac{|A_p \cap D_j| + |A_q \cap D_j|}{|A_p| + |A_q|}.
 \end{aligned}
 \tag{32}$$

Theorem 7 naturally comes from Theorems 2–6. The three bounds in Equations (30)–(32) hierarchically integrate previous three bounds at meso-middle and micro-bottom, and thus they become three new uncertainty measures at macro-top ($U/IND(A), U/IND(A)$) to correspondingly restrict $H(D/A)$. They are supplemented into the bottom in the previous bound table: Table 5.

Theorem 8. *At macro-top, the double-granule conditional-entropy has granulation non-monotonicity. That is, $U/IND(A) \succeq U/IND(B)$ cannot necessarily achieve*

$$\text{either } H(D/A) \geq H(D/B) \text{ or } H(D/A) \leq H(D/B),$$

and both cases can practically exist. In addition, the matched double bounds $\underline{H}(D/A)$ and $\overline{H}(D/A)$ (Equations (30) and (31)) also have the granulation non-monotonicity, and they cannot theoretically acquire

$$\begin{aligned}
 &\text{either } \underline{H}(D/A) \geq \underline{H}(D/B) \text{ or } \underline{H}(D/A) \leq \underline{H}(D/B), \\
 &\text{either } \overline{H}(D/A) \geq \overline{H}(D/B) \text{ or } \overline{H}(D/A) \leq \overline{H}(D/B).
 \end{aligned}$$

At macro-top, the double-granule conditional-entropy completely breaks away from the condition granule dependence to establish the condition granulation description, so it becomes a powerful type of information measure for knowledge-based uncertainty representation. In terms of condition granulation, its non-monotonicity is finally revealed in Theorem 8, and the relevant evidence will be provided in the later example and experiment. Moreover, this fundamental non-monotonicity conclusion embodies information uncertainty, and it can be induced or explained by the previous complexity mechanism at micro-bottom and meso-middle. Based on macro-top and its granulation mechanism, the related three bounds (Equations (30)–(32)) and their monotonicity/non-monotonicity can be practically observed, and thus we also obtain the granulation non-monotonicity for $\underline{H}(D/A)$ and $\overline{H}(D/A)$; however, the case of upper bound $H^*(D/A)$ becomes a remaining problem.

4. Decision Table Example

In this section, the above theoretical constructions and properties are illustrated by a decision table example. By extracting a part of VOTING data set (which comes from UCI database [65]), we provide a practical decision table ($U, C \cup D$) in Table 6 with

$$|U| = 8, |C| = 11, |D| = 1.$$

Table 6. A decision table.

U	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	D
x_1	2	2	4	4	4	3	4	4	4	2	4	1
x_2	2	2	4	4	2	2	4	4	4	2	3	1
x_3	3	4	3	4	2	4	2	4	4	2	2	0
x_4	2	4	2	3	2	4	2	4	2	2	4	0
x_5	4	4	2	4	2	4	3	4	4	4	3	0
x_6	2	4	2	4	2	2	2	4	4	4	4	0
x_7	2	2	4	4	2	2	2	3	4	4	2	0
x_8	2	2	4	4	2	2	2	4	4	3	4	1

According to this decision table,

$$U/IND(D) = \{D_1 = \{x_1, x_2, x_8\}, D_2 = \{x_3, x_4, x_5, x_6, x_7\}\}$$

provides $m = |U/IND(D)| = 2$. As an example, $A = \{c_1, c_2, c_3, c_4, c_5\}$ is chosen to generate condition granulation

$$U/IND(A) = \{A_1 = \{x_1\}, A_2 = \{x_2, x_7, x_8\}, A_3 = \{x_3\}, A_4 = \{x_4\}, A_5 = \{x_5\}, A_6 = \{x_6\}\},$$

where $n = |U/IND(A)| = 6$. By virtue of three-level granular structures (Table 1), double-granule conditional-entropies and their three bounds are calculated by relevant algorithms and definitions, and they are compactly listed in Tables 7 and 8, respectively. The measures at micro-bottom, meso-middle, macro-top have numbers 36, 6, 1, respectively, and they correspond to the central 6×6 matrix, marginal 6-dimensional vector, lower-right-corner 1 digit, respectively. In part, we provide some processes of entropy calculation as follows.

$$\begin{aligned} -P(D_1/A_1)\log_2 P(D_1/A_1) &= 0, -P(D_2/A_1)\log_2 P(D_2/A_1) = 0, \\ -P(D_1/A_2)\log_2 P(D_1/A_2) &= 0.3900, -P(D_2/A_2)\log_2 P(D_2/A_2) = 0.5283, \\ H_{(A_1,A_1)}(D/A) &= \frac{1}{1+1}(0+0) + \frac{1}{1+1}(0+0) = 0, \\ H_{(A_1,A_2)}(D/A) &= \frac{1}{1+3}(0+0) + \frac{3}{1+3}(0.3900+0.5283) = 0.6887; \\ H_{(A_1)}(D/A) &= H_{(A_1,A_1)}(D/A) + H_{(A_1,A_2)}(D/A) + \dots + H_{(A_1,A_6)}(D/A) \\ &= 0 + 0.6887 + 0 + 0 + 0 + 0 = 0.6887; \\ H(D/A) &= H_{(A_1)}(D/A) + H_{(A_2)}(D/A) + \dots + H_{(A_6)}(D/A) \\ &= 0.6887 + 4.3619 + 0.6887 + 0.6887 + 0.6887 + 0.6887 = 7.8055. \end{aligned} \tag{33}$$

By Tables 7 and 8, we can make relevant verification analyses. First, entropies and bounds naturally present hierarchical integration relationships from micro-bottom to meso-middle to macro-top. Indeed, conditional-entropies are correspondingly restricted by three bounds. Moreover, the two types of upper bounds exactly have no necessary size relationships, and a part but powerful proof is provided as follows:

$$\begin{cases} \bar{H}_{(A_1,A_2)}(D/A) = 0.9183 > 0.8113 = H_{(A_1,A_2)}^*(D/A), \\ \bar{H}_{(A_1,A_3)}(D/A) = 0 < 1 = H_{(A_1,A_3)}^*(D/A); \end{cases}$$

$$\begin{cases} \bar{H}_{(A_1)}(D/A) = 0.9183 < 4.8113 = H_{(A_1)}^*(D/A), \\ \bar{H}_{(A_2)}(D/A) = 6.4281 > 5.7296 = H_{(A_2)}^*(D/A). \end{cases}$$

Table 7. Information values of double-granule conditional-entropies in the example.

