



# Entropy based optimal scale selection and attribute reduction in multi-scale interval-set decision tables

Zhen-Huang Xie<sup>1,2</sup> · Wei-Zhi Wu<sup>1,2</sup> · Lei-Xi Wang<sup>1,2</sup> · Anhui Tan<sup>1,2</sup>

Received: 2 October 2023 / Accepted: 22 December 2023 / Published online: 20 February 2024  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

## Abstract

One of the current research gaps in multi-scale data analysis is studying information systems characterized by attributes with interval sets as attribute values and multiple scales. To address this gap, we first introduce the concepts of a multi-scale interval-set information system (MISIS) and a multi-scale interval-set decision table (MISDT). We then define the similarity relation between objects in an MISIS and the corresponding rough approximations. We further propose the positive region optimal scale, the modified conditional entropy optimal scale, and the positive complementary conditional entropy optimal scale in an MISDT. We examine the relationships among these optimal scales in consistent and inconsistent MISDTs and show that the positive region optimal scale and the modified conditional entropy optimal scale are equivalent in a consistent MISDT, while in an inconsistent MISDT, the positive region optimal scale is the same as the positive complementary conditional entropy optimal scale, and the modified conditional entropy optimal scale is not greater than the positive complementary conditional entropy optimal scale. Based on the optimal scale, we also develop attribute reduction approaches in MISDTs. Finally, through experimental analysis of data on the UCI dataset, we verify the effectiveness and reasonableness of our proposed methods.

**Keywords** Attribute reduction · Granular computing · Information entropy · Multi-scale information systems · Optimal scale

## 1 Introduction

Granular computing (GrC) [16, 21, 42, 43] is crucial for knowledge discovery in big data. The primary idea of GrC is to decompose large data sets or complex problems into granules through information granulation [44], and to integrate

and represent the information at different granularity levels to solve complex problems effectively.

One of the fundamental models in GrC is rough set theory, initiated by Pawlak [19], which shows promise in handling uncertain information [1, 2, 20]. This theory specifically focuses on information systems (ISs) [18], which consist of datasets composed of objects and their associated attribute values. For a target set of objects in an IS, we can represent it by a pair of operators called upper and lower approximations. We can find a similar way of expressing information in the interval sets [40]. An interval set is an extended interval number. The main idea of an interval set is to represent uncertain information by using two definite sets as upper and lower bounds. In real life, achieving an accurate representation of information or data is challenging due to the uncertainty and redundancy of the information and data volume. For example, in the course selection of students [48], we take the set of compulsory courses as the lower bound and the set of required courses and elective subjects as the upper bound. Therefore, the course selection of students constitutes an interval set. There are many similar situations. To represent ISs in which attribute values

✉ Wei-Zhi Wu  
wuwz@zjou.edu.cn

Zhen-Huang Xie  
xiezh1120@163.com

Lei-Xi Wang  
wangleixi22@163.com

Anhui Tan  
tananhui86@163.com

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

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

are interval sets, Yao and Liu [41] proposed the concept of an interval-set information system (ISIS), which has gained significant attention in recent research [10, 14, 26, 45–48].

Nevertheless, current equivalence relations used in ISISs impose on strict conditions that limit the comprehensive representation of relationships between objects. To overcome this limitation, Wang and Yue [26] proposed fuzzy dominance relations to define four uncertainty measures. They further explored the relationships among these uncertainty measures. Similarly, Zhang et al. [48] introduced similarity relations for objects in an ISIS.

In real-life situations, information is understood at various levels and depths. Objects in ISs may own different attribute values at different scales. To describe such multi-scale information, Wu and Leung [31] employed rough set methods to investigate knowledge discovery in multi-scale information systems (MISs) and multi-scale decision tables (MDTs), known as Wu–Leung model. In an MIS, there is a granular information transformation mapping to ease the transformation of object attribute values across different scales. Thus, the selection of an optimal scale is a key step for final decision in MDTs and has become a main direction in multi-scale data analysis [4, 11, 17, 25, 27, 33, 35, 49, 51]. Wu and Leung [32] defined seven types of optimal scales in an MDT and examined their relationships. These optimal scales are equivalent when an MDT is consistent, but they may be no longer equivalent in an inconsistent MDT. On the base of the Wu–Leung model, Li and Hu [9] proposed a generalized MIS, which relaxes the requirement that all attributes in MIS have the same scale level. A key step to rule acquisition in a generalized MDT is to seek for an optimal scale for each attribute to determine a single scale decision table with some requirement for final decision, which is called the optimal scale combination selection. For example, Wu and Leung [34] discussed seven optimal scale combinations in a generalized MDT. In addition, the optimal scales or optimal scale combinations based on matrix [6] and cost [50] were studied.

Information entropy [23] is an important measure to evaluate uncertainty of information. Various information entropies were used to study attribute reduction and rule acquisition in rough set data analysis [7, 8, 22, 37–39, 52]. Recently, information entropies have also been utilized to select optimal scales in multi-scale data analysis [3, 30]. Bao et al. [3] introduced a method of optimal scale combination, using Shannon entropy as the guiding principle that maintains the measurement of uncertainty in knowledge. Wang et al. [30] introduced multi-scale fuzzy entropy to integrate granular information across different scales.

With the increasing application of rough set data analysis, attribute values in data sets were emerged as sets [24], intervals [5], and fuzzy numbers [29]. Various types of attribute values were also found in multi-scale data analysis, see e.g. [12, 13, 28]. Despite the increasing interest of researchers in

multi-scale data analysis models and ISISs, the issue of attribute multi-scale problems in ISISs has received limited attention. Furthermore, the problem of entropy optimal scale selection is also worth investigating. These facts have motivated our research on knowledge discovery in multi-scale interval-set information systems (MISISs). In [36], Xie et al. introduced the notion of MISISs and defined optimal scale in multi-scale interval-set decision tables (MISDTs). The main objective of this paper is to study further the selection of entropy optimal scales and the corresponding attribute reductions in MISDTs.

In the next section, we review some basic notions related to ISISs and information entropies. In Sect. 3, we introduce MISISs with the representations of information granules as well as rough approximations in MISISs. In Sect. 4, we give some concepts with their properties in MISDTs. In Sect. 5, we discuss optimal scale selection and attribute reduction in MISDTs. In Sect. 6, we give experiments evaluating algorithms. We summarize this research in Sect. 7.

## 2 Preliminaries

In this section, we overview the fundamental concepts of interval sets, interval-set information systems and information entropies. Throughout this paper, we adopt  $C(n, m) = \frac{n!}{m!(n-m)!}$  to represent the combination number formula and  $|\cdot|$  to denote the cardinality of a set.

### 2.1 Interval sets

**Definition 1** [40] Let  $V$  be a finite set and  $2^V$  the power set of  $V$ , a closed interval set on  $2^V$  is defined as follows:

$$\mathcal{X} = [X_1, X_2] = \{X \in 2^V | X_1 \subseteq X \subseteq X_2\}, \quad (1)$$

where  $X_1$  and  $X_2$  are the lower and upper bounds of  $\mathcal{X}$ , respectively, and satisfy  $X_1 \subseteq X_2 \subseteq V$ . Note that  $[X_1, X_2]$  degenerates into an ordinary set  $X_1$  when  $X_1 = X_2$ .

In this paper, we use  $\mathcal{I}(2^V)$  to represent the set of all interval sets on the power set of a finite set  $V$ . For two interval sets  $\mathcal{X} = [X_1, X_2]$  and  $\mathcal{Y} = [Y_1, Y_2]$  on  $2^V$ , the interval-set intersection and union operations are defined, respectively, as follows [40]:

$$\mathcal{X} \cap \mathcal{Y} = [X_1 \cap Y_1, X_2 \cap Y_2] = \{X \cap Y | X \in \mathcal{X}, Y \in \mathcal{Y}\}, \quad (2)$$

$$\mathcal{X} \sqcup \mathcal{Y} = [X_1 \cup Y_1, X_2 \cup Y_2] = \{X \cup Y | X \in \mathcal{X}, Y \in \mathcal{Y}\}. \quad (3)$$

**Definition 2** [48] Let  $\mathcal{X} = [X_1, X_2]$  and  $\mathcal{Y} = [Y_1, Y_2]$  be two interval sets on the power set of a finite set  $V$ , the similarity degree of  $\mathcal{X}$  and  $\mathcal{Y}$  is defined as follows:

$$GS(\mathcal{X}, \mathcal{Y}) = \frac{1}{2}(PD_{\mathcal{X}-\mathcal{Y}} + PD_{\mathcal{Y}-\mathcal{X}}), \tag{4}$$

where

$$PD_{\mathcal{X}-\mathcal{Y}} = \frac{1}{2} \left( \frac{|X_1 \cap Y_1|}{|X_1|} + \frac{|X_2 \cap Y_2|}{|X_2|} \right). \tag{5}$$

Moreover, note that the function  $GS(\mathcal{X}, \mathcal{Y})$  is symmetric. We assume that  $\mathcal{X} \cap \mathcal{Y} = [Z_1, Z_2]$ , then Eq. (4) can be represented as follows:

$$GS(\mathcal{X}, \mathcal{Y}) = \frac{1}{4} \sum_{i=1}^2 \left( \frac{|Z_i|}{|X_i|} + \frac{|Z_i|}{|Y_i|} \right). \tag{6}$$

### 2.2 Interval-set information systems

**Definition 3** [18] An information system (IS) is a pair  $S = (O, C)$ , where  $O = \{o_1, o_2, \dots, o_n\}$  is a nonempty finite set of objects called the universe of discourse, and  $C = \{c_1, c_2, \dots, c_m\}$  a nonempty finite set of attributes such that  $c_j : O \rightarrow V_j$  for all  $c_j \in C$ , where  $V_j$  is the domain of attribute  $c_j$ .

**Definition 4** [41] An interval-set information system (ISIS) is a pair  $S = (O, C)$ , where  $O = \{o_1, o_2, \dots, o_n\}$  is the universe of discourse, and  $C = \{c_1, c_2, \dots, c_m\}$  the nonempty finite set of attributes such that  $c_j : O \rightarrow \mathcal{I}(2^{V_j})$  for all  $c_j \in C$ , i.e.  $c_j(o) = [A_j^1(o), A_j^2(o)] \in \mathcal{I}(2^{V_j})$ ,  $o \in O$ , where  $V_j$  is the domain of attribute  $c_j$ , and  $A_j^1(o) \subseteq A_j^2(o) \subseteq V_j$ .

In Definition 4,  $c_j(o) = [A_j^1(o), A_j^2(o)]$  indicates that, under attribute  $c_j$ , object  $o$  must contain  $A_j^1(o)$  and possibly  $A_j^2(o)$ . For example, Table 1 depicts an ISIS, in which  $O = \{o_1, o_2, \dots, o_8\}$  represents eight residents, with  $c_1$  and  $c_2$  show the abilities of listening and writing, respectively. The attribute values “E”, “G”, “R”, “V”, “T” mean “English”, “German”, “Russian”, “Ukrainian”, “Tibetan”, respectively. We can observe that the attribute value of  $o_3$  under  $c_1$  is  $[\{E, G\}, \{E, G, T\}]$ . The attribute value  $[\{E, G\}, \{E, G, T\}]$  indicates that in terms of the ability of listening, resident  $o_3$  has definitely mastered English and German, while possibly done Tibetan.

Let  $S = (O, C)$  be an ISIS,  $\tau \in [0, 1]$ , and  $B \subseteq C$ , a similarity relation w.r.t.  $\tau$  and  $B$  is defined as follows [48]:

$$R_B^\tau = \{(o, p) \in O \times O \mid GS(c(o), c(p)) \geq \tau, \forall c \in B\}. \tag{7}$$

Based on the similarity relation  $R_B^\tau$ , it can be inferred that the similarity class of each object  $o \in O$  w.r.t.  $\tau$  and  $B$  is denoted as [48]  $SC_B^\tau(o) = \{p \mid (o, p) \in R_B^\tau, p \in O\}$ . Additionally,

**Table 1** An interval-set information system

$O$	$c_1$	$c_2$
$o_1$	$[\{R\}, \{R, G\}]$	$[\{E\}, \{G, E\}]$
$o_2$	$[\{V\}, \{T, V\}]$	$[\{V\}, \{V, R\}]$
$o_3$	$[\{E, G\}, \{E, G, T\}]$	$[\{R, G\}, \{R, G\}]$
$o_4$	$[\emptyset, \{E\}]$	$[\{T, V\}, \{T, V\}]$
$o_5$	$[\{T\}, \{T\}]$	$[\{E\}, \{G, E\}]$
$o_6$	$[\{E, G\}, \{E, G, R\}]$	$[\{R\}, \{R, G\}]$
$o_7$	$[\{T\}, \{T, V\}]$	$[\{E\}, \{E\}]$
$o_8$	$[\{T\}, \{T\}]$	$[\{G, E\}, \{G, E\}]$

the granular structure w.r.t.  $\tau$  and  $B$  is defined as [48]  $SC_B^\tau = \{SC_B^\tau(o_1), SC_B^\tau(o_2), \dots, SC_B^\tau(o_n)\}$ .

It is obvious that  $o \in SC_B^\tau(o)$  for all  $o \in O$ , i.e.,  $\bigcup_{o \in O} SC_B^\tau(o) = O$  and  $SC_B^\tau(o) \neq \emptyset$ , so the granular structure  $SC_B^\tau$  forms a covering on  $O$ .

An interval-set decision table (ISDT) [48] is a pair  $S = (O, C \cup \{d\})$ , where  $(O, C)$  is an ISIS, and  $d \notin C$  the decision such that  $d : O \rightarrow V_d$ , where  $V_d$  is the domain of  $d$ . A binary relation on objects induced by  $d$  is defined as follows:

$$R_d = \{(o, p) \in O \times O \mid d(o) = d(p)\}. \tag{8}$$

Then, the  $R_d$  constitutes a partition  $O/R_d$  on  $O$ , written as  $O/R_d = \{R_d(o) \mid o \in O\} = \{D_1, D_2, \dots, D_r\}$ , where  $R_d(o) = \{p \mid (o, p) \in R_d, p \in O\}$ .

**Definition 5** Let  $S = (O, C \cup \{d\})$  be an ISDT and  $\tau \in [0, 1]$ , if  $SC_C^\tau(o) \subseteq R_d(o)$  for all  $o \in O$ , then we say that  $S$  is consistent. Otherwise,  $S$  is inconsistent.

**Definition 6** [48] Let  $S = (O, C \cup \{d\})$  be an ISDT,  $\tau \in [0, 1]$ , and  $B, F \subseteq C$ , if  $SC_B^\tau(o) \subseteq SC_F^\tau(o)$  for all  $o \in O$ , then we say that  $SC_B^\tau$  is finer than  $SC_F^\tau$  and mark it as  $SC_B^\tau \preceq SC_F^\tau$ .

**Proposition 1** [48] Let  $S = (O, C \cup \{d\})$  be an ISDT,  $\tau \in [0, 1]$ , and  $B, F \subseteq C$ . If  $F \subseteq B$ , then

- (1)  $R_B^\tau \subseteq R_F^\tau$ ,
- (2)  $SC_B^\tau \preceq SC_F^\tau$ .

