International Journal of Machine Learning and Cybernetics (2021) 12:1427–1437
https://doi.org/10.1007/s13042-020-01243-y

ORIGINAL ARTICLE

Entropy based optimal scale combination selection for generalized multi-scale information tables

Han Bao[1,2] · **Wei-Zhi Wu**[1,2] · **Jia-Wen Zheng**[1,2] · **Tong-Jun Li**[1,2]

Received: 25 June 2020 / Accepted: 19 November 2020 / Published online: 2 January 2021

Abstract

In many real-life applications, data are often hierarchically structured at different levels of granulations. A multi-scale information table is a special hierarchical data set in which each object can take on as many values as there are scales under the same attribute. An important issue in such a data set is to select optimal scale combination in order to keep certain condition for final decision. In this paper, by employing Shannon's entropy, we study the selection of optimal scale combination to maintain uncertain measure of a knowledge from a generalized multi-scale information table. We first review the concept of entropy and its basic properties in information tables. We then introduce the notion of scale combinations in a generalized multi-scale information table. We further define entropy optimal scale combination in generalized multi-scale information tables and generalized multi-scale decision tables. Finally, we examine relationship between the entropy optimal scale combination and the classical optimal scale combination. We show that, in either a generalized multi-scale information table or a consistent generalized multi-scale decision table, the entropy optimal scale combination and the classical optimal scale combination are equivalent. And in an inconsistent generalized multi-scale decision table, a scale combination is generalized decision optimal if and only if it is a generalized decision entropy optimal.

Keywords Entropy · Granular computing · Information tables · Multi-scale information tables · Scale combinations

1 Introduction

Granular computing (GrC), which imitates human being's thinking, is currently a vivid direction in the research fields of artificial intelligence. Its basic computing unit is called a granule which is a class of objects or a collection of entities grouped together by the criteria of indistinguishability, similarity or functionality [36]. The main objective of research in GrC is the construction, interpretation, representation of granules, the selection of optimal granularities and the discovery of IF-THEN rules hidden in granular variables. It has been shown great promise as a new way for data mining and knowledge discovery in the context of big data [4, 16].

As a granular computing model, rough set theory, which was initiated by Pawlak [21], is a formal tool for modelling and processing insufficient and incomplete information. It has been successfully applied in the research fields of artificial intelligence such as pattern recognition, machine learning, and knowledge discovery in databases. The basic idea of rough set data analysis is to unravel knowledge in the form of a set of decision rules from an attribute-value representation model called an information system or an information table. In the traditional rough set data analysis, most of data sets are single scale information tables, that is, each object under each attribute in information tables can only take on one value. However, in many real-life data sets, an object can take on as many values as there are scales under the same attribute. Hence, to discover knowledge in hierarchically

✉ Wei-Zhi Wu
wuwz@zjou.edu.cn

Han Bao
995400054@qq.com

Jia-Wen Zheng
645958552@qq.com

Tong-Jun Li
litj@zjou.edu.cn

1 School of Mathematics, Physics and Information Science, Zhejiang Ocean University, Zhoushan, Zhejiang 316022, PR China

2 Key Laboratory of Oceanographic Big Data Mining and Application of Zhejiang Province, Zhejiang Ocean University, Zhoushan, Zhejiang 316022, PR China

organized information is an important issue in real-life data mining. To this end, Qian et al. developed a rough set data analysis approach called multi-granulation rough set model to study knowledge acquisition in hierarchically data sets [22, 23]. The multi-granulation rough set model concentrates on the selection of attributes with different combinations [13]. On the other hand, by introducing the notion of multi-scale information tables, Wu and Leung [31] proposed a novel rough-set data analysis model which has been named as the Wu-Leung model in [11, 12]. In this model, a multi-scale information table is an extension of information system in which each object under the same attribute can take on different values at different scales. Since the Wu-Leung model was introduced, the study of knowledge discovery in multi-scale information tables in rough set theory has attracted increasing attention [3, 5–9, 18, 25, 33, 34]. In this model, a key issue is to select an optimal scale with various requirements, which can be used to construct a proper sub-table in terms of all attributes restricted to the same level of granularity, for final decision or classification [6, 11, 18, 25, 29, 32–34, 38]. In the Wu-Leung model, a basic assumption on multi-scale information tables is that all attributes are granulated with the same number of scales. However, in many practical applications, different attributes may be measured by different numbers of scales/granulaities. Therefore, the assumption of the Wu-Leung model may limit the applicability of multi-scale information tables. Recently, Li and Hu [11, 12] introduced a new rough set data analysis model in which different numbers of scales can be used to measure different attributes. We call such a data set a generalized multi-scale information table or a generalized multi-scale decision table.

To unravel decision rules in such a data set, there are two key issues that need to be resolved, one is the choice of a suitable decision table from the given system, called optimal scale combination selection, in order to keep certain condition for final decision, and the other is the selection of minimal attributes in the chosen decision table, known as attribute reduction, in order to maintain the same condition for the extraction of decision rules. Since there are many rough set approaches to analyze attribute reduction and extract decision rules in decision tables, the first issue is of particular importance. In fact, Li and Hu [11] explored a complement model and a lattice model to analyze optimal scale selection in generalized multi-scale decision tables. By introducing the notion of scale combinations in a generalized multi-scale information table, Li et al. [12] defined six types of optimal scale combinations, namely the positive-region (respectively, lower-approximation, upper-approximation, generalized-decision, distribution and maximum-distribution) optimal scale combination, in inconsistent multi-scale decision tables and designed some algorithms to calculate these optimal scale combinations. Xu et al. [35] employed the Dempster-Shafer theory of evidence to characterize optimal scale combinations in consistent generalized multi-scale decision table. Recently, Wu et al. [30] further clarified relationships among seven types of optimal scale combinations in inconsistent generalized multi-scale decision tables.

The concept of entropy, proposed by Shannon in [24], is an important measure to evaluate uncertainty of a system. It has been successfully used to characterize information contents in rough set theory with applications in various information systems (see e.g. [1, 2, 14, 15, 17, 19, 26–28, 37–39]). It is well-known that the entropy reduct in information systems or decision tables is an important concept in rough set data analysis. In generalized multi-scale information tables or generalized multi-scale decision tables, to unravel a set of decision rules keeping information uncertainty, one has to select a proper sub-table in order to maintain entropy unchanged before making decision. To this end, in the present paper, we propose the notion of entropy optimal scale combinations in generalized multi-scale information/decision tables and examine the relationship between the new concept and the classical notion of optimal scale combination.

In the next section, we first review some preliminary concepts relevant to information tables, decision tables, and information entropy with basic properties. In Sect. 3, we introduce scale combinations and their properties in generalized multi-scale information tables. In Sect. 4, we analyze entropy based optimal scale combinations in generalized multi-scale information tables and generalized multi-scale decision tables. Relationships between the new type of optimal scale combination and the classical one are examined. We then summarize the paper with an outlook for further study in Sect. 5.

2 Preliminaries

Throughout this paper, for a given nonempty set U, $\mathcal{P}(U)$ denotes the power set of U. For $X \in \mathcal{P}(U)$, the complement of X in U is denoted as $\sim X$, i.e. $\sim X = U - X = \{x \in U | x \notin X\}$.

2.1 Information system and decision tables

Data set in rough set theory are usually described as an object-attribute value table called information system or information table. It provides a convenient tool for the representation of objects in terms of their attribute values.

Definition 1 An information system is a 2-tuple (U, A), where $U = \{x_1, x_2, \ldots, x_n\}$ is a non-empty and finite set of objects called the universe of discourse, and $A = \{a_1, a_2, \ldots, a_m\}$ is a non-empty and finite set of attributes such that $a : U \to V_a$, for any $a \in A$, i.e. $a(x) \in V_a$, $x \in U$, where $V_a = \{a(x)|x \in U\}$ is called the domain of a.

