International Journal of Approximate Reasoning 187 (2025) 109569

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

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

Optimal scale combination selection based on a monotonic variable precision multi-scale rough set model

Ruili Guo [a,b,c], Qinghua Zhang [b,c,*], Yunlong Cheng [d], Ying Yang [a,b,c], Hang Zhong [a,b,c]

[a] School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China
[b] Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunication, Chongqing, 400065, China
[c] Key Laboratory of Cyberspace Big Data Intelligent Security, Ministry of Education, Chongqing, 400065, China
[d] School of Science, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China

ARTICLE INFO

Keywords:
Granular computing
Multi-scale decision system
Optimal scale combination selection
Variable precision rough set
Information measurement

ABSTRACT

Most existing generalized multi-scale rough set models (GMRSMs) are based on Pawlak's rough set, which lacks fault tolerance and thus limits their generalization ability. To improve generalization, the variable precision generalized multi-scale rough set model (VPGMRSM) was proposed. However, this model disrupts the monotonicity of the positive region, posing challenges for optimal scale combination (OSC) selection. To address these issues, a monotonic VPGMRSM is proposed in this paper through a two-stage approximation process. The proposed model preserves the monotonicity of the GMRSM and the fault tolerance of the VPGMRSM, and is further applied to OSC selection. First, the non-monotonicity of the positive region in the original VPGMRSM is analyzed. Then, a monotonic VPGMRSM is proposed, whose information measurements are proven to satisfy the monotonicity lacking in the original model. Second, an extended definition of OSC is proposed based on the positive region in the new model, which significantly simplifies and improves the efficiency of the OSC selection process. Third, two OSC selection algorithms are proposed: one based on binary search to find a single OSC, and the other based on three-way decision theory to identify all OSCs. Finally, the experimental results validate the monotonicity of the positive region in the new model and demonstrate that the proposed algorithms are not only suitable for VPGMRSMs, but also effectively reduce the computation time.

1. Introduction

Granular computing (GrC) [1] is an emerging computing paradigm that emulates human cognitive processes for knowledge representation and discovery. By introducing granular structures, GrC provides effective solutions to complex real-world problems. In the era of big data, GrC shifts traditional computational models by leveraging hierarchical granularity to formalize information processing, demonstrating clear advantages in data mining applications [2,3]. Building upon the foundational framework of GrC, researchers have established extended theories including rough set theory (RST) [4], three-way decision theory [5–7], quotient space theory [8], fuzzy set theory [9], concept-cognitive learning [10,11] and so on.

Among these, RST, proposed by Pawlak [4], is a fundamental theory in GrC to deal with imprecise and uncertain information [12]. Classical RST, which is known as Pawlak's RST, uses indiscernibility relations to partition the universe of discourse into equivalence

* Corresponding author.
E-mail address: zhangqh@cqupt.edu.cn (Q. Zhang).

https://doi.org/10.1016/j.ijar.2025.109569
Received 27 May 2025; Received in revised form 10 August 2025; Accepted 1 September 2025

Available online 10 September 2025

classes. These classes serve as atomic information granules to construct upper and lower approximations. The lower approximation consists of equivalence classes entirely within the target concept, while the upper approximation includes equivalence classes that partially intersect with the target concept [4]. RST has contributed significantly to knowledge discovery through approaches such as attribute reduction [13,14], uncertainty measures [15,16], and rule induction [17]. However, classical RST lacks fault tolerance, which makes it highly sensitive to noise. To address this issue, extensions such as probabilistic RST [18,19] and variable precision RST [20–22] have been proposed for better noise handling. However, the above RSTs typically operate under a fixed scale for each attribute.

In practical applications, objects usually have multiple values under the same attribute. For example, a student's performance can be described by levels, numerical scores, or simple pass/fail labels. Therefore, the representation of information granules and knowledge discovery in such multi-scale datasets emerge as critical challenges. To overcome these challenges, Wu and Leung [23] proposed the Wu-Leung model based on multi-scale decision systems (MDSs), where each object can have multiple values for each attribute. However, the requirement that all attributes share the same number of scales limits its practical applicability. To overcome this limitation, Li and Hu [24] proposed the GMRSM, which is based on generalized multi-scale decision systems (GMDSs). This model allows each attribute to have a different number of scales, thereby eliminating the structural constraint imposed by the Wu–Leung model. Subsequently, various extended multi-scale rough set models have been proposed, such as incomplete multi-scale rough set model [25], fuzzy multi-scale rough set model [26], hybrid multi-scale rough set models [27] and set-valued multi-scale rough set model [28]. However, these works were still based on classical RST. To enhance their ability to handle datasets with noise, researchers have proposed various extensions. For example, Wu and Leung [29] integrated the probabilistic rough set model into MDSs to explore the OSC problem. Based on this, Niu et al. [30] incorporated variable precision into the GMRSM and proposed the VPGMRSM, in which four OSC types were defined and the relationships among them were systematically analyzed. Furthermore, Gong and Li [31] introduced the cover VPGMRSM, within which they investigated OSC selection methods and designed the corresponding knowledge acquisition rules. These models significantly improve robustness to noise, but they also introduce new challenges.

Information measurements play a crucial role in OSC selection, which aims to identify the coarsest scale combination that preserves optimal discriminative ability. Typically, the finest scale combination induces the finest partition, containing the richest knowledge and offering the highest discrimination. Based on information measurements, Li et al. [32] proposed six OSC types in the GMRSM, including upper approximation, lower approximation, positive region, distribution, maximal distribution and generalized decision. Building on this, Wu and Leung [33] summarized and extended seven OSC types in the GMRSM. Moreover, various OSCs have been introduced based on entropy [34–36], granularity description accuracy [37], and cost [38–40].

However, Wang et al. [16] identified a key challenge in probabilistic rough set models: information measurements are not always monotonic with respect to (w.r.t.) partition granularity. This issue becomes particularly severe in the VPGMRSM. In analyzing the relationship between the maximum distribution OSC and the lower approximation OSC, Niu et al. [30] pointed out that monotonicity holds only under two strict conditions: (1) maximum distribution consistency, and (2) a strictly controlled range of the parameter β. This implies that unless both conditions are simultaneously satisfied, monotonicity generally does not hold. Consequently, the positive region may decrease as the scale combination becomes finer, contradicting intuitive expectations. This non-monotonicity complicates OSC selection, as verifying whether a scale combination is an OSC requires checking all coarser combinations to ensure their information measurements (e.g., positive region) differ from that of the finest scale combination. In contrast, monotonic information measurements eliminate such exhaustive checks, greatly reducing computational complexity. Therefore, how to address information measurements in the VPGMRSM that maintain monotonicity even when ideal conditions are not fully satisfied remains a critical issue.

OSC selection is a core issue of knowledge discovery in GMRSMs. Li and Hu proposed the lattice model [24] for OSC selection in the GMRSM, which is an exhaustive algorithm. Inspired by the local strategy for OSC selection in MDSs proposed by Gu and Wu [41], Li and Hu further developed the stepwise OSC selection algorithm [32] for GMRSM, which can quickly obtain an OSC. Building on these foundational studies, Hao et al. [42] proposed an OSC selection method for object updates, addressing the OSC selection issue in dynamic MDSs. Wang et al. [27] designed corresponding OSC selection algorithms to handle different types of data in hybrid GMRSMs. In addition, Cheng et al. [43] integrated OSC selection with attribute reduction using three-way decision theory and proposed an extended stepwise OSC selection algorithm. Other OSC selection algorithms have also been explored, including heuristic search [44–47], incremental learning [28,48], and matrix-based methods [49,50], and so on. However, existing OSC selection methods often face the challenge of scale combination explosion. Moreover, most of these methods are designed for GMRSMs and inherently rely on the monotonicity of information measurements. In contrast, information measurements in VPGMRSMs are non-monotonic, making it infeasible to use local OSC. As a result, directly applying GMRSM-based algorithms may fail to obtain the coarsest scale combination consistent with the finest; instead, they can only identify scale combinations that are consistent with the finest. Therefore, designing efficient and accurate OSC selection methods for VPGMRSMs remains an open and challenging problem.

In the VPGMRSM, information measurements are generally non-monotonic unless two strict conditions are simultaneously satisfied. This non-monotonicity significantly increases the complexity of OSC selection, rendering existing efficient local search algorithms inapplicable. To overcome these limitations, this paper proposes an improved monotonic model based on a two-stage approximation process, which fundamentally alleviates the challenges caused by non-monotonic information measurements. Based on this model, a new definition of OSC and the corresponding OSC selection algorithms are proposed to enhance the effectiveness and adaptability of the application of VPGMRSMs. The main contributions are as follows.

- A monotonic variable precision generalized multi-scale rough set model (MVPGMRSM) is proposed, which not only maintains the monotonicity of information measurements of the GMRSM but also incorporates the fault tolerance capability of the VPGMRSM.

Table 1
List of abbreviations.

Abbreviation	Full name
OSC	Optimal scale combination
RST	Rough set theory
SDS	Single-scale decision system
MDS	Multi-scale decision system
GMDS	Generalized multi-scale decision system
GMRSM	Generalized multi-scale rough set model
VPGMRSM	Variable precision generalized multi-scale rough set model
MVPGMRSM	Monotonic variable precision generalized multi-scale rough set model

- A local OSC definition based on the monotonic positive region in the MVPGMRSM is proposed. This definition adopts a local strategy, enabling the identification of OSCs without checking all coarser scale combinations, thereby significantly improving the efficiency of OSC selection in VPGMRSMs.
- Two OSC selection algorithms are proposed for MVPGMRSM. The first algorithm uses the idea of binary search [51] to find a single OSC, allowing for quick results even with large-scale data. The second algorithm uses three-way decision theory to accurately identify all OSCs.

The remaining sections of this paper are organized as follows: Section 2 introduces RST, GMDS, GMRSM, and VPGMRSM. Section 3 analyzes the non-monotonicity issue in the VPGMRSM and proposes the MVPGMRSM. Section 4 presents the OSC definition based on the MVPGMRSM and proposes two OSC selection algorithms. Section 5 presents the results of related experiments. Finally, the paper concludes in Section 6.

2. Preliminaries

Table 1 lists the abbreviations used in this paper. Following that, this section provides a concise overview of key concepts related to RST, GMDS, GMRSM, and VPGMRSM.

2.1. Pawlak's rough set

Definition 1. [4] A single-scale decision system (SDS) can be represented as a tuple $S = (U, A \cup D)$, where $U = \{x_1, x_2, \dots, x_n\}$ represents the universe of discourse, and $A = \{a_1, a_2, \dots, a_m\}$ is a non-empty finite set of attributes, called the condition attribute set. The set of decision attributes is denoted as $D = \{d\}$, satisfying $A \cap D = \emptyset$. Let $a_j : U \to V_j$ be the mapping for each $j = 1, 2, \dots, m$, where $V_j = \{a_j(x) \mid x \in U\}$ represents the domain of a_j. Similarly, each $d \in D$ defines a mapping $d : U \to V_d$, where V_d represents the domain of d.

Let $B \subseteq A$ and $X \subseteq U$. For any $B \subseteq A$, the equivalence relation R_B can be defined as follows:

$$R_B = \{(x, y) \in U \times U \mid a(x) = a(y), \forall a \in B\}, \tag{1}$$

where R_B partitions U into equivalence classes, denoted as U/B, that is:

$$U/B = \{[x]_B \mid x \in U\}, \tag{2}$$

where $[x]_B = \{y \in U \mid (x, y) \in R_B\}$ represents the equivalence class that includes the element x.

The lower approximation $\underline{R_B}(X)$ and upper approximation $\overline{R_B}(X)$ of X w.r.t. B are defined as follows:

$$\begin{aligned} \underline{R_B}(X) &= \{x \in U \mid [x]_B \subseteq X\}, \\ \overline{R_B}(X) &= \{x \in U \mid [x]_B \cap X \neq \emptyset\}. \end{aligned} \tag{3}$$

Consequently, the universe of discourse U can be divided into three distinct regions: the positive region $POS_B(X) = \underline{R_B}(X)$, the negative region $NEG_B(X) = U - \overline{R_B}(X)$, and the boundary region $BND_B(X) = \overline{R_B}(X) - \underline{R_B}(X)$.

Let $U/D = D_1, D_2, \dots, D_r$ denote the partition of U induced by d. The equivalence relation induced by d is defined as:

$$R_d = \{(x, y) \in U \times U \mid d(x) = d(y)\}. \tag{4}$$

The SDS $S = (U, A \cup D)$ is said to be consistent if $R_A \subseteq R_d$; otherwise, it is inconsistent.

For an SDS $S = (U, A \cup D)$, $Z \subseteq A$, the generalized decision value of x w.r.t. Z is defined as:

$$\partial_Z(x) = \left\{d(y) \mid y \in [x]_Z\right\}, \quad x \in U, \tag{5}$$

and ∂_Z denotes the generalized decision function w.r.t. Z. It follows that S is consistent if and only if $|\partial_A(x)| = 1$ for all $x \in U$, where $|*|$ is the cardinality of a set.

Table 2
A GMDS and its scale transformation process.

U	a_1					a_2					a_3			d
	a_1^1	$\xleftarrow{g_1^{2,1}}$	a_1^2	$\xleftarrow{g_1^{3,2}}$	a_1^3	a_2^1	$\xleftarrow{g_2^{2,1}}$	a_2^2	$\xleftarrow{g_2^{3,2}}$	a_2^3	a_3^1	$\xleftarrow{g_3^{2,1}}$	a_3^2	
x_1	1		1		55	2		2		65	2		90	-
x_2	2		2		60	2		2		70	2		85	+
x_3	2		2		65	2		2		70	1		50	+
x_4	2		2		65	2		2		70	1		50	+
x_5	2		2		65	2		2		70	1		50	-
x_6	2		2		65	2		3		80	1		55	+
x_7	2		2		65	2		3		80	1		55	+
x_8	2		2		70	2		3		80	2		90	+
x_9	2		2		70	2		4		95	1		55	-
x_{10}	2		2		70	2		4		95	1		55	+
x_{11}	2		3		75	1		1		50	1		55	-
x_{12}	2		3		75	2		2		65	1		50	+
x_{13}	2		3		75	2		3		75	2		85	+
x_{14}	2		3		80	2		2		70	2		85	-
x_{15}	2		4		90	2		3		75	2		80	+
x_{16}	2		4		95	2		4		90	2		80	+

2.2. Generalized multi-scale decision system

Definition 2. [24] A GMDS can be represented as a tuple $S = (U, A \cup D)$, where $U = \{x_1, x_2, \dots, x_n\}$ represents the universe of discourse, and $A = \{a_1, a_2, \dots, a_m\}$ is a non-empty finite set of conditional attributes. Each attribute $a_j \in A$ is a multi-scale attribute, meaning that for each object in U, the attribute a_j takes different values under different scales. The decision attribute set, denoted as $D = \{d\}$, which is a single-scale attribute.

Suppose that a_j has I_j levels of scale, and V_j^k is the domain of attribute a_j at the k-th scale. Then, a GMDS can be represented as:

$$S = (U, \{a_j^k \mid k = 1, 2, \dots, I_j, j = 1, 2, \dots, m\} \cup D), \tag{6}$$

where a_j^k is a surjective mapping from U to V_j^k. In this paper, it is stipulated that as the superscript k of a_j^k increases, the scale of a_j becomes finer. Therefore, for $I_j \geq k \geq 2$, there exists a surjective mapping $g_j^{k,k-1} : V_j^k \to V_j^{k-1}$ that satisfies $a_j^{k-1}(x) = g_j^{k,k-1}(a_j^k(x))$, where $g_j^{k,k-1}$ is the granularity information transformation function. Table 2 presents the scale transformation process of the condition attributes.

Definition 3. [24] Consider a GMDS $S = (U, A \cup D)$. For any attribute $a_j \in A$, if it takes the l_j-th scale, where $1 \leq l_j \leq I_j$ ($j = 1, 2, \dots, m$), then $K = (l_1, l_2, \dots, l_m)$ is called a scale combination of S. The corresponding attribute set under this combination is represented as $A^K = (a_1^{l_1}, a_2^{l_2}, \dots, a_m^{l_m})$, and the SDS formed under K is represented as $S^K = (U, A^K \cup D)$. The set of scale combinations in S is called the scale collection of S and is represented as Ω^+:

$$\Omega^+ = \{(l_1, l_2, \dots, l_m) \mid 1 \leq l_j \leq I_j, j = 1, 2, \dots, m\}. \tag{7}$$

Evidently, a GMDS S can be decomposed into $\prod_{j=1}^{m} I_j$ SDSs. Notably, when all attributes share the same number of scale levels, that is, $I_1 = I_2 = \cdots = I_m = I$, the GMDS degenerates to the MDS.

