International Journal of Machine Learning and Cybernetics (2022) 13:1115–1129
https://doi.org/10.1007/s13042-021-01438-x

ORIGINAL ARTICLE

Evidence theory based optimal scale selection for multi-scale ordered decision systems

Jia-Wen Zheng[1,2] · **Wei-Zhi Wu**[1,2] · **Han Bao**[1,2] · **An-Hui Tan**[1,2]

Received: 2 June 2021 / Accepted: 26 September 2021 / Published online: 23 October 2021

Abstract

In real data sets, objects are usually measured by multiple scales under the same attribute. Many information systems are given dominance relations on account of various factors which make classical equivalence relations change accordingly. This paper investigates the optimal scale selection for multi-scale ordered decision systems based on evidence theory. Five concepts of optimal scales related to rough set theory and the Dempster–Shafer theory of evidence in multi-scale ordered information/decision systems are first defined. Relationships are then clarified among ≥-optimal scale, ≥-lower approximation and ≥-upper approximation optimal scales as well as ≥-belief and ≥-plausibility optimal scales in multi-scale ordered information systems and consistent multi-scale ordered decision systems respectively. Finally, in inconsistent multi-scale ordered decision systems, by introducing a notion of ≥-generalized decision optimal scale, relationships among different types of optimal scales are also examined.

Keywords Evidence theory · Granular computing · Multi-scale ordered decision systems · Rough sets

1 Introduction

Granular computing (GrC), which imitates human being's thinking, is a very active research direction in the field of data mining and intelligent information processing [3, 4, 18, 20]. Among many research methods, the theory of rough sets, proposed by Pawlak, plays an important role in promoting and developing granular computing [21, 23–25]. This theory has been used to deal with imprecise and uncertain knowledge in intelligent systems, and has been widely applied in medical diagnosis, geological remote sensing, machine learning, data analysis, pattern recognition, expert systems and so on.

The data sets processed in rough set theory are called information systems (information tables or object-attribute value tables). An attribute corresponding to each object in an information system analyzed by the traditional rough set only takes on a unique value. Such an information system reflects the object information at a fixed scale, which is called a single-scale information system. However, in practical problems, multi-scale data widely exists, for example, there are percentile systems and grade systems in test scores, the general grade of a paper can be specified by impact factors or the grade of journal where it publishes. People may decide which level of data analysis is suitable according to their own reality. This phenomenon is caused by the difference in cost and effect of data display at different scale levels. As we all known, cost of data presentation at coarse scale level is low because coarse scale generally means that data provides less information, while the cost of data presentation at fine granularity level is high since fine scale usually implies that data provides more information. Based on this observation, Wu and Leung [31] introduced a novel rough set data analysis model named Wu–Leung model in [17]. In this model, the notion of a multi-scale information system is used to represent data set having different levels of

✉ Wei-Zhi Wu
wuwz@zjou.edu.cn

Jia-Wen Zheng
zjw11101110@163.com

An-Hui Tan
tananhui86@163.com

1 School of Information Engineering, Zhejiang Ocean University, Zhoushan 316022, Zhejiang, People's Republic of China

2 Key Laboratory of Oceanographic Big Data Mining and Application of Zhejiang Province, Zhejiang Ocean University, Zhoushan 316022, Zhejiang, People's Republic of China

granulations. Since then approach to reduce knowledge and induce decision rules on this model has become an attractive topic in rough set theory [1, 14, 16, 19, 29, 32, 34, 37, 39]. In such a model, an important issue is to select an optimal scale which connects with a single decision system from the given multi-scale decision system for final classification or decision-making. In fact, Wu and Leung investigated optimal scale selection in multi-scale decision systems by employing the standard Pawlak rough set model and a dual probabilistic rough set model respectively [32]. Xie et al. proposed three types of rules and corresponding optimal scale selections in multi-scale formal decision contexts and applied them to smart city [34]. Huang et al. employed multi-scale decision information into generalized multi-scale decision systems [16]. Huang et al. developed two novel types of multi-granulation decision-theoretic rough sets on the basis of intuitionistic fuzzy inclusion measures for acquiring knowledge from multi-scale intuitionistic fuzzy information tables [14]. Zhan et al. established group decision-making idea on Wu-Leung multi-scale information systems from the perspective of multi-expert group decision-making [37]. Li et al. constructed a rough set model based on multi-scale covering [19].

In real life applications, an information system may involve criteria, that is, attributes with preference-ordered domains. In such a case, the original rough set theory is not effective for discovering and unraveling knowledge in the data set. To address this issue, Greco et al. proposed the dominance-based rough set approach (DRSA) which is an extension of Pawlak rough set data analysis model [8–11]. The DRSA replaces indiscernibility relations with dominance ones, in which conditional attributes are criteria and decision attributes have preference relations. Unlike an indiscernibility relation which gives a partition of objects, a dominance relation renders a covering of objects. In [26], Shao and Zhang made use of the concept of dominance relation to discuss attribute reduction and to derive decision rules in incomplete ordered information systems. Susmaga introduced reducts and constructs in the context of DRSA [28]. Hu and Zhang explored dynamic dominance-based multi-granulation rough sets approaches to ordered data [13].

Dempster–Shafer theory of evidence is another significant method for dealing with uncertainty in information systems [5, 27]. A basic representation in this theory is a belief structure which consists of a family of subsets called focal elements, with associated individual positive weights summing to one. Two important numerical measures derived from the belief structure are a dual pair of belief and plausibility functions. Rough set theory and Dempster–Shafer evidence theory are closely related. Various belief structures are associated with various rough approximation spaces such that the dual pairs of different lower and upper approximation operators induced by rough approximation spaces can be used to interpret the corresponding belief and plausibility functions induced by belief structures. Yao and Lingras examined algebraic aspects involved in the interpretation belief functions in the theory of rough sets in the classical environment [36]. Wu et al. established the relationship between rough set theory and the Dempster–Shafer theory of evidence for various situations [33]. Zhang et al. were the first ones to define concepts of belief reduct and plausibility reduct in complete information systems [38]. Chen et al. used evidence theory to develop numerical algorithms of attribute reduction with neighborhood-covering rough sets [2]. Huang et al. employed the DRSA to process composite ordered decision systems including categorical, numerical, set-valued, interval-valued, and missing attributes, and dynamic maintenance of lower and upper approximations under the attribute generalization [15].

Belief and plausibility functions in the evidence theory have also been used to characterize attribute reductions in ordered information systems. For example, Xu et al. employed the evidence theory to attribute reduction in ordered information systems without decision [35]. Du and Hu further explored evidence theory to complete ordered decision systems and incomplete ordered decision systems respectively [6, 7]. Wu et al. introduced concept of multi-scale ordered decision systems using multi-scale ordered granular labeled structures [30]. And, Gu et al. discussed knowledge approximations in multi-scale ordered information systems [12]. However, the idea of the optimal scale selection has not been taken into account to multi-scale ordered information systems, which is the main objective of our study. Since evidence theory can be used to apply knowledge acquisition in data set, see e.g., Lingras and Yao [22] employed two different generalizations of rough set models to generate plausibilistic rules with incomplete databases instead of probabilistic rules generated by a Pawlak's rough set model with complete decision tables, before knowledge discovery in multi-scale ordered decision systems, the study of optimal scale selection characterized by the Dempster-Shafer of evidence becomes a necessity.

To facilitate our discussion, some preliminary notions relevant to ordered information systems and ordered decision systems are reviewed in the next section. In Sect. 3, concepts of multi-scale ordered information/decision systems with basic properties are introduced, and belief and plausibility functions in multi-scale ordered information systems are constructed. In Sect. 4, several types of optimal scales in multi-scale ordered information systems are defined and their relationships are examined. In Sect. 5, $\geq$-belief and $\geq$-plausibility optimal scales in consistent and inconsistent multi-scale ordered decision systems are respectively investigated and relationships among new optimal scales and some existing ones are clarified. We then summarize the paper with an outlook for further study in Sect. 6.

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. $\sim X = U - X = \{x \in U | x \notin X\}$.

An information system is a 2-tuple (U, A), where $U = \{x_1, x_2, \dots, x_n\}$ is a non-empty and finite set of objects called the universe of discourse, and $A = \{a_1, a_2, \dots, a_m\}$ 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.

In an information system (U, A), if the domain of attribute a is ordered according to decreasing or increasing preference, then the attribute a is a criterion. Let us define the pre-ordered relation by an outranking relation $\geq_a$ on the domain of a, then $y \geq_a x$ means that y is at least as good as x with respect to criterion a, or say that y dominates x. For a subset of attributes $B \subseteq A$, $y \geq_B x$ means that $y \geq_a x$ for all $a \in B$, and that is to say y dominates x with respect to all attributes in B.

Definition 1 [26] (U, A) is called an ordered information system if all attributes are criteria.

Let (U, A) be an ordered information system, for $B \subseteq A$, denote

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

$R_B^{\geq}$ is called a dominance relation about attribute set B, $(y, x) \in R_B^{\geq}$ means that y dominates x with respect to B.

The inverse relation of $R_B^{\geq}$ is denoted by $R_B^{\leq}$:

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

The granules of knowledge induced by the dominance relation $R_B^{\geq}$ are the set of objects dominating x,

$$[x]_B^{\geq} = \{y \in U | (y, x) \in R_B^{\geq}\} = \{y \in U | a(y) \geq a(x), \forall a \in B\}.$$

And the granules of knowledge induced by the dominance relation $R_B^{\leq}$ are the set of objects dominated by x,

$$[x]_B^{\leq} = \{y \in U | (y, x) \in R_B^{\leq}\} = \{y \in U | a(y) \leq a(x), \forall a \in B\}.$$

Then $[x]_B^{\geq}$ and $[x]_B^{\leq}$ are referred to as a dominating set and a dominated set with respect to B respectively.

Denote

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

then $U/R_B^{\geq}$ constitutes a covering of U, that is to say for every $x \in U$, we have that $[x]_B^{\geq} \neq \emptyset$ and $\bigcup_{x \in U} [x]_B^{\geq} = U$. Similarly, $U/R_B^{\leq}$ also forms a covering of U.

For simplicity and without any loss of generality, in what follows we only consider dominance relation with increasing preferences and discuss the properties of dominating set.

For $X \subseteq U$, and $B \subseteq A$, the lower and upper approximations of X with respect to the dominance relation $R_B^{\geq}$ are defined as follows [9, 10]:

$$\underline{R_B^{\geq}}(X) = \{x \in U | [x]_B^{\geq} \subseteq X\},$$

$$\overline{R_B^{\geq}}(X) = \{x \in U | [x]_B^{\geq} \cap X \neq \emptyset\}.$$

Definition 2 [10] $(U, C \cup \{d\})$ is referred to as an ordered decision system if all conditional attributes are criteria and $d \notin C$ is an overall preference called the decision, that is, the domain of decision criterion is sorted in increasing or decreasing order.

Denote

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

then $R_d^{\geq}$ is a dominance relation with respect to decision criterion d. If $(y, x) \in R_d^{\geq}$, then y dominates x. Denote

$$[x]_d^{\geq} = \{y \in U | (y, x) \in R_d^{\geq}\} = \{y \in U | d(y) \geq d(x)\},$$

then $[x]_d^{\geq}$ can be interpreted as a set of objects whose decision values are at least as good as x in the universe of discourse, and is referred to as a decision dominance class for short. And $U/R_d^{\geq} = \{[x]_d^{\geq} | x \in U\}$ is used to represent the set of knowledge granules related to the dominance relation, then any element in $U/R_d^{\geq}$ is a decision dominance class.

An ordered decision system $S = (U, C \cup \{d\})$ is referred to as $\geq$-consistent if $R_C^{\geq} \subseteq R_d^{\geq}$, otherwise it is $\geq$-inconsistent.

3 Multi-scale ordered information systems with belief structures

In this section, we mainly review basic concepts of multi-scale ordered information systems and evidence theory to prepare for the following research.

