International Journal of Approximate Reasoning 161 (2023) 108983

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

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

Optimal scale selection and knowledge discovery in generalized multi-scale decision tables

Xiaoyan Zhang *, Yuyang Huang

College of Artificial Intelligence, Southwest University, 400715 Chongqing, People's Republic of China

A R T I C L E I N F O

Article history:
Received 22 March 2023
Received in revised form 6 July 2023
Accepted 7 July 2023
Available online 13 July 2023

Keywords:
Granular computing
Rough set
Multi-scale decision tables
Scale combinations

A B S T R A C T

In the real world, each attribute of data sets has multiple scales, whether it is a conditional attribute or a decision attribute. Knowledge mined from data based on the optimal scale can well meet the needs of real life. Hence, there is no denying that optimal scale selection is an problem to be solved urgently in the knowledge discovery field of multi-scale decision tables. The optimal scale combines coarser condition attributes with the finer decision attribute, so as to achieve a balance between efficiency and accuracy. With the aim of selecting the optimal scale combination, we firstly explore some related properties and theorems of generalized multi-scale decision tables. Then we define the optimal scale in generalized multi-scale decision tables and propose two algorithms for optimal scale selection. In addition, a knowledge acquisition algorithm and a multi-scale rough set classifier framework are proposed. Finally, numerical experiments are performed on some open data sets to test the effectiveness of the algorithms.

1. Introduction

Granular computing is a novel approach that simulates the human mind to solve some complex problems. Its basic idea comes from fuzzy information granulation proposed by Zadeh [42,43], and it was Lin [15] who explicitly proposed the expression "granular computing." Lin [16] and Yao [39,40] were the first to discuss in detail the basic problems, research methods, and potential applications of granular computing. Although granular computing is a new research field, its basic ideas and methods are embodied in many disciplines. For example, Pawlak [18] came up with rough set theory in 1982, Wille [28] proposed formal concept analysis in 1982, and Zhang et al. proposed quotient space theory in 1990. These theories have played a positive role in promoting the continuous progress of granular computing. In addition, the application potential of granular computing has also received attention from scholars, and some challenging scientific problems have evolved in big data processing.

Many granular computing models and methods have been proposed in specific application background, such as fuzzy set theory [4,33], three-way decisions [5,6,8], and concept-cognitive learning [13,34,35]. A rough set, use of which is a very effective method to handle complex information systems, has a significant role in the progress of granular computing research. Pawlak's classical rough set model describes knowledge particles with equivalence classes given by equivalence relations, and deduces decision or classification rules by feature selection while the positive region remains unchanged. Rough set theory is widely used in the field of intelligent information processing, involving an expert system, image processing, pattern

* Corresponding author.
E-mail addresses: zxy19790915@163.com (X. Zhang), 1483912367@qq.com (Y. Huang).

https://doi.org/10.1016/j.ijar.2023.108983

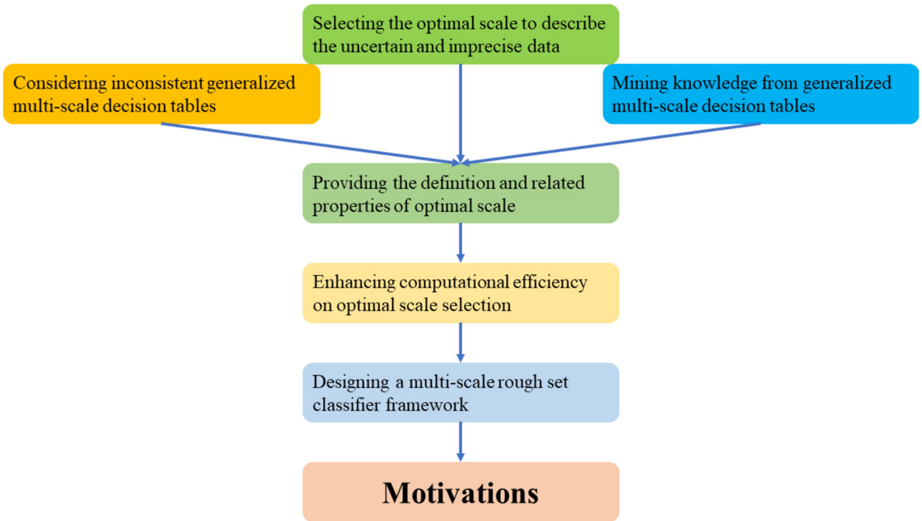

Fig. 1. Motivation for our work.

recognition, decision analysis, and so on. In the field of data mining, which includes many technologies, the application of rough set theory mainly centers on the following aspects: feature selection [14,38], knowledge discovery [7,32], incremental algorithms [36,37,41], and information fusion [44]. Another hot research topic of granular computing is multi-granularity computing, which mainly considers how to divide the domain (research object) multiple times through multiple relations, so that problems (such as rule extraction and optimal granularity selection) can be solved from multiple granularity levels. In rough set theory, much research has been done on multi-granularity problems. According to different starting points, it can be approximately divided into two classifications: multi-granulation rough sets [19–23] and multi-scale rough sets [25–27,29].

Since Wu and Leung [31] came up with multi-scale rough sets, scholars have conducted much research, which mainly focuses on the optimal scale selection and rule extraction. Wu and Leung [30] considered optimal granularity selection and rule extraction in coordinated, uncoordinated, and incomplete multi-granularity labeling decision systems, and compared the similarities and differences among the results of optimal scale selection. Gu and Wu [7] discussed local scale selection and rule extraction of multi-scale decision systems. In the multi-scale decision system, She et al. [24] further considered the problem of local rule extraction and performed simulation experiments to demonstrate the effectiveness of the local strategy. In addition, considering that the multi-scale decision system proposed by Wu and Leung requires the same number of scales for each conditional attribute, it is too harsh, so Li and Hu [11] relaxed the restriction of a multi-scale conditional attribute (i.e., the number of scales for each conditional attribute can be different in pairs). On this basis, Huang et al. [9] further relaxed the restriction of decision attributes and proposed a generalized multi-scale information system with multi-scale decision attributes. The scales of each attribute can be randomly combined to produce a single-scale decision system in the generalized multi-scale information system. In optimal scale selection, the best single-scale decision system is selected from multiple single-scale decision systems to get the best result for the problem under consideration. Scale combinations increase explosively as the conditional attributes and the number of scales per conditional attribute increase, making calculation very time-consuming. For increased calculation efficiency, Li et al. [12] offered a new optimal scale selection algorithm based on the importance of multi-scale attributes and the method of stepwise regression; that is, heuristic thinking was adopted to solve the optimal scale combination. Considering that multi-scale decision systems are often in a dynamic update environment, Hao et al. [10] came up with an optimal scale updating method for multi-scale decision systems when objects are updated, which solved the partial and molecular problem of optimal scale selection for dynamic multi-scale decision systems. In addition, multi-scale rough sets have also been studied in conjunction with some other theories, such as entropy [1], three-way decision theory [17], and evidence theory [45]. However, the optimal scale selection of inconsistent generalized multi-scale information tables with multi-scale decision attributes is not considered in the present study. Meanwhile, the optimal scale selection algorithm for consistent generalized multi-scale information tables with multi-scale decision attributes is also not efficient. Therefore, it is necessary to design efficient optimal scale selection algorithms for generalized multi-scale information tables. More importantly, rules are extracted from the decision table on the basis of optimal scales and are applied to the frame of the multi-scale rough set classifier. As Fig. 1 shows, we have performed an intensive study in this field.

This article has three major contributions. Firstly, some related properties and theorems of generalized multi-scale information tables with multi-scale decision attributes are presented. Secondly, the definition and properties of optimal scale are given, and two optimal scale selection algorithms for generalized multi-scale decision tables (GMSDTs) are proposed.

Thirdly, a knowledge acquisition algorithm that is able to obtain decision rules effectively and a multi-scale rough set classifier framework are proposed.

The rest of this article is organized as follows: We succinctly review the concepts of rough sets and multi-scale information tables in Section 2. We introduce generalized multi-scale information tables with multi-scale decision attributes and discuss some related properties and theorems in Section 3. In Section 4 we define the optimal scale of the GMSDTs and give two algorithms for optimal scale selection. On this basis, a knowledge acquisition algorithm and a multi-scale rough set classifier framework are proposed in Section 5. In Section 6, some related experimental tests, results, and conclusions are provided.

2. Preliminaries

We succinctly review several basic concepts related to rough sets and multi-scale information tables in this section. Furthermore, some related properties and theorems of generalized multi-scale information decision tables with multi-scale decision attributes are explored for the purpose of better describing the research in this article.

Let U be a nonempty finite set referred to as the universe of discourse, and let R be an equivalence relationship on U. Then, for any $X \subseteq U$, the lower and upper approximations of X are defined as

$$\underline{R}(X) = \bigcup \left\{ [X]_R \mid [X]_R \subseteq X \right\},$$
$$\overline{R}(X) = \bigcup \left\{ [X]_R \mid [X]_R \cap X \neq \varnothing \right\},$$

where $[x]_R = \{y \in U \mid (x, y) \in R\}$ is an equivalence class containing x.

A multi-scale information table is a tuple $S = (U, AT)$, where $U = (x_1, x_2, \ldots, x_n)$ is a universe of discourse and $AT = (a_1, a_2, \ldots, a_n)$ is a nonempty and finite set of attributes, where each attribute has I scales. Then a multi-scale information table can be represented as $\left(U, \left\{a_j^k \mid k = 1, 2, \ldots, I;\ j = 1, 2, \ldots, m\right\}\right)$.

Let $S = (U, AT \cup D) = \left(U, \left\{a_j^k \mid k = 1, 2, \ldots, I;\ j = 1, 2, \ldots, m\right\} \cup \{d\}\right)$ be a multi-scale decision table (MSDT). Attribute a_j is restricted on its l_jth scale $(j = 1, 2, \ldots, m)$. The index set $(l_1, l_2, \ldots, l_m)$ is called a scale combination of condition attributes, and the family of all scale combinations is denoted as $\mathscr{L} = \left\{(l_1, l_2, \ldots, l_m) \mid l_j \in \{1, 2, \ldots, I\},\quad j = 1, \ldots, m\right\}$.

Assume that $\mathscr{L}$ is the family of all scale combinations of condition attributes, $L_1 = \left\{l_1^1, l_2^1, \ldots, l_m^1\right\}$, $L_2 = \left\{l_1^2, l_2^2, \ldots, l_m^2\right\} \in \mathscr{L}$. If $l_j^1 \leqslant l_j^2$, then L_2 is said to be coarser than L_1, and this is denoted as $L_1 \preceq L_2$. Moreover, if $L_1 \preceq L_2$ and there exists $k \in \{1, 2, \ldots, m\}$ such that $l_k^1 < l_k^2$, then L_2 is said to be strictly coarser than L_1, and this is denoted as $L_1 \prec L_2$.

A GMSDT is a triple $S = (U, AT \cup D)$, where $(U, AT) = \left(U, \left\{a_j^k \mid k = 1, 2, \ldots, I_j;\ j = 1, 2, \ldots, m\right\}\right)$ is a multi-scale information table, $D = \{d\}$ is a nonempty and finite set of decision attributes, and d has n scales $\left\{d^t \mid t = 1, 2, \ldots, n\right\}$. Then a GMSDT can be represented as a table $S = (U, AT \cup D) = \left(U, \left\{a_j^k \mid k = 1, 2, \ldots, I_j;\ j = 1, 2, \ldots, m\right\} \cup \left\{d^t \mid t = 1, 2, \ldots, n\right\}\right)$.