Since the scale of attribute a_j becomes finer as the superscript k of a_j^k increases, the finest scale combination is $I = (I_1, I_2, \dots, I_m)$, and the coarsest scale combination is $CSC = (1, 1, \dots, 1)$. The GMDS $S = (U, A \cup D)$ is called consistent if its corresponding SDS $S^I = (U, A^I \cup D)$ is consistent; otherwise, it is inconsistent.

Definition 4. [24] Let $K = (l_1, l_2, \dots, l_m)$ and $K' = (l_1', l_2', \dots, l_m') \in \Omega^+$. If $l_j \geq l_j'$ for all $j \in \{1, 2, \dots, m\}$, then K is regarded as finer than K', denoted as $K \succeq K'$, while K' is regarded as coarser than K, denoted as $K' \preceq K$. Furthermore, if $K \succeq K'$ and there exists at least one j such that $l_j > l_j'$, then K is regarded as strictly finer than K', denoted as $K \succ K'$, and correspondingly, K' is regarded as strictly coarser than K, denoted as $K' \prec K$. If K and K' satisfy $K \succeq K'$ or $K' \succeq K$, they are considered comparable; otherwise, they are incomparable.

2.3. Generalized multi-scale rough set model

Definition 5. [24] Consider a GMDS $S = (U, A \cup D)$. For any scale combination $K \in \Omega^+$, the equivalence relation induced by A^K is defined as:

$$R_{A^K} = \{(x,y) \in U \times U \mid a_j^{l_j}(x) = a_j^{l_j}(y), j = 1,2,\ldots,m\}. \tag{8}$$

Based on R_{A^K}, and $X \subseteq U$, the partition of U can be defined as: $U/A^K = \{[x]_{A^K} \mid x \in U\}$, where $[x]_{A^K} = \{y \mid (x,y) \in [x]_{A^K}\}$. The lower approximation $\underline{R_{A^K}}(X)$ and the upper approximation $\overline{R_{A^K}}(X)$ are defined as:

$$\begin{aligned} &\underline{R_{A^K}}(X) = \{x \in U \mid [x]_{A^K} \subseteq X\}, \\ &\overline{R_{A^K}}(X) = \{x \in U \mid [x]_{A^K} \cap X \neq \emptyset\}. \end{aligned} \tag{9}$$

For $U/D = \{D_1, D_2, \ldots, D_r\}$, then $\underline{R_{A^K}}(D) = \bigcup_{i=1}^{r} \underline{R_{A^K}}(D_i)$, $\overline{R_{A^K}}(D) = \bigcup_{i=1}^{r} \overline{R_{A^K}}(D_i)$.

Then, the positive region of D w.r.t. A^K is defined as:

$$POS_{A^K}(D) = \underline{R_{A^K}}(D) = \bigcup_{i=1}^{r} \underline{R_{A^K}}(D_i). \tag{10}$$

Proposition 1. *[24,33] Let $S = (U, A \cup D)$ be a GMDS. Suppose $K, K' \in \Omega^+$ and $K \succeq K'$. Then, the following properties hold:*

(1) $R_{A^K} \subseteq R_{A^{K'}}$,

(2) $[x]_{A^K} \subseteq [x]_{A^{K'}}$,

(3) $POS_{A^K}(X) \supseteq POS_{A^{K'}}(X)$,

(4) $POS_{A^K}(D) \supseteq POS_{A^{K'}}(D)$.

2.4. Variable precision generalized multi-scale rough set model

In the GMRSM, the definitions of the upper and lower approximations lack fault tolerance, which reduces their robustness when handling noisy data in real-world applications. To address this limitation and enhance the model's generalization ability, variable precision is incorporated into the GMRSM.

Definition 6. [30] Consider a GMDS $S = (U, A \cup D)$, where $K \in \Omega^+$ and $0.5 < \beta \leq 1$. The β upper approximation and the β lower approximation of X w.r.t. A^K are defined as:

$$\begin{aligned} &\underline{R^{\beta}_{A^K}}(X) = \{x \in U \mid P(X \mid [x]_{A^K}) \geq \beta\}, \\ &\overline{R^{\beta}_{A^K}}(X) = \{x \in U \mid P(X \mid [x]_{A^K}) > 1 - \beta\}, \end{aligned} \tag{11}$$

where $P(X \mid [x]_{A^K}) = \frac{|[x]_{A^K} \cap X|}{|[x]_{A^K}|}$.

For $U/D = \{D_1, D_2, \ldots, D_r\}$, then $\underline{R^{\beta}_{A^K}}(D) = \bigcup_{i=1}^{r} \underline{R^{\beta}_{A^K}}(D_i)$, $\overline{R^{\beta}_{A^K}}(D) = \bigcup_{i=1}^{r} \overline{R^{\beta}_{A^K}}(D_i)$, and $POS^{\beta}_{A^K}(D) = \underline{R^{\beta}_{A^K}}(D)$.

When $\beta = 1$, Definition 6 will degenerate to Definition 5. That is, $\underline{R^1_{A^K}}(X) = \underline{R_{A^K}}(X)$, $\overline{R^1_{A^K}}(X) = \overline{R_{A^K}}(X)$, $\underline{R^1_{A^K}}(D) = \underline{R_{A^K}}(D)$, $\overline{R^1_{A^K}}(D) = \overline{R_{A^K}}(D)$. Therefore, it can be considered that the VPGMRSM is an extension of the GMRSM, and the GMRSM is a specific instance of the VPGMRSM. In this paper, this VPGMRSM is also referred to as the classical VPGMRSM.

3. A new variable precision generalized multi-scale rough set model

This section first examines the non-monotonicity issue in the classical VPGMRSM and analyzes the challenges it poses for OSC selection. Then, a new monotonic VPGMRSM is proposed, and its corresponding properties are demonstrated.

3.1. Non-monotonicity issue in the variable precision generalized multi-scale rough set model

Proposition 1 indicates that as the scale combination becomes finer, the corresponding partition also becomes finer, leading to a larger positive region. Thus, in the GMRSM, the positive region increases monotonically with finer scale combinations. However, in the classical VPGMRSM, such monotonicity requires two strict conditions: (1) maximum distribution consistency, and (2) a strictly controlled range of the parameter β. As these conditions are rarely met in practice, the positive region typically lacks monotonicity.

In the VPGMRSM, consider a GMDS $S = (U, A \cup D)$, where $0.5 < \beta \leq 1$, $K, K' \in \Omega^+$ satisfy $K \succeq K'$, $X \subseteq U$, and $U/D = \{D_1, D_2, \ldots, D_r\}$; the following properties do not necessarily hold:

(1) $POS^{\beta}_{A^{K'}}(X) \subseteq POS^{\beta}_{A^K}(X)$,

(2) $POS^{\beta}_{A^{K'}}(D) \subseteq POS^{\beta}_{A^K}(D)$.

An example is presented to demonstrate that the aforementioned properties do not necessarily hold.

Example 1. Table 2 is a GMDS, assume that $\beta = 0.7$, $K = (2,2,2)$, and $K' = (1,1,1)$. It is evident that $K \succ K'$. Let $X = \{x_2, x_3, x_4, x_6, x_7, x_8, x_{10}, x_{12}, x_{13}, x_{15}, x_{16}\}$.

The equivalence classes of U induced by R_{A^K} and $R_{A^{K'}}$ are:

$$U/A^K = \{\{x_1\}, \{x_2\}, \{x_3, x_4, x_5\}, \{x_6, x_7\}, \{x_8\}, \{x_9, x_{10}\}, \{x_{11}\}, \{x_{12}\}, \{x_{13}\},$$
$$\{x_{14}\}, \{x_{15}\}, \{x_{16}\}\},$$
$$U/A^{K'} = \{\{x_1\}, \{x_2, x_8, x_{13}, x_{14}, x_{15}, x_{16}\}, \{x_3, x_4, x_5, x_6, x_7, x_9, x_{10}, x_{12}\}, \{x_{11}\}\}.$$

The β positive regions of X w.r.t. A^K and $A^{K'}$ are:

$$POS^{0.7}_{A^K}(X) = \underline{R^{0.7}_{A^K}}(X) = \{x_2, x_6, x_7, x_8, x_{12}, x_{13}, x_{15}, x_{16}\},$$
$$POS^{0.7}_{A^{K'}}(X) = \underline{R^{0.7}_{A^{K'}}}(X) = \{x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}\}.$$

It is worth noting that, although $K \succ K'$, the property $POS^{\beta}_{A^{K'}}(X) \subseteq POS^{\beta}_{A^K}(X)$ do not hold.

The decision equivalence classes are:

$$U/D = \{\{x_2, x_3, x_4, x_6, x_7, x_8, x_{10}, x_{12}, x_{13}, x_{15}, x_{16}\}, \{x_1, x_5, x_9, x_{11}, x_{14}\}\}.$$

The β positive regions of D w.r.t. A^K and $A^{K'}$ are:

$$POS^{0.7}_{A^K}(D) = \underline{R^{0.7}_{A^K}}(D) = \{x_1, x_2, x_6, x_7, x_8, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}\},$$
$$POS^{0.7}_{A^{K'}}(D) = \underline{R^{0.7}_{A^{K'}}}(D) = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}\}.$$

It is clear that, although $K \succ K'$, the property $POS^{\beta}_{A^{K'}}(D) \subseteq POS^{\beta}_{A^K}(D)$ do not hold.

Example 1 illustrates that the positive region defined in Definition 6 is non-monotonic without strict conditions. Specifically, the positive region of a finer scale combination may be smaller than that of a coarser one. This poses challenges for OSC selection. OSC refers to the coarsest scale combination that maintains consistency with the finest scale combination I. However, due to the lack of monotonicity, the use of a local strategy is not feasible in the OSC selection process. When the positive region of a given scale combination is consistent with that of I, it is necessary to check all strictly coarser combinations to ensure that they do not share the same positive region. Only after this verification can the given combination be considered an OSC. This verification step significantly increases computational cost and time complexity, thus reducing practical efficiency.

3.2. A monotonic variable precision generalized multi-scale rough set model

The classical VPGMRSM improves noise tolerance through variable precision, but this causes non-monotonicity in the positive region, increasing the computational cost of OSC selection. In contrast, the positive region in the GMRSM is monotonic, facilitating efficient OSC selection, but lacks fault tolerance. Based on this observation, MVPGMRSM is proposed, which combines the fault tolerance of VPGMRSM with the monotonicity of GMRSM, aiming to enhance generalization while maintaining monotonicity, thereby improving the efficiency of OSC selection in VPGMRSMs.

Definition 7. Consider a GMDS $S = (U, A \cup D)$, where $K \in \Omega^+$, $0.5 < \beta \leq 1$, and $I = (I_1, I_2, \ldots, I_m)$. The β lower approximation and the β upper approximation of X w.r.t. A^K are defined as:

$$\begin{aligned} \underline{R^{\beta}_{A^K}}(X)' &= \underline{R_{A^K}}\left(\underline{R^{\beta}_{A^I}}(X)\right), \\ \overline{R^{\beta}_{A^K}}(X)' &= \overline{R_{A^K}}\left(\overline{R^{\beta}_{A^I}}(X)\right). \end{aligned} \tag{12}$$

For $U/D = \{D_1, D_2, \ldots, D_r\}$, the β lower approximation and the β upper approximation of D w.r.t. A^K are defined as:

$$\begin{aligned} \underline{R^{\beta}_{A^K}}(D)' &= \bigcup_{i=1}^{r}\left(\underline{R^{\beta}_{A^K}}(D_i)'\right), \\ \overline{R^{\beta}_{A^K}}(D)' &= \bigcup_{i=1}^{r}\left(\overline{R^{\beta}_{A^K}}(D_i)'\right). \end{aligned} \tag{13}$$

The definition of the β positive region of D w.r.t. A^K is as follows:

$$POS^{\beta}_{A^K}(D)' = \underline{R^{\beta}_{A^K}}(D)'. \tag{14}$$

The design of the MVPGMRSM aims to integrate the advantages of both the VPGMRSM and the GMRSM, maintaining fault tolerance while ensuring the monotonicity of information measurements. In this new model, the computation of information measurements differs from that in the classical VPGMRSM. Specifically, the upper and lower approximations for any scale combination K are computed through a two-stage process. In the first stage, the β upper and lower approximations are calculated under the finest scale combination I using the method of the VPGMRSM, which enables the model to capture the most expressive and fault-tolerant knowledge. In the second stage, these results are further refined under scale combination K using the GMRSM method, thus preserving the monotonicity of information measurements. This two-stage approximation structure not only enhances the robustness of the model but also alleviates the computational burden caused by non-monotonicity in OSC selection. Next, the important properties of the MVPGMRSM are presented.

Theorem 1. *Consider a GMDS $S = (U, A \cup D)$, where $0.5 < \beta \leq 1$, $I = (I_1, \ldots, I_m)$, $K, K' \in \Omega^+$ satisfy $K \succeq K'$, and $X \subseteq U$. Then, the following properties hold:*

(1) $\underline{R^{\beta}_{A^{K'}}}(X)' \subseteq \underline{R^{\beta}_{A^K}}(X)'$,

(2) $\overline{R^{\beta}_{A^{K'}}}(X)' \supseteq \overline{R^{\beta}_{A^K}}(X)'$,

(3) $POS^{\beta}_{A^{K'}}(X)' \subseteq POS^{\beta}_{A^K}(X)'$.

Proof. (1) As $K \succeq K'$, it can be deduced from Proposition 1 that $[x]_{A^K} \subseteq [x]_{A^{K'}}$.

For any $x \in \underline{R_{A^{K'}}}(\underline{R^{\beta}_{A^I}}(X))$, it follows that $[x]_{A^{K'}} \subseteq \underline{R^{\beta}_{A^I}}(X)$. As $[x]_{A^K} \subseteq [x]_{A^{K'}}$, it follows that $[x]_{A^K} \subseteq \underline{R^{\beta}_{A^I}}(X)$, implying $x \in \underline{R_{A^K}}(\underline{R^{\beta}_{A^I}}(X))$. Thus, $\underline{R_{A^K}}(\underline{R^{\beta}_{A^I}}(X)) \supseteq \underline{R_{A^{K'}}}(\underline{R^{\beta}_{A^I}}(X))$. Furthermore, from Definition 7, the following conclusion can be drawn:

$$\underline{R^{\beta}_{A^{K'}}}(X)' = \underline{R_{A^{K'}}}(\underline{R^{\beta}_{A^I}}(X)) \subseteq \underline{R_{A^K}}(\underline{R^{\beta}_{A^I}}(X)) = \underline{R^{\beta}_{A^K}}(X)'.$$

(2) For any $x \in \overline{R_{A^K}}(\underline{R^{\beta}_{A^I}}(X))$, it follows that $[x]_{A^K} \cap \underline{R^{\beta}_{A^I}}(X) \neq \emptyset$. As $[x]_{A^K} \subseteq [x]_{A^{K'}}$, it follows that $[x]_{A^{K'}} \cap \underline{R^{\beta}_{A^I}}(X) \neq \emptyset$, implying $x \in \overline{R_{A^{K'}}}(\underline{R^{\beta}_{A^I}}(X))$. Thus, $\overline{R_{A^K}}(\underline{R^{\beta}_{A^I}}(X)) \subseteq \overline{R_{A^{K'}}}(\underline{R^{\beta}_{A^I}}(X))$. Furthermore, from Definition 7, the following conclusion can be drawn:

$$\overline{R^{\beta}_{A^{K'}}}(X)' = \overline{R_{A^{K'}}}(\underline{R^{\beta}_{A^I}}(X)) \supseteq \overline{R_{A^K}}(\underline{R^{\beta}_{A^I}}(X)) = \overline{R^{\beta}_{A^K}}(X)'.$$

(3) Since $POS^{\beta}_{A^K}(X)' = \underline{R^{\beta}_{A^K}}(X)'$, this proof is similarly to (1). □

Theorem 2. *Consider a GMDS $S = (U, A \cup D)$, where $0.5 < \beta \leq 1$, $I = (I_1, I_2, \ldots, I_m)$, $K, K' \in \Omega^+$ satisfy $K \succeq K'$, and $U/D = \{D_1, D_2, \ldots, D_r\}$. Then, the following properties hold:*

(1) $\underline{R^{\beta}_{A^{K'}}}(D)' \subseteq \underline{R^{\beta}_{A^K}}(D)'$,

(2) $\overline{R^{\beta}_{A^{K'}}}(D)' \supseteq \overline{R^{\beta}_{A^K}}(D)'$,

(3) $POS^{\beta}_{A^{K'}}(D)' \subseteq POS^{\beta}_{A^K}(D)'$.