### 2.3 Information entropies

Entropy measurement has become an important method to describe uncertainty in rough set theory. In this subsection, to better quantitatively analyze the information in ISISs, we review two types of conditional entropies.

**Definition 7** [48] Let  $S = (O, C \cup \{d\})$  be an ISDT,  $\tau \in [0, 1]$ , and  $B \subseteq C$ , a modified conditional entropy w.r.t.  $B$  and  $\tau$  is defined as follows:

$$H^\tau(d|B) = - \sum_{i=1}^n \sum_{j=1}^t \frac{|SC_B^\tau(o_i) \cap D_j|}{n^2} \log_2 \left( \frac{|SC_B^\tau(o_i) \cap D_j|}{|SC_B^\tau(o_i)|} \right), \tag{9}$$

where  $n$  and  $t$  are the cardinalities of  $O$  and  $O/R_d$ , respectively. In this paper, we define  $\frac{|SC_B^\tau(o_i) \cap D_j|}{n^2} \log_2 \left( \frac{|SC_B^\tau(o_i) \cap D_j|}{|SC_B^\tau(o_i)|} \right) = 0$  when  $|SC_B^\tau(o_i) \cap D_j| = 0$ .

According to Definition 7, the range of the modified conditional entropy is  $[0, \log_2 n]$ .

**Definition 8** [15] Let  $O$  be a nonempty finite set, with  $\mathcal{Q} = \{Q_1, Q_2, \dots, Q_{n_1}\}$  and  $\mathcal{P} = \{P_1, P_2, \dots, P_{n_2}\}$  as two partitions of  $O$ , a complementary conditional entropy of  $\mathcal{Q}$  about  $\mathcal{P}$  is defined as follows:

$$CH(\mathcal{Q}|\mathcal{P}) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \frac{|Q_i \cap P_j|}{|O|} \frac{|Q_i^c \cap P_j|}{|O|}, \tag{10}$$

where  $Q_i^c = O - Q_i = \{o \in O | o \notin Q_i\}$  is the complement of  $Q_i$  in  $O$ .

### 3 Multi-scale information systems and multi-scale interval-set information systems

In practical applications, attribute values of an object may exhibit variations across different scales. Therefore, Wu and Leung introduced MISs [31]. To facilitate the description in this paper, we assume that each attribute has  $L$  scale levels.

**Definition 9** [31] An MIS is a pair  $S = (O, C)$ , where  $O = \{o_1, o_2, \dots, o_n\}$  is the universe of discourse and  $C = \{c_1, c_2, \dots, c_m\}$  the set of attributes. Both  $O$  and  $C$  are nonempty finite sets, and each attribute in  $C$  has  $L$  scale levels. Therefore, an MIS can be represented as follows:

$$S = (O, C) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\}), \tag{11}$$

where  $c_j^l : O \rightarrow V_j^l$  and  $V_j^l$  is the domain of attribute  $c_j$  at the  $l$ -th scale. For each  $l \in \{1, 2, \dots, L - 1\}$  and  $j \in \{1, 2, \dots, m\}$ , there exists a surjective mapping  $g_j^{l,l+1} : V_j^l \rightarrow V_j^{l+1}$  such that  $c_j^{l+1} = g_j^{l,l+1} \circ c_j^l$ , i.e.

$$c_j^{l+1}(o) = g_j^{l,l+1}(c_j^l(o)), o \in O, \tag{12}$$

where  $g_j^{l,l+1}$  is called a granular information transformation mapping.

For  $l \in \{1, 2, \dots, L\}$ , we denote the attribute set under the  $l$ -th scale as  $C^l = \{c_j^l | j = 1, 2, \dots, m\}$ . Therefore, an MIS can be decomposed into  $L$  single-scale sub-ISs, denoted as  $S^l = (O, C^l), l = 1, 2, \dots, L$ .

**Definition 10** [31] A multi-scale decision table (MDT) is a pair  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$ , where  $(O, C) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\})$  is an MIS, and  $d \notin \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\}$  the decision such that  $d : O \rightarrow V_d$ , where  $V_d$  is the domain of  $d$ .

When all attribute values in an MIS are interval sets, we refer to this system as a multi-scale interval-set information system (MISIS).

**Definition 11** An MISIS is a pair  $S = (O, C)$ , where  $O = \{o_1, o_2, \dots, o_n\}$  is the universe of discourse, and  $C = \{c_1, c_2, \dots, c_m\}$  the set of attributes. Both  $O$  and  $C$  are nonempty finite sets, and each attribute in  $C$  has  $L$  scale levels, i.e.,  $C$  can be represented as  $\{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\}$  such that  $c_j^l : O \rightarrow \mathcal{I}(2^{V_j^l})$  for all  $j \in \{1, 2, \dots, m\}$  and  $l \in \{1, 2, \dots, L\}$ , i.e.  $c_j^l(o) = [A_{j,l}^1(o), A_{j,l}^2(o)]$ , where  $V_j^l$  is the domain of the attribute  $c_j$  at the  $l$ -th scale, and  $A_{j,l}^1(o) \subseteq A_{j,l}^2(o) \subseteq V_j^l$ .

For each  $l \in \{1, 2, \dots, L - 1\}$  and  $j \in \{1, 2, \dots, m\}$ , there exist two surjective mappings, one is  $h_j^{l,l+1} : V_j^l \rightarrow V_j^{l+1}$  such that

$$v_j^{l+1} = h_j^{l,l+1}(v_j^l), v_j^l \in V_j^l, \tag{13}$$

and the other is  $\psi_j^{l,l+1} : \mathcal{I}(2^{V_j^l}) \rightarrow \mathcal{I}(2^{V_j^{l+1}})$  such that  $c_j^{l+1} = \psi_j^{l,l+1} \circ c_j^l$ , i.e.

$$\begin{aligned} c_j^{l+1}(o) &= \psi_j^{l,l+1}(c_j^l(o)) \\ &= \psi_j^{l,l+1}([A_{j,l}^1(o), A_{j,l}^2(o)]) \\ &= [h_j^{l,l+1}(A_{j,l}^1(o)), h_j^{l,l+1}(A_{j,l}^2(o))] \\ &= [A_{j,l+1}^1(o), A_{j,l+1}^2(o)], o \in O, \end{aligned} \tag{14}$$

where  $h_j^{l,l+1}$  is the granular information transformation mapping and  $\psi_j^{l,l+1}$  is the domain information transformation mapping.

**Remark 1** Let  $S = (O, C) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\})$  be an MISIS. For any  $l \in \{1, 2, \dots, L - 1\}$  and  $c_j \in C$ , we employ the granular information transformation

mapping  $h_j^{l,l+1}$  to assign every element in  $V_j^l$  to its corresponding element in  $V_j^{l+1}$ , with the cardinality of  $V_j^{l+1}$  must not surpass the cardinality of  $V_j^l$ . We use the domain information transformation mapping  $\psi_j^{l,l+1}$  to map each element from attribute value  $c_j^l(o) = [A_{j,l}^1(o), A_{j,l}^2(o)]$  of object  $o$  under the  $l$ -th scale of attribute  $c_j$  to attribute value  $c_j^{l+1}(o) = [A_{j,l+1}^1(o), A_{j,l+1}^2(o)]$  under the  $(l + 1)$ -th scale. The cardinality of  $A_{j,l+1}^1(o)$  must not exceed the cardinality of  $A_{j,l}^1(o)$ , and the cardinality of  $A_{j,l+1}^2(o)$  can not be greater than the cardinality of  $A_{j,l}^2(o)$ .

For any  $l \in \{1, 2, \dots, L\}$ ,  $c_j \in C$  and  $k \leq |V_j^l|$ , the elements in  $V_j^l$  can generate  $C(|V_j^l|, k)$  different subsets with a cardinality of  $k$ , with each subset serving as an upper bound for a different interval set. According to Definition 1, there are at most  $2^k$  different lower bounds for an upper bound of cardinality  $k$ . Therefore, we conclude that

$$|\mathcal{I}(2^{V_j^l})| = \sum_{k=0}^{|V_j^l|} 2^k C(|V_j^l|, k).$$

**Proposition 2** Let  $S = (O, C) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\})$  be an MISIS. For  $l \in \{1, 2, \dots, L - 1\}$  and  $c_j \in C$ , we have

- (1)  $|V_j^{l+1}| \leq |V_j^l|$
- (2)  $|\mathcal{I}(2^{V_j^{l+1}})| \leq |\mathcal{I}(2^{V_j^l})|$ .

**Proof** (1) For any  $v^{l+1}, w^{l+1} \in V_j^{l+1}$ , if  $v^{l+1} \neq w^{l+1}$ , according to the property of surjective mappings, there exist  $v^l, w^l \in V_j^l$ , where  $v^l \neq w^l$ , such that  $v^{l+1} = h^{l,l+1}(v^l)$  and  $w^{l+1} = h^{l,l+1}(w^l)$ , therefore  $|V_j^{l+1}| \leq |V_j^l|$ .

(2) For any  $k \leq |V_j^{l+1}|$ , according to (1), we have  $2^k C(|V_j^{l+1}|, k) \leq 2^k C(|V_j^l|, k)$ , then

$$\sum_{k=0}^{|V_j^{l+1}|} 2^k C(|V_j^{l+1}|, k) \leq \sum_{k=0}^{|V_j^l|} 2^k C(|V_j^l|, k) \leq \sum_{k=0}^{|V_j^l|} 2^k C(|V_j^l|, k).$$

Therefore, we conclude that  $|\mathcal{I}(2^{V_j^{l+1}})| \leq |\mathcal{I}(2^{V_j^l})|$ . □

Like MISs, an MISIS can be decomposed into  $L$  single-scale sub-ISISs, denoted as  $S^l = (O, C^l), l = 1, 2, \dots, L$ .

For an MISIS, the similarity degree between objects may decrease as the scale becomes coarse. For example, in Table 2,  $o_3$  and  $o_6$  are two objects in  $O$ , where  $c_1^1(o_3) = [\{E, G\}, \{E, G, T\}]$ ,  $c_1^2(o_3) = [\{Wb\}, \{Wb, Tb\}]$ ,  $c_1^1(o_6) = [\{E, G\}, \{E, G, R\}]$ , and  $c_1^2(o_6) = [\{Wb\}, \{Wb, Eb\}]$ . Then we have

$$GS(c_1^1(o_3), c_1^1(o_6)) = \frac{5}{6} > \frac{3}{4} = GS(c_1^2(o_3), c_1^2(o_6)).$$

Therefore, if the threshold  $\tau$  in Eq. (7) remains unchanged, there is a possibility of smaller similarity classes as the scale becomes coarse. To address this issue, we define a similarity relation that changes with the scale.

**Proposition 3** Let  $S = (O, C) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\})$  be an MISIS and  $B \subseteq C$ . For  $l \in \{1, 2, \dots, L\}$  and  $\tau^l \in (0, 1]$ , we define

$$R_{B^l}^{\tau^l} = \{(o, p) \in O \times O | GS(c^l(o), c^l(p)) \geq \tau^l, \forall c \in B\}. \tag{15}$$

If  $\tau^{l+1} \leq \lambda^{l+1} \leq \tau^l$  for all  $l \in \{1, 2, \dots, L - 1\}$ , where

$$\lambda^{l+1} = \min \left( \min_{c \in C} \left( \min_{(o,p) \in R_{C^l}^{\tau^l}} (GS(c^{l+1}(o), c^{l+1}(p))) \right), \tau^l \right), \tag{16}$$

then we have  $R_{C^l}^{\tau^l} \subseteq R_{C^{l+1}}^{\tau^{l+1}}$ .

**Proof** For any  $(o, p) \in R_{C^l}^{\tau^l}$  and  $c \in C$ , we have

$$GS(c^{l+1}(o), c^{l+1}(p)) \geq \min_{a \in C} \left( \min_{(x,y) \in R_{C^l}^{\tau^l}} (GS(a^{l+1}(x), a^{l+1}(y))) \right),$$

i.e.,  $GS(c^{l+1}(o), c^{l+1}(p)) \geq \lambda^{l+1} \geq \tau^{l+1}$ , hence  $(o, p) \in R_{C^{l+1}}^{\tau^{l+1}}$ , therefore  $R_{C^l}^{\tau^l} \subseteq R_{C^{l+1}}^{\tau^{l+1}}$ . □

**Table 2** A multi-scale interval-set information system

$O$	$c_1^1$	$c_1^2$	$c_2^1$	$c_2^2$
$o_1$	$[\{R\}, \{R, G\}]$	$[\{Eb\}, \{Eb, Wb\}]$	$[\{E\}, \{G, E\}]$	$[\{Wb\}, \{Wb\}]$
$o_2$	$[\{V\}, \{T, V\}]$	$[\{Eb\}, \{Tb, Eb\}]$	$[\{V\}, \{V, R\}]$	$[\{Eb\}, \{Eb\}]$
$o_3$	$[\{E, G\}, \{E, G, T\}]$	$[\{Wb\}, \{Wb, Tb\}]$	$[\{R, G\}, \{R, G\}]$	$[\{Eb, Wb\}, \{Eb, Wb\}]$
$o_4$	$[\emptyset, \{E\}]$	$[\emptyset, \{Wb\}]$	$[\{T, V\}, \{T, V\}]$	$[\{Tb, Eb\}, \{Tb, Eb\}]$
$o_5$	$[\{T\}, \{T\}]$	$[\{Tb\}, \{Tb\}]$	$[\{E\}, \{G, E\}]$	$[\{Wb\}, \{Wb\}]$
$o_6$	$[\{E, G\}, \{E, G, R\}]$	$[\{Wb\}, \{Wb, Eb\}]$	$[\{R\}, \{R, G\}]$	$[\{Eb\}, \{Eb, Wb\}]$
$o_7$	$[\{T\}, \{T, V\}]$	$[\{Tb\}, \{Tb, Eb\}]$	$[\{E\}, \{E\}]$	$[\{Wb\}, \{Wb\}]$
$o_8$	$[\{T\}, \{T\}]$	$[\{Tb\}, \{Tb\}]$	$[\{G, E\}, \{G, E\}]$	$[\{Wb\}, \{Wb\}]$



**Remark 2** In Proposition 3,  $\lambda^{l+1}$  denotes the maximum value of threshold  $\tau^{l+1}$ , ensuring that the similarity relations between objects at the  $l$ -th scale hold at the  $(l + 1)$ -th scale. For example, suppose that in Table 2, the set of attributes only includes  $c_1$ , the universe of discourse only consists of objects  $o_3$  and  $o_6$ , i.e.,  $C = \{c_1\}$ ,  $O = \{o_3, o_6\}$ , and  $\tau^1 = \frac{5}{6}$ . Objects  $o_3$  and  $o_6$  have the similarity relation  $R_{C^1}^{\tau^1}$  at the 1-th scale since  $GS(c_1^1(o_3), c_1^1(o_6)) = \frac{5}{6} \geq \tau^1$ . By Eq. (16), we obtain  $\lambda^2 = \frac{3}{4}$ . If  $\tau^2 \leq \lambda^2$ , then  $o_3$  and  $o_6$  remain similarity relation  $R_{C^2}^{\tau^2}$  at the 2-th scale.

