On rule acquisition in incomplete multi-scale decision tables

Wei-Zhi Wu [a,b,*], Yuhua Qian [c], Tong-Jun Li [a,b], Shen-Ming Gu [a,b]

[a] *School of Mathematics, Physics and Information Science, Zhejiang Ocean University, Zhoushan, Zhejiang 316022, PR China*

[b] *Key Laboratory of Oceanographic Big Data Mining & Application of Zhejiang Province, Zhejiang Ocean University, Zhoushan, Zhejiang 316022, PR China*

[c] *School of Computer and Information Technology, Shanxi University, Taiyuan, Shanxi 030006, PR China*

Abstract

Granular computing and acquisition of if-then rules are two basic issues in knowledge representation and data mining. A rough set approach to knowledge discovery in incomplete multi-scale decision tables from the perspective of granular computing is proposed in this paper. The concept of incomplete multi-scale information tables in the context of rough sets is first introduced. Information granules at different levels of scales in incomplete multi-scale information tables are then described. Lower and upper approximations with reference to different levels of scales in incomplete multi-scale information tables are also defined and their properties are examined. Optimal scale selection with various requirements in incomplete multi-scale decision tables are further discussed. Relationships among different notions of optimal scales in incomplete multi-scale decision tables are presented. Finally, knowledge acquisition in the sense of rule induction in consistent and inconsistent incomplete multi-scale decision tables are explored.

Key words: Belief functions; Granular computing; Incomplete information tables; Multi-scale decision tables; Rough sets

* Corresponding author at: School of Mathematics, Physics and Information Science, Zhejiang Ocean University, Zhoushan, Zhejiang 316022, PR China. Tel.: +86 580 8180928; Fax: +86 580 2550029.

Email addresses: wuwz@zjou.edu.cn (Wei-Zhi Wu), jinchengqyh@126.com (Yuhua Qian), litj@zjou.edu.cn (Tong-Jun Li), gsm@zjou.edu.cn (Shen-Ming Gu).

1. Introduction

With the development of computer science and internet technology, Big Data has been one of the current and future research frontiers. The discovery of non-trivial, previously unknown, and potentially useful knowledge from databases is of importance in the processing and utilization of large-scale information. An important task of knowledge discovery is to establish relations among granules such as classes, clusters, sets, groups, concepts, etc. As an approach for knowledge representation and data mining, *Granular computing* (GrC), which can reduce the data size into different level of granularity, may be a potential technique to explore Big Data [1]. The purpose of GrC is to seek for an approximation scheme which can effectively solve a complex problem at a certain level of granulation. The root of GrC comes from the concept of "information granularity" proposed by Zadeh in the context of fuzzy set theory [59, 60]. Since its conception, GrC has become a fast growing field of research in the scope of both applied and theoretical information sciences [30–33, 48, 52, 53, 61].

A primitive notion in GrC is an *information granule* or simply a granule which is a clump of objects (points) drawn together by the criteria of indistinguishability, similarity or functionality [2, 33, 60]. A granule may be interpreted as one of the numerous small particles forming a larger unit. Alternatively, a granule may be considered as a localized view or a specific aspect of a large unit satisfying a given specification. The set of granules provides a representation of the unit with respect to a particular level of granularity. The process of constructing information granules is called *information granulation*. It granulates a universe of discourse into a family of disjoint or overlapping granules. Thus one of the main directions in the study of GrC is the construction, interpretation, representation of granules, and the search for relations among granules represented as IF-THEN rules having granular variables and granular values [48, 62].

Rough set theory is perhaps one of the most advanced approaches that popularizes GrC (see e.g. [11–13, 16, 17, 22, 23, 34, 36, 52, 54, 58]). It was initiated by Pawlak [28] as a formal tool for modelling and processing incomplete information. The basic notions of rough set theory are lower and upper approximations constructed by an approximation space. When the rough set approach is used to extract decision rules from a given information table, two types of decision rules may be unravelled. Based on the lower approximation of a decision class, certain information can be discovered and certain rules can be derived whereas, by using the upper approximation of a decision class, uncertain or partially certain information can be discovered and possible rules may be induced.

Various approaches using Pawlak's rough set model have been proposed to induce decision rules from data sets taking the form of decision tables. A typical and rampant existence of data set is called incomplete information tables in which attribute values for some objects are unknown (missing, null) [14]. Many authors employed the extensions of Pawlak's rough set model to reason in incomplete information tables [3–7, 10, 14–17, 20, 21, 24, 25, 27, 43–45, 50, 51]. For example, Lingras and Yao [24] employed two different generalizations of rough set models to generate plausibilistic rules with incomplete databases instead of probabilistic rules generated by a Pawlak's rough set model with complete decision tables. Greco et al. [6], Grzymala-Busse [7], and Kryszkiewicz [14, 15] used similarity relations in incomplete information tables with missing values. By analyzing similarity classes defined by Kryszkiewicz, Leung and Li [16] introduced the concept of maximal consistent block technique for rule acquisition in incomplete information tables. To unravel certain and possible decision rules in incomplete information tables, Leung et al. [17] developed new rough set approximations based on a new information structure called labelled blocks. In [45], Wu explored knowledge reduction approaches by employing the Dempster-Shafer theory of evidence in incomplete decision tables. With reference to keeping the lower approximation and upper approximation of the decision classification in the context of maximal consistent blocks, Qian et al. [36] introduced a discernibility matrix approach to calculate a lower approximation reduct and an upper approximation reduct in inconsistent incomplete decision tables.

The Pawlak's rough set model and its extensive rough set models are mainly concerned with the approximations of sets described by a single binary relation on the universe of discourse. Qian et al. [37] extended Pawlak's rough set model to a multi-granulation rough set model, where the set approximations are defined by using multi-equivalence relations on the universe of discourse. In [35], Qian et al. illuminated several basic views for establishing a multi-granulation rough set model in the context of incomplete information tables. In view of the rough-set data analysis, the multi-granulation rough set models proposed in [35, 37] are in fact obtained by adding/deleting attributes in the information tables. It is well-known that, in a Pawlak information table, each object under each attribute can only take on one value. We call such information table a single scale information table. However, objects are usually measured at different scales under the same attribute [18]. Thus, in many real-life multi-scale information tables, an object can take on as many values as there are scales under the same attribute. For example, maps can be hierarchically organized into different scales, from large to small and vice versa. The political subdivision of China at the top level has 34 provinces, autonomous regions, and directly-governed city regions. Under each province, there are many prefecture-level cities. And, under each prefecture-level city, there are several counties, so on and so forth down the hierarchy. With respect to different scales, a point in space may be located in a province, or in a prefecture-level city, or in a county, etc. Another example is that the examination results of mathematics for students can be recorded as natural numbers between 0 to 100, and it can also be graded as "Excellent", "Good", "Moderate", "Bad", and "Unacceptable". Sometimes, if needed, it might be graded into two values, "Pass" and "Fail". Hence, how to discover knowledge in hierarchically organized information tables is of particular importance in real-life data mining. In [46], Wu and Leung introduced the notion of *multi-scale information tables* from the perspective of granular computing, represented the structure of and relationships among information granules, and analyzed knowledge acquisition in multi-scale decision tables under different levels of granularity. In a multi-scale information table, each object under each attribute is represented by different scales at different levels of granulations having a granular information transformation from a finer to a coarser labelled value. In [8], Gu and Wu proposed algorithms to knowledge acquisition in multi-scale decision tables. Wu and Leung [47] further investigated the optimal scale selection for choosing a proper decision table with some requirement for final decision or classification.

However, the unravelling of rules in multi-scale information tables in which attribute values for some objects are unknown is crucial in decision making. Such a system is called an *incomplete multi-scale information table*. Hence, the main objective of this paper is to study the description of information granules and knowledge acquisition in incomplete multi-scale information tables from the perspective of granular computing.

For a given incomplete multi-scale information table, there are two key issues crucial to the discovery of knowledge in the sense of granular IF-THEN rules. One is the optimal scale selection for choosing a proper incomplete decision table with reference to some requirement for final decision or classification, and the other is knowledge reduction by reducing attributes in the selected decision table to maintain structure consistency for the induction of concise decision rules.

In the next section, we introduce some basic notions related to incomplete information tables, incomplete decision tables, rough set approximations, and belief and plausibility functions in the Dempster-Shafer theory of evidence. The concepts of incomplete multi-scale information tables and the corresponding rough set approximations are explored in Section 3. In Section 4, we investigate optimal scale selection and rule acquisition in incomplete multi-scale decision tables. We then conclude the paper with a summary and outlook for further research in Section 5.

2. Preliminaries

In this section we recall some basic notions and previous results which will be used in the later parts of this paper.

Throughout this paper, for a nonempty set U, the class of all subsets of U is denoted by $\mathcal{P}(U)$. For $X \in \mathcal{P}(U)$, we denote the complement of X in U as $\sim X$, i.e. $\sim X = U - X = \{x \in U | x \notin X\}$, when X is a finite set, the cardinality of X is denoted as $|X|$.

2.1. *Incomplete information tables and rough approximations*

The notion of information tables (sometimes called information systems, data tables, attribute-value systems, knowledge representation systems etc.) provides a convenient tool for the representation of objects in terms of their attribute values.

An *information table* (IT) S is a pair (U, A), where $U = \{x_1, x_2, \ldots, x_n\}$ is a nonempty finite set of objects called the universe of discourse and $A = \{a_1, a_2, \ldots, a_m\}$ is a nonempty finite set of attributes such that $a : U \to V_a$ for any $a \in A$, i.e., $a(x) \in V_a$, where V_a is called the domain of attribute a.

We see that in an IT information about any object is uniquely determined by the values of all the attributes, that is, for each object $x \in U$ and each attribute $a \in A$, there exists unique $v \in V_a$ such that $a(x) = v$, such a system is also called a *complete information table* (CIT). However, it may happen that our knowledge is not complete and we are not able to state with certainty what is the value taken by a given attribute $a \in A$ for a given object $x \in U$, that is, the precise value of the attribute for the object in the information table may be unknown, i.e. missing or null, in such a case, we will denote the value by $*$. Such a system is referred to as an *incomplete information table* (IIT). Any domain value different from $*$ will be called regular.

For an IIT $S = (U, A)$ and $B \subseteq A$, one can define a binary relation on U as follows [14]:

$$R_B = \{(x, y) \in U \times U | a(x) = a(y), \text{ or } a(x) = *, \text{ or } a(y) = *, \forall a \in B\}. \quad (1)$$

Obviously, R_B is reflexive and symmetric, that is, it is a similarity relation. The concept of a similarity relation has a wide variety of applications in classification [15–17, 21]. It can easily be seen that

Table 1
An exemplary incomplete information table

Car	Price	Mileage	Size	Max-Speed
x_1	High	Low	Full	Low
x_2	Low	$*$	Full	Low
x_3	$*$	$*$	Compact	Low
x_4	High	$*$	Full	High
x_5	$*$	$*$	Full	High
x_6	Low	High	Full	$*$

$$R_B = \bigcap_{a \in B} R_{\{a\}}. \tag{2}$$

Denote $S_B(x) = \{y \in U | (x, y) \in R_B\}$, $S_B(x)$ is called the similarity class of x w.r.t. B in S, the family of all similarity classes w.r.t. B is denoted by U/R_B, i.e., $U/R_B = \{S_B(x) | x \in U\}$.

A similarity relation in an IIT renders a covering of universe of discourse. Information about similarity can be represented using similarity classes for each object. It is apparent that an object of a given similarity class may also be similar to an object in another similarity class. Therefore, the basic granules of knowledge in an IIT are essentially overlapping.

By the definition of similarity class, it can easily be obtained the following:

Proposition 1 [5] *Let $S = (U, A)$ be an IIT. If $C \subseteq B \subseteq A$, then $S_B(x) \subseteq S_C(x)$ for all $x \in U$.*

Example 1 *Table 1 depicts an IIT $S = (U, A)$ with missing values containing information about cars which is a modification in* [14]. *From Table 1 we have:*

$U = \{x_1, x_2, x_3, x_4, x_5, x_6\}$, $A = \{P, M, S, X\}$,

where "P", "M", "S", "X" stand for "Price", "Mileage", "Size", "Max-Speed", respectively. The attribute domains are as follows:

$V_P = \{High,\ Low\}$, $V_M = \{High,\ Low\}$, $V_S = \{Full,\ Compact\}$, *and* $V_X = \{High,\ Low\}$.

The similarity classes determined by A are as follows:

$S_A(x_1) = \{x_1\}$, $S_A(x_2) = \{x_2, x_6\}$, $S_A(x_3) = \{x_3\}$,

$S_A(x_4) = \{x_4, x_5\}$, $S_A(x_5) = \{x_4, x_5, x_6\}$, $S_A(x_6) = \{x_2, x_5, x_6\}$.