Proof. (1) Following the proof of (1) in Theorem 1, for any $D_i \in U/D$, it can be obtained: $\underline{R_{A^{K'}}}(\underline{R^{\beta}_{A^I}}(D_i)) \subseteq \underline{R_{A^K}}(\underline{R^{\beta}_{A^I}}(D_i))$.

Thus: $\bigcup_{i=1}^{r} \underline{R_{A^{K'}}}(\underline{R^{\beta}_{A^I}}(D_i)) \subseteq \bigcup_{i=1}^{r} \underline{R_{A^K}}(\underline{R^{\beta}_{A^I}}(D_i))$.

Which implies:

$$\underline{R^{\beta}_{A^{K'}}}(D)' \subseteq \underline{R^{\beta}_{A^K}}(D)'.$$

Properties (2) and (3) can be proved similarly to (1). □

Theorems 1 and 2 show that as the scale combination K becomes finer, the lower approximation, the upper approximation, and the positive region in the MVPGMRSM exhibit monotonicity.

The following example shows the MVPGMRSM construction process and illustrates the monotonicity of the positive region.

Example 2. (Continued from Example 1) From Table 2, $I = (3, 2, 2)$ is obtained. It can be derived through the calculation that:

$$U/A^I = \{\{x_1\}, \{x_2\}, \{x_3, x_4, x_5\}, \{x_6, x_7\}, \{x_8\}, \{x_9, x_{10}\}, \{x_{11}\}, \{x_{12}\}, \{x_{13}\}, \{x_{14}\}, \{x_{15}\}, \{x_{16}\}\}.$$

By Definition 6, the β lower approximation of X w.r.t. A^I is:

$$\underline{R^{0.7}_{A^I}}(X) = \{x_2, x_6, x_7, x_8, x_{12}, x_{13}, x_{15}, x_{16}\}.$$

By Definition 7, the β positive regions of X w.r.t. A^K and $A^{K'}$ are:

$$POS^{0.7}_{A^K}(X)' = \underline{R^{0.7}_{A^K}}(X)' = \{x_2, x_6, x_7, x_8, x_{12}, x_{13}, x_{15}, x_{16}\},$$
$$POS^{0.7}_{A^{K'}}(X)' = \underline{R^{0.7}_{A^{K'}}}(X)' = \{\emptyset\}.$$

As evidenced by the results, it is clear that $K \succeq K' \Rightarrow POS^{\beta}_{A^{K'}}(X) \subseteq POS^{\beta}_{A^K}(X)$ (as shown in Theorem 1).

By Definition 7, the β positive regions of D w.r.t. A^K and $A^{K'}$ are:

$$POS^{0.7}_{A^K}(D)' = \underline{R^{0.7}_{A^K}}(D)' = \{x_1, x_2, x_6, x_7, x_8, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}\},$$
$$POS^{0.7}_{A^{K'}}(D)' = \underline{R^{0.7}_{A^{K'}}}(D)' = \{x_1, x_{11}\}.$$

Therefore, it can be inferred that $K \succeq K' \Rightarrow POS^{\beta}_{A^{K'}}(D) \subseteq POS^{\beta}_{A^K}(D)$ (as shown in Theorem 2).

The following two theorems explain the relationship among the lower approximation, upper approximation, positive region, and finest scale combination I in the MVPGMRSM and classical VPGMRSM.

Theorem 3. *Consider a GMDS $S = (U, A \cup D)$, where $I = (I_1, I_2, \ldots, I_m) \in \Omega^+$, $0.5 < \beta \leq 1$, and $X \subseteq U$. Then, the following properties hold:*

(1) $\underline{R^{\beta}_{A^I}}(X)' = \underline{R^{\beta}_{A^I}}(X)$,

(2) $\overline{R^{\beta}_{A^I}}(X)' = \overline{R^{\beta}_{A^I}}(X)$,

(3) $POS^{\beta}_{A^I}(X)' = POS^{\beta}_{A^I}(X)$.

Proof. (1) For any $x \in \underline{R_{A^I}}(\underline{R^{\beta}_{A^I}}(X))$, it can be deduced that $x \in \underline{R^{\beta}_{A^I}}(X)$. Therefore, $\underline{R_{A^I}}(\underline{R^{\beta}_{A^I}}(X)) \subseteq \underline{R^{\beta}_{A^I}}(X)$. For any $x \in \underline{R^{\beta}_{A^I}}(X)$, it holds that $[x]_{A^I} \subseteq \underline{R^{\beta}_{A^I}}(X)$. By the definition of lower approximation, it can be deduced that $x \in \underline{R_{A^I}}(\underline{R^{\beta}_{A^I}}(X))$. Thus, $\underline{R^{\beta}_{A^I}}(X) \subseteq \underline{R_{A^I}}(\underline{R^{\beta}_{A^I}}(X))$. Hence, $\underline{R_{A^I}}(\underline{R^{\beta}_{A^I}}(X)) = \underline{R^{\beta}_{A^I}}(X)$.

Then,

$$\underline{R^{\beta}_{A^I}}(X)' = \underline{R^{\beta}_{A^I}}(X).$$

Properties (2) and (3) can be proved similarly to (1). □

Theorem 4. *Consider a GMDS $S = (U, A \cup D)$, where $I = (I_1, I_2, \ldots, I_m) \in \Omega^+$, $0.5 < \beta \leq 1$, and $U/D = \{D_1, D_2, \ldots, D_r\}$. Then, the following properties hold:*

(1) $\underline{R^{\beta}_{A^I}}(D)' = \underline{R^{\beta}_{A^I}}(D)$,

(2) $\overline{R^{\beta}_{A^I}}(D)' = \overline{R^{\beta}_{A^I}}(D)$,

(3) $POS^{\beta}_{A^I}(D)' = POS^{\beta}_{A^I}(D)$.

Proof. (1) Following the proof of (1) in Theorem 3, for any $D_i \in U/D$, it can be obtained: $\underline{R^{\beta}_{A^I}}(D_i)' = \underline{R^{\beta}_{A^I}}(D_i)$.

Thus: $\bigcup_{i=1}^{r} \underline{R^{\beta}_{A^I}}(D_i)' = \bigcup_{i=1}^{r} \underline{R^{\beta}_{A^I}}(D_i)$.

Which implies:

$$\underline{R^{\beta}_{A^I}}(D)' = \underline{R^{\beta}_{A^I}}(D).$$

Properties (2) and (3) can be proved similarly to (1). □

From Theorems 3 and 4, it follows that the lower approximation, upper approximation, and the positive region of I in the MVPGMRSM are the same as those in the classical VPGMRSM.

The following two theorems illustrate the relationship between the information measurements in the MVPGMRSM and those in the GMRSM when $\beta = 1$.

Theorem 5. *Consider a GMDS $S = (U, A \cup D)$, where $K \in \Omega^+$, $I = (I_1, \dots, I_m)$, $X \subseteq U$, and $\beta = 1$. Then, the following properties hold:*

(1) $\underline{R^1_{A^K}}(X)' = \underline{R_{A^K}}(X)$,

(2) $\overline{R^1_{A^K}}(X)' = \overline{R_{A^K}}(X)$,

(3) $POS^1_{A^K}(X)' = POS_{A^K}(X)$.

Proof. According to Theorem 3, it holds that $\underline{R^1_{A^I}}(X)' = \underline{R^1_{A^I}}(X)$ and $\overline{R^1_{A^I}}(X)' = \overline{R^1_{A^I}}(X)$. Consequently, $\underline{R^1_{A^I}}(X)' = \underline{R_{A^I}}(X)$, $\overline{R^1_{A^I}}(X)' = \overline{R_{A^I}}(X)$. Therefore, properties (1) and (2) only need to be proved as follows:

(1) $\underline{R_{A^K}}\left(\underline{R_{A^I}}(X)\right) = \underline{R_{A^K}}(X)$,

(2) $\overline{R_{A^K}}\left(\overline{R_{A^I}}(X)\right) = \overline{R_{A^K}}(X)$.

(1) Since $I \succeq K$, it can be directly verified that $\xi([x]_{A^K}) = \{[y]_{A^I} : [y]_{A^I} \subseteq [x]_{A^K}\}$ constitutes a partition of $[x]_{A^K}$.

For any $x \in \underline{R_{A^K}}\left(\underline{R_{A^I}}(X)\right)$, then $[x]_{A^K} \subseteq \underline{R_{A^I}}(X)$, and since $\underline{R_{A^I}}(X) \subseteq X$, it follows that $[x]_{A^K} \subseteq X$, which implies $x \in \underline{R_{A^K}}(X)$.

Conversely, for any $x \in \underline{R_{A^K}}(X)$, the inclusion $[x]_{A^K} \subseteq X$ holds. By the partition property, it holds that $[x]_{A^K} = \bigcup\{[y]_{A^I} : [y]_{A^I} \in \xi([x]_{A^K})\}$, which leads to $[y]_{A^I} \subseteq [x]_{A^K} \subseteq X$ for all $[y]_{A^I} \in \xi([x]_{A^K})$. Consequently, $[y]_{A^I} \subseteq \underline{R_{A^I}}(X)$ for all $[y]_{A^I} \in \xi([x]_{A^K})$.

Assume, for the sake of contradiction, that $x \notin \underline{R_{A^K}}\left(\underline{R_{A^I}}(X)\right)$, which implies $[x]_{A^K} \nsubseteq \underline{R_{A^I}}(X)$. Two cases arise:

- If $[x]_{A^K} \cap \underline{R_{A^I}}(X) = \emptyset$, then $[x]_{A^I} \subseteq [x]_{A^K}$ leads to $[x]_{A^I} \cap \underline{R_{A^I}}(X) = \emptyset$, implying $[x]_{A^I} \nsubseteq \underline{R_{A^I}}(X)$, which contradicts $[x]_{A^I} \subseteq \underline{R_{A^I}}(X)$.
- If $[x]_{A^K} \cap \underline{R_{A^I}}(X) \neq \emptyset$, then there exists $[y]_{A^I} \in \xi([x]_{A^K})$ such that $[y]_{A^I} \cap \underline{R_{A^I}}(X) \neq \emptyset$. This implies $[y]_{A^I} \nsubseteq \underline{R_{A^I}}(X)$, which contradicts $[y]_{A^I} \subseteq \underline{R_{A^I}}(X)$ for all $[y]_{A^I} \in \xi([x]_{A^K})$.

Therefore, the assumption must be false, and it follows that $x \in \underline{R_{A^K}}\left(\underline{R_{A^I}}(X)\right)$.

In conclusion, $\underline{R_{A^K}}\left(\underline{R_{A^I}}(X)\right) = \underline{R_{A^K}}(X)$, which implies:

$$\underline{R^1_{A^K}}(X)' = \underline{R_{A^K}}(X).$$

(2) For any $x \in \overline{R_{A^K}}(X)$. By definition, this implies that the equivalence class $[x]_{A^K}$ has a non-empty intersection with X, then $[x]_{A^K} \cap X \neq \emptyset$. Since $[x]_{A^K} = \bigcup\left\{[y]_{A^I} : [y]_{A^I} \in \xi([x]_{A^K})\right\}$, there must exist at least one block $[y]_{A^I} \in \xi([x]_{A^K})$ such that $[y]_{A^I} \cap X \neq \emptyset$.

Assume, for the sake of contradiction, that $x \notin \overline{R_{A^K}}\left(\overline{R_{A^I}}(X)\right)$, which implies that $[x]_{A^K} \cap \overline{R_{A^I}}(X) = \emptyset$. Since $[x]_{A^K} = \bigcup\left\{[y]_{A^I} : [y]_{A^I} \in \xi([x]_{A^K})\right\}$, it follows that every $[y]_{A^I} \in \xi([x]_{A^K})$ is disjoint from $\overline{R_{A^I}}(X)$, then $[y]_{A^I} \cap \overline{R_{A^I}}(X) = \emptyset$. However, this contradicts our earlier observation that some $[y]_{A^I} \in \xi([x]_{A^K})$ intersects with X, and thus must be contained in $\overline{R_{A^I}}(X)$. Therefore, the assumption is invalid, and it follows that $x \in \overline{R_{A^K}}\left(\overline{R_{A^I}}(X)\right)$.

For the reverse inclusion, for any $x \in \overline{R_{A^K}}\left(\overline{R_{A^I}}(X)\right)$, which implies $[x]_{A^K} \cap \overline{R_{A^I}}(X) \neq \emptyset$. Since $[x]_{A^K} = \bigcup\left\{[y]_{A^I} : [y]_{A^I} \in \xi([x]_{A^K})\right\}$, there exists $[y]_{A^I} \in \xi([x]_{A^K})$ satisfying $[y]_{A^I} \cap \overline{R_{A^I}}(X) \neq \emptyset$. As $\overline{R_{A^I}}(X)$ includes all the equivalence classes that intersect with X, it follows that $[y]_{A^I} \cap X \neq \emptyset$.

Suppose, to the contrary, that $x \notin \overline{R_{A^K}}(X)$ implies $[x]_{A^K} \cap X = \emptyset$. From the decomposition of $[x]_{A^K}$ into blocks of $\xi([x]_{A^K})$, it can be concluded that $[y]_{A^I} \cap X = \emptyset$ for all such $[y]_{A^I}$. This again contradicts the earlier deduction that there exists $[y]_{A^I} \in \xi([x]_{A^K})$ such that $[y]_{A^I} \cap X \neq \emptyset$.

Therefore, the contradiction implies $x \in \overline{R_{A^K}}(X)$.

In conclusion, $\overline{R_{A^K}}\left(\overline{R_{A^I}}(X)\right) = \overline{R_{A^K}}(X)$, which implies:

$$\overline{R^1_{A^K}}(X)' = \overline{R_{A^K}}(X).$$

(3) Since $POS^1_{\underline{A^K}}(X)' = \underline{R^1_{A^K}}(X)'$, this proof is similarly to (1). □

Theorem 6. *Consider a GMDS $S = (U, A \cup D)$, where $K \in \Omega^+$, $I = (I_1, \ldots, I_m)$, $U/D = \{D_1, D_2, \ldots, D_r\}$, and $\beta = 1$. Then, the following properties hold:*

(1) $\underline{R^1_{A^K}}(D)' = \underline{R_{A^K}}(D)$,

(2) $\overline{R^1_{A^K}}(D)' = \overline{R_{A^K}}(D)$,

(3) $POS^1_{\underline{A^K}}(D)' = POS_{A^K}(D)$.

Proof. (1) Following the proof of (1) in Theorem 5, for any $D_i \in U/D$, it can be obtained: $\underline{R^1_{A^K}}(D_i)' = \underline{R_{A^K}}(D_i)$.

Thus: $\bigcup_{i=1}^r \underline{R^1_{A^K}}(D_i)' = \bigcup_{i=1}^r \underline{R_{A^K}}(D)$.

Which implies:

$$\underline{R^1_{A^K}}(D)' = \underline{R_{A^K}}(D).$$

Properties (2) and (3) can be proved similarly to (1). □

Theorems 5 and 6 show that when $\beta = 1$, the upper approximation, lower approximation, and positive region in the GMRSM are special cases of those in the MVPGMRSM. This implies that MVPGMRSM degenerates into GMRSM when fault tolerance is absent, which indicates that the information measurements in the MVPGMRSM are consistent extensions of those in the GMRSM.

The following two theorems explain the relationship between the probabilistic threshold β and the information measurements in the MVPGMRSM.

Theorem 7. *Consider a GMDS $S = (U, A \cup D)$, where $K \in \Omega^+$, $X \subseteq U$, and $0.5 < \beta_1 \le \beta_2 \le 1$. Then, the following properties hold:*

(1) $\underline{R^{\beta_2}_{A^K}}(X)' \subseteq \underline{R^{\beta_1}_{A^K}}(X)'$,

(2) $\overline{R^{\beta_2}_{A^K}}(X)' \subseteq \overline{R^{\beta_1}_{A^K}}(X)'$,

(3) $POS^{\beta_2}_{\underline{A^K}}(X)' \subseteq POS^{\beta_1}_{\underline{A^K}}(X)'$.

Proof. (1) For any $0.5 < \beta_1 \le \beta_2 \le 1$, according to Definition 6, it is easy to obtain that $\underline{R^{\beta_2}_{A^I}}(X) \subseteq \underline{R^{\beta_1}_{A^I}}(X)$.