Based on the above discussion, hereafter we always assume that

- (1)  $\tau^{l+1} \leq \lambda^{l+1} \leq \tau^l$  for all  $l \in \{1, 2, \dots, L - 1\}$ ,
- (2)  $\tau^l \in (0, 1]$  for all  $l \in \{1, 2, \dots, L\}$ .

**Proposition 4** Let  $S = (O, C) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\})$  be an MISIS, for  $l \in \{1, 2, \dots, L\}$ , and  $B \subseteq C$ , we have

- (1)  $R_{B^l}^{\tau^l}$  is reflexive, i.e.,  $(o, o) \in R_{B^l}^{\tau^l}$  for all  $o \in O$ ,
- (2)  $R_{B^l}^{\tau^l}$  is symmetric, i.e., if  $(p, o) \in R_{B^l}^{\tau^l}$ , then  $(o, p) \in R_{B^l}^{\tau^l}$ .

By adjusting the threshold, we define a similarity relation that adapts to the scale change, like Eq. (7), we then give the concepts of similarity class and granular structure in an MISIS, and we can directly derive the monotonicity of these concepts with scale according to Proposition 3.

**Proposition 5** Let  $S = (O, C) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\})$  be an MISIS, for  $l \in \{1, 2, \dots, L\}$ , and  $B \subseteq C$ , we define

$$SC_{B^l}^{\tau^l}(o) = \{p | (o, p) \in R_{B^l}^{\tau^l}, p \in O\}, \forall o \in O, \tag{17}$$

$$SC_{B^l}^{\tau^l} = \{SC_{B^l}^{\tau^l}(o_1), SC_{B^l}^{\tau^l}(o_2), \dots, SC_{B^l}^{\tau^l}(o_n)\}. \tag{18}$$

Then, for  $l \in \{1, 2, \dots, L - 1\}$ , we have

- (1)  $SC_{C^l}^{\tau^l}(o) \subseteq SC_{C^{l+1}}^{\tau^{l+1}}(o)$  for all  $o \in O$ ,
- (2)  $SC_{C^l}^{\tau^l} \leq SC_{C^{l+1}}^{\tau^{l+1}}$ .

Proposition 5 establishes the monotonicity of the granular structure in an MISIS, which indicates that, as the scale increases, the covering on the universe becomes coarse.

**Definition 12** Let  $S = (O, C) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\})$  be an MISIS, for  $l \in \{1, 2, \dots, L\}$ ,  $T \subseteq O$  and  $B \subseteq C$ , the lower and upper approximations of  $T$  w.r.t.  $B^l$  and  $\tau^l$  are defined as follows:

$$\underline{R}_{B^l}^{\tau^l}(T) = \{o \in O | SC_{B^l}^{\tau^l}(o) \subseteq T\}, \tag{19}$$

$$\overline{R}_{B^l}^{\tau^l}(T) = \{o \in O | SC_{B^l}^{\tau^l}(o) \cap T \neq \emptyset\}. \tag{20}$$

**Proposition 6** Let  $S = (O, C) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\})$  be an MISIS and  $T \subseteq O$ . For  $l \in \{1, 2, \dots, L - 1\}$ , we have

- (1)  $\underline{R}_{C^{l+1}}^{\tau^{l+1}}(T) \subseteq \underline{R}_{C^l}^{\tau^l}(T) \subseteq T$ ,
- (2)  $T \subseteq \overline{R}_{C^l}^{\tau^l}(T) \subseteq \overline{R}_{C^{l+1}}^{\tau^{l+1}}(T)$ .

### 4 Multi-scale interval-set decision Tables

In this section, we introduce basic notions of multi-scale interval-set decision tables (MISDTs) and present their properties.

**Definition 13** An MISDT is a pair  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$ , where  $(O, C) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\})$  is an MISIS, and  $d \notin \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\}$  a decision with a single scale such that  $d : O \rightarrow V_d$ , where  $V_d$  is the domain of  $d$ .

**Remark 3** For an MISDT  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$ , it can be decomposed into  $L$  sub-ISDTs with the same decision, denoted as  $S^l = (O, C^l \cup \{d\})$ ,  $l = 1, 2, \dots, L$ .

**Definition 14** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be an MISDT,  $S$  is said to be consistent if  $S^l = (O, C^l \cup \{d\})$  is consistent, i.e.,  $SC_{C^l}^{\tau^l}(o) \subseteq R_d(o)$  for all  $o \in O$ . Otherwise,  $S$  is inconsistent.

**Definition 15** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be an MISDT, for  $l \in \{1, 2, \dots, L\}$ , and  $B \subseteq C$ , the positive region of  $d$  w.r.t.  $B^l$  and  $\tau^l$  is defined as follows:

$$Pos_{B^l}^{\tau^l}(O/R_d) = \bigcup_{1 \leq j \leq t} R_{B^l}^{\tau^l}(D_j), \tag{21}$$

where  $t$  is the cardinality of  $O/R_d$ .

**Proposition 7** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be an MISDT. For  $l \in \{1, 2, \dots, L - 1\}$ , we have

$$Pos_{C^{l+1}}^{\tau^{l+1}}(O/R_d) \subseteq Pos_{C^l}^{\tau^l}(O/R_d).$$

Definition 15 of the positive region is employed for the qualitative analysis in MISDTs. Meanwhile, information entropy is a quantitative analysis method widely used in the uncertainty measurement of rough sets. In what follows, we study the monotonicity of the modified conditional entropy in Definition 7 and its related properties.

**Proposition 8** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be an MISDT, for  $l \in \{1, 2, \dots, L\}$ , and  $B \subseteq C$ , we define

$$H^{\tau^l}(d|B^l) = - \sum_{i=1}^n \sum_{j=1}^t \frac{|SC_{B^l}^{\tau^l}(o_i) \cap D_j|}{n^2} \log_2 \left( \frac{|SC_{B^l}^{\tau^l}(o_i) \cap D_j|}{|SC_{B^l}^{\tau^l}(o_i)|} \right), \tag{22}$$

where  $n$  and  $t$  are the cardinalities of  $O$  and  $O/R_d$ , respectively. Then, for  $l \in \{1, 2, \dots, L - 1\}$ , we have  $H^{\tau^l}(d|C^l) \leq H^{\tau^{l+1}}(d|C^{l+1})$ .

**Proof** On the one hand, according to [48], we see that  $f(x, y) = -x \log_2 \frac{x}{x+y}$  is a monotonically increasing function

when  $x > 0$  and  $y \geq 0$ . And Eq. (22) can be expressed as follows:

$$H^{\tau^l}(d|C^l) = \sum_{i=1}^n \sum_{j=1}^t \frac{1}{n^2} f(|SC_{C^l}^{\tau^l}(o_i) \cap D_j|, |SC_{C^l}^{\tau^l}(o_i) \cap D_j^c|). \tag{23}$$

On the other hand, according to Proposition 5, for any  $o_i \in O$  and  $D_j \in O/R_d$ , we have

$$|SC_{C^l}^{\tau^l}(o_i) \cap D_j| \leq |SC_{C^{l+1}}^{\tau^{l+1}}(o_i) \cap D_j|, \tag{24}$$

and

$$|SC_{C^l}^{\tau^l}(o_i) \cap D_j^c| \leq |SC_{C^{l+1}}^{\tau^{l+1}}(o_i) \cap D_j^c|. \tag{25}$$

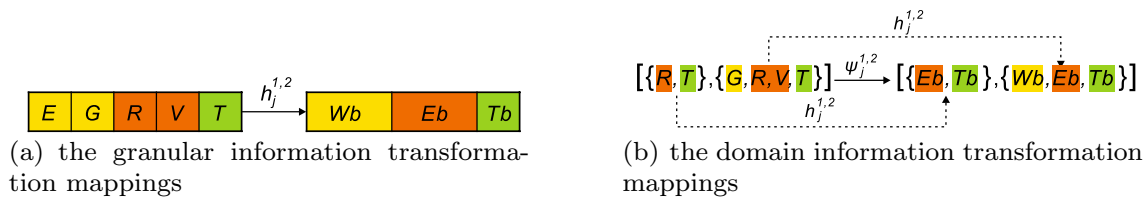
Hence  $0 \leq f(|SC_{C^l}^{\tau^l}(o_i) \cap D_j|, |SC_{C^l}^{\tau^l}(o_i) \cap D_j^c|) \leq f(|SC_{C^{l+1}}^{\tau^{l+1}}(o_i) \cap D_j|, |SC_{C^{l+1}}^{\tau^{l+1}}(o_i) \cap D_j^c|)$ . Therefore, by Eq. (23), we obtain that  $H^{\tau^l}(d|C^l) \leq H^{\tau^{l+1}}(d|C^{l+1})$ . □

**Proposition 9** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be an MISDT. For  $l_1, l_2 \in \{1, 2, \dots, L\}$ , if  $SC_{C^{l_1}}^{\tau^{l_1}} = SC_{C^{l_2}}^{\tau^{l_2}}$ , then  $H^{\tau^{l_1}}(d|C^{l_1}) = H^{\tau^{l_2}}(d|C^{l_2})$ .

**Example 1** Table 3 is an MISDT  $S = (O, C \cup \{d\})$ , in which the MISIS  $(O, C)$  is Table 2,  $O = \{o_1, o_2, \dots, o_8\}$  represents eight residents, with  $c_1$  and  $c_2$  show the abilities of listening and writing, respectively. The attribute values “E”, “G”, “R”, “V”, “T”, “Eb”, “Wb”, “Tb” mean “English”, “German”, “Russian”, “Ukrainian”, “Tibetan”, “East Slavic branch”, “West Germanic branch”, “Tibetan branch”, respectively. The granular information transformation mappings of the MISDT in Table 3 are as follows:

**Table 3** A multi-scale interval-set decision table

$O$	$c_1^1$	$c_1^2$	$c_2^1$	$c_2^2$	$d$
$o_1$	[[R], {R, G}]	[[Eb], {Eb, Wb}]	[[E], {G, E}]	[[Wb], {Wb}]	1
$o_2$	[[V], {T, V}]	[[Eb], {Tb, Eb}]	[[V], {V, R}]	[[Eb], {Eb}]	0
$o_3$	[[E, G], {E, G, T}]	[[Wb], {Wb, Tb}]	[[R, G], {R, G}]	[[Eb, Wb], {Eb, Wb}]	1
$o_4$	[∅, {E}]	[∅, {Wb}]	[[T, V], {T, V}]	[[Tb, Eb], {Tb, Eb}]	0
$o_5$	[[T], {T}]	[[Tb], {Tb}]	[[E], {G, E}]	[[Wb], {Wb}]	0
$o_6$	[[E, G], {E, G, R}]	[[Wb], {Wb, Eb}]	[[R], {R, G}]	[[Eb], {Eb, Wb}]	1
$o_7$	[[T], {T, V}]	[[Tb], {Tb, Eb}]	[[E], {E}]	[[Wb], {Wb}]	0
$o_8$	[[T], {T}]	[[Tb], {Tb}]	[[G, E], {G, E}]	[[Wb], {Wb}]	0



**Fig. 1** Conversion flowchart of the granular information transformation mappings and the domain information transformation mappings

$$h_j^{1,2}(v) = \begin{cases} Wb, & v \in \{G, E\}, \\ Eb, & v \in \{V, R\}, j = 1, 2. \\ Tb, & v \in \{T\}, \end{cases}$$

Figure 1 shows the flowcharts of granular information transformation mappings and domain information transformation mappings from the 1-th scale to the 2-th scale in Table 3, respectively, where  $j = 1, 2$ . We assume that  $\tau^1 = \tau^2 = 0.5$ . It can be calculated that

$$\begin{aligned} O/R_d &= \{\{o_1, o_3, o_6\}, \{o_2, o_4, o_5, o_7, o_8\}\}, \\ SC_{C^1}^{\tau^1} &= \{\{o_1\}, \{o_2\}, \{o_3, o_6\}, \{o_4\}, \{o_5, o_7, o_8\}, \\ &\quad \{o_3, o_6\}, \{o_5, o_7, o_8\}, \{o_5, o_7, o_8\}\}, \\ SC_{C^2}^{\tau^2} &= \{\{o_1\}, \{o_2\}, \{o_3, o_6\}, \{o_4\}, \{o_5, o_7, o_8\}, \\ &\quad \{o_3, o_6\}, \{o_5, o_7, o_8\}, \{o_5, o_7, o_8\}\}. \end{aligned}$$

And,

$$\begin{aligned} Pos_{C^1}^{\tau^1}(O/R_d) &= O, H^{\tau^1}(d|C^1) = 0, \\ Pos_{C^2}^{\tau^2}(O/R_d) &= O, H^{\tau^2}(d|C^2) = 0. \end{aligned}$$

### 5 Optimal scale selection and attribute reduction in MISDTs

In this section, we discuss optimal scale selection with attribute reduction in MISDTs. It is well known that entropy is a tool for quantitatively analyzing and measuring uncertainty in information. We aim to investigate optimal scale selection and attribute reduction from the perspective of information uncertainty and quantitative analysis. To achieve this, we introduce the concept of entropy in the process of optimal scale selection and attribute reduction. Furthermore, this section suggests choosing the optimal scale according to the positive region.

**Definition 16** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be an MISDT. For  $l \in \{1, 2, \dots, L\}$ , we say that

(1)  $S^l = (O, C^l \cup \{d\})$  is positive region consistent to  $S$  if  $Pos_{C^l}^{\tau^l}(O/R_d) = Pos_{C^1}^{\tau^1}(O/R_d)$ . If  $S^l$  is positive region consistent to  $S$  and  $S^{l+1}$  is not positive region consistent to  $S$  (if  $l + 1 \leq L$ ), then  $l$  is the positive region optimal scale (we hereafter represent it using  $l_{Pos}^{\tau^l}$ ) of  $S$ .

(2)  $S^l = (O, C^l \cup \{d\})$  is modified conditional entropy consistent to  $S$  if  $H^{\tau^l}(d|C^l) = H^{\tau^1}(d|C^1)$ . If  $S^l$  is modified conditional entropy consistent to  $S$  and  $S^{l+1}$  is not modified conditional entropy consistent to  $S$  (if  $l + 1 \leq L$ ), then  $l$  is the modified conditional entropy optimal scale (we hereafter depict it using  $l_H^{\tau^l}$ ) of  $S$ .

**Remark 4** According to Definition 16, we explain the optimal scale as follows:

(1) The positive region optimal scale  $l_{Pos}^{\tau^l}$  is the coarsest scale to ensure that the positive region of  $S^{l_{Pos}^{\tau^l}}$  is the same as the positive region of  $S^1$ .

(2) The modified conditional entropy optimal scale  $l_H^{\tau^l}$  is the coarsest scale which guarantees that the modified conditional entropy w.r.t.  $\tau^{l_H^{\tau^l}}$  and  $C^{l_H^{\tau^l}}$  is equal to the modified conditional entropy w.r.t.  $\tau^1$  and  $C^1$ .