For each non-empty subset $B \subseteq A$, denote

$$R_B = \{(x, y) \in U \times U|a(x) = a(y), \forall a \in B\}.$$

R_B is called the indiscernibility relation induced by attributes B and it granulates U into disjointed sets

$$U/R_B = \{[x]_B|x \in U\},$$

where $[x]_B$ is the equivalence class of x w.r.t. (with respect to) attributes B, i.e.,

$$[x]_B = \{y \in U|(x, y) \in R_B\}.$$

From the perspective of granular computing, an equivalence class $[x]_B$ is a granule consisting of indistinguishable elements determined by attributes B and an attribute set B granulates U into a family of disjointed granules U/R_B which are basic elements for the approximation of any subset of U.

A decision table, also called a decision information system, is a 2-tuple $S = (U, C \cup \{d\})$, where (U, C) is an information table, C is called the conditional attribute set, and $d \notin C$ is a special attribute called the decision attribute which can be seen as a mapping $d : U \to V_d$. Without loss of generality, we assume in this paper that $V_d = \{1, 2, \dots, r\}$. The decision attribute d also determines an equivalence relation:

$$R_d = \{(x, y) \in U \times U|d(x) = d(y)\}.$$

And it partitions U into disjointed decision classes $U/R_d = \{D_1, D_2, \dots, D_r\}$, where $D_j = \{x \in U|d(x) = j\}$, and $j \in \{1, 2, \dots, r\}$.

If $R_C \subseteq R_d$, then the decision table $S = (U, C \cup \{d\})$ is said to be consistent, and it is inconsistent otherwise.

For any $B \subseteq C$, denote

$$\partial_B(x) = \{d(y)|y \in [x]_B\},\ x \in U,$$

$\partial_B(x)$ is called the generalized decision value of x w.r.t. B in S ([10]) and ∂_B the generalized decision function w.r.t. B in S. It is easy to see that S is consistent if and only if $|\partial_C(x)| = 1$ for all $x \in U$, and S is inconsistent otherwise.

2.2 Partitions and information entropies

Definition 2 [24] Let U be a non-empty finite set and $\mathcal{X} = \{X_1, X_2, \dots, X_t\}$ a partition on U. The information entropy of $\mathcal{X}$, denoted by $H(\mathcal{X})$, is defined as follows:

$$H(\mathcal{X}) = -\sum_{i=1}^{t} P(X_i)\log_2 P(X_i),$$

where $P(X_i) = |X_i|/|U|$ and $|X|$ is the cardinality of X.

Remark 1 If (U, A) is an information table, $B \subseteq A$ and $\mathcal{X}$ the partition generated by the equivalence relation R_B, then we will write $H(B)$ instead of $H(\mathcal{X})$.

Definition 3 Let U be a non-empty set, and $\mathcal{X}$ and $\mathcal{Y}$ two partitions on U. If for each $X \in \mathcal{X}$, there exists $Y \in \mathcal{Y}$ such that $X \subseteq Y$, then we say that $\mathcal{X}$ is finer than $\mathcal{Y}$ or $\mathcal{Y}$ is coarser than $\mathcal{X}$, and is denoted as $\mathcal{X} \preceq \mathcal{Y}$. Furthermore, if there exist $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ such that $X \subset Y$, then we say that $\mathcal{X}$ is strictly finer than $\mathcal{Y}$, and is denoted as $\mathcal{X} \prec \mathcal{Y}$.

Definition 4 Let U be a non-empty and finite set, and $\mathcal{X} = \{X_1, X_2, \dots, X_t\}$ and $\mathcal{Y} = \{Y_1, Y_2, \dots, Y_s\}$ two partitions on U. The conditional entropy of partition $\mathcal{Y}$ with reference to partition $\mathcal{X}$, denoted as $H(\mathcal{Y}|\mathcal{X})$, is defined as follows:

$$H(\mathcal{Y}|\mathcal{X}) = \sum_{i=1}^{t} P(X_i)H(\mathcal{Y}|X_i),$$

where $H(\mathcal{Y}|X_i) = -\sum_{j=1}^{s} P(Y_j|X_i)\log_2 P(Y_j|X_i)$, and $P(Y_j|X_i) = |Y_j \cap X_i|/|X_i|$.

Remark 2 Similar to Remark 1, if (U, A) is an information table, $B, C \subseteq A$, and $\mathcal{Y}$ and $\mathcal{X}$ are generated by the equivalence relations R_B and R_C respectively, then we will write $H(B|C)$ instead of $H(\mathcal{Y}|\mathcal{X})$.

Proposition 1 [20] *Let U be a non-empty finite set, and $\mathcal{X} = \{X_1, X_2, \dots, X_t\}$ and $\mathcal{Y} = \{Y_1, Y_2, \dots, Y_s\}$ two partitions on U. Then*

$$\mathcal{X} \preceq \mathcal{Y} \Rightarrow H(\mathcal{X}) \geq H(\mathcal{Y}).$$

Proposition 2 [27] *Let (U, A) be an information table, and $B, C \subseteq A$. If $B \subseteq C$ and $H(B) = H(C)$, then $R_B = R_C$.*

According to Proposition 2, it can directly be obtained the following proposition.

Proposition 3 *Let U be a non-empty finite set, and $\mathcal{X}$ and $\mathcal{Y}$ two partitions on U. If $\mathcal{X} \preceq \mathcal{Y}$ and $H(\mathcal{X}) = H(\mathcal{Y})$, then $\mathcal{X} = \mathcal{Y}$.*

Proposition 4 [15] *Let U be a non-empty finite set, and $\mathcal{X}$ and $\mathcal{Y}$ two partitions on U. Denote*

$$\mathcal{Z} = \mathcal{X} \cap \mathcal{Y} = \{Z \neq \emptyset|\exists X \in \mathcal{X}, \exists Y \in \mathcal{Y}, Z = X \cap Y\},$$

then

(1) $H(\mathcal{Z}) = H(\mathcal{X}) + H(\mathcal{Y}|\mathcal{X}) = H(\mathcal{Y}) + H(\mathcal{X}|\mathcal{Y})$.
(2) $H(\mathcal{X}) \geq H(\mathcal{X}|\mathcal{Y})$.

By Definition 4, the following Propositions 5 and 6 can be obtained immediately.

Proposition 5 *Let U be a non-empty finite set, and* $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ *three partitions on U. Then*

$\mathcal{X} \preceq \mathcal{Y} \Rightarrow H(\mathcal{Z}|\mathcal{X}) \leq H(\mathcal{Z}|\mathcal{Y})$.

Proposition 6 *Let U be a non-empty finite set, and* $\mathcal{X}$ *and* $\mathcal{Y}$ *two partitions on U. Then*

$\mathcal{X} \preceq \mathcal{Y} \Rightarrow H(\mathcal{Y}|\mathcal{X}) = 0$.

3 Generalized multi-scale information tables and generalized multi-scale decision tables

In a traditional information table, an object has a uniquely determined value at each attribute. In many practical applications, an object may have different attribute values according to different levels of scales under the same attribute. In [31], Wu and Leung introduced the concept of multi-scale information table to represent such a data set.

Definition 5 A tuple $S = (U, A)$ is called a multi-scale information table, where $U = \{x_1, x_2, \dots, x_n\}$ is a non-empty and finite set of objects called the universe of discourse and $A = \{a_1, a_2, \dots, a_m\}$ is a non-empty and finite set of attributes, and each $a_j \in A$ is a multi-scale attribute, i.e., for each object x_i in U, attribute value $a_j(x_i)$ may take on different values at different scales.