Definition 1 *Let $S = (U, A)$ be an IIT, $B \subseteq A$, and $X \subseteq U$, one can characterize X by a pair of lower and upper approximations w.r.t. B:*

$$\underline{R_B}(X) = \{x \in U | S_B(x) \subseteq X\}, \quad \overline{R_B}(X) = \{x \in U | S_B(x) \cap X \neq \emptyset\}. \tag{3}$$

$\underline{R_B}(X)$ and $\overline{R_B}(X)$ are, respectively, referred to as the lower and upper approximations of X w.r.t. the knowledge generated by attributes B. X is said to be definable w.r.t. the knowledge generated by attributes B if $\underline{R_B}(X) = \overline{R_B}(X)$. The pair $(\underline{R_B}(X), \overline{R_B}(X))$ is called the rough set of X w.r.t. B.

Given a subset $X \subseteq U$, by using the lower and upper approximations, the universe of discourse can be divided into three pair-wise disjoint regions, namely, the positive, the boundary, and the negative regions:

$\mathrm{POS}_B(X) = \underline{R_B}(X)$,

$\mathrm{BN}_B(X) = \overline{R_B}(X) - \underline{R_B}(X)$,

$\mathrm{NEG}_B(X) = \sim \overline{R_B}(X) = U - \overline{R_B}(X)$.

Objects in the positive region $\mathrm{POS}_B(X)$ can be with certainty classified as elements of X on the basis of knowledge generated by B, objects in the negative region $\mathrm{NEG}_B(X)$ can be with certainty classified as elements not belonging to X on the basis of knowledge generated by B, and objects in the boundary region $\mathrm{BN}_B(X)$ can only be classified possibly as elements of X on the basis of knowledge generated by B.

The *accuracy* of the rough set approximation is defined as follows:

$$\alpha_B(X) = \frac{|\underline{R_B}(X)|}{|\overline{R_B}(X)|}, \tag{4}$$

where for the empty set $\emptyset$, we define $\alpha_B(\emptyset) = 1$. Clearly, $0 \leq \alpha_B(X) \leq 1$. If X is definable w.r.t. B, then $\alpha_B(X) = 1$.

Since a similarity relation is reflexive and symmetric, the approximations have following properties [56]:

Proposition 2 *Let (U, A) be an IIT, $B, C \subseteq A$, then: $\forall X, Y \in \mathcal{P}(U)$,*

(1) $\underline{R_B}(X) \subseteq X \subseteq \overline{R_B}(X)$,

(2) $\underline{R_B}(X) = \sim \overline{R_B}(\sim X)$,

(3) $\underline{R_B}(U) = \overline{R_B}(U) = U$, $\underline{R_B}(\emptyset) = \overline{R_B}(\emptyset) = \emptyset$,

(4) $\underline{R_B}(X \cap Y) = \underline{R_B}(X) \cap \underline{R_B}(Y)$, $\overline{R_B}(X \cup Y) = \overline{R_B}(X) \cup \overline{R_B}(Y)$,

(5) $X \subseteq Y \Longrightarrow \underline{R_B}(X) \subseteq \underline{R_B}(Y)$, $\quad \overline{R_B}(X) \subseteq \overline{R_B}(Y)$,

(6) $X \subseteq \underline{R_B}(\overline{R_B}(X))$, $\quad \overline{R_B}(\underline{R_B}(X)) \subseteq X$,

(7) $C \subseteq B \Longrightarrow \underline{R_C}(X) \subseteq \underline{R_B}(X)$, $\overline{R_B}(X) \subseteq \overline{R_C}(X)$.

Example 2 *In Example 1, if we set $X = \{x_2, x_5, x_6\}$, then we can obtain $\underline{R_A}(X) = \{x_2, x_6\}$, and $\overline{R_A}(X) = \{x_2, x_4, x_5, x_6\}$.*

2.2. *Belief structures and belief functions*

The Dempster-Shafer theory of evidence, also called the "evidence theory" or the "belief function theory", is treated as a promising method of dealing with uncertainty in intelligent systems. In this section we present results related to evidence theory and rough approximations in IITs.

The basic representational structure in the Dempster-Shafer theory of evidence is a belief structure [38].

Definition 2 *Let U be a non-empty finite set, a set function $m : \mathcal{P}(U) \to [0, 1]$ is referred to as a basic probability assignment or mass distribution, if it satisfies axioms:*

$$\text{(M1) } m(\emptyset) = 0, \quad \text{(M2) } \sum_{X \subseteq U} m(X) = 1.$$

The value $m(X)$ represents the degree of belief that a specific element of U belongs to set X, but not to any particular subset of X. A set $X \in \mathcal{P}(U)$ with nonzero basic probability assignment is referred to as a *focal element* of m. We denote by $\mathcal{M}$ the family of all focal elements of m. The pair $(\mathcal{M}, m)$ is called a *belief structure* on U.

Associated with each belief structure, a pair of belief and plausibility functions can be derived [38].

Definition 3 *Let* $(\mathcal{M}, m)$ *be a belief structure on* U*. A set function* $\mathrm{Bel}: \mathcal{P}(U) \to [0,1]$ *is referred to as a belief function on* U *if*

$$\mathrm{Bel}(X) = \sum_{\{Y \subseteq U | Y \subseteq X\}} m(Y), \quad \forall X \in \mathcal{P}(U). \tag{5}$$

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

$$\mathrm{Pl}(X) = \sum_{\{Y \subseteq U | Y \cap X \neq \emptyset\}} m(Y), \quad \forall X \in \mathcal{P}(U). \tag{6}$$

In the literature, one can find several interesting properties of belief and plausibility functions.

Proposition 3 [38] *If* Bel *and* Pl *are belief and plausibility functions derived from a belief structure* $(\mathcal{M}, m)$ *on* U*, then*

(1) $\mathrm{Bel}(\emptyset) = \mathrm{Pl}(\emptyset) = 0$,

(2) $\mathrm{Bel}(U) = \mathrm{Pl}(U) = 1$,

(3) $X \subseteq Y \subseteq U \Longrightarrow \mathrm{Bel}(X) \leq \mathrm{Bel}(Y)$, $\mathrm{Pl}(X) \leq \mathrm{Pl}(Y)$,

(4) $\forall \{Y_1, Y_2, \ldots, Y_k\} \subseteq \mathcal{P}(U)$,

$$\mathrm{Bel}\big(\bigcup_{i=1}^{k} Y_i\big) \geq \sum \big\{(-1)^{|J|+1} \mathrm{Bel}\big(\bigcap_{i \in J} Y_i\big) \mid \emptyset \neq J \subseteq \{1, 2, \ldots, k\}\big\},$$

$$\mathrm{Pl}\big(\bigcap_{i=1}^{k} Y_i\big) \leq \sum \big\{(-1)^{|J|+1} \mathrm{Pl}\big(\bigcup_{i \in J} Y_i\big) \mid \emptyset \neq J \subseteq \{1, 2, \ldots, k\}\big\},$$

(5) $\mathrm{Bel}(X) + \mathrm{Bel}(\sim X) \leq 1,\ \mathrm{Pl}(X) + \mathrm{Pl}(\sim X) \geq 1, \quad \forall X \subseteq U$,

(6) $\mathrm{Bel}(X) = 1 - \mathrm{Pl}(\sim X), \forall X \subseteq U$,

(7) $\mathrm{Bel}(X) \leq \mathrm{Pl}(X), \forall X \subseteq U$.

For a given belief function, the basic probability assignment can be calculated by Möbius transformation [38]:

$$m(X) = \sum_{D \subseteq X} (-1)^{|X-D|} \mathrm{Bel}(D), \quad X \subseteq U. \tag{7}$$

There are strong connections between rough set theory and the Dempster-Shafer theory of evidence. The following Theorem 1 states that a pair of lower and upper approximations induced from a reflexive approximation space can interpret a pair of belief and plausibility functions derived from a special kind of belief structure [49, 57].

Theorem 1 *Let* (U, R) *be a reflexive approximation space, i.e.,* U *is a nonempty finite set and* R *a reflexive relation on* U*, for any* $X \subseteq U$*, the lower and upper approximations of* X *w.r.t.* (U, R) *are defined as follows:*

$$\underline{R}(X) = \{x \in U | R_s(x) \subseteq X\}, \quad \overline{R}(X) = \{x \in U | R_s(x) \cap X \neq \emptyset\}, \tag{8}$$

where $R_s(x) = \{y \in U | (x, y) \in R\}$ *is the successor neighborhood of* x *w.r.t.* R*. Denote*

$$\mathrm{Bel}(X) = P(\underline{R}(X)), \qquad \mathrm{Pl}(X) = P(\overline{R}(X)), \tag{9}$$

where $P(X) = |X|/|U|$. Then Bel *and* Pl *are belief and plausibility functions on U respectively, and the corresponding basic probability assignment is*

$$m(Y) = P(j(Y)), \quad Y \in \mathcal{P}(U), \tag{10}$$

where $j(Y) = \{u \in U | R_s(u) = Y\}$. Conversely, if Bel *and* Pl *are a dual pair of belief and plausibility functions induced from a belief structure $(\mathcal{M}, m)$ on U where $m(X)$ is equivalent to a rational number with $|U|$ as its denominator for all $X \in \mathcal{P}(U)$, then there exists a reflexive approximation space (U, R) such that the induced qualities of lower and upper approximations satisfy*

$$P(\underline{R}(X)) = \mathrm{Bel}(X), \qquad P(\overline{R}(X)) = \mathrm{Pl}(X), \quad \forall X \subseteq U. \tag{11}$$

Since a similarity relation is reflexive, by Theorem 1, we can conclude the following theorem, which shows that the pair of lower and upper approximations w.r.t. an attribute set in an IIT generate a dual pair of belief and plausibility functions.

Theorem 2 *Let (U, A) be an IIT and $B \subseteq A$. For any $X \subseteq U$, denote*

$$\mathrm{Bel}_B(X) = P(\underline{R_B}(X)), \qquad \mathrm{Pl}_B(X) = P(\overline{R_B}(X)). \tag{12}$$

Then Bel_B *and* Pl_B *are a dual pair of belief and plausibility functions on U, and the corresponding basic probability assignment is*

$$m_{_B}(Y) = P(j_{_B}(Y)), \quad Y \in \mathcal{P}(U), \tag{13}$$

where $j_{_B}(Y) = \{u \in U | S_B(u) = Y\}$.

The initial results from rough set theory into evidence theory are presented by Skowron [39–41]. Theorem 1, which includes results from evidence theory into rough set theory, was first described by Yao [57]. The results of Theorem 2 can be found in [41]. In literature, $P(\underline{R_B}(X))$ and $P(\overline{R_B}(X))$ are referred to as qualities of lower and upper approximations of X w.r.t. B [41, 57], and they are called random qualities of lower and upper approximations of X w.r.t. B when P is an arbitrary probability on U [49]. The quality of lower approximation of X w.r.t. B is the probability of set of all certainly classified objects by attributes from B, whereas the quality of upper approximation of X w.r.t. B is the probability of set of all possibly classified objects by attributes from B.

Combining Theorem 2 and Proposition 2, we obtain the following:

Proposition 4 *Let (U, A) be an IIT. If $C \subseteq B \subseteq A$, then, for any $X \subseteq U$,*

$$\mathrm{Bel}_C(X) \le \mathrm{Bel}_B(X) \le P(X) \le \mathrm{Pl}_B(X) \le \mathrm{Pl}_C(X). \tag{14}$$

2.3. *Incomplete decision tables and decision rules*

A *decision table* (DT) (also called a decision system) is a system $S = (U, C \cup \{d\})$ where (U, C) is an information table, $d \notin C$ and d is a complete attribute called decision, in this case C is called the conditional attribute set, d is a mapping $d : U \to V_d$ from the universe U into the value set V_d, we assume, without any loss of generality, that $V_d = \{1, 2, \ldots, r\}$. Define

$$R_d = \{(x, y) \in U \times U | d(x) = d(y)\}. \tag{15}$$

Then R_d is an equivalence relation and it forms a partition $U/R_d = \{D_1, D_2, \ldots, D_r\} = \{[x]_d | x \in U\}$ of U into decision classes, where $D_j = \{x \in U | d(x) = j\}, j = 1, 2, \ldots, r$, and $[x]_d = \{y \in U | (x, y) \in R_d\}$. If (U, C) is a CIT, then S is referred to as a *complete decision table* (CDT), and if (U, C) is an IIT, then S is called an *incomplete decision table* (IDT).

Let $S = (U, C \cup \{d\})$ be an IDT and $B \subseteq C$, denote

$$\partial_B(x) = \{d(y) | y \in S_B(x)\}, \quad x \in U. \tag{16}$$

$\partial_B(x)$ is called the *generalized decision* of x w.r.t. B in S. S is said to be *consistent* if $|\partial_C(x)| = 1$ for all $x \in U$, otherwise it is *inconsistent.* In a consistent IDT $S = (U, C \cup \{d\})$, we have $R_C \subseteq R_d$, i.e., $S_C(x) \subseteq [x]_d$ for all $x \in U$, alternatively, $\partial_C(x) = \{d(x)\}$ for all $x \in U$.