#### 5.1 Optimal scale selection in consistent MISDTs

**Proposition 10** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be an MISDT. For  $l \in \{1, 2, \dots, L\}$ ,  $H^{\tau^l}(d|C^l) = 0$  iff  $S^l$  is consistent.

**Proof** “ $\Rightarrow$ ” Assume that  $H^{\tau^l}(d|C^l) = 0$ . By Proof of Proposition 8, for any  $o \in O$  and  $D \in O/R_d$ , we have  $f(|SC_{C^l}^{\tau^l}(o) \cap D|, |SC_{C^l}^{\tau^l}(o) \cap D^c|) = 0$ . Hence, there are two, and only two outcomes, one is  $|SC_{C^l}^{\tau^l}(o) \cap D| = 0$  and  $|SC_{C^l}^{\tau^l}(o) \cap D^c| = |SC_{C^l}^{\tau^l}(o)|$ , and the other is



$|SC_{C^l}^{\tau^l}(o) \cap D| = |SC_{C^l}^{\tau^l}(o)|$  and  $|SC_{C^l}^{\tau^l}(o) \cap D^c| = 0$ . It follows that  $SC_{C^l}^{\tau^l}(o) \subseteq R_d(o)$  for all  $o \in O$ , i.e.,  $S^l$  is consistent.

“ $\Leftarrow$ ” It is a direct consequence of Definition 5 and Proof of Proposition 8.  $\square$

**Proposition 11** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be an MISDT. For  $l \in \{1, 2, \dots, L\}$ , we have

$$H^{\tau^l}(d|C^l) = 0 \Leftrightarrow \overline{R}_{C^l}^{\tau^l}(D) - \underline{R}_{C^l}^{\tau^l}(D) = \emptyset, \forall D \in O/R_d.$$

**Proof** “ $\Rightarrow$ ” Assume that  $H^{\tau^l}(d|C^l) = 0$ . For any  $D \in O/R_d$  and  $o \in \overline{R}_{C^l}^{\tau^l}(D)$ , according to Proof of Proposition 10, we have  $SC_{C^l}^{\tau^l}(o) \subseteq R_d(o) = D$ , i.e.,  $o \in \underline{R}_{C^l}^{\tau^l}(D)$ , which means  $\overline{R}_{C^l}^{\tau^l}(D) \subseteq \underline{R}_{C^l}^{\tau^l}(D)$ . By Proposition 6, we have  $\overline{R}_{C^l}^{\tau^l}(D) = D = \underline{R}_{C^l}^{\tau^l}(D)$ . Therefore, we conclude that  $\overline{R}_{C^l}^{\tau^l}(D) - \underline{R}_{C^l}^{\tau^l}(D) = \emptyset$ .

“ $\Leftarrow$ ” Assume that  $\overline{R}_{C^l}^{\tau^l}(D) - \underline{R}_{C^l}^{\tau^l}(D) = \emptyset$  for all  $D \in O/R_d$ .

For any  $o \in O$ , according to Proposition 6, there exists  $D \in O/R_d$  such that  $R_d(o) = D = \overline{R}_{C^l}^{\tau^l}(D)$ . Since  $\overline{R}_{C^l}^{\tau^l}(D) = D$ , for any  $D_1 \in O/R_d - \{D\}$ , we have  $o \notin D_1 = \overline{R}_{C^l}^{\tau^l}(D_1)$ , i.e.,  $SC_{C^l}^{\tau^l}(o) \cap D_1 = \emptyset$ , hence  $SC_{C^l}^{\tau^l}(o) \subseteq D = R_d(o)$ , which means  $S^l$  is consistent, thus  $H^{\tau^l}(d|C^l) = 0$ .  $\square$

**Proposition 12** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be an MISDT. For  $l \in \{1, 2, \dots, L\}$ , we have

$$H^{\tau^l}(d|C^l) = 0 \Leftrightarrow Pos_{C^l}^{\tau^l}(O/R_d) = O.$$

By Proposition 12, we can obtain Theorem 1.

**Theorem 1** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be a consistent MISDT, we have

$$l_{Pos}^{\tau^l} = l_H^{\tau^l}. \tag{26}$$

**Algorithm 1** Calculate the modified conditional entropy optimal scale in an MISDT

---

```

Input: An MISDT  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$ .
Output: The modified conditional entropy optimal scale  $l_H^{\tau^l}$ .
1:  $l \leftarrow 1$ ;
2:  $\tau^1 \leftarrow 0.5$ ;
3: Compute  $O/R_d$ ;
4: Compute  $SC_{C^1}^{\tau^1}$ ;
5: Compute  $H^{\tau^1}(d|C^1)$ ;
6:  $Temp \leftarrow H^{\tau^1}(d|C^1)$ ;
7: while  $Temp = H^{\tau^1}(d|C^1)$  do
8:    $l \leftarrow l + 1$ ;
9:   if  $l = L + 1$  then
10:     break;
11:   end if
12:   Compute  $\lambda^l$ ;
13:    $\tau^l \leftarrow \lambda^l$ ;
14:   Compute  $SC_{C^l}^{\tau^l}$ ;
15:   Compute  $H^{\tau^l}(d|C^l)$ ;
16:    $Temp \leftarrow H^{\tau^l}(d|C^l)$ ;
17: end while
18:  $l_H^{\tau^l} \leftarrow l - 1$ ;
19: return  $l_H^{\tau^l}$ .

```

---

Algorithm 1 is to search for the modified conditional entropy optimal scale in an MISDT. The time complexity (TC) of calculating both  $SC_{C^l}^{\tau^l}$  and  $O/R_d$  in Algorithm 1 is  $O(n^2)$ . The TC of step 5 is  $O(n^2)$  because each object itself forms a decision class and a similarity class in the worst case. The worst case for step 12 is all similarity classes are the universe itself. Therefore, the TC of step 12 is  $O(mn^2)$ . Additionally, in the worst case, the modified conditional entropy optimal scale is the coarsest. Consequently,  $O(Lmn^2)$  is the overall TC of Algorithm 1.

**Example 2** In Example 1, it is easy to observe that  $S^1$  is consistent. It can be calculated that  $Pos_{C^1}^{\tau^1}(O/R_d) = Pos_{C^2}^{\tau^2}(O/R_d)$  and  $H^{\tau^1}(d|C^1) = H^{\tau^2}(d|C^2)$ , we then conclude that  $I_{Pos}^{\tau^1} = I_H^{\tau^1} = 2$ .

### 5.2 Optimal scale selection in inconsistent MISDTs

In this subsection, we discuss the optimal scales in an inconsistent MISDT. Firstly, by strengthening the conditions of Proposition 9, we obtain Proposition 13.

**Proposition 13** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be an inconsistent MISDT, for  $l_1, l_2 \in \{1, 2, \dots, L\}$  and  $l_1 \leq l_2$ , we have

(1)  $H^{\tau^{l_1}}(d|C^{l_1}) = H^{\tau^{l_2}}(d|C^{l_2})$  iff  $SC_{C^{l_1}}^{\tau^{l_1}}(o) = SC_{C^{l_2}}^{\tau^{l_2}}(o)$  for all  $o \in \{p \in O | p \notin Pos_{C^{l_2}}^{\tau^{l_2}}(O/R_d)\}$ ,

(2) If  $H^{\tau^{l_1}}(d|C^{l_1}) = H^{\tau^{l_2}}(d|C^{l_2})$ , then  $Pos_{C^{l_1}}^{\tau^{l_1}}(O/R_d) = Pos_{C^{l_2}}^{\tau^{l_2}}(O/R_d)$ .

Proposition 13 shows that, in an inconsistent MISDT, if  $S^l$  is modified conditional entropy consistent to  $S$ , then  $S^l$  must be positive region consistent to  $S$ . Therefore, we obtain Theorem 2.

**Theorem 2** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be an inconsistent MISDT, we have

$$I_H^{\tau^l} \leq I_{Pos}^{\tau^l}. \tag{27}$$

According to Theorem 2, the modified conditional entropy optimal scale in an inconsistent MISDT is not greater than the positive region optimal scale. For this reason, we propose a new entropy based on Eq. (10) to quantify the amount of information in each scale and select a coarser scale to simplify the complexity.

**Definition 17** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be an MISDT and  $B \subseteq C$ . For  $l \in \{1, 2, \dots, L\}$ , the positive complementary conditional entropy is defined as follows:

$$PCH^{\tau^l}(d|B^l) = \frac{1}{th} \sum_{o \in Pos_{C^l}^{\tau^l}(O/R_d)} \sum_{j=1}^t \frac{|SC_{B^l}^{\tau^l}(o) \cap D_j|}{|D_j|} \frac{|SC_{B^l}^{\tau^l}(o) \cap D_j^c|}{|D_j^c|}, \tag{28}$$

where  $t$  and  $h$  are, respectively, the cardinalities of  $O/R_d$  and  $Pos_{C^l}^{\tau^l}(O/R_d)$ . Additionally, we define  $PCH^{\tau^l}(d|B^l) = 0$  when  $|O/R_d| = 1$  and specify  $PCH^{\tau^l}(d|B^l) = 1$  when  $|Pos_{C^l}^{\tau^l}(O/R_d)| = 0$ . Note that in the following we always assume that the positive region of  $d$  w.r.t.  $C^l$  and  $\tau^l$  is a nonempty set.

**Proposition 14** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be an MISDT, for  $l \in \{1, 2, \dots, L-1\}$ , we have

$$PCH^{\tau^l}(d|C^l) \leq PCH^{\tau^{l+1}}(d|C^{l+1}).$$

**Proof** Let  $g(x, y) = xy$  be a monotonically increasing function for  $x \geq 0$  and  $y \geq 0$ . In this case, Eq. (28) can be expressed as follows:

$$PCH^{\tau^l}(d|C^l) = \frac{1}{th} \sum_{o \in Pos_{C^l}^{\tau^l}(O/R_d)} \sum_{j=1}^t \frac{g(|SC_{C^l}^{\tau^l}(o) \cap D_j|, |SC_{C^l}^{\tau^l}(o) \cap D_j^c|)}{|D_j| |D_j^c|}. \tag{29}$$

Then, for any  $o \in Pos_{C^l}^{\tau^l}(O/R_d)$  and  $D \in O/R_d$ , Eqs. (24) and (25) hold. According to the monotonicity of  $g(x, y)$ , we have

$$g(|SC_{C^l}^{\tau^l}(o) \cap D|, |SC_{C^l}^{\tau^l}(o) \cap D^c|) \leq g(|SC_{C^{l+1}}^{\tau^{l+1}}(o) \cap D|, |SC_{C^{l+1}}^{\tau^{l+1}}(o) \cap D^c|).$$

Therefore, we conclude that  $PCH^{\tau^l}(d|C^l) \leq PCH^{\tau^{l+1}}(d|C^{l+1})$ .  $\square$

**Proposition 15** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be an MISDT. For  $l \in \{1, 2, \dots, L\}$ , if  $S^l$  is consistent, then  $PCH^{\tau^l}(d|C^l) = 0$ .

**Proof** Since  $S^l$  is consistent, by Propositions 7, 10, and 12, we have  $Pos_{C^l}^{\tau^l}(O/R_d) = Pos_{C^l}^{\tau^l}(O/R_d) = O$ .

Hence, for any  $o \in O$  and  $D \in O/R_d$ , we have  $g(|SC_{C^l}^{\tau^l}(o) \cap D|, |SC_{C^l}^{\tau^l}(o) \cap D^c|) = 0$ . Thus, by Eq. (29), we conclude that  $PCH^{\tau^l}(d|C^l) = 0$ .  $\square$

**Proposition 16** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be an inconsistent MISDT, for  $l_1, l_2 \in \{1, 2, \dots, L\}$  and  $l_1 \leq l_2$ , if  $PCH^{\tau^{l_1}}(d|C^{l_1}) = PCH^{\tau^{l_2}}(d|C^{l_2})$ , then  $Pos_{C^{l_1}}^{\tau^{l_1}}(O/R_d) = Pos_{C^{l_2}}^{\tau^{l_2}}(O/R_d)$ .

**Proof** Since  $PCH^{\tau^{l_1}}(d|C^{l_1}) = PCH^{\tau^{l_2}}(d|C^{l_2})$ , on the one hand, for any  $o \in Pos_{C^{l_1}}^{\tau^{l_1}}(O/R_d)$ , by Proof of Proposition 14, we have  $g(|SC_{C^{l_2}}^{\tau^{l_2}}(o) \cap D|, |SC_{C^{l_2}}^{\tau^{l_2}}(o) \cap D^c|) = g(|SC_{C^{l_1}}^{\tau^{l_1}}(o) \cap D|, |SC_{C^{l_1}}^{\tau^{l_1}}(o) \cap D^c|) = 0$  for all  $D \in O/R_d$ , i.e.,  $SC_{C^{l_2}}^{\tau^{l_2}}(o) \subseteq R_d(o)$ , which means  $o \in Pos_{C^{l_2}}^{\tau^{l_2}}(O/R_d)$ , hence  $Pos_{C^{l_1}}^{\tau^{l_1}}(O/R_d) \subseteq Pos_{C^{l_2}}^{\tau^{l_2}}(O/R_d)$ . On the other hand, by Proposition 7, we have  $Pos_{C^{l_2}}^{\tau^{l_2}}(O/R_d) \subseteq Pos_{C^{l_1}}^{\tau^{l_1}}(O/R_d)$ . Thus, we conclude that  $Pos_{C^{l_1}}^{\tau^{l_1}}(O/R_d) = Pos_{C^{l_2}}^{\tau^{l_2}}(O/R_d)$ .  $\square$

**Definition 18** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be an MISDT. For  $l \in \{1, 2, \dots, L\}$ , if  $PCH^{\tau^l}(d|C^l) = PCH^{\tau^1}(d|C^1)$ , then  $S^l = (U, C^l \cup \{d\})$  is said to be positive complementary conditional entropy consistent to  $S$ . If  $S^l$  is positive complementary conditional entropy consistent to  $S$  and  $S^{l+1}$  is not positive complementary conditional entropy consistent to  $S$  (if  $l + 1 \leq L$ ), then  $l$  is called the positive complementary conditional entropy optimal scale (we hereafter represent it using  $l_{PCH}^{\tau^l}$ ) of  $S$ .

**Remark 5** Based on Definition 18, we obtain that the positive complementary conditional entropy optimal scale  $l_{PCH}^{\tau^l}$  is the coarsest scale to ensure that the positive complementary conditional entropy w.r.t.  $\tau_{PCH}^{\tau^l}$  and  $C_{PCH}^{\tau^l}$  is equal to the positive complementary conditional entropy w.r.t.  $\tau^1$  and  $C^1$ .

**Proposition 17** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be an inconsistent MISDT. If  $l \in \{1, 2, \dots, l_{PCH}^{\tau^1}\}$ , then  $PCH^{\tau^l}(d|C^l) = 0$ .