In [31], a basic assumption on the multi-scale information table $S = (U, A)$ is that all attributes possess the same number, I, of scales. Such a multi-scale information table can be represented as a system $(U, \{a_j^k | k = 1, 2, \dots, I, j = 1, 2, \dots, m\})$, where $a_j^k : U \to V_j^k$ is a surjective mapping and V_j^k is the domain of the k-th scale attribute a_j^k. For $k \in \{1, 2, \dots, I-1\}$, there exists a surjective mapping $g_j^{k,k+1} : V_j^k \to V_j^{k+1}$ such that $a_j^{k+1} = g_j^{k,k+1} \circ a_j^k$, i.e.

$$a_j^{k+1}(x) = g_j^{k,k+1}(a_j^k(x)), \quad x \in U,$$

where $g_j^{k,k+1}$ is referred to as a granular information transformation.

Definition 6 [31] A multi-scale decision table is a system $(U, C \cup \{d\}) = (U, \{a_j^k | k = 1, 2, \dots, I, \quad j = 1, 2, \dots, m\} \cup \{d\})$, where $(U, C) = (U, \{a_j^k | k = 1, 2, \dots, I, j = 1, 2, \dots, m\})$ is a multi-scale information table and $d \notin \{a_j^k | k = 1, 2, \dots, I, j = 1, 2, \dots, m\}$, $d : U \to V_d$, is a special single-scale attribute called the decision.

For a given multi-scale decision table $S = (U, C \cup \{d\}) = (U, \{a_j^k | k = 1, 2, \dots, I, j = 1, 2, \dots, m\} \cup \{d\})$, it can be divided into I decision sub-tables $S^k = (U, C^k \cup \{d\})$, $k = 1, 2, \dots, I$, with the same decision attribute, where $C^k = \{a_j^k | j = 1, 2, \dots, m\}$.

Evidently, the assumption of Wu-Leung model on multi-scale information tables that all attributes are granulated with the same number of scales may restrict its applications. Based on the observation, Li and Hu [11] recently developed a generalization of Wu-Leung model that different attributes have different numbers of scales. We call the system defined by Li and Hu a generalized multi-scale information table.

Definition 7 [11] A generalized multi-scale information table is a tuple $S = (U, A) = (U, \{a_j^k | k = 1, 2, \dots, I_j, j = 1, 2, \dots, m\})$, where $U = \{x_1, x_2, \dots, x_n\}$ is a non-empty and finite set of objects called the universe of discourse, and $A = \{a_1, a_2, \dots, a_m\}$ is a non-empty and finite set of attributes such that, for each $j \in \{1, 2, \dots, m\}$, $a_j^k : U \to V_j^k$ is a surjective mapping and V_j^k is the domain of the k-th scale attribute a_j^k, and for $1 \leq k \leq I_j - 1$, there exists a surjective mapping $g_j^{k,k+1} : V_j^k \to V_j^{k+1}$ such that $a_j^{k+1} = g_j^{k,k+1} \circ a_j^k$. And, a generalized multi-scale decision table is a system $S = (U, C \cup \{d\}) = (U, \{a_j^k | k = 1, 2, \dots, I_j, j = 1, 2, \dots, m\} \cup \{d\})$, where $(U, C) = (U, \{a_j^k | k = 1, 2, \dots, I_j, j = 1, 2, \dots, m\})$ is a generalized multi-scale information table and $d \notin \{a_j^k | k = 1, 2, \dots, I_j, j = 1, 2, \dots, m\}$, $d : U \to V_d$, is a special single-scale attribute called the decision.

We can see that if $I_1 = I_2 = \cdots = I_m = I$ in Definition 7, then the generalized multi-scale information table and generalized multi-scale decision table are, respectively, degenerated to a multi-scale information table and a multi-scale decision table in Definitions 5 and 6.

According to [31], we can see that the main approach to knowledge acquisition in a generalized multi-scale decision table is to select an appropriate single-scale decision table determined by a special scale for each attribute. Thus, how to select a suitable scale for each attribute for final decision is an important task in a generalized multi-scale decision table. In [11], Li and Hu defined the notion of scale combination in a generalized multi-scale information table to realize such task.

Definition 8 For a generalized multi-scale information table $S = (U, A) = (U, \{a_j^k | k = 1, 2, \dots, I_j, j = 1, 2, \dots, m\})$, where attribute a_j has I_j levels of scales, $j = 1, 2, \dots, m$. If attributes $a_1, a_2, \dots, a_m$ restricted to their l_j-th scale, $j = 1, 2, \dots, m$, respectively, form a single-scale information table S^L, where $L = (l_1, l_2, \dots, l_m)$, then the index set $(l_1, l_2, \dots, l_m)$ is called a scale combination of S^L in S. The family of all scale combinations in S, called the scale collection of S, is denoted as $\mathcal{L}$, i.e. $\mathcal{L} = \{(l_1, l_2, \dots, l_m) | l_j \in \{1, 2, \dots, I_j\}, j = 1, 2, \dots, m\}$.

For an $L = (l_1, l_2, \dots, l_m) \in \mathcal{L}$, let $A^L = \{a_1^{l_1}, a_2^{l_2}, \dots, a_m^{l_m}\}$, then $S^L = (U, A^L)$ is a single-scale information system corresponding to the scale combination L.

Definition 9 [11] For a generalized multi-scale information table $S = (U, A)$ and $L_1 = (l_1^1, l_2^1, \dots, l_m^1), L_2 = (l_1^2, l_2^2, \dots, l_m^2) \in \mathcal{L}$, if $l_j^1 \le l_j^2$ for all $j \in \{1, 2, \dots, m\}$, then L_1 is said to be weaker than L_2, or L_2 is stronger than L_1, denoted as $L_1 \preceq L_2$. Moreover, if there is a $j \in \{1, 2, \dots, m\}$ such that $l_j^1 < l_j^2$, then L_1 is said to be strictly weaker than L_2, or L_2 is strictly stronger than L_1, denoted as $L_1 \prec L_2$.

According to Li and Hu in [11], a scale combination is mainly used to select an appropriate decision table from the given generalized multi-scale decision table for decision making.

Proposition 7 [30] *For a generalized multi-scale information table $S = (U, A)$, and $L_1 = (l_1^1, l_2^1, \dots, l_m^1), L_2 = (l_1^2, l_2^2, \dots, l_m^2) \in \mathcal{L}$, define $L_1 \wedge L_2 = (l_1^1 \wedge l_1^2, l_2^1 \wedge l_2^2, \dots, l_m^1 \wedge l_m^2)$ and $L_1 \vee L_2 = (l_1^1 \vee l_1^2, l_2^1 \vee l_2^2, \dots, l_m^1 \vee l_m^2)$, where $l_j^1 \wedge l_j^2 = \min\{l_j^1, l_j^2\}$ and $l_j^1 \vee l_j^2 = \max\{l_j^1, l_j^2\}$ for all $j \in \{1, 2, \dots, m\}$. Then*

$$L_1 \preceq L_2 \Leftrightarrow L_1 \wedge L_2 = L_1 \Leftrightarrow L_1 \vee L_2 = L_2.$$

And $(\mathcal{L}, \preceq, \wedge, \vee)$ is a bounded lattice with the maximal element $(I_1, I_2, \dots, I_m)$ and the minimal element $(1, 1, \dots, 1)$.

For $B \subseteq A$ and $L = (l_1, l_2, \dots, l_m) \in \mathcal{L}$, denote by L_B the restriction of L on B (for example, if $A = \{a_1, a_2, a_3, a_4\}$, $B = \{a_1, a_2\}$, and $L = (2, 1, 1, 4) \in \mathcal{L}$, then $L_B = (2, 1)$) and denote $\mathcal{L}_B = \{L_B | L \in \mathcal{L}\}$. Clearly, $\mathcal{L}_B$ is the scale collection of the multi-scale information table (U, B). Denote

$$R_{B^L} = \{(x, y) \in U \times U | b^l(x) = b^l(y), \forall b^l \in B^{L_B}\}.$$

Obviously, R_{B^L} is the indiscernibility relation induced by attributes B under the scale combination L. It partitions U into equivalence classes $U/R_{B^L} = \{[x]_{B^L} | x \in U\}$, where $[x]_{B^L} = \{y \in U | (x, y) \in R_{B^L}\}$.