U	A_1	A_2	A_3	A_4	A_5	A_6	Meso-Middle
A_1	0	0.6887	0	0	0	0	0.6887
A_2	0.6887	0.9183	0.6887	0.6887	0.6887	0.6887	4.3619
A_3	0	0.6887	0	0	0	0	0.6887
A_4	0	0.6887	0	0	0	0	0.6887
A_5	0	0.6887	0	0	0	0	0.6887
A_6	0	0.6887	0	0	0	0	0.6887
Meso-Middle	0.6887	4.3619	0.6887	0.6887	0.6887	0.6887	Macro-Top: 7.8055

Table 8. Three bounds of double-granule conditional-entropies in the example.

U	A_1	A_2	A_3	A_4	A_5	A_6	Meso-Middle
A_1	[0, 0] 0	[0.1722, 0.9183] 0.8113	[0, 0] 1	[0, 0] 1	[0, 0] 1	[0, 0] 1	[0.1722, 0.9183] 4.8113
A_2	[0.1722, 0.9183] 0.8113	[0.3444, 1.8366] 0.9183	[0.1722, 0.9183] 1	[0.1722, 0.9183] 1	[0.1722, 0.9183] 1	[0.1722, 0.9183] 1	[1.2053, 6.4281] 5.7296
A_3	[0, 0] 1	[0.1722, 0.9183] 1	[0, 0] 0	[0, 0] 0	[0, 0] 0	[0, 0] 0	[0.1722, 0.9183] 2.0000
A_4	[0, 0] 1	[0.1722, 0.9183] 1	[0, 0] 0	[0, 0] 0	[0, 0] 0	[0, 0] 0	[0.1722, 0.9183] 2.0000
A_5	[0, 0] 1	[0.1722, 0.9183] 1	[0, 0] 0	[0, 0] 0	[0, 0] 0	[0, 0] 0	[0.1722, 0.9183] 2.0000
A_6	[0, 0] 1	[0.1722, 0.9183] 1	[0, 0] 0	[0, 0] 0	[0, 0] 0	[0, 0] 0	[0.1722, 0.9183] 2.0000
Meso-Middle	[0.1722, 0.9183] 4.8113	[1.2053, 6.4281] 5.7296	[0.1722, 0.9183] 2.0000	[0.1722, 0.9183] 2.0000	[0.1722, 0.9183] 2.0000	[0.1722, 0.9183] 2.0000	Macro-Top: [2.0662, 11.0196] 18.5049

Finally, the granulation non-monotonicity at macro-top (Theorem 8) is verified. For this, we chose a natural attribute-addition chain:

$$\{c_1\} \subset \{c_1, c_2\} \subset \cdots \subset \{c_1, c_2, \cdots, c_{11}\}.$$

CA_k ($k \in \{1, 2, \cdots, 11\}$) denotes the attribute subset in the chain, and its granulation is represented by

$$U/IND(CA_k) = \{CA_{k,1}, \cdots, CA_{k,p}, \cdots, CA_{k,|U/IND(CA_k)|}\}.$$

In other words, $CA_{k,p}$ corresponds to the k th chain element CA_k to represent the p th condition granule in partition $U/IND(CA_k)$. According to the subset chain, Table 9 provides double-granule conditional-entropies, including both part values at micro-bottom ($CA_{k,p}, CA_{k,q}$), meso-middle ($CA_{k,p}, U/IND(CA_k)$) and all values (as well as the three bounds) at macro-top ($U/IND(CA_k), U/IND(CA_k)$). As a supporting detail, previous Tables 7 and 8 actually embrace the chain element CA_5 and its partition $U/IND(CA_5) = \{\{x_1\}, \{x_2, x_7, x_8\}, \{x_3\}, \{x_4\}, \{x_5\}, \{x_6\}\}$, while double-granule conditional-entropies regarding attribute subset $CA_2 = \{c_1, c_2\}$ and corresponding condition granulation

$$U/IND(CA_2) = \{CA_{2,1} = \{x_1, x_2, x_7, x_8\}, CA_{2,2} = \{x_3\}, CA_{2,3} = \{x_4, x_6\}, CA_{2,4} = \{x_5\}\}$$

are supplemented in Table 10 for better observation and illustration.

- (1) Since different chain subsets may have different equivalence partitions and granule numbers, the measures at micro-bottom and meso-middle consider condition granules to have a distinctive number and difficult correspondence. Table 9 focuses on the small and the same granule number, but relevant granules have different connotations. For example, the granules of the first one — $CA_{k,1}$ ($k = 1, 2, \cdots, 11$)—may be different. Thus, we cannot acquire the so-called granulation non-monotonicity assertion because of granulation incompleteness, although the values at micro-bottom and meso-middle actually exhibit a kind of non-monotonic change in Table 9.
- (2) In contrast, macro-top offers the complete condition granulation, so we can effectively focus on value monotonicity/non-monotonicity for both double-granule conditional-entropies and their three bounds. Observing the bottom part of Table 9 in the enlargement chain direction, we can discover that the three types of information measures are all non-monotonic, i.e.,

$$H(D/CA_k), \underline{H}(D/CA_k), \overline{H}(D/CA_k) \text{ (except } H^*(D/CA_k)\text{)}.$$

More vividly, the entropy and its three bounds regarding the chain are depicted in Figure 3, so the related granulation non-monotonicity becomes clearer. For example, the macro entropy value $H(D/CA_k)$ first increases and then decreases in the addition chain direction. Moreover, Table 9 and Figure 3 reflect the restriction properties of three bounds.