3.1 Multi-scale ordered information systems

Definition 3 [31] A tuple $S = (U, A)$ is called a multi-scale information system, where $U = \{x_1, x_2, \dots, x_n\}$ is a non-empty and finite set of objects called the universe of discourse, and $A = \{a_1, a_2, \dots, a_m\}$ 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.

A basic assumption on the multi-scale information system $S=(U,A)$ is that all attributes possess the same number I of scales. Such a system can be represented as a table $(U,\{a_j^k \mid k=1,2,\dots,I,j=1,2,\dots,m\})$, 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 function.

Denote $A^k = \{a_j^k \mid j=1,2,\dots,m\}$, $k=1,2,\dots,I$, then a multi-scale information system $S=(U,A)$ can be decomposed into I information systems $S^k=(U,A^k)$, $k=1,2,\dots,I$.

Definition 4 [32] Let U be a nonempty set, and $\mathcal{A}_1$ and $\mathcal{A}_2$ two partitions of U. If for each $A_1 \in \mathcal{A}_1$, there exist $A_2 \in \mathcal{A}_2$ such that $A_1 \subseteq A_2$, then we say that $\mathcal{A}_1$ is finer than $\mathcal{A}_2$ or $\mathcal{A}_2$ is coarser than $\mathcal{A}_1$, and is denoted as $\mathcal{A}_1 \sqsubseteq \mathcal{A}_2$. Furthermore, if there exist $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$ such that $A_1 \subset A_2$, then we say that $\mathcal{A}_1$ is strictly finer than $\mathcal{A}_2$, and is denoted as $\mathcal{A}_1 \sqsubset \mathcal{A}_2$.

Definition 5 [12, 30] For a multi-scale information system (U,A), if (U,A^1) is an ordered information system and granular information transformation functions $g_j^{1,2}$, $j=1,2,\dots,m$, are order preserving, then (U,A^2) is also an ordered information system. If $g_j^{k,k+1}$, $j=1,2,\dots,m$, $k=1,2,\dots,I-1$, are order preserving, then (U,A^k), $k=2,3,\dots,I$, are ordered information systems. If (U,A^k), $k=1,2,\dots,I$, are all ordered information systems, then (U, A) is referred to as a multi-scale ordered information system.

Proposition 1 [12, 30] *Let* $S=(U,A)=(U,\{a_j^k \mid k=1,2,\dots,I,j=1,2,\dots,m\})$ *be a multi-scale ordered information system,* $B \subseteq A$, *and* $k \in \{1,2,\dots,I\}$, *denote*

$$R_{B^k}^{\geq} = \{(y,x) \in U \times U \mid a^k(y) \geq a^k(x), \forall a \in B\},$$

$$[x]_{B^k}^{\geq} = \{y \in U \mid (y,x) \in R_{B^k}^{\geq}\} = \{y \in U \mid a^k(y) \geq a^k(x), \forall a \in B\},$$

$$U/R_{B^k}^{\geq} = \{[x]_{B^k}^{\geq} \mid x \in U\},$$

then

1. $R_{B^1}^{\geq} \subseteq R_{B^2}^{\geq} \subseteq \cdots \subseteq R_{B^I}^{\geq}$,
2. $[x]_{B^1}^{\geq} \subseteq [x]_{B^2}^{\geq} \subseteq \cdots \subseteq [x]_{B^I}^{\geq}$, $x \in U$,
3. $U/R_{B^1}^{\geq} \sqsubseteq U/R_{B^2}^{\geq} \sqsubseteq \cdots \sqsubseteq U/R_{B^I}^{\geq}$.

Definition 6 Let $S=(U,A)=(U,\{a_j^k \mid k=1,2,\dots,I,j=1,2,\dots,m\})$ be a multi-scale ordered information system, $X \in \mathcal{P}(U)$, and $k \in \{1,2,\dots,I\}$, the lower and upper approximations of X with respect to the dominance relation $R_{A^k}^{\geq}$ are defined as follows:

$$\underline{R_{A^k}^{\geq}}(X) = \{x \in U \mid [x]_{A^k}^{\geq} \subseteq X\},$$

$$\overline{R_{A^k}^{\geq}}(X) = \{x \in U \mid [x]_{A^k}^{\geq} \cap X \neq \emptyset\}.$$

Proposition 2 [12] *Let* $S=(U,A)=(U,\{a_j^k \mid k=1,2,\dots,I,j=1,2,\dots,m\})$ *be a multi-scale ordered information system,* $k \in \{1,2,\dots,I\}$, *and* $X,Y \in \mathcal{P}(U)$, *then the lower and upper approximations in Definition 6 satisfy the following properties*:

1. $\underline{R_{A^k}^{\geq}}(X) = \sim \overline{R_{A^k}^{\geq}}(\sim X)$, $\overline{R_{A^k}^{\geq}} = \sim \underline{R_{A^k}^{\geq}}(\sim X)$.
2. $\underline{R_{A^k}^{\geq}}(\emptyset) = \overline{R_{A^k}^{\geq}}(\emptyset) = \emptyset$, $\underline{R_{A^k}^{\geq}}(U) = \overline{R_{A^k}^{\geq}}(U) = U$.
3. $\underline{R_{A^k}^{\geq}}(X \cap Y) = \underline{R_{A^k}^{\geq}}(X) \cap \underline{R_{A^k}^{\geq}}(Y)$, $\overline{R_{A^k}^{\geq}}(X \cup Y) = \overline{R_{A^k}^{\geq}}(X) \cup \overline{R_{A^k}^{\geq}}(Y)$.
4. $\underline{R_{A^k}^{\geq}}(X \cup Y) \supseteq \underline{R_{A^k}^{\geq}}(X) \cup \underline{R_{A^k}^{\geq}}(Y)$, $\overline{R_{A^k}^{\geq}}(X \cap Y) \subseteq \overline{R_{A^k}^{\geq}}(X) \cap \overline{R_{A^k}^{\geq}}(Y)$.
5. $\underline{R_{A^k}^{\geq}}(X) \subseteq X \subseteq \overline{R_{A^k}^{\geq}}(X)$.
6. If $X \subseteq Y$, then $\underline{R_{A^k}^{\geq}}(X) \subseteq \underline{R_{A^k}^{\geq}}(Y)$, $\overline{R_{A^k}^{\geq}}(X) \subseteq \overline{R_{A^k}^{\geq}}(Y)$.
7. $\underline{R_{A^k}^{\geq}}(X) = \underline{R_{A^k}^{\geq}}(\underline{R_{A^k}^{\geq}}(X))$, $\overline{R_{A^k}^{\geq}}(X) = \overline{R_{A^k}^{\geq}}(\overline{R_{A^k}^{\geq}}(X))$.
8. $\underline{R_{A^{k+1}}^{\geq}}(X) \subseteq \underline{R_{A^k}^{\geq}}(X)$, $\overline{R_{A^k}^{\geq}}(X) \subseteq \overline{R_{A^{k+1}}^{\geq}}(X)$.

Example 1 Table 1 shows an example of a multi-scale ordered information system for scenic spot rating, $S=(U,A)=(U,\{a_j^k \mid k=1,2,3,j=1,2\})$, where $U=\{x_1,x_2,\dots,x_6\}$ represents the collection of scenic spots, $A=\{a_1,a_2\}$, attributes a_1 and a_2 all have three levels of scale, respectively represent "tour service" and "characteristic culture" among many criteria for creating scenic spots. The attribute values "very satisfactory (4)", "satisfactory (3)", "fair (2)" and "unsatisfactory (1)" represent the description of rating under the first scale. The attribute values "Good (A)", "General (B)" and "Poor (C)" indicate the description of rating under the second scale. The attribute

Table 1 A multi-scale ordered information system

U	a_1^1	a_1^2	a_1^3	a_2^1	a_2^2	a_2^3
x_1	1	C	N	2	B	N
x_2	1	C	N	4	A	Y
x_3	2	B	N	1	C	N
x_4	3	B	N	4	A	Y
x_5	4	A	Y	1	C	N
x_6	4	A	Y	3	B	N

values "Acceptable (Y)" and "Unacceptable (N)" reveal the description of rating under the third scale. The domains of attributes satisfy the following total ordered relation: $4 > 3 > 2 > 1, A > B > C, Y > N$.

The dominance classes of each scale can be calculated as follows:

The first scale:

$$[x_1]^{\geq}_{A^1} = \{x_1, x_2, x_4, x_6\}, [x_2]^{\geq}_{A^1} = \{x_2, x_4\},$$
$$[x_3]^{\geq}_{A^1} = \{x_3, x_4, x_5, x_6\},$$
$$[x_4]^{\geq}_{A^1} = \{x_4\}, [x_5]^{\geq}_{A^1} = \{x_5, x_6\}, [x_6]^{\geq}_{A^1} = \{x_6\}.$$

The second scale:

$$[x_1]^{\geq}_{A^2} = \{x_1, x_2, x_4, x_6\}, [x_2]^{\geq}_{A^2} = \{x_2, x_4\},$$
$$[x_3]^{\geq}_{A^2} = \{x_3, x_4, x_5, x_6\},$$
$$[x_4]^{\geq}_{A^2} = \{x_4\}, [x_5]^{\geq}_{A^2} = \{x_5, x_6\}, [x_6]^{\geq}_{A^2} = \{x_6\}.$$

The third scale:

$$[x_1]^{\geq}_{A^3} = U, [x_2]^{\geq}_{A^3} = \{x_2, x_4\}, [x_3]^{\geq}_{A^3} = U,$$
$$[x_4]^{\geq}_{A^3} = \{x_2, x_4\}, [x_5]^{\geq}_{A^3} = \{x_5, x_6\}, [x_6]^{\geq}_{A^3} = \{x_5, x_6\}.$$

Let $X = \{x_2, x_6\}$, then $\underline{R^{\geq}_{A^1}}(X) = \{x_6\}$, $\underline{R^{\geq}_{A^3}}(X) = \emptyset$, $\overline{R^{\geq}_{A^1}}(X) = \{x_1, x_2, x_3, x_5, x_6\}$, $\overline{R^{\geq}_{A^3}}(X) = U$. Obviously we have

$$\underline{R^{\geq}_{A^3}}(X) \subseteq \underline{R^{\geq}_{A^1}}(X) \subseteq X \subseteq \overline{R^{\geq}_{A^1}}(X) \subseteq \overline{R^{\geq}_{A^3}}(X).$$

Definition 7 A multi-scale ordered decision system is a system $(U, C \cup \{d\}) = (U, \{a^k_j | k = 1, 2, \ldots, I, \quad j = 1, 2, \ldots, m\} \cup \{d\})$, where $(U, C) = (U, \{a^k_j | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\})$ is a multi-scale ordered information system, $d \notin C$, $d : U \to V_d$, is a special single-scale attribute called the decision and the domain V_d of d is a total ordered set.

For a given multi-scale ordered decision system $S = (U, C \cup \{d\}) = (U, \{a^k_j | k = 1, 2, \ldots, I$, $j = 1, 2, \ldots, m\} \cup \{d\})$, it can be divided into I decision subsystems $S^k = (U, C^k \cup \{d\}), k = 1, 2, \ldots, I$, with the same decision criterion, where $C^k = \{a^k_j | j = 1, 2, \ldots, m\}$.

In a multi-scale ordered decision system $S = (U, C \cup \{d\})$, assume $V_d = \{1, 2, \ldots, s\}$. Denote $D_p = \{x \in U | d(x) = p\}$, then $\mathcal{D} = \{D_p | p \in V_d\}$ is a partition of U. For any $p, q \in V_d$, if $q > p$, then objects belong to D_q are preferred to objects from D_p and not the other way around. Let

$$D^{\geq}_p = \bigcup_{q \geq p} D_q, D^{\leq}_p = \bigcup_{q \leq p} D_q, p, q \in V_d.$$

The statement $x \in D^{\geq}_p$ means that x belongs to at least class D_p, and denote $\mathcal{D}^{\geq} = \{D^{\geq}_p | p \in V_d\}$. It is not difficult to see that $D^{\geq}_1 = D^{\leq}_s = U$, $D^{\geq}_s = D_s$, $D^{\leq}_1 = D_1$, $D^{\geq}_p$ and $D^{\leq}_{p-1}$ are complementary to each other for $p = 2, 3, \ldots, s$.