Let $S = (U, AT \cup D) = \left(U, \left\{a_j^k \mid k = 1, 2, \ldots, I_j;\ j = 1, 2, \ldots, m\right\} \bigcup \left\{d^t \mid t = 1, 2, \ldots, n\right\}\right)$ be a GMSDT, where $a_j^k : U \to V_j^k$, where V_j^k is the domain of the kth scale of attribute a_j. For $j = 1, 2, \ldots, m$, $1 \leqslant k \leqslant k - 1$, there exists a surjective mapping $g_j^{k,k+1} : V_j^k \to V_j^{k+1}$ such that $a_j^{k+1} : g_j^{k,k+1} \circ a_j^k$, i.e., $a_j^{k+1}(x) = g_j^{k,k+1}\left(a_j^k(x)\right)$, $x \in U$, where $g_j^{k,k+1}$ is called a granular transformation mapping of condition attribute $d^t : U \to V_d^t$, where V_d^t is the domain of the tth scale of decision attribute d. For $t = 1, 2, \ldots, n$, there exists a surjective mapping $h^{t,t+1} : V_d^t \to V_d^{t+1}$ such that $d^{t+1} = h^{t,t+1} \circ d^t$, i.e., $d^{t+1} = h^{t,t+1}\left(d^t(x)\right)$, $x \in U$, where $h^{t,t+1}$ is called a scale transformation mapping of decision attributes.

Example 1. As shown in Table 1, this is an instance of GMSDTs, where $U = \{x_1, x_2, \ldots, x_5\}$ represents five students. a_1 represents the scores of courses, and has two scales. Different scales indicate the scores recorded in different scoring systems. a_2 represents the evaluation of moral level, and has two scales. d has three scales, and represents the comprehensive evaluation of a student. In Table 1, $V_1^1 = \{95, 90, 85, 80, 75, 70\}$, $V_1^2 = \{A, B, C\}$, $g_1^{1,2}(95, 90) = A$, $g_1^{1,2}(85, 80) = B$, and $g_1^{1,2}(75, 70) = C$; $V_2^1 = \{L, M, N\}$, $V_2^2 = \{P, Q\}$, $g_2^{1,2}(L) = Q$, and $g_2^{1,2}(M, N) = P$; $V_d^1 = \{1, 2, 3, 4, 5\}$, $V_d^2 = \{G, X, W\}$, $V_d^3 = \{1, 0\}$, $h^{1,2}(1, 2) = G$, $h^{1,2}(3, 4) = X$, $h^{1,2}(5) = W$, $h^{2,3}(G, X) = 1$, and $h^{2,3}(W) = 0$.

Let $S = (U, AT \cup D) = \left(U, \left\{a_j^k \mid k = 1, 2, \ldots, I_j;\ j = 1, 2, \ldots, m\right\} \cup \left\{d^t \mid t = 1, 2, \ldots, n\right\}\right)$ be a GMSDT. $Q = (K, t) = (\{k_1, k_2, \ldots, k_m\}, t)$ is called a scale selection of S if the condition attribute a_j is restricted on its k_jth scale $(1 \le j \le m)$ and the decision attribute d is restricted on the tth scale $(1 \le t \le n)$. Each scale selection $Q = (K, t) = (\{k_1, k_2, \ldots, k_m\}, t)$ forms a single-scale decision table $S^Q = \left(U, AT^K \cup d^t\right)$, where $AT^K = \left\{a_1^{k_1}, a_2^{k_2}, \ldots, a_m^{k_m}\right\}$.

Property 1. Assuming $S = (U, AT) = \left(U, \left\{a_j^k \mid k = 1, 2, \ldots, I_j;\ j = 1, 2, \ldots, m\right\}\right)$ is a generalized multi-scale information table and $L_1 = \left\{l_1^1, l_2^1, \ldots, l_m^1\right\}$, $L_2 = \left\{l_1^2, l_2^2, \ldots, l_m^2\right\} \in L$, the following properties are true for any $X \subseteq U$:

Table 1
Scale transformation process.

U/C	a_1			a_2			d				
	a_1^1	$\xrightarrow{g_1^{12}}$	a_1^2	a_2^1	$\xrightarrow{g_2^{12}}$	a_2^2	d^1	$\xrightarrow{h^{12}}$	d^2	$\xrightarrow{h^{23}}$	d^3
x_1	95		A	L		Q	1		G		1
x_2	90		A	M		P	1		G		1
x_3	85		B	N		P	2		G		1
x_4	80		B	L		Q	3		X		1
x_5	75		C	M		P	4		X		1
x_6	70		C	N		P	5		W		0

(1) $L_1 \preceq L_2 \Rightarrow R_{AT^{L_1}} \subseteq R_{AT^{L_2}}$.
(2) $L_1 \preceq L_2 \Rightarrow U/R_{AT^{L_1}} \subseteq U/R_{AT^{L_2}}$.
(3) $L_1 \preceq L_2 \Rightarrow \underline{R}_{AT^{L_1}}(X) \supseteq \underline{R}_{AT^{L_2}}(X)$.
(4) $L_1 \preceq L_2 \Rightarrow \overline{R}_{AT^{L_1}}(X) \subseteq \overline{R}_{AT^{L_2}}(X)$.
(5) $L_1 \preceq L_2 \Rightarrow BND\left(R_{AT^{L_1}}, X\right) \subseteq BND\left(R_{AT^{L_2}}, X\right)$.
(6) If $L_0 = (1, 1, \ldots, 1)$ and $L' = (I_1, I_2, \ldots, I_n)$, for any $L \in \mathscr{L}$ we have $BND\left(R_{AT^{L_0}}, X\right) \subseteq BND\left(R_{AT^{L}}, X\right) \subseteq BND\left(R_{AT^{L'}}, X\right)$.

In Property 1, properties (1) and (2) reveal that the finer the scale is, the more precise the divided universes are, thus increasing the equivalence relation and equivalence class. Properties (3) and (4) make clear that the finer the scale is, the closer the lower approximation and upper approximation of X are to X. Property (5) shows that boundary regions of X increase while the scale becomes coarser. As can be seen from property (6), $BND\left(R_{AT^{L}}, X\right)$ can achieve its minimum value and its maximum value in the case of $L = (1, 1, \ldots, 1)$ and $L = (I_1, I_2, \ldots, I_n)$.

Let $D = \{d\}$ be a decision attribute set. Then the positive and boundary regions of $D = \{d\}$ relative to R_{AT^L} are defined as

$$POS\left(R_{AT^L}, D\right) = \bigcup_{D_i \in U/R_D} \underline{R}_{AT^L}(D_i),$$

$$BND\left(R_{AT^L}, D\right) = \bigcup_{D_i \in U/R_D} \left(\overline{R}_{AT^L}(D_i) - \underline{R}_{AT^L}(D_i)\right).$$

Property 2. Let $S = (U, AT \cup D) = \left(U, \left\{a_j^k \mid k = 1, 2, \ldots, I_j;\ j = 1, 2, \ldots, m\right\} \cup \left\{d^t \mid t = 1, 2, \ldots, n\right\}\right)$ be a GMSDT and $L_1, L_2 \in \mathscr{L}$. Then we have the following properties:
(1) $L_1 \preceq L_2 \Rightarrow BND\left(R_{AT^{L_1}}, D_t\right) \subseteq BND\left(R_{AT^{L_2}}, D_t\right)$, $t = 1, 2, \ldots, n$.
(2) $t_1 < t_2 \Rightarrow BND\left(R_{AT^{L}}, D_{t_1}\right) \supseteq BND\left(R_{AT^{L}}, D_{t_2}\right)$.
(3) If $L_0 = (1, 1, \ldots, 1)$, then $BND\left(R_{AT^{L_0}}, D_n\right) \subseteq BND\left(R_{AT^{L}}, D_t\right)$, $L \in \mathscr{L}$, $t = 1, 2, \ldots, n$.
(4) If $L' = (I_1, I_2, \ldots, I_n)$, then $BND\left(R_{AT^{L'}}, D_1\right) \supseteq BND\left(R_{AT^{L}}, D_t\right)$, $L \in \mathscr{L}$, $t = 1, 2, \ldots, n$.

In Property 2, property (1) reveals that the finer the condition attributes are, the smaller the corresponding boundary regions are. Property (2) makes clear that the finer the decision attributes are, the larger the corresponding regions are. Moreover, property (3) indicates that the corresponding boundary regions achieve their maximum value when the condition attributes are finest and the decision attributes are coarsest. Property (4) shows that the corresponding boundary regions achieve their maximum value when the condition attributes are coarsest and the decision attributes are finest.

3. Optimal scale selection for GMSDTs

People tend to draw strong conclusions from weak conditions for many mathematical models. If conditions are too weak and conclusions are too strong, conclusions may not be reliable. It is important to strike a balance between efficiency and precision in terms of conditions and conclusions, and this problem is actually the optimal scale selection in the GMSDT. We first define the optimal scale of the GMSDT and discuss its related properties in this section. Two optimal scale selection algorithms are then provided. One can quickly find an optimal scale, and the other can find all the optimal scales.

Let $S = (U, AT \cup D) = \left(U, \left\{a_j^k \mid k = 1, 2, \ldots, I_j;\ j = 1, 2, \ldots, m\right\} \cup \left\{d^t \mid t = 1, 2, \ldots, n\right\}\right)$ be a GMSDT, and let $L_0 = (1, 1, \ldots, 1)$ be the finest scale of the condition attribute. If $BND\left(AT^{L_0}, D_n\right) = \varnothing$, i.e., S is consistent relative to the decision attribute d^n with the coarsest scale, then S is considered to be consistent; otherwise, S is inconsistent.

Definition 1. Let $S=(U,AT\cup D)=\left(U,\left\{a_j^k\mid k=1,2,\ldots,I_j;\,j=1,2,\ldots,m\right\}\cup\left\{d^t\mid t=1,2,\ldots,n\right\}\right)$ be a consistent GMSDT, $L=(l_1,l_2,\ldots,l_m)$, and $t\in\{1,2,\ldots,n\}$. If the following conditions are satisfied, then $Q=(L,t)=(\{l_1,l_2,\ldots,l_m\},t)$ is called the global optimal scale of S, and $AT^L\bigcup\left\{d^t\right\}=\left(a_1^{l_1},a_2^{l_2},\ldots,a_m^{l_m},d^t\right)$ is the global optimal attribute set of S:
(1) $BND\left(AT^L,D_t\right)=\varnothing$.
(2) For any $H=(h_1,h_2,\ldots,h_m)\in\mathscr{L}$, and $t'\in\{1,2,\ldots,n\}$, if $L\preceq H$ and $t'\le t$ (the equal sign can only have one at most), we have $BND\left(AT^H,D_{t'}\right)\neq\varnothing$.

Definition 2. Let $S=(U,AT\cup D)=\left(U,\left\{a_j^k\mid k=1,2,\ldots,I_j;\,j=1,2,\ldots,m\right\}\cup\left\{d^t\mid t=1,2,\ldots,n\right\}\right)$ be an inconsistent GMSDT, $L_0=(1,1,\ldots,1)$, $L=(l_1,l_2,\ldots,l_m)$, and $t\in\{1,2,\ldots,n\}$. If the following conditions are satisfied, then $Q=(L,t)=(\{l_1,l_2,\ldots,l_m\},t)$ is called the global optimal scale of S, and $AT^L\bigcup\left\{d^t\right\}=\left(a_1^{l_1},a_2^{l_2},\ldots,a_m^{l_m},d^t\right)$ is the global optimal attribute set of S:
(1) $BND\left(AT^L,D_t\right)=BND\left(AT^{L_0},D_n\right)$.
(2) For any $H=(h_1,h_2,\ldots,h_m)\in\mathscr{L}$, and $t'\in\{1,2,\ldots,n\}$, if $L\preceq H$ and $t'\le t$ (the equal sign can only have one at most), we have $BND\left(AT^{L_0},D_n\right)\subset BND\left(AT^H,D_{t'}\right)$.

By Definitions 1 and 2, it can be seen that the so-called global optimal scale selection is to find scale combinations in which $\sum_{j=1}^m l_j-t$ of the scale $Q=(L,t)=(\{l_1,l_2,\ldots,l_m\},t)$ is as large as possible under the condition that $BND\left(AT^L,D_t\right)=BND\left(AT^{L_0},D_n\right)$.