Table 9. Double-granule conditional-entropies based on an attribute-enlargement chain in the example.

Level	Measure	CA ₁	CA ₂	CA ₃	CA ₄	CA ₅	CA ₆	CA ₇	CA ₈	CA ₉	CA ₁₀	CA ₁₁
Micro-Bottom	$H_{(CA_{k,1},CA_{k,1})}(D/CA_k)$	1.0000	0.8113	0.8113	0.8113	0	0	0	0	0	0	0
	$H_{(CA_{k,1},CA_{k,2})}(D/CA_k)$	0.8571	0.6490	0.6490	0.6490	0.6887	0.6887	1	1	1	1	1
	$H_{(CA_{k,1},CA_{k,3})}(D/CA_k)$	0.8571	0.5409	0.5409	0.6490	0	0	0	0	0	0	0
	$H_{(CA_{k,2},CA_{k,2})}(D/CA_k)$	0	0	0	0	0.9183	0.9183	0	0	0	0	0
	$H_{(CA_{k,2},CA_{k,3})}(D/CA_k)$	0	0	0	0	0.6887	0.6887	0	0	0	0	0
Meso-Middle	$H_{(CA_{k,1})}(D/CA_k)$	2.7143	2.6502	2.6502	3.4074	0.6887	0.6887	0.6667	0	0	0	0
	$H_{(CA_{k,2})}(D/CA_k)$	0.8571	0.6490	0.6490	0.6490	4.3619	4.3619	0.6667	0	0	0	0
	$H_{(CA_{k,3})}(D/CA_k)$	0.8571	0.5409	0.5409	0.6490	0.6887	0.6887	0.6667	0	0	0	0
Macro-Top	$H(D/CA_k)$	4.4286	4.4891	4.4891	6.0035	7.8055	7.8055	9.0000	0	0	0	0
	$\underline{H}(D/CA_k)$	2.2500	1.6226	1.6226	2.0282	2.0282	2.0282	1.7500	0	0	0	0
	$\overline{H}(D/CA_k)$	6.0000	6.4902	6.4902	8.1128	11.0196	11.0196	14.0000	0	0	0	0
	$H^*(D/CA_k)$	4.9409	6.6951	6.6951	8.5789	18.5409	18.5409	28.0196	30	30	30	30

Table 10. Double-granule conditional-entropies regarding $CA_2 = \{c_1, c_2\}$ in the example.

U	CA _{2,1}	CA _{2,2}	CA _{2,3}	CA _{2,4}	Meso-Middle
CA _{2,1}	0.8113	0.6490	0.5409	0.6490	2.6502
CA _{2,2}	0.6490	0	0	0	0.6490
CA _{2,3}	0.5409	0	0	0	0.5409
CA _{2,4}	0.6490	0	0	0	0.6490
Meso-Middle	2.6502	0.6490	0.5409	0.6490	Macro-Top: 4.4891

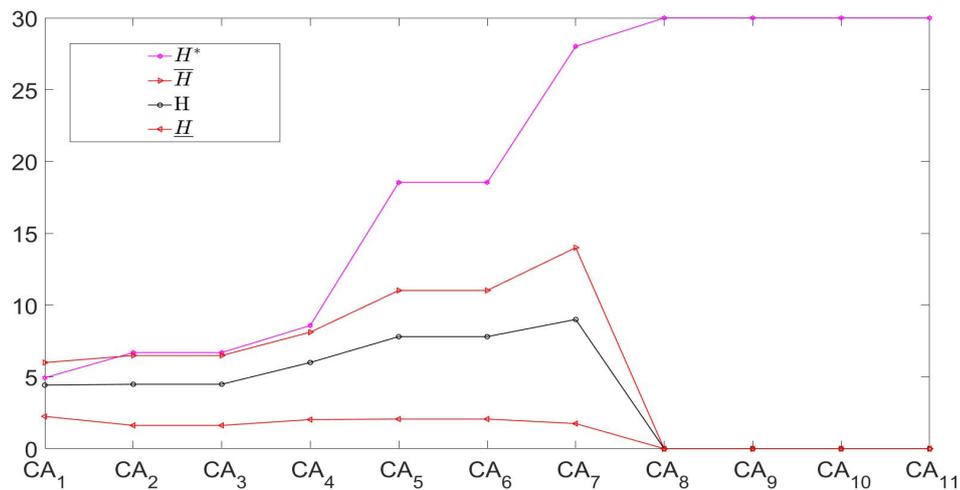


Figure 3. Macro-top’s double-granule conditional-entropies and their three bounds based on an attribute-enlargement chain in the example.

5. Data Experiments

In this section, the above theoretical results and their effectiveness are verified by data experiments. The new measures are mainly suitable for categorical (or nominal) data, which are usually used in the traditional rough set theory, and thus we adopt three relevant data sets from the UCI Machine Learning Repository [65], whose concrete descriptions on decision table $(U, C \cup D)$ are given in Table 11.

Table 11. Three UCI data sets.