Definition 8 Let $S = (U, C \cup \{d\}) = (U, \{a^k_j | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\} \cup \{d\})$ be a multi-scale ordered decision system, and $k \in \{1, 2, \ldots, I\}$. Then for $1 \leq p \leq s$, the lower and upper approximations of $D^{\geq}_p$ with respect to $R^{\geq}_{C^k}$ are respectively defined as follows:

$$\underline{R^{\geq}_{C^k}}\left(D^{\geq}_p\right) = \{x \in U | [x]^{\geq}_{C^k} \subseteq D^{\geq}_p\},$$

$$\overline{R^{\geq}_{C^k}}\left(D^{\geq}_p\right) = \{x \in U | [x]^{\geq}_{C^k} \cap D^{\geq}_p \neq \emptyset\}.$$

Similarly, the lower and upper approximations of $D^{\leq}_p$ with respect to $R^{\leq}_{C^k}$ are defined as follows:

$$\underline{R^{\leq}_{C^k}}(D^{\leq}_p) = \{x \in U | [x]^{\leq}_{C^k} \subseteq D^{\leq}_p\},$$

$$\overline{R^{\leq}_{C^k}}(D^{\leq}_p) = \{x \in U | [x]^{\leq}_{C^k} \cap D^{\leq}_p \neq \emptyset\}.$$

Proposition 3 *Rough approximations* $\underline{R^{\geq}_{C^k}}(D^{\geq}_p)$, $\overline{R^{\geq}_{C^k}}(D^{\geq}_p)$, $\underline{R^{\leq}_{C^k}}(D^{\leq}_p)$, $\overline{R^{\leq}_{C^k}}(D^{\leq}_p)$ *satisfy the following properties*:

1. $\underline{R^{\geq}_{C^k}}(D^{\geq}_p) \subseteq D^{\geq}_p \subseteq \overline{R^{\geq}_{C^k}}(D^{\geq}_p)$, $\underline{R^{\leq}_{C^k}}(D^{\leq}_p) \subseteq D^{\leq}_p \subseteq \overline{R^{\leq}_{C^k}}(D^{\leq}_p)$.
2. $\underline{R^{\geq}_{C^k}}(D^{\geq}_p) = U - \overline{R^{\leq}_{C^k}}(D^{\leq}_{p-1})$, $\underline{R^{\leq}_{C^k}}(D^{\leq}_p) = U - \overline{R^{\geq}_{C^k}}(D^{\geq}_{p+1})$.
3. $\underline{R^{\geq}_{C^{k+1}}}(D^{\geq}_p) \subseteq \underline{R^{\geq}_{C^k}}(D^{\geq}_p)$, $\underline{R^{\leq}_{C^{k+1}}}(D^{\leq}_p) \subseteq \underline{R^{\leq}_{C^k}}(D^{\leq}_p)$, $\overline{R^{\geq}_{C^k}}(D^{\geq}_p) \subseteq \overline{R^{\geq}_{C^{k+1}}}(D^{\geq}_p)$, $\overline{R^{\leq}_{C^k}}(D^{\leq}_p) \subseteq \overline{R^{\leq}_{C^{k+1}}}(D^{\leq}_p)$.

3.2 Belief and plausibility functions in multi-scale ordered information systems

In this section, we review some basic notions of the Dempster-Shafer theory of evidence.

Definition 9 [5, 27] Let U be a non-empty finite universe of discourse, a set function $m : \mathcal{P}(U) \to [0, 1]$ is referred to as a mass function or a basic probability assignment if it satisfies following axioms (M1) and (M2):

(M1) $m(\emptyset) = 0$,

(M2) $\sum_{X \subseteq U} m(X) = 1$.

A set $X \in \mathcal{P}(U)$ with $m(X) > 0$ is referred to as a focal element. We denote by $\mathcal{M}$ the family of all focal elements of m. Then the pair $(\mathcal{M}, m)$ is said to be a belief structure on U.

Associated with each belief structure, a pair of belief and plausibility functions can be derived.

Definition 10 [5, 27] Let $(\mathcal{M}, m)$ be a belief structure on U. A set function $\mathrm{Bel} : \mathcal{P}(U) \to [0, 1]$ is referred to as a belief function on U if

$$\mathrm{Bel}(X) = \sum_{Y \subseteq X} m(Y), \forall X \in \mathcal{P}(U).$$

A set function $\mathrm{Pl} : \mathcal{P}(U) \to [0, 1]$ is referred to as a plausibility function on U if

$$\mathrm{Pl}(X) = \sum_{Y \cap X \neq \emptyset} m(Y), \forall X \in \mathcal{P}(U).$$

The relationships between rough set theory and evidence theory have been widely studied. The probabilities of lower and upper approximations of the universe of discourse in information systems are a dual pair of belief and plausibility measures of the set.

Theorem 1 [36] *Let $S = (U, A)$ be an information system, for any $X \subseteq U$, and $B \subseteq A$, denote*

$$\mathrm{Bel}_B(X) = P(\underline{R_B}(X)), \mathrm{Pl}_B(X) = P(\overline{R_B}(X)),$$

where $P(X) = |X|/|U|$, and $|X|$ is the cardinality of X. Then Bel_B and Pl_B are a dual pair of belief and plausibility functions on U, and the corresponding mass function is

$$m_B(Y) = \begin{cases} P(Y), & \text{if } Y \in U/R_B; \\ 0, & \text{otherwise .} \end{cases}$$

From [35], we see that the pair of lower and upper approximate operators in ordered information systems generates a pair of belief and plausibility functions respectively. Extend the result to multi-scale ordered information systems, we can conclude following theorem.

Theorem 2 *Let $S = (U, A) = (U, \{a_j^k | k = 1, 2, \dots, I, j = 1, 2, \dots, m\})$ be a multi-scale ordered information system, $k \in \{1, 2, \dots, I\}$, and $X \subseteq U$, denote*

$$\mathrm{Bel}^{\geq}_{A^k}(X) = P(\underline{R^{\geq}_{A^k}}(X)), \mathrm{Pl}^{\geq}_{A^k}(X) = P(\overline{R^{\geq}_{A^k}}(X)).$$

Then $\mathrm{Bel}^{\geq}_{A^k}$ and $\mathrm{Pl}^{\geq}_{A^k}$ are a dual pair of belief and plausibility functions on U, and the corresponding mass function is

$$m^{\geq}_{A^k}(X) = \begin{cases} |h^{\geq}_{A^k}(X)|/|U|, & \text{if } X \in U/R^{\geq}_{A^k}; \\ 0, & \text{otherwise .} \end{cases}$$

where $h^{\geq}_{A^k}(X) = \{x \in U | [x]^{\geq}_{A^k} = X\}$.

Combining Theorem 2 and Proposition 2, we have following corollary.

Corollary 1 *Let $S = (U, A) = (U, \{a_j^k | k = 1, 2, \dots, I, j = 1, 2,$ $S = (U, A) = (U, \{a_j^k | k = 1, 2, \dots, I, j = 1, 2, \dots, m\})$ be a multi-scale ordered information system, and $k \in \{1, 2, \dots, I-1\}$, then belief and plausibility functions, $\mathrm{Bel}^{\geq}_{A^k}$ and $\mathrm{Pl}^{\geq}_{A^k}$, satisfy following property:*

$$\mathrm{Bel}^{\geq}_{A^{k+1}}(X) \leq \mathrm{Bel}^{\geq}_{A^k}(X) \leq \frac{|X|}{|U|} \leq \mathrm{Pl}^{\geq}_{A^k}(X) \leq \mathrm{Pl}^{\geq}_{A^{k+1}}(X), \forall X \subseteq U.$$

4 Optimal scale selection for multi-scale ordered information systems

In this section, the concepts of $\geq$-optimal scale, $\geq$-belief and $\geq$-plausibility optimal scales as well as $\geq$-lower approximation and $\geq$-upper approximation optimal scales in multi-scale ordered information systems are defined, and their relationships are clarified.

Definition 11 Let $S = (U, A) = (U, \{a_j^k | k = 1, 2, \dots, I, j = 1, 2, \dots, m\})$ be a multi-scale ordered information system, $k \in \{1, 2, \dots, I\}$, and $U/R^{\geq}_{A^1} = \{C^{\geq}_1, C^{\geq}_2, \dots, C^{\geq}_t\}$, we say that

1. S^k is $\geq$-consistent with S if $R^{\geq}_{A^k} = R^{\geq}_{A^1}$. And, k is referred to as the $\geq$-optimal scale of S if S^k is $\geq$-consistent with S and S^{k+1} is not $\geq$-consistent with S (if $k+1 \leq I$).
2. S^k is $\geq$-lower approximation consistent with S if $\underline{R^{\geq}_{A^k}}(X) = \underline{R^{\geq}_{A^1}}(X)$ for all $X \in U/R^{\geq}_{A^1}$. And, k is referred to as the $\geq$-lower approximation optimal scale of S if S^k is $\geq$-lower approximation consistent with S and S^{k+1} is not $\geq$-lower approximation consistent with S (if $k+1 \leq I$).
3. S^k is $\geq$-upper approximation consistent with S if $\overline{R^{\geq}_{A^k}}(X) = \overline{R^{\geq}_{A^1}}(X)$ for all $X \in U/R^{\geq}_{A^1}$. And, k is referred to as the $\geq$-upper approximation optimal scale of S if S^k is $\geq$-upper approximation consistent with S and S^{k+1} is not $\geq$-upper approximation consistent with S (if $k+1 \leq I$).
4. S^k is $\geq$-belief consistent with S if $\mathrm{Bel}^{\geq}_{A^k}(X) = \mathrm{Bel}^{\geq}_{A^1}(X)$ for all $X \in U/R^{\geq}_{A^1}$. And, k is referred to as the $\geq$-belief optimal scale of S if S^k is $\geq$-belief consistent with S and S^{k+1} is not $\geq$-belief consistent with S (if $k+1 \leq I$).
5. S^k is $\geq$-plausibility consistent with S if $\mathrm{Pl}^{\geq}_{A^k}(X) = \mathrm{Pl}^{\geq}_{A^1}(X)$ for all $X \in U/R^{\geq}_{A^1}$. And, k is referred to as the $\geq$-plausibility optimal scale of S if S^k is $\geq$-plausibility consistent with S and S^{k+1} is not $\geq$-plausibility consistent with S (if $k+1 \leq I$).

By Theorem 2 and Corollary 1, we can obtain following Theorem 3 which shows that $\geq$-belief optimal scale and $\geq$-lower approximation optimal scale are equivalent concepts in multi-scale ordered information systems. Moreover, $\geq$-plausibility optimal scale and $\geq$-upper approximation optimal scale are also equivalent concepts.

Theorem 3 *Let* $S=(U,A)=(U,\{a_j^k \mid k=1,2,\dots,I, j=1,2,\dots,m\})$ *be a multi-scale ordered information system, and* $k\in\{1,2,\dots,I\}$, *then*

1. S^k is $\geq$-belief consistent with S if and only if S^k is $\geq$-lower approximation consistent with S.
2. k is the $\geq$-belief optimal scale of S if and only if k is the $\geq$-lower approximation optimal scale of S.
3. S^k is $\geq$-plausibility consistent with S if and only if S^k is $\geq$-upper approximation consistent with S.
4. k is the $\geq$-plausibility optimal scale of S if and only if k is the $\geq$-upper approximation optimal scale of S.

Theorem 4 *Let* $S=(U,A)=(U,\{a_j^k \mid k=1,2,\dots,I, j=1,2,\dots,m\})$ *be a multi-scale ordered information system, and* $k\in\{1,2,\dots,I\}$, *then*

1. S^k is $\geq$-consistent with S if and only if S^k is $\geq$-belief consistent with S.
2. k is the $\geq$-optimal scale of S if and only if k is the $\geq$-belief optimal scale of S.