In the discussion to follow, the symbols $\wedge$ and $\vee$ denote the logical connectives "and" (conjunction) and "or"(disjunction), respectively. Any attribute-value pair $(a, v), v \in V_a, a \in B, B \subseteq C$, is called a *B-atomic property.* Any B-atomic property or conjunction of different B-atomic properties is called a *B-descriptor.* Let t be a B-descriptor, the attribute set occurring in t is denoted by $B(t)$. If (a, v) is an atomic property occurring in t, we simply say that $(a, v) \in t$. The set of objects having descriptor t is called the *support* of t and is denoted by $\|t\|$, i.e., $\|t\| = \{x \in U | v = a(x), \ \forall (a, v) \in t\}$. If t and s are two atomic properties, then it can be observed that $\|t \wedge s\| = \|t\| \cap \|s\|$ and $\|t \vee s\| = \|t\| \cup \|s\|$.

For $B \subseteq C$, we denote

$$\text{DES}(B) = \{t | t \text{ is a } B\text{-descriptor and } \|t\| \neq \emptyset\}. \tag{17}$$

For any $t \in \text{DES}(B)$, if $B(t) = B$, then t is called a *full B-descriptor.* Denote

$$\text{FDES}(B) = \{t | t \text{ is a full } B\text{-descriptor}\}. \tag{18}$$

Example 3 *In Example 1, let* $B = \{S, X\}$*, then all B-atomic properties are* (S, Full)*,* $(S, \text{Compact})$*,* (X, Low)*, and* (X, High).

The set of all B-descriptors is $\{(S, \text{Full}), (S, \text{Compact}), (X, \text{Low}), (X, \text{High}), (S, \text{Full}) \wedge (X, \text{Low}), (S, \text{Full}) \wedge (X, \text{High}), (S, \text{Compact}) \wedge (X, \text{Low}), (S, \text{Compact}) \wedge (X, \text{High})\}$.

And it can be derived from Table 1 that

$$\begin{aligned} \text{DES}(B) &= \{(S, \text{Full}), (S, \text{Compact}), (X, \text{Low}), (X, \text{High}), (S, \text{Full}) \wedge (X, \text{Low}), \\ &\quad (S, \text{Compact}) \wedge (X, \text{Low}), (S, \text{Full}) \wedge (X, \text{High})\}. \\ \text{FDES(B)} &= \{(S, \text{Full}) \wedge (X, \text{Low}), (S, \text{Compact}) \wedge (X, \text{Low}), (S, \text{Full}) \wedge (X, \text{High})\}. \end{aligned}$$

For any B-descriptor t, denote

$$\partial(t) = \{d(y) | y \in \|t\|\}, \tag{19}$$

$\partial(t)$ is called the *generalized decision* of t in S. Any (d, w), $w \in \partial(t)$, is referred to as a *generalized decision descriptor* of t.

If $|(\partial(t))| = 1$, we then say that descriptor t is *consistent*, otherwise it is *inconsistent.*

The following proposition can easily be concluded [17]:

Proposition 5 *Let $S = (U, C \cup \{d\})$ be an IDT, $w \in V_d$, and $t \in \mathrm{DES}(C)$. Then*
(1) $\|t\| \subseteq \|(d, w)\|$ *iff* $\partial(t) = \{w\}$,
(2) $\|t\| \cap \|(d, w)\| \neq \emptyset$ *iff* $w \in \partial(t)$.

Thus, we can see that an IDT S is consistent iff $|\partial(t)| = 1$ for all $t \in \mathrm{FDES}(C)$.

For $X \subseteq U$ and $B \subseteq C$, it can be verified that

$$\begin{aligned} \underline{R_B}(X) &= \cup\{\|t\| \mid \|t\| \subseteq X, t \in \mathrm{FDES}(B)\}, \\ \overline{R_B}(X) &= \cup\{\|t\| \mid \|t\| \cap X \neq \emptyset, t \in \mathrm{FDES}(B)\}. \end{aligned} \tag{20}$$

Let $w \in V_d$ and $t \in \mathrm{FDES}(B)$. If $\|t\| \subseteq \underline{R_B}(\|(d, w)\|)$ (resp. $\|t\| \subseteq \overline{R_B}(\|(d, w)\|)$), then we call t a *lower* (resp. an *upper*) *approximation B-descriptor* of (d, w). The set of all lower (resp. upper) approximation B-descriptors of (d, w) is denoted by $\underline{R_B}((d, w))$ (resp. $\overline{R_B}((d, w))$). And also, if $\|t\| \subseteq \mathrm{BN}_B(\|(d, w)\|)$, then t is referred to as a *boundary descriptor* of (d, w) w.r.t. B. The set of all boundary descriptors of (d, w) w.r.t. B is denoted by $\mathrm{BNDES}_B((d, w))$.

Proposition 6 below shows that the approximations of decision classes can be expressed by means of the generalized decision [17].

Proposition 6 *Let $S = (U, C \cup \{d\})$ be an IDT. If $w \in V_d$, t is a C-descriptor, and $B \subseteq C$, then*
(1) $\underline{R_B}(\|(d, w)\|) = \cup\{\|t\| \mid t \in \mathrm{FDES}(B), \partial(t) = \{w\}\}$,
(2) $\overline{R_B}(\|(d, w)\|) = \cup\{\|t\| \mid t \in \mathrm{FDES}(B), w \in \partial(t)\}$,
(3) $\underline{R_B}((d, w)) = \{ t \in \mathrm{FDES}(B) \mid \partial(t) = \{w\}\}$,
(4) $\overline{R_B}((d, w)) = \{t \in \mathrm{FDES}(B) \mid w \in \partial(t)\}$.

The knowledge hidden in an IDT $S = (U, C \cup \{d\})$ may be discovered and expressed in the form of *decision rules*: $t \to s$, where $t = \wedge(a, v), a \in B \subseteq C$, and $s = (d, w), w \in V_d$, t and s are, respectively, called the *condition* and *decision* parts of the rule.

We say that an object $x \in U$ *supports a rule* $t \to s$ in the IDT S iff $x \in \|t\| \cap \|s\|$.

A decision rule $t \to s = (d, w)$ is referred to as *certain* in the IDT S iff $\|t\| \neq \emptyset$ and $\|t\| \subseteq \|s\|$, in such a case, we denote $t \Rightarrow s$ instead of $t \to s$.

A decision rule $t \to s = (d, w)$ is referred to as *possible* in the IDT S iff $\|t\| \not\subseteq \|s\|$ and $\|t\| \cap \|s\| \neq \emptyset$.

By using the lower and upper approximations of decision classes w.r.t. a conditional attribute subset, one can acquire certain decision rules and possible decision rules from an IDT.

With each decision rule $t \to s = (d, w)$ in an IDT S, we associate a quantitative measure, called the *certainty*, of the rule in S and is defined by [17, 29]:

$$Cer(t \to s) = \frac{|\ \|t\| \cap \|s\|\ |}{|\ \|t\|\ |}. \tag{21}$$

The quantity $Cer(t \to s)$ shows the degree to which objects supporting descriptor t also support decision s in S. If $Cer(t \to s) = \alpha$, then $(100\alpha)\%$ of objects supporting t also support s in S.

The following Proposition 7 shows that the types of decision rules can be expressed by means of the certainty factors of the rules as well as the lower and the upper approximations of each decision class w.r.t the set of conditional attributes in an IDT.

Proposition 7 *Let $S = (U, C \cup \{d\})$ be an IDT. If $w \in V_d$, t is a C-descriptor, and $s = (d, w)$, then the decision rule $t \to s$*

$$
\begin{aligned}
&(1) \text{ is certain in } S && \Longleftrightarrow \|t\| \subseteq \underline{R_{C(t)}}(\|(d,w)\|) \\
& && \Longleftrightarrow t \in \underline{R_{C(t)}}((d,w)) \\
& && \Longleftrightarrow \partial(t) = \{w\} \\
& && \Longleftrightarrow Cer(t \to (d,w)) = 1. \\
&(2) \text{ is a possible rule in } S && \Longleftrightarrow \|t\| \subseteq \mathrm{BN}_{C(t)}(\|(d,w)\|) \\
& && \Longleftrightarrow t \in \mathrm{BNDES}_{C(t)}((d,w)) \\
& && \Longleftrightarrow w \in \partial(t) \text{ and } \mid \partial(t) \mid \geq 2 \\
& && \Longleftrightarrow 0 < Cer(t \to (d,w)) < 1.
\end{aligned}
$$

One can acquire certainty decision rules from consistent IDTs and uncertainty decision rules from inconsistent IDTs. In fact, if $|\partial_C(x)| = 1$, then the decision rule corresponding to (or supported by the objects in) the similarity class $S_C(x)$ is certain, otherwise, $|\partial_C(x)| \geq 2$, the decision rule corresponding to the similarity class $S_C(x)$ is uncertain.

We can see that the lower and upper approximations divide the universe of objects into three pair-wise disjoint regions: the lower approximation as the positive region, the complement of the upper approximation as the negative region, and the difference between the upper and lower approximations as the boundary region. Observing that rules constructed from the three regions are associated with different actions and decisions, by employing probabilistic rough sets and Bayesian decision theory, Yao [55] proposed a new notion of three-way decision rules in which a positive rule makes a decision of acceptance, a negative rule makes a decision of rejection, and a boundary rule makes a decision of abstaining.

A decision rule with too long a description means high prediction cost. To acquire concise decision rules from IDTs, knowledge reduction is needed. It is well-known that not all conditional attributes are necessary to depict the decision attribute before decision rules are generated. Thus knowledge reduction by reducing attributes is one of the main problems in the study of rough set theory and it is performed in information tables by means of the notion of a reduct (see e.g. [9, 11, 19, 26, 28, 42, 45, 48]).

Definition 4 *Let $S = (U, C \cup \{d\})$ be an IDT and $B \subseteq C$. Then*

(1) *B is referred to as a lower approximation consistent set of S if $\underline{R_B}(D) = \underline{R_C}(D)$ for all $D \in U/R_d$. If B is a lower approximation consistent set of S and no proper subset of B is a lower approximation consistent set of S, then B is referred to as a lower approximation reduct of S.*

(2) *B is referred to as an upper approximation consistent set of S if $\overline{R_B}(D) = \overline{R_C}(D)$ for all $D \in U/R_d$. If B is an upper approximation consistent set of S and no proper subset of B is an upper approximation consistent set of S, then B is referred to as an upper approximation reduct of S.*

(3) *B is referred to as a generalized decision consistent set of S if $\partial_B(x) = \partial_C(x)$ for all $x \in U$. If B is a generalized decision consistent set of S and no proper subset of B is*

a generalized decision consistent set of S, then B is referred to as a generalized decision reduct of S.

(4) *B is referred to as a belief consistent set of S if* $\mathrm{Bel}_B(D) = \mathrm{Bel}_C(D)$ *for all $D \in U/R_d$. If B is a belief consistent set of S and no proper subset of B is a belief consistent set of S, then B is referred to as a belief reduct of S.*

(5) *B is referred to as a plausibility consistent set of S if* $\mathrm{Pl}_B(D) = \mathrm{Pl}_C(D)$ *for all $D \in U/R_d$. If B is a plausibility consistent set of S and no proper subset of B is a plausibility consistent set of S, then B is referred to as a plausibility reduct of S.*

By Definition 4, we see that a lower approximation reduct in an IDT is a minimal attribute subset to preserve the lower approximations of the decision classes w.r.t. the full conditional attribute set; An upper approximation reduct in an IDT is a minimal attribute subset to preserve the upper approximations of the decision classes w.r.t. the full conditional attribute set; A generalized decision reduct in an IDT is a minimal attribute subset to preserve the generalized decisions of the objects in the similarity classes generated by the full conditional attribute set. And a belief (resp. plausibility) reduct in an IDT is a minimal attribute subset to keep the degree of belief (resp. plausibility) of each decision class w.r.t. the full conditional attribute set.

It has been proved [45] that an attribute subset $B \subseteq C$ in an IDT $S = (U, C \cup \{d\})$ is a lower approximation reduct of S iff it is a belief reduct of S; and B is an upper approximation reduct of S iff it is a generalized decision reduct of S iff it is a plausibility reduct of S. So, in fact, there are only two different types of reducts in Definition 4, namely, the lower approximation reduct (belief reduct) and the upper approximation reduct (generalized decision reduct, plausibility reduct), which are related to certain and possible decision rules, respectively.

More specifically, if $S = (U, C \cup \{d\})$ is a consistent IDT, that is, $R_C \subseteq R_d$, then it is easy to see that $\partial_C(x) = \{d(x)\}$ for all $x \in U$. In such case, a subset $B \subseteq C$ is a consistent set of S iff $R_B \subseteq R_d$ and B is a reduct of the consistent IDT S iff B is a minimal attribute set keeping the subtable $(U, B \cup \{d\})$ consistent, i.e. $R_B \subseteq R_d$. In this case, all the five types of reducts in Definition 4 are equivalent (see [45]), in the literature, it is called a relative reduct.