**Proof** For  $l \in \{1, 2, \dots, l_{PCH}^{\tau^1}\}$ , according to Definition 18 and Proposition 14, we have  $PCH^{\tau^1}(d|C^1) = PCH^{\tau^l}(d|C^l)$ . Then, by Proof of Proposition 15, we have  $g(|SC_{C^1}^{\tau^1}(o) \cap D|, |SC_{C^1}^{\tau^1}(o) \cap D^c|) = 0$  for all  $o \in Pos_{C^1}^{\tau^1}(O/R_d)$  and  $D \in O/R_d$ , which means  $PCH^{\tau^1}(d|C^1) = 0$ . Therefore,  $PCH^{\tau^l}(d|C^l) = 0$ .  $\square$

**Proposition 18** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be an inconsistent MISDT. If  $l \in \{1, 2, \dots, l_{PCH}^{\tau^1}\}$ , then  $PCH^{\tau^l}(d|C^l) \leq H^{\tau^l}(d|C^l)$ .

**Proof** It is evident.  $\square$

By using Propositions 12, 13, and 16, we can obtain the relationship between  $H^{\tau^l}(d|C^l)$  and  $PCH^{\tau^l}(d|C^l)$  as Proposition 19.

**Proposition 19** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be an MISDT. For  $l, l_1, l_2 \in \{1, 2, \dots, L\}$ , we have

(1) If  $S$  is consistent, then  $H^{\tau^l}(d|C^l) = 0$  iff  $PCH^{\tau^l}(d|C^l) = 0$ ,

(2) If  $H^{\tau^{l_1}}(d|C^{l_1}) = H^{\tau^{l_2}}(d|C^{l_2})$ , then  $PCH^{\tau^{l_1}}(d|C^{l_1}) = PCH^{\tau^{l_2}}(d|C^{l_2})$ ,

(3) If  $l_1 < l_2$  and  $0 < PCH^{\tau^{l_1}}(d|C^{l_1}) < PCH^{\tau^{l_2}}(d|C^{l_2})$ , then  $H^{\tau^{l_1}}(d|C^{l_1}) < H^{\tau^{l_2}}(d|C^{l_2})$ .

**Proof** (1) Assume that  $S$  is consistent, according to Definition 14 and Proof of Proposition 15, we obtain  $Pos_{C^l}^{\tau^l}(O/R_d) = O$  and  $PCH^{\tau^l}(d|C^l) = 0$ . If  $PCH^{\tau^l}(d|C^l) = 0$ , then Proposition 16 implies that  $Pos_{C^l}^{\tau^l}(O/R_d) = O$ . As a result,  $S^l$  is consistent, and it follows that  $H^{\tau^l}(d|C^l) = 0$ . Conversely, if  $H^{\tau^l}(d|C^l) = 0$ , then by Proposition 12, we have  $Pos_{C^l}^{\tau^l}(O/R_d) = O$ . Therefore, based on Proof of Proposition 15, we deduce that  $PCH^{\tau^l}(d|C^l) = 0$ .

(2) It is direct consequence of Proposition 13 and Proof of Proposition 14.

(3) Assume that  $0 < PCH^{\tau^{l_1}}(d|C^{l_1}) < PCH^{\tau^{l_2}}(d|C^{l_2})$ . Based on Proof of Proposition 14, there exists at least one  $o \in U$  such that  $SC_{C^{l_1}}^{\tau^{l_1}}(o) \subset SC_{C^{l_2}}^{\tau^{l_2}}(o)$  and  $o \notin Pos_{C^{l_2}}^{\tau^{l_2}}(O/R_d)$ . Therefore, by Proposition 13, we have  $H^{\tau^{l_1}}(d|C^{l_1}) < H^{\tau^{l_2}}(d|C^{l_2})$ .  $\square$

**Theorem 3** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be an inconsistent MISDT, we have

$$l_{PCH}^{\tau^l} = l_{Pos}^{\tau^l}. \tag{30}$$

**Proof** Suppose  $l = l_{PCH}^{\tau^l}$ , we have  $PCH^{\tau^1}(d|C^1) = PCH^{\tau^l}(d|C^l)$ . By Proposition 16, we obtain  $Pos_{C^1}^{\tau^1}(O/R_d) = Pos_{C^l}^{\tau^l}(O/R_d)$ , hence  $l_{PCH}^{\tau^1} \leq l_{Pos}^{\tau^1}$ . Now, we are to prove  $l_{Pos}^{\tau^1} \leq l_{PCH}^{\tau^1}$ . Suppose  $l = l_{Pos}^{\tau^1}$ , we see that  $Pos_{C^1}^{\tau^1}(O/R_d) = Pos_{C^l}^{\tau^l}(O/R_d)$ . Then, by Proof of Proposition 15, we observe that  $PCH^{\tau^1}(d|C^1) = 0 = PCH^{\tau^l}(d|C^l)$ , thus  $l_{Pos}^{\tau^1} \leq l_{PCH}^{\tau^1}$ . Therefore,  $l_{Pos}^{\tau^1} = l_{PCH}^{\tau^1}$ .  $\square$

**Table 4** An inconsistent multi-scale interval-set decision table

$O$	$c_1^1$	$c_1^2$	$c_2^1$	$c_2^2$	$d$
$o_1$	[[ $R$ ], { $R, V$ }]	[[ $Eb$ ], { $Eb$ }]	[[ $R$ ], { $R, V$ }]	[[ $Eb$ ], { $Eb$ }]	1
$o_2$	[[ $E, G$ ], { $E, G$ }]	[[ $Wb$ ], { $Wb$ }]	[[ $E, G$ ], { $E, G$ }]	[[ $Wb$ ], { $Wb$ }]	3
$o_3$	[[ $R$ ], { $R, E$ }]	[[ $Eb$ ], { $Wb, Eb$ }]	[[ $R, T$ ], { $R, E, T$ }]	[[ $Eb, Tb$ ], { $Wb, Eb, Tb$ }]	3
$o_4$	[[ $E, V$ ], { $E, V$ }]	[[ $Wb, Eb$ ], { $Wb, Eb$ }]	[[ $E, V$ ], { $E, V$ }]	[[ $Wb, Eb$ ], { $Wb, Eb$ }]	1
$o_5$	[[ $G, V$ ], { $G, V, E$ }]	[[ $Wb, Eb$ ], { $Wb, Eb$ }]	[[ $G, V$ ], { $G, V, E$ }]	[[ $Wb, Eb$ ], { $Wb, Eb$ }]	3
$o_6$	[[ $T$ ], { $T$ }]	[[ $Tb$ ], { $Tb$ }]	[[ $V, T$ ], { $V, T$ }]	[[ $Eb, Tb$ ], { $Eb, Tb$ }]	2
$o_7$	[[ $G, R$ ], { $G, R$ }]	[[ $Wb, Eb$ ], { $Wb, Eb$ }]	[[ $G, R$ ], { $G, R$ }]	[[ $Wb, Eb$ ], { $Wb, Eb$ }]	1
$o_8$	[[ $E$ ], { $E, G$ }]	[[ $Wb$ ], { $Wb$ }]	[[ $E, V$ ], { $E, G, V$ }]	[[ $Wb, Eb$ ], { $Wb, Eb$ }]	2

**Remark 6** According to Theorems 2 and 3, we conclude that  $l_H^l \leq l_{Pos}^l = l_{PCH}^l$  in an inconsistent MISDT.

**Algorithm 2** Calculate the positive complementary conditional entropy optimal scale in an MISDT

Algorithm 2 is utilized to seek for the positive complementary conditional entropy optimal scale when the positive region under the finest scale is a nonempty set. In step 6, the TC of identifying the positive region is  $O(n^2)$ . The TC of steps 7 and 18 is  $O(n^2)$  because the universe is equal to the positive region in the worst case. Consequently, the overall TC is  $O(Lmn^2)$ .

---

**Input:** An MISDT  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$ .

**Output:** The positive complementary conditional entropy optimal scale  $l_{PCH}^l$ .

```

1:  $Temp1 \leftarrow 0$ ;
2:  $l \leftarrow 1$ ;
3:  $\tau^1 \leftarrow 0.5$ ;
4: Compute  $O/R_d$ ;
5: Compute  $SC_{C^1}^{\tau^1}$ ;
6: Compute  $Pos_{C^1}^{\tau^1}(O/R_d)$ ;
7: Compute  $PCH^{\tau^1}(d|C^1)$ ;
8: while  $Temp1 = PCH^{\tau^1}(d|C^1)$  do
9:    $l \leftarrow l + 1$ ;
10:  if  $l = L + 1$  then
11:    break;
12:  end if
13:  Compute  $\lambda^l$ ;
14:   $\tau^l \leftarrow \lambda^l$ ;
15:  for  $o \in Pos_{C^1}^{\tau^1}(O/R_d)$  do
16:    Compute  $SC_{C^l}^{\tau^l}(o)$ ;
17:  end for
18:  Compute  $PCH^{\tau^l}(d|C^l)$ ;
19:   $Temp1 \leftarrow PCH^{\tau^l}(d|C^l)$ ;
20: end while
21:  $l_{PCH}^l \leftarrow l - 1$ ;
22: return  $l_{PCH}^l$ .

```

---

**Example 3** Table 4 depicts an inconsistent MISDT. We assume that  $\tau^1 = \tau^2 = 0.5$ . It can be calculated that

$$\begin{aligned} O/R_d &= \{\{o_1, o_4, o_7\}, \{o_2, o_3, o_5\}, \{o_6, o_8\}\}, \\ SC_{C^1}^{\tau^1} &= \{\{o_1, o_3, o_7\}, \{o_2, o_4, o_5, o_7, o_8\}, \{o_1, o_3\}, \\ &\quad \{o_2, o_4, o_5, o_8\}, \{o_2, o_4, o_5\}, \{o_6\}, \{o_1, o_2, o_7\}, \{o_2, o_4, o_8\}\}, \\ SC_{C^2}^{\tau^2} &= \{\{o_1, o_3, o_4, o_5, o_7\}, \{o_2, o_4, o_5, o_7, o_8\}, \{o_1, o_3, o_4, o_5, o_7\}, \\ &\quad O - \{o_6\}, O - \{o_6\}, \{o_6\}, O - \{o_6\}, \{o_2, o_4, o_5, o_7, o_8\}\}. \end{aligned}$$

And,

$$\begin{aligned} Pos_{C^1}^{\tau^1}(O/R_d) &= \{o_6\}, H^{\tau^1}(d|C^1) = 0.45, PCH^{\tau^1}(d|C^1) = 0, \\ Pos_{C^2}^{\tau^2}(O/R_d) &= \{o_6\}, H^{\tau^2}(d|C^2) = 0.86, PCH^{\tau^2}(d|C^2) = 0. \end{aligned}$$

Since  $Pos_{C^2}^{\tau^2}(O/R_d) = Pos_{C^1}^{\tau^1}(O/R_d)$  and  $PCH^{\tau^2}(d|C^2) = PCH^{\tau^1}(d|C^1)$ , we obtain that  $l_{Pos}^{\tau^1} = l_{PCH}^{\tau^1} = 2$ . Subsequently, notice that  $H^{\tau^2}(d|C^2) \neq H^{\tau^1}(d|C^1)$ , we conclude that  $l_H^{\tau^1} = 1$ . Thus,  $l_H^{\tau^1} \leq l_{Pos}^{\tau^1} = l_{PCH}^{\tau^1}$ .

### 5.3 Attribute reduction in MISDTs

In [48], the reduction of single-scale decision tables based on modified conditional entropy was discussed. In this subsection, based on the positive region optimal scale and the positive complementary conditional entropy optimal scale, we study attribute reduction in MISDTs.

**Definition 19** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be an MISDT and  $\emptyset \neq B \subseteq C$ . For the positive complementary conditional entropy optimal scale  $l_{PCH}^{\tau^l}$  and the positive region optimal scale  $l_{Pos}^{\tau^l}$ , we say that

(1)  $B$  is an  $l_{PCH}^{\tau^l}$ -reduct if  $PCH^{\tau^l, l_{PCH}}(d|C^{l_{PCH}}) = PCH^{\tau^l, l_{PCH}}(d|B^{l_{PCH}})$  and  $PCH^{\tau^l, l_{PCH}}(d|C^{l_{PCH}}) \neq PCH^{\tau^l, l_{PCH}}(d|\{B - c\}^{l_{PCH}})$  for all  $c \in B$ .

(2)  $B$  is an  $l_{Pos}^{\tau^l}$ -reduct if  $Pos_{C^{l_{Pos}}}^{\tau^l, l_{Pos}}(O/R_d) = Pos_{B^{l_{Pos}}}^{\tau^l, l_{Pos}}(O/R_d)$  and  $Pos_{C^{l_{Pos}}}^{\tau^l, l_{Pos}}(O/R_d) \neq Pos_{\{B - c\}^{l_{Pos}}}^{\tau^l, l_{Pos}}(O/R_d)$  for all  $c \in B$ .

**Proposition 20** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be an MISDT. For  $B \subseteq C$  and  $l \in \{1, 2, \dots, L\}$ , we have

$$(1) Pos_{B^l}^{\tau^l}(O/R_d) \subseteq Pos_{C^l}^{\tau^l}(O/R_d),$$

$$(2) PCH^{\tau^l}(d|C^l) \leq PCH^{\tau^l}(d|B^l).$$

**Proof** (1) It follows immediately from Proposition 1 and Definition 15.

(2) For any  $o \in Pos_{C^l}^{\tau^l}(O/R_d)$ , and  $R_d(o) = D \in O/R_d$ , according to Propositions 1 and 7, we have  $SC_{C^l}^{\tau^l}(o) \subseteq SC_{B^l}^{\tau^l}(o)$ , and  $Pos_{B^l}^{\tau^l}(O/R_d) \subseteq Pos_{C^l}^{\tau^l}(O/R_d) \subseteq Pos_{C^l}^{\tau^l}(O/R_d)$ . When  $o \in Pos_{C^l}^{\tau^l}(O/R_d) - Pos_{B^l}^{\tau^l}(O/R_d)$ , we have  $SC_{C^l}^{\tau^l}(o) \not\subseteq D$ , which means

$$\begin{aligned} 0 &< g(|SC_{C^l}^{\tau^l}(o) \cap D|, |SC_{C^l}^{\tau^l}(o) \cap D^c|) \\ &\leq g(|SC_{B^l}^{\tau^l}(o) \cap D|, |SC_{B^l}^{\tau^l}(o) \cap D^c|). \end{aligned}$$

When  $o \in Pos_{C^l}^{\tau^l}(O/R_d) - Pos_{B^l}^{\tau^l}(O/R_d)$ , we have  $SC_{C^l}^{\tau^l}(o) \subseteq D$

$$\begin{aligned} 0 &= g(|SC_{C^l}^{\tau^l}(o) \cap D|, |SC_{C^l}^{\tau^l}(o) \cap D^c|) \\ &< g(|SC_{B^l}^{\tau^l}(o) \cap D|, |SC_{B^l}^{\tau^l}(o) \cap D^c|). \end{aligned}$$

When  $o \in Pos_{B^l}^{\tau^l}(O/R_d)$ , we have  $SC_{B^l}^{\tau^l}(o) \subseteq D$ , which shows that

$$\begin{aligned} 0 &= g(|SC_{C^l}^{\tau^l}(o) \cap D|, |SC_{C^l}^{\tau^l}(o) \cap D^c|) \\ &= g(|SC_{B^l}^{\tau^l}(o) \cap D|, |SC_{B^l}^{\tau^l}(o) \cap D^c|). \end{aligned}$$

Therefore, by Proof of Proposition 14, we have  $PCH^{\tau^l}(d|C^l) \leq PCH^{\tau^l}(d|B^l)$ .  $\square$

Propositions 21 and 22 can be derived similar to Propositions 15, 16, and 17.