Proof □

1. "$\Rightarrow$". Assume that S^k is $\geq$-consistent with S. For any $X\in U/R^{\geq}_{A^1}$, since $[x]^{\geq}_{A^k}=[x]^{\geq}_{A^1}$ for all $x\in U$, we can have

$$[x]^{\geq}_{A^k}\subseteq[x]^{\geq}_{A^1}\iff[x]^{\geq}_{A^1}\subseteq[x]^{\geq}_{A^1}.$$

Then by the definition of lower approximation, we obtain

$$x\in\underline{R^{\geq}_{A^k}}(X)\iff x\in\underline{R^{\geq}_{A^1}}(X), x\in U.$$

Therefore $\underline{R^{\geq}_{A^k}}(X)=\underline{R^{\geq}_{A^1}}(X)$ for all $X\in U/R^{\geq}_{A^1}$. By Theorem 3, it follows that $\mathrm{Bel}^{\geq}_{A^k}(X)=\mathrm{Bel}^{\geq}_{A^1}(X)$ for all $X\in U/R^{\geq}_{A^1}$, i.e. S^k is $\geq$-belief consistent with S.

"$\Leftarrow$". Assume that S^k is $\geq$-belief consistent with S, that is

$$\mathrm{Bel}^{\geq}_{A^k}(X)=\mathrm{Bel}^{\geq}_{A^1}(X),\forall X\in U/R^{\geq}_{A^1},$$

i.e.

$$\mathrm{Bel}^{\geq}_{A^k}\left([x]^{\geq}_{A^1}\right)=\mathrm{Bel}^{\geq}_{A^1}\left([x]^{\geq}_{A^1}\right),\forall x\in U,$$

then we have

$$\frac{\left|\underline{R^{\geq}_{A^k}}\left([x]^{\geq}_{A^1}\right)\right|}{|U|}=\frac{\left|\underline{R^{\geq}_{A^1}}\left([x]^{\geq}_{A^1}\right)\right|}{|U|}.$$

Since $\underline{R^{\geq}_{A^k}}([x]^{\geq}_{A^1})\subseteq\underline{R^{\geq}_{A^1}}([x]^{\geq}_{A^1})$, we obtain $\underline{R^{\geq}_{A^k}}([x]^{\geq}_{A^1})=\underline{R^{\geq}_{A^1}}([x]^{\geq}_{A^1})$ for all $x\in U$. By the definition of lower approximation, we have

$$\{y\in U\mid[y]^{\geq}_{A^k}\subseteq[x]^{\geq}_{A^1}\}=\{y\in U\mid[y]^{\geq}_{A^1}\subseteq[x]^{\geq}_{A^1}\},\forall x\in U.$$

That is to say

$$[y]^{\geq}_{A^k}\subseteq[x]^{\geq}_{A^1}\iff[y]^{\geq}_{A^1}\subseteq[x]^{\geq}_{A^1},\forall x,y\in U.$$

Let $y=x$, we then conclude

$$[x]^{\geq}_{A^k}\subseteq[x]^{\geq}_{A^1}\iff[x]^{\geq}_{A^1}\subseteq[x]^{\geq}_{A^1}.$$

Hence, we have $[x]^{\geq}_{A^k}\subseteq[x]^{\geq}_{A^1}$ for all $x\in U$. By Proposition 1, $[x]^{\geq}_{A^1}\subseteq[x]^{\geq}_{A^k}$, therefore $[x]^{\geq}_{A^k}=[x]^{\geq}_{A^1}$ for all $x\in U$, i.e. S^k is $\geq$-consistent with S.

2. It follows immediately from (1).

□

Definition 12 Let $S=(U,A)=(U,\{a_j^k \mid k=1,2,\dots,I, j=1,2,\dots,m\})$ be a multi-scale ordered information system, $k\in\{1,2,\dots,I\}$, and $U/R^{\geq}_{A^1}=\{C_1^{\geq},C_2^{\geq},\dots,C_t^{\geq}\}$, the $\geq$-belief sum and $\geq$-plausibility sum of S are defined as follows:

$$M=\sum_{i=1}^{t}\mathrm{Bel}^{\geq}_{A^1}(C_i^{\geq}), M'=\sum_{i=1}^{t}\mathrm{Pl}^{\geq}_{A^1}(C_i^{\geq}).$$

Theorem 5 *Let* $S=(U,A)=(U,\{a_j^k \mid k=1,2,\dots,I, j=1,2,\dots,m\})$ *be a multi-scale ordered information system, and* $k\in\{1,2,\dots,I\}$, *then*

1. S^k is $\geq$-consistent with S if and only if $\sum_{i=1}^{t}\mathrm{Bel}^{\geq}_{A^k}(C_i^{\geq})=M$.
2. k is the $\geq$-optimal scale of S if and only if $\sum_{i=1}^{t}\mathrm{Bel}^{\geq}_{A^k}(C_i^{\geq})=M$ and $\sum_{i=1}^{t}\mathrm{Bel}^{\geq}_{A^{k+1}}(C_i^{\geq})<M$ (if $k+1\leq I$).

Proof 1. "$\Rightarrow$". Assume that S^k is $\geq$-consistent with S, by Theorem 3, we have S^k is $\geq$-belief consistent with S, and $\mathrm{Bel}^{\geq}_{A^k}(C_i^{\geq})=\mathrm{Bel}^{\geq}_{A^1}(C_i^{\geq})$, $i=1,2,\dots,t$, therefore $\sum_{i=1}^{t}\mathrm{Bel}^{\geq}_{A^k}(C_i^{\geq})=\sum_{i=1}^{t}\mathrm{Bel}^{\geq}_{A^1}(C_i^{\geq})=M$.

"$\Leftarrow$". Assume that $\sum_{i=1}^{t}\mathrm{Bel}^{\geq}_{A^k}(C_i^{\geq})=M$, by Corollary 1, we have $\mathrm{Bel}^{\geq}_{A^k}(C_i^{\geq})\leq \mathrm{Bel}^{\geq}_{A^1}(C_i^{\geq})$, $i=1,2,\ldots,t$. Then $M=\sum_{i=1}^{t}\mathrm{Bel}^{\geq}_{A^k}(C_i^{\geq})\leq \sum_{i=1}^{t}\mathrm{Bel}^{\geq}_{A^1}(C_i^{\geq})=M$, therefore $\mathrm{Bel}^{\geq}_{A^k}(C_i^{\geq})=\mathrm{Bel}^{\geq}_{A^1}(C_i^{\geq})$, for all $i=1,2,\ldots,t$, i.e. S^k is $\geq$-belief consistent with S. By Theorem 3, S^k is $\geq$-consistent with S.
2. It follows immediately from (1) and Definition 11.

$\square$

Theorem 5 indicates that the $\geq$-optimal scale can be characterized by sum of belief measures.

Theorem 6 *Let* $S=(U,A)=(U,\{a_j^k \mid k=1,2,\ldots,I, j=1,2,\ldots,m\})$ *be a multi-scale ordered information system, and* $k\in\{1,2,\ldots,I\}$, *then*

1. S^k is $\geq$-plausibility consistent with S if and only if $\sum_{i=1}^{t}\mathrm{Pl}^{\geq}_{A^k}(C_i^{\geq})=M'$.
2. k is the $\geq$-plausibility optimal scale of S if and only if $\sum_{i=1}^{t}\mathrm{Pl}^{\geq}_{A^k}(C_i^{\geq})=M'$ and $\sum_{i=1}^{t}\mathrm{Pl}^{\geq}_{A^{k+1}}(C_i^{\geq})>M'$.

Proof It is similar to the proof of Theorem 4. $\square$

Theorem 7 *Let* $S=(U,A)=(U,\{a_j^k \mid k=1,2,\ldots,I, j=1,2,\ldots,m\})$ *be a multi-scale ordered information system, and* $k\in\{1,2,\ldots,I\}$. *If* S^k *is* $\geq$*-consistent with* S, *then* S^k *is* $\geq$*-plausibility consistent with* S.

Proof Assume that S^k is $\geq$-consistent with S. For any $X\in U/R^{\geq}_{A^1}$, since $[x]^{\geq}_{A^k}=[x]^{\geq}_{A^1}$ for all $x\in U$, we have

$$[x]^{\geq}_{A^k}\cap X\neq\emptyset \Longleftrightarrow [x]^{\geq}_{A^1}\cap X\neq\emptyset.$$

Then, by the definition of upper approximation, we obtain

$$x\in\overline{R^{\geq}_{A^k}}(X)\Longleftrightarrow x\in\overline{R^{\geq}_{A^1}}(X), \forall x\in U,$$

so $\overline{R^{\geq}_{A^k}}(X)=\overline{R^{\geq}_{A^1}}(X)$ for all $X\in U/R^{\geq}_{A^1}$. Therefore, by Theorem 3, we conclude $\mathrm{Pl}^{\geq}_{A^k}(X)=\mathrm{Pl}^{\geq}_{A^1}(X)$ for all $X\in U/R^{\geq}_{A^1}$, i.e. S^k is $\geq$-plausibility consistent with S. $\square$

Theorem 7 shows that in multi-scale ordered information systems the $\geq$-optimal scale is finer than the $\geq$-plausibility optimal scale. The converse of Theorem 7 is not always true, as we can see in the following example.

Example 2 (Continued from Example 1) Let $U/R^{\geq}_{A^1}=\{C_1^{\geq},C_2^{\geq},\ldots,C_6^{\geq}\}$, we can calculate the belief and plausibility measures as follows:

$$
\begin{aligned}
&\underline{R^{\geq}_{A^1}}(C_1^{\geq})=\underline{R^{\geq}_{A^2}}(C_1^{\geq})=\{x_1,x_2,x_4,x_6\},\underline{R^{\geq}_{A^3}}(C_1^{\geq})=\{x_2,x_4\},\\
&\underline{R^{\geq}_{A^1}}(C_2^{\geq})=\underline{R^{\geq}_{A^2}}(C_2^{\geq})=\underline{R^{\geq}_{A^3}}(C_2^{\geq})=\{x_2,x_4\},\\
&\underline{R^{\geq}_{A^1}}(C_3^{\geq})=\underline{R^{\geq}_{A^2}}(C_3^{\geq})=\{x_3,x_4,x_5,x_6\},\underline{R^{\geq}_{A^3}}(C_3^{\geq})=\{x_5,x_6\},\\
&\underline{R^{\geq}_{A^1}}(C_4^{\geq})=\underline{R^{\geq}_{A^2}}(C_4^{\geq})=\{x_4\},\underline{R^{\geq}_{A^3}}(C_4^{\geq})=\emptyset,\\
&\underline{R^{\geq}_{A^1}}(C_5^{\geq})=\underline{R^{\geq}_{A^2}}(C_5^{\geq})=\underline{R^{\geq}_{A^3}}(C_5^{\geq})=\{x_5,x_6\},\\
&\underline{R^{\geq}_{A^1}}(C_6^{\geq})=\underline{R^{\geq}_{A^2}}(C_6^{\geq})=\{x_6\},\underline{R^{\geq}_{A^3}}(C_6^{\geq})=\emptyset,\\
&\overline{R^{\geq}_{A^1}}(C_1^{\geq})=\overline{R^{\geq}_{A^2}}(C_1^{\geq})=\overline{R^{\geq}_{A^3}}(C_1^{\geq})=U,\\
&\overline{R^{\geq}_{A^1}}(C_2^{\geq})=\overline{R^{\geq}_{A^2}}(C_2^{\geq})=\overline{R^{\geq}_{A^3}}(C_2^{\geq})=\{x_1.x_2,x_3,x_4\},\\
&\overline{R^{\geq}_{A^1}}(C_3^{\geq})=\overline{R^{\geq}_{A^2}}(C_3^{\geq})=\overline{R^{\geq}_{A^3}}(C_3^{\geq})=U,\\
&\overline{R^{\geq}_{A^1}}(C_4^{\geq})=\overline{R^{\geq}_{A^2}}(C_4^{\geq})=\overline{R^{\geq}_{A^3}}(C_4^{\geq})=\{x_1.x_2,x_3,x_4\},\\
&\overline{R^{\geq}_{A^1}}(C_5^{\geq})=\overline{R^{\geq}_{A^2}}(C_5^{\geq})=\overline{R^{\geq}_{A^3}}(C_5^{\geq})=\{x_1.x_3,x_5,x_6\},\\
&\overline{R^{\geq}_{A^1}}(C_6^{\geq})=\overline{R^{\geq}_{A^2}}(C_6^{\geq})=\overline{R^{\geq}_{A^3}}(C_6^{\geq})=\{x_1.x_3,x_5,x_6\}.
\end{aligned}
$$