As for an inconsistent IDT $S = (U, C \cup \{d\})$, for any $B \subseteq C$, we define a binary relation R_d^B on U as follows:

$$R_d^B = \{(x, y) \in U \times U | \partial_{_B}(x) = \partial_{_B}(y)\}. \tag{22}$$

It can easily be verified that R_d^B is an equivalence relation on U and $(U, C \cup \{\partial_C\})$ is a consistent IDT. Moreover, it can be checked that $\partial_{_B}(x) = \partial_{_C}(x)$ for all $x \in U$ iff $R_B \subseteq R_d^C$, that is, B is a generalized decision consistent set of the inconsistent IDT $(U, C \cup \{d\})$ iff B is a (lower approximation) consistent set of the consistent IDT $(U, C \cup \{\partial_C\})$, and consequently, B is a generalized decision reduct of the inconsistent IDT $(U, C \cup \{d\})$ iff B is a (lower approximation) reduct of consistent IDT $(U, C \cup \{\partial_C\})$. Thus, we can calculate reducts in the consistent IDT $(U, C \cup \{\partial_C\})$ in stead of computing generalized decision reducts in the inconsistent IDT $(U, C \cup \{d\})$.

Computing reducts in an IDT can also be translated into the computation of prime implicants of a Boolean function. For the detail, we refer the reader to [14, 15, 36].

3. Incomplete multi-scale information tables

In an information table, each object can only take on one value under each attribute. However, in some real-life applications, one has to make decision with different levels of scales. That is, an object may take on different values under the same attribute, depending on at which scale it is measured. In this section we introduce the concept of incomplete multi-scale information tables from the perspective of granular computing.

Definition 5 [46] *A multi-scale information table is a tuple $S = (U, A)$, where*

- $U = \{x_1, x_2, \ldots, x_n\}$ *is a non-empty, finite set of objects called the universe of discourse;*
- $A = \{a_1, a_2, \ldots, a_m\}$ *is a non-empty, finite set of attributes, and each $a_j \in A$ is a multi-scale attribute, i.e., for the same object in U, attribute a_j can take on different values at different scales.*

In the discussion to follow, we always assume that all the attributes have the same number I of levels of scales. Hence, a multi-scale information table can be represented as a table $(U, \{a_j^k | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\})$, where $a_j^k : U \to V_j^k$ is a surjective mapping and V_j^k is the domain of the k-th scale attribute a_j^k. For $k \in \{1, 2, \ldots, I-1\}$, there exists a surjective mapping $g_j^{k,k+1} : V_j^k \to V_j^{k+1}$ such that $a_j^{k+1} = g_j^{k,k+1} \circ a_j^k$, i.e.

$$a_j^{k+1}(x) = g_j^{k,k+1}(a_j^k(x)), \quad x \in U, \tag{23}$$

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

For $k \in \{1, 2, \ldots, I\}$, we denote $A^k = \{a_j^k | j = 1, 2, \ldots, m\}$. Then a multi-scale information table $S = (U, A)$ can be decomposed into I ITs $S^k = (U, A^k), k = 1, 2, \ldots, I$.

If $S^1 = (U, A^1)$ is an IIT, then $(U, \{a_j^k | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\})$ is called an *incomplete multi-scale information table*, in such a case, the granular information transformation mappings are defined as follows:

$$a_j^{k+1}(x) = \begin{cases} *, & \text{if } a_j^k(x) = *, \\ g_j^{k,k+1}(a_j^k(x)), & \text{otherwise.} \end{cases} \tag{24}$$

where $k = 1, 2, \ldots, I-1, x \in U, j = 1, 2, \ldots, m$. Hence, an incomplete multi-scale information table $S = (U, A)$ can be decomposed into I IITs $S^k = (U, A^k), k = 1, 2, \ldots, I$.

Definition 6 *Let U be a nonempty set, and $\mathcal{A}_1$ and $\mathcal{A}_2$ two coverings of U. If for each $A_1 \in \mathcal{A}_1$, there exists $A_2 \in \mathcal{A}_2$ such that $A_1 \subseteq A_2$, then we say that $\mathcal{A}_1$ is finer than $\mathcal{A}_2$ or $\mathcal{A}_2$ is coarser than $\mathcal{A}_1$, and is denoted as $\mathcal{A}_1 \sqsubseteq \mathcal{A}_2$.*

The following Proposition 8 can easily be concluded.

Proposition 8 *Let $S = (U, A) = (U, \{a_j^k | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\})$ be an incomplete multi-scale information table, and $B \subseteq A$, for $k \in \{1, 2, \ldots, I\}$, denote*

$$\begin{aligned} &R_{B^k} = \{(x, y) \in U \times U | a^k(x) = a^k(y), \ \textit{or } a(x) = *, \ \textit{or } a(y) = *, \forall a \in B\}, \\ &S_{B^k}(x) = \{y \in U | (x, y) \in R_{B^k}\}, \\ &U/R_{B^k} = \{S_{B^k}(x) | x \in U\}. \end{aligned} \tag{25}$$

Then

$$\begin{aligned}&R_{B^1} \subseteq R_{B^2} \subseteq \cdots \subseteq R_{B^I},\\ &S_{_{B^1}}(x) \subseteq S_{_{B^2}}(x) \subseteq \cdots \subseteq S_{_{B^I}}(x),\ x \in U,\\ &U/R_{B^1} \sqsubseteq U/R_{B^2} \sqsubseteq \cdots \sqsubseteq U/R_{B^I}.\end{aligned} \tag{26}$$

Let $S = (U, A) = (U, \{a_j^k | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\})$ be an incomplete multi-scale information table, $B \subseteq A$, $k \in \{1, 2, \ldots, I\}$, and $X \subseteq U$, the lower and upper approximations of X w.r.t. B^k are defined as follows:

$$\underline{R_{B^k}}(X) = \{x \in U | S_{B^k}(x) \subseteq X\}, \quad \overline{R_{B^k}}(X) = \{x \in U | S_{B^k}(x) \cap X \neq \emptyset\}. \tag{27}$$

The following Proposition 9 presents some properties of set approximations with different scales.

Proposition 9 *Let $S = (U, A) = (U, \{a_j^k | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\})$ be an incomplete multi-scale information table, $B \subseteq A$, and $k \in \{1, 2, \ldots, I\}$, then: $\forall X, Y \in \mathcal{P}(U)$,*

(1) $\underline{R_{B^k}}(X) = \sim \overline{R_{B^k}}(\sim X)$,
(2) $\overline{R_{B^k}}(X) = \sim \underline{R_{B^k}}(\sim X)$,
(3) $\underline{R_{B^k}}(\emptyset) = \overline{R_{B^k}}(\emptyset) = \emptyset$,
(4) $\underline{R_{B^k}}(U) = \overline{R_{B^k}}(U) = U$,
(5) $\underline{R_{B^k}}(X \cap Y) = \underline{R_{B^k}}(X) \cap \underline{R_{B^k}}(Y)$,
(6) $\overline{R_{B^k}}(X \cup Y) = \overline{R_{B^k}}(X) \cup \overline{R_{B^k}}(Y)$,
(7) $X \subseteq Y \Longrightarrow \underline{R_{B^k}}(X) \subseteq \underline{R_{B^k}}(Y)$,
(8) $X \subseteq Y \Longrightarrow \overline{R_{B^k}}(X) \subseteq \overline{R_{B^k}}(Y)$,
(9) $\underline{R_{B^k}}(X \cup Y) \supseteq \underline{R_{B^k}}(X) \cup \underline{R_{B^k}}(Y)$,
(10) $\overline{R_{B^k}}(X \cap Y) \subseteq \overline{R_{B^k}}(X) \cap \overline{R_{B^k}}(Y)$,
(11) $\underline{R_{B^k}}(X) \subseteq X \subseteq \overline{R_{B^k}}(X)$,
(12) $\underline{R_{B^{k+1}}}(X) \subseteq \underline{R_{B^k}}(X)$, *where* $k \in \{1, 2, \ldots, I-1\}$,
(13) $\overline{R_{B^k}}(X) \subseteq \overline{R_{B^{k+1}}}(X)$, *where* $k \in \{1, 2, \ldots, I-1\}$.

For $B \subseteq A$ and $X \subseteq U$, since $U/R_{B^1} \sqsubseteq U/R_{B^2} \sqsubseteq \cdots \sqsubseteq U/R_{B^I}$, according to Proposition 9, we can obtain a nested sequence of set approximations as follows:

$$\begin{aligned}&\underline{R_{B^I}}(X) \subseteq \underline{R_{B^{I-1}}}(X) \subseteq \cdots \subseteq \underline{R_{B^2}}(X) \subseteq \underline{R_{B^1}}(X) \subseteq X,\\ &X \subseteq \overline{R_{B^1}}(X) \subseteq \overline{R_{B^2}}(X) \subseteq \cdots \subseteq \overline{R_{B^{I-1}}}(X) \subseteq \overline{R_{B^I}}(X).\end{aligned} \tag{28}$$

Therefore, we have nested sequences of the positive, the boundary, and the negation regions:

$$\begin{aligned}&\mathrm{POS}_{B^I}(X) \subseteq \mathrm{POS}_{B^{I-1}}(X) \subseteq \cdots \subseteq \mathrm{POS}_{B^2}(X) \subseteq \mathrm{POS}_{B^1}(X),\\ &\mathrm{BN}_{B^1}(X) \subseteq \mathrm{BN}_{B^2}(X) \subseteq \cdots \subseteq \mathrm{BN}_{B^{I-1}}(X) \subseteq \mathrm{BN}_{B^I}(X),\\ &\mathrm{NEG}_{B^I}(X) \subseteq \mathrm{NEG}_{B^{I-1}}(X) \subseteq \cdots \subseteq \mathrm{NEG}_{B^2}(X) \subseteq \mathrm{NEG}_{B^1}(X).\end{aligned} \tag{29}$$

Consequently, we obtain a sequence of accuracies for approximations w.r.t. different scales:

$$\alpha_{_{B^I}} \leq \alpha_{_{B^{I-1}}} \leq \cdots \leq \alpha_{_{B^2}} \leq \alpha_{_{B^1}}. \tag{30}$$

Table 2
An incomplete multi-scale decision table with three levels of scales

U	a_1^1	a_1^2	a_1^3	a_2^1	a_2^2	a_2^3	a_3^1	a_3^2	a_3^3	d
x_1	1	S	N	2	S	N	3	M	Y	1
x_2	2	S	N	2	S	N	3	M	Y	1
x_3	2	S	N	$*$	$*$	$*$	4	L	Y	2
x_4	3	M	Y	5	L	Y	1	S	N	2
x_5	3	M	Y	4	L	Y	4	L	Y	2
x_6	4	L	Y	4	L	Y	$*$	$*$	$*$	2
x_7	5	L	Y	4	L	Y	3	M	Y	2
x_8	1	S	N	5	L	Y	3	M	Y	2

By employing Theorem 2 and inclusion relation (28), we can conclude following:

Proposition 10 *Let $S = (U, A) = (U, \{a_j^k | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\})$ be an incomplete multi-scale information table, and $B \subseteq A$, for $k \in \{1, 2, \ldots, I\}$, denote*

$$\begin{aligned} \mathrm{Bel}_{B^k}(X) &= P(\underline{R_{B^k}}(X)) = \frac{|\underline{R_{B^k}}(X)|}{|U|}, \ \forall X \in \mathcal{P}(U), \\ \mathrm{Pl}_{B^k}(X) &= P(\overline{R_{B^k}}(X)) = \frac{|\overline{R_{B^k}}(X)|}{|U|}, \ \forall X \in \mathcal{P}(U). \end{aligned} \tag{31}$$

Then $\mathrm{Bel}_{B^k} : \mathcal{P}(U) \to [0,1]$ and $\mathrm{Pl}_{B^k} : \mathcal{P}(U) \to [0,1]$ are a dual pair of belief and plausibility functions on U, and the corresponding basic probability assignment $m_{_{B^k}} : \mathcal{P}(U) \to [0,1]$ is

$$m_{_{B^k}}(Y) = \begin{cases} P(j_{_{B^k}}(Y)) = \dfrac{|j_{_{B^k}}(Y)|}{|U|}, & \text{if } Y \in U/R_{_{B^k}}, \\ 0, & \text{otherwise.} \end{cases} \tag{32}$$

where $j_{_{B^k}}(Y) = \{u \in U | S_{B^k}(u) = Y\}$. Moreover, the belief and plausibility functions satisfy the following properties:

(1) $\mathrm{Bel}_{B^I}(X) \le \mathrm{Bel}_{B^{I-1}}(X) \le \cdots \le \mathrm{Bel}_{B^2}(X) \le \mathrm{Bel}_{B^1}(X) \le P(X)$,

(2) $P(X) \le \mathrm{Pl}_{B^1}(X) \le \mathrm{Pl}_{B^2}(X) \le \cdots \le \mathrm{Pl}_{B^{I-1}}(X) \le \mathrm{Pl}_{B^I}(X)$,

(3) $C \subseteq B \subseteq A \Longrightarrow \mathrm{Bel}_{C^k}(X) \le \mathrm{Bel}_{B^k}(X) \le P(X) \le \mathrm{Pl}_{B^k}(X) \le \mathrm{Pl}_{C^k}(X)$.