Then, $\underline{R_{A^K}}\left(\underline{R^{\beta_2}_{A^I}}(X)\right) \subseteq \underline{R_{A^K}}\left(\underline{R^{\beta_1}_{A^I}}(X)\right)$,

That is,

$$\underline{R^{\beta_2}_{A^K}}(X)' \subseteq \underline{R^{\beta_1}_{A^K}}(X)'.$$

Properties (2) and (3) can be proved similarly to (1). □

Theorem 8. *Consider a GMDS $S = (U, A \cup D)$, where $K \in \Omega^+$, $U/D = \{D_1, D_2, \ldots, D_r\}$, and $0.5 < \beta_1 \le \beta_2 \le 1$. Then, the following properties hold:*

(1) $\underline{R^{\beta_2}_{A^K}}(D)' \subseteq \underline{R^{\beta_1}_{A^K}}(D)'$,

(2) $\overline{R^{\beta_2}_{A^K}}(D)' \subseteq \overline{R^{\beta_1}_{A^K}}(D)'$,

(3) $POS^{\beta_2}_{\underline{A^K}}(D)' \subseteq POS^{\beta_1}_{\underline{A^K}}(D)'$.

Proof. (1) Following the proof of (1) in Theorem 7, for any $D_i \in U/D$, it can be obtained: $\underline{R_{A^{K'}}}(\underline{R^{\beta_2}_{A^I}}(D_i)) \subseteq \underline{R_{A^K}}(\underline{R^{\beta_1}_{A^I}}(D_i))$.

Thus: $\bigcup_{i=1}^r \underline{R_{A^{K'}}}(\underline{R^{\beta_2}_{A^I}}(D_i)) \subseteq \bigcup_{i=1}^r \underline{R_{A^K}}(\underline{R^{\beta_1}_{A^I}}(D_i))$.

Which implies:

$$\underline{R^{\beta_2}_{A^{K'}}}(D)' \subseteq \underline{R^{\beta_1}_{A^K}}(D)'.$$

Properties (2) and (3) can be proved similarly to (1). □

Based on Theorems 7 and 8, it can be deduced that as β increases, the upper approximation, lower approximation, and positive region in the MVPGMRSM exhibit a monotonically decreasing trend.

The following proposition presents the fundamental properties of the MVPGMRSM.

Proposition 2. *Consider a GMDS $S = (U, A \cup D)$, where $K \in \Omega^+$, $0.5 < \beta \leq 1$, $X \subseteq U$, and $Y \subseteq U$. Then, the following properties hold:*

(1) $\underline{R^{\beta}_{A^K}}(\emptyset)' = \overline{R^{\beta}_{A^K}}(\emptyset)' = \emptyset$,

(2) $\underline{R^{\beta}_{A^K}}(U)' = \overline{R^{\beta}_{A^K}}(U)' = U$,

(3) $X \subseteq Y \Rightarrow \underline{R^{\beta}_{A^K}}(X)' \subseteq \underline{R^{\beta}_{A^K}}(Y)'$,

(4) $X \subseteq Y \Rightarrow \overline{R^{\beta}_{A^K}}(X)' \subseteq \overline{R^{\beta}_{A^K}}(Y)'$,

(5) $\underline{R^{\beta}_{A^K}}(X)' \subseteq \overline{R^{\beta}_{A^K}}(X)'$.

According to Definition 7, the above properties can be easily derived. The discussion of the important properties of the MVPGMRSM in this section shows that this model not only retains the fundamental characteristics of the classical VPGMRSM but also preserves the monotonicity of the positive region.

4. Optimal scale combination selection in the MVPGMRSM

OSC selection is a critical and widely studied issue in VPGMRSMs. However, the non-monotonicity of the positive region leads to significant computational costs, and existing algorithms in GMRSMs are not applicable. To address these challenges, this section first introduces the definition of OSC in the classical VPGMRSM. Subsequently, an extended definition of local OSC is proposed based on the monotonic positive region in the MVPGMRSM. Then, a comparison of the time complexity between the two definitions is presented. Finally, two OSC selection algorithms are proposed based on the local OSC definition.

4.1. Optimal scale combination based on positive region

To address the OSC selection challenges caused by the non-monotonicity of the positive region in the VPGMRSM, this subsection first introduces the OSC definition and analyzes its computational limitations. To mitigate the high computational cost, a new OSC definition is further proposed based on the monotonic positive region of the MVPGMRSM. Then, the time complexity of searching for an OSC is compared under both definitions.

Definition 8. [30] Consider a GMDS $S = (U, A \cup D)$, where $K^* \in \Omega^+$, $0.5 < \beta \leq 1$, and $I = (I_1, I_2, \ldots, I_m)$. Then, K^* is called a β OSC of S under the VPGMRSM if and only if the following conditions hold:

$$\begin{aligned} &(1)\quad POS^{\beta}_{A^{K^*}}(D) = POS^{\beta}_{A^I}(D);\\ &(2)\quad \forall K \in \Omega^+, K \prec K^*, \text{ then } POS^{\beta}_{A^K}(D) \neq POS^{\beta}_{A^I}(D). \end{aligned} \tag{15}$$

In Definition 8, condition (1), known as the consistency criterion, ensures that K^* is β positive region consistent with the finest scale combination I. Therefore, a scale combination satisfying condition (1) is a consistent scale combination; otherwise, it is an inconsistent scale combination. Condition (2) guarantees that all strictly coarser scale combinations are inconsistent, making K^* the coarsest among all consistent combinations. However, due to the non-monotonicity of the positive region in the VPGMRSM, verifying condition (2) requires checking all strictly coarser scale combinations, which leads to an exponential number of comparisons in the worst case. Under the assumption of only considering the number of consistency checks, taking the selection of a β OSC as an example, if the number of attributes is m and the number of scales for each attribute does not exceed s, the time complexity for the number of required consistency checks is $O(s^m)$.

To reduce the computational burden associated with β OSC selection in the VPGMRSM, a local β OSC definition is proposed based on the MVPGMRSM, where the positive region is monotonic.

Definition 9. Let $K = (l_1, \ldots, l_m) \in \Omega^+$. For any attribute a_j $(1 \leq j \leq m)$ with $l_j > 1$, the subscale combination K'_j is defined as the result of coarsening the scale of a_j by one level:

$$K'_j = (l_1, \ldots, l_{j-1}, l_j - 1, l_{j+1}, \ldots, l_m). \tag{16}$$

The set of all valid subscale combinations of K in Ω^+ is defined as:

$$Child_{\Omega^+}(K) = \left\{ K'_j \in \Omega^+ \;\middle|\; K'_j = (l_1, \ldots, l_{j-1}, l_j - 1, l_{j+1}, \ldots, l_m), j = 1, 2, \ldots, m \right\}. \tag{17}$$

Table 3
Comparison of three models.

Features	MVPGMRSM	VPGMRSM	GMRSM
Fault tolerance	✓	✓	✗
Monotonicity	✓	✗	✓
Support for local OSC	✓	✗	✓
Applicability of existing algorithms	✓	✗	✓

Definition 10. Consider a GMDS $S = (U, A \cup D)$, where $K^* \in \Omega^+$, $0.5 < \beta \leq 1$, and $I = (I_1, I_2, \ldots, I_m)$. Then, K^* is called a local β OSC of S under the MVPGMRSM if and only if the following conditions hold:

(1) $POS^{\beta}_{A^{K^*}}(D)' = POS^{\beta}_{A^I}(D)'$,

(2) $\forall K \in Child_{\Omega^+}(K^*)$, then $POS^{\beta}_{A^K}(D)' \neq POS^{\beta}_{A^I}(D)'$.

Due to the monotonicity of the positive region in the MVPGMRSM, it is sufficient to check only $Child_{\Omega^+}(K^*)$, rather than all strictly coarser scale combinations. This significantly reduces the computational burden, since monotonicity ensures that if a subscale combination K' is not a consistent scale combination, then any scale combination coarser than K' must also be inconsistent. Therefore, condition (2) guarantees that K^* is the coarsest consistent scale combination. Given that each iteration only checks $O(m)$ immediate subscale combinations and each attribute can be coarsened up to s times, the worst-case number of consistency checks required during the selection of a local β OSC is $O(m \cdot s)$.

In the classical VPGMRSM, the non-monotonic positive region requires checking all strictly coarser scale combinations, leading to exponential time complexity $O(s^m)$, where m is the number of attributes and s is the maximum scale levels per attribute. In contrast, local OSC selection in the MVPGMRSM exploits monotonicity and only needs to verify immediate subscale combinations, reducing complexity to $O(m \cdot s)$. Therefore, MVPGMRSM adopts a local strategy to effectively reduce the computational cost caused by non-monotonicity in the VPGMRSM. To more intuitively demonstrate the advantages of MVPGMRSM, Table 3 provides a brief comparison of the models. As shown in the table, the proposed MVPGMRSM combines the advantages of both VPGMRSM and GMRSM.

4.2. Optimal scale combination selection algorithms

In this subsection, algorithms for OSC selection based on the MVPGMRSM are proposed. The lattice model [24] and the extended stepwise OSC selection algorithm [44] are classical methods for OSC selection in GMRSMs. However, due to the non-monotonicity of the positive region in the VPGMRSM, these methods cannot be directly applied. Although the OSCs obtained by these methods satisfy the consistency condition, they are not guaranteed to be the coarsest scale combinations. Moreover, the lattice model identifies all OSCs through exhaustive enumeration, which makes it unsuitable for large-scale datasets. In contrast, the extended stepwise OSC selection algorithm can obtain an OSC in large-scale datasets, but since it examines scale combinations sequentially from fine to coarse, it may lead to unnecessary computations. Based on these observations, this paper proposes two improved OSC selection algorithms in the MVPGMRSM: Algorithm 1 introduces binary search to quickly find an OSC; Algorithm 2 applies three-way decision theory generates all OSCs.

Algorithm 1 combines a heuristic method with binary search. It starts with an initial middle scale combination mid and iteratively refines it to obtain the final mid. For example, in Table 2, $I = (3, 3, 2)$ and $CSC = (1, 1, 1)$, then initial $mid = \lceil (3+1, 3+1, 2+1)/2 \rceil = (2, 2, 2)$. If mid is a consistent scale combination, it is updated as $mid = \lceil (mid + CSC)/2 \rceil$; otherwise, it is updated as $mid = \lceil (mid + I)/2 \rceil$, until obtaining the final mid. Then, the optimal scale for each attribute in mid is determined stepwise. The algorithm is composed of two primary parts. The first part uses binary search to determine the final mid. The second part adjusts the scale of each attribute to find the OSC. The consistency check essentially involves partitioning equivalence classes, which is equivalent to computing U/A^{sc}, and the time complexity of $O(m|U|)$.

In the first part of Algorithm 1, the search space is halved in each iteration, resulting in at most $O(\log K)$ iterations, where K denotes the maximum scale value in I. For instance, in Table 2, $I = (3, 3, 2)$, hence $K = 3$. Since each iteration requires consistency check, the complexity is $O(\log K \cdot m|U|)$. In the second part, in the worst case, adjusts the scale of each attribute from I_i to CSC_i, resulting in a complexity of $O(|U| \cdot (I_i - CSC_i))$. Based on the above analysis, the time complexity of Algorithm 1 is $O\left(m|U| \sum_{i=1}^{m} (I_i - CSC_i)\right)$, where $|U|$ denotes the cardinality of the universe of discourse, and m is the total number of attributes.

Definition 11. Consider a GMDS $S = (U, A \cup D)$, where $K \in \Omega^+$. The upper bound and lower bound of K are defined as:

$$UB_{\Omega^+}(K) = \{K' \mid K' \in \Omega^+, K' \succ K\},$$
$$LB_{\Omega^+}(K) = \{K' \mid K' \in \Omega^+, K' \prec K\}.$$

Theorem 9. *Consider a GMDS $S = (U, A \cup D)$, where $K \in \Omega^+$ and $0.5 < \beta \leq 1$. Then:*

(1) *If K is a local β OSC of S in Ω^+, then for any $K' \in (UB_{\Omega^+}(K) \cup LB_{\Omega^+}(K))$, K' is not a local β OSC of S,*

Algorithm 1 Binary Search for finding a local β OSC.

Input: A GMDS $S = (U, A \cup D)$, precision β, finest scale combination I, coarsest scale combination CSC.
Output: A local β OSC and inconsistent scale combination set ISC.

$ISC \leftarrow \emptyset$;
$lower \leftarrow CSC$;
$upper \leftarrow I$;
while true **do**
  $mid \leftarrow \lceil (lower + upper)/2 \rceil$;
  Compute $POS^{\beta}_{A^{mid}}(D)'$;
  if S^{mid} *is* β *positive region consistent* **then**
    $upper \leftarrow mid$;
  else
    $lower \leftarrow mid$;
    $ISC \leftarrow ISC \cup \{mid\}$;
  end if
  $new\_mid \leftarrow \lceil (lower + upper)/2 \rceil$;
  if $new\_mid == mid$ **then**
    break;
  end if
end while
$OSC \leftarrow mid$;
for i *from* 1 *to* m **do**
  while $OSC_i > CSC_i$ **do** // while the scale of the i-th attribute in OSC is greater than in CSC.
    $sc \leftarrow OSC$;
    $sc \leftarrow (sc_1, \ldots, sc_{i-1}, sc_i - 1, sc_{i+1}, \ldots, sc_m)$; // the scale level of the i-th attribute in sc is decreased by one.
    Compute $POS^{\beta}_{A^{sc}}(D)'$;
    if S^{sc} *is* β *positive region consistent* **then**
      $OSC_i \leftarrow sc_i$; // update the scale of the i-th attribute in OSC with that in SC.
    else
      $ISC \leftarrow ISC \cup \{sc\}$;
      break;
    end if
  end while
end for
return OSC **and** ISC.

(2) *If K is an inconsistent scale combination in Ω^+, then for any $K' \in LB_{\Omega^+}(K)$, K' is not a local β OSC of S.*

Proof. (1) If $K' \in UB_{\Omega^+}(K)$, then $K' \succ K$, indicating that K' is strictly finer than the OSC K. By Theorem 2, the property $POS^{\beta}_{A^K}(D)' \subseteq POS^{\beta}_{A^{K'}}(D)'$ holds. Since K is the OSC, it follows that $POS^{\beta}_{A^K}(D)' = POS^{\beta}_{A^I}(D)'$. It is evident that $K' \preceq I$, and from Theorem 2, it follows that $POS^{\beta}_{A^{K'}}(D)' \subseteq POS^{\beta}_{A^I}(D)'$. Consequently, $POS^{\beta}_{A^{K'}}(D)' = POS^{\beta}_{A^I}(D)'$, demonstrating that K' is a consistent scale combination. By Definition 10, K' does not satisfy the conditions for being a local β OSC of S.

If $K' \in LB_{\Omega^+}(K)$, then $K' \prec K$, which means that K' is strictly coarser than the β OSC K. Since K is the β OSC, by Definition 10, the condition $POS^{\beta}_{A^{K'}}(D)' \neq POS^{\beta}_{A^I}(D)'$ indicates that K' is an inconsistent scale combination. As a result, K' is not an OSC of S. Thus, for any $K' \in UB_{\Omega^+}(K) \cup LB_{\Omega^+}(K)$, K' does not satisfy the conditions for being a local β OSC of S.

(2) If $K' \in LB_{\Omega^+}(K)$, then $K' \prec K$. By Theorem 2, the property $POS^{\beta}_{A^{K'}}(D)' \subseteq POS^{\beta}_{A^K}(D)'$ is satisfied. Since K is an inconsistent scale combination, it follows that $POS^{\beta}_{A^{K'}}(D)' \subseteq POS^{\beta}_{A^K}(D)' \subseteq POS^{\beta}_{A^I}(D)'$, which shows that K' is also an inconsistent scale combination. Therefore, for any $K' \in LB_{\Omega^+}(K)$, K' does not satisfy the conditions for being a local β OSC of S. □

Definition 1 shows that the universe of discourse U can be divided into three regions based on the three-way decision theory. Similarly, applying the three-way decision theory and Theorem 9, the set Ω^+ can also be divided into three regions.

Definition 12. Consider a GMDS $S = (U, A \cup D)$, where $0.5 < \beta \leq 1$. Let $K^* \in \Omega^+$ be a local β OSC of S, and ISC represent the set of inconsistent scale combinations in Ω^+. Then, the partitioning of Ω^+ is defined as follows:

$$\begin{aligned} POS &= \{K^*\}, \\ NEG &= UB_{\Omega^+}(K^*) \cup LB_{\Omega^+}(K^*) \cup ISC \cup \bigcup_{K \in ISC} LB_{\Omega^+}(K), \\ BND &= \Omega^+ - POS - NEG. \end{aligned} \tag{18}$$

According to Definition 12, the set Ω^+ can be partitioned into three regions. The positive region (POS) contains the known OSC. The negative region (NEG) includes scale combinations that are strictly coarser or finer than the known OSC, inconsistent combinations, and those strictly coarser than inconsistent ones. The boundary region (BND) consists of all remaining scale combinations.