Label	Name	$ U $	$ C $	$ U/IND(C) $	$ D $	$ U/IND(D) $
(1)	VOTING	435	16	342	1	2
(2)	SPECT	187	22	169	1	2
(3)	Tic-Tac-Toe	958	9	958	1	2

Similar to the above example, we also adopt the attribute-addition chain

$$CA_1 = \{c_1\} \subset \dots \subset CA_{|C|} = \{c_1, c_2, \dots, c_{|C|}\} \tag{34}$$

and its relevant symbol such as

$$U/IND(CA_k) = \{CA_{k,1}, \dots, CA_{k,p}, \dots, CA_{k,|U/IND(CA_k)|}\}.$$

Note that this attribute-subset sequence (Equation (34)) can deeply and typically probe the hierarchical knowledge-granulation within a framework of the complete lattice $(2^C, \subseteq)$. As a representative manifestation, we provide two typical results regarding the first chain element $CA_1 = \{c_1\}$ and the last one $CA_{|C|} = C$.

- (1) Regarding VOTING, $\{c_1\}$ and C induce three and 342 granules, respectively, and relevant double-granule conditional-entropies and three bounds are provided in Tables 12 and 13, respectively.
- (2) Regarding SPECT, $\{c_1\}$ and C produce two and 169 granules, respectively, and relevant three-level measures and three bounds are provided in Tables 14 and 15, respectively.
- (3) Regarding Tic-Tac-Toe, $\{c_1\}$ and C determine three and 958 granules, respectively, and relevant entropies and bounds are provided in Tables 16 and 17, respectively.

Table 12. Double-granule conditional-entropies in the VOTING data set.

$U/IND(CA_1)$	$CA_{1,1}$	$CA_{1,2}$	$CA_{1,3}$	Meso-Middle	...	$U/IND(CA_{16})$	$CA_{16,1}$...	$CA_{16,342}$	Meso-Middle
$CA_{1,1}$	0.9867	0.9782	0.8369	2.8018	...	$CA_{16,1}$	0	...	0	0
$CA_{1,2}$	0.9782	0.8113	0.6578	2.4473	...	\vdots	\vdots	\ddots	\vdots	\vdots
$CA_{1,3}$	0.8369	0.6578	0.6479	2.1427	...	$CA_{16,342}$	0	...	0	0
Meso-Middle	2.8018	2.4473	2.1427	Macro-Top: 7.3918	...	Meso-Middle	0	...	0	Macro-Top: 0

Table 13. Three information bounds in the VOTING data set.

$U/IND(CA_1)$	$CA_{1,1}$	$CA_{1,2}$	$CA_{1,3}$	Meso-Middle	...	$U/IND(CA_{16})$	$CA_{16,1}$...	$CA_{16,342}$	Meso-Middle
$CA_{1,1}$	[1.0706, 1.9734]	[0.5577, 1.7980]	[0.8139, 1.6346]	[2.4422, 5.4060]	...	$CA_{16,1}$	[0,0] 0	...	[0,0] 0	[0,0] 221.3143
$CA_{1,2}$	[0.5577, 1.7980]	[0.0448, 1.6226]	[0.3009, 1.4592]	[0.9034, 4.8798]	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
$CA_{1,3}$	[0.8139, 1.6346]	[0.3009, 1.4592]	[0.5571, 1.2959]	[1.6719, 4.3898]	...	$CA_{16,342}$	[0,0] 0	...	[0,0] 0	[0,0] 221.3143
Meso-Middle	[2.4422, 5.4060] 2.9436	[0.9034, 4.8798] 2.4630	[1.6719, 4.3898] 2.2724	Macro-Top: [5.0174, 14.6755] 7.6790	...	Meso-Middle	[0,0] 221.3143	...	[0,0] 221.3143	Macro-Top: [0,0] 50132

Table 14. Double-granule conditional-entropies in the SPECT data set.

$U/IND(CA_1)$	$CA_{1,1}$	$CA_{1,2}$	Meso-Middle	...	$U/IND(CA_{22})$	$CA_{22,1}$...	$CA_{22,169}$	Meso-Middle
$CA_{1,1}$	0.2108	0.3815	0.5924	...	$CA_{22,1}$	0	...	0	1.5335
					\vdots	\vdots	\ddots	\vdots	\vdots
$CA_{1,2}$	0.3815	0.5399	0.9215	...	$CA_{22,169}$	0	...	0	1.5335
Meso-Middle	0.5924	0.9215	Macro-Top: 1.5139	...	Meso-Middle	1.5335	...	1.5335	Macro-Top: 513.0879

Table 15. Three information bounds in the SPECT data set.

$U/IND(CA_1)$	$CA_{1,1}$	$CA_{1,2}$	Meso-Middle	...	$U/IND(CA_{22})$	$CA_{22,1}$...	$CA_{22,169}$	Meso-Middle
$CA_{1,2}$	[0.1015, 0.4217]	[0.1908, 0.7508]	[0.2922, 1.1725]	...	$CA_{22,1}$	[0,0] 0	...	[0,0] 0	[0.0332, 1.9457] 8.8982
					\vdots	\vdots	\ddots	\vdots	\vdots
$CA_{1,2}$	[0.1908, 0.7508]	[0.2801, 1.0799]	[0.4708, 1.8306]	...	$CA_{22,169}$	[0,0] 0	...	[0,0] 0	[0.0332, 1.9457] 161.2140
Meso-Middle	[0.2922, 1.1725] 0.6138	[0.4708, 1.8306] 0.9429	Macro-Top: [0.7631, 14.6755] 1.5566	...	Meso-Middle	[0.0332, 1.9457] 8.8982	...	[0.0332, 1.9457] 161.2140	Macro-Top: [11.2085, 657.6332] 2867

Table 16. Double-granule conditional-entropies in the Tic-Tac-Toe data set.