Proposition 8 [30] *If $S = (U, A)$ is a generalized multi-scale*

(1) $L_1 \preceq L_2 \Rightarrow R_{B^{L_1}} \subseteq R_{B^{L_2}}$,

(2) $L_1 \preceq L_2 \Rightarrow [x]_{B^{L_1}} \subseteq [x]_{B^{L_2}}, \forall x \in U$,
(3) $L_1 \preceq L_2 \Rightarrow U/R_{B^{L_1}} \preceq U/R_{B^{L_2}}$,
(4) $B \subseteq C \subseteq A \Rightarrow R_{C^L} \subseteq R_{B^L}$.

4 Optimal scale combination in generalized multi-scale decision tables

Knowledge acquisition in a generalized multi-scale decision table is a generalization of the one in a multi-scale decision table. We know that not all decision tables determined by different scale combinations meet the requirement of the decision table at the finest scale. Therefore, it is critical to select an optimal scale of combinations which can be used to generate a suitable decision table before decision rules are produced. In this section, we discuss the selection of optimal scale combinations under the requirement of entropy unchanged in generalized multi-scale information/decision tables.

4.1 Optimal scale combination selection in generalized multi-scale information tables

In this subsection, we will use the information entropy to characterize optimal scale combinations in generalized multi-scale information tables.

Definition 10 For a generalized multi-scale information table $S = (U, A) = (U, \{a_j^k | k = 1, 2, \dots, I_j, j = 1, 2, \dots, m\})$, $L \in \mathcal{L}$, and $A^{L_0} = \{a_j^1 | j = 1, 2, \dots, m\}$, we say that

(1) $S^L = (U, A^L)$ is consistent to S if $R_{A^L} = R_{A^{L_0}}$. And, L is said to be an optimal scale combination of S if S^L is consistent to S and S^K is not consistent to S for all $K \in \mathcal{L}$ with $L \prec K$ (if $\{K \in \mathcal{L} | L \prec K\} \neq \emptyset$).
(2) $S^L = (U, A^L)$ is entropy consistent to S if $H(A^L) = H(A^{L_0})$. And, L is said to be an entropy optimal scale combination of S if S^L is entropy consistent to S and S^K is not entropy consistent to S for all $K \in \mathcal{L}$ with $L \prec K$ (if $\{K \in \mathcal{L} | L \prec K\} \neq \emptyset$).

Theorem 1 *Let $S = (U, A) = (U, \{a_j^k | k = 1, 2, \dots, I_j, j = 1, 2, \dots, m\})$ be a generalized multi-scale information table, and $L \in \mathcal{L}$, then*

(1) *S^L is consistent to S if and only if S^L is entropy consistent to S.*
(2) *L is an optimal scale combination of S if and only if L is an entropy optimal scale combination of S.*

Proof

(1) "$\Rightarrow$". Assume that S^L is consistent to S, that is, $R_{A^L} = R_{A^{L_0}}$. It is easy to see that the partitions derived from these two equivalence relations are the same, consequently, their corresponding probability distribu-

tions are also the same, so $H(A^L) = H(A^{L_0})$. Thus S^L is entropy consistent to S.

"⇐". Assume that S^L is entropy consistent to S, that is $H(A^L) = H(A^{L_0})$.

It is obvious that $L_0 \preceq L$, combining with Proposition 8, we have $R_{A^{L_0}} \subseteq R_{A^L}$. Then, by Proposition 2, we see that $R_{A^L} = R_{A^{L_0}}$, thus we conclude that S^L is consistent to S.

(2) It follows immediately from (1). □

Theorem 2 *Let* $S = (U, A) = (U, \{a_j^k | k = 1, 2, \dots, I_j, j = 1, 2, \dots, m\})$ *be a generalized multi-scale information table, and* $L \in \mathcal{L}$, *then L is an optimal scale combination of S if and only if the following conditions hold*:

(1) $H(A^L) = H(A^{L_0})$, *where* $A^{L_0} = \{a_j^1 | j = 1, 2, \dots, m\}$.
(2) $H(A^K) < H(A^L)$ *for all* $K \in \mathcal{L}$ *with* $L \prec K$ (*if* $\{K \in \mathcal{L} | L \prec K\} \neq \emptyset$).

Proof "⇒". If L is an optimal scale combination of S, that is, S^L is consistent to S, then, by Theorem 1, we see that L is an entropy optimal scale combination of S, of course, S^L is entropy consistent to S, and hence $H(A^L) = H(A^{L_0})$. For any $K \in \mathcal{L}$ with $L \prec K$, on the one hand, by Propositions 1 and 8, we have $R_{A^L} \subseteq R_{A^K}$ and $H(A^K) \leq H(A^L)$. On the other hand, since L is an optimal scale combination of S and $L \prec K$, by Theorem 1, we see that S^K is not entropy consistent to S, i.e., $H(A^K) \neq H(A^{L_0})$. Thus we have proved that $H(A^K) < H(A^L)$.

"⇐". It is straightforward. □

Example 1 Table 1 is an example of a generalized multi-scale information table $S = (U, A)$, where $U = \{x_1, x_2, \dots, x_6\}$, $A = \{a_1, a_2\}$, attribute a_1 has three levels of scale while a_2 has two. Thus, (U, C) has six scale combinations $L_1 = (1, 1)$, $L_2 = (2, 1)$, $L_3 = (3, 1)$, $L_4 = (1, 2)$, $L_5 = (2, 2)$, and $L_6 = (3, 2)$.

We can calculate that

Table 1 A generalized multi-scale information table

U	a_1^1	a_1^2	a_1^3	a_2^1	a_2^2
x_1	1	S	Y	2	Y
x_2	1	S	Y	2	Y
x_3	3	M	N	1	N
x_4	2	L	N	1	N
x_5	2	L	N	1	N
x_6	4	M	N	3	Y

$$U/R_{A^{L_1}} = \{\{x_1, x_2\}, \{x_3\}, \{x_4, x_5\}, \{x_6\}\} = U/R_{A^{L_2}} = U/R_{A^{L_4}} = U/R_{A^{L_5}},$$
$$U/R_{A^{L_3}} = \{\{x_1, x_2\}, \{x_3, x_4, x_5\}, \{x_6\}\} = U/R_{A^{L_6}} \neq U/R_{A^{L_1}},$$
$$H(A^{L_1}) = -(\frac{2}{6}\log_2\frac{2}{6} + \frac{1}{6}\log_2\frac{1}{6} + \frac{2}{6}\log_2\frac{2}{6} + \frac{1}{6}\log_2\frac{1}{6}) = 1.918$$
$$= H(A^{L_2}) = H(A^{L_4}) = H(A^{L_5}),$$
$$H(A^{L_3}) = -(\frac{2}{6}\log_2\frac{2}{6} + \frac{3}{6}\log_2\frac{3}{6} + \frac{1}{6}\log_2\frac{1}{6}) = 1.459$$
$$= H(A^{L_6}) < H(A^{L_4}),$$

It is obvious that S^{L_2}, S^{L_4}, and S^{L_5} are consistent to S, and they are also entropy consistent to S. Notice that $L_6 = (3, 2)$ is the unique scale combination which is strictly stronger than L_5, and S^{L_6} is not consistent to S, then $L_5 = (2, 2)$ is an optimal scale combination of S, it is also an entropy optimal scale combination of S.

4.2 Optimal scale combination selection in consistent generalized multi-scale decision tables

In this subsection, we will examine that entropy optimal scale combinations can be employed to characterize optimal scale combinations in consistent generalized multi-scale decision tables.