Algorithm 2 Combining three-way decision to obtain all OSCs.

Input: A GMDS $S = (U, A \cup D)$ and precision β.
Output: All optimal scale combinations $OSCs$.

$OSCs \leftarrow$ OSC, $ISCs \leftarrow$ ISC; (Algorithm 1)
$POS \leftarrow OSCs$, $NEG \leftarrow ISCs$, $BND \leftarrow \emptyset$;
Initialize Ω^+;
for all $K \in \Omega^+$ **do** (Definition 12)
  if $K \in POS$ **or** $K \in NEG$ **then**
    continue;
  end if
  if $K \succ$ OSC **or** $K \prec$ OSC **or** $K \in LB_{\Omega^+}(ISCs)$ **then**
    $NEG \leftarrow NEG \cup \{K\}$;
  else
    $BND \leftarrow BND \cup \{K\}$;
  end if
end for
for i *from* 1 **to** size(BND, 1) **do**
  $sc \leftarrow BND[i]$;
  Compute $POS^{\beta}_{A^{sc}}(D)'$;
  if S^{sc} *is* β *positive region consistent* **then**
    $POS \leftarrow POS \cup \{sc\}$;
  end if
end for
$OSCs \leftarrow POS$;
for all $K \in OSCs$ **do**
  for all $scale \in OSCs - \{K\}$ **do**
    if $\exists\, scale \in OSCs$ subject to $K \preceq scale$ **then**
      $OSCs \leftarrow OSCs - \{scale\}$;
    end if
  end for
end for
return $OSCs$.

Algorithm 2 is based on Definition 12 to partition Ω^+ into three regions, and the time complexity of this partitioning process is $O\left(\prod_{i=1}^{m} I_i\right)$. After partitioning, Algorithm 2 first selects the scale combinations in the BND that satisfy the consistency criterion and adds them to the OSC set. Then, it examines the OSC set to ensure that only incomparable scale combinations remain. This process determines all OSCs, and the time complexity is $O(m|U|\prod_{i=1}^{m} I_i)$.

Fig. 1 shows the situation of the three regions obtained by partitioning the scale combinations from Table 2. In the figure, Algorithm 1 identifies $(2, 2, 1)$ as the β OSC, and $\{(1, 2, 2), (2, 1, 2)\}$ as inconsistent scale combinations ISC. The gray region represents the NEG, while the yellow region represents the BND.

Next, an example illustrates the feasibility of Algorithms 1 and 2.

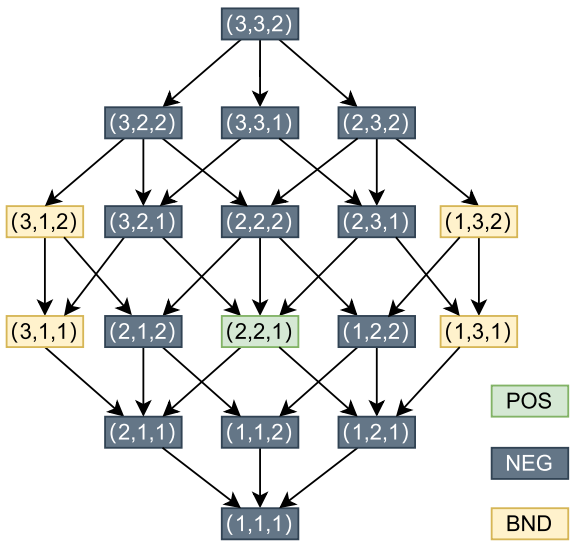


Fig. 1. Using Definition 12 to divide scale combinations into three regions. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

Example 3. (Continued from Example 2) Let $\beta = 0.7$, the finest scale combination $I = (3,3,2)$, the coarsest scale combination $CSC = (1,1,1)$ and $POS^{0.7}_{A^I}(D)' = \{x_1, x_2, x_6, x_7, x_8, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}\}$.

The initial $mid = (2,2,2)$, by Definition 7, the β positive region of D w.r.t. A^{mid} is:

$$POS^{0.7}_{A^{mid}}(D)' = \{x_1, x_2, x_6, x_7, x_8, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}\}.$$

Thus, S^{mid} is β positive region consistent. Therefore, the updated $new\_mid = (2,2,2)$. As $mid = new\_mid$, the current $OSC = (2,2,2)$.

First, checking attribute a_1, since the scale for a_1 at OSC is 2, now by Definition 7, calculate the β positive region of D w.r.t. $A^{(1,2,2)}$:

$$POS^{0.7}_{A^{(1,2,2)}}(D)' = \{x_1, x_6, x_7, x_8, x_{11}, x_{13}, x_{15}, x_{16}\}.$$

Thus, $S^{(1,2,2)}$ is β positive region consistent, so the scale for attribute a_1 can not be coarser. Thus, the optimal scale for a_1 is 2. And $(1,2,2)$ is added to the inconsistent scale combination set ISC.

Next, checking attribute a_2, now by Definition 7, calculate the β positive region of D w.r.t. $A^{(2,1,2)}$:

$$POS^{0.7}_{A^{(2,1,2)}}(D)' = \{x_1, x_2, x_8, x_{11}, x_{12}, x_{15}, x_{16}\}.$$

Thus, $S^{(2,1,2)}$ is inconsistent, so the scale for attribute a_2 can not be coarser. Thus, the optimal scale for a_2 is 2. And $(2,1,2)$ is added to the inconsistent scale combination set ISC.

For attribute a_3, now by Definition 7, calculate the β positive region of D w.r.t. $A^{(2,2,1)}$:

$$POS^{0.7}_{A^{(2,2,1)}}(D)' = \{x_1, x_2, x_6, x_7, x_8, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}\}.$$

Thus, $S^{(2,2,1)}$ is β positive region consistent, and 1 is already the coarsest scale, so the optimal scale the scale for attribute a_3 is 1. Finally, by using Algorithm 1, the results are $OSC = (2,2,1)$ and the inconsistent scale combination set $ISC = \{(1,2,2),(2,1,2)\}$. Now Algorithm 2 is used to obtain BND:

$$BND = \{(3,1,1),(1,3,1),(3,1,2),(1,3,2)\}.$$

By Definition 7, calculate the β positive regions for each scale combination in BND:

$$POS^{0.7}_{A^{(3,1,1)}}(D)' = \{x_1, x_2, x_8, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}\} \neq POS^{0.7}_{A^I}(D)',$$
$$POS^{0.7}_{A^{(1,3,1)}}(D)' = \{x_1, x_6, x_7, x_8, x_{11}, x_{12}, x_{13}, x_{15}, x_{16}\} \neq POS^{0.7}_{A^I}(D)',$$
$$POS^{0.7}_{A^{(3,1,2)}}(D)' = \{x_1, x_2, x_6, x_7, x_8, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}\} = POS^{0.7}_{A^I}(D)',$$
$$POS^{0.7}_{A^{(1,3,2)}}(D)' = \{x_1, x_6, x_7, x_8, x_{11}, x_{12}, x_{13}, x_{15}, x_{16}\} \neq POS^{0.7}_{A^I}(D)'.$$

Thus, $POS = \{(2,2,1),(3,1,2)\}$.

Finally, after checking the scale combinations in POS, the OSC set is output:

$$OSCs = \{(2,2,1),(3,1,2)\}.$$

This example shows that Algorithm 1 only needs four consistency checks to find the β OSC. This reduces the number of evaluations compared to the extended stepwise OSC selection algorithm. Similarly, Algorithm 2 also uses four consistency checks to find all β OSCs. This greatly lowers the computational cost compared to the lattice model. Since both algorithms leverage the monotonicity of the positive region in the MVPGMRSM, they not only achieve high computational efficiency but also ensure the correctness of the obtained OSCs.

5. Experiments and analysis

This section conducts a series of experiments to evaluate the effectiveness of the MVPGMRSM proposed in Section 3 and the algorithms proposed in Section 4. 12 publicly available datasets from the UCI repository are utilized in the experiments. All experimental programs are implemented in MATLAB and executed on a personal computer. Table 4 outlines the operating environment, while Table 5 provides detailed information about the 12 datasets.

As the datasets listed in Table 5 are SDSs, the transformation into corresponding GMDSs is performed using the method proposed in [32], and the core steps briefly described below.

Step 1: Given an attribute a_j, define its value set as $V_j = \{a_j(x) \mid x \in U\}$. And δ_j denotes the standard deviation of V_j, while m_j represents its minimum value. Then, for any $x \in U$, the attribute value at the finest scale combination is computed as: $a_j^{I_j}(x) = \left\lfloor \frac{a_j(x) - m_j}{\delta_j} \right\rfloor$, where $\lfloor y \rfloor$ denotes the largest integer less than or equal to y.

Table 4
Experimental environment.

Component	Model	Parameter
Platform	MATLAB	R2016b
Operating system	Windows 11	64 bit
CPU	Intel Core i7-10700	2.90 GHz
Memory	DDR4	16 GB; 3200 MHz
Hard disk	TOSHIBA DT01ACA100	1TB

Table 5
Datasets description.

No.	Datasets	Rename	Instances	Features	Classes
1	Seeds	Seeds	210	7	3
2	Glass Identification	Glass	214	9[a]	6
3	Vertebral column (3C)	3C	310	6	3
4	Ecoli	Ecoli	336	7	8
5	Auto MPG	Auto	392[b]	7	3
6	Contraceptive Method Choice	CMC	1473	9	3
7	Yeast	Yeast	1484	8	10
8	Wireless Indoor Localization	WIL	2000	7	4
9	Pen-Based Recognition of Handwritten Digits	Pendigits	10992	16	10
10	Polish Companies Bankruptcy	Polish	10503	20[c]	2
11	Magic Gamma Telescope	Magic	19020	10	2
12	Letter Recognition	Letter	20000	16	26

[a] Drop the first feature column (ID number).
[b] After deleting objects with missing values.
[c] After deleting feature columns with missing values.

Step 2: Based on $a_j^{I_j}(x)$, attribute values for each scale combination can be obtained by merging values from top to bottom. For each scale, if the number of equivalence classes exceeds 3, a new scale is added. The new scale is formed by merging values from the previous scale (taking the minimum value during the merge), and this new scale becomes the $(I_j - 1)$-th scale. This process continues until the number of equivalence classes reaches three.

Step 3: For each attribute, **Step 1** and **Step 2** are applied to obtain a complete GMDS.

5.1. Monotonicity experiment

In this subsection, the monotonicity of the positive region in the MVPGMRSM (proposed in Section 3) is verified by comparison with the classical VPGMRSM. The number of elements in the positive region of D is calculated during the stepwise refinement of the scale combination K, from coarse to fine.

The refinement begins with the coarsest scale combination. At each step, one attribute is randomly selected, and its scale value is increased by one. This continues until the finest scale combination I is reached. For example, starting from $(1,1,1)$, if the selected attributes are a_2, a_1, a_1, and a_3, the scale combinations become: $(1,1,1) \to (1,2,1) \to (2,2,1) \to (3,2,1) \to (3,2,2)$. Since $(3,2,2)$ is the finest scale, the stepwise refinement process from coarse to fine is complete. In the experiments, $\beta = 0.7$.

The experimental results are presented in two figures. The results for datasets 1–9 are presented in Fig. 2. For datasets 10–12, the full transformation from coarse to fine scales is too large to observe directly. To improve visibility, Fig. 3 provides two subfigures for each dataset: one shows the complete transformation, and the other zooms in on specific regions for clearer observation. In all subfigures, the horizontal axis represents the scale combinations from coarse to fine, where larger numbers indicate finer scale combinations. For example, the leftmost number 1 in the figure denotes the coarsest scale combination CSC, while the rightmost and largest number represents the finest scale combination I. The vertical axis represents the number of elements in the positive region. In the figures, CPOS denotes the number of elements in the β positive region of the classical VPGMRSM, while MPOS denotes the corresponding number for MVPGMRSM.

It can be observed from Figs. 2 and 3 that CPOS does not exhibit monotonicity in any dataset without strict conditions. Its trend is unstable as the scale combination changes from coarse to fine. In certain datasets, such as Glass, 3C, Auto, and Polish, CPOS at coarser scale combinations is even larger than at the finest scale combination I. In contrast, MPOS exhibits monotonicity. It follows the expected rule that the number of elements in the positive region increases as the scale combination becomes finer, and I yields the largest positive region. Moreover, it can be observed that MPOS and CPOS are equal at I. These results confirm the correctness of Theorems 1, 2, 3 and 4.

5.2. Optimal scale combination selection experiment

This subsection first applies Algorithm 1, the extended stepwise optimal scale combination (ESOSC) in [44], and the positive region stepwise optimal scale combination (PSOSC) in [27] to obtain OSCs. Both ESOSC and PSOSC are OSC selection methods originally

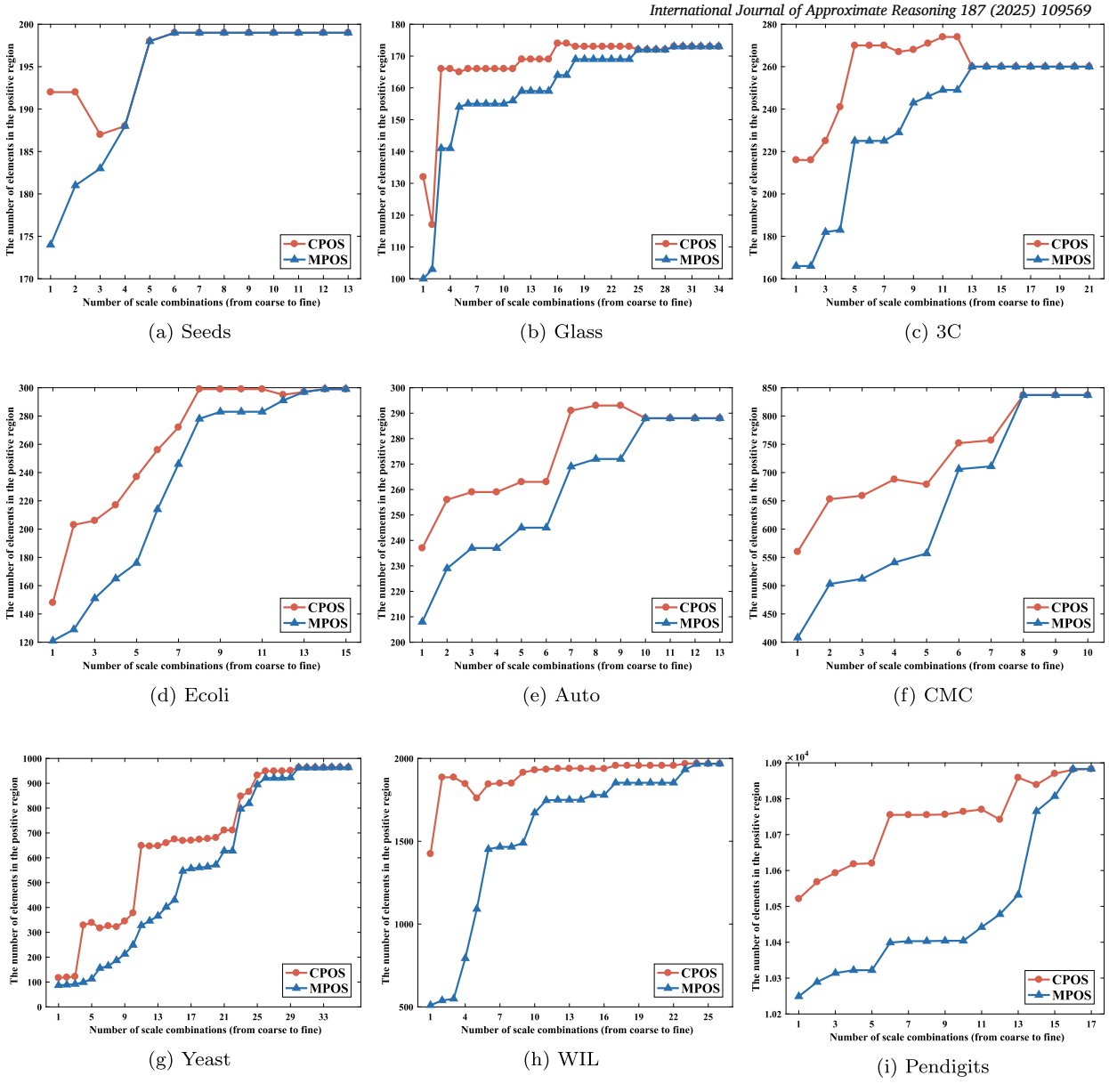


Fig. 2. Positive regions on datasets 1-9.

designed for GMRSMs. In this experiment, they are applied using the positive region defined in the classical VPGMRSM, with the aim of demonstrating that OSC selection methods designed for GMRSMs cannot be directly applied to VPGMRSMs. Subsequently, the effectiveness of Algorithms 1 and 2 in reducing computational costs is evaluated. In the experiments, $\beta = 0.7$.