Definition 7 *A system $S = (U, C \cup \{d\}) = (U, \{a_j^k | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\} \cup \{d\})$ is referred to as an incomplete multi-scale decision table, where $(U, C) = (U, \{a_j^k | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\})$ is an incomplete multi-scale information table and $d \notin \{a_j^k | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\}$, $d : U \to V_d$, is a special attribute called the decision. Such a system can be decomposed into I IDTs $S^k = (U, \{a_j^k | j = 1, 2, \ldots, m\} \cup \{d\}) = (U, C^k \cup \{d\})$, $k = 1, 2, \ldots, I$, with the same decision d. An incomplete multi-scale decision table S is referred to as consistent if the IDT under the first (finest) level of scale, $S^1 = (U, \{a_j^1 | j = 1, 2, \ldots, m\} \cup \{d\}) = (U, C^1 \cup \{d\})$, is consistent, and S is called inconsistent if S^1 is an inconsistent IDT.*

Example 4 *Table 2 is an example of an incomplete multi-scale decision table $S = (U, \{a_j^k | k = 1, 2, 3, j = 1, 2, 3\} \cup \{d\})$, where $U = \{x_1, x_2, \ldots, x_8\}$, $C = \{a_1, a_2, a_3\}$. The*

Table 3
The incomplete decision table with the first level of scale of Table 2

U	a_1^1	a_2^1	a_3^1	d
x_1	1	2	3	1
x_2	2	2	3	1
x_3	2	$*$	4	2
x_4	3	5	1	2
x_5	3	4	4	2
x_6	4	4	$*$	2
x_7	5	4	3	2
x_8	1	5	3	2

Table 4
The incomplete decision table with the second level of scale of Table 2

U	a_1^2	a_2^2	a_3^2	d
x_1	S	S	M	1
x_2	S	S	M	1
x_3	S	$*$	L	2
x_4	M	L	S	2
x_5	M	L	L	2
x_6	L	L	$*$	2
x_7	L	L	M	2
x_8	S	L	M	2

Table 5
The incomplete decision table with the third level of scale of Table 2

U	a_1^3	a_2^3	a_3^3	d
x_1	N	N	Y	1
x_2	N	N	Y	1
x_3	N	$*$	Y	2
x_4	Y	Y	N	2
x_5	Y	Y	Y	2
x_6	Y	Y	$*$	2
x_7	Y	Y	Y	2
x_8	N	Y	Y	2

table has three levels of scales, where "S", "M", "L", "Y", and "N" stand for, respectively, "Small", "Medium", "Large", "Yes", and "No". For these levels of granularities, the system is associated with three IDTs which are described as Tables 3–5, respectively. It can easily be checked that $R_{C^1} \subseteq R_d$, thus S is a consistent incomplete multi-scale decision table.

4. Knowledge discovery in incomplete multi-scale decision tables

In this section we discuss knowledge discovery with rough set approach in incomplete multi-scale decision tables.

4.1. *Rule acquisition in consistent incomplete multi-scale decision tables*

In this subsection we investigate knowledge acquisition in the sense of rule induction from a consistent incomplete multi-scale decision table.

Let $S = (U, C \cup \{d\}) = (U, \{a_j^k | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\} \cup \{d\})$ be a consistent incomplete multi-scale decision table which has I levels of scales. For $1 \leq i < k \leq I$, if $S^k = (U, C^k \cup \{d\}) = (U, \{a_j^k | j = 1, 2, \ldots, m\} \cup \{d\})$ is a consistent IDT, i.e. $R_{_{C^k}} \subseteq R_d$, then, by Proposition 8, we can observe that $R_{_{C^i}} \subseteq R_{_{C^k}} \subseteq R_d$. Hence, $S^i = (U, C^i \cup \{d\}) = (U, \{a_j^i | j = 1, 2, \ldots, m\} \cup \{d\})$ is also a consistent IDT.

Definition 8 *Let $S = (U, C \cup \{d\}) = (U, \{a_j^k | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\} \cup \{d\})$ be a consistent incomplete multi-scale decision table which has I levels of scales, the k-th level of scale is said to be optimal if S^k is consistent and S^{k+1} (if there exists $k + 1$) is inconsistent.*

According to Definition 8, we can see that the optimal scale of a consistent incomplete multi-scale decision table is the best scale for decision making or classification in the multi-scale decision table. And k is the optimal scale iff k is the maximal number such that S^k is a consistent IDT.

Lemma 1 *Let $S = (U, C \cup \{d\}) = (U, \{a_j^k | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\} \cup \{d\})$ be a consistent incomplete multi-scale decision table, for $k \in \{1, 2, \ldots, I\}$, if $S^k = (U, C^k \cup \{d\}) = (U, \{a_j^k | j = 1, 2, \ldots, m\} \cup \{d\})$ is a consistent IDT, i.e., $R_{C^k} \subseteq R_d$, then*

(1) $\underline{R_{C^k}}(D) = D$ *for all* $D \in U/R_d$.

(2) $\overline{R_{C^k}}(D) = D$ *for all* $D \in U/R_d$.

Proof. (1) For each $D \in U/R_d$, by Proposition 9, we have $\underline{R_{C^k}}(D) \subseteq D$. On the other hand, for any $x \in D$, clearly, $[x]_d = D$. Since $S^k = (U, C^k \cup \{d\})$ is consistent, we have $S_{C^k}(x) \subseteq [x]_d = D$. By the definition of lower approximation, we conclude $x \in \underline{R_{C^k}}(D)$. Consequently, $D \subseteq \underline{R_{C^k}}(D)$. Thus $\underline{R_{C^k}}(D) = D$.

(2) For each $D \in U/R_d$, by Proposition 9, we have $D \subseteq \overline{R_{C^k}}(D)$. Conversely, assume that $x \in \overline{R_{C^k}}(D)$, by the definition of upper approximation, we have $S_{C^k}(x) \cap D \neq \emptyset$. For any $y \in S_{C^k}(x) \cap D$, it is easy to observe that $[y]_d = D$. Since R_{C^k} is symmetric, from $y \in S_{C^k}(x)$ we have $x \in S_{C^k}(y)$. Since $S^k = (U, C^k \cup \{d\})$ is consistent, we obtain $S_{C^k}(y) \subseteq [y]_d$, hence $x \in S_{C^k}(y) \subseteq [y]_d = D$. It follows that $\overline{R_{C^k}}(D) \subseteq D$. Thus $\overline{R_{C^k}}(D) = D$. □

Theorem 3 *Let $S = (U, C \cup \{d\}) = (U, \{a_j^k | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\} \cup \{d\})$ be a consistent incomplete multi-scale decision table which has I levels of scales. For $k \in \{1, 2, \ldots, I\}$, then the following statements are equivalent:*

(1) $S^k = (U, C^k \cup \{d\}) = (U, \{a_j^k | j = 1, 2, \ldots, m\} \cup \{d\})$ *is a consistent IDT, i.e.,* $R_{C^k} \subseteq R_d$.

(2) $\sum_{D \in U/R_d} \mathrm{Bel}_{C^k}(D) = 1$.

(3) $\sum_{D \in U/R_d} \mathrm{Pl}_{C^k}(D) = 1$.

Proof. "(1) $\Rightarrow$ (2)" Since $S^k = (U, C^k \cup \{d\})$ is consistent, by Lemma 1, we have $\underline{R_{C^k}}(D) = D$ for all $D \in U/R_d$. Therefore,

$$\sum_{D\in U/R_d} \mathrm{Bel}_{C^k}(D) = \sum_{D\in U/R_d} \frac{|\underline{R_{C^k}}(D)|}{|U|} = \sum_{D\in U/R_d} \frac{|D|}{|U|} = 1. \tag{33}$$

"(2) $\Rightarrow$ (1)" Assume that $\sum_{D\in U/R_d} \mathrm{Bel}_{C^k}(D) = 1$. Since S is a consistent incomplete multi-scale decision table, by the definition, we see that $S^1 = (U, C^1 \cup \{d\}) = (U, \{a_j^1 | j = 1, 2, \ldots, m\} \cup \{d\})$ is a consistent IDT. Hence, $\sum_{D\in U/R_d} \mathrm{Bel}_{C^1}(D) = 1$. That is to say,

$$\sum_{D\in U/R_d} \mathrm{Bel}_{C^k}(D) = \sum_{D\in U/R_d} \mathrm{Bel}_{C^1}(D) = 1. \tag{34}$$

By Proposition 10, we observe that $\mathrm{Bel}_{C^k}(D) \leq \mathrm{Bel}_{C^1}(D)$ for all $D \in U/R_d$, then, by Eq. (34), we obtain

$$\mathrm{Bel}_{C^k}(D) = \mathrm{Bel}_{C^1}(D) = \frac{|D|}{|U|}, \quad \forall D \in U/R_d. \tag{35}$$

It follows that

$$|\underline{R_{C^k}}(D)| = |\underline{R_{C^1}}(D)| = |D|, \quad \forall D \in U/R_d. \tag{36}$$

According to Proposition 9, we have $\underline{R_{C^k}}(D) \subseteq D$, by Eq. (36), we then conclude that $\underline{R_{C^k}}(D) = D$ for all $D \in U/R_d$, i.e., for any $x \in U$, we have $\underline{R_{C^k}}([x]_d) = [x]_d$. Therefore, for any $y \in [x]_d$, we obtain $y \in \underline{R_{C^k}}([x]_d)$. By the definition of lower approximation, we conclude $S_{C^k}(y) \subseteq [x]_d$. Let $y = x$, we get $S_{C^k}(x) \subseteq [x]_d$. Thus we have proved that $R_{C^k} \subseteq R_d$, which means that $S^k = (U, C^k \cup \{d\})$ is a consistent IDT.

"(1) $\Rightarrow$ (3)" Since $S^k = (U, C^k \cup \{d\})$ is consistent, by Lemma 1, we have $\overline{R_{C^k}}(D) = D$ for all $D \in U/R_d$. Therefore,

$$\sum_{D\in U/R_d} \mathrm{Pl}_{C^k}(D) = \sum_{D\in U/R_d} \frac{|\overline{R_{C^k}}(D)|}{|U|} = \sum_{D\in U/R_d} \frac{|D|}{|U|} = 1. \tag{37}$$

"(3) $\Rightarrow$ (1)" Assume that $\sum_{D\in U/R_d} \mathrm{Pl}_{C^k}(D) = 1$. Since S is a consistent incomplete multi-scale decision table, by the definition, we see that $S^1 = (U, C^1 \cup \{d\}) = (U, \{a_j^1 | j = 1, 2, \ldots, m\} \cup \{d\})$ is a consistent IDT. Hence, $\sum_{D\in U/R_d} \mathrm{Pl}_{C^1}(D) = 1$, that is,

$$\sum_{D\in U/R_d} \mathrm{Pl}_{C^k}(D) = \sum_{D\in U/R_d} \mathrm{Pl}_{C^1}(D) = 1. \tag{38}$$

By Proposition 10, we see that $\mathrm{Pl}_{C^1}(D) \leq \mathrm{Pl}_{C^k}(D)$ for all $D \in U/R_d$, then, by Eq. (38), we obtain

$$\mathrm{Pl}_{C^k}(D) = \mathrm{Pl}_{C^1}(D) = \frac{|D|}{|U|}, \quad \forall D \in U/R_d. \tag{39}$$

It follows that

$$|\overline{R_{C^k}}(D)| = |\overline{R_{C^1}}(D)| = |D|, \quad \forall D \in U/R_d. \tag{40}$$

According to Proposition 9, we see that

$$D \subseteq \overline{R_{C^1}}(D) \subseteq \overline{R_{C^k}}(D), \quad \forall D \in U/R_d. \tag{41}$$

Combining Eq. (40) and inclusion relation (41), we conclude

$$D = \overline{R_{C^1}}(D) = \overline{R_{C^k}}(D), \quad \forall D \in U/R_d. \tag{42}$$

For any $x \in U$, we select $D \in U/R_d$ such that $x \in D$, clearly, $[x]_d = D$. For any $y \in S_{C^k}(x)$, from the symmetry of R_{C^k}, we have $x \in S_{C^k}(y)$. Hence $S_{C^k}(y) \cap [x]_d \neq \emptyset$, by the definition of upper approximation, we get $y \in \overline{R_{C^k}}([x]_d)$. Consequently, by Eq. (42), we conclude $y \in [x]_d$. Thus we have proved that $S_{C^k}(x) \subseteq [x]_d$ for all $x \in U$, i.e., $R_{C^k} \subseteq R_d$, which means that $S^k = (U, C^k \cup \{d\})$ is a consistent IDT. □

According to Theorem 3 and Proposition 10, we can conclude following:

Theorem 4 *Let $S = (U, C \cup \{d\}) = (U, \{a_j^k | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\} \cup \{d\})$ be a consistent incomplete multi-scale decision table which has I levels of scales. For $k \in \{1, 2, \ldots, I\}$, the following statements are equivalent:*

(1) *the k-th level of scale is the optimal scale.*

(2)

$$\sum_{D \in U/R_d} \mathrm{Bel}_{C^k}(D) = 1. \tag{43}$$

And (if there is $k+1 \leq I$)

$$\sum_{D \in U/R_d} \mathrm{Bel}_{C^{k+1}}(D) < 1. \tag{44}$$

(3)

$$\sum_{D \in U/R_d} \mathrm{Pl}_{C^k}(D) = 1. \tag{45}$$

And (if there is $k+1 \leq I$)

$$\sum_{D \in U/R_d} \mathrm{Pl}_{C^{k+1}}(D) > 1. \tag{46}$$

Theorem 4 shows that, in a consistent incomplete multi-scale decision table, k is the optimal scale iff k is the maximum number such that the sum of degrees of belief (and, the sum of degrees of plausibility) of all decision classes in S^k is 1.