$U/IND(CA_1)$	$CA_{1,1}$	$CA_{1,2}$	$CA_{1,3}$	Meso-Middle	...	$U/IND(CA_9)$	$CA_{9,1}$...	$CA_{9,958}$	Meso-Middle
$CA_{1,1}$	0.8742	0.9248	0.8794	2.6784	...	$CA_{9,1}$	0	...	0	0
$CA_{1,2}$	0.9248	0.9881	0.9509	2.8638	...	\vdots	\vdots	\ddots	\vdots	\vdots
$CA_{1,3}$	0.8794	0.9509	0.8901	2.7203	...	$CA_{9,958}$	0	...	0	0
Meso-Middle	2.6784	2.8638	2.7203	Macro-Top: 8.2625	...	Meso-Middle	0	...	0	Macro-Top: 0

Table 17. Three information bounds in the Tic-Tac-Toe data set.

$U/IND(CA_1)$	$CA_{1,1}$	$CA_{1,2}$	$CA_{1,3}$	Meso-Middle	...	$U/IND(CA_9)$	$CA_{9,1}$...	$CA_{9,958}$	Meso-Middle
$CA_{1,1}$	[0.3814, 1.7483]	[0.3635, 1.8622]	[0.2859, 1.7642]	[1.0308, 5.3748]	...	$CA_{9,1}$	[0,0] 0	...	[0,0] 0	[0,0] 332
$CA_{1,2}$	[0.3635, 1.8622]	[0.3455, 1.9762]	[0.2680, 1.8781]	[0.9770, 5.7165]	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
$CA_{1,3}$	[0.2859, 1.7642]	[0.2680, 1.8781]	[0.1905, 1.7801]	[0.7444, 5.4224]	...	$CA_{9,958}$	[0,0] 0	...	[0,0] 0	[0,0] 626
Meso-Middle	[1.0308, 5.3748] 2.6940	[0.9770, 5.7165] 2.8913	[0.7444, 5.4224] 2.7342	Macro-Top: [2.7522, 16.5138] 8.3178	...	Meso-Middle	[0,0] 332	...	[0,0] 626	Macro-Top: [0,0] 415664

From the perspective of macro-top, double-granule conditional-entropies and their three information bounds based on the attribute-enlargement chain are finally summarized in Figure 4. These tables and figures can be utilized to effectively verify all previous conclusions, including the hierarchy, algorithm, restriction, and non-monotonicity. In particular, double-granule conditional-entropies are confined by three bounds, thus supporting the boundedness (Theorems 2, 3, 5 and 7); moreover, the entropies and their matched double-bounds fluctuate up and down, thus proving relevant granulation non-monotonicity (Theorem 8).

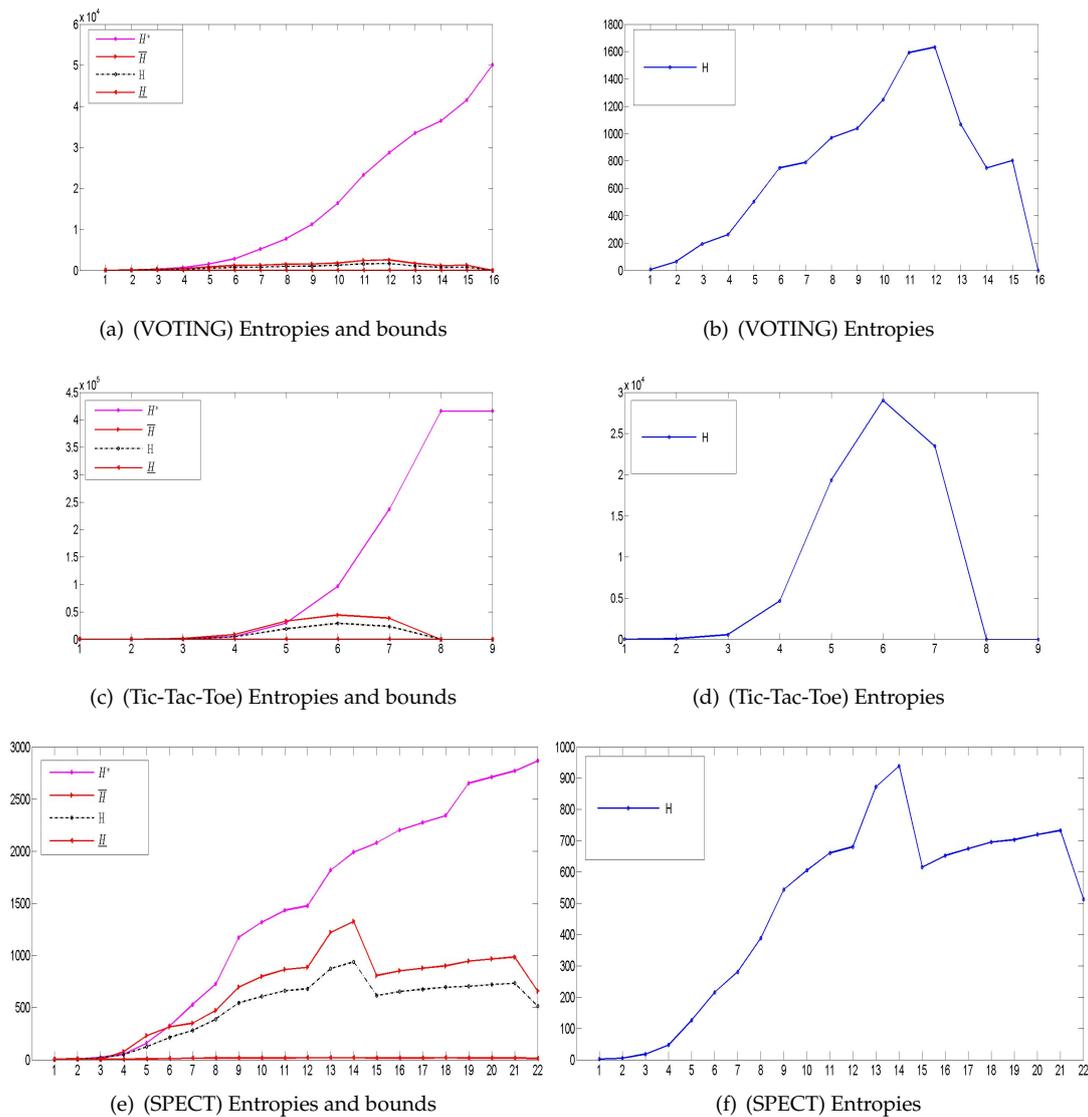


Figure 4. Macro-top’s double-granule conditional-entropies and their three information bounds based on an attribute-enlargement chain in data experiments.