Definition 11 For a generalized multi-scale decision table $S = (U, C \cup \{d\}) = (U, \{a_j^k | k = 1, 2, \dots, I_j, j = 1, 2, \dots, m\} \cup \{d\})$, if the decision table under the first (finest) level of scale combination L_0, $S^{L_0} = (U, \{a_j^1 | j = 1, 2, \dots, m\} \cup \{d\}) = (U, C^{L_0} \cup \{d\})$, is consistent, i.e., $R_{C^{L_0}} \subseteq R_d$, then we say that S is consistent, and it is inconsistent otherwise.

Definition 12 Let $S = (U, C \cup \{d\}) = (U, \{a_j^k | k = 1, 2, \dots, I_j, j = 1, 2, \dots, m\} \cup \{d\})$ be a consistent generalized multi-scale decision table , and $L \in \mathcal{L}$, we say that

(1) $S^L = (U, C^L \cup \{d\})$ is consistent to S if $R_{C^L} \subseteq R_d$. And, L is said to be an optimal scale combination of S if S^L is consistent to S and S^K is not consistent to S for all $K \in \mathcal{L}$ with $L \prec K$ (if $\{K \in \mathcal{L} | L \prec K\} \neq \emptyset$).
(2) $S^L = (U, C^L \cup \{d\})$ is entropy consistent to S if $H(d|C^L) = H(d|C^{L_0})$, where $C^{L_0} = \{a_j^1 | j = 1, 2, \dots, m\}$. And, L is said to be an entropy optimal scale combination of S if S^L is entropy consistent to S and S^K is not entropy consistent to S for all $K \in \mathcal{L}$ with $L \prec K$ (if $\{K \in \mathcal{L} | L \prec K\} \neq \emptyset$).

According to Definition 12, we see that an optimal scale combination of a consistent generalized multi-scale decision table is the best scale combination for decision making or classification in the generalized multi-scale decision table. And L is an optimal scale combination if and only if L is a maximal element in $\mathcal{L}$ such that S^L is a consistent decision table.

Theorem 3 *Let* $S=(U,C\cup\{d\})=(U,\{a_j^k|k=1,2,\dots,I_j,j=1,2,\dots,m\}\cup\{d\})$ *be a consistent generalized multi-scale decision table and* $L\in\mathcal{L}$, *then*

(1) *S^L is consistent to S if and only if S^L is entropy consistent to S.*
(2) *L is an optimal scale combination of S if and only if L is an entropy optimal scale combination of S.*

Proof

(1) "⇒". Assume that S^L is consistent to S. Since S is consistent, we have $R_{C^{L_0}}\subseteq R_d$, where $C^{L_0}=\{a_j^1|j=1,2,\dots,m\}$. Notice that S^L is consistent to S, that is, $R_{C^L}\subseteq R_d$. Then, by Proposition 6, we conclude $H(d|C^L)=H(d|C^{L_0})=0$. It follows that S^L is entropy consistent to S.
"⇐". Assume that S^L is entropy consistent to S. On the one hand, since S is consistent, we have $R_{C^{L_0}}\subseteq R_d$, combining with Proposition 6, we then conclude $H(d|C^{L_0})=0$. Since S^L is entropy consistent to S, we have $H(d|C^L)=H(d|C^{L_0})=0$.
If S^L is not consistent to S, that is, $R_{C^L}\nsubseteq R_d$, then there exits $Y_j\in U/R_d$ and $X_i\in U/R_{C^L}$ such that $Y_j\cap X_i\neq\emptyset$ and $Y_j\cap X_i\neq X_i$, and hence $\log_2 P(Y_j|X_i)=\log_2\frac{|Y_j\cap X_i|}{X_i}\neq 0$. Consequently, $H(d|C^L)\geq -P(X_i)P(Y_j|X_i)\log_2 P(Y_j|X_i)>0$ which contradicts with $H(d|C^L)=0$. Thus, we have proved that S^L is consistent to S.
(2) It follows immediately from (1).

□

Theorem 4 *Let* $S=(U,C\cup\{d\})=(U,\{a_j^k|k=1,2,\dots,I_j,j=1,2,\dots,m\}\cup\{d\})$ *be a consistent generalized multi-scale decision table and* $L\in\mathcal{L}$, *then L is an optimal scale combination of S if and only if the following conditions hold*:

(1) $H(d|C^L)=0$.
(2) $H(d|C^K)>0$ *for all* $K\in\mathcal{L}$ *with* $L\prec K$ (*if* $\{K\in\mathcal{L}|L\prec K\}\neq\emptyset$).

Proof It follows immediately from Theorem 3. □

Example 2 Table 2 is an example of a consistent generalized multi-scale decision table $S=(U,C\cup\{d\})$, where $U=\{x_1,x_2,\dots,x_8\}$, $C=\{a_1,a_2\}$, attribute a_1 has three levels of scale while a_2 has two. Thus, (U,C) has six scale combinations $L_1=(1,1)$, $L_2=(2,1)$, $L_3=(3,1)$, $L_4=(1,2)$, $L_5=(2,2)$, and $L_6=(3,2)$.

We can calculate that

$$U/R_d=\{\{x_1,x_2,x_5,x_6\},\{x_3,x_4,x_7,x_8\}\},$$
$$U/R_{C^{L_1}}=\{\{x_1,x_2\},\{x_3,x_4\},\{x_5,x_6\},\{x_7,x_8\}\}=U/R_{C^{L_4}},$$
$$U/R_{C^{L_2}}=\{\{x_1,x_2,x_5,x_6\},\{x_3,x_4\},\{x_7,x_8\}\}=U/R_{C^{L_3}}=U/R_{C^{L_5}},$$
$$U/R_{C^{L_6}}=\{\{x_1,x_2,x_5,x_6,x_7,x_8\},\{x_3,x_4\}\},$$
$$H(d|C^{L_1})=-[\frac{2}{8}(\frac{2}{2}\log_2\frac{2}{2})+\frac{2}{8}(\frac{2}{2}\log_2\frac{2}{2})+\frac{2}{8}(\frac{2}{2}\log_2\frac{2}{2})+\frac{2}{8}(\frac{2}{2}\log_2\frac{2}{2})]$$
$$=H(d|C^{L_4})=0,$$
$$H(d|C^{L_2})=-[\frac{4}{8}(\frac{4}{4}\log_2\frac{4}{4})+\frac{2}{8}(\frac{2}{2}\log_2\frac{2}{2})+\frac{2}{8}(\frac{2}{2}\log_2\frac{2}{2})]=H(d|C^{L_3})$$
$$=H(d|C^{L_5})=0,$$
$$H(d|C^{L_6})=-[\frac{6}{8}(\frac{4}{6}\log_2\frac{4}{6}+\frac{2}{6}\log_2\frac{2}{6})+\frac{2}{8}(\frac{2}{2}\log_2\frac{2}{2})]=0.689>H(d|C^{L_5}),$$

It is obvious that $R_{C^{L_1}}=R_{C^{L_4}}\subseteq R_d$, $R_{C^{L_2}}=R_{C^{L_3}}=R_{C^{L_5}}\subseteq R_d$, and $R_{C^{L_6}}\nsubseteq R_d$. So S^{L_2}, S^{L_3}, S^{L_4} and S^{L_5} are consistent to S, and they are also entropy consistent to S. $L_5=(2,2)$ is a scale combination such that S^{L_5} is consistent to S, $L_6=(3,2)$ is the unique scale combination which is strictly stronger than L_5. Notice that L_6 is a scale combination such that S^{L_6} is not consistent to S. Therefore, $L_5=(2,2)$ is an optimal scale combination of S, it is also an entropy optimal scale combination of S.