Example 5 *In Example 4, since*

$\sum_{D \in U/R_d} \mathrm{Bel}_{C^1}(D) = \sum_{D \in U/R_d} \mathrm{Bel}_{C^2}(D) = 1,$

$\sum_{D \in U/R_d} \mathrm{Bel}_{C^3}(D) = 5/8 < 1,$

$\sum_{D \in U/R_d} \mathrm{Pl}_{C^1}(D) = \sum_{D \in U/R_d} \mathrm{Pl}_{C^2}(D) = 1,$

$\sum_{D \in U/R_d} \mathrm{Pl}_{C^3}(D) = 11/8 > 1,$

by Theorem 4, we conclude that the second scale is optimal in making decision d. On the other hand, it can be calculated that $R_{C^1} \subseteq R_d$ and $R_{C^2} \subseteq R_d$, however, $R_{C^3} \not\subseteq R_d$, therefore, the second is indeed the optimal scale of S.

Definition 9 *Let* $S = (U, C \cup \{d\}) = (U, \{a_j^k | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\} \cup \{d\})$ *be a consistent incomplete multi-scale decision table which has* I *levels of scales, for* $k \in \{1, 2, \ldots, I\}$*, we assume that* $S^k = (U, \{a_j^k | j = 1, 2, \ldots, m\} \cup \{d\})$ *is a consistent IDT. For* $B \subseteq C$*, if* $R_{B^k} \subseteq R_d$*, then* B *is referred to as a* k*-scale consistent set of* S*. If* B *is a* k*-scale consistent set of* S *and no proper subset of* B *is a* k*-scale consistent set of* S*, then* B *is referred to as a* k*-scale reduct of* S*, that is to say, a* k*-scale reduct of* S *is a minimal set of attributes* $B \subseteq C$ *such that* $R_{B^k} \subseteq R_d$*.*

Using Proposition 8, we can easily conclude the following:

Proposition 11 *Let* $S = (U, C \cup \{d\})$ *be a consistent incomplete multi-scale decision table. For* $1 \leq i < k \leq I$*, we assume that* $S^k = (U, \{a_j^k | j = 1, 2, \ldots, m\} \cup \{d\})$ *is a consistent IDT. If* $B \subseteq C$ *is a* k*-scale consistent set of* S*, then* B *is also an* i*-scale consistent set of* S*.*

Proposition 11 shows that if an attribute subset B is a consistent set of a coarser consistent IDT, then it must be a consistent set of a finer consistent IDT. This means that if an attribute set $B \subseteq C$ is a reduct of a coarser consistent IDT, then there must exist $E \subseteq B$ such that E is a reduct of a finer consistent IDT. Formally, we summarize the result as follows:

Proposition 12 *Let* $S = (U, C \cup \{d\}) = (U, \{a_j^k | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\} \cup \{d\})$ *be a consistent incomplete multi-scale decision table. For* $1 \leq i < k \leq I$*, we assume that* $S^k = (U, \{a_j^k | j = 1, 2, \ldots, m\} \cup \{d\})$ *is a consistent IDT. If* $B^k \subseteq C^k$ *is a reduct of* S^k*, then there exists an* $E \subseteq B$ *such that* E *is an* i*-scale reduct of* S*, i.e.,* E^i *is a reduct of* S^i*.*

After we select the optimal scale k, based on computing reducts of the kth IDT S^k, we can unravel the set of decision rules hidden in S.

Example 6 *In Example 4, since the second is the optimal scale for making decision, by using approaches in* [14,15,17,36,45]*, we can calculate that* $\{a_2^2, a_3^2\}$ *is the unique reduct of the consistent IDT* $S^2 = (U, \{a_1^2, a_2^2, a_3^2\} \cup \{d\})$*. Furthermore, based on the reduct* $\{a_2^2, a_3^2\}$*, we can obtain all local reducts for all decision classes, we then derive the certain decision rules from Table 4 as follows:*

$$
\begin{array}{llll}
r_1^2: (a_2^2, \mathrm{S}) \wedge & (a_3^2, \mathrm{M}) & \Longrightarrow (d, 1) & \textit{supported by } x_1, x_2 \\
r_2^2: & (a_3^2, \mathrm{L}) & \Longrightarrow (d, 2) & \textit{supported by } x_3, x_5, x_6 \\
r_3^2: & (a_3^2, \mathrm{S}) & \Longrightarrow (d, 2) & \textit{supported by } x_4, x_6 \\
r_4^2: (a_2^2, \mathrm{L}) & & \Longrightarrow (d, 2) & \textit{supported by } x_3, x_4, x_5, x_6, x_7, x_8
\end{array}
$$

According to Proposition 11*,* $\{a_2^1, a_3^1\}$ *is also a consistent set of* $S^1 = (U, \{a_1^1, a_2^1, a_3^1\} \cup \{d\})$*, it can be verified that* $\{a_2^1, a_3^1\}$ *is a reduct of* S^1*, by calculating all local reducts for all decision classes in* S^1*, we can derive the following certain decision rules from* S^1*:*

$$
\begin{aligned}
&r_1^1:\ (a_2^1,2)\wedge\ (a_3^1,3) \Longrightarrow (d,1) && \textit{supported by } x_1,x_2\\
&r_2^1: \qquad\qquad\quad (a_3^1,4) \Longrightarrow (d,2) && \textit{supported by } x_3,x_5,x_6\\
&r_3^1: \qquad\qquad\quad (a_3^1,1) \Longrightarrow (d,2) && \textit{supported by } x_4,x_6\\
&r_4^1:\ (a_2^1,5) \qquad\qquad\ \Longrightarrow (d,2) && \textit{supported by } x_3,x_4,x_8\\
&r_5^1:\ (a_2^1,4) \qquad\qquad\ \Longrightarrow (d,2) && \textit{supported by } x_3,x_5,x_6,x_7
\end{aligned}
$$

Using the same consistent set or reduct of IDTs at different scales, it can easily be verified that a set of decision rules derived from a coarser IDT are more general than that from a finer one. In Example 6, we can see that the decision rule r_4^2 is more general than any decision rule in $\{r_2^1, r_3^1, r_4^1, r_5^1\}$.

In fact, intuitively, we have the following granular transformation from the finer to the coarser scale:

$$
\begin{aligned}
(a_2^1,2)\wedge\ (a_3^1,3) &\xrightarrow{g_{12}} (a_2^2,\mathrm{S})\wedge\ (a_3^2,\mathrm{M})\\
(a_3^1,4) &\xrightarrow{g_{12}} (a_3^2,\mathrm{L})\\
(a_3^1,1) &\xrightarrow{g_{12}} (a_3^2,\mathrm{S})\\
(a_2^1,5) &\xrightarrow{g_{12}} (a_2^2,\mathrm{L})\\
(a_2^1,4) &\xrightarrow{g_{12}} (a_2^2,\mathrm{L})
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\|(a_2^1,2)\wedge(a_3^1,3)\| \subseteq \|(a_2^2,\mathrm{S})\wedge(a_3^2,\mathrm{M})\| &\subseteq \|(d,1)\|,\\
\|(a_3^1,4)\| \subseteq \|(a_3^2,\mathrm{L})\| &\subseteq \|(d,2)\|,\\
\|(a_3^1,1)\| \subseteq \|(a_3^2,\mathrm{L})\| &\subseteq \|(d,2)\|,\\
\|(a_2^1,5)\| \subseteq \|(a_2^2,\mathrm{L})\| &\subseteq \|(d,2)\|,\\
\|(a_2^1,4)\| \subseteq \|(a_2^2,\mathrm{L})\| &\subseteq \|(d,2)\|.
\end{aligned}
$$

4.2. *Rule acquisition in inconsistent incomplete multi-scale decision tables*

In this subsection we investigate knowledge acquisition in the sense of rule induction from an inconsistent incomplete multi-scale decision table.

Let $S = (U, C\cup\{d\}) = (U, \{a_j^k | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\}\cup\{d\})$ be an incomplete multi-scale decision table which has I levels of scales. For $1 \le i < k \le I$, if $(U, \{a_j^i | j = 1, 2, \ldots, m\}\cup\{d\})$ is an inconsistent IDT, then it can easily be observed that $(U, \{a_j^k | j = 1, 2, \ldots, m\}\cup\{d\})$ is also an inconsistent IDT.

For $k \in \{1, 2, \ldots, I\}$, and $X \subseteq U$, define

$$\underline{R_{C^k}}(X) = \{x\in U | S_{C^k}(x) \subseteq X\}, \quad \overline{R_{C^k}}(X) = \{x\in U | S_{C^k}(x)\cap X \neq \emptyset\}, \tag{47}$$

where

$$R_{C^k} = \{(x, y) \in U\times U | a^k(x) = a^k(y), \text{or } a^k(x) = *, \text{or } a^k(y) = *, \forall a \in C\}$$

and

$$S_{C^k}(x) = \{y \in U | (x, y) \in R_{C^k}\}.$$

Denote

$$\begin{aligned}
L_{C^k}(d) &= \big(\underline{R_{C^k}}(D_1), \underline{R_{C^k}}(D_2), \ldots, \underline{R_{C^k}}(D_r)\big),\\
H_{C^k}(d) &= \big(\overline{R_{C^k}}(D_1), \overline{R_{C^k}}(D_2), \ldots, \overline{R_{C^k}}(D_r)\big),\\
\mathrm{Bel}_{C^k}(d) &= \big(\mathrm{Bel}_{C^k}(D_1), \mathrm{Bel}_{C^k}(D_2), \ldots, \mathrm{Bel}_{C^k}(D_r)\big),\\
\mathrm{Pl}_{C^k}(d) &= \big(\mathrm{Pl}_{C^k}(D_1), \mathrm{Pl}_{C^k}(D_2), \ldots, \mathrm{Pl}_{C^k}(D_r)\big),\\
\partial_{C^k}(x) &= \{d(y) | y \in S_{C^k}(x)\},\ x \in U,
\end{aligned}$$

where $\mathrm{Bel}_{C^k}(D_j) = P(\underline{R_{C^k}}(D_j)) = \frac{|\underline{R_{C^k}}(D_j)|}{|U|}$, and $\mathrm{Pl}_{C^k}(D_j) = P(\overline{R_{C^k}}(D_j)) = \frac{|\overline{R_{C^k}}(D_j)|}{|U|}$, $j = 1, 2, \ldots, r$.