Table 6 presents the OSCs obtained by Algorithm 1, ESOSC, and PSOSC. The OSCs that differ between Algorithm 1 and ESOSC are highlighted in bold in the table. Notably, some attribute scales in PSOSC are shown as 0, which occurs because PSOSC is a reduction algorithm where 0 indicates removable attributes. Due to space limitations and the complexity of calculating the strictly coarser scale combinations of the OSC, consider the OSC of the Magic dataset, obtained through Algorithm 1, which is $(2, 3, 3, 3, 5, 9, 8, 10, 2, 4)$. Therefore, the number of scale combinations for this OSC and its strictly coarser combinations is $2 \times 3 \times 3 \times \cdots \times 4$, leading to a graph that is too dense for effective observation and analysis. Consequently, Figs. 4, 5 and 6 display the OSCs and their corresponding positive regions for the Auto, 3C, Ecoli, and Glass datasets.

Similar results are observed across all datasets. Specifically, Fig. 4 shows the results from ESOSC, Fig. 5 illustrates the results from PSOSC and Fig. 6 presents the results from Algorithm 1. In all figures, the horizontal axis represents the scale combinations (starting from the OSC, followed by strict coarser scale combinations), and the vertical axis indicates the size of the positive region. Data points where the positive region is equal to or larger than that of the OSC are marked in dark blue, while those smaller than the OSC are marked in light blue.

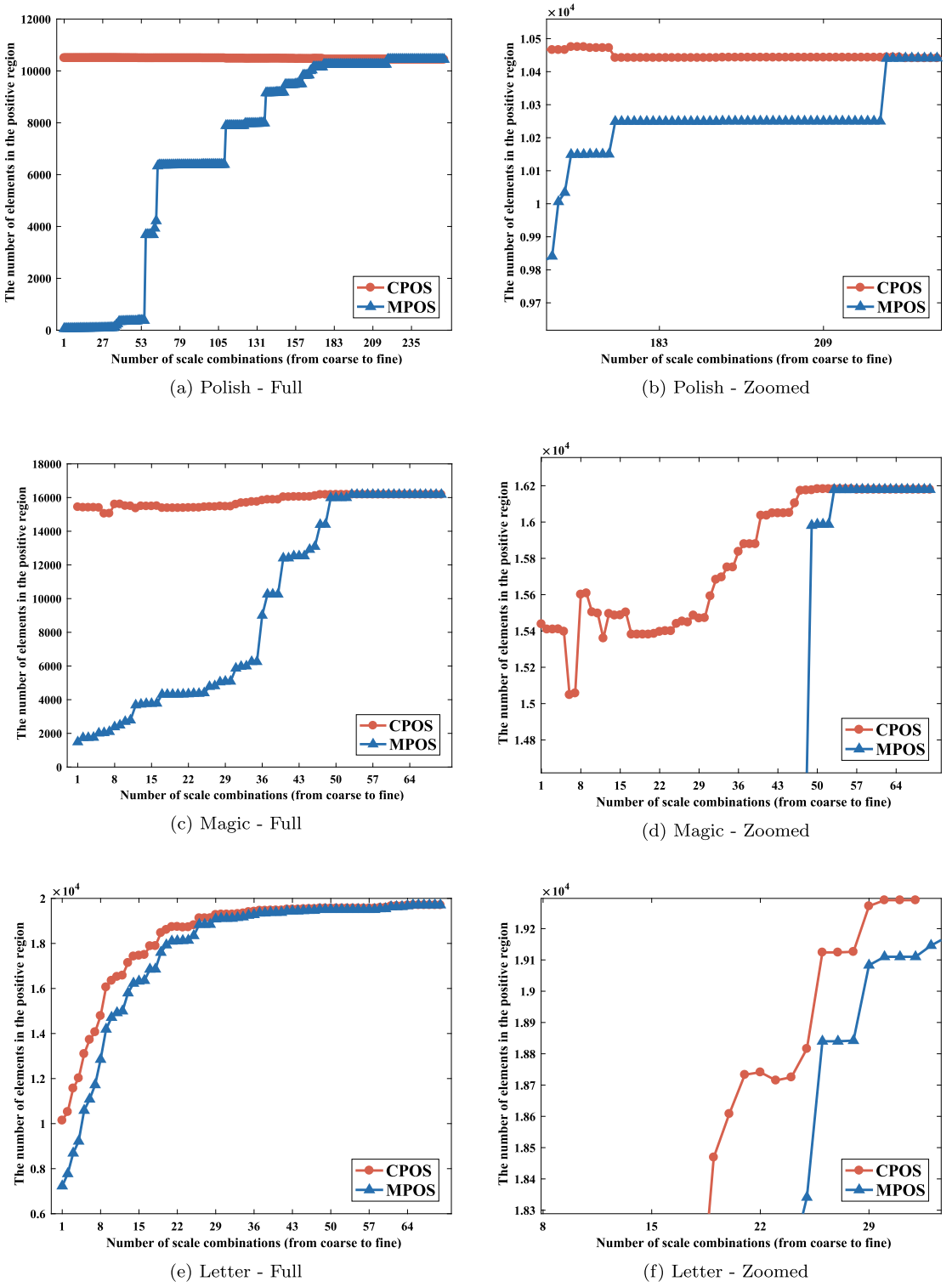
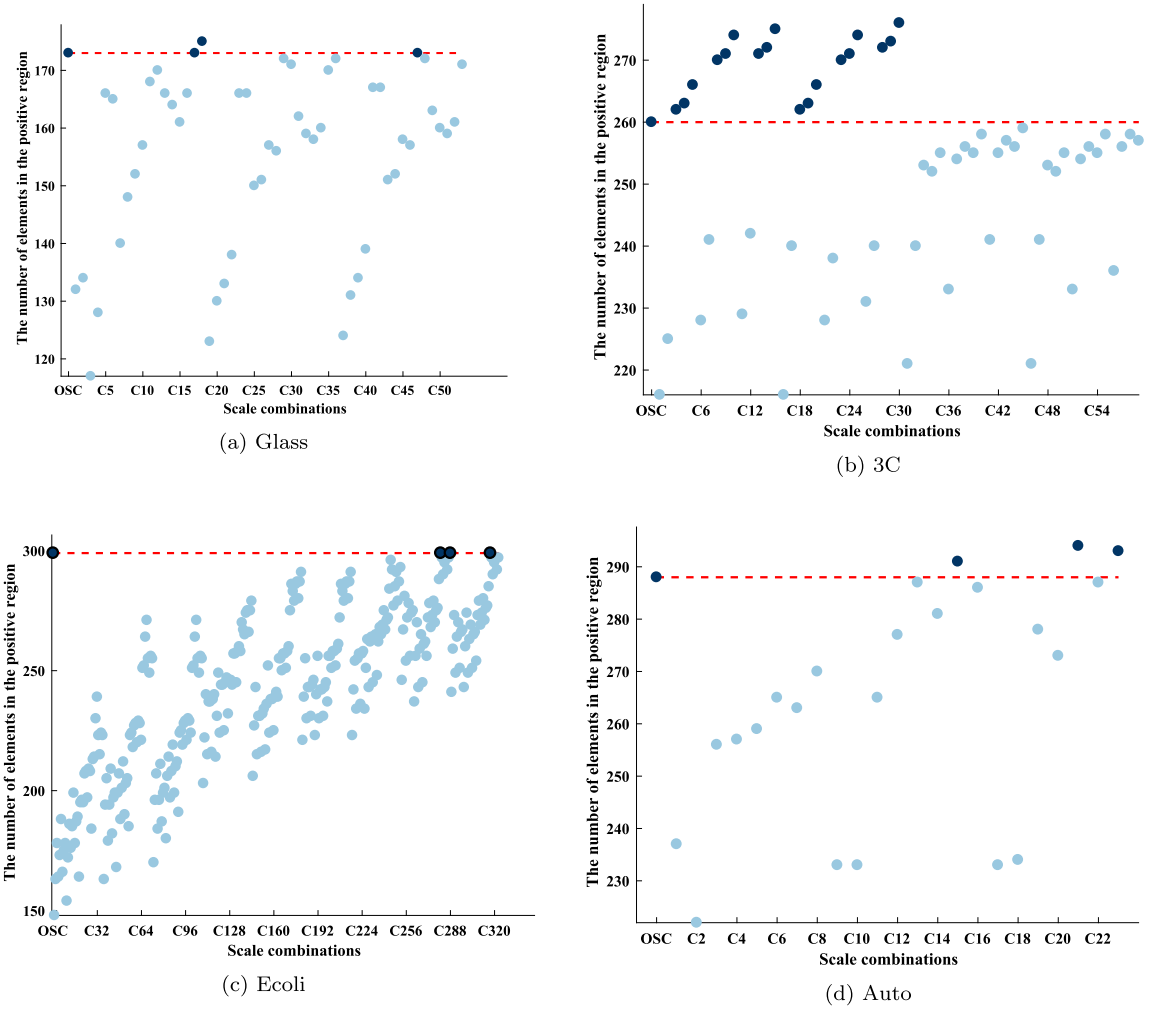


Fig. 3. Positive regions on datasets 10-12.

Table 6
OSCs obtained by three algorithms.

No.	Scales	Algorithm 1	ESOSC	PSOSC
1	2×2×3×3×2×4×3	**(1,1,3,1,1,2,2)**	(1,1,3,1,1,2,1)	(1,0,3,1,1,2,1)
2	6×6×2×5×6×5×3	(3,3,1,1,3,1,2,1,1)	(3,3,1,1,3,1,2,1,1)	(3,3,1,1,3,0,2,1,1)
3	5×4×4×4×5×4	(1,2,2,3,5,1)	(1,2,2,3,5,1)	(1,2,2,3,5,1)
4	3×4×1×1×6×3×3	(3,3,1,1,4,3,3)	(3,3,1,1,4,3,3)	(3,3,1,0,4,3,3)
5	3×1×2×3×3×5×2	(3,1,1,1,1,4,2)	(3,1,1,1,1,4,2)	(3,1,1,1,1,4,2)
6	3×1×2×5×1×1×2×2×1	(3,1,2,3,1,1,2,2,1)	(3,1,2,3,1,1,2,2,1)	(3,1,2,3,1,1,2,2,1)
7	5×6×6×6×1×1×10×8	(4,5,4,5,1,1,8,6)	(4,5,4,5,1,1,8,6)	(4,5,4,5,1,1,8,6)
8	4×7×5×4×4×4×4	(2,6,4,2,3,2,3)	(2,6,4,2,3,2,3)	(2,6,4,2,3,2,3)
9	1×5×2×4×1×2×2×2×1×2 ×1×2×3×2×1×1	(1,5,2,4,1,2,2,2,1,1,1, 1,3,2,1,1)	(1,5,2,4,1,2,2,2,1,1,1, 1,3,2,1,1)	(1,5,2,4,1,2,2,2,1,1,1, 1,3,2,1,1)
10	19×4×6×10×18×6×19×18 ×17×22×7×9×19×15×6× 24×4×29×16×8	**(10,1,1,8,1,4,9,9,8, 12,6,6,8,3,1,16,1, 13,10,2)**	(10,1,1,8,1,4,9,9,8,12, 6,6,8,3,1,16,1,13,10,1)	(10,0,0,8,0,4,9,9,8,12, 6,6,8,3,0,16,0,13,10,0)
11	6×11×6×3×5×15×10×16× 2×5	(2,3,3,3,5,9,8,10,2,4)	(2,3,3,3,5,9,8,10,2,4)	(2,3,3,3,5,9,8,10,2,4)
12	6×3×6×5×5×6×5×4×5×5 ×4×6×5×8×4×8	**(4,3,3,3,4,4,5,3,3,3, 3,5,3,6,2,6)**	(4,3,3,3,4,4,5,3,3,3,2, 5,3,6,2,6)	(4,3,3,3,4,4,5,3,3,3,2, 5,3,6,2,6)



(a) Glass (b) 3C

(c) Ecoli (d) Auto

Fig. 4. Positive regions of the OSC obtained by ESOSC and its all strictly coarser scale combinations.

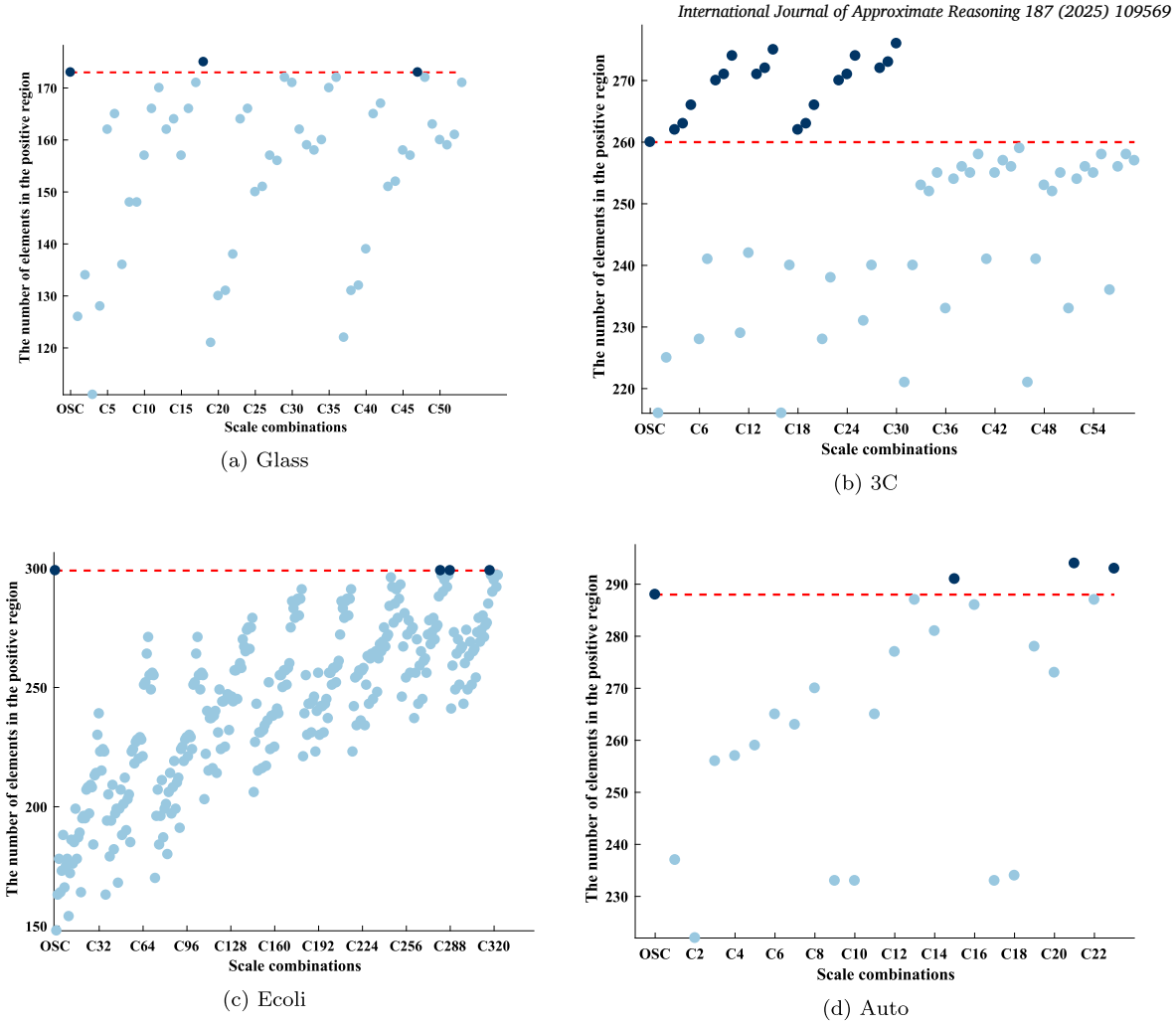


Fig. 5. Positive regions of the OSC obtained by PSOSC and its strictly coarser scale combinations.