$$\mathrm{Bel}^{\geq}_{A^1}(C_1^{\geq})=\frac{|\underline{R^{\geq}_{A^1}}(C_1^{\geq})|}{|U|}=\mathrm{Bel}^{\geq}_{A^2}(C_1^{\geq})=\frac{2}{3}\neq\frac{1}{3}=\mathrm{Bel}^{\geq}_{A^3}(C_1^{\geq}),$$

$$\mathrm{Bel}^{\geq}_{A^1}(C_2^{\geq})=\mathrm{Bel}^{\geq}_{A^2}(C_2^{\geq})=\mathrm{Bel}^{\geq}_{A^3}(C_2^{\geq})=\frac{1}{3},$$

$$\mathrm{Bel}^{\geq}_{A^1}(C_3^{\geq})=\mathrm{Bel}^{\geq}_{A^2}(C_3^{\geq})=\frac{2}{3}\neq\frac{1}{3}=\mathrm{Bel}^{\geq}_{A^3}(C_3^{\geq}),$$

$$\mathrm{Bel}^{\geq}_{A^1}(C_4^{\geq})=\mathrm{Bel}^{\geq}_{A^2}(C_4^{\geq})=\frac{1}{6}\neq 0=\mathrm{Bel}^{\geq}_{A^3}(C_4^{\geq}),$$

$$\mathrm{Bel}^{\geq}_{A^1}(C_5^{\geq})=\mathrm{Bel}^{\geq}_{A^2}(C_5^{\geq})=\mathrm{Bel}^{\geq}_{A^3}(C_5^{\geq})=\frac{1}{3},$$

$$\mathrm{Bel}^{\geq}_{A^1}(C_6^{\geq})=\mathrm{Bel}^{\geq}_{A^2}(C_6^{\geq})=\frac{1}{6}\neq 0=\mathrm{Bel}^{\geq}_{A^3}(C_6^{\geq}),$$

$$\sum_{i=1}^{6}\mathrm{Bel}^{\geq}_{A^1}(C_i^{\geq})=\sum_{i=1}^{6}\mathrm{Bel}^{\geq}_{A^2}(C_i^{\geq})=\frac{7}{3}>\frac{4}{3}=\sum_{i=1}^{6}\mathrm{Bel}^{\geq}_{A^3}(C_i^{\geq}).$$

$$
\begin{aligned}
&\mathrm{Pl}^{\geq}_{A^1}(C_1^{\geq})=\mathrm{Pl}^{\geq}_{A^2}(C_1^{\geq})=\mathrm{Pl}^{\geq}_{A^3}(C_1^{\geq})=\frac{\left|\overline{R^{\geq}_{A^1}}(C_1^{\geq})\right|}{|U|}=1,\\
&\mathrm{Pl}^{\geq}_{A^1}(C_2^{\geq})=\mathrm{Pl}^{\geq}_{A^2}(C_2^{\geq})=\mathrm{Pl}^{\geq}_{A^3}(C_2^{\geq})=\frac{2}{3},\\
&\mathrm{Pl}^{\geq}_{A^1}(C_3^{\geq})=\mathrm{Pl}^{\geq}_{A^2}(C_3^{\geq})=\mathrm{Pl}^{\geq}_{A^3}(C_3^{\geq})=1,\\
&\mathrm{Pl}^{\geq}_{A^1}(C_4^{\geq})=\mathrm{Pl}^{\geq}_{A^2}(C_4^{\geq})=\mathrm{Pl}^{\geq}_{A^3}(C_4^{\geq})=\frac{2}{3},\\
&\mathrm{Pl}^{\geq}_{A^1}(C_5^{\geq})=\mathrm{Pl}^{\geq}_{A^2}(C_5^{\geq})=\mathrm{Pl}^{\geq}_{A^3}(C_5^{\geq})=\frac{2}{3},\\
&\mathrm{Pl}^{\geq}_{A^1}(C_6^{\geq})=\mathrm{Pl}^{\geq}_{A^2}(C_6^{\geq})=\mathrm{Pl}^{\geq}_{A^3}(C_6^{\geq})=\frac{2}{3},
\end{aligned}
$$

$$\sum_{i=1}^{6}\mathrm{Pl}^{\geq}_{A^1}(C_i^{\geq})=\sum_{i=1}^{6}\mathrm{Pl}^{\geq}_{A^2}(C_i^{\geq})=\sum_{i=1}^{6}\mathrm{Pl}^{\geq}_{A^3}(C_i^{\geq})=\frac{14}{3}.$$

It is obvious that $R^{\geq}_{A^1} = R^{\geq}_{A^2} \neq R^{\geq}_{A^3}$, so $k = 2$ is both the $\geq$-optimal scale and the $\geq$-belief optimal scale of S. However, the $\geq$-plausibility optimal scale of S is $k = 3$.

5 Optimal scale selection for multi-scale ordered decision systems

In this section, we define some concepts of optimal scales in multi-scale ordered decision systems and discuss their relationships.

Definition 13 Let $S = (U, C \cup \{d\}) = (U, \{a^k_j | k = 1, 2, \dots, I, j = 1, 2, \dots, m\} \cup \{d\})$ be a multi-scale ordered decision system, $k \in \{1, 2, \dots, I\}$, $D^{\geq}_p = \bigcup_{q \geq p} D_q$, $p, q \in V_d$, and $\mathcal{D}^{\geq} = \{D^{\geq}_p | p \in V_d\}$, we say that

1. S^k is $\geq$-lower approximation consistent with S if $\underline{R^{\geq}_{C^k}}(D^{\geq}_p) = \underline{R^{\geq}_{C^1}}(D^{\geq}_p)$ for all $D^{\geq}_p \in \mathcal{D}^{\geq}$. And, k is referred to as the $\geq$-lower approximation optimal scale of S if S^k is $\geq$-lower approximation consistent with S and S^{k+1} is not $\geq$-lower approximation consistent with S (if $k + 1 \leq I$).
2. S^k is $\geq$-upper approximation consistent with S if $\overline{R^{\geq}_{C^k}}(D^{\geq}_p) = \overline{R^{\geq}_{C^1}}(D^{\geq}_p)$ for all $D^{\geq}_p \in \mathcal{D}^{\geq}$. And, k is referred to as the $\geq$-upper approximation optimal scale of S if S^k is $\geq$-upper approximation consistent with S and S^{k+1} is not $\geq$-upper approximation consistent with S (if $k + 1 \leq I$).
3. S^k is $\geq$-belief consistent with S if $\mathrm{Bel}^{\geq}_{C^k}(D^{\geq}_p) = \mathrm{Bel}^{\geq}_{C^1}(D^{\geq}_p)$ for all $D^{\geq}_p \in \mathcal{D}^{\geq}$. And, k is referred to as the $\geq$-belief optimal scale of S if S^k is $\geq$-belief consistent with S and S^{k+1} is not $\geq$-belief consistent with S (if $k + 1 \leq I$).
4. S^k is $\geq$-plausibility consistent with S if $\mathrm{Pl}^{\geq}_{C^k}(D^{\geq}_p) = \mathrm{Pl}^{\geq}_{C^1}(D^{\geq}_p)$ for all $D^{\geq}_p \in \mathcal{D}^{\geq}$. And, k is referred to as the $\geq$-plausibility optimal scale of S if S^k is $\geq$-plausibility consistent with S and S^{k+1} is not $\geq$-plausibility consistent with S (if $k + 1 \leq I$).

According to Theorem 2 and Corollary 1, the following theorem can directly be obtained.

Theorem 8 *Let $S = (U, C \cup \{d\}) = (U, \{a^k_j | k = 1, 2, \dots, I, j = 1, 2, \dots, m\} \cup \{d\})$ be a multi-scale ordered decision system, and $k \in \{1, 2, \dots, I\}$, then*

1. S^k is $\geq$-belief consistent with S if and only if S^k is $\geq$-lower approximation consistent with S,
2. k is the $\geq$-belief optimal scale of S if and only if k is the $\geq$-lower approximation optimal scale of S.
3. S^k is $\geq$-plausibility consistent with S if and only if S^k is $\geq$-upper approximation consistent with S,
4. k is the $\geq$-plausibility optimal scale of S if and only if k is the $\geq$-upper approximation optimal scale of S.

Definition 14 Let $S = (U, C \cup \{d\}) = (U, \{a^k_j | k = 1, 2, \dots, I, j = 1, 2, \dots, m\} \cup \{d\})$ be a multi-scale ordered decision system, and $k \in \{1, 2, \dots, I\}$, the $\geq$-belief sum and $\geq$-plausibility sum of S are respectively defined as follows:

$$M = \sum_{p=1}^{s} \mathrm{Bel}^{\geq}_{C^1}\left(D^{\geq}_p\right), M' = \sum_{p=1}^{s} \mathrm{Pl}^{\geq}_{C^1}\left(D^{\geq}_p\right).$$

Theorem 9 *Let $S = (U, C \cup \{d\}) = (U, \{a^k_j | k = 1, 2, \dots, I, j = 1, 2, \dots, m\} \cup \{d\})$ be a multi-scale ordered decision system, and $k \in \{1, 2, \dots, I\}$, then*

1. S^k is $\geq$-belief consistent with S if and only if $\sum_{p=1}^{s} \mathrm{Bel}^{\geq}_{C^k}(D^{\geq}_p) = M$.
2. k is the $\geq$-belief optimal scale of S if and only if $\sum_{p=1}^{s} \mathrm{Bel}^{\geq}_{C^k}(D^{\geq}_p) = M$ and $\sum_{p=1}^{s} \mathrm{Bel}^{\geq}_{C^{k+1}}(D^{\geq}_p) < M$ (if $k + 1 \leq I$).
3. S^k is $\geq$-plausibility consistent with S if and only if $\sum_{p=1}^{s} \mathrm{Pl}^{\geq}_{C^k}(D^{\geq}_p) = M'$.
4. k is the $\geq$-plausibility optimal scale of S if and only if $\sum_{p=1}^{s} \mathrm{Pl}^{\geq}_{C^k}(D^{\geq}_p) = M'$ and $\sum_{p=1}^{s} \mathrm{Pl}^{\geq}_{C^{k+1}}(D^{\geq}_p) > M'$ (if $k + 1 \leq I$).

Proof It is similar to the proof of Theorem 4. □

5.1 Optimal scale selection in consistent multi-scale ordered decision systems

For a multi-scale ordered decision system $S = (U, C \cup \{d\}) = (U, \{a^k_j | k = 1, 2, \dots, I, j = 1, 2, \dots, m\} \cup \{d\})$, S is called $\geq$-consistent if the decision system at the first (finest) scale is $\geq$-consistent, otherwise it is $\geq$-inconsistent. In addition, for $1 \leq i \leq k \leq I$, if $S^k = (U, C^k \cup \{d\}) = (U, \{a^k_j | j = 1, 2, \dots, m\} \cup \{d\})$ is a consistent ordered decision system, that is $R^{\geq}_{C^k} \subseteq R^{\geq}_d$, by Proposition 1, we can obtain $R^{\geq}_{C^i} \subseteq R^{\geq}_{C^k} \subseteq R^{\geq}_d$, thus S^i is also a consistent ordered decision system.

Definition 15 Let $S = (U, C \cup \{d\}) = (U, \{a^k_j | k = 1, 2, \dots, I, j = 1, 2, \dots, m\} \cup \{d\})$ be a consistent multi-scale ordered decision system, and $k \in \{1, 2, \dots, I\}$, k is referred to as the $\geq$-optimal scale if S^k is $\geq$-consistent with S and S^{k+1} is not $\geq$-consistent with S (if $k + 1 \leq I$).