6. Conclusions

The information measures implement fundamental uncertainty measurement in rough set theory and granular computing. The local conditional-entropies have the second-order feature, but they are limited to micro-bottom for describing discernibility matrix and reduction core [18]. In this paper, double-granule conditional-entropies achieve corresponding improvements of hierarchical/conditional granulation, and thus they become broader measures with uncertainty representation and information processing. They focus more on the double-granule interaction rather than granule-union locality, which is used in local conditional-entropies [18]. This strategy

directly utilizes the second-order mechanism to implement more systematic and robust uncertainty measurements, especially when compared to the current mainstream of first-order information measures. In our studies, double-granule conditional-entropies and their hierarchies, granulation, algorithms, bounds, and non-monotonicity are acquired and verified at three-level granular structures (i.e., micro-bottom, meso-middle, macro-top), and these results underlie both the efficiency in information processing and effectiveness in knowledge-based data analyses. Furthermore, their future developments and in-depth applications can be explored as follows.

- (1) In contrast to the relevant technology in [56], the hierarchical granulation of three-level granular structures focuses on the conditional granulation and relevant number, and it can be generalized for granular computing.
- (2) The double-granule conditional-entropies and their three bounds become new types of information measures with the second-order feature. In contrast to the traditional first-order entropy system, their description power and application advantage need further practical verification.
- (3) The double-granule conditional-entropies have three-restrictive bounds and granulation non-monotonicity, which have been experimentally verified by a granulation-hierarchical sequence (i.e., Equation (34)). These results are worth deeply utilizing in uncertainty measurement and data mining.
- (4) The double-granule conditional-entropies originate from the local conditional-entropies to carry a potential and distinctive advantage of discernibility matrix representation, and they also have the complete conditional granulation to have application prospects in knowledge reasoning or acquisition. Both their relationships with the discernibility matrix and their functions on attribute reduction need be deeply researched by promoting the previous studies in [18].

Author Contributions: T.M. conceived the algorithms and implemented the experiments; X.Z. mined the hierarchies and properties; Z.M. analyzed the results. All authors read and revised the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grants 61673285 and 11671284), the Sichuan Science and Technology Project of China (Grants 19YYJC2845 and 2017JY0197), and the Sichuan Youth Science and Technology Foundation of China (Grant 2017JQ0046).

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

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pawlak, Z. Rough set. *Int. J. Comput. Inf. Sci.* **1982**, *11*, 38–39. [[CrossRef](#)]
2. Raza, M.S.; Qamar, U. Redefining core preliminary concepts of classic rough set theory for feature selection. *Eng. Appl. Artif. Intell.* **2017**, *65*, 375–387. [[CrossRef](#)]
3. Saha, I.; Sarkar, J.P.; Maulik, U. Integrated rough fuzzy clustering for categorical data analysis. *Fuzzy Sets Syst.* **2019**, *361*, 1–32. [[CrossRef](#)]
4. Qian, Y.H.; Liang, X.Y.; Wang, Q.; Liang, J.Y.; Liu, B.; Andrzej, S.; Yao, Y.Y.; Ma, J.M.; Dang, C.Y. Local rough set: A solution to rough data analysis in big data. *Int. J. Approx. Reason.* **2018**, *97*, 38–63. [[CrossRef](#)]
5. Hu, M.J.; Yao, Y.Y. Structured approximations as a basis for three-way decisions in rough set theory. *Knowl.-Based Syst.* **2019**, *165*, 92–109.
6. Yang, X.B.; Liang, S.C.; Yu, H.L.; Gao, S.; Qian, Y.H. Pseudo-label neighborhood rough set: Measures and attribute reductions. *Int. J. Approx. Reason.* **2019**, *105*, 112–129. [[CrossRef](#)]
7. Wang, Z.H.; Feng, Q.R.; Wang, H. The lattice and matroid representations of definable sets in generalized rough sets based on relations. *Inf. Sci.* **2019**, *485*, 505–520. [[CrossRef](#)]
8. Luo, C.; Li, T.R.; Chen, H.M.; Fujita, H.; Zhang, Y. Incremental rough set approach for hierarchical multicriteria classification. *Inf. Sci.* **2018**, *429*, 72–87. [[CrossRef](#)]
9. Yao, Y.Y.; Zhang, X.Y. Class-specific attribute reducts in rough set theory. *Inf. Sci.* **2017**, *418–419*, 601–618.