$L_{C^k}(d)$ and $H_{C^k}(d)$ are referred to as the *lower approximation distribution* and *upper approximation distribution* of decision classes U/R_d under the k-th scale in S, respectively. $\mathrm{Bel}_{C^k}(d)$ and $\mathrm{Pl}_{C^k}(d)$ are said to be the *belief distribution* and *plausibility distribution* of decision classes U/R_d under the k-th scale in S, respectively. And $\partial_{C^k}(x)$ is the *generalized decision values* of object x under the k-th scale in S. According to Proposition 8, it is easy to see that

$$\partial_{C^1}(x) \subseteq \partial_{C^2}(x) \subseteq \cdots \subseteq \partial_{C^{I-1}}(x) \subseteq \partial_{C^I}(x),\ x \in U. \tag{48}$$

Definition 10 *Let $S = (U, C \cup \{d\}) = (U, \{a_j^k | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\} \cup \{d\})$ be an inconsistent incomplete multi-scale decision table which has I levels of scales. For $k \in \{1, 2, \ldots, I\}$, we say that*

(1) $S^k = (U, C^k \cup \{d\}) = (U, \{a_j^k | j = 1, 2, \ldots, m\} \cup \{d\})$ *is lower approximation consistent to S if $L_{C^k}(d) = L_{C^1}(d)$. And, the kth level of scale is said to be the lower approximation optimal scale of S if S^k is lower approximation consistent to S and S^{k+1} (if there is $k+1$) is not lower approximation consistent to S.*

(2) $S^k = (U, C^k \cup \{d\}) = (U, \{a_j^k | j = 1, 2, \ldots, m\} \cup \{d\})$ *is upper approximation consistent to S if $H_{C^k}(d) = H_{C^1}(d)$. And, the kth level of scale is said to be the upper approximation optimal scale of S if S^k is upper approximation consistent to S and S^{k+1} (if there is $k+1$) is not upper approximation consistent to S.*

(3) $S^k = (U, C^k \cup \{d\}) = (U, \{a_j^k | j = 1, 2, \ldots, m\} \cup \{d\})$ *is belief consistent to S if* $\mathrm{Bel}_{C^k}(d) = \mathrm{Bel}_{C^1}(d)$. *And, the kth level of scale is said to be the belief optimal scale of S if S^k is belief consistent to S and S^{k+1} (if there is $k+1$) is not belief consistent to S.*

(4) $S^k = (U, C^k \cup \{d\}) = (U, \{a_j^k | j = 1, 2, \ldots, m\} \cup \{d\})$ *is plausibility consistent to S if* $\mathrm{Pl}_{C^k}(d) = \mathrm{Pl}_{C^1}(d)$. *And, the kth level of scale is said to be the plausibility optimal scale of S if S^k is plausibility consistent to S and S^{k+1} (if there is $k+1$) is not plausibility consistent to S.*

(5) $S^k = (U, C^k \cup \{d\}) = (U, \{a_j^k | j = 1, 2, \ldots, m\} \cup \{d\})$ *is generalized decision consistent to S if $\partial_{C^k}(x) = \partial_{C^1}(x)$ for all $x \in U$. And, the kth level of scale is said to be the generalized decision optimal scale of S if S^k is generalized decision consistent to S and S^{k+1} (if there is $k+1$) is not generalized decision consistent to S.*

In an inconsistent incomplete multi-scale decision table which has I levels of scales, it can be observed that $S^k = (U, \{a_j^k | j = 1, 2, \ldots, m\} \cup \{d\})$ is an inconsistent IDT for all $k \in \{1, 2, \ldots, I\}$. Moreover, we can see that

• S^k is lower approximation consistent to S iff S^k preserves the lower approximations of all decision classes of the finest scale IDT S^1, in this case, an object supports a certain decision rule derived from S^1 iff it supports a certain decision rule derived from S^k. And k is the lower approximation optimal scale of S iff k is the maximal number such that S^k preserves the lower approximations of all decision classes of S^1.

• S^k is upper approximation consistent to S iff S^k preserves the upper approximations of all decision classes of the finest scale IDT S^1, in this case, an object supports a possible rule derived from S^1 iff it supports a possible rule derived from S^k. And k is the upper approximation optimal scale of S iff k is the maximal number such that S^k preserves the upper approximations of all decision classes of S^1.

• S^k is belief consistent to S iff S^k preserves the same belief degree of each decision class in the finest scale IDT S^1. And k is the belief optimal scale of S iff k is the maximal number such that S^k preserves the same belief degree of each decision class in S^1.

• S^k is plausibility consistent to S iff S^k preserves the same plausibility degree of each decision class in the finest scale IDT S^1. And k is the plausibility optimal scale of S iff k is the maximal number such that S^k preserves the same plausibility degree of each decision class in S^1.

• S^k is generalized decision consistent to S iff S^k keeps the generalized decision values of the finest scale IDT S^1. And k is the generalized decision optimal scale of S iff k is the maximal number such that S^k keeps the generalized decision values of S^1.

It is important to clarify the interrelationships among the types of optimal scale in Definition 10.

Theorem 5 *Let $S = (U, C \cup \{d\}) = (U, \{a_j^k | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\} \cup \{d\})$ be an inconsistent incomplete multi-scale decision table which has I levels of scales. For $k \in \{1, 2, \ldots, I\}$, then the following statements are equivalent:*

(1) $L_{C^k}(d) = L_{C^1}(d)$.

(2) $\mathrm{Bel}_{C^k}(d) = \mathrm{Bel}_{C^1}(d)$.

(3) $\sum_{j=1}^r \mathrm{Bel}_{C^k}(D_j) = \sum_{j=1}^r \mathrm{Bel}_{C^1}(D_j)$.

Proof. "(1)⇒(2)". By the definitions, we have

$$\begin{aligned} L_{C^k}(d) = L_{C^1}(d) &\Longrightarrow \underline{R_{C^k}}(D_j) = \underline{R_{C^1}}(D_j),\ \forall j \in \{1, 2, \ldots, r\}, \\ &\Longrightarrow P(\underline{R_{C^k}}(D_j)) = P(\underline{R_{C^1}}(D_j)),\ \forall j \in \{1, 2, \ldots, r\}, \\ &\Longrightarrow \mathrm{Bel}_{C^k}(D_j) = \mathrm{Bel}_{C^1}(D_j),\ \forall j \in \{1, 2, \ldots, r\}, \\ &\Longrightarrow \mathrm{Bel}_{C^k}(d) = \mathrm{Bel}_{C^1}(d). \end{aligned}$$

"(2)⇒(3)". It is obvious.

"(3)⇒(1)". Since $\sum_{j=1}^r \mathrm{Bel}_{C^k}(D_j) = \sum_{j=1}^r \mathrm{Bel}_{C^1}(D_j)$, we have

$$\sum_{j=1}^{r} |\underline{R_{C^k}}(D_j)| = \sum_{j=1}^{r} |\underline{R_{C^1}}(D_j)|. \tag{49}$$

By Proposition 9, we observe that

$$\underline{R_{C^k}}(D_j) \subseteq \underline{R_{C^1}}(D_j),\ \forall j \in \{1, 2, \ldots, r\}, \tag{50}$$

then

$$|\underline{R_{C^k}}(D_j)| \leq |\underline{R_{C^1}}(D_j)|,\ \forall j \in \{1, 2, \ldots, r\}. \tag{51}$$

Hence, according to inequality (51), Eq. (49) implies that

$$|\underline{R_{C^k}}(D_j)| = |\underline{R_{C^1}}(D_j)|,\ \forall j \in \{1, 2, \ldots, r\}. \tag{52}$$

In terms of inclusion relation (50), we must have

$$\underline{R_{C^k}}(D_j) = \underline{R_{C^1}}(D_j),\ \forall j \in \{1, 2, \ldots, r\}. \tag{53}$$

It follows that $L_{C^k}(d) = L_{C^1}(d)$. □

Theorem 5 shows that the k-th level of scale is the lower approximation optimal scale of S iff it is the belief optimal scale of S. Moreover, we can easily conclude following:

Theorem 6 *Let $S = (U, C \cup \{d\}) = (U, \{a_j^k | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\} \cup \{d\})$ be an inconsistent incomplete multi-scale decision table which has I levels of scales. For $k \in \{1, 2, \ldots, I\}$, the k-th level of scale is the lower approximation optimal scale of S iff*

$$\sum_{j=1}^{r} \mathrm{Bel}_{C^k}(D_j) = \sum_{j=1}^{r} \mathrm{Bel}_{C^1}(D_j). \tag{54}$$

And (if there is $k + 1 \leq I$)

$$\sum_{j=1}^{r} \mathrm{Bel}_{C^{k+1}}(D_j) < \sum_{j=1}^{r} \mathrm{Bel}_{C^1}(D_j). \tag{55}$$

Theorem 7 *Let $S = (U, C \cup \{d\}) = (U, \{a_j^k | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\} \cup \{d\})$ be an inconsistent incomplete multi-scale decision table which has I levels of scales. For $k \in \{1, 2, \ldots, I\}$, then the following statements are equivalent:*

(1) $H_{C^k}(d) = H_{C^1}(d)$.
(2) $\mathrm{Pl}_{C^k}(d) = \mathrm{Pl}_{C^1}(d)$.
(3) $\sum_{j=1}^{r} \mathrm{Pl}_{C^k}(D_j) = \sum_{j=1}^{r} \mathrm{Pl}_{C^1}(D_j)$.
(4) $\partial_{C^k}(x) = \partial_{C^1}(x),\ \forall x \in U$.

Proof. "(1)⇒(2)". By the definitions, we have

$$\begin{aligned} H_{C^k}(d) = H_{C^1}(d) &\Longrightarrow \overline{R_{C^k}}(D_j) = \overline{R_{C^1}}(D_j),\ \forall j \in \{1, 2, \ldots, r\}, \\ &\Longrightarrow P(\overline{R_{C^k}}(D_j)) = P(\overline{R_{C^1}}(D_j)),\ \forall j \in \{1, 2, \ldots, r\}, \\ &\Longrightarrow \mathrm{Pl}_{C^k}(D_j) = \mathrm{Pl}_{C^1}(D_j),\ \forall j \in \{1, 2, \ldots, r\}, \\ &\Longrightarrow \mathrm{Pl}_{C^k}(d) = \mathrm{Pl}_{C^1}(d). \end{aligned}$$

"(2)⇒(3)". It is obvious.

"(3)⇒(1)". Since $\sum_{j=1}^{r} \mathrm{Pl}_{C^k}(D_j) = \sum_{j=1}^{r} \mathrm{Pl}_{C^1}(D_j)$, we have

$$\sum_{j=1}^{r} |\overline{R_{C^k}}(D_j)| = \sum_{j=1}^{r} |\overline{R_{C^1}}(D_j)|. \tag{56}$$

By Proposition 9, we see that

$$\overline{R_{C^1}}(D_j) \subseteq \overline{R_{C^k}}(D_j),\ \forall j \in \{1, 2, \ldots, r\}, \tag{57}$$

then we conclude that

$$|\overline{R_{C^1}}(D_j)| \leq |\overline{R_{C^k}}(D_j)|,\ \forall j \in \{1, 2, \ldots, r\}. \tag{58}$$

Hence, according to inequality (58), Eq. (56) implies that

$$|\overline{R_{C^k}}(D_j)| = |\overline{R_{C^1}}(D_j)|, \ \forall j \in \{1, 2, \ldots, r\}. \tag{59}$$

In terms of inclusion relation (57), we must have

$$\overline{R_{C^k}}(D_j) = \overline{R_{C^1}}(D_j), \ \forall j \in \{1, 2, \ldots, r\}. \tag{60}$$

It follows that $H_{C^k}(d) = H_{C^1}(d)$.

"(4)⇒(1)". Assume that $\partial_{C^k}(x) = \partial_{C^1}(x)$ for all $x \in U$. For any $D_j \in U/R_d$ and any $y \in \overline{R_{C^k}}(D_j)$, by the definition of upper approximation, we have $S_{C^k}(y) \cap D_j \neq \emptyset$. We select $u \in U$ such that $[u]_d = D_j$. Let $d(u) = w$, obviously, $d(u') = w$ for all $u' \in D_j$. Then we can find $z \in S_{C^k}(y)$ such that $z \in [u]_d$. Clearly, $d(z) = w \in \partial_{C^k}(y)$. Since $\partial_{C^k}(x) = \partial_{C^1}(x)$ for all $x \in U$, we have $\partial_{C^k}(y) = \partial_{C^1}(y)$. Thus $w \in \partial_{C^1}(y)$, consequently, we can find $y' \in S_{C^1}(y)$ such that $d(y') = w$. It is easy to observe that $y' \in [u]_d$. Hence we conclude $S_{C^1}(y) \cap [u]_d \neq \emptyset$, by the definition of upper approximation, we deduce $y \in \overline{R_{C^1}}(D_j)$. It follows that

$$\overline{R_{C^k}}(D_j) \subseteq \overline{R_{C^1}}(D_j), \quad \forall D_j \in U/R_d. \tag{61}$$

On the other hand, by Proposition 9, we have

$$\overline{R_{C^1}}(D_j) \subseteq \overline{R_{C^k}}(D_j), \quad \forall D_j \in U/R_d. \tag{62}$$

Combining inclusion relations (61) and (62), we conclude $\overline{R_{C^k}}(D_j) = \overline{R_{C^1}}(D_j)$ for all $D_j \in U/R_d$, that is, $H_{C^k}(d) = H_{C^1}(d)$.