4.3 Optimal scale combination selection in inconsistent generalized multi-scale decision tables

In this subsection, we will discuss entropy based optimal scale combinations in inconsistent generalized multi-scale decision tables.

For an inconsistent generalized multi-scale decision table $S=(U,C\cup\{d\})=(U,\{a_j^k|k=1,2,\dots,I_j,j=1,2,\dots,m\}\cup\{d\})$ and $L\in\mathcal{L}$, if $(U,C^L\cup\{d\})$ is an inconsistent decision table, then

Table 2 A consistent generalized multi-scale decision table

U	a_1^1	a_1^2	a_1^3	a_2^1	a_2^2	d
x_1	1	S	M	1	Y	1
x_2	1	S	M	1	Y	1
x_3	2	H	Y	2	N	2
x_4	2	H	Y	2	N	2
x_5	3	S	M	1	Y	1
x_6	3	S	M	1	Y	1
x_7	4	L	M	3	Y	2
x_8	4	L	M	3	Y	2

it can easily be observed that $(U, C^K \cup \{d\})$ is also an inconsistent decision table for all $K \in \mathcal{L}$ with $L \prec K$. We denote

$$\partial_{C^L}(x) = \{d(y) | y \in [x]_{C^L}\},\ x \in U,$$

then $\partial_{C^L}(x)$ is the generalized decision value of x in decision table $(U, C^L \cup \{d\})$ which is determined by the scale combination L.

Remark 3 For an inconsistent decision table $(U, C^L \cup \{d\})$, where $L \in \mathcal{L}$, $\partial_{C^{L_0}}$ is the generalized decision under the first (finest) level of scale combination L_0, denote

$$R_{\partial_{C^{L_0}}} = \{(x, y) \in U \times U | \partial_{C^{L_0}}(x) = \partial_{C^{L_0}}(y)\}.$$

Then $R_{\partial_{C^{L_0}}}$ is an equivalence relation induced by the generalized decision $\partial_{C^{L_0}}$. It can be proved that $R_{C^{L_0}} \subseteq R_{\partial_{C^{L_0}}}$.

Now, we use the generalized decision $\partial_{C^{L_0}}$ to replace the decision d in decision table $(U, C^L \cup \{d\})$, then it can be verified that $(U, C^L \cup \{\partial_{C^{L_0}}\})$ is a consistent decision table.

Definition 13 For a generalized multi-scale decision table $S = (U, C \cup \{d\}) = (U, \{a_j^k | k = 1, 2, \ldots, I_j, j = 1, 2, \ldots, m\} \cup \{d\})$ and $L \in \mathcal{L}$, we say that

(1) $S^L = (U, C^L \cup \{d\})$ is generalized decision consistent to S if $\partial_{C^L}(x) = \partial_{C^{L_0}}(x)$ for all $x \in U$. And, L is said to be a generalized decision optimal scale combination of S if S^L is generalized decision consistent to S and S^K is not generalized decision consistent to S for all $K \in \mathcal{L}$ with $L \prec K$ (if $\{K \in \mathcal{L} | L \prec K\} \neq \emptyset$).

(2) $S^L = (U, C^L \cup \{d\})$ is entropy consistent to S if $H(d|C^L) = H(d|C^{L_0})$. And, L is said to be an entropy optimal scale combination of S if S^L is entropy consistent to S and S^K is not entropy consistent to S for all $K \in \mathcal{L}$ with $L \prec K$ (if $\{K \in \mathcal{L} | L \prec K\} \neq \emptyset$).

(3) $S^L = (U, C^L \cup \{d\})$ is generalized decision entropy consistent to S if $H(\partial_{C^{L_0}} | C^L) = H(\partial_{C^{L_0}} | C^{L_0})$. And, L is said to be a generalized decision entropy optimal scale combination of S if S^L is generalized decision entropy consistent to S and S^K is not generalized decision entropy consistent to S for all $K \in \mathcal{L}$ with $L \prec K$ (if $\{K \in \mathcal{L} | L \prec K\} \neq \emptyset$).

Example 3 Table 3 is an example of an inconsistent generalized multi-scale decision table $S = (U, C \cup \{d\})$, where $U = \{x_1, x_2, \ldots, x_8\}$, $C = \{a_1, a_2\}$, attribute a_1 has three levels of scale while a_2 has two. Thus, (U, C) has six scale combinations $L_1 = (1, 1)$, $L_2 = (2, 1)$, $L_3 = (3, 1)$, $L_4 = (1, 2)$, $L_5 = (2, 2)$, and $L_6 = (3, 2)$. Columns 8-10 in Table 3 are the generalized decision values with scale combinations L_1, L_5 and L_6, respectively.

We can calculate that

$$U/R_d = \{\{x_1, x_4, x_7\}, \{x_2, x_3, x_5, x_6, x_8\}\},$$

$$U/R_{C^{L_1}} = \{\{x_1, x_2\}, \{x_3, x_4, x_5\}, \{x_6\}, \{x_7\}, \{x_8\}\} = U/R_{C^{L_2}} = U/R_{C^{L_4}},$$

$$U/R_{C^{L_3}} = \{\{x_1, x_2\}, \{x_3, x_4, x_5\}, \{x_6\}, \{x_7, x_8\}\},$$

$$U/R_{C^{L_5}} = \{\{x_1, x_2, x_3, x_4, x_5\}, \{x_6\}, \{x_7\}, \{x_8\}\},$$

$$U/R_{C^{L_6}} = \{\{x_1, x_2, x_3, x_4, x_5\}, \{x_6\}, \{x_7, x_8\}\},$$

$$H(d|C^{L_1}) = -\frac{2}{8}(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}) - \frac{3}{8}(\frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log_2\frac{2}{3}) - \frac{1}{8}(\frac{1}{1}\log_2\frac{1}{1})$$
$$- \frac{1}{8}(\frac{1}{1}\log_2\frac{1}{1}) - \frac{1}{8}(\frac{1}{1}\log_2\frac{1}{1}) = 0.594 = H(d|C^{L_2}) = H(d|C^{L_4}),$$

$$H(d|C^{L_3}) = -\frac{2}{8}(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}) - \frac{3}{8}(\frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log_2\frac{2}{3})$$
$$- \frac{2}{8}(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}) - \frac{1}{8}(\frac{1}{1}\log_2\frac{1}{1}) = 0.844,$$

$$H(d|C^{L_5}) = -\frac{5}{8}(\frac{2}{5}\log_2\frac{2}{5} + \frac{3}{5}\log_2\frac{3}{5}) - \frac{1}{8}(\frac{1}{1}\log_2\frac{1}{1}) - \frac{1}{8}(\frac{1}{1}\log_2\frac{1}{1}) - \frac{1}{8}(\frac{1}{1}\log_2\frac{1}{1})$$
$$= 0.607,$$

$$H(d|C^{L_6}) = -\frac{5}{8}(\frac{2}{5}\log_2\frac{2}{5} + \frac{3}{5}\log_2\frac{3}{5}) - \frac{1}{8}(\frac{1}{1}\log_2\frac{1}{1}) - \frac{2}{8}(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2})$$
$$= 0.857,$$

It can be seen that scale combinations $L_2 = (2, 1)$ and $L_4 = (1, 2)$ are entropy consistent to S, and those combinations which are strictly stronger than L_2 and L_4 are not entropy consistent to S. So both $L_2 = (2, 1)$ and $L_4 = (1, 2)$ are entropy optimal scale combinations of S. On the other hand, we can calculate that $\partial_{C^{L_1}}(x_i) = \partial_{C^{L_2}}(x_i) = \partial_{C^{L_4}}(x_i) = \partial_{C^{L_5}}(x_i)$ for all $x_i \in U$. That is, decision tables S^{L_2}, S^{L_4}, and S^{L_5} are generalized decision consistent to S. Notice that $L_6 = (3, 2)$ is the unique scale combination which is strictly stronger than L_5 and S^{L_6} is not consistent to S. Therefore, $L_5 = (2, 2)$ is a generalized decision optimal scale combination of S.