Definition 3. Let $S=(U,AT\cup D)=\left(U,\left\{a_j^k\mid k=1,2,\ldots,I_j;\,j=1,2,\ldots,m\right\}\cup\left\{d^t\mid t=1,2,\ldots,n\right\}\right)$ be a GMSDT, $L_0=(1,1,\ldots,1)$, $L=(l_1,l_2,\ldots,l_m)$, and $t\in\{1,2,\ldots,n\}$. If the following conditions are satisfied, then $Q=(L,t)=(\{l_1,l_2,\ldots,l_m\},t)$ is called the local optimal scale of S relative to the decision attribute d^t, and $AT^L\bigcup\left\{d^t\right\}=\left(a_1^{l_1},a_2^{l_2},\ldots,a_m^{l_m},d^t\right)$ is the local optimal attribute set of S:
(1) $BND\left(AT^L,D_t\right)=BND\left(AT^{L_0},D_t\right)$.
(2) For any $H=(h_1,h_2,\ldots,h_m)\in\mathscr{L}$, if $L\prec H$, we have $BND\left(AT^{L_0},D_t\right)\subset BND\left(AT^H,D_t\right)$.

In real life, sometimes we care only about the information hidden in the MSDT under a certain scale of decision attributes. Therefore, the definition of local optimal scale is given. Obviously, when a GMSDT degenerates to an MSDT, the local optimal scale is the global optimal scale.

Theorem 1. *Let $S=(U,AT\cup\{d\})=\left(U,\left\{a_j^k\mid k=1,2,\ldots,I_j;\,j=1,2,\ldots,m\right\}\cup\left\{d^t\mid t=1,2,\ldots,n\right\}\right)$ be a GMSDT, $L_0=(1,1,\ldots,1)$, $L=(l_1,l_2,\ldots,l_m)$, and $t\in\{1,2,\ldots,n\}$. If $BND\left(AT^L,D_t\right)=BND\left(AT^{L_0},D_n\right)$, then for any $H=(h_1,h_2,\ldots,h_m)\in\mathscr{L}$, and $t'\in\{1,2,\ldots,n\}$, if $H\preceq L$ and $t'\ge t$ (the equal sign can only have one at most), $Q=\left(H,t'\right)=\left(\{h_1,h_2,\ldots,h_m\},t'\right)$ is not the global optimal scale of S and $C^H\bigcup\left\{d^{t'}\right\}=\left(a_1^{l_1},a_2^{l_2},\ldots,a_m^{l_m},d^{t'}\right)$ is not the global optimal attribute set of S.*

Proof. Assume that $Q=\left(H,t'\right)=\left(\{h_1,h_2,\ldots,h_m\},t'\right)$ is the global optimal scale of S.

However, there exists $H\preceq L$ and $t'\ge t$ (the equal sign can only have one at most) such that $BND\left(AT^L,D_t\right)=BND\left(AT^{K_0},D_n\right)$, which clearly contradicts the assumption that $Q=\left(H,t'\right)=\left(\{h_1,h_2,\ldots,h_m\},t'\right)$ is the global optimal scale of S. Therefore, $Q=\left(H,t'\right)=\left(\{h_1,h_2,\ldots,h_m\},t'\right)$ is not the global optimal scale of S and $AT^H\bigcup\left\{d^{t'}\right\}=\left(a_1^{l_1},a_2^{l_2},\ldots,a_m^{l_m},d^{t'}\right)$ is not the global optimal attribute set of S.

Theorem 2. *Let $S=(U,AT\cup\{d\})=\left(U,\left\{a_j^k\mid k=1,2,\ldots,I_j;\,j=1,2,\ldots,m\right\}\cup\left\{d^t\mid t=1,2,\ldots,n\right\}\right)$ be a GMSDT, $L_0=(1,1,\ldots,1)$, $L=(l_1,l_2,\ldots,l_m)$, and $t\in\{1,2,\ldots,n\}$. If $BND\left(AT^{L_0},D_n\right)\subset BND\left(AT^L,D_t\right)$, then for any $H=(h_1,h_2,\ldots,h_m)\in\mathscr{L}$, and $t'\in\{1,2,\ldots,n\}$, if $L\preceq H$ and $t\ge t'$ (the equal sign can only have one at most), $Q=\left(H,t'\right)=\left(\{h_1,h_2,\ldots,h_m\},t'\right)$ is not the global optimal scale of S and $AT^H\bigcup\left\{d^t\right\}=\left(a_1^{l_1},a_2^{l_2},\ldots,a_m^{l_m},d^{t'}\right)$ is not the global optimal attribute set of S.*

Proof. Assume that $Q=\left(H,t'\right)=\left(\{h_1,h_2,\ldots,h_m\},t'\right)$ is the global optimal scale of S.

We obtain $BND\left(AT^H,D_{t'}\right)=BND\left(AT^{L_0},D_n\right)$ by the definition of the global optimal scale. Because $BND\left(AT^{L_0},D_n\right)\subset BND\left(AT^L,D_t\right)$ and $L\preceq H$ and $t\ge t'$ (the equal sign can only have one at most), we know that $BND\left(AT^{L_0},D_n\right)\subset BND\left(AT^L,D_t\right)\subseteq BND\left(AT^H,D_{t'}\right)$ by Property 2. $BND\left(AT^H,D_{t'}\right)=BND\left(AT^{L_0},D_n\right)$ obviously contradicts $BND\left(AT^{L_0},\right.$

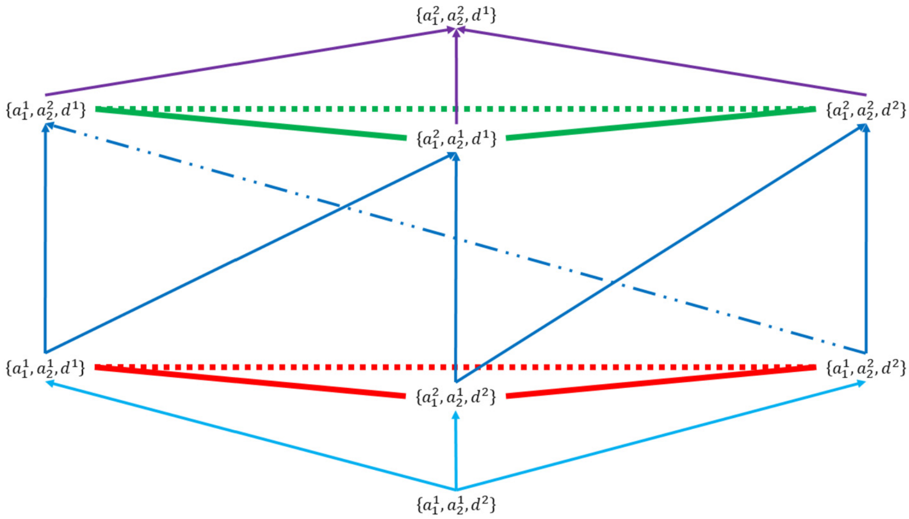

Fig. 2. The lattice structure of scale combination in Example 2.

$D_n) \subset BND\left(AT^H, D_{t'}\right)$. Therefore, $Q = \left(H, t'\right) = \left(\{h_1, h_2, \ldots, h_m\}, t'\right)$ is not the global optimal scale of S and $AT^H \bigcup \left\{d^{t'}\right\} = \left(a_1^{l_1}, a_2^{l_2}, \ldots, a_m^{l_m}, d^{t'}\right)$ is not the global optimal attribute set of S.

Theorem 3. *Let $S = (U, AT \cup \{d\}) = \left(U, \left\{a_j^k \mid k = 1, 2, \ldots, I_j;\ j = 1, 2, \ldots, m\right\} \cup \left\{d^t \mid t = 1, 2, \ldots, n\right\}\right)$ be a GMSDT, $L_0 = (1, 1, \ldots, 1)$, $L = (l_1, l_2, \ldots, l_m)$, and $t \in \{1, 2, \ldots, n\}$. If $Q = (H, t) = (\{h_1, h_2, \ldots, h_m\}, t)$ is not the local optimal scale of S and $AT^H \bigcup \left\{d^t\right\} = \left(a_1^{l_1}, a_2^{l_2}, \ldots, a_m^{l_m}, d^t\right)$ is not the local optimal attribute set of S, then $Q = (H, t) = (\{h_1, h_2, \ldots, h_m\}, t)$ is not the global optimal scale of S and $AT^H \bigcup \left\{d^t\right\} = \left(a_1^{l_1}, a_2^{l_2}, \ldots, a_m^{l_m}, d^t\right)$ is not the global optimal attribute set of S.*

Proof. Assume that $Q = (H, t) = (\{h_1, h_2, \ldots, h_m\}, t)$ is the global optimal scale of S.

By the definition of the global optimal scale, we know that:

(1) $BND\left(AT^H, D_t\right) = BND\left(AT^{L_0}, D_n\right)$.

(2) For any $K = (k_1, k_2, \ldots, k_m) \in \mathscr{L}$, and $t' \in \{1, 2, \ldots, n\}$, if $H \preceq K$ and $t' \leq t$ (the equal sign can only have one at most), we have $BND\left(AT^{L_0}, D_n\right) \subset BND\left(AT^K, D_{t'}\right)$.

Because $BND\left(AT^{L_0}, D_n\right) \subseteq BND\left(AT^{L_0}, D_t\right) \subseteq BND\left(AT^H, D_t\right)$ by Property 2, we know that $BND\left(AT^{L_0}, D_n\right) = BND\left(AT^{L_0}, D_t\right) = BND\left(AT^H, D_t\right)$.

Let $t' = t$. For any $K = (k_1, k_2, \ldots, k_m) \in \mathscr{L}$, if $H \prec K$, we have $BND\left(AT^{L_0}, D_n\right) = BND\left(AT^{L_0}, D_t\right) = BND\left(AT^H, D_t\right) \subset BND\left(AT^K, D_t\right)$. This clearly contradicts the fact that $Q = (H, t) = (\{h_1, h_2, \ldots, h_m\}, t)$ is not the local optimal scale of S and $AT^H \bigcup \left\{d^t\right\} = \left(a_1^{l_1}, a_2^{l_2}, \ldots, a_m^{l_m}, d^t\right)$ is not the local optimal attribute set of S. Therefore, $Q = (H, t) = (\{h_1, h_2, \ldots, h_m\}, t)$ is not the global optimal scale of S and $AT^H \bigcup \left\{d^t\right\} = \left(a_1^{l_1}, a_2^{l_2}, \ldots, a_m^{l_m}, d^t\right)$ is not the global optimal attribute set of S.

Example 2 *(Continued from Example 1).* Assume that $S = (U, AT \cup \{d\}) = \left(U, \left\{a_1^1, a_1^2, a_2^1, a_2^2\right\} \cup \left\{d^1, d^2\right\}\right)$ is a GMSDT, condition attributes a_1 and a_2 have two scales, and decision attributes have two scales; hence, S consists of eight scale combinations that form a lattice structure, as shown in Fig. 2.

As shown in Fig. 2, $\left\{a_1^2, a_2^2, d^1\right\}$ at the top has maximum boundary regions and $\left\{a_1^1, a_2^1, d^2\right\}$ at the bottom has minimum boundary regions. In addition, if $BND\left(\left\{a_1^2, a_2^2\right\}, d_1\right) = BND\left(\left\{a_1^1, a_2^1\right\}, d_2\right)$, then $\left\{a_1^2, a_2^2, d^1\right\}$ is the global optimal scale. From bottom to top in Fig. 2, the scale gets coarser and coarser, and the boundary regions get larger and larger. In any case, $\left\{a_1^1, a_2^1, d^2\right\}$ at the bottom always satisfies $BND\left(\left\{a_1^1, a_2^1\right\}, d_2\right) = BND\left(\left\{a_1^1, a_2^1\right\}, d_2\right)$, which ensures the presence of the global optimal scale.