As shown in Figs. 4 and 5, the OSCs obtained by ESOSC and PSOSC are not the coarsest consistent scale combinations. Their strictly coarser counterparts have positive regions that are equal to or larger to those of the OSCs, which violates the definition of OSC. In contrast, Fig. 6 demonstrates that the OSCs obtained by Algorithm 1 conform to the definition of the OSC. Therefore, the OSC selection methods in GMRSMs cannot be applied directly to VPGMRSMs. By comparison, Algorithm 1 is better suited for VPGMRSMs and enables the effective selection of OSCs.

To demonstrate that the OSC selection methods proposed in this paper (Algorithms 1 and 2) can effectively reduce computational complexity, a comparison is made with Algorithm 1, ESOSC, PSOSC, Reduct-based extended stepwise optimal scale combination (RESOSC) in [43], Random-order extended stepwise optimal scale combination (ROESOSC) in [47], the stepwise optimal scale combination (SOSC) in [32], and the global optimal scale combination (GOSC) in [45], all of which aim to quickly select a single OSC. Furthermore, Algorithm 2 is compared with the lattice model (LM) in [24], with the aim of demonstrating that Algorithm 2 is capable of obtaining all OSCs. Considering that the primary purpose of this experiment is to highlight the time efficiency of the proposed algorithms and ensure that all methods can effectively obtain β OSCs, the β positive region in the MVPGMRSM is used as the information measurement in the consistency criterion for all algorithms.

LM is essentially an enumeration method, so it is not suitable for large datasets, and in practical applications, it is usually unnecessary to obtain all OSCs. Therefore, in the experiments, datasets 1-12 are used to select an OSC, while datasets 1-9 are used to select all OSCs.

Table 7 presents the runtime required for Algorithm 1 and the other six algorithms to obtain an β OSC. The runtime for each algorithm is the average of 10 repeated runs, with the shortest runtime for each dataset highlighted in bold. Table 7 reveals that Algorithm 1 achieves the shortest running time on most datasets, significantly outperforming the SOSC and GOSC algorithms, especially on large-scale datasets. PSOSC, RESOSC, and ROESOSC are all reduction algorithms, which generally require more time than ESOSC. However, compared to ESOSC, Algorithm 1 demonstrates better computational efficiency, and this advantage becomes increasingly evident with larger datasets. This is mainly because Algorithm 1 adopts a binary search strategy, first narrowing the interval where

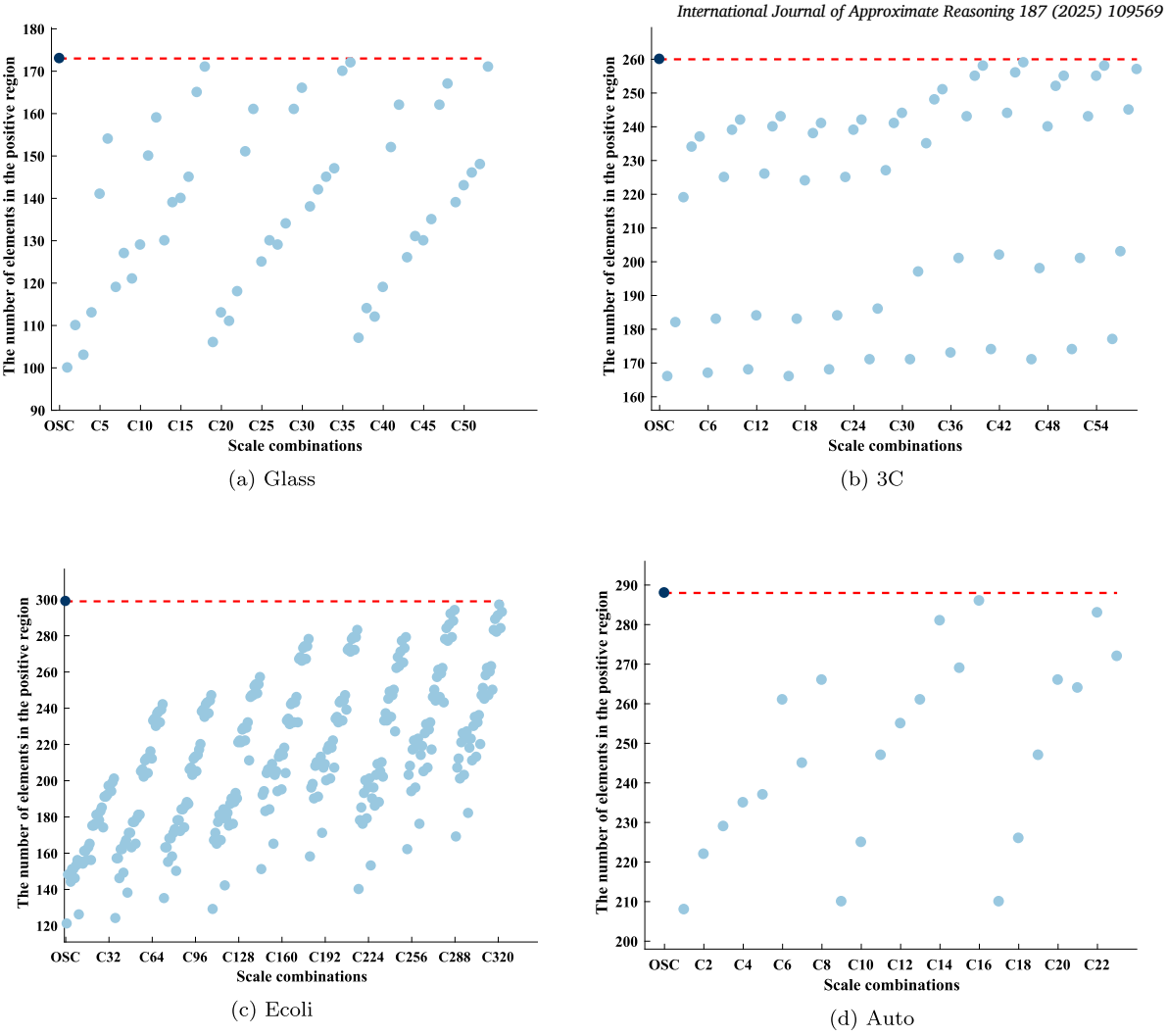


Fig. 6. Positive regions of the OSC obtained by Algorithm 1 and its all strictly coarser scale combinations.

Table 7
Running time of the seven algorithms to obtain an OSC.

No.	Time (s)						
	Algorithm 1	ESOSC	PSOSC	RESOSC	ROESOSC	SOSC	GOSC
1	0.0290	**0.0254**	0.0384	0.0378	0.0392	0.1186	0.0571
2	**0.0613**	0.0703	0.1011	0.1004	0.0986	0.4454	0.4733
3	0.0665	**0.0498**	0.0656	0.0668	0.0678	0.2382	0.1397
4	**0.0689**	0.1314	0.1653	0.1682	0.1688	0.4321	0.1234
5	**0.0337**	0.0421	0.0597	0.0629	0.0600	0.1604	0.0684
6	**0.1357**	0.2724	0.3979	0.4046	0.4088	0.8380	0.2558
7	**0.5317**	1.3333	1.5667	1.5791	1.5943	4.4023	1.9567
8	**0.5230**	0.5667	0.7145	0.7111	0.7021	2.2527	1.4012
9	**6.5642**	14.8323	21.7697	21.8025	21.6225	43.0984	12.2197
10	**5.4620**	6.8924	7.4401	7.2456	7.1559	36.6631	68.8935
11	**16.4815**	20.7653	23.3594	23.4774	23.4363	68.5799	67.6216
12	**108.1311**	145.5896	184.2947	183.8706	184.1281	535.1585	594.1721

OSC may occur and then selecting the optimal scale for each attribute within this interval, thus reducing the number of consistency checks.

Table 8 presents the runtimes and β OSCs obtained by Algorithm 2 and LM, with the shortest runtime highlighted in bold. As shown in Table 8, although Algorithm 2 and LM share the same time complexity, Algorithm 2 achieves higher time efficiency and

Table 8
Performance comparison of Algorithm 2 and LM.

No.	Algorithm 2		LM	
	Time (s)	OSCs	Time (s)	OSCs
1	**1.0414**	(1,1,3,1,1,2,2) (1,2,3,1,1,2,1) (1,1,3,2,1,2,1)	1.9297	(1,1,3,1,1,2,2) (1,2,3,1,1,2,1) (1,1,3,2,1,2,1)
2	**1908.7682**	(3,3,1,1,3,1,2,1,1)	> two hours	(3,3,1,1,3,1,2,1,1)
3	**15.4329**	(1,2,2,3,5,1) (1,2,3,2,5,1) (2,2,1,3,5,1)	21.7891	(1,2,2,3,5,1) (1,2,3,2,5,1) (2,2,1,3,5,1)
4	**1.7516**	(3,3,1,1,4,3,3)	3.7803	(3,3,1,1,4,3,3)
5	**1.1087**	(3,1,1,1,1,4,2)	1.5940	(3,1,1,1,1,4,2)
6	**0.5527**	(3,1,2,3,1,1,2,2,1)	1.7713	(3,1,2,3,1,1,2,2,1)
7	**2187.1351**	(4,5,4,5,1,1,8,6)	> two hours	(4,5,4,5,1,1,8,6)
8	**485.3824**	(2,6,4,2,3,2,3)	679.6218	(2,6,4,2,3,2,3)
9	**115.7064**	(1,5,2,4,1,2,2,2,1,1,1,1,3,2,1,1)	> one hour	(1,5,2,4,1,2,2,2,1,1,1,1,3,2,1,1)

shorter runtimes. Furthermore, it is observed that the OSCs obtained by Algorithm 2 are consistent with those of LM. Therefore, Algorithm 2 can efficiently identify all OSCs.

5.3. Impact analysis of β values on OSCs and classification experiment

This subsection first analyzes the impact of different β values on OSCs, followed by classification experiments based on these OSCs. Specifically, β values are selected from the interval $(0.5, 1]$ with a step size of 0.1. Since the endpoint 0.5 cannot be used, 0.51 is chosen as the starting value. The final set of selected β values is: (0.51, 0.6, 0.7, 0.8, 0.9, 1).

In the classification experiments, the performance of the following types of scale combinations was compared: (1) MOSC, the β OSC obtained by Algorithm 1 based on the monotonic β positive region in the MVPGMRSM; (2) VSC, the scale combination obtained by ESOSC based on the non-monotonic β positive region in the VPGMRSM; (3) FSC, the finest scale combination; and (4) CSC, the coarsest scale combination.

Due to space limitations, Table 9 presents the scale combinations obtained from the Seeds, Glass, WIL, and Polish datasets under six different β values. If the scale combination corresponding to a given β differs from that of the previous β, it is highlighted in bold. For example, in the Seeds dataset, the MOSC at $\beta = 0.8$ differs from that at $\beta = 0.7$, and is therefore marked in bold.

Although Table 9 only presents the results of four datasets, similar patterns are observed across all 12 datasets:

- With decreasing β values, both MOSC and VSC tend to become coarser, as observed in datasets such as Seeds, WIL, and Polish. This is because smaller β values provide higher fault tolerance, allowing consistency to be maintained even under coarser scale combinations. This finding further confirms that the proposed variable precision model possesses fault tolerance. Compared with the GMRSM model, the proposed model can better adapt to real-world datasets with noise or uncertainty. This flexibility enables the model to exhibit greater robustness and adaptability when dealing with complex or incomplete information, effectively avoiding the misjudgments or performance degradation that traditional models tend to suffer from in the presence of noisy data.
- When $\beta = 1$, MOSC and VSC are identical. This is an expected result, as both MVPGMRSM and VPGMRSM degenerate into the GMRSM when $\beta = 1$, leading to the same OSC. Further comparison between MOSC and VSC shows that, at the same β value, some datasets produce identical results (e.g., Glass and WIL), while others (e.g., Seeds and Polish) exhibit differences.
- In certain datasets, MOSCs obtained under different β values may be incomparable or completely identical. For example, in the Polish dataset, the MOSCs at $\beta = 0.9$ and $\beta = 1$ are incomparable, while in the Glass dataset, the six MOSCs remain exactly the same with different values of β.

In the final classification experiments, the performance of MOSC and VSC was obtained by averaging the classification accuracies under the six β values. To evaluate this performance, several widely used machine learning algorithms were employed, including logistic regression (LR), support vector machine (SVM), and k-nearest neighbors (KNN). A 10-fold cross-validation method was applied, with the accuracy for each experiment represented by the average of the 10-fold results and the standard deviation (Sd) of the test accuracy also reported.

Tables 10 to 12 display the results achieved using the LR, SVM, and KNN classifiers, respectively. The final row summarizes the average classification accuracy and Sd, where the highest classification accuracy in each row is highlighted in bold. The following observations can be made from Tables 10 to 12:

- MOSC achieves the best classification performance on most datasets and consistently achieves the highest average accuracy. This indicates that, in real-world scenarios where data often contain noise, OSCs derived from the positive region of the MVPGMRSM are more robust and effective for classification tasks. The accuracy of MOSC highlights the advantages of incorporating fault tolerance and monotonicity into the model.

Table 9
Scale combinations under different β values.

Dataset	β	MOSC	VSC
Seeds	0.51	(1,1,3,1,1,2,2)	(1,1,3,1,1,2,1)
	0.6	(1,1,3,1,1,2,2)	(1,1,3,1,1,2,1)
	0.7	(1,1,3,1,1,2,2)	(1,1,3,1,1,2,1)
	0.8	**(1,1,3,1,2,2,2)**	**(1,1,3,1,2,2,2)**
	0.9	(1,1,3,1,2,2,2)	(1,1,3,1,2,2,2)
	1	(1,1,3,1,2,2,2)	(1,1,3,1,2,2,2)
Glass	0.51	(3,3,1,1,3,1,2,1,1)	(3,3,1,1,3,1,2,1,1)
	0.6	(3,3,1,1,3,1,2,1,1)	(3,3,1,1,3,1,2,1,1)
	0.7	(3,3,1,1,3,1,2,1,1)	(3,3,1,1,3,1,2,1,1)
	0.8	(3,3,1,1,3,1,2,1,1)	(3,3,1,1,3,1,2,1,1)
	0.9	(3,3,1,1,3,1,2,1,1)	(3,3,1,1,3,1,2,1,1)
	1	(3,3,1,1,3,1,2,1,1)	(3,3,1,1,3,1,2,1,1)
WIL	0.51	(2,6,3,2,3,2,3)	(2,6,3,2,3,2,3)
	0.6	(2,6,3,2,3,2,3)	(2,6,3,2,3,2,3)
	0.7	**(2,6,4,2,3,2,3)**	**(2,6,4,2,3,2,3)**
	0.8	(2,6,4,2,3,2,3)	(2,6,4,2,3,2,3)
	0.9	(2,6,4,2,3,2,3)	(2,6,4,2,3,2,3)
	1	(2,6,4,2,3,2,3)	(2,6,4,2,3,2,3)
Polish	0.51	(10,1,1,7,1,4,8,9,8,12,6,6, 8,3,1,16,1,13,10,2)	(10,1,1,7,1,4,8,9,8,11,6,6, 8,2,1,15,1,13,10,1)
	0.6	(10,1,1,7,1,4,8,9,8,12,6,6, 8,3,1,16,1,13,10,2)	(10,1,1,7,1,4,8,9,8,11,6,6, 8,2,1,15,1,13,10,1)
	0.7	**(10,1,1,8,1,4,9,9,8,12,6,6, 8,3,1,16,1,13,10,2)**	**(10,1,1,8,1,4,9,9,8,12,6,6, 8,3,1,16,1,13,10,1)**
	0.8	**(10,1,1,8,1,4,9,9,8,12,6,6, 8,3,1,16,1,16,10,2)**	**(10,1,1,8,1,4,9,9,8,12,6,6, 8,3,1,16,1,16,10,1)**
	0.9	**(10,1,1,8,1,4,9,9,8,12,6,7, 8,3,1,16,1,16,10,2)**	**(10,1,1,8,1,4,9,9,8,12,6,7, 8,3,1,16,1,16,10,1)**
	1	**(10,1,1,8,1,4,10,10,9,12,6,7, 9,3,1,15,1,16,10,2)**	**(10,1,1,8,1,4,10,10,9,12,6,7, 9,3,1,15,1,16,10,2)**

Table 10
Classification accuracy of scale combinations with LR classifier (%).