**Proposition 21** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be an MISDT. For  $B \subseteq C$  and  $l \in \{1, 2, \dots, L\}$ , if  $PCH^{\tau^l}(d|C^l) = PCH^{\tau^l}(d|B^l)$ , then  $Pos_{C^l}^{\tau^l}(O/R_d) = Pos_{B^l}^{\tau^l}(O/R_d)$ .

**Proposition 22** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be an MISDT. For  $B \subseteq C$  and  $l \in \{1, 2, \dots, L\}$ , if  $Pos_{C^l}^{\tau^l}(O/R_d) = Pos_{B^l}^{\tau^l}(O/R_d) \neq \emptyset$ , then  $PCH^{\tau^l}(d|B^l) = 0$ .

**Proposition 23** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be an MISDT. For  $l_1, l_2 \in \{1, 2, \dots, L\}$  and  $A, B \subseteq C$ , if  $SC_{B^{l_1}}^{\tau^{l_1}}(o) = O$  for all  $o \in Pos_{C^{l_1}}^{\tau^{l_1}}(O/R_d)$ , then  $PCH^{\tau^{l_2}}(d|A^{l_2}) \leq PCH^{\tau^{l_1}}(d|B^{l_1})$ .

**Table 5** A description of data sets

ID	Data sets	Samples	Attributes	Decision classes
1	Breast cancer	286	9	2
2	Breast tissue	106	9	6
3	Ecoli	336	7	8
4	Fertility	100	9	2
5	Glass identification	214	9	7
6	Iris	150	4	3
7	Wholesale customers (region)	440	6	3
8	Vertebral column	310	6	3

**Proof** Since  $SC_{B^1}^{\tau^1}(o) = O$  for all  $o \in Pos_{C^1}^{\tau^1}(O/R_d)$ , we obtain that  $|SC_{A^2}^{\tau^2}(o) \cap D^c| \leq |D^c| = |SC_{B^1}^{\tau^1}(o) \cap D^c|$  and  $|SC_{A^2}^{\tau^2}(o) \cap D| \leq |D| = |SC_{B^1}^{\tau^1}(o) \cap D|$  for all  $D \in O/R_d$ .

Then, by Proof of Proposition 14, we conclude that  $PCH^{\tau^2}(d|A^2) \leq PCH^{\tau^1}(d|B^1) = 1$ . □

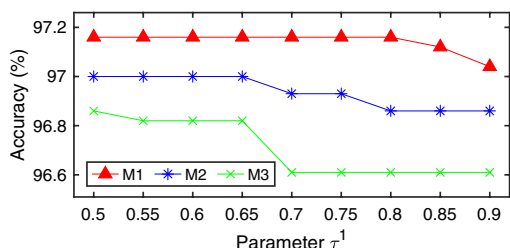
**Remark 7** By Propositions 22 and 23, we see that the domain of positive complementary conditional entropy is  $[0, 1]$ .

According to Propositions 16, 17, 21, and 22, we can obtain the relationship between an  $l_{Pos}^{\tau^1}$ -reduct and an  $l_{PCH}^{\tau^1}$ -reduct as Theorem 4.

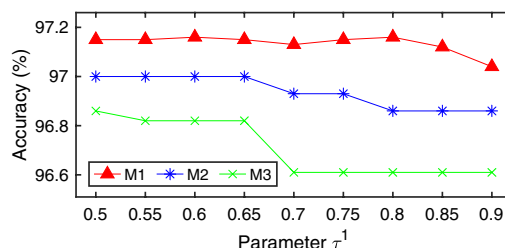
**Theorem 4** Let  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$  be an MISDT. For  $B \subseteq C$ ,  $B$  is an  $l_{Pos}^{\tau^1}$ -reduct iff  $B$  is an  $l_{PCH}^{\tau^1}$ -reduct.

**Table 6** A description of  $k$  value selection for data sets

Data sets	Breast cancer	Breast tissue	Ecoli	Fertility	Glass identification	Iris	Wholesale customers (region)	Vertebral column
$k$	4	7	5	5	5	6	14	7



(a) the modified conditional entropy



(b) the positive complementary conditional entropy

**Fig. 2** A comparison of the performance of optimal scales obtained from three similarity benchmarks in the dataset Breast cancer



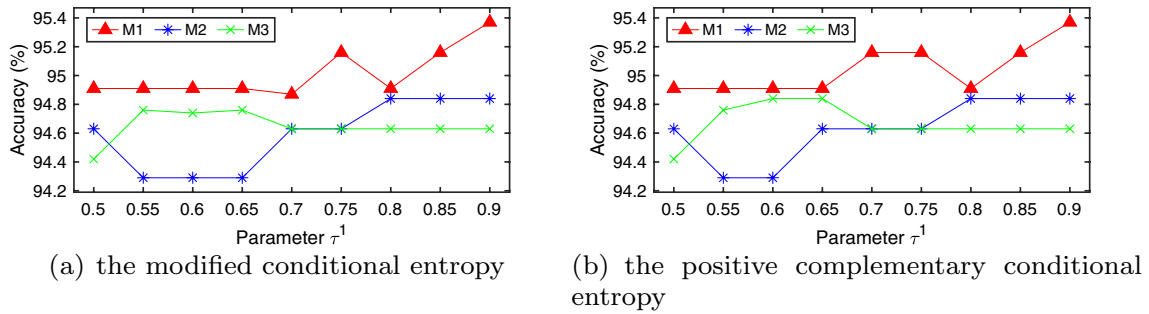


Fig. 3 A comparison of the performance of optimal scales obtained from three similarity benchmarks in the dataset Breast tissue

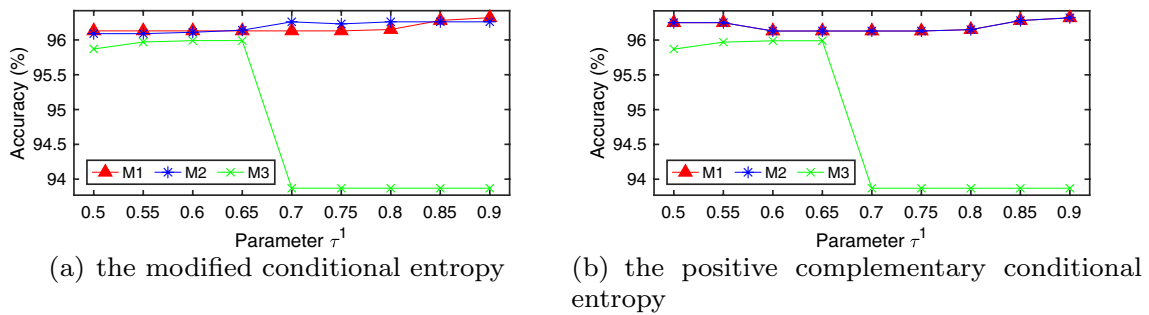


Fig. 4 A comparison of the performance of optimal scales obtained from three similarity benchmarks in the dataset Ecoli

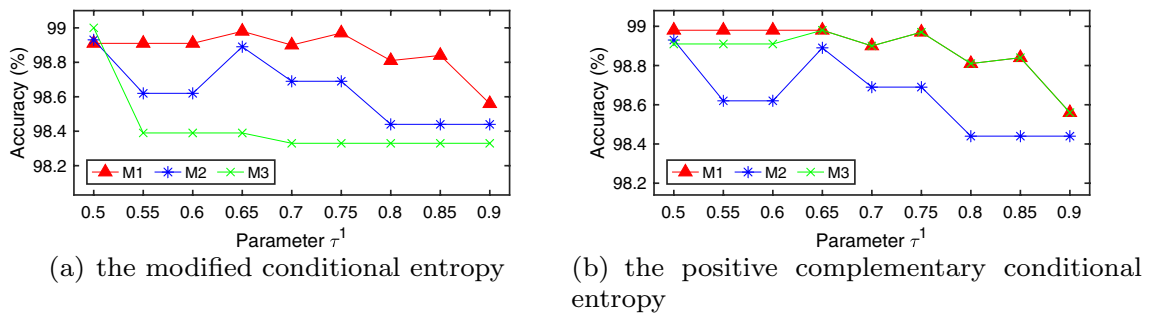


Fig. 5 A comparison of the performance of optimal scales obtained from three similarity benchmarks in the dataset Fertility

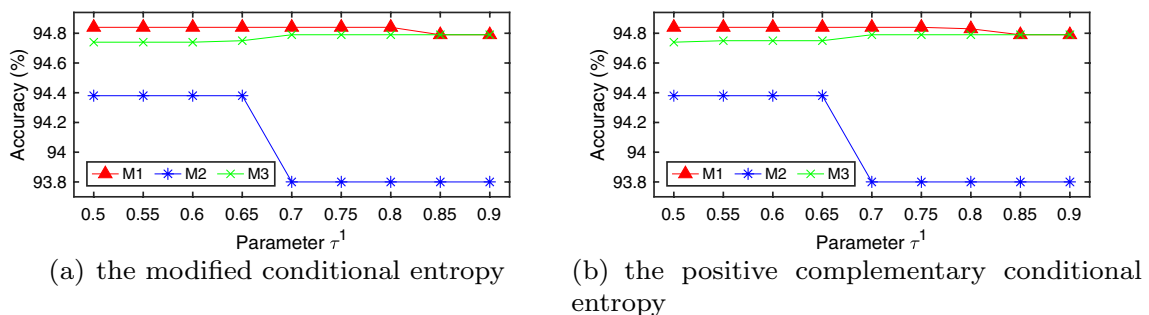


Fig. 6 A comparison of the performance of optimal scales obtained from three similarity benchmarks in the dataset Glass identification

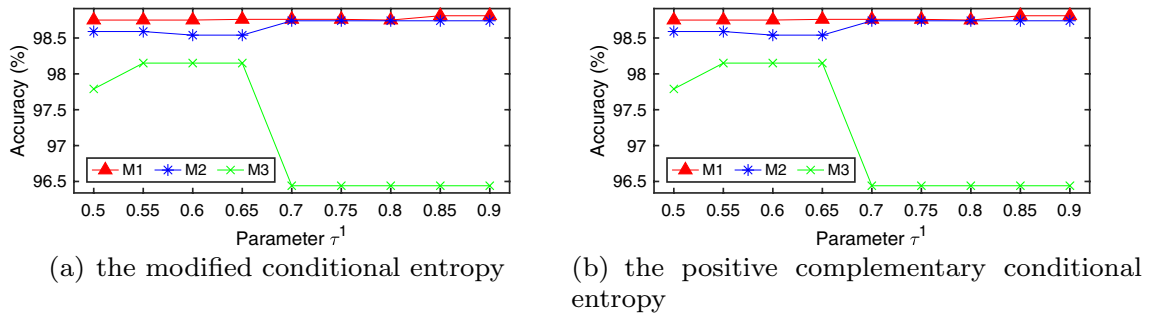


Fig. 7 A comparison of the performance of optimal scales obtained from three similarity benchmarks in the dataset Iris

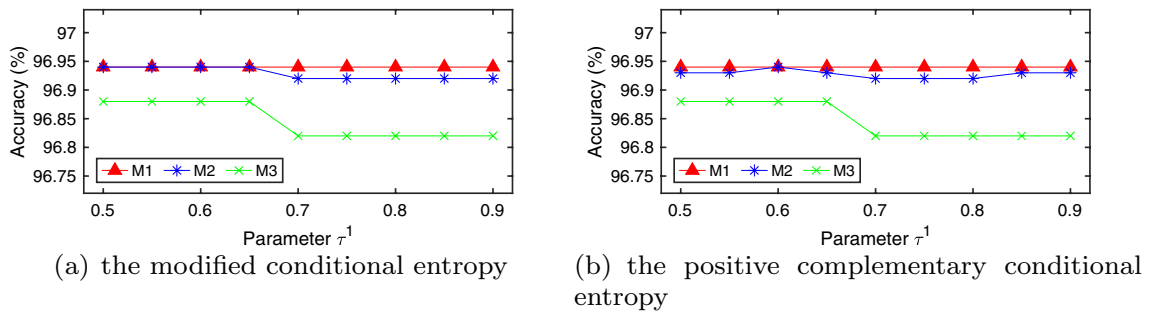


Fig. 8 A comparison of the performance of optimal scales obtained from three similarity benchmarks in the dataset Wholesale customers (region)

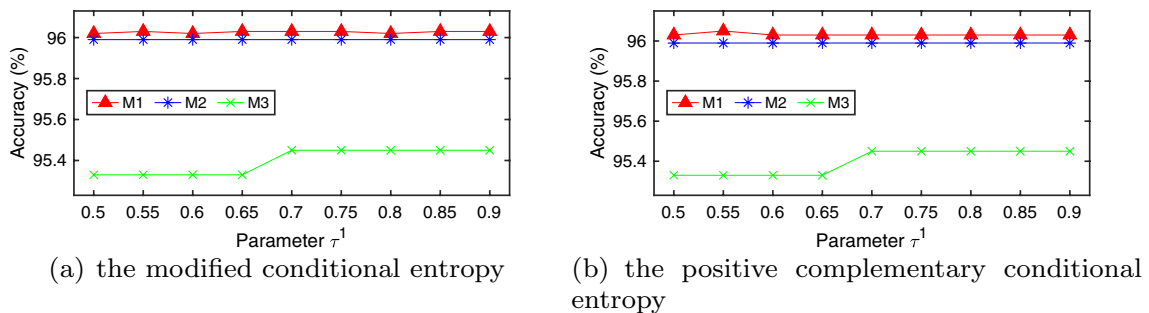


Fig. 9 A comparison of the performance of optimal scales obtained from three similarity benchmarks in the dataset Vertebral column

**Table 7** A comparison of classification performance at various scales

Data sets	$l_{Fin}$	$l_H^l$	$l_{PCH}^l$	$l_{Coa}$
Breast cancer	<b>97.16±3.69</b>	<b>97.16±3.69</b>	97.15±3.7	96.69±4.23
Breast tissue	94.84±2.33	<b>94.91±2.36</b>	<b>94.91±2.36</b>	94.84±3.07
Ecoil	96.13±3.93	96.13±3.93	<b>96.25±3.83</b>	93.48±4.18
Fertility	98.89±1.31	98.91±1.26	<b>98.98±1.1</b>	98.56±1.32
Glass identification	<b>94.84±3.17</b>	<b>94.84±3.17</b>	<b>94.84±3.17</b>	93.85±3.28
Iris	<b>98.75±0.93</b>	<b>98.75±0.93</b>	<b>98.75±0.93</b>	96±4.06
Wholesale customers (region)	<b>96.94±3.97</b>	<b>96.94±3.97</b>	<b>96.94±3.97</b>	96.92±3.98
Vertebral column	96.02±2.09	96.02±2.09	<b>96.03±2.09</b>	95.52±3.36
Average	96.7±2.68	96.71±2.68	<b>96.73±2.64</b>	95.73±3.44

**Table 8** A description of the optimal scale of the data sets

Data sets	Breast cancer	Breast tissue	Ecoli	Fertility	Glass identification	Iris	Wholesale customers (region)	Vertebral column
$l_H^l$	1±0	2.2±0.4	1±0	2.3±0.64	1±0	1±0	2.2±0.4	1±0
$l_{PCH}^l$	1.2±0.4	2.2±0.4	1.6±0.49	3.5±0.81	1±0	1.1±0.3	2.2±0.4	1.3±0.46

**Algorithm 3** Calculate an  $l_{PCH}^l$ -reduct in an MISDT

---

**Input:** An MISDT  $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, \dots, m, l = 1, 2, \dots, L\} \cup \{d\})$ .