There is no difference between local optimal scale selection and global optimal scale selection, so we propose two global optimal scale selection algorithms. Algorithm 1 can quickly find a global optimal scale, and Algorithm 2 can find all the global optimal scales.

Algorithm 1 takes $AT^{L_0} \bigcup \left\{d^n\right\}$ as the starting point to search upward for a scale combination that meets the definition of the global optimal scale. $Queue$ is a set of scale combinations, and is used to determine whether OS satisfies the definition

Table 2
A GMSDT.

U/C	a_1		a_2		a_3		d	
	a_1^1	a_1^2	a_2^1	a_2^2	a_3^1	a_3^2	d^1	d^2
x_1	1	S	1	Y	1	G	1	+
x_2	1	S	1	Y	1	G	1	+
x_3	1	S	2	Y	1	G	2	+
x_4	2	S	1	Y	2	G	2	+
x_5	3	L	3	N	3	F	1	+
x_6	3	L	3	N	3	F	3	−
x_7	4	L	4	N	4	F	2	+
x_8	4	L	4	N	4	F	4	−
x_9	3	L	4	N	5	B	3	−
x_{10}	4	L	3	N	5	B	3	−
x_{11}	3	L	3	N	6	B	4	−
x_{12}	3	L	3	N	6	B	4	−

Algorithm 1: Global optimal scale selection of GMSDTs.

Input: A GMSDT $S=(U, AT\cup\{d\})$.
Output: A global optimal scale OS.
calculate $BND\left(R_{AT^{L_0}}, D_n\right)$ and $BND\left(R_{AT^{L'}}, D_1\right)$;
if $BND\left(R_{AT^{L_0}}, D_n\right) \neq BND\left(R_{AT^{L'}}, D_1\right)$ **then**
  $OS = AT^{L_0} \bigcup \{d^n\}$
  $Queue = InitQueue(OS)$
  while $Queue \neq \varnothing$ **do**
    if $BND\left(R_{AT^H}, D_t\right) = BND\left(R_{AT^{L_0}}, D_n\right)$ **then**
      $OS = AT^H \bigcup \{d^t\}$
      $Queue = InitQueue(OS)$
    else
      $Queue = Queue - AT^H \bigcup \{d^t\}$
    end if
  end while
else
  $OS = AT^{L'} \bigcup \{d^1\}$
end if
return OS;

of the global optimal scale. Algorithm 1 can quickly find a global optimal scale that satisfies the definition. The worst-case time complexity of Algorithm 1 is $O\left(\left(\sum_{j=1}^{m} I_j + n\right) \times |U|^2\right)$, where $|U|$ is the number of objects.

Algorithm 2 needs to traverse all scale combinations to find all global optimal scales. Nevertheless, Theorems 1 and 2 allow us to reduce some unnecessary calculations when judging whether some scale combinations meet the definition of global optimal scales. OSS is a set of scale combinations whose $BND\left(AT^L, D_t\right) = BND\left(AT^{L_0}, D_n\right)$. T is a set of scale combinations $BND\left(AT^{L_0}, D_n\right) \subset BND\left(AT^L, D_t\right)$. Algorithm 2 can find all the global optimal scales in the lattice structure of scale combinations. The worst-case time complexity of Algorithm 2 is $O\left(\left(\prod_{j=1}^{m} I_j \times n\right) \times |U|^2\right)$, where $|U|$ is the number of objects.

Example 3. Let $S=(U, AT\cup\{d\}) = \left(U, \left\{a_1^1, a_1^2, a_2^1, a_2^2, a_3^1, a_3^2\right\} \cup \left\{d^1, d^2\right\}\right)$ be a GMSDT, as given in Table 2, where $U = \{x_1, x_2, \ldots, x_{12}\}$ represents 12 bank customers. a_1, a_2, and a_3 stand for consumption, savings, and credit, respectively. Each attribute has two scales, where S, L, Y, N, G, F, and B, respectively, represent small, large, yes, no, good, fair, and bad. d represents different levels of bank customers.

We now find a global optimal scale for Example 3 on the basis of Algorithm 1.

(1) It can be calculated that $BND\left(\left\{a_1^1, a_2^1, a_3^1\right\}, d_2\right) = \{x_5, x_6, x_7, x_8\}$; therefore, S is inconsistent.

(2) Let $OS = \left\{a_1^1, a_2^1, a_3^1, d^2\right\}$. Then $Queue = \left\{\left\{a_1^2, a_2^1, a_3^1, d^2\right\}, \left\{a_1^1, a_2^2, a_3^1, d^2\right\}, \left\{a_1^1, a_2^1, a_3^2, d^2\right\}, \left\{a_1^1, a_2^1, a_3^1, d^1\right\}\right\}$.

(3) It is easy to obtain $BND\left(\left\{a_1^2, a_2^1, a_3^1\right\}, d_2\right) = \{x_5, x_6, x_7, x_8\}$. Hence, if we let $OS = \left\{a_1^2, a_2^1, a_3^1, d^2\right\}$, then $Queue = \left\{\left\{a_1^2, a_2^2, a_3^1, d^2\right\}, \left\{a_1^2, a_2^1, a_3^2, d^2\right\}, \left\{a_1^2, a_2^1, a_3^1, d^1\right\}\right\}$.

(4) It is easy to obtain $BND\left(\left\{a_1^2, a_2^2, a_3^1\right\}, d_2\right) = \{x_5, x_6, x_7, x_8\}$. Hence, if we let $OS = \left\{a_1^2, a_2^2, a_3^1, d^2\right\}$, then $Queue = \left\{\left\{a_1^2, a_2^2, a_3^2, d^2\right\}, \left\{a_1^2, a_2^2, a_3^1, d^1\right\}\right\}$.

(5) It is easy to obtain $BND\left(\left\{a_1^2, a_2^2, a_3^2\right\}, d_2\right) = \{x_5, x_6, x_7, x_8\}$. Hence, if we let $OS = \left\{a_1^2, a_2^2, a_3^2, d^2\right\}$, then $Queue = \left\{\left\{a_1^2, a_2^2, a_3^2, d^1\right\}\right\}$.

(6) We have $BND\left(\left\{a_1^2, a_2^2, a_3^2\right\}, d_2\right) = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\}$. By Definition 2, $(\{2,2,2\}2)$ is the global optimal scale of S and $\left\{a_1^2, a_2^2, a_3^2, d^2\right\}$ is the global optimal attribute set of S.

Algorithm 2: Global optimal scale selection of GMSDTs.

Input: (1) A GMSDT $S=(U, AT \cup \{d\})$; (2) a global optimal scale OS.
Output: The global optimal scales OSS.

initialize $Queue$;
if $OS \neq AT^{L'} \bigcup \{d^1\}$ **then**
  calculate $BND\left(R_{AT^{L_0}}, D_n\right)$;
  $OSS \leftarrow OS$
  $T \leftarrow OS$
  for each $item \in Queue$ **do**
    if there exists $scale$ in T subject to $scale \preceq item$ **then**
      continue
    end if
    if there exists $scale$ in OSS subject to $item \preceq scale$ **then**
      continue
    end if
    calculate boundary regions of the scale $item$, denote as B
    if $B = BND\left(R_{AT^{L_0}}, D_n\right)$ **then**
      $OSS \leftarrow item$
    else
      $T \leftarrow item$
    end if
  end for
else
  $OSS \leftarrow AT^{L'} \bigcup \{d^1\}$
end if
for each $item \in OSS$ **do**
  for each $scale \in OSS - item$ **do**
    if there exists $scale$ in OSS subject to $item \preceq scale$ **then**
      $OSS = OSS - item$
    end if
  end for
end for
return OSS;

Furthermore, we now find all the global optimal scales for Example 2 on the basis of Algorithm 2.

It can be calculated that $BND\left(\left\{a_1^1, a_2^1, a_3^1\right\}, d_2\right) \neq BND\left(\left\{a_1^2, a_2^2, a_3^2\right\}, d_1\right)$. Therefore, let $OSS = \left\{\left\{a_1^2, a_2^2, a_3^2, d^2\right\}\right\}$ and $T = \left\{\left\{a_1^2, a_2^2, a_3^2, d^2\right\}\right\}$. Then we can obtain three global optimal scales of S by traversing all scale combinations in the lattice structure of scale combinations. They are $(\{2,2,2\}\, 2)$, $(\{1,1,2\}\, 1)$, and $(\{2,1,1\}\, 1)$. Hence, $\left\{a_1^2, a_2^2, a_3^2, d^2\right\}$, $\left\{a_1^1, a_2^1, a_3^2, d^1\right\}$, and $\left\{a_1^2, a_2^1, a_3^1, d^1\right\}$ are the global optimal attribute sets of S.

We now discuss feature selection of the optimal scale.

Definition 4. Let U be a nonempty finite set, where $\mathcal{X} = \{X_1, X_2, \ldots, X_n\}$ is a partition of U. The information entropy of $\mathcal{X}$ is defined as

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

where $P(X_i) = \frac{|X_i|}{|U|}$.

Definition 5. Let U be a nonempty finite set, where $\mathcal{X} = \{X_1, X_2, \ldots, X_n\}$ and $\mathcal{Y} = \{Y_1, Y_2, \ldots, Y_m\}$ are two partitions of U. The conditional entropy for partitioning $\mathcal{Y}$ with respect to partitioning $\mathcal{X}$ is defined as

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

where $H(\mathcal{Y} \mid X_i) = -\sum_{j=1}^{m} P\left(Y_j \mid X_i\right) \log_2 P\left(Y_j \mid X_i\right)$ and $P\left(Y_j \mid X_i\right) = \frac{\left|Y_j \cap X_i\right|}{X_i}$.

Definition 6. Let $S = (U, AT \cup D) = \left(U, \left\{a_j^k \mid k = 1, 2, \ldots, I_j; j = 1, 2, \ldots, m\right\} \cup \left\{d^t \mid t = 1, 2, \ldots, n\right\}\right)$ be a GMSDT, $L_0 = (1, 1, \ldots, 1)$, $L = (l_1, l_2, \ldots, l_m)$, and $t \in \{1, 2, \ldots, n\}$. $OS = (L, t) = (\{l_1, l_2, \ldots, l_m\}, t)$ is the global optimal scale of S. If $S \subseteq L$, $H\left(d^t \mid [x]_S\right) = H\left(d^t \mid [x]_L\right)$. If for all $s \in S$, $H\left(d^t \mid [x]_{S-\{s\}}\right) \neq H\left(d^t \mid [x]_L\right)$, then S is the optimal subset of the global optimal scale L.

4. Rule extraction and rough set classifier design

The classification problem is not only a core problem of supervised learning; it is also a problem that we often encounter in daily life. To achieve accurate classification, so as to better handle practical problems, what we should do is select the optimal scale to fully mine the hidden information in decision tables. In this section we propose a knowledge acquisition algorithm based on rough set theory to extract decision rules from decision tables. Furthermore, a multi-scale rough set classifier framework based on a ruleset is designed.

Algorithm 3: Decision rule acquisition.

Input: The optimal attribute set $AT^L \bigcup \{d^t\}$;
Output: Decision rules;
1: **for** each $class \in U/R_{d^t}$ **do**
2: **for** each $x \in class$ **do**
3: Output $\left(x, [x]_{AT^L} \bigcap class\right)$;
4: $class \leftarrow class - \left([x]_{AT^L} \bigcap class\right)$;
5: **end for**
6: **end for**

The decision rules can be abstracted by the optimal scale selected from the GMSDT. Hence, we provide a decision rule extraction method, as described in Algorithm 3. Not only the decision rules for each object but also the number of supported objects for each rule is output when the set of optimal scales for S is acquired in Algorithm 3. The time complexity of Algorithm 3 is $O\left(|U|^2\right)$, where $|U|$ is the number of objects.