"(1)⇒(4)". Assume that $H_{C^k}(d) = H_{C^1}(d)$, that is,

$$\overline{R_{C^k}}(D_j) = \overline{R_{C^1}}(D_j), \ \forall D_j \in U/R_d. \tag{63}$$

For any $x \in U$ and $w \in \partial_{C^k}(x)$, there exists $y \in S_{C^k}(x)$ such that $w = d(y)$. Let $D_w = \{z \in U | d(z) = w\}$. Clearly, $D_w \in U/R_d$ and $y \in D_w$. Hence

$$S_{C^k}(x) \cap D_w \neq \emptyset, \tag{64}$$

by the definition of upper approximation, we have $x \in \overline{R_{C^k}}(D_w)$. Since $\overline{R_{C^k}}(D_w) = \overline{R_{C^1}}(D_w)$, we obtain $x \in \overline{R_{C^1}}(D_w)$, by the definition of upper approximation again, we conclude $S_{C^1}(x) \cap D_w \neq \emptyset$, consequently, there exists $z \in S_{C^1}(x)$ such that $z \in D_w$. Obviously, $d(z) = w$, thus $w \in \partial_{C^1}(x)$. It follows that

$$\partial_{C^k}(x) \subseteq \partial_{C^1}(x). \tag{65}$$

On the other hand, by Proposition 8, we have $S_{C^1}(x) \subseteq S_{C^k}(x)$. Hence

$$\partial_{C^1}(x) \subseteq \partial_{C^k}(x). \tag{66}$$

Combining inclusion relations (65) and (66), we conclude that $\partial_{C^k}(x) = \partial_{C^1}(x)$ for all $x \in U$. □

Theorem 7 shows that, in an inconsistent incomplete multi-scale decision table, the kth level of scale is the upper approximation optimal scale iff it is the plausibility optimal scale iff it is the generalized decision optimal scale, in other words, all the upper approximation optimal scale, the plausibility optimal scale, and the generalized decision optimal scale are the same. Similar to Theorem 6, we have following:

Table 6
An inconsistent incomplete multi-scale decision table

U	a_1^1	a_1^2	a_1^3	a_2^1	a_2^2	a_2^3	a_3^1	a_3^2	a_3^3	a_4^1	a_4^2	a_4^3	d	∂_{C^1}
x_1	1	S	Y	1	E	Y	3	G	Y	4	L	N	1	{1}
x_2	2	S	Y	1	E	Y	2	G	Y	4	L	N	1	{1}
x_3	3	M	Y	3	G	Y	2	G	Y	3	M	Y	1	{1}
x_4	3	M	Y	2	G	Y	2	G	Y	3	M	Y	1	{1}
x_5	$*$	$*$	$*$	2	G	Y	2	G	Y	3	M	Y	1	{1, 2}
x_6	$*$	$*$	$*$	3	G	Y	2	G	Y	3	M	Y	1	{1, 2}
x_7	4	L	N	2	G	Y	$*$	$*$	$*$	3	M	Y	2	{1, 2}
x_8	5	L	N	3	G	Y	$*$	$*$	$*$	3	M	Y	2	{1, 2}
x_9	6	L	N	2	G	Y	1	E	Y	2	S	Y	2	{2}
x_{10}	6	L	N	3	G	Y	1	E	Y	1	S	Y	2	{2}

Theorem 8 *Let $S = (U, C \cup \{d\}) = (U, \{a_j^k | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\} \cup \{d\})$ be an inconsistent incomplete multi-scale decision table which has I levels of scales. For $k \in \{1, 2, \ldots, I\}$, then the kth level of scale is the generalized decision optimal scale of S iff*

$$\sum_{j=1}^{r} \mathrm{Pl}_{C^k}(D_j) = \sum_{j=1}^{r} \mathrm{Pl}_{C^1}(D_j). \tag{67}$$

And (if there is $k + 1 \leq I$)

$$\sum_{j=1}^{r} \mathrm{Pl}_{C^{k+1}}(D_j) > \sum_{j=1}^{r} \mathrm{Pl}_{C^1}(D_j). \tag{68}$$

Definition 11 *Let $S = (U, C \cup \{d\}) = (U, \{a_j^k | k = 1, 2, \ldots, I, j = 1, 2, \ldots, m\} \cup \{d\})$ be an inconsistent incomplete multi-scale decision table which has I levels of scales. Let $k \in \{1, 2, \ldots, I\}$, for $B \subseteq C$, if $\partial_{B^k}(x) = \partial_{C^1}(x)$ for all $x \in U$, then B is referred to as a k-scale consistent set of S. If B is a k-scale consistent set of S and no proper subset of B is a k-scale consistent set of S, then B is referred to as a k-scale reduct of S.*

If B is a k-scale reduct of S, then it is easy to see that $S^k = (U, \{a_j^k | j = 1, 2, \ldots, m\} \cup \{d\})$ must be consistent to S. According to Definition 11, we see that $B \subseteq C$ is a k-scale reduct of S iff it is a minimal set of attributes to keep the generalized decisions of full attributes at the k-th level of scale, i.e. $\partial_{B^k}(x) = \partial_{C^k}(x)$ for all $x \in U$.

Example 7 *Table 6 depicts an example of an inconsistent incomplete multi-scale decision table $S = (U, \{a_j^k | k = 1, 2, 3, j = 1, 2, 3, 4\} \cup \{d\})$, where $U = \{x_1, x_2, \ldots, x_{10}\}$, $C = \{a_1, a_2, a_3, a_4\}$. The table has three levels of scales, where "S", "M", "L", "E", "G", "Y", and "N" stand for, respectively, "Small", "Medium", "Large", "Excellent", "Good", "Yes", and "No". For these levels of scales, the table is associated with three inconsistent IDTs which are depicted as Tables 7–9, respectively.*

It is easy to observe that $D_1 = \{x \in U | d(x) = 1\} = \{x_1, x_2, \ldots, x_6\}$ and $D_2 = \{x \in U | d(x) = 2\} = \{x_7, x_8, x_9, x_{10}\}$. Then it can be calculated that

$$\mathrm{Bel}_{C^1}(D_1) + \mathrm{Bel}_{C^1}(D_2) = \mathrm{Bel}_{C^2}(D_1) + \mathrm{Bel}_{C^2}(D_2) = 6/10,$$

Table 7
The incomplete decision table at the first level of scale

U	a_1^1	a_2^1	a_3^1	a_4^1	d	∂_{C^1}
x_1	1	1	3	4	1	{1}
x_2	2	1	2	4	1	{1}
x_3	3	3	2	3	1	{1}
x_4	3	2	2	3	1	{1}
x_5	*	2	2	3	1	{1, 2}
x_6	*	3	2	3	1	{1, 2}
x_7	4	2	*	3	2	{1, 2}
x_8	5	3	*	3	2	{1, 2}
x_9	6	2	1	2	2	{2}
x_{10}	6	3	1	1	2	{2}

Table 8
The incomplete decision table at the second level of scale

U	a_1^2	a_2^2	a_3^2	a_4^2	d	∂_{C^2}
x_1	S	E	G	L	1	{1}
x_2	S	E	G	L	1	{1}
x_3	M	G	G	M	1	{1}
x_4	M	G	G	M	1	{1}
x_5	*	G	G	M	1	{1, 2}
x_6	*	G	G	M	1	{1, 2}
x_7	L	G	*	M	2	{1, 2}
x_8	L	G	*	M	2	{1, 2}
x_9	L	G	E	S	2	{2}
x_{10}	L	G	E	S	2	{2}

and

$$\mathrm{Bel}_{C^3}(D_1) + \mathrm{Bel}_{C^3}(D_2) = 4/10 < 6/10.$$

Hence, by Theorem 6, we see that $k = 2$ is the lower approximation optimal scale of S.

Similarly, since

$$\mathrm{Pl}_{C^1}(D_1) + \mathrm{Pl}_{C^1}(D_2) = \mathrm{Pl}_{C^2}(D_1) + \mathrm{Pl}_{C^2}(D_2) = 14/10,$$

and

$$\mathrm{Pl}_{C^3}(D_1) + \mathrm{Pl}_{C^3}(D_2) = 16/10 > 14/10,$$

by Theorem 8, we conclude that $k = 2$ is the generalized decision optimal scale of S. We can see directly from Tables 7-9 that $\partial_{C^2}(x) = \partial_{C^1}(x)$ for all $x \in U$, and $\partial_{C^3}(x_9) \neq \partial_{C^1}(x_9)$. So S^3 is not consistent to S and the second level of scale is optimal for making decision of S.

By using approaches in [17, 45], we can conclude that the inconsistent IDT $S^2 = (U, C^2 \cup \{d\})$ has two generalized decision reducts $B_1^2 = \{a_1^2, a_4^2\}$ and $B_2^2 = \{a_1^2, a_3^2\}$.

Table 9
The incomplete decision table at the third level of scale

U	a_1^3	a_2^3	a_3^3	a_4^3	d	∂_{C^3}
x_1	Y	Y	Y	N	1	{1}
x_2	Y	Y	Y	N	1	{1}
x_3	Y	Y	Y	Y	1	{1}
x_4	Y	Y	Y	Y	1	{1}
x_5	*	Y	Y	Y	1	{1, 2}
x_6	*	Y	Y	Y	1	{1, 2}
x_7	N	Y	*	Y	2	{1, 2}
x_8	N	Y	*	Y	2	{1, 2}
x_9	N	Y	Y	Y	2	{1, 2}
x_{10}	N	Y	Y	Y	2	{1, 2}

On the other hand, for $B \subseteq C$, similar to Eq. (22) in Subsection 2.3, we define an equivalence relation $R_d^{B^2}$ on U as follows:

$$R_d^{B^2} = \{(x, y) \in U \times U | \partial_{B^2}(x) = \partial_{B^2}(y)\}. \tag{69}$$

Then, according to Subsection 2.3, we can see that $(U, C^2 \cup \{\partial_{C^2}\}) = (U, C^2 \cup \{\partial_{C^1}\})$ is a consistent IDT. It can be observed that an attribute subset $B^2 \subseteq C^2$ is a generalized decision consistent set of the inconsistent IDT $(U, C^2 \cup \{d\})$ iff B^2 is a consistent set of the consistent IDT $(U, C^2 \cup \{\partial_{C^1}\})$. Hence, $B^2 \subseteq C^2$ is a generalized decision reduct of the inconsistent IDT $(U, C^2 \cup \{d\})$ iff B is a reduct of the consistent IDT $(U, C^2 \cup \{\partial_{C^1}\})$. That is, we can find generalized decision reducts of $(U, C^2 \cup \{d\})$ via calculating reducts of $(U, C^2 \cup \{\partial_{C^1}\})$. And, by employing the approach in Subsection 4.1, we can calculate all reducts of $(U, C^2 \cup \{\partial_{C^1}\})$. In fact, similar to Subsection 4.1, we can calculate that the consistent IDT $(U, C^2 \cup \{\partial_{C^1}\})$ has two reducts $B_1^2 = \{a_1^2, a_4^2\}$ and $B_2^2 = \{a_1^2, a_3^2\}$, thus, the inconsistent IDT $S^2 = (U, C^2 \cup \{d\})$ has two generalized decision reducts $B_1^2 = \{a_1^2, a_4^2\}$ and $B_2^2 = \{a_1^2, a_3^2\}$.

Based on the generalized decision reduct $\{a_1^2, a_4^2\}$, and by further calculating all local generalized decision reducts of objects, we can derive decision rules hidden in the inconsistent IDT $(U, C^2 \cup \{d\})$ as follows:

Certain rules*:*

$$r_1^2 : (a_1^2, \mathrm{S}) \Longrightarrow (d, 1) \text{ with certainty 1, supported by } x_1, x_2, x_5, x_6$$
$$r_2^2 : (a_4^2, \mathrm{L}) \Longrightarrow (d, 1) \text{ with certainty 1, supported by } x_1, x_2$$
$$r_3^2 : (a_1^2, \mathrm{M}) \Longrightarrow (d, 1) \text{ with certainty 1, supported by } x_3, x_4, x_5, x_6$$
$$r_4^2 : (a_4^2, \mathrm{S}) \Longrightarrow (d, 2) \text{ with certainty 1, supported by } x_9, x_{10}$$

Possible rules*:*

$$r_5^2 : (a_1^2, \mathrm{L}) \wedge (a_4^2, \mathrm{M}) \longrightarrow (d, 1) \text{ with certainty 1/2, supported by } x_5, x_6$$
$$r_5^{2\prime} : (a_1^2, \mathrm{L}) \wedge (a_4^2, \mathrm{M}) \longrightarrow (d, 2) \text{ with certainty 1/2, supported by } x_7, x_8.$$

It can be verified that $\{a_1^1, a_4^1\}$ is also a reduct of $S^1 = (U, C^1 \cup \{d\})$, then, by using $\{a_1^1, a_4^1\}$, we can further calculate all local generalized decision reducts of all objects in $S^1 = (U, C^1 \cup \{d\})$, therefore, we can derive the decision rules in the inconsistent IDT $(U, C^1 \cup \{d\})$ (at the first level of scale) as follows:

Certain rules*:*

$r_1^1 : (a_1^1, 1) \Longrightarrow (d, 1)$ *with certainty 1, supported by* x_1, x_5, x_6

$r_2^1 : (a_1^1, 2) \Longrightarrow (d, 1)$ *with certainty 1, supported by* x_2, x_5, x_6

$r_3^1 : (a_1^1, 3) \Longrightarrow (d, 1)$ *with certainty 1, supported by* x_3, x_4, x_5, x_6

$r_4^1 : (a_4^1, 4) \Longrightarrow (d