**Output:** An  $l_{PCH}^l$ -reduct.

- 1:  $l \leftarrow l_{PCH}^l$ ;
- 2:  $A \leftarrow \emptyset$ ;
- 3:  $B \leftarrow \{c_j^l | j = 1, 2, \dots, m\}$ ;
- 4: **Compute**  $PCH^{\tau^l}(d|B^l)$ ;
- 5:  $temp2 \leftarrow PCH^{\tau^l}(d|B^l)$ ;
- 6: **while**  $PCH^{\tau^l}(d|B^l) = temp2$  **do**
- 7:      $A \leftarrow B$ ;
- 8:      $B \leftarrow B - \{c_j^l\}$ ;
- 9:     **if**  $B = \emptyset$  **then**
- 10:         break;
- 11:     **end if**
- 12:     **for**  $o \in Pos_{C_1}^{\tau^l}(O/R_d)$  **do**
- 13:         **Compute**  $SC_{B^l}^{\tau^l}(o)$ ;
- 14:     **end for**
- 15:     **Compute**  $PCH^{\tau^l}(d|B^l)$ ;
- 16: **end while**
- 17:  $l_{PCH}^l\text{-reduct} \leftarrow A$ ;
- 18: **return**  $l_{PCH}^l\text{-reduct}$ .

---

Compared to  $l_{Pos}^l$ -reduct, the advantage of  $l_{PCH}^l$ -reduct lies in its ability to measure the information content of a dataset as a whole and represent the overall information content of the data with a single numerical value. Algorithm 3 is used to find an  $l_{PCH}^l$ -reduct. The TC of steps 12 to 14 is  $O(n^2m)$ . Additionally, in the worst case,

each attribute forms an  $l_{PCH}^l$ -reduct. Therefore, the overall TC of Algorithm 3 is  $O(n^2m^2)$ .

**Example 4** In Example 3, we have  $l_{Pos}^l = l_{PCH}^l = 2$ . It can be calculated that

**Table 9** A description of the  $I_{PCH}^r$ -reduct classification performance

Data sets	$I_{PCH}^r$ -reduct
Breast cancer	97.79±3.77
Breast tissue	94.2±2.26
Ecoil	96.25±3.83
Fertility	98.79±1.38
Glass identification	94.73±2.43
Iris	96.94±3.97
Wholesale customers (region)	96.54±4.09
Vertebral column	96.73±2.64
Average	96.5±3.05

$$SC_{\{c_1\}^2}^{\tau^2} = \{ \{o_1, o_3, o_4, o_5, o_7\}, \{o_2, o_4, o_5, o_7, o_8\}, \{o_1, o_3, o_4, o_5, o_7\}, O - \{o_6\}, O - \{o_6\}, \{o_6\}, O - \{o_6\}, \{o_2, o_4, o_5, o_7, o_8\} \},$$

$$SC_{\{c_2\}^2}^{\tau^2} = \{ O - \{o_2\}, \{o_2, o_4, o_5, o_7, o_8\}, O - \{o_2\}, O, O, O - \{o_2\}, O, O \}.$$

And,

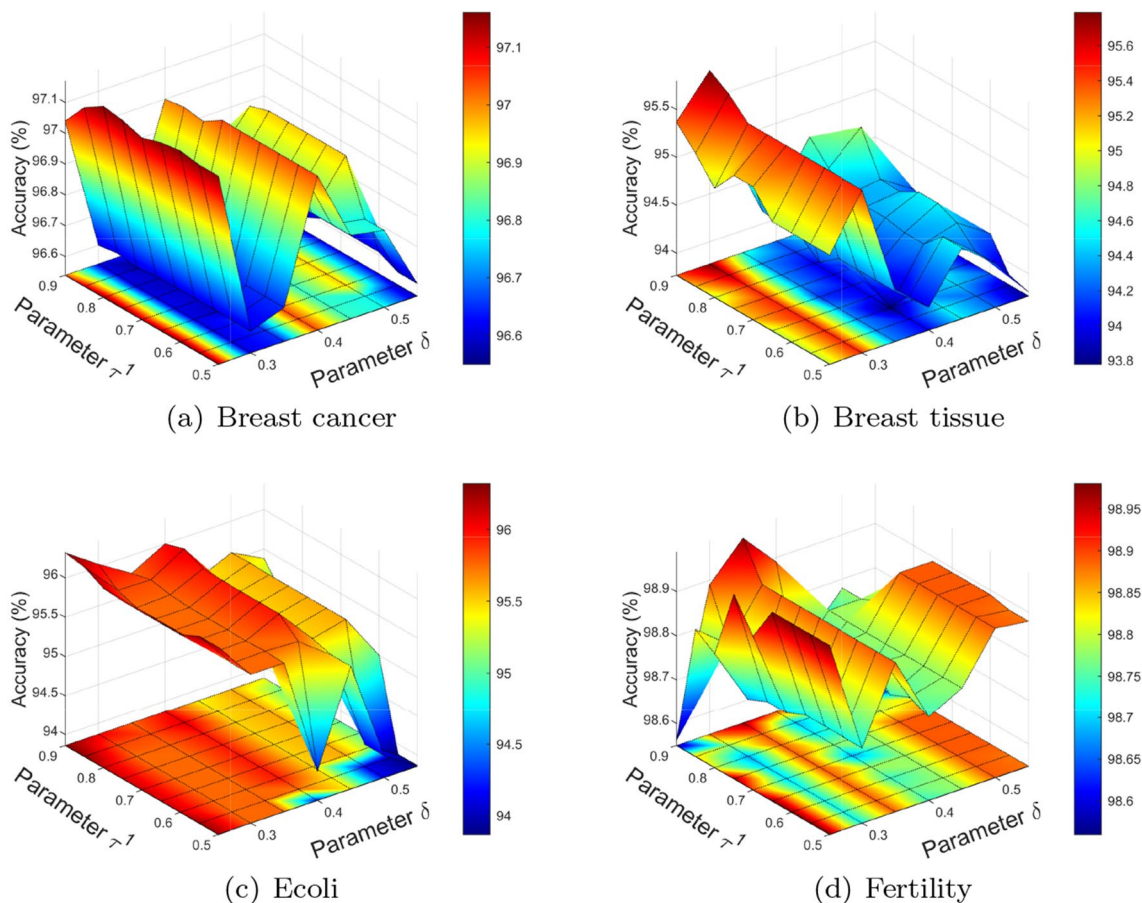
$$Pos_{\{c_1\}^2}^{\tau^2}(O/R_d) = \{o_6\}, PCH^{\tau^2}(d|\{c_1\}^2) = 0,$$

$$Pos_{\{c_2\}^2}^{\tau^2}(O/R_d) = \emptyset, PCH^{\tau^2}(d|\{c_2\}^2) = 0.77.$$

Since  $Pos_{C^2}^{\tau^2}(O/R_d) = Pos_{\{c_1\}^2}^{\tau^2}(O/R_d)$  and  $PCH^{\tau^2}(d|\{c_1\}^2) = PCH^{\tau^2}(d|C^2)$ , we conclude that  $\{c_1\}$  is both an  $I_{Pos}^r$ -reduct and an  $I_{PCH}^r$ -reduct.

### 6 Experiments

In this section, we validate the feasibility and effectiveness of the proposed MISDT model, as well as its corresponding optimal scale selection algorithm and attribute reduction algorithm with experiments. All experimental hardware setups are configured as Windows 10, Intel(R) Core(TM) i7-6700HQ CPU @ 2.60GHz, and 8GB memory.



**Fig. 10** The impact of different parameter settings on the optimal scale performance

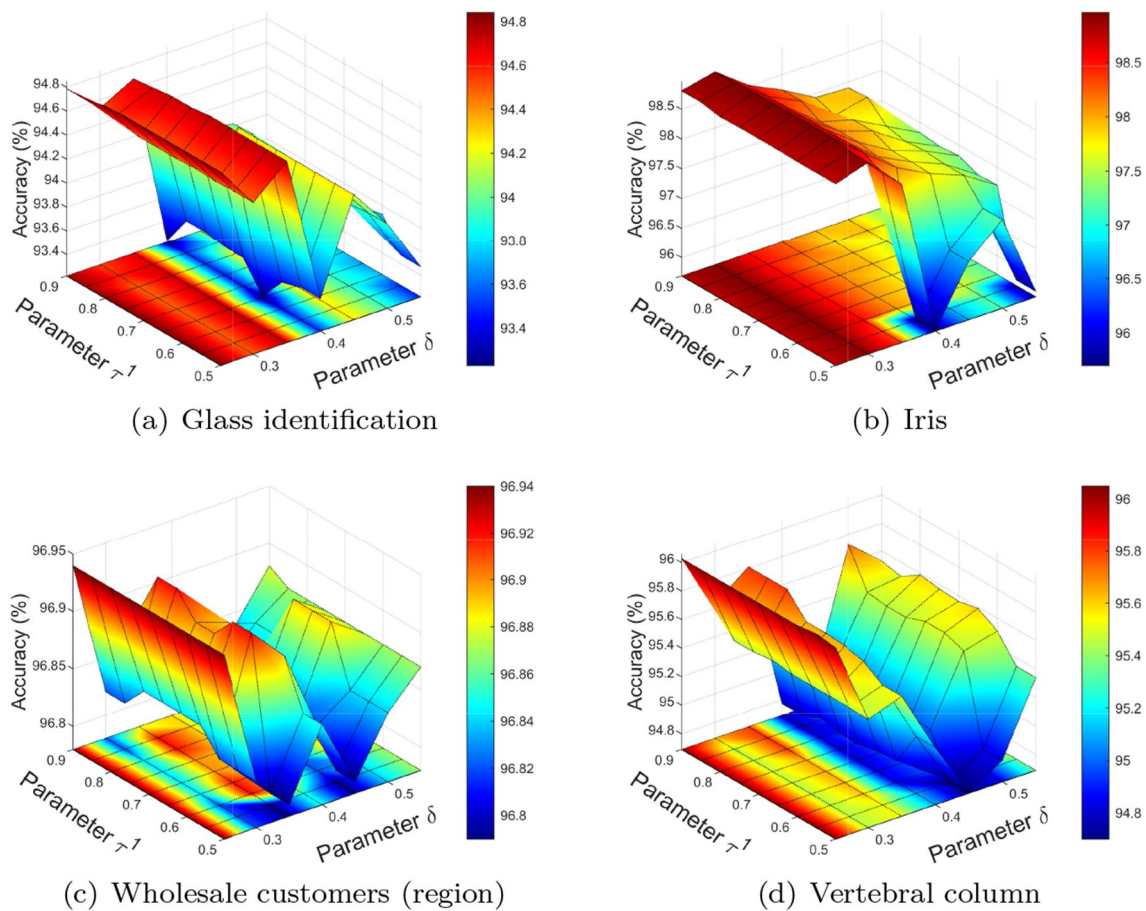


Fig. 11 The impact of different parameter settings on the optimal scale performance

The software environment for executing the algorithm is MATLAB R2019a.

The experimental data is obtained from 8 datasets in the UCI database, as shown in Table 5. Since no existing MISDTs are available in the datasets, it is necessary to preprocess the data and construct the corresponding MISDTs.

The construction of MISDTs is based on the data construction methods analogized in [9] and [47]. The specific construction processes are as follows:

**step 1.** If an object’s attribute value is a semantic data value, it must be converted to an integer value.

**step 2.** Following the construction method of MDTs in [9], we determine the domain of any attribute under any scale. In this experiment, we first standardize all attribute values of the initial ISs, and we then adjust the cardinality to 8 within the domain of each attribute at the finest scale. As the scale becomes coarse by one unit, the cardinality of the domain of each attribute decreases by one unit until they reach 3, completing the construction of domains of each attribute under each scale. Hence, in this experiment, the

scales of each conditional attribute in all the datasets are uniformly set to 6.

**step 3.** Following the construction method of ISDTs in [47], we construct an ISIS for each IS that has been standardized at the finest scale. In construction, a fixed number of missing values are randomly assigned to an original IS to create interval sets. Here, the parameter  $\delta$  represents the extent of missing data, where, for example,  $\delta = 0.25$  indicates a 25% loss of items  $c(o)$  in the original information system.

**step 4.** We achieve the interval set transformation at each scale using the domain information transformation mapping  $\psi_j^{l,l+1}$  defined in Definition 11, which results in an MISDT. Moreover, we describe the parameter  $\tau^l$  as the threshold in the similarity relation  $R_{C_l}^{\tau^l}$  at the finest scale.

Our experiments utilize Algorithms 1 and 2 to obtain the modified conditional entropy optimal scale and the positive complementary conditional entropy optimal scale, respectively. Then, we obtain an  $I_{PCH}^{\tau^l}$ -reduct through Algorithm 3. In this experiment, we employ the ten-fold cross-validation

method and the K-Nearest Neighbor (KNN) classifier to evaluate the classification performance of two optimal scales and  $I_{PCH}^l$ -reduct. To obtain the best classification accuracy in the KNN classifier, we select the  $k$  values for each dataset, as shown in Table 6. Additionally, Eq. (4) is employed to construct the distance function  $Dis(\mathcal{X}, \mathcal{Y})$  between any two interval sets  $\mathcal{X}$  and  $\mathcal{Y}$  on the power set of a finite set  $V$ . The specific expression of the distance function  $Dis(\mathcal{X}, \mathcal{Y})$  is as follows:

$$Dis(\mathcal{X}, \mathcal{Y}) = 1 - GS(\mathcal{X}, \mathcal{Y}). \quad (31)$$

### 6.1 Similarity measures comparison experiments

In [48], the authors compared the effects of different similarity measures on the classification performance of attribute reduction. However, there needs to be more research on the impact of distinct similarity measures on the classification performance of the optimal scale in MISDTs. To address this gap, we compare the performance of entropy optimal scales with three similarity measures. These similarity measures are derived from Eq. (4), [14] and [46]. For any two interval sets  $\mathcal{X} = [X_1, X_2]$  and  $\mathcal{Y} = [Y_1, Y_2]$  on the power set of a finite set  $V$ , the similarity degree in [46] is denoted as follows:

$$GS(\mathcal{X}, \mathcal{Y}) = \frac{|X_1 \cap Y_1|}{2|X_1 \cup Y_1|} + \frac{|X_2 \cap Y_2|}{2|X_2 \cup Y_2|}, \quad (32)$$

while the similarity degree in [14] is described as follows:

$$GS(\mathcal{X}, \mathcal{Y}) = \frac{1}{3} \sum_{i=1}^3 \frac{|U_i(\mathcal{X}) \cap U_i(\mathcal{Y})|}{|U_i(\mathcal{X}) \cup U_i(\mathcal{Y})|}, \quad (33)$$

where  $U_1(\mathcal{X}) = X_1$ ,  $U_2(\mathcal{X}) = X_2 - X_1$ , and  $U_3(\mathcal{X}) = O - X_2$ .