Example 4 *(Continued from Example 3).* According to Algorithms 1 and 2, the optimal scale $(\{2,2,2\}2)$ is selected from the GMSDT given in Table 2, and the following decision rules can be acquired with Algorithm 3:
(1) Ruleset obtained from the first scale $(\{1,1,1\}2)$ of S, deterministic rule set are as follows:
$r_1 : \left(a_1^1, 1\right) \wedge \left(a_2^1, 1\right) \wedge \left(a_3^1, 1\right) \Rightarrow \left(d^2, +\right)$, supported by x_1, x_2;
$r_2 : \left(a_1^1, 1\right) \wedge \left(a_2^1, 2\right) \wedge \left(a_3^1, 1\right) \Rightarrow \left(d^2, +\right)$, supported by x_3;
$r_3 : \left(a_1^1, 2\right) \wedge \left(a_2^1, 1\right) \wedge \left(a_3^1, 2\right) \Rightarrow \left(d^2, +\right)$, supported by x_4;
$r_4 : \left(a_1^1, 3\right) \wedge \left(a_2^1, 4\right) \wedge \left(a_3^1, 5\right) \Rightarrow \left(d^2, -\right)$, supported by x_9;
$r_5 : \left(a_1^1, 4\right) \wedge \left(a_2^1, 3\right) \wedge \left(a_3^1, 5\right) \Rightarrow \left(d^2, -\right)$, supported by x_{10};
$r_6 : \left(a_1^1, 3\right) \wedge \left(a_2^1, 3\right) \wedge \left(a_3^1, 6\right) \Rightarrow \left(d^2, -\right)$, supported by x_{11}, x_{12}.
And possible ruleset are
$r_7 : \left(a_1^1, 3\right) \wedge \left(a_2^1, 3\right) \wedge \left(a_3^1, 3\right) \Rightarrow \left(d^2, +\right)$, supported by x_5;
$r_7' : \left(a_1^1, 3\right) \wedge \left(a_2^1, 3\right) \wedge \left(a_3^1, 3\right) \Rightarrow \left(d^2, -\right)$, supported by x_6;
$r_8 : \left(a_1^1, 4\right) \wedge \left(a_2^1, 4\right) \wedge \left(a_3^1, 4\right) \Rightarrow \left(d^2, +\right)$, supported by x_7;
$r_8' : \left(a_1^1, 4\right) \wedge \left(a_2^1, 4\right) \wedge \left(a_3^1, 4\right) \Rightarrow \left(d^2, -\right)$, supported by x_8.
(2) Ruleset obtained from the optimal scale $(\{2,2,2\}2)$ of S, deterministic rule set are
$r_1 : \left(a_1^2, S\right) \wedge \left(a_2^2, Y\right) \wedge \left(a_3^2, G\right) \Rightarrow \left(d^2, +\right)$, supported by x_1, x_2, x_3, x_4;
$r_2 : \left(a_1^2, L\right) \wedge \left(a_2^2, N\right) \wedge \left(a_3^2, B\right) \Rightarrow \left(d^2, -\right)$, supported by x_9, x_{10}, x_{11}, x_{12}.
And possible ruleset are
$r_3 : \left(a_1^2, L\right) \wedge \left(a_2^2, N\right) \wedge \left(a_3^2, F\right) \Rightarrow \left(d^2, +\right)$, supported by x_5, x_7;
$r_3' : \left(a_1^2, L\right) \wedge \left(a_2^2, N\right) \wedge \left(a_3^2, F\right) \Rightarrow \left(d^2, -\right)$, supported by x_6, x_8.

According to Example 4, rulesets extracted from the optimal scale are more concise than those extracted from the first scale and each rule has more supporting objects. Occam's razor principle is called entities should not be multiplied unnecessarily. If this principle is adopted, the rulesets extracted from the optimal scale are considered to be more efficient and accurate than the rulesets extracted from other scales.

However, to design a rough set classifier, it is not enough to rely only on the ruleset extracted from the optimal scale. For example, neither the deterministic rule nor the possible rule can determine the category of the object $\left(a_1^2, S\right) \wedge \left(a_2^2, N\right) \wedge \left(a_3^2, F\right)$. We need to consider the decision rules that apply to objects whose categories cannot be determined with either of the above-mentioned rules, so as to realize the automatic classification of objects to be classified.

There are three situations for the objects to be classified:
(1) Deterministic rules can be used to determine the category of objects such as $\left(a_1^2, S\right) \wedge \left(a_2^2, Y\right) \wedge \left(a_3^2, G\right)$. The category is $\left(d^2, +\right)$.
(2) Possible rules can be used to determine the category of objects such as $\left(a_1^2, L\right) \wedge \left(a_2^2, N\right) \wedge \left(a_3^2, F\right)$. Select the possible rule with the maximum number of supported objects to judge the category of objects.
(3) Neither deterministic rules nor possible rules can judge the category of objects such as $\left(a_1^2, L\right) \wedge \left(a_2^2, N\right) \wedge \left(a_3^2, G\right)$. Calculate the distance of the object from the antecedent of all deterministic and possible rules in the ruleset. Select the rule

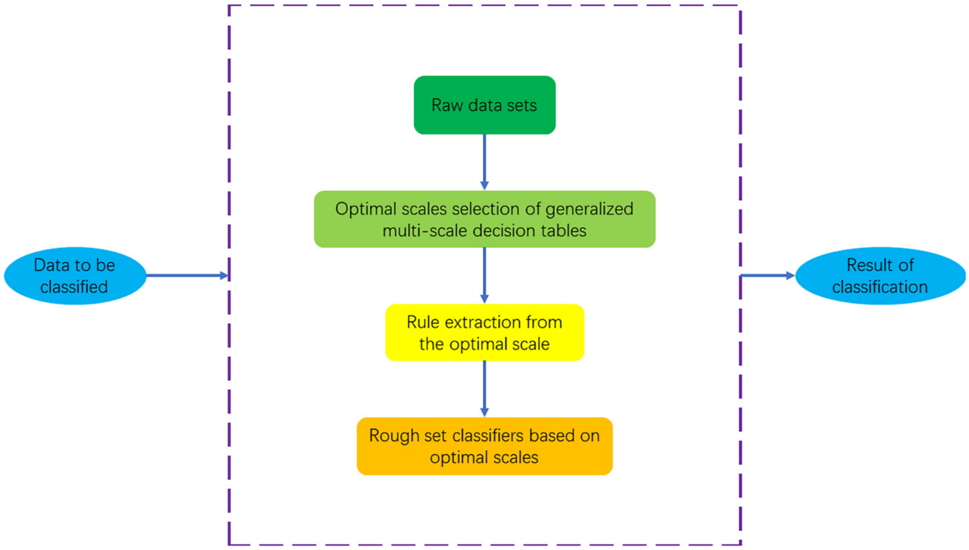

Fig. 3. A multi-scale rough set classifier framework.

Table 3
Operating environment.

Component	Model	Parameter
CPU	AMD Ryzen 7 6800H	3.2 GHz
Platform	Python	3.9
Operating system	Windows 11	64 bit
Memory	DDR5	16 GB; 6400 MHz
Hard disk	MTFDKBA512TFH	512 GB

whose antecedent is closest to the object. Select the category with the maximum number of supported objects to judge the category of objects. The distance used here is the Hamming distance, and the formula is $d(x, y) = \frac{1}{N}\sum_i 1_{x_i \neq x_j}$.

It is not difficult to see that (1) and (2) are two special cases when the Hamming distance is 0, and (3) is a reasonable generalization of (1) and (2).

Example 5 *(Continued from Example 4).* Use the ruleset obtained from the optimal scale $(\{2, 2, 2\}2)$ of S to determine the category of the object $x : \left(a_1^2, L\right) \wedge \left(a_2^2, N\right) \wedge \left(a_3^2, G\right)$.

Let the antecedent of rules r_1, r_2, r_3, r_3' be a_1, a_2, a_3, a_3'. By calculation, we can get $d(x, a_1) = 1, d(x, a_2) = 2, d(x, a_3) = 2, d\left(x, a_3'\right) = 2$. Because $d(x, a_1) < d(x, a_2) = d(x, a_3) = d\left(x, a_3'\right)$, select the rule r_2, r_3, r_3'. The objects that support category $\left(d^2, -\right)$ are $\{x_6, x_8, x_9, x_{10}, x_{11}, x_{12}\}$. The objects that support category $\left(d^2, +\right)$ are $\{x_5, x_7\}$. Apparently, category $\left(d^2, -\right)$ has the largest number of supported objects. Therefore, we judge that object x belongs to category $\left(d^2, -\right)$.

Using the method outlined above, we can design a rough set classifier to realize the automatic classification of objects. For GMSDTs, we can select multiple optimal scales through the optimal scale selection algorithms proposed in this article. Because each optimal scale can train a rough set classifier, a multi-scale rough set classifier can be constructed by our taking the mode of predicted results. Fig. 3 shows the framework for a multi-scale rough set classifier.

5. Experiments and analysis

5.1. Data description

All programs for the experiments were written in Python and were executed on a personal computer. The details of the operating environment are shown in Table 3.

Twelve open data sets from the University of California, Irvine were selected to verify the algorithms presented in this article. Details of these 12 open data sets are shown in Table 4.

The method of creating GMSDTs in [9] is used to transform standard decision tables into GMSDTs. For continuous attributes, we first use the k-means clustering algorithm to discretize the attributes. The cluster core is used as the new attribute value. Finally, the method of creating GMSDTs in [9] is used to transform the single-scale attribute into the multi-scale attribute. For data sets with more than 20 attributes, the clustering number K of each attribute is uniformly specified as 6.

Table 4
Description of data sets.

Data set	Instances	Features	Classes
Iris	150	4	3
Wholesale customers (region)	440	6	3
Contraceptive Method Choice	1473	9	3
Car Evaluation	1728	6	4
Estimation of obesity levels based on eating habits and physical condition	2111	16	7
Abalone	4177	8	9
Wine Quality (white)	4898	11	7
Shill Bidding Dataset	6321	9	2
Nursery	12960	8	5
Mice Protein Expression	1080	80	8
MEU-Mobile KSD	2856	71	56
Polish companies bankruptcy data	10503	64	2

Table 5
Optimal scale combination and running time for three data sets with Algorithm 1.

Data set	Scales	Optimal scale combination	Running time (s)
Mice Protein Expression	$5 \times 5 \times \cdots \times 1 \times 1$	$(5, 5, 5, 5, 5, \ldots, 5, 1, 1, 1, 1)$	0.805225
MEU-Mobile KSD	$5 \times 5 \times \cdots \times 5 \times 1$	$(5, 5, 5, 5, 5, \ldots, 5, 1, 5, 5, 1)$	500.804162
Polish companies bankruptcy data	$5 \times 5 \times \cdots \times 5 \times 1$	$(5, 5, 5, 5, 5, \ldots, 5, 5, 5, 5, 1)$	15609.695373

Table 6
Running time for nine data sets with Algorithm 2.

Data set	Running time (s)		
	This work	Method in [9]	Method in [11]
Iris	**2.522581**	14.318465	13.844829
Wholesale customers (region)	**0.600713**	140.869354	142.895350
Contraceptive Method Choice	**2.057566**	200.788399	229.648538
Car Evaluation	**10.353220**	82.681253	80.569686
Estimation of obesity levels based on eating habits and physical condition	**23.779038**	29118.080091	42730.959870
Abalone	**4.085046**	3164.573462	3368.164955
Wine Quality (white)	**539.554701**	Long time	Long time
Shill Bidding Dataset	**250.044801**	1185.044555	1097.402318
Nursery	**144.847004**	1605.351084	1188.747542

5.2. Experimental results

Algorithms 1 and 2 can be used to acquire the global optimal scale. The experimental results are displayed in Tables 5–7.