It can be seen that the ≥-optimal scale of a consistent multi-scale ordered decision system is the coarsest scale for making decision or classification. That is to say, k is the ≥-optimal scale of S if and only if it is the maximum number such that S^k keeps ≥-consistent.

Theorem 10 *Let* $S = (U, C \cup \{d\}) = (U, \{a_j^k | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\} \cup \{d\})$ *be a consistent multi-scale ordered decision system, and* $k \in \{1, 2, \ldots, I\}$, *then* S^k *is ≥-consistent if and only if* $\underline{R^{\geq}_{C^k}}(D_p^{\geq}) = D_p^{\geq}$ *for all* $D_p^{\geq} \in \mathcal{D}^{\geq}$.

Proof "⇒". For any $D_p^{\geq} \in \mathcal{D}^{\geq}$, on one hand, by Proposition 2, we have $\underline{R^{\geq}_{C^k}}(D_p^{\geq}) \subseteq D_p^{\geq}$. On the other hand, for any $x \in D_p^{\geq}$, obviously $[x]_d^{\geq} \subseteq D_p^{\geq}$. Since S^k is ≥-consistent, we have $[x]_{C^k}^{\geq} \subseteq [x]_d^{\geq}$, that is $x \in \underline{R^{\geq}_{C^k}}(D_p^{\geq})$, so $D_p^{\geq} \subseteq \underline{R^{\geq}_{C^k}}(D_p^{\geq})$. Therefore $\underline{R^{\geq}_{C^k}}(D_p^{\geq}) = D_p^{\geq}$ for all $D_p^{\geq}$.

"⇐". Assume that $R^{\geq}_{C^k} \nsubseteq R^{\geq}_d$, that is, there exists an $x \in U$ such that $[x]_{C^k}^{\geq} \nsubseteq [x]_d^{\geq}$. Then there exists a $y \in U$ such that $y \in [x]_{C^k}^{\geq}$ but $y \notin [x]_d^{\geq}$. Let $d(x) = p$, clearly $x \in D_p^{\geq}$ and $y \notin D_p^{\geq}$. Since $\underline{R^{\geq}_{C^k}}(D_p^{\geq}) = D_p^{\geq}$, we have $x \in \underline{R^{\geq}_{C^k}}(D_p^{\geq})$. By the definition of lower approximation, we conclude $[x]_{C^k}^{\geq} \subseteq D_p^{\geq}$, thus $y \in D_p^{\geq}$, a contradiction. Therefore $R^{\geq}_{C^k} \subseteq R^{\geq}_d$, i.e. S^k is ≥-consistent. □

Theorem 11 *Let* $S = (U, C \cup \{d\}) = (U, \{a_j^k | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\} \cup \{d\})$ *be a consistent multi-scale ordered decision system, and* $k \in \{1, 2, \ldots, I\}$, *then*

1. S^k is ≥-consistent if and only if S^k is ≥-belief consistent with S.
2. k is the ≥-optimal scale of S if and only if k is the ≥-belief optimal scale of S.

Proof It follows immediately from Theorems 8 and 10. □

Example 3 As shown in Table 2, considering the multi-scale ordered decision system given by adding the decision criterion d_1 to Table 1, $S = (U, C \cup \{d_1\}) = (U, \{a_j^k | k = 1, 2, 3, j = 1, 2\} \cup \{d_1\})$, where decision criterion indicates the rating of scenic spots, and there are three grades of scenic spots, namely, "A-level scenic spots (1)", "AA-level scenic spots (2)" and "AAA-level scenic spots (3)". The domain of decision criterion satisfies the following total ordered relation: 3 > 2 > 1. (In Example 5, an inconsistent multi-scale ordered decision system is considered, and only decision criterion is changed. To save space, the decision criterion d_2 is listed in Table 2 as well.)

We calculate the decision dominance classes corresponding to d_1 and the lower and upper approximations at each scale as follows:

$$D_1^{\geq} = [x_1]_{d_1}^{\geq} = [x_3]_{d_1}^{\geq} = U,$$
$$D_2^{\geq} = [x_2]_{d_1}^{\geq} = [x_5]_{d_1}^{\geq} = \{x_2, x_4, x_5, x_6\},$$
$$D_3^{\geq} = [x_4]_{d_1}^{\geq} = [x_6]_{d_1}^{\geq} = \{x_4, x_6\}.$$

$$\begin{aligned}
&\underline{R^{\geq}_{C^1}}\left(D_1^{\geq}\right) = \underline{R^{\geq}_{C^2}}\left(D_1^{\geq}\right) = \underline{R^{\geq}_{C^3}}\left(D_1^{\geq}\right) = U,\\
&\underline{R^{\geq}_{C^1}}\left(D_2^{\geq}\right) = \underline{R^{\geq}_{C^2}}\left(D_2^{\geq}\right) = \underline{R^{\geq}_{C^3}}\left(D_2^{\geq}\right) = \{x_2, x_4, x_5, x_6\},\\
&\underline{R^{\geq}_{C^1}}\left(D_3^{\geq}\right) = \underline{R^{\geq}_{C^2}}\left(D_3^{\geq}\right) = \{x_4, x_6\}, \underline{R^{\geq}_{C^3}}\left(D_3^{\geq}\right) = \emptyset;\\
&\overline{R^{\geq}_{C^1}}\left(D_1^{\geq}\right) = \overline{R^{\geq}_{C^2}}\left(D_1^{\geq}\right) = \overline{R^{\geq}_{C^3}}\left(D_1^{\geq}\right) = U,\\
&\overline{R^{\geq}_{C^1}}\left(D_2^{\geq}\right) = \overline{R^{\geq}_{C^2}}\left(D_2^{\geq}\right) = \overline{R^{\geq}_{C^3}}\left(D_2^{\geq}\right) = U,\\
&\overline{R^{\geq}_{C^1}}\left(D_3^{\geq}\right) = \overline{R^{\geq}_{C^2}}\left(D_3^{\geq}\right) = \overline{R^{\geq}_{C^3}}\left(D_3^{\geq}\right) = U.
\end{aligned}$$

$$\sum_{p=1}^{3} \mathrm{Bel}^{\geq}_{C^1}\left(D_p^{\geq}\right) = \sum_{p=1}^{3} \mathrm{Bel}^{\geq}_{C^2}\left(D_p^{\geq}\right) = 2 > \frac{5}{3} = \sum_{p=1}^{3} \mathrm{Bel}^{\geq}_{C^3}\left(D_p^{\geq}\right),$$

$$\sum_{p=1}^{3} \mathrm{Pl}^{\geq}_{C^1}\left(D_p^{\geq}\right) = \sum_{p=1}^{3} \mathrm{Pl}^{\geq}_{C^2}\left(D_p^{\geq}\right) = \sum_{p=1}^{3} \mathrm{Pl}^{\geq}_{C^3}\left(D_p^{\geq}\right) = 3.$$

Since $R^{\geq}_{C^2} \subseteq R^{\geq}_{d_1}$, and $R^{\geq}_{C^3} \nsubseteq R^{\geq}_{d_1}$, we conclude that $k = 2$ is the ≥-optimal scale of S. According to the above calculation, $k = 2$ is also the ≥-belief optimal scale of S. However, the ≥-plausibility optimal scale of S is $k = 3$.

Table 2 A multi-scale ordered decision system

U	a_1^1	a_1^2	a_1^3	a_2^1	a_2^2	a_2^3	d_1	d_2	∂_{C^1}	∂_{C^3}
x_1	1	C	N	2	B	N	1	1	{1, 2, 3}	{1, 2, 3}
x_2	1	C	N	4	A	Y	2	2	{1, 2}	{1, 2}
x_3	2	B	N	1	C	N	1	2	{1, 2, 3}	{1, 2, 3}
x_4	3	B	N	4	A	Y	3	1	{1}	{1, 2}
x_5	4	A	Y	1	C	N	2	3	{3}	{3}
x_6	4	A	Y	3	B	N	3	3	{3}	{3}

Example 4 Table 3 shows an example of a new multi-scale ordered decision system for scenic spot rating, $S=(U, C\cup\{d_1\})=(U,\{a_j^k|k=1,2,3, j=1,2\}\cup\{d_1\})$, where $U=\{x_1,x_2,\dots,x_6\}$ represents the collection of scenic spots, $A=\{a_1,a_2\}$, attribute a_1 and a_2 all have three levels of scale, respectively represent "tour service" and "characteristic culture" among many criteria for creating scenic spots. The attribute values "very satisfactory (4)", "satisfactory (3)", "fair (2)" and "unsatisfactory (1)" represent the description of rating under the first scale. The attribute values "Good (A)", "General (B)" and "Poor (C)" signify the description of rating under the second scale. The attribute values "Acceptable (Y)" and "Unacceptable (N)" express the description of rating under the third scale. The decision criterion indicates the rating of scenic spots, and there are three levels, namely, "A-level scenic spots (1)", "AA-level scenic spots (2)" and "AAA-level scenic spots (3)". The domains of conditional and decision attributes satisfy the following total ordered relations: $4>3>2>1$, $A>B>C$, $Y>N$, $3>2>1$. (In Example 6, an inconsistent multi-scale decision system is considered, and only the decision criterion is changed. To save space, the decision criterion d_2 is listed in Table 3.)

The dominance classes of each scale are calculated as follows:

The first scale:

$$[x_1]^{\geq}_{C^1}=\{x_1,x_2,x_4,x_6\},[x_2]^{\geq}_{C^1}=\{x_2\},[x_3]^{\geq}_{C^1}=\{x_3,x_4,x_5,x_6\},$$
$$[x_4]^{\geq}_{C^1}=\{x_4\},[x_5]^{\geq}_{C^1}=\{x_5,x_6\},[x_6]^{\geq}_{C^1}=\{x_6\}.$$

The second scale:

$$[x_1]^{\geq}_{C^2}=\{x_1,x_2,x_4,x_6\},[x_2]^{\geq}_{C^2}=\{x_2,x_4\},[x_3]^{\geq}_{C^2}=\{x_3,x_4,x_5,x_6\},$$
$$[x_4]^{\geq}_{C^2}=\{x_4\},[x_5]^{\geq}_{C^2}=\{x_5,x_6\},[x_6]^{\geq}_{C^2}=\{x_6\}.$$

The third scale:

$$[x_1]^{\geq}_{C^3}=U,[x_2]^{\geq}_{C^3}=\{x_2,x_4\},[x_3]^{\geq}_{C^3}=U,$$
$$[x_4]^{\geq}_{C^3}=\{x_2,x_4\},[x_5]^{\geq}_{C^3}=\{x_5,x_6\},[x_6]^{\geq}_{C^3}=\{x_5,x_6\}.$$

We calculate the decision dominance classes corresponding to d_1 and the lower and upper approximations at each scale as follows:

$$D_1^{\geq}=[x_1]^{\geq}_{d_1}=[x_3]^{\geq}_{d_1}=U,$$
$$D_2^{\geq}=[x_2]^{\geq}_{d_1}=[x_5]^{\geq}_{d_1}=\{x_2,x_4,x_5,x_6\},$$
$$D_3^{\geq}=[x_4]^{\geq}_{d_1}=[x_6]^{\geq}_{d_1}=\{x_4,x_6\}.$$
$$\underline{R^{\geq}_{C^1}}(D_1^{\geq})=\underline{R^{\geq}_{C^2}}(D_1^{\geq})=\underline{R^{\geq}_{C^3}}(D_1^{\geq})=U,$$
$$\underline{R^{\geq}_{C^1}}(D_2^{\geq})=\underline{R^{\geq}_{C^2}}(D_2^{\geq})=\underline{R^{\geq}_{C^3}}(D_2^{\geq})=\{x_2,x_4,x_5,x_6\},$$
$$\underline{R^{\geq}_{C^1}}(D_3^{\geq})=\underline{R^{\geq}_{C^2}}(D_3^{\geq})=\{x_4,x_6\},\underline{R^{\geq}_{C^3}}(D_3^{\geq})=\emptyset;$$
$$\overline{R^{\geq}_{C^1}}(D_1^{\geq})=\overline{R^{\geq}_{C^2}}(D_1^{\geq})=U,$$
$$\overline{R^{\geq}_{C^1}}(D_2^{\geq})=\overline{R^{\geq}_{C^2}}(D_2^{\geq})=U,$$
$$\overline{R^{\geq}_{C^1}}(D_3^{\geq})=\{x_1,x_3,x_4,x_5,x_6\},\overline{R^{\geq}_{C^2}}(D_3^{\geq})=U.$$
$$\sum_{p=1}^{3}\mathrm{Bel}^{\geq}_{C^1}\left(D_p^{\geq}\right)=\sum_{p=1}^{3}\mathrm{Bel}^{\geq}_{C^2}\left(D_p^{\geq}\right)=2>\frac{5}{3}=\sum_{p=1}^{3}\mathrm{Bel}^{\geq}_{C^3}\left(D_p^{\geq}\right),$$
$$\sum_{p=1}^{3}\mathrm{Pl}^{\geq}_{C^1}\left(D_p^{\geq}\right)=\frac{17}{6}<3=\sum_{p=1}^{3}\mathrm{Pl}^{\geq}_{C^2}\left(D_p^{\geq}\right).$$

Since $R^{\geq}_{C^2}\subseteq R^{\geq}_{d_1}$, and $R^{\geq}_{C^3}\nsubseteq R^{\geq}_{d_1}$, we conclude that $k=2$ is the $\geq$-optimal scale of S. According to the above calculation, $k=2$ is also the $\geq$-belief optimal scale of S. However, the $\geq$-plausibility optimal scale of S is $k=1$.