Additionally, we provide specific experimental results depicted in Figs. 2, 3, 4, 5, 6, 7, 8, 9, in which the parameter  $\tau^1 \in \{0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9\}$  and the parameter  $\delta = 0.25$ . In the legends of Figs. 2, 3, 4, 5, 6, 7, 8, 9, M1, M2, and M3 indicate the classification accuracy achieved by the optimal scale obtained via formulas Eqs. (4), (32), and (33) as similarity measures, respectively. Figures 2, 3, 4, 5, 6, 7, 8, 9 demonstrate that using the similarity in Eq. (4) as a benchmark, the positive complementary conditional entropy optimal scale yields the highest classification accuracy for most datasets. Similarly, the performance evaluation of the modified conditional entropy optimal scale also leads to the same conclusion. Figures 2, 3, 4, 5, 6, 7, 8, 9 suggest that the similarity formula used in this paper can better evaluate the relationship between different interval sets, thus we obtain an appropriate scale for knowledge discovery

and rule extraction. In addition, in Figs. 6 and 7, the classification accuracy based on the optimal scales of two entropies is the same, which indicates positive complementary conditional entropy simplifies information complexity while ensuring classification capability.

### 6.2 Optimal scale performance experiments

For each dataset, we compare the classification accuracies under the finest scale  $l_{Fin}$ , the coarsest scale  $l_{Coa}$ , the modified conditional entropy optimal scale  $l_H^l$ , and the positive complementary conditional entropy optimal scale  $l_{PCH}^l$ .

Experimental results are shown in Table 7, where the parameters used are  $\delta = 0.25$  and  $\tau^1 = 0.5$ . From Table 7, the proposed optimal scales perform well on most datasets. Among the eight datasets, the positive complementary conditional entropy optimal scale achieves the highest classification accuracy in seven datasets. In comparison, the modified conditional entropy optimal scale achieves the highest classification accuracy in five datasets. Experimental results indicate that the optimal scale balances the fitting effect on training data and the generalization ability, resulting in higher classification accuracy on test datasets.

Additionally, under the conditions of  $\delta = 0.25$  and  $\tau^1 = 0.5$ , Table 8 illustrates the average optimal scale obtained through ten-fold cross-validation, while Table 9 shows the classification accuracy achieved through an  $l_{PCH}^l$ -reduct.

### 6.3 Parameter evaluation experiments

Finally, we discuss the impact of parameters  $\delta$  and  $\tau^1$  on the performance of the positive complementary conditional entropy optimal scale. Figures 10 and 11 describe the effect of parameter variations. The x-axis represents parameter  $\tau^1$ , ranging from 0.5 to 0.9. The y-axis indicates parameter  $\delta$ , ranging from 0.25 to 0.55. Furthermore, the z-axis expresses the classification accuracy obtained using the KNN classifier. From Figs. 10 and 11, we can observe that parameter  $\delta$  significantly impacts the classification accuracy across all datasets. However, parameter  $\tau^1$  minimizes classification accuracy for specific datasets, as shown in Fig. 10a. Moreover, different combinations of parameter values result in varying classification accuracy. For instance, in Fig. 11a, the classification accuracy reaches its minimum value of 93.23% when  $\delta = 0.4$  and  $\tau^1 = 0.65$ , and the classification accuracy reaches its maximum value of 95.79% when  $\delta = 0.3$  and  $\tau^1 = 0.9$  in Fig. 10b.



## 7 Conclusion

Multi-scale data analysis has attracted significant attention in the research of GrC. However, the study of information systems that involve multiple scales and interval sets of attribute values still needs to be explored. In this paper, we have studied optimal scale selection and attribute reduction in MISDTs. We have first introduced the concept of an MISIS, and have defined a scale-adaptive similarity relation for each attribute set with each scale. In consistent MISDTs, we have proposed two optimal scale selection methods based on the positive region and modified conditional entropy, and examined their equivalence through rigorous proof. For inconsistent MISDTs, we have defined a new entropy measure called positive complementary conditional entropy, established the positive complementary conditional entropy optimal scale, and investigated its relationship with those above two optimal scales. Based on an optimal scale, we have also developed attribute reduction methods in MISDTs. Finally, we have conducted experiments on our proposed algorithms to verify three points, i.e., the reasonableness of the similarity formula used in this paper, the advantages of the effectiveness of the proposed optimal scale, and the productivity of the presented algorithm.

Our future research will consider optimal scale selection by integrating various measures of uncertainty and will focus on rule extraction in MISDTs. Moreover, we will aim to investigate knowledge acquisition in generalized MISDTs and incomplete MISDTs.

**Acknowledgements** This work was supported by grants from the National Natural Science Foundation of China (grant numbers 12371466, 61976194 and 62076221).

**Author Contributions** ZX: data curation, methodology, validation, writing-original draft. WW: validation, writing-review & editing. LW: investigation, software, visualization. AT: data curation, formal analysis.

**Data availability** Enquiries about data availability should be forwarded to the authors.

## Declarations

**Conflict of interest** The authors declare no competing interests.

## References

- Al-shami TM (2021) An improvement of rough sets' accuracy measure using containment neighborhoods with a medical application. *Inform Sci* 569:110–124
- An S, Guo XY, Wang CZ, Guo G, Dai JH (2023) A soft neighborhood rough set model and its applications. *Inform Sci* 624:185–199
- Bao H, Wu WZ, Zheng JW, Li TJ (2021) Entropy based optimal scale combination selection for generalized multi-scale information tables. *Int J Mach Learn Cybernet* 12:1427–1437
- Chen YS, Li JH, Li JJ, Chen DX, Lin RD (2023) Sequential 3WD-based local optimal scale selection in dynamic multi-scale decision information systems. *Int J Approx Reason* 152:221–235
- Dai JH, Wang ZY, Huang WY (2023) Interval-valued fuzzy discernibility pair approach for attribute reduction in incomplete interval-valued information systems. *Inform Sci* 642:119215
- Huang XF, Zhan JM, Xu ZS, Fujita H (2023) A prospect-regret theory-based three-way decision model with intuitionistic fuzzy numbers under incomplete multi-scale decision information systems. *Expert Syst Appl* 214:119144
- Huang YY, Guo KJ, Yi XW, Li Z, Li TR (2022) Matrix representation of the conditional entropy for incremental feature selection on multi-source data. *Inform Sci* 591:263–286
- Ji X, Li J, Yao S, Zhao P (2023) Attribute reduction based on fusion information entropy. *Int J Approx Reason* 160:108949
- Li F, Hu BQ (2017) A new approach of optimal scale selection to multi-scale decision tables. *Inform Sci* 381:193–208
- Li HX, Wang MH, Zhou XZ, Zhao JB (2012) An interval set model for learning rules from incomplete information table. *Int J Approx Reason* 53:24–37
- Li JH, Feng Y (2023) Update of optimal scale in dynamic multi-scale decision information systems. *Int J Approx Reason* 152:310–324
- Li RK, Yang JL, Zhang XY (2023) Optimal scale selection based on three-way decisions with decision-theoretic rough sets in multi-scale set-valued decision tables. *Int J Mach Learn Cybernet* 14:3719–3736
- Li WK, Li JJ, Huang JX, Dai WZ, Zhang XP (2021) A new rough set model based on multi-scale covering. *Int J Mach Learn Cybernet* 12:243–256
- Liang DC, Fu YY, Xu ZS (2022) Novel AQM analysis approach based on similarity and dissimilarity measures of interval set for multi-expert multi-criterion decision making. *Int J Approx Reason* 142:266–289
- Liang JY, Chin KS, Dang CY, Yam RCM (2002) A new method for measuring uncertainty and fuzziness in rough set theory. *Int J General Syst* 31:331–342
- Liao WH, Jiang WG, Huang ZQ (2022) Spatiotemporal variations of eco-environment in the Guangxi Beibu Gulf Economic Zone based on remote sensing ecological index and granular computing. *J Geographical Sci* 32:1813–1830
- Luo C, Li TR, Huang YY, Fujita H (2019) Updating three-way decisions in incomplete multi-scale information systems. *Inform Sci* 476:274–289
- Pawlak Z (1981) Information systems theoretical foundations. *Inform Syst* 6:205–218
- Pawlak Z (1982) Rough sets. *Int J Comput Inform Sci* 11:341–356
- Qi GA, Yang B, Li W (2023) Some neighborhood-related fuzzy covering-based rough set models and their applications for decision making. *Inform Sci* 621:799–843
- Qin JD, Martínez L, Pedrycz W, Ma XY, Liang YY (2023) An overview of granular computing in decision-making: extensions, applications, and challenges. *Inform Fus* 98:101833
- Qu LD, He JL, Zhang GQ, Xie NX (2022) Entropy measure for a fuzzy relation and its application in attribute reduction for heterogeneous data. *Appl Soft Comput* 118:108455
- Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27:379–423
- Singh S, Shreevastava S, Som T, Somani G (2020) A fuzzy similarity-based rough set approach for attribute selection in set-valued information systems. *Soft Comput* 24:4675–4691

25. Sun Y, Wu WZ, Wang X (2023) Maximal consistent block based optimal scale selection for incomplete multi-scale information systems. *Int J Mach Learn Cybernet* 14:1797–1809
26. Wang H, Yue HB (2016) Entropy measures and granularity measures for interval and set-valued information systems. *Soft Comput* 20:3489–3495
27. Wang JB, Wu WZ, Tan AH (2022) Multi-granulation-based knowledge discovery in incomplete generalized multi-scale decision systems. *Int J Mach Learn Cybernet* 13:3963–3979
28. Wang WJ, Huang B, Wang TX (2023) Optimal scale selection based on multi-scale single-valued neutrosophic decision-theoretic rough set with cost-sensitivity. *Int J Approx Reason* 155:132–144
29. Wang WJ, Zhan JM, Zhang C, Herrera-Viedma E, Kou G (2023) A regret-theory-based three-way decision method with a priori probability tolerance dominance relation in fuzzy incomplete information systems. *Inform Fus* 89:382–396
30. Wang ZH, Chen HM, Yuan Z, Wan JH, Li TR (2023) Multi-scale fuzzy entropy-based feature selection. *IEEE Trans Fuzzy Syst* 31:3248–3262
31. Wu WZ, Leung Y (2011) Theory and applications of granular labelled partitions in multi-scale decision tables. *Inform Sci* 181:3878–3897
32. Wu WZ, Leung Y (2013) Optimal scale selection for multi-scale decision tables. *Int J Approx Reason* 54:1107–1129
33. Wu WZ, Qian YH, Li TJ, Gu SM (2017) On rule acquisition in incomplete multi-scale decision tables. *Inform Sci* 378:282–302
34. Wu WZ, Leung Y (2020) A comparison study of optimal scale combination selection in generalized multi-scale decision tables. *Int J Mach Learn Cybernet* 11:961–972
35. Wu WZ, Niu DR, Li JH, Li TJ (2023) Rule acquisition in generalized multi-scale information systems with multi-scale decisions. *Int J Approx Reason* 154:56–71
36. Xie ZH, Wu WZ, Wang LX (2023) Optimal scale selection in multi-scale interval-set decision tables. *2023 International Conference on Machine Learning and Cybernetics*. Adelaide, Australia, pp 310–314
37. Xu JC, Qu KL, Meng XR, Sun YH, Hou QC (2022) Feature selection based on multiview entropy measures in multiperspective rough set. *Int J Intell Syst* 37:7200–7234
38. Xu JC, Meng XR, Qu KL, Sun YH, Hou QC (2023) Feature selection using relative dependency complement mutual information in fitting fuzzy rough set model. *Appl Intell* 53:18239–18262
39. Xun YL, Yin QX, Zhang JF, Yang HF, Cui XH (2021) A novel discretization algorithm based on multi-scale and information entropy. *Appl Intell* 51:991–1009
40. Yao YY (1993) Interval-set algebra for qualitative knowledge representation. In: *Proceedings of the 5th International Conference on Computing and Information*, Sudbury, ON, pp 370–374
41. Yao YY, Liu Q (1999) A generalized decision logic in interval-set-valued information tables. In: *International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing*, Berlin, Heidelberg, pp 285–293
42. Yao YY (2001) Information granulation and rough set approximation. *Int J Intell Syst* 16:87–104
43. Yao YY (2005) Perspectives of granular computing. *International Conference on Granular Computing*. Beijing, China, pp 85–90
44. Zadeh LA (1979) Fuzzy sets and information granularity. In: Gupta N, Ragade R, Yager RR (eds) *Advances in Fuzzy Set Theory and Applications*. North-Holland, Amsterdam, pp 3–18
45. Zhang HY, Yang SY, Ma JM (2016) Ranking interval sets based on inclusion measures and applications to three-way decisions. *Knowledge-Based Systems* 91:62–70
46. Zhang QH, Wang J, Wang GY, Hu F (2014) Approximation set of the interval set in Pawlak's space. *Sci World J* 2014:1–12
47. Zhang YM, Jia XY, Tang ZM, Long XZ (2019) Uncertainty measures for interval set information tables based on interval  $\delta$ -similarity relation. *Inform Sci* 501:272–292
48. Zhang YM, Jia XY, Tang ZM (2021) Information-theoretic measures of uncertainty for interval-set decision tables. *Inform Sci* 577:81–104
49. Zhang XY, Huang YY (2023) Optimal scale selection and knowledge discovery in generalized multi-scale decision tables. *Int J Approx Reason* 161:108983
50. Zhang XQ, Zhang QH, Cheng YL, Wang GY (2020) Optimal scale selection by integrating uncertainty and cost-sensitive learning in multi-scale decision tables. *Int J Mach Learn Cybernet* 11:1095–1114
51. Zheng JW, Wu WZ, Bao H, Tan AH (2022) Evidence theory based optimal scale selection for multi-scale ordered decision systems. *Int J Mach Learn Cybernet* 13:1115–1129
52. Zou L, Ren SY, Sun YB, Yang XH (2023) Attribute reduction algorithm of neighborhood rough set based on supervised granulation and its application. *Soft Comput* 27:1565–1582

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.