The algorithms in [9,11] find all optimal scale combinations by traversing all scale combinations, and their outer time complexity is $O\left(\left(\prod_{j=1}^{m} I_j \times n\right)\right)$. With increasing number of attributes in the data set, the time complexity increases exponentially. Considering that it is not necessary to find all optimal scale combinations in practical applications, Algorithm 1, which can find an optimal scale combination, is proposed in this article. Its outer time complexity is $O\left(\left(\sum_{j=1}^{m} I_j + n\right)\right)$, and it can handle big data sets with variation in a number of attributes. From Table 5, we can see a global optimal scale combination obtained by Algorithm 1 and the corresponding running time.

From Table 6, we can see the running time of Algorithm 2 in this article and of the algorithms in [9,11] for global optimal scale selection. It should be noted that the time efficiency of Algorithm 2 in this article is much greater than that of the algorithms in [9,11] even though the time complexity is the same.

We can see that the global optimal scale combination can be acquired by Algorithms 1 and 2 from Table 7. It should be noted that there is more than one global optimal scale combination in some cases.

From Fig. 4, we know that the fewer the condition attributes, the shorter the running time of the global optimal scale selection, while the running time of the global optimal scale selection is not necessarily related to the size of the dataset.

In the next experiment, we apply four diverse kinds of classifiers to compare the performance for the global optimal scale (GOS), the first scale (FirS), the final scale (FinS), the optimal subset about the global optimal scale (OSGOS), the optimal subset about the first scale (OSFirS), and the optimal subset about the final scale (OSFinS). They are the k-nearest-neighbor (KNN) classifier, the classification and regression trees (CART) classifier, the support vector machine (SVM) classifier, and the rough set classifier proposed in this article.

The experiments in this article use fivefold cross-validation. The experimental results are given in Tables 8–11. All data are averaged from five independent experiments.

Table 7
Optimal scale combination for nine data sets.

Data set	Scales	Optimal scale combination
Iris	$5 \times 5 \times 5 \times 5 \times 1$	(2, 3, 3, 4, 1) (5, 3, 4, 3, 1) (1, 3, 5, 4, 1) (3, 3, 5, 3, 1) (5, 2, 3, 4, 1) (5, 3, 2, 4, 1)
Wholesale customers (region)	$4 \times 4 \times 4 \times 4 \times 4 \times 4 \times 1$	(1, 1, 1, 1, 1, 1, 1)
Contraceptive Method Choice	$2 \times 3 \times 3 \times 7 \times 1 \times 1 \times 3 \times 3 \times 1 \times 1$	(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)
Car Evaluation	$3 \times 3 \times 3 \times 2 \times 2 \times 2 \times 3$	(1, 2, 1, 1, 1, 2, 3) (1, 1, 1, 1, 1, 1, 1)
Estimation of obesity levels based on eating habits and physical condition	$1 \times 3 \times 3 \times 7 \times 1 \times 1 \times 2 \times 3 \times 3 \times 1 \times 2 \times 1 \times 3 \times 2 \times 3 \times 1 \times 3$	(1, 1, 2, 3, 1, 1, 1, 2, 3, 1, 1, 1, 1, 1, 1, 1, 2)
Abalone	$1 \times 3 \times 3 \times 4 \times 3 \times 3 \times 3 \times 3 \times 4$	(1, 1, 1, 1, 1, 1, 1, 1, 4)
Wine Quality (white)	$5 \times 6 \times 6 \times 2 \times 2 \times 8 \times 2 \times 2 \times 6 \times 4 \times 3 \times 3$	(2, 2, 2, 1, 1, 1, 1, 1, 2, 1, 1, 3)
Shill Bidding Dataset	$3 \times 3 \times 1 \times 3 \times 3 \times 1 \times 3 \times 1 \times 2 \times 1$	(3, 3, 1, 3, 3, 1, 2, 2, 1, 1) (2, 3, 1, 3, 3, 1, 2, 3, 1, 1) (3, 3, 1, 2, 3, 1, 3, 2, 1, 1) (2, 3, 1, 2, 3, 1, 3, 3, 1, 1) (2, 3, 1, 3, 3, 1, 3, 1, 1, 1)
Nursery	$2 \times 3 \times 2 \times 3 \times 2 \times 1 \times 2 \times 2 \times 3$	(1, 3, 1, 1, 1, 1, 2, 1, 3) (1, 2, 1, 1, 1, 1, 2, 1, 2) (1, 1, 1, 1, 1, 1, 2, 1, 1)

5.3. Statistical analysis

In this subsection, the Friedman test [3] and corresponding post hoc tests [2] are applied to reveal the differences for different scales. By the Friedman test, we have that

$$F_F = \frac{(M-1)\,\chi_F^2}{M\,(s-1)-\chi_F^2} \sim F\,(s-1,(s-1)\,(M-1))\,,$$
$$\chi_F^2 = \frac{12M}{s\,(s-1)}\left(\sum_{i=1}^{s} R_i^2 - \frac{s\,(s+1)^2}{4}\right),$$

where M and s are the number of datasets and compared scales, respectively, and R_i is the mean ranking of classification accuracy.

According to the procedure for the statistical test in this article, the mean ranking can be computed for each model. Table 12 gives the mean raking and the corresponding F_F and P values for three scales under four classifiers.

Table 12 shows that the P values under the four classifiers are all well below the selected significance level $\alpha = 0.01$; therefore, one can reject the original hypothesis. There are significant differences in classification accuracy among these scales under the KNN classifier, CART classifier, SVM classifier, and rough set classifier.

In this case, the Bonferroni-Dunn statistical test is used to uncover the statistical discrepancy of different scales. The critical value can be calculated in accordance with the following formula:

$$CD_\alpha = q_\alpha \sqrt{\frac{s\,(s+1)}{6M}}.$$

If the difference between the average ranking of two scales is greater than CD_α, then they will be considered to be markedly different. In Fig. 5, a special form of graph is introduced to visually show the differences between different scales. If there is a linear connection between the two scales, then they are not significantly different from each other.

From Fig. 5(a) and 5(b), for the KNN classifier and the CART classifier, we can see that there is a significant difference between OSGOS and FirS, while GOS, FinS, OSFirS, and OSFinS have no significant difference with FirS. From Fig. 5(c), for the SVM classifier, it can be seen that there is no significant difference between GOS, OSFirS, FirS, FinS, and OSFinS. It can be seen from Fig. 5(d), for the rough set classifier, that there is no significant difference between GOS, OSFirS, FirS, and FinS, while OSGOS has a significant difference with FirS and FinS. In conclusion, OSGOS is superior to GOS, OSFirS, FirS, FinS, and OSFinS from the results of the two statistical tests.

6. Conclusion

Research on multi-scale rough sets has achieved fruitful results. However, there has not been much research on GMSDTs with multi-scale decision attributes, especially for inconsistent GMSDTs. In this article, we introduced GMSDTs with multi-scale decision attributes. Some of the related properties and theorems were presented. What is more, the definition of

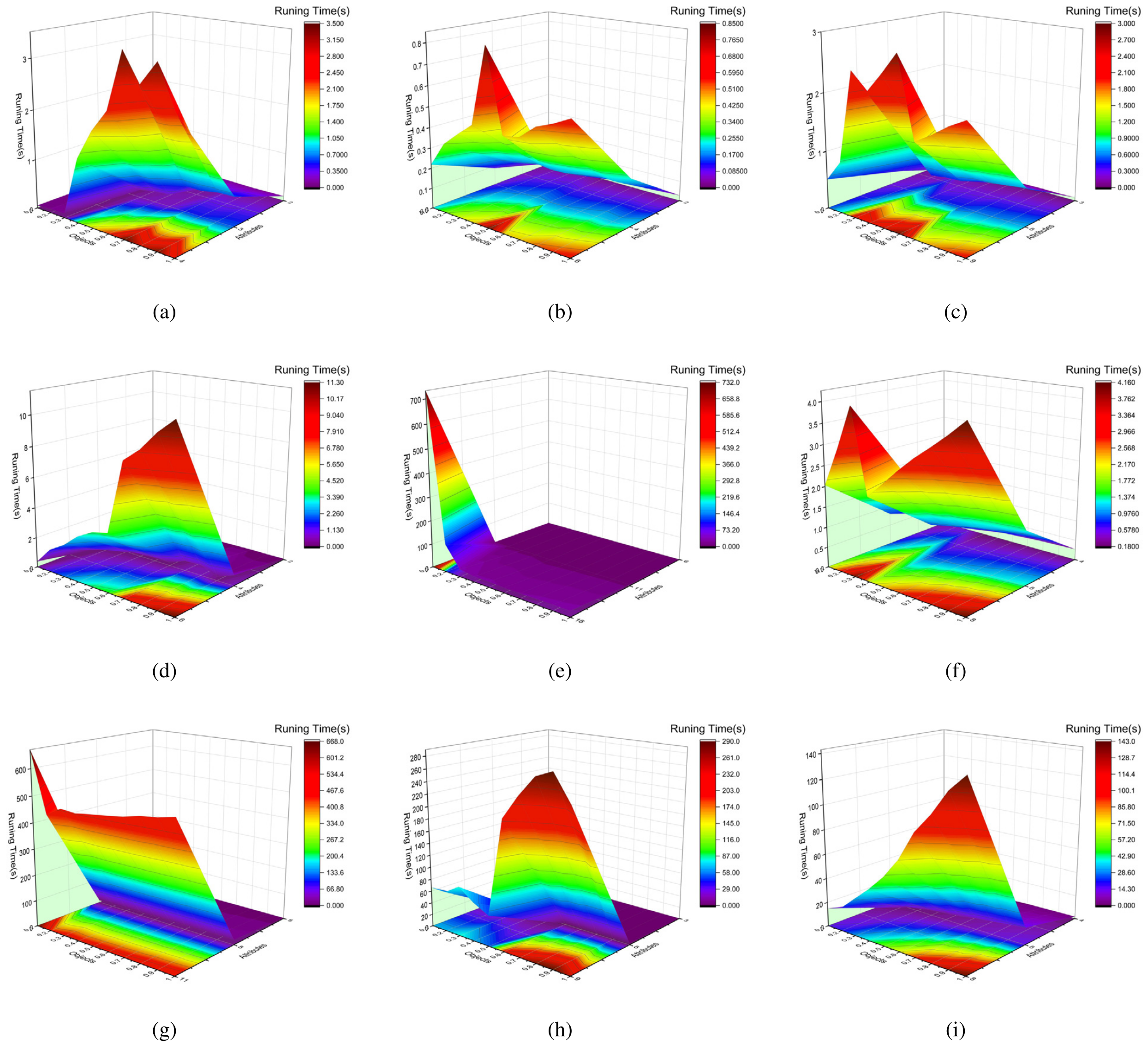

Fig. 4. Experiments with different sizes of the universe: (a) Iris; (b) Wholesale customers (region); (c) Contraceptive Method Choice; (d) Car Evaluation; (e) Estimation of obesity levels based on eating habits and physical condition; (f) Abalone; (g) Wine Quality (white); (h) Shill Bidding Dataset; (i) Nursery. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

Table 8
Classification accuracy (%) of six different optimal subset methods applied on 12 data sets obtained with the KNN classifier.