10. Zhang, X.Y.; Yang, J.L.; Tang, L.Y. Three-way class-specific attribute reducts from the information viewpoint. *Inf. Sci.* **2018**. [[CrossRef](#)]
11. Ma, X.A.; Yao, Y.Y. Three-way decision perspectives on class-specific attribute reducts. *Inf. Sci.* **2018**, *450*, 227–245. [[CrossRef](#)]
12. Miao, D.Q.; Zhao, Y.; Yao, Y.Y.; Li, H.X.; Xu, F.F. Relative reducts in consistent and inconsistent decision tables of the Pawlak rough set model. *Inf. Sci.* **2009**, *179*, 4140–4150. [[CrossRef](#)]
13. Lang, G.M.; Cai, M.J.; Fujita, H.; Xiao, Q.M. Related families-based attribute reduction of dynamic covering decision information systems. *Knowl.-Based Syst.* **2018**, *162*, 161–173.
14. Gao, C.; Lai, Z.H.; Zhou, J.; Wen, J.J.; Wong, W.K. Granular maximum decision entropy-based monotonic uncertainty measure for attribute reduction. *Int. J. Approx. Reason.* **2019**, *104*, 9–24. [[CrossRef](#)]
15. Wang, C.Z.; Shi, Y.P.; Fan, X.D.; Shao, M.W. Attribute reduction based on k-nearest neighborhood rough sets. *Int. J. Approx. Reason.* **2019**, *106*, 18–31. [[CrossRef](#)]
16. Wei, W.; Wu, X.Y.; Liang, J.Y.; Cui, J.B.; Sun, Y.J. Discernibility matrix based incremental attribute reduction for dynamic data. *Knowl.-Based Syst.* **2018**, *140*, 142–157.
17. Ma, F.M.; Ding, M.W.; Zhang, T.F.; Cao, J. Compressed binary discernibility matrix based incremental attribute reduction algorithm for group dynamic data. *Neurocomputing* **2019**, *344*, 20–27.
18. Nie, H.M.; Zhou, J.Q. A new discernibility matrix and the computation of a core. *J. Sichuan Univ. (Nat. Sci. Ed.)* **2007**, *44*, 277–283. (In Chinese)
19. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
20. Shiraz, R.K.; Fukuyama, H.; Tavarna, M.; Caprio, D.D. An integrated data envelopment analysis and free disposal hull framework for cost-efficiency measurement using rough sets. *Appl. Soft Comput.* **2016**, *46*, 204–219. [[CrossRef](#)]
21. Liang, J.Y.; Shi, Z.Z.; Li, D.Y.; Wierman, M.J. Information entropy, rough entropy and knowledge granularity in incomplete information systems. *Int. J. Gen. Syst.* **2006**, *35*, 641–654. [[CrossRef](#)]
22. Wei, W.; Liang, J.Y. Information fusion in rough set theory: An overview. *Inf. Fusion* **2019**, *48*, 107–118. [[CrossRef](#)]
23. Hu, Q.H.; Yu, D.R.; Xie, Z.X.; Liu, J.F. Fuzzy probabilistic approximation spaces and their information measures. *IEEE Trans. Fuzzy Syst.* **2006**, *14*, 191–201.
24. Dai, J.H.; Wei, B.J.; Zhang, X.H. Uncertainty measurement for incomplete interval-valued information systems based on α -weak similarity. *Knowl.-Based Syst.* **2017**, *136*, 159–171.
25. Chen, Y.M.; Xue, Y.; Ma, Y.; Xu, F.F. Measures of uncertainty for neighborhood rough sets. *Knowl.-Based Syst.* **2017**, *120*, 226–235.
26. Miao, D.Q. Rough Set Theory and Its Application in Machine Learning. Ph.D. Thesis, Institute of Automation, The Chinese Academy of Sciences, Beijing, China, 1997. (In Chinese)
27. Wang, G.Y.; Zhao, J.; An, J.J.; Wu, Y. A comparative study of algebra viewpoint and information viewpoint in attribute reduction. *Fundam. Inf.* **2005**, *68*, 289–301.
28. Jiang, F.; Sui, Y.F.; Zhou, L. A relative decision entropy-based feature selection approach. *Pattern Recognit.* **2015**, *48*, 2151–2163. [[CrossRef](#)]
29. Slezak, D. Approximate entropy reducts. *Fundam. Inf.* **2002**, *53*, 365–390.
30. Qian, W.B.; Shu, W.H. Mutual information criterion for feature selection from incomplete data. *Neurocomputing* **2015**, *168*, 210–220. [[CrossRef](#)]
31. Liang, J.Y.; Chin, K.S.; Dang, C.Y.; Yam, R.C.M. A new method for measuring uncertainty and fuzziness in rough set theory. *Int. J. Gen. Syst.* **2002**, *31*, 331–342. [[CrossRef](#)]
32. Qian, Y.H.; Liang, J.Y. Combination entropy and combination granulation in rough set theory. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.* **2008**, *16*, 179–193.
33. Hu, Q.H.; Che, X.J.; Zhang, L. Rank entropy-based decision trees for monotonic classification. *IEEE Trans. Knowl. Data Eng.* **2012**, *24*, 2052–2064. [[CrossRef](#)]
34. Dai, J.H.; Xu, Q.; Wang, W.T.; Tian, H.W. Conditional entropy for incomplete decision systems and its application in data mining. *Int. J. Gen. Syst.* **2012**, *41*, 713–728. [[CrossRef](#)]
35. Sun, L.; Zhang, X.Y.; Xu, J.C.; Zhang, S.G. An attribute reduction method using neighborhood entropy measures in neighborhood rough sets. *Entropy* **2019**, *21*, 155. [[CrossRef](#)]
36. Chen, D.G.; Yang, W.X.; Li, F.C. Measures of general fuzzy rough sets on a probabilistic space. *Inf. Sci.* **2008**, *178*, 3177–3187. [[CrossRef](#)]