Theorem 11 shows that the concept of $\geq$-optimal scale is equivalent to that of $\geq$-belief optimal scale in consistent multi-scale ordered decision systems. Examples 3 and 4 indicate that there is no static relationship between $\geq$-optimal scale and $\geq$-plausibility optimal scale.

Table 3 A multi-scale ordered decision system

U	a_1^1	a_1^2	a_1^3	a_2^1	a_2^2	a_2^3	d_1	d_2	∂_{C^1}	∂_{C^3}
x_1	1	C	N	2	B	N	1	1	{1, 2, 3}	{1, 2, 3}
x_2	1	C	N	4	A	Y	2	2	{2}	{1, 2}
x_3	2	B	N	1	C	N	1	2	{1, 2, 3}	{1, 2, 3}
x_4	2	B	N	3	A	Y	3	1	{1}	{1}
x_5	3	A	Y	1	C	N	2	3	{3}	{3}
x_6	4	A	Y	2	B	N	3	3	{3}	{3}

5.2 Optimal scale selection in inconsistent multi-scale ordered decision systems

For an inconsistent multi-scale ordered decision system $S=(U,C\cup\{d\})=(U,\{a_j^k|k=1,2,\dots,I,j=1,2,\dots,m\}\cup\{d\})$, for any $k\in\{1,2,\dots,I\}$, obviously, $S^k=(U,C^k\cup\{d\})$ is $\geq$-inconsistent.

Definition 16 Let $S=(U,C\cup\{d\})=(U,\{a_j^k|k=1,2,\dots,I,$ $j=1,2,\dots,m\}\cup\{d\})$ be a multi-scale ordered decision system, and $k\in\{1,2,\dots,I\}$, denote

$$\partial^{\geq}_{C^k}(x)=\left\{d(y)|y\in[x]^{\geq}_{C^k}\right\},x\in U,$$

then $\partial^{\geq}_{C^k}(x)$ is referred to as the generalized decision value of object x under the k-th scale in S.

Definition 17 Let $S=(U,C\cup\{d\})=(U,\{a_j^k|k=1,2,\dots,I,$ $S=(U,C\cup\{d\})=(U,\{a_j^k|k=1,2,\dots,I,j=1,2,\dots,m\}\cup\{d\})$ be an inconsistent multi-scale ordered decision system, and $k\in\{1,2,\dots,I\}$, we say that S^k is $\geq$-generalized decision consistent with S if $\partial^{\geq}_{C^k}(x)=\partial^{\geq}_{C^1}(x)$ for all $x\in U$. And k is referred to as the $\geq$-generalized decision optimal scale of S if S^k is $\geq$-generalized decision consistent with S and S^{k+1} is not $\geq$-generalized decision consistent with S (if $k+1\leq I$).

In an inconsistent multi-scale ordered decision system with I scales, it is obvious that S^k is $\geq$-generalized decision consistent with S if and only if S^k keeps the same generalized decision values as the first (finest) scale ordered decision system S^1. k is the $\geq$-generalized decision optimal scale of S if and only if k is the maximum number such that S^k keeps the same generalized decision values at the finest scale $k=1$.

Theorem 12 *Let* $S=(U,C\cup\{d\})=(U,\{a_j^k|k=1,2,\dots,I,$ $j=1,2,\dots,m\}\cup\{d\})$ *be an inconsistent multi-scale ordered decision system, and* $k\in\{1,2,\dots,I\}$, *if* S^k *is* $\geq$*-generalized decision consistent with* S, *then*

1. S^k is $\geq$-belief consistent with S.
2. S^k is $\geq$-plausibility consistent with S.

Proof 1. By Proposition 3, we have $\underline{R^{\geq}_{C^k}}(D^{\geq}_p)\subseteq\underline{R^{\geq}_{C^1}}(D^{\geq}_p)$ for all $D^{\geq}_p$, what we need to prove is that $\underline{R^{\geq}_{C^1}}(D^{\geq}_p)\subseteq\underline{R^{\geq}_{C^k}}(D^{\geq}_p)$, i.e. $[x]^{\geq}_{C^1}\subseteq D^{\geq}_p\Longrightarrow[x]^{\geq}_{C^k}\subseteq D^{\geq}_p$, $\forall x\in U$. For any $x\in U$ with $[x]^{\geq}_{C^1}\subseteq D^{\geq}_p$, take $d(x)=p$, since $V_d=\{1,2,\dots,s\}$, we have $\partial^{\geq}_{C^1}(x)\subseteq\{p,p+1,\dots,s\}$. If $[x]^{\geq}_{C^k}\nsubseteq D^{\geq}_p$, then there exists $y\in[x]^{\geq}_{C^k}$ such that $d(y)<p$, which implies that $\partial^{\geq}_{C^1}(x)\neq\partial^{\geq}_{C^k}(x)$, a contradiction. Hence, we obtain that

$$\underline{R^{\geq}_{C^1}}\left(D^{\geq}_p\right)\subseteq\underline{R^{\geq}_{C^k}}\left(D^{\geq}_p\right),$$

thus $\underline{R^{\geq}_{C^1}}(D^{\geq}_p)=\underline{R^{\geq}_{C^k}}(D^{\geq}_p)$, $\forall D^{\geq}_p\in\mathcal{D}^{\geq}$, i.e. S^k is $\geq$-belief consistent with S.

2. For any $D^{\geq}_p\in\mathcal{D}^{\geq}$, assume $y\in\overline{R^{\geq}_{C^k}}(D^{\geq}_p)$, select an $x\in U$ such that $d(x)=p$, then $[x]^{\geq}_d=D^{\geq}_p$, we have $[y]^{\geq}_{C^k}\cap[x]^{\geq}_d\neq\emptyset$. Let $z\in[y]^{\geq}_{C^k}\cap[x]^{\geq}_d$, and set $d(z)=w\geq p$, obviously $w\in\partial^{\geq}_{C^k}(y)$. Since S^k is $\geq$-generalized decision consistent with S, we have $\partial^{\geq}_{C^k}(y)=\partial^{\geq}_{C^1}(y)$. Hence $w\in\partial^{\geq}_{C^1}(y)$. We can find $y'\in[y]^{\geq}_{C^1}$ such that $d(y')=w$, thus $y'\in D^{\geq}_p$. Consequently, $[y]^{\geq}_{C^1}\cap[x]^{\geq}_d\neq\emptyset$, by the definition of upper approximation, $y\in\overline{R^{\geq}_{C^1}}(D^{\geq}_p)$. Hence we conclude

$$\overline{R^{\geq}_{C^k}}\left(D^{\geq}_p\right)\subseteq\overline{R^{\geq}_{C^1}}\left(D^{\geq}_p\right),$$

and by Proposition 3, we have $\overline{R^{\geq}_{C^1}}(D^{\geq}_p)\subseteq\overline{R^{\geq}_{C^k}}(D^{\geq}_p)$ for all $D^{\geq}_p$. Therefore, $\overline{R^{\geq}_{C^k}}(D^{\geq}_p)=\overline{R^{\geq}_{C^1}}(D^{\geq}_p)$, $\forall D^{\geq}_p\in\mathcal{D}^{\geq}$, i.e. S^k is $\geq$-plausibility consistent with S.

□

Theorem 12 shows that the $\geq$-generalized decision optimal scale is finer than the $\geq$-belief and $\geq$-plausibility optimal scales in inconsistent multi-scale ordered decision systems, and the converse does not hold in general. Moreover, there is no static relationship between $\geq$-belief and $\geq$-plausibility optimal scales, which is confirmed in the following examples.

Example 5 As shown in Table 2, considering the multi-scale ordered decision system given by adding decision criterion d_2 to the multi-scale ordered information system of scenic spot rating, $S=(U,C\cup\{d_2\})=(U,\{a_j^k|k=1,2,3,j=1,2\}\cup\{d_2\})$.

We calculate the decision dominance classes corresponding to d_2 and the lower and upper approximations at each scale as follows:

$$D_1^{\geq} = [x_1]_{d_2}^{\geq} = [x_4]_{d_2}^{\geq} = U,$$
$$D_2^{\geq} = [x_2]_{d_2}^{\geq} = [x_3]_{d_2}^{\geq} = \{x_2, x_3, x_5, x_6\},$$
$$D_3^{\geq} = [x_5]_{d_2}^{\geq} = [x_6]_{d_2}^{\geq} = \{x_5, x_6\}.$$
$$\underline{R_{C^1}^{\geq}}(D_1^{\geq}) = \underline{R_{C^2}^{\geq}}(D_1^{\geq}) = \underline{R_{C^3}^{\geq}}(D_1^{\geq}) = U,$$
$$\underline{R_{C^1}^{\geq}}(D_2^{\geq}) = \underline{R_{C^2}^{\geq}}(D_2^{\geq}) = \underline{R_{C^3}^{\geq}}(D_2^{\geq}) = \{x_5, x_6\},$$
$$\underline{R_{C^1}^{\geq}}(D_3^{\geq}) = \underline{R_{C^2}^{\geq}}(D_3^{\geq}) = \underline{R_{C^3}^{\geq}}(D_3^{\geq}) = \{x_5, x_6\};$$
$$\overline{R_{C^1}^{\geq}}(D_1^{\geq}) = \overline{R_{C^2}^{\geq}}(D_1^{\geq}) = \overline{R_{C^3}^{\geq}}(D_1^{\geq}) = U,$$
$$\overline{R_{C^1}^{\geq}}(D_2^{\geq}) = \overline{R_{C^2}^{\geq}}(D_2^{\geq}) = \{x_1, x_2, x_3, x_5, x_6\}, \overline{R_{C^3}^{\geq}}(D_2^{\geq}) = U,$$
$$\overline{R_{C^1}^{\geq}}(D_3^{\geq}) = \overline{R_{C^2}^{\geq}}(D_3^{\geq}) = \overline{R_{C^3}^{\geq}}(D_3^{\geq}) = \{x_1, x_3, x_5, x_6\}.$$
$$\sum_{p=1}^{3} \mathrm{Bel}_{C^1}^{\geq}\left(D_p^{\geq}\right) = \sum_{p=1}^{3} \mathrm{Bel}_{C^2}^{\geq}\left(D_p^{\geq}\right) = \sum_{p=1}^{3} \mathrm{Bel}_{C^3}^{\geq}\left(D_p^{\geq}\right) = \frac{6+2+2}{6} = \frac{5}{3},$$
$$\sum_{p=1}^{3} \mathrm{Pl}_{C^1}^{\geq}\left(D_p^{\geq}\right) = \sum_{p=1}^{3} \mathrm{Pl}_{C^2}^{\geq}\left(D_p^{\geq}\right) = \frac{5}{2} < \frac{8}{3} = \sum_{p=1}^{3} \mathrm{Pl}_{C^3}^{\geq}\left(D_p^{\geq}\right).$$

Since $\partial_{C^1}^{\geq}(x) = \partial_{C^2}^{\geq}(x)$ for all $x \in U$, and $\partial_{C^2}^{\geq}(x_2) \neq \partial_{C^3}^{\geq}(x_2)$, we conclude that $k = 2$ is the $\geq$-generalized decision optimal scale of S. According to the above calculation, $k = 2$ is also the $\geq$-plausibility optimal scale of S. However, the $\geq$-belief optimal scale of S is $k = 3$.