Data set	FirS	FinS	GOS	OSFirS	OSFinS	OSGOS
Iris	96.00 ± 0.33	90.67 ± 1.11	96.17 ± 0.17	93.67 ± 1.94	81.83 ± 2.00	$\mathbf{96.17 \pm 0.41}$
Wholesale customers (region)	70.80 ± 0.79	70.64 ± 3.33	70.80 ± 0.79	$\mathbf{70.97 \pm 1.58}$	70.07 ± 2.64	$\mathbf{70.97 \pm 1.58}$
Contraceptive Method Choice	51.63 ± 0.94	43.51 ± 1.12	51.63 ± 0.94	$\mathbf{52.31 \pm 1.33}$	44.06 ± 1.17	$\mathbf{52.31 \pm 1.33}$
Car Evaluation	98.70 ± 0.18	72.66 ± 0.58	93.05 ± 0.56	98.39 ± 0.24	71.55 ± 0.90	$\mathbf{98.83 \pm 0.18}$
Estimation of obesity levels based on eating habits and physical condition	91.68 ± 0.25	61.59 ± 0.61	$\mathbf{92.16 \pm 0.37}$	91.77 ± 0.44	61.09 ± 1.48	92.04 ± 0.12
Abalone	88.20 ± 0.46	49.90 ± 0.93	88.20 ± 0.46	$\mathbf{88.29 \pm 0.32}$	49.92 ± 2.11	$\mathbf{88.29 \pm 0.32}$
Wine Quality (white)	61.35 ± 0.46	44.63 ± 0.80	$\mathbf{62.13 \pm 0.61}$	61.65 ± 0.72	44.80 ± 0.75	61.80 ± 0.29
Shill Bidding Dataset	96.86 ± 0.08	98.16 ± 0.09	98.38 ± 0.06	97.38 ± 0.24	98.30 ± 0.14	$\mathbf{98.64 \pm 0.13}$
Nursery	98.91 ± 0.05	84.95 ± 0.29	$\mathbf{99.25 \pm 0.09}$	98.63 ± 0.08	84.35 ± 0.27	98.99 ± 0.06
Mice Protein Expression	99.07 ± 0.19	96.09 ± 0.31	99.07 ± 0.19	$\mathbf{99.21 \pm 0.22}$	96.46 ± 0.7	$\mathbf{99.21 \pm 0.22}$
MEU-Mobile KSD	61.20 ± 0.60	53.84 ± 0.52	61.26 ± 0.56	64.23 ± 0.29	54.1 ± 0.59	$\mathbf{64.79 \pm 0.41}$
Polish companies bankruptcy data	95.26 ± 0.10	95.21 ± 0.10	95.31 ± 0.12	95.31 ± 0.12	95.29 ± 0.09	$\mathbf{95.33 \pm 0.13}$

Table 9
Classification accuracy (%) of six different optimal subset methods applied on 12 data sets obtained with the CART classifier.

Data set	FirS	FinS	GOS	OSFirS	OSFinS	OSGOS
Iris	93.67 ± 1.00	84.00 ± 2.13	93.83 ± 0.41	94.67 ± 1.45	76.50 ± 5.78	$\mathbf{95.33 \pm 1.13}$
Wholesale customers (region)	72.10 ± 0.46	71.31 ± 1.02	72.10 ± 0.46	$\mathbf{72.33 \pm 1.06}$	71.36 ± 0.79	$\mathbf{72.33 \pm 1.06}$
Contraceptive Method Choice	43.26 ± 0.56	42.70 ± 0.79	43.26 ± 0.56	$\mathbf{43.28 \pm 0.61}$	42.87 ± 0.50	$\mathbf{43.28 \pm 0.61}$
Car Evaluation	96.09 ± 0.18	69.97 ± 0.49	96.19 ± 0.19	96.15 ± 0.28	70.02 ± 0.32	$\mathbf{96.32 \pm 0.13}$
Estimation of obesity levels based on eating habits and physical condition	84.92 ± 0.50	42.50 ± 0.40	85.26 ± 0.29	85.06 ± 0.60	42.82 ± 0.32	$\mathbf{85.41 \pm 0.28}$
Abalone	88.24 ± 0.12	45.29 ± 0.29	88.24 ± 0.12	$\mathbf{88.31 \pm 0.31}$	45.42 ± 0.42	$\mathbf{88.31 \pm 0.31}$
Wine Quality (white)	44.86 ± 0.34	44.64 ± 0.31	$\mathbf{45.00 \pm 0.19}$	44.81 ± 0.29	44.71 ± 0.19	45.00 ± 0.21
Shill Bidding Dataset	97.31 ± 0.04	97.23 ± 0.05	$\mathbf{97.38 \pm 0.06}$	97.24 ± 0.04	97.29 ± 0.01	97.32 ± 0.14
Nursery	97.43 ± 0.09	66.27 ± 0.20	97.47 ± 0.06	97.44 ± 0.05	66.30 ± 0.18	$\mathbf{97.51 \pm 0.06}$
Mice Protein Expression	100.0 ± 0.00	100.0 ± 0.00	100.0 ± 0.00	100.0 ± 0.00	100.0 ± 0.00	100.0 ± 0.00
MEU-Mobile KSD	2.01 ± 0.04	2.01 ± 0.08	2.03 ± 0.05	2.04 ± 0.06	2.0 ± 0.04	$\mathbf{2.07 \pm 0.10}$
Polish companies bankruptcy data	95.28 ± 0.07	95.30 ± 0.11	95.32 ± 0.11	95.29 ± 0.07	95.28 ± 0.05	$\mathbf{95.39 \pm 0.08}$

Table 10
Classification accuracy (%) of six different optimal subset methods applied on 12 data sets obtained with the SVM classifier.

Data set	FirS	FinS	GOS	OSFirS	OSFinS	OSGOS
Iris	96.33 ± 0.41	89.50 ± 0.85	96.50 ± 0.62	96.33 ± 1.55	80.33 ± 2.15	$\mathbf{97.17 \pm 1.13}$
Wholesale customers (region)	72.33 ± 0.52	71.36 ± 0.33	72.33 ± 0.52	$\mathbf{73.82 \pm 0.66}$	71.54 ± 1.10	$\mathbf{73.82 \pm 0.66}$
Contraceptive Method Choice	55.43 ± 0.99	46.83 ± 1.22	55.43 ± 0.99	$\mathbf{55.99 \pm 0.25}$	46.66 ± 0.41	$\mathbf{55.99 \pm 0.25}$
Car Evaluation	$\mathbf{99.77 \pm 0.15}$	73.47 ± 0.24	97.55 ± 0.17	98.93 ± 0.13	73.79 ± 0.57	98.99 ± 0.29
Estimation of obesity levels based on eating habits and physical condition	92.50 ± 0.26	65.33 ± 0.40	92.65 ± 0.24	92.40 ± 0.44	64.76 ± 0.86	$\mathbf{92.96 \pm 0.61}$
Abalone	88.16 ± 0.13	54.27 ± 0.31	88.16 ± 0.13	$\mathbf{88.30 \pm 0.19}$	53.68 ± 0.42	$\mathbf{88.30 \pm 0.19}$
Wine Quality (white)	57.50 ± 0.33	51.76 ± 0.75	53.73 ± 0.47	56.19 ± 0.70	47.64 ± 0.85	$\mathbf{57.77 \pm 1.15}$
Shill Bidding DataSet	98.89 ± 0.08	98.58 ± 0.09	99.49 ± 0.08	99.05 ± 0.24	99.22 ± 0.17	$\mathbf{99.59 \pm 0.05}$
Nursery	99.95 ± 0.02	86.17 ± 0.25	99.97 ± 0.01	98.16 ± 0.32	98.11 ± 0.26	$\mathbf{99.01 \pm 0.24}$
Mice Protein Expression	99.86 ± 0.09	$\mathbf{100.0 \pm 0.00}$	99.86 ± 0.09	99.91 ± 0.05	$\mathbf{100.0 \pm 0.00}$	99.91 ± 0.05
MEU-Mobile KSD	73.98 ± 0.23	68.51 ± 0.21	74.2 ± 0.52	80.56 ± 0.44	67.42 ± 0.77	$\mathbf{80.73 \pm 0.29}$
Polish companies bankruptcy data	96.07 ± 0.25	96.07 ± 0.2	$\mathbf{96.33 \pm 0.27}$	95.95 ± 0.13	96.00 ± 0.28	96.17 ± 0.20

Table 11
Classification accuracy (%) of six different optimal subset methods applied on 12 data sets obtained with the rough set classifier.

Data set	FirS	FinS	GOS	OSFirS	OSFinS	OSGOS
Iris	92.13 ± 4.61	88.53 ± 5.17	93.33 ± 4.71	92.13 ± 5.32	79.20 ± 4.65	$\mathbf{94.13 \pm 2.54}$
Wholesale customers (region)	59.32 ± 4.82	$\mathbf{69.64 \pm 3.78}$	59.32 ± 4.82	61.32 ± 5.36	69.55 ± 3.62	61.32 ± 5.36
Contraceptive Method Choice	47.81 ± 1.75	45.23 ± 2.52	47.81 ± 1.75	$\mathbf{48.38 \pm 2.88}$	46.64 ± 2.39	$\mathbf{48.38 \pm 2.88}$
Car Evaluation	96.79 ± 0.89	72.44 ± 2.35	98.82 ± 0.50	98.55 ± 0.62	73.31 ± 2.49	$\mathbf{98.83 \pm 0.47}$
Estimation of obesity levels based on eating habits and physical condition	89.92 ± 1.65	63.58 ± 2.19	90.11 ± 1.26	90.10 ± 1.40	64.16 ± 1.77	$\mathbf{90.17 \pm 1.20}$
Abalone	86.61 ± 0.92	52.84 ± 1.20	86.61 ± 0.92	$\mathbf{86.91 \pm 0.84}$	52.74 ± 1.66	$\mathbf{86.91 \pm 0.84}$
Wine Quality (white)	64.61 ± 1.43	46.18 ± 1.28	$\mathbf{64.83 \pm 1.14}$	64.41 ± 1.66	46.24 ± 1.40	64.74 ± 0.94
Shill Bidding Dataset	97.37 ± 0.42	98.00 ± 0.34	98.57 ± 0.27	98.16 ± 0.32	98.11 ± 0.26	$\mathbf{99.01 \pm 0.24}$
Nursery	98.50 ± 0.28	85.73 ± 0.69	$\mathbf{99.92 \pm 0.10}$	98.20 ± 0.24	85.96 ± 0.71	98.71 ± 0.14
Mice Protein Expression	98.98 ± 0.52	97.43 ± 1.01	98.98 ± 0.52	$\mathbf{99.00 \pm 0.69}$	97.50 ± 1.01	$\mathbf{99.00 \pm 0.69}$
MEU-Mobile KSD	52.59 ± 2.13	49.26 ± 2.40	52.33 ± 1.68	63.37 ± 1.92	49.09 ± 1.58	$\mathbf{63.78 \pm 1.39}$
Polish companies bankruptcy data	94.26 ± 0.49	94.27 ± 0.5	95.24 ± 0.51	94.27 ± 0.41	94.32 ± 0.44	$\mathbf{95.24 \pm 0.44}$

Table 12
Friedman test.