Table 3 An inconsistent generalized multi-scale decision table

U	a_1^1	a_1^2	a_1^3	a_2^1	a_2^2	d	$\partial_{C^{L_1}}$	$\partial_{C^{L_5}}$	$\partial_{C^{L_6}}$
x_1	1	S	Y	1	Y	1	{1, 2}	{1, 2}	{1, 2}
x_2	1	S	Y	1	Y	2	{1, 2}	{1, 2}	{1, 2}
x_3	2	S	Y	2	Y	2	{1, 2}	{1, 2}	{1, 2}
x_4	2	S	Y	2	Y	1	{1, 2}	{1, 2}	{1, 2}
x_5	2	S	Y	2	Y	2	{1, 2}	{1, 2}	{1, 2}
x_6	3	L	N	1	Y	2	{2}	{2}	{2}
x_7	3	L	N	3	N	1	{1}	{1}	{1, 2}
x_8	4	M	N	3	N	2	{2}	{2}	{1, 2}

Example 3 indicates that, in an inconsistent generalized multi-scale decision table, the concept of generalized decision optimal scale combination is not equivalent to the one of entropy optimal scale combination.

Theorem 5 *Let* $S=(U,C\cup\{d\})=(U,\{a_j^k|k=1,2,\dots,I_j,$ $j=1,2,\dots,m\}\cup\{d\})$ *be an inconsistent generalized multi-scale decision table and* $L\in\mathcal{L}$, *then*

(1) S^L *is generalized decision consistent to* S *if and only if* S^L *is generalized decision entropy consistent to* S.
(2) L *is a generalized decision optimal scale combination of* S *if and only if* L *is a generalized decision entropy optimal scale combination of* S.

Proof

(1) Denote

$$S_\partial=(U,C\cup\{\partial_{C^{L_0}}\})=(U,\{a_j^k|k=1,2,\dots,I_j,j=1,2,\dots,m\}\cup\{\partial_{C^{L_0}}\}),$$

Then, by Remark 3, we see that S_∂ is a consistent generalized multi-scale decision table. Moreover, it can be verified that an inconsistent decision table S^L is generalized decision consistent to the inconsistent generalized multi-scale decision table S if and only if the corresponding decision table S_∂^L is consistent to the consistent generalized multi-scale decision table S_∂. Thus, by Theorem 3, we conclude that S^L is generalized decision consistent to S if and only if S^L is generalized decision entropy consistent to S.
(2) It follows immediately from (1).

□

According to Theorem 5, and similar to Theorem 4, we can obtain following Theorem 6.

Theorem 6 *Let* $S=(U,C\cup\{d\})=(U,\{a_j^k|k=1,2,\dots,I_j,j=1,2,\dots,m\}\cup\{d\})$ *be an inconsistent generalized multi-scale decision table and* $L\in\mathcal{L}$, *then* L *is a generalized decision optimal scale combination of* S *if and only if the following conditions hold*:

(1) $H(\partial_{C^{L_0}}|C^L)=0$.
(2) $H(\partial_{C^{L_0}}|C^K)>0$ *for all* $K\in\mathcal{L}$ *with* $L\prec K$ (*if* $\{K\in\mathcal{L}|L\prec K\}\neq\emptyset$).

Example 4 Table 4 is an example of an inconsistent generalized multi-scale decision table $S=(U,C\cup\{d\})$, where $U=\{x_1,x_2,\dots,x_8\}$, $C=\{a_1,a_2\}$, and attribute a_1 has three levels of scale while a_2 has two. Thus, (U,C) has six scale combinations $L_1=(1,1),L_2=(2,1),L_3=(3,1),L_4=(1,2),$ $L_5=(2,2)$, and $L_6=(3,2)$. Column 8 in Table 4 is the generalized decision values with the finest scale combination L_1.

We can calculate that

$$U/R_d=\{\{x_1,x_5\},\{x_2,x_3,x_4,x_6,x_7,x_8\}\},$$
$$U/R_{\partial_{C^{L_1}}}=\{\{x_1,x_2,x_5,x_6\},\{x_3,x_4,x_7,x_8\}\},$$
$$U/R_{C^{L_1}}=\{\{x_1,x_2\},\{x_3,x_4\},\{x_5,x_6\},\{x_7,x_8\}\}=U/R_{C^{L_4}},$$
$$U/R_{C^{L_2}}=\{\{x_1,x_2,x_5,x_6\},\{x_3,x_4\},\{x_7,x_8\}\}=U/R_{C^{L_3}}=U/R_{C^{L_5}},$$
$$U/R_{C^{L_6}}=\{\{x_1,x_2,x_5,x_6,x_7,x_8\},\{x_3,x_4\}\},$$
$$H(\partial_{C^{L_1}}|C^{L_1})=-[\frac{2}{8}(\frac{2}{2}\log_2\frac{2}{2})+\frac{2}{8}(\frac{2}{2}\log_2\frac{2}{2})+\frac{2}{8}(\frac{2}{2}\log_2\frac{2}{2})+\frac{2}{8}(\frac{2}{2}\log_2\frac{2}{2})]$$
$$=H(\partial_{C^{L_1}}|C^{L_4})=0,$$
$$H(\partial_{C^{L_1}}|C^{L_2})=-[\frac{4}{8}(\frac{4}{4}\log_2\frac{4}{4})+\frac{2}{8}(\frac{2}{2}\log_2\frac{2}{2})+\frac{2}{8}(\frac{2}{2}\log_2\frac{2}{2})]=H(\partial_{C^{L_1}}|C^{L_3})$$
$$=H(\partial_{C^{L_1}}|C^{L_5})=0,$$
$$H(\partial_{C^{L_1}}|C^{L_6})=-\frac{6}{8}(\frac{4}{6}\log_2\frac{4}{6}+\frac{2}{6}\log_2\frac{2}{6})-\frac{2}{8}(\frac{2}{2}\log_2\frac{2}{2})=0.689$$
$$>H(\partial_{C^{L_1}}|C^{L_5}),$$

It is obvious that $R_{C^{L_1}}\nsubseteq R_d$, $R_{C^{L_2}}=R_{C^{L_3}}=R_{C^{L_5}}\subseteq R_{\partial_{C^{L_1}}}$, $R_{C^{L_1}}=R_{C^{L_4}}\subseteq R_{\partial_{C^{L_1}}}$, and $R_{C^{L_6}}\nsubseteq R_{\partial_{C^{L_1}}}$. Hence S^{L_2}, S^{L_3}, S^{L_4} and S^{L_5} are generalized decision consistent to S, and they are also generalized decision entropy consistent to S. We see that $L_5=(2,2)$ is a scale combination such that S^{L_5} is generalized decision consistent to S. And $L_6=(3,2)$ is the unique scale combination which is strictly stronger than L_5. Notice that S^{L_6} is not generalized decision consistent to S, thus, $L_5=(2,2)$ is a generalized decision optimal scale combination of S, it is also a generalized decision entropy optimal scale combination of S.