37. Mi, J.S.; Leung, Y.; Wu, W.Z. An uncertainty measure in partition-based fuzzy rough sets. *Int. J. Gen. Syst.* **2005**, *34*, 77–90. [[CrossRef](#)]
38. Hu, Q.H.; Zhang, L.; Zhang, D.; Pan, W.; An, S.; Pedrycz, W. Measuring relevance between discrete and continuous features based on neighborhood mutual information. *Expert Syst. Appl.* **2011**, *38*, 10737–10750. [[CrossRef](#)]
39. Zhao, J.Y.; Zhang, Z.L.; Han, C.Z.; Zhou, Z.F. Complement information entropy for uncertainty measure in fuzzy rough set and its applications. *Soft Comput.* **2015**, *19*, 1997–2010. [[CrossRef](#)]
40. Deng, X.F.; Yao, Y.Y. A multifaceted analysis of probabilistic three-way decisions. *Fundam. Inf.* **2014**, *132*, 291–313.
41. Deng, X.F.; Yao, Y.Y. An information-theoretic interpretation of thresholds in probabilistic rough sets. *Lect. Notes Comput. Sci.* **2012**, *7414*, 369–378.
42. Ma, X.A.; Wang, G.Y.; Yu, H.; Li, T.R. Decision region distribution preservation reduction in decision-theoretic rough set model. *Inf. Sci.* **2014**, *278*, 614–640. [[CrossRef](#)]
43. Zadeh, L.A. Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets Syst.* **1997**, *90*, 111–127. [[CrossRef](#)]
44. Yao, Y.Y. A triarchic theory of granular computing. *Granul. Comput.* **2016**, *1*, 145–157. [[CrossRef](#)]
45. Skowron, A.; Stepaniuk, J.; Swiniarski, R. Modeling rough granular computing based on approximation spaces. *Inf. Sci.* **2012**, *184*, 20–43. [[CrossRef](#)]
46. Chiaselotti, G.; Gentile, T.; Infusino, F. Granular computing on information tables: Families of subsets and operators. *Inf. Sci.* **2018**, *442–443*, 72–102.
47. Eissa, M.M.; Elmogy, M.; Hashem, M. Rough-granular computing knowledge discovery models for medical classification. *Egypt. Inf. J.* **2016**, *17*, 265–272. [[CrossRef](#)]
48. Qian, Y.H.; Zhang, H.; Sang, Y.L.; Liang, J.Y. Multigranulation decision-theoretic rough sets. *Int. J. Approx. Reason.* **2014**, *55*, 225–237. [[CrossRef](#)]
49. Li, J.H.; Mei, C.L.; Xu, W.H.; Qian, Y.H. Concept learning via granular computing: A cognitive viewpoint. *Inf. Sci.* **2015**, *298*, 447–467. [[CrossRef](#)]
50. Wang, G.Y.; Ma, X.A.; Yu, H. Monotonic uncertainty measures for attribute reduction in probabilistic rough set model. *Int. J. Approx. Reason.* **2015**, *59*, 41–67. [[CrossRef](#)]
51. Jia, X.Y.; Shang, L.; Zhou, B.; Yao, Y.Y. Generalized attribute reduct in rough set theory. *Knowl.-Based Syst.* **2016**, *91*, 204–218.
52. Zhang, X.Y.; Miao, D.Q. Double-quantitative fusion of accuracy and importance: Systematic measure mining, benign integration construction, hierarchical attribute reduction. *Knowl.-Based Syst.* **2016**, *91*, 219–240.
53. Calvanese, D.; Dumas, M.; Laurson, U.; Maggi, F.M.; Montali, M.; Teinmaa, I. Semantics analysis and simplification of DMN decision tables. *Inf. Syst.* **2018**, *78*, 112–125. [[CrossRef](#)]
54. Liu, G.L.; Hua, Z.; Zou, J.Y. Local attribute reductions for decision tables. *Inf. Sci.* **2018**, *422*, 204–217. [[CrossRef](#)]
55. Ge, H.; Li, L.S.; Xu, Y.; Yang, C.J. Quick general reduction algorithms for inconsistent decision tables. *Int. J. Approx. Reason.* **2017**, *82*, 56–80. [[CrossRef](#)]
56. Zhang, X.Y.; Miao, D.Q. Three-layer granular structures and three-way informational measures of a decision table. *Inf. Sci.* **2017**, *412–413*, 67–86.
57. Wang, J.; Tang, L.Y.; Zhang, X.Y.; Luo, Y.Y. Three-way weighted combination-entropies based on three-layer granular structures. *Appl. Math. Nonlinear Sci.* **2017**, *2*, 329–340.
58. Yao, Y.Y. An outline of a theory of three-way decisions. In *Rough Sets and Current Trends in Computing, Proceedings of the International Conference on Rough Sets and Current Trends in Computing, Chengdu, China, 17–20 August 2012*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 1–17.
59. Yao, Y.Y. Three-way decision and granular computing. *Int. J. Approx. Reason.* **2018**, *103*, 107–123. [[CrossRef](#)]
60. Fard, A.M.F.; Hajaghaei-Keshteli, M. A tri-level location-allocation model for forward/reverse supply chain. *Appl. Soft Comput.* **2018**, *62*, 328–346. [[CrossRef](#)]
61. Fathollahi-Fard, A.M.; Hajaghaei-Keshteli, M.; Mirjalili, S. Hybrid optimizers to solve a tri-level programming model for a tire closed-loop supply chain network design problem. *Appl. Soft Comput.* **2018**, *70*, 701–722. [[CrossRef](#)]
62. Gu, Y.; Cai, X.J.; Han, D.R.; Wang, D.Z.W. A tri-level optimization model for a private road competition problem with traffic equilibrium constraints. *Eur. J. Operat. Res.* **2019**, *273*, 190–197. [[CrossRef](#)]

63. Ye, D.Y.; Chen, Z.J. A new discernibility matrix and the computation of a core. *Acta Electr. Sin.* **2002**, *30*, 1086–1088. (In Chinese)
64. Zhang, X.Y.; Miao, D.Q. Quantitative/qualitative region-change uncertainty/certainty in attribute reduction: Comparative region-change analyses based on granular computing. *Inf. Sci.* **2016**, *334–335*, 174–204.
65. Dua, D.; Graff, C. *UCI Machine Learning Repository*; University of California, School of Information and Computer Science: Irvine, CA, USA, 2019. Available online: <http://archive.ics.uci.edu/ml> (accessed on 3 July 2019).



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).