Classifier	Mean ranking						F_F	P
	FirS	FinS	GOS	OSFirS	OSFinS	OSGOS		
KNN	3.75	5.42	2.54	2.71	5.17	1.42	43.44	3.00×10^{-8}
CART	3.71	5.33	2.46	2.79	5	1.46	43.31	3.20×10^{-8}
SVM	3.33	4.88	2.92	3.17	5.04	1.67	28.41	3.03×10^{-5}
Rough set	4	5.17	2.67	2.92	4.75	1.5	33.64	2.81×10^{-6}

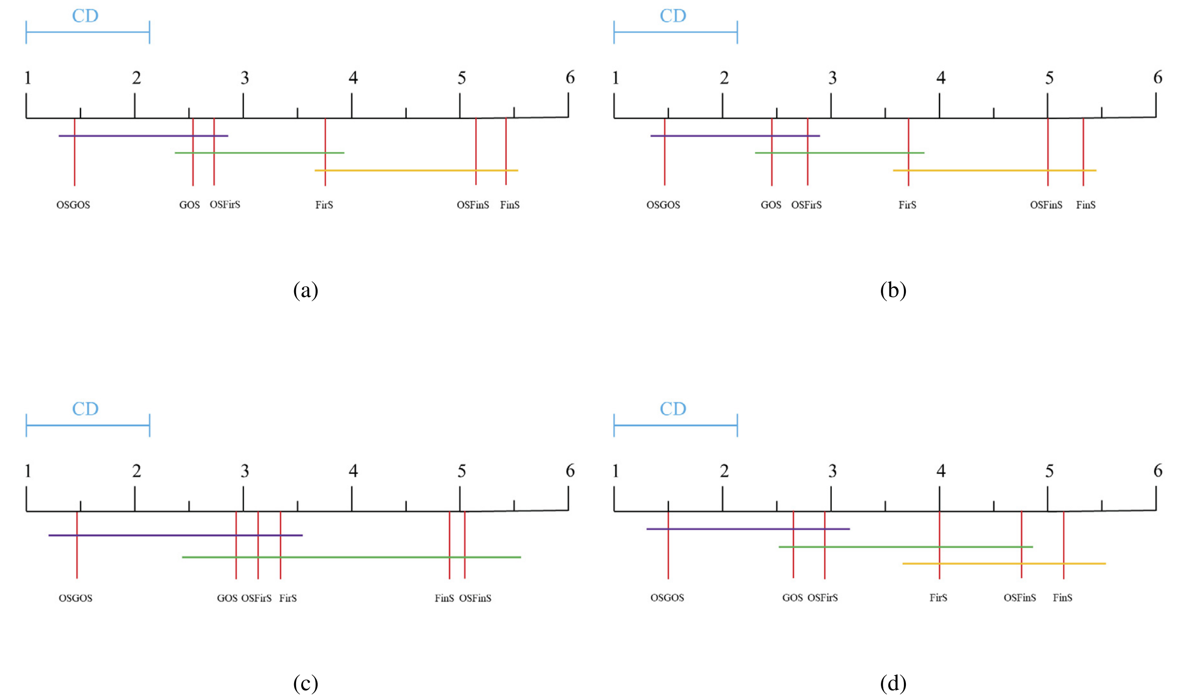

Fig. 5. Critical distance (CD) diagram for the Bonferroni-Dunn test: (a) KNN classifier; (b) CART classifier; (c) SVM classifier; (d) rough set classifier.

optimal scale in GMSDTs and two algorithms for optimal scale selection were proposed. The algorithms were proved to be efficient by experiments and can handle data sets with a large number of attributes: however, the running time of the algorithms still needs to be improved. It is well known that classical rough set theory can deal only with symbolic data. If classical rough set theory is used to handle continuous data, it is necessary to discretize the data first. The optimal scale selection method proposed in this article can be used to determine the number of categories for discretization of continuous attributes. Besides, a knowledge acquisition algorithm and a multi-scale rough set classifier framework were proposed. The above-mentioned work solves the partial molecular problem of optimal scale selection and knowledge discovery in GMSDTs with multi-scale decision attributes. But there are still many challenging problems to be addressed before these problems can be completely solved, which will be considered in the future, such as optimal scale selection and knowledge discovery in incomplete GMSDTs with multi-scale decision attributes, an optimal scale updating method for dynamic MSDTs, and computational efficiency of large-scale data sets.

Future research will focus on optimal scale selection and knowledge discovery in incomplete generalized multi-scale tables with multi-scale decision attributes. Besides, more attention will be focused on the application of these models, especially the most common classification problems in life.

CRediT authorship contribution statement

Xiaoyan Zhang: Conceptualization, funding acquisition, investigation, methodology, project administration, supervision, validation. Yuyang Huang: Data curation, methodology, software, visualization, writing—original draft, writing—review and editing.

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 article.

Data availability

No data was used for the research described in the article.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 61976245).

References

[1] H. Bao, W.Z. Wu, J.W. Zheng, T.J. Li, Entropy based optimal scale combination selection for generalized multi-scale information tables, Int. J. Mach. Learn. Cybern. 12 (2021) 1427–1437.
[2] O.J. Dunn, Multiple comparisons among means, J. Am. Stat. Assoc. 56 (1961) 52–64.
[3] M. Friedman, A comparison of alternative tests of significance for the problem of m rankings, Ann. Math. Stat. 11 (1940) 86–92.
[4] D.D. Guo, W.H. Xu, Fuzzy-based concept-cognitive learning: an investigation of novel approach to tumor diagnosis analysis, Inf. Sci. 639 (2023) 118998, https://doi.org/10.1016/j.ins.2023.118998.
[5] D.D. Guo, C.M. Jiang, R.X. Sheng, S.S. Liu, A novel outcome evaluation model of three-way decision: a change viewpoint, Inf. Sci. 607 (2022) 1089–1110.
[6] D.D. Guo, C.M. Jiang, P. Wu, Three-way decision based on confidence level change in rough set, Int. J. Approx. Reason. 143 (2022) 57–77.
[7] S.M. Gu, W.Z. Wu, On knowledge acquisition in multi-scale decision systems, Int. J. Mach. Learn. Cybern. 4 (2013) 477–486.
[8] B.Q. Hu, Three-way decisions space and three-way decisions, Inf. Sci. 281 (2014) 21–52.
[9] Z.H. Huang, J.J. Li, W.Z. Dai, R.D. Lin, Generalized multi-scale decision tables with multi-scale decision attributes, Int. J. Approx. Reason. 115 (2019) 194–208.
[10] C. Hao, J.H. Li, M. Fan, W. Liu, E.C.C. Tsang, Optimal scale selection in dynamic multi-scale decision tables based on sequential three-way decisions, Inf. Sci. 415–416 (2017) 213–232.
[11] F. Li, B.Q. Hu, A new approach of optimal scale selection to multi-scale decision tables, Inf. Sci. 381 (2017) 193–208.
[12] 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.
[13] J.H. Li, C.L. Mei, W.H. Xu, Y.H. Qian, Concept learning via granular computing: a cognitive viewpoint, Inf. Sci. 298 (2015) 447–467.
[14] W.T. Li, H.X. Zhou, W.H. Xu, X.Z. Wang, W. Pedrycz, Interval dominance-based feature selection for interval-valued ordered data, IEEE Trans. Neural Netw. Learn. Syst. (2022), https://doi.org/10.1109/TNNLS.2022.3184120.
[15] T.Y. Lin, Granular computing: from rough sets and neighborhood systems to information granulation and computing with words, in: European Congress on Intelligent Techniques and Soft Computing, September 8-12, 1997, pp. 1602–1606.
[16] T.Y. Lin, Granular computing on binary relations II: rough set representations and belief functions, in: Rough Sets in Knowledge Discovery, 1998, pp. 121–140.
[17] C. Luo, T.R. Li, Y.Y. Huang, H. Fujita, Updating three-way decisions in incomplete multi-scale information systems, Inf. Sci. 476 (2018) 274–289.
[18] Z. Pawlak, Rough sets, Int. J. Comput. Inf. Sci. 11 (1982) 341–356.
[19] Y.H. Qian, J.Y. Liang, C.Y. Dang, Incomplete multi-granulation rough set, IEEE Trans. Syst. Man Cybern. 40 (2010) 420–431.
[20] Y.H. Qian, J.Y. Liang, C.Y. Dang, Knowledge structure, knowledge granulation and knowledge distance in a knowledge base, Int. J. Approx. Reason. 50 (2009) 174–188.
[21] Y.H. Qian, S.Y. Li, J.Y. Liang, Z.Z. Shi, F. Wang, Pessimistic rough set based decisions: a multi-granulation fusion strategy, Inf. Sci. 264 (2014) 196–210.
[22] Y.H. Qian, J.Y. Liang, Y.Y. Yao, C.Y. Dang, MGRS: a multi-granulation rough set, Inf. Sci. 180 (2010) 949–970.
[23] Y.H. Qian, H. Zhang, Y.L. Sang, J.Y. Liang, Multigranulation decision-theoretic rough sets, Int. J. Approx. Reason. 55 (2014) 225–237.
[24] Y.H. She, J.H. Li, H.L. Yang, A local approach to rule induction in multi-scale decision tables, Knowl.-Based Syst. 89 (2015) 398–410.
[25] A.H. Tan, J.J. Li, Y.J. Lin, G.P. Lin, Matrix-based set approximations and reductions in covering decision information systems, Int. J. Approx. Reason. 59 (2015) 68–80.
[26] A.H. Tan, W.Z. Wu, Y.Z. Tao, On the belief structures and reductions of multi-granulation spaces with decisions, Int. J. Approx. Reason. 88 (2017) 39–52.
[27] A.H. Tan, W.Z. Wu, Y.H. Qian, J.Y. Liang, J.K. Chen, J.J. Li, Intuitionistic fuzzy rough set-based granular structures and attribute subset selection, IEEE Trans. Fuzzy Syst. 27 (2019) 527–539.
[28] R. Wille, Restructuring lattice theory: an approach based on hierarchies of concepts, Ordered Sets 83 (1982) 445–470.
[29] 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.
[30] W.Z. Wu, Y. Leung, Optimal scale selection for multi-scale decision tables, Int. J. Approx. Reason. 54 (2013) 1107–1129.
[31] W.Z. Wu, Y. Leung, Theory and applications of granular labeled partitions in multi-scale decision tables, Inf. Sci. 181 (2011) 3878–3897.
[32] W.Z. Wu, Y.H. Qian, T.J. Li, S.M. Gu, On rule acquisition in incomplete multi-scale decision tables, Inf. Sci. 378 (2017) 282–302.
[33] W.H. Xu, D.D. Guo, Y.H. Qian, W.P. Ding, Two-way concept-cognitive learning method: a fuzzy-based progressive learning, IEEE Trans. Fuzzy Syst. 31 (2023) 1885–1899, https://doi.org/10.1109/TFUZZ.2022.3216110.
[34] W.H. Xu, D.D. Guo, J.S. Mi, Y.H. Qian, K.Y. Zheng, W.P. Ding, Two-way concept-cognitive learning via concept movement viewpoint, IEEE Trans. Neural Netw. Learn. Syst. (2023), https://doi.org/10.1109/TNNLS.2023.3235800.
[35] 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 (2016) 366–379.
[36] W.H. Xu, Y.Z. Pan, X.W. Chen, Y.H. Qian, W.P. Ding, A novel dynamic fusion approach using information entropy for interval-valued ordered datasets, IEEE Trans. Big Data 9 (2023) 845–859, https://doi.org/10.1109/TBDATA.2022.3215494.
[37] W.H. Xu, K.H. Yuan, W.T. Li, Dynamic updating approximations of local generalized multigranulation neighborhood rough set, Appl. Intell. 52 (2022) 9148–9173.
[38] W.H. Xu, K.H. Yuan, W.T. Li, W.P. Ding, An emerging fuzzy feature selection method using composite entropy-based uncertainty measure and data distribution, IEEE Trans. Emerg. Top. Comput. Intell. 7 (2023) 76–88.
[39] Y.Y. Yao, Granular computing using neighborhood systems, in: Advances in Soft Computing, Springer-Verlag, London, 1999, pp. 539–553.
[40] Y.Y. Yao, Information granulation and rough set approximation, Int. J. Intell. Syst. 16 (2001) 87–104.
[41] K.H. Yuan, W.H. Xu, W.T. Li, W.P. Ding, An incremental learning mechanism for object classification based on progressive fuzzy three-way concept, Inf. Sci. 584 (2022) 127–147.
[42] L.A. Zadeh, Fuzzy sets and information granularity, in: Advances in Fuzzy Set Theory and Applications, North-Holland, Amsterdam, 1979, pp. 3–18.
[43] L.A. Zadeh, Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, Fuzzy Sets Syst. 90 (1997) 111–127.
[44] X.Y. Zhang, D.D. Guo, W.H. Xu, Two-way concept-cognitive learning with multi-source, Cogn. Comput. (2023), https://doi.org/10.1007/s12559-023-10107-w.
[45] J.W. Zheng, W.Z. Wu, H. Bao, A.H. Tan, Evidence theory based optimal scale selection for multi-scale ordered decision systems, Int. J. Mach. Learn. Cybern. 13 (2022) 1115–1129.