Example 6 As shown in Table 3, considering the multi-scale ordered decision system given by adding decision criterion d_2 to the multi-scale ordered information system of scenic spot rating, $S = (U, C \cup \{d_2\}) = (U, \{a_j^k | k = 1, 2, 3, j = 1, 2\} \cup \{d_2\})$.

We calculate the decision dominance classes corresponding to d_2 and the lower and upper approximations at each scale as follows:

$$D_1^{\geq} = [x_1]_{d_2}^{\geq} = [x_4]_{d_2}^{\geq} = U,$$
$$D_2^{\geq} = [x_2]_{d_2}^{\geq} = [x_3]_{d_2}^{\geq} = \{x_2, x_3, x_5, x_6\},$$
$$D_3^{\geq} = [x_5]_{d_2}^{\geq} = [x_6]_{d_2}^{\geq} = \{x_5, x_6\}.$$
$$\underline{R_{C^1}^{\geq}}(D_1^{\geq}) = \underline{R_{C^2}^{\geq}}(D_1^{\geq}) = U,$$
$$\underline{R_{C^1}^{\geq}}(D_2^{\geq}) = \{x_2, x_5, x_6\}, \underline{R_{C^2}^{\geq}}(D_2^{\geq}) = \{x_5, x_6\},$$
$$\underline{R_{C^1}^{\geq}}(D_3^{\geq}) = \underline{R_{C^2}^{\geq}}(D_3^{\geq}) = \{x_5, x_6\};$$
$$\overline{R_{C^1}^{\geq}}(D_1^{\geq}) = \overline{R_{C^2}^{\geq}}(D_1^{\geq}) = \overline{R_{C^3}^{\geq}}(D_1^{\geq}) = U,$$
$$\overline{R_{C^1}^{\geq}}(D_2^{\geq}) = \overline{R_{C^2}^{\geq}}(D_2^{\geq}) = \{x_1, x_2, x_3, x_5, x_6\}, \overline{R_{C^3}^{\geq}}(D_2^{\geq}) = U,$$
$$\overline{R_{C^1}^{\geq}}(D_3^{\geq}) = \overline{R_{C^2}^{\geq}}(D_3^{\geq}) = \overline{R_{C^3}^{\geq}}(D_3^{\geq}) = \{x_1, x_3, x_5, x_6\}.$$

$$\sum_{p=1}^{3} \mathrm{Bel}_{C^1}^{\geq}\left(D_p^{\geq}\right) = \frac{6+3+2}{6} = \frac{11}{6} > \frac{6+2+2}{6} = \frac{5}{3} = \sum_{p=1}^{3} \mathrm{Bel}_{C^2}^{\geq}\left(D_p^{\geq}\right).$$
$$\sum_{p=1}^{3} \mathrm{Pl}_{C^1}^{\geq}\left(D_p^{\geq}\right) = \sum_{p=1}^{3} \mathrm{Pl}_{C^2}^{\geq}\left(D_p^{\geq}\right) = \frac{5}{2} < \frac{8}{3} = \sum_{p=1}^{3} \mathrm{Pl}_{C^3}^{\geq}\left(D_p^{\geq}\right).$$

Since $\partial_{C^1}^{\geq}(x_2) \neq \partial_{C^2}^{\geq}(x_2)$, we conclude that $k = 1$ is the $\geq$-generalized decision optimal scale of S. According to the above calculation, $k = 1$ is also the $\geq$-belief optimal scale of S. However, the $\geq$-plausibility optimal scale of S is $k = 2$.

6 Conclusion

As we all know, rough set theory is an effective method to study inaccurate data sets, and DRSA is an important branch of rough set data analysis because it considers people's preferences in practical application. Combining the Dempster–Shafer theory of evidence with rough set theory in the DRSA can analyze and characterize quantitative uncertainty of knowledge. In this paper, we have studied optimal scale selections in multi-scale ordered information/decision systems. Based on lower and upper approximations in rough set theory and belief and plausibility functions in the Dempster-Shafer theory of evidence, we have defined several types of optimal scales in both of multi-scale ordered information systems and multi-scale ordered decision systems. We have proved that the ≥-optimal scale is equivalent to the ≥-belief optimal scale both in multi-scale ordered information systems and consistent multi-scale ordered decision systems. Though a ≥-belief optimal scale is finer than a ≥-plausibility optimal scale in the former, there is no static relationship between them in the latter. Furthermore, in inconsistent multi-scale ordered decision systems, we have introduced the notion of ≥-generalized decision optimal scale. We have examined that the ≥-generalized decision optimal scale is finer than the ≥-belief and ≥-plausibility optimal scales, the converse is not true in general, and there is no static relationship between ≥-belief and ≥-plausibility optimal scales, that is, they are different concepts. Besides, whether in multi-scale ordered information systems or multi-scale ordered decision systems, the ≥-lower approximation optimal scale is equivalent to the ≥-belief optimal scale, and the ≥-upper approximation optimal scale is equivalent to the ≥-plausibility optimal scale. This work lays out a foundation for knowledge acquisition of multi-scale ordered decision systems based on evidence theory. For further study, on one hand, the obtained results can be used to apply rule acquisition and decision making in multi-scale ordered information/decision systems. On the other hand, we will investigate optimal scale selection and attribute reduction based on evidence theory with rule acquisition approaches on more complex and more generalized information systems such as incomplete and set-valued multi-scale ordered information/decision systems.

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

References

1. Bao H, Wu WZ, Zheng JW, Li TJ (2021) Entropy based optimal scale combination selection for generalized multi-scale information tables. Int J Mach Learn Cybern 12(5):1427–1437
2. Chen DG, Li WL, Zhang X, Kwong S (2014) Evidence-theory-based numerical algorithms of attribute reduction with neighborhood-covering rough sets. Int J Approx Reason 55(3):908–923
3. Chen DG, Xu WH, Li JH, Hu QH (2020) Basic course of granular computing (in Chinese). Science Press, Beijing
4. Chen CLP, Zhang CY (2014) Data-intensive applications, challenges, techniques and technologies: a survey on big data. Inf Sci 275:314–347
5. Dempster AP (1967) Upper and lower probabilities induced by a multivalued mapping. Ann Inst Stat Math 38(2):325–339
6. Du WS, Hu BQ (2016) Attribute reduction in ordered decision tables via evidence theory. Inf Sci 364–365:91–110
7. Du WS, Hu BQ (2016) Dominance-based rough set approach to incomplete ordered information systems. Inf Sci 346–347:106–129
8. Greco S, Matarazzo B, Slowinski R (1999) Rough approximation of a preference relation by dominance relations. Eur J Oper Res 117(1):63–83
9. Greco S, Matarazzo B, Slowinski R (2001) Rough sets theory for multicriteria decision analysis. Eur J Oper Res 129(1):1–47
10. Greco S, Matarazzo B, Slowinski R (2002) Rough approximation by dominance relations. Int J Intell Syst 17(2):153–171
11. Greco S, Matarazzo B, Slowinski R (2007) Dominance-based rough set approach as a proper way of handling graduality in rough set theory. Transactions on Rough Sets VII. Lect Notes Comput Sci 4400:36–52
12. Gu SM, Wu Y, Wu WZ, Li TJ (2014) Knowledge approximations in multi-scale ordered information systems. In: Proceedings of International Conference on Rough Sets and Knowledge Technology, October 24-26, 2014, Shanghai, China. Lecture Notes in Computer Science, vol 8818. Springer, Berlin, pp 525–534
13. Hu CX, Zhang L (2021) Dynamic dominance-based multigranulation rough sets approaches with evolving ordered data. Int J Mach Learn Cybern 12(1):17–38
14. 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
15. Huang QQ, Li TR, Huang YY, Yang X, Fujita H (2020) Dynamic dominance rough set approach for processing composite ordered data. Knowl Based Syst 187:104829
16. 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
17. Li F, Hu BQ (2017) A new approach of optimal scale selection to multi-scale decision tables. Inf Sci 381:193–208
18. Li JH, Wang F, Wu WZ, Xu WH, Yang XB, She YH (2021) Review of multi-granularity data analysis methods based on granular computing (in Chinese). J Data Acquis Process 36(3):418–435
19. Li WK, Li JJ, Huang JX, Dai WZ, Zhang XP (2021) A new rough set model based on multi-scale covering. Int J Mach Learn Cybern 12(1):243–256
20. Liang JY, Qian YH, Li DY, Hu QH (2015) Theory and method of granular computing for big data mining (in Chinese). Sci Sin Inf 45(11):1355–1369
21. Lin TY, Yao YY, Zadeh LA (2002) Data mining, rough sets and granular computing. Physica-Verlag, Heidelberg
22. Lingras PJ, Yao YY (1998) Data mining using extensions of the rough set model. J Am Soc Inf Sci 49(5):415–422
23. Miao DQ, Li DY, Yao YY (2011) Uncertainty and granular computing (in Chinese). Science Press, Beijing
24. Pawlak Z (1991) Rough sets: theoretical aspects of reasoning about data. Kluwer Academic Publishers, Boston
25. Pedrycz W, Skowron A, Kreinovich V (2008) Handbook of granular computing. Wiley, New York
26. Shao MW, Zhang WX (2005) Dominance relation and rules in an incomplete ordered information system. Int J Intell Syst 20(1):13–27
27. Shafer G (1976) A mathematical theory of evidence. Princeton University Press, Princeton
28. Susmaga R (2014) Reducts and constructs in classic and dominance-based rough sets approach. Inf Sci 271:45–64
29. Wang HR, Li WT, Zhan T, Yuan KH, Hu XC (2021) Multi-granulation-based optimal scale selection in multi-scale information systems. Comput Electr Eng 92:107107
30. Wu WZ, Gu SM, Wang X (2015) Information granules in multi-scale ordered information systems. In: Proceeding of the 2015

International Conference on Machine Learning and Cybernetics, Guangzhou, China, July 12-15:182–187
31. Wu WZ, Leung Y (2011) Theory and applications of granular labelled partitions in multi-scale decision tables. Inf Sci 181(18):3878–3897
32. Wu WZ, Leung Y (2013) Optimal scale selection for multi-scale decision tables. Int J Approx Reason 54(8):1107–1129
33. Wu WZ, Leung Y, Zhang WX (2002) Connections between rough set theory and Dempster–Shafer theory of evidence. Int J Gen Syst 31(4):405–430
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 WH, Zhang XY, Zhong JM, Zhang WX (2010) Attribute reduction in ordered information systems based on evidence theory. Knowl Inf Sci 25(1):169–184
36. Yao YY, Lingras PJ (1998) Interpretation of belief functions in the theory of rough sets. Inf Sci 104(1–2):81–106
37. Zhan JM, Zhang K, Wu WZ (2021) An investigation on Wu-Leung multi-scale information systems and multi-expert group decision-making. Expert Syst Appl 170:114542
38. Zhang M, Xu LD, Zhang WX, Li HZ (2003) A rough set approach to knowledge reduction based on inclusion degree and evidence reasoning theory. Expert Syst 20(5):298–304
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(5):1095–1114