5 Conclusion

A generalized multi-scale information table is an extension of information system in which different numbers of scales can be used to measure different attributes. An important issue of knowledge acquisition in a given generalized multi-scale information/decision table is to choose a suitable information/decision table with some requirements

Table 4 An inconsistent generalized multi-scale decision table

U	a_1^1	a_1^2	a_1^3	a_2^1	a_2^2	d	$\partial_{C^{L_1}}$
x_1	1	S	N	1	Y	1	{1,2}
x_2	1	S	N	1	Y	2	{1,2}
x_3	2	M	Y	2	N	2	{2}
x_4	2	M	Y	2	N	2	{2}
x_5	3	S	N	1	Y	1	{1,2}
x_6	3	S	N	1	Y	2	{1,2}
x_7	4	L	N	3	Y	2	{2}
x_8	4	L	N	3	Y	2	{2}

for final decision or classification. Such a process is called optimal scale combination selection. By using information entropy defined by Shannon, we have investigated optimal scale combination selection in generalized multi-scale information tables and generalized multi-scale decision tables respectively. We have clarified that, both in a generalized multi-scale information table and a consistent generalized multi-scale decision table, a scale combination is optimal if and only if it is entropy optimal. Though, in inconsistent generalized multi-scale decision tables, there are no static relationship between entropy optimal scale combination and (classical) optimal scale combination, that is, they are different notions, we have examined that a scale combination is generalized decision entropy optimal if and only if it is generalized decision optimal. Thus the uncertainty measure of entropy can be used to characterize optimal scale combinations in generalized multi-scale information/decision tables. For future study, we will study optimal scale combination selection and attribute reduction based on entropy with rule acquisition approaches in more complicate data sets such as incomplete, ordered, or set-valued multi-scale information tables.

Acknowledgements This work was supported by grants from the National Natural Science Foundation of China (grant numbers 61976194, 41631179, 62076221 and 61773349) and the Zhejiang Provincial Natural Science Foundation of China (grant number LY18F030017).

References

1. Beaubouef T, Petry FE, Arora G (1998) Information-theoretic measures of uncertainty for rough sets and rough relational databases. Inf Sci 109:185–195
2. Bianucci B, Cattaneo G, Ciucci D (2007) Entropies and co-entropies of coverings with application to incomplete information systems. Fundam Inf 75:77–105
3. Chen YS, Li JJ, Huang JX (2019) Matrix method for the optimal scale selection of multi-scale information decision systems. Mathematics 7:290
4. Chen PCL, Zhang CY (2014) Data intensive applications, challenges, techniques and technologies: a survey on big data. Inf Sci 275:314–347
5. Gu SM, Wu WZ (2013) On knowledge acquisition in multi-scale decision systems. Int J Mach Learn Cybern 4:477–486
6. Hao C, Li JH, Fan M, Liu WQ, Tsang ECC (2017) Optimal scale selection in dynamic multi-scale decision tables based on sequential three-way decisions. Inf Sci 415:213–232
7. Huang ZH, Li JJ, Dai WZ, Lin RD (2019) Generalized multi-scale decision tables with multi-scale decision attributes. Int J Approx Reason 115:194–208
8. Huang B, Li HX, Feng GF, Zhou XZ (2019) Dominance-based rough sets in multi-scale intuitionistic fuzzy decision tables. Appl Math Comput 348:487–512
9. Huang B, Wu WZ, Yan JJ, Li HX, Zhou XZ (2020) Inclusion measure-based multi-granulation decision-theoretic rough sets in multi-scale intuitionistic fuzzy information tables. Inf Sci 507:421–448
10. Komorowski J, Pawlak Z, Polkowski L, Skowron A (1999) Rough sets: tutorial. In: Pal SK, Skowron A (eds) Rough fuzzy hybridization, a new trend in decision making. Springer, Berlin, pp 3–98
11. Li F, Hu BQ (2017) A new approach of optimal scale selection to multi-scale decision tables. Inf Sci 381:193–208
12. Li F, Hu BQ, Wang J (2017) Stepwise optimal scale selection for multi-scale decision tables via attribute significance. Knowl-Based Syst 129:4–16
13. Li JH, Ren Y, Mei CL, Qian YH, Yang XB (2016) A comparative study of multigranulation rough sets and concept lattices via rule acquisition. Knowl-Based Syst 91:152–164
14. Li ZW, Zhang GQ, Wu WZ, Xie NX (2020) Measures of uncertainty for knowledge bases. Knowl Inf Syst 62:611–637
15. Liang JY, Li DY (2005) Uncertainty and knowledge acquisition in information systems (in Chinese). Science Press, Beijing
16. Liang JY, Qian YH, Li DY, Hu QH (2015) Theory and method of granular computing for big data mining. Sci Sin Inf 45:1355–1369
17. Liang JY, Shi ZZ (2004) The information entropy, rough entropy and knowledge granulation in rough set theory. Int J Uncertain, Fuzz Knowl-Based Syst 12:37–46
18. Luo C, Li TR, Huang YY, Fujita H (2019) Updating three-way decisions in incomplete multi-scale information systems. Inf Sci 476:274–289
19. Miao DQ, Li DY, Yao YY et al (2011) Uncertainty and granular computing. Science Press, Beijing
20. Miao DQ, Wang J (2011) On the relationships between information entropy and roughness of knowledge in rough set theory. Pattern Recognit Artif Intell 11:34–40
21. Pawlak Z (1991) Rough sets: theoretical aspects of reasoning about data. Kluwer Academic Publishers, Boston
22. Qian YH, Liang JY, Dang CY (2010) Incomplete multigranulation rough set. IEEE Trans Syst, Man Cybern-Part A 40:420–431
23. Qian YH, Liang JY, Yao YY, Dang CY (2010) MGRS: a multi-granulation rough set. Inf Sci 180:949–970
24. Shannon CE (1948) The mathematical theory of communication. Bell Syst Tech J 27:373–423
25. She YH, Li JH, Yang HL (2015) A local approach to rule induction in multi-scale decision tables. Knowl-Based Syst 89:398–410
26. Sun L, Zhang XY, Qian YH, Xu JC, Zhang SG (2019) Feature selection using neighborhood entropy-based uncertainty measures for gene expression data classification. Inf Sci 502:18–41
27. Wang GY (2003) Rough reduction in algebra view and information view. Int J Intell Syst 18:679–688
28. Wang GY, Yu H, Yang DC (2002) Decision table reduction based on conditional information entropy. Chin J Comput 25:759–766
29. Wu WZ, Chen CJ, Li TJ, Xu YH (2016) Comparative study on optimal granularities in inconsistent multi-granular labeled decision systems. Pattern Recognit Artif Intell 29:1103–1111
30. Wu WZ, Leung Y (2020) A comparison study of optimal scale combination selection in generalized multi-scale decision tables. Int J Mach Learn Cyber 11:961–972
31. Wu WZ, Leung Y (2011) Theory and applications of granular labelled partitions in multi-scale decision tables. Inf Sci 181:3878–3897
32. Wu WZ, Leung Y (2013) Optimal scale selection for multi-scale decision tables. Int J Approx Reason 54:1107–1129
33. Wu WZ, Qian YH, Li TJ, Gu SM (2017) On rule acquisition in incomplete multi-scale decision tables. Inf Sci 378:282–302
34. Xie JP, Yang MH, Li JH, Zheng Z (2018) Rule acquisition and optimal scale selection in multi-scale formal decision contexts and their applications to smart city. Future Gener Comput Syst 83:564–581

35. Xu YH, Wu WZ, Tan AH (2017) Optimal scale selections in consistent generalized multi-scale decision tables. In: Proceedings of International Joint Conference on Rough Sets, July 3-7, 2017, Olsztyn, Poland. Lecture Notes in Artificial Intelligence. Springer, Berlin, 10313: 185–198
36. Zadeh LA (1997) Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets Syst 90:111–127
37. Zhang GQ, Li ZW, Wu WZ, Liu XF, Xie NX (2018) Information structures and uncertainty measures in a fully fuzzy information system. Int J Approx Reason 101:119–149
38. Zhang QH, Wang GY, Hu J (2013) Multigranularity knowledge acquisition and uncertainty measurement (in Chinese). Science Press, Beijing
39. Zhang XQ, Zhang QH, Cheng YL, Wang GY (2020) Optimal scale selection by integrating uncertainty and cost-sensitive learning in multi-scale decision tables. Int J Mach Learn Cybern 11:1095–1114