No.	MOSC	VSC	FSC	CSC
1	$\mathbf{91.93 \pm 5.35}$	90.26 ± 4.36	90.72 ± 5.12	91.08 ± 4.31
2	$\mathbf{64.78 \pm 9.49}$	61.56 ± 8.45	61.21 ± 6.66	61.20 ± 5.23
3	$\mathbf{81.33 \pm 5.91}$	80.77 ± 6.67	80.87 ± 7.13	76.97 ± 6.26
4	$\mathbf{85.96 \pm 5.48}$	85.32 ± 5.42	84.77 ± 5.22	70.34 ± 5.12
5	68.67 ± 8.65	67.75 ± 8.77	68.12 ± 4.88	$\mathbf{70.21 \pm 5.11}$
6	$\mathbf{49.76 \pm 2.48}$	49.66 ± 3.92	49.55 ± 3.43	48.55 ± 4.56
7	$\mathbf{56.93 \pm 3.78}$	56.17 ± 3.24	56.67 ± 3.42	49.65 ± 3.24
8	$\mathbf{96.44 \pm 1.46}$	96.22 ± 1.55	96.04 ± 1.05	82.96 ± 1.33
9	94.76 ± 0.76	95.12 ± 0.53	$\mathbf{95.16 \pm 0.55}$	91.05 ± 0.67
10	$\mathbf{95.29 \pm 0.05}$	$\mathbf{95.29 \pm 0.05}$	$\mathbf{95.29 \pm 0.05}$	$\mathbf{95.29 \pm 0.05}$
11	$\mathbf{77.19 \pm 1.21}$	77.12 ± 0.91	77.16 ± 1.02	76.33 ± 0.82
12	76.12 ± 1.09	$\mathbf{76.33} \pm 1.18$	76.12 ± 0.95	52.44 ± 1.10
Average	$\mathbf{78.26 \pm 3.81}$	77.63 ± 3.75	77.64 ± 3.29	72.17 ± 3.15

- The classification performance of VSC is generally lower than that of MOSC, while CSC shows significantly poorer accuracy compared to the other three scale combinations. This comparison suggests that CSC cannot capture enough detail for a reliable classification. Therefore, OSC selection plays a critical role in practical applications.

6. Conclusion

The monotonicity of information measurements is crucial for VPGMRSMs and OSC selection. However, information measurements in the classical VPGMRSM generally lack monotonicity, exhibiting it only under strict conditions; therefore, directly using its positive region for OSC selection poses challenges. Furthermore, current OSC selection algorithms are not applicable to VPGMRSMs. To address these issues, a new monotonic VPGMRSM is proposed, and the OSC is defined based on the monotonic positive region. Then, two OSC selection algorithms are proposed. Experimental results confirm the monotonicity of the positive region in the MVPGMRSM and demonstrate that the proposed algorithms are not only well-suited for VPGMRSMs, but also effective in reducing computational costs.

Table 11
Classification accuracy of scale combinations with SVM classifier (%).

No.	MOSC	VSC	FSC	CSC
1	92.88 ± 4.77	92.65 ± 5.86	92.86 ± 5.66	**93.26 ± 6.13**
2	65.16 ± 7.98	**65.32 ± 8.76**	64.64 ± 7.65	61.29 ± 6.34
3	**81.46 ± 4.33**	80.94 ± 5.52	80.78 ± 4.18	75.88 ± 5.83
4	**85.82 ± 5.67**	85.79 ± 5.12	84.33 ± 5.89	69.23 ± 5.77
5	**71.96 ± 5.32**	71.64 ± 7.96	70.14 ± 6.33	70.23 ± 4.39
6	49.82 ± 5.16	49.24 ± 4.33	48.44 ± 4.72	**51.33 ± 3.83**
7	55.56 ± 3.44	55.44 ± 3.65	**55.82 ± 2.23**	47.88 ± 2.04
8	**96.42 ± 0.92**	96.36 ± 0.98	96.34 ± 0.98	83.44 ± 1.31
9	**95.66 ± 0.60**	95.32 ± 0.58	95.38 ± 0.44	91.22 ± 0.72
10	95.27 ± 0.07	95.27 ± 0.06	**95.28 ± 0.07**	95.27 ± 0.06
11	**77.42 ± 0.96**	77.12 ± 0.97	77.24 ± 0.88	76.50 ± 0.97
12	**76.88 ± 1.02**	76.77 ± 0.96	76.42 ± 1.02	53.22 ± 1.12
Average	**78.69 ± 3.35**	78.49 ± 3.73	78.14 ± 3.34	72.40 ± 3.21

Table 12
Classification accuracy of scale combinations with KNN classifier (%).

No.	MOSC	VSC	FSC	CSC
1	**89.69 ± 4.55**	89.41 ± 5.78	88.48 ± 5.22	89.33 ± 6.71
2	**62.85 ± 7.54**	60.02 ± 8.13	62.30 ± 3.83	60.81 ± 7.86
3	**77.44 ± 8.42**	74.50 ± 5.41	74.22 ± 6.43	67.50 ± 5.88
4	80.82 ± 6.77	78.88 ± 5.57	**81.56 ± 5.62**	63.41 ± 4.17
5	**71.92 ± 5.24**	70.67 ± 5.66	71.44 ± 3.21	70.25 ± 4.23
6	**46.86 ± 2.78**	46.47 ± 2.34	46.03 ± 2.77	45.55 ± 3.52
7	**47.84 ± 3.23**	46.76 ± 2.97	47.55 ± 2.28	26.42 ± 1.83
8	**94.72 ± 1.83**	93.88 ± 1.45	93.82 ± 1.66	65.32 ± 1.76
9	96.26 ± 0.55	96.26 ± 0.60	**96.36 ± 0.62**	89.24 ± 0.79
10	**95.29 ± 0.06**	95.26 ± 0.04	95.27 ± 0.06	**95.29 ± 0.05**
11	**78.50 ± 0.82**	78.47 ± 0.75	78.33 ± 0.76	73.65 ± 1.22
12	85.66 ± 0.72	85.86 ± 0.47	**86.40 ± 0.88**	50.14 ± 0.80
Average	**77.32 ± 3.54**	76.37 ± 3.26	76.81 ± 2.78	66.41 ± 3.23

However, while addressing the non-monotonicity of information measurements and the OSC selection problem in VPGMRSMs, this paper also recognizes some limitations. For example, the proposed MVPGMRSM is based on RST, which opens new research directions for exploring information theory methods in the future. Moreover, although Algorithm 2 reduces computational costs, its time complexity remains relatively high. Therefore, subsequent research aims to enhance the model through information theory and optimize the OSC selection algorithm's computational efficiency, thereby enabling broader real-world applications.

CRediT authorship contribution statement

Ruili Guo: Writing – review & editing, Writing – original draft, Validation, Methodology, Data curation, Conceptualization. **Qinghua Zhang:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition. **Yunlong Cheng:** Writing – review & editing, Investigation, Formal analysis, Conceptualization. **Ying Yang:** Writing – review & editing, Formal analysis. **Hang Zhong:** Investigation, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is supported by National Key Research and Development Program of China (No. 2021YFF0704101), Chongqing Natural Science Foundation Joint Fund for Innovation and Development Project (No. CSTB2023NSCQ-LZX0164), National Natural Science Foundation of China (No. 62276038, No. 62221005, No. 62576056).

Data availability

Data will be made available on request.

References

[1] L.A. Zadeh, Fuzzy logic = computing with words, IEEE Trans. Fuzzy Syst. 4 (2) (1996) 103–111.
[2] S. Acharjee, A. Oza, U. Gogoi, Topological aspects of granular computing, in: Advances in Topology and Their Interdisciplinary Applications, Springer Nature, Singapore, Singapore, 2023, pp. 217–228.
[3] X.Y. Zhang, D.D. Guo, W.H. Xu, Two-way concept-cognitive learning with multi-source fuzzy context, Cogn. Comput. 15 (5) (2023) 1526–1548.
[4] Z. Pawlak, Rough sets, Int. J. Comput. Inf. Sci. 11 (1982) 341–356.
[5] Y.Y. Yao, Three-way decisions with probabilistic rough sets, Inf. Sci. 180 (3) (2010) 341–353.
[6] Y.Y. Yao, Three-way decision and granular computing, Int. J. Approx. Reason. 103 (2018) 107–123.
[7] Y.Y. Yao, Three-way granular computing, rough sets, and formal concept analysis, Int. J. Approx. Reason. 116 (2020) 106–125.
[8] Q.H. Zhang, K. Xu, G.Y. Wang, Fuzzy equivalence relation and its multigranulation spaces, Inf. Sci. 346 (2016) 44–57.
[9] L.A. Zadeh, Fuzzy sets, Inf. Control 8 (3) (1965) 338–353.
[10] J.H. Li, C.L. Mei, W.H. Xu, et al., Concept learning via granular computing: a cognitive viewpoint, Inf. Sci. 298 (2015) 447–467.
[11] W.H. Xu, W.T. Li, Granular computing approach to two-way learning based on formal concept analysis in fuzzy datasets, IEEE Trans. Cybern. 46 (2) (2016) 366–379.
[12] Z. Pawlak, R. Sowinski, Rough set approach to multi-attribute decision analysis, Eur. J. Oper. Res. 72 (3) (1994) 443–459.
[13] H. Li, D.Y. Li, Y.H. Zhai, et al., A novel attribute reduction approach for multi-label data based on rough set theory, Inf. Sci. 367 (2016) 827–847.
[14] Y.Y. Yao, Y. Zhao, Attribute reduction in decision-theoretic rough set models, Inf. Sci. 178 (17) (2008) 3356–3373.
[15] J.C. Xu, C.S. Zhou, S.H. Xu, et al., Feature selection based on multi-perspective entropy of mixing uncertainty measure in variable-granularity rough set, Appl. Intell. 54 (1) (2024) 147–168.
[16] G.Y. Wang, X.A. Ma, H. Yu, Monotonic uncertainty measures for attribute reduction in probabilistic rough set model, Int. J. Approx. Reason. 59 (2015) 41–67.
[17] X.P. Zhang, J.J. Li, W.K. Li, A new mechanism of rule acquisition based on covering rough sets, Appl. Intell. 52 (11) (2022) 12369–12381.
[18] W.M. Ma, B.Z. Sun, Probabilistic rough set over two universes and rough entropy, Int. J. Approx. Reason. 53 (4) (2012) 608–619.
[19] Z. Pawlak, S.K.M. Wong, W. Ziarko, et al., Rough sets: probabilistic versus deterministic approach, Int. J. Man-Mach. Stud. 29 (1) (1988) 81–95.
[20] W. Ziarko, Variable precision rough set model, J. Comput. Syst. Sci. 46 (1) (1993) 39–59.
[21] D.D. Zou, Y.L. Xu, L.Q. Li, et al., A novel granular variable precision fuzzy rough set model and its application in fuzzy decision system, Soft Comput. 27 (13) (2023) 8897–8918.
[22] D.D. Zou, Y.L. Xu, L.Q. Li, et al., Novel variable precision fuzzy rough sets and three-way decision model with three strategies, Inf. Sci. 629 (2023) 222–248.
[23] W.Z. Wu, Y. Leung, Theory and applications of granular labelled partitions in multi-scale decision tables, Inf. Sci. 181 (18) (2011) 3878–3897.
[24] F. Li, B.Q. Hu, A new approach of optimal scale selection to multi-scale decision tables, Inf. Sci. 381 (2017) 193–208.
[25] W.Z. Wu, Y.Y. Qian, T.J. Li, et al., On rule acquisition in incomplete multi-scale decision tables, Inf. Sci. 378 (2017) 282–302.
[26] B. Huang, H.X. Li, G.F. Feng, et al., Double-quantitative rough sets, optimal scale selection and reduction in multi-scale dominance if decision tables, Int. J. Approx. Reason. 130 (2021) 170–191.
[27] W.L. Xi, W.W. Zhi, X.Z. Huang, et al., Optimal scale combination selection in generalized multi-scale hybrid decision systems, Inf. Sci. 689 (2025) 121429.
[28] Y.D. Huang, Y.J. Zhang, J.F. Xu, Incremental approaches for optimal scale selection in dynamic multi-scale set-valued decision tables, Int. J. Mach. Learn. Cybern. 14 (6) (2023) 2251–2270.
[29] W.Z. Wu, Y. Leung, Optimal scale selection for multi-scale decision tables, Int. J. Approx. Reason. 54 (8) (2013) 1107–1129.
[30] D.R. Niu, W.Z. Wu, T.J. Li, Variable precision based optimal scale combinations in generalized multi-scale decision systems, Int. J. Pattern Recognit. Artif. Intell. 32 (2019) 965–974.
[31] Z.T. Gong, W.T. Li, Multi-scale variable precision covering rough sets and its applications, Appl. Intell. 53 (24) (2023) 31018–31032.
[32] F. Li, B.Q. Hu, J. Wang, Stepwise optimal scale selection for multi-scale decision tables via attribute significance, Knowl.-Based Syst. 129 (2017) 4–16.
[33] W.Z. Wu, Y. Leung, A comparison study of optimal scale combination selection in generalized multi-scale decision tables, Int. J. Mach. Learn. Cybern. 11 (2020) 961–972.
[34] H. Bao, W.Z. Wu, J.W. Zheng, et al., Entropy based optimal scale combination selection for generalized multi-scale information tables, Int. J. Mach. Learn. Cybern. 12 (2021) 1427–1437.
[35] D.X. Chen, J.J. Li, R.D. Lin, et al., Information entropy and optimal scale combination in multi-scale covering decision systems, IEEE Access 8 (2020) 182908–182917.
[36] Z.H. Xie, W.Z. Wu, L.X. Wang, et al., Entropy based optimal scale selection and attribute reduction in multi-scale interval-set decision tables, Int. J. Mach. Learn. Cybern. 15 (2024) 3005–3026.
[37] Q. Wan, J.H. Li, L. Wei, et al., Optimal granule level selection: a granule description accuracy viewpoint, Int. J. Approx. Reason. 116 (2020) 85–105.
[38] X.Q. Zhang, Q.H. Zhang, Y.L. Cheng, et al., Optimal scale selection by integrating uncertainty and cost-sensitive learning in multi-scale decision tables, Int. J. Mach. Learn. Cybern. 11 (2020) 1095–1114.
[39] Q.H. Zhang, X.Q. Zhang, G.H. Pang, Cost-sensitive optimal scale combination in multi-scale decision systems, Control Decis. 36 (10) (2021) 2369–2378.
[40] J.M. Wu, D.Y. Liu, Z.H. Huang, et al., Multi-scale decision systems with test cost and applications to three-way multi-attribute decision-making, Appl. Intell. 54 (4) (2024) 3591–3605.
[41] S.M. Gu, W.W. Zhi, On knowledge acquisition in multi-scale decision systems, Int. J. Mach. Learn. Cybern. 4 (5) (2013) 477–486.
[42] C. Hao, J.H. Li, M. Fan, et al., Optimal scale selection in dynamic multi-scale decision tables based on sequential three-way decisions, Inf. Sci. 415 (2017) 213–232.
[43] Y.L. Cheng, Q.H. Zhang, G.Y. Wang, et al., Optimal scale selection and attribute reduction in multi-scale decision tables based on three-way decision, Inf. Sci. 541 (2020) 36–59.
[44] Y.L. Cheng, Q.H. Zhang, G.Y. Wang, Optimal scale combination selection for multi-scale decision tables based on three-way decision, Int. J. Mach. Learn. Cybern. 12 (2021) 281–301.
[45] X.Y. Zhang, Y.Y. Huang, Optimal scale selection and knowledge discovery in generalized multi-scale decision tables, Int. J. Approx. Reason. 161 (2023) 108983.
[46] Y. Yang, Q.H. Zhang, F. Zhao, et al., Optimal scale combination selection based on genetic algorithm in generalized multi-scale decision systems for classification, Inf. Sci. 693 (2025) 121685.
[47] Q.H. Zhang, C.Y. Long, Z. Fan, et al., Optimal scale combination selection integrating three-way decision with Hasse diagram, IEEE Trans. Neural Netw. Learn. Syst. 33 (8) (2022) 3675–3689.
[48] C. Luo, T.R. Li, Y.Y. Huang, et al., Updating three-way decisions in incomplete multi-scale information systems, Inf. Sci. 476 (2019) 274–289.
[49] Y.S. Chen, J.J. Li, J.X. Huang, Matrix method for the optimal scale selection of multi-scale information decision systems, Mathematics 7 (3) (2019) 290–307.
[50] W.K. Li, J.X. Huang, J.J. Li, et al., Matrix representation of optimal scale for generalized multi-scale decision table, J. Ambient Intell. Humaniz. Comput. 12 (2021) 8549–8559.
[51] B. Crawford, R. Soto, G. Astorga, et al., Putting continuous metaheuristics to work in binary search spaces, Complexity 2017 (2017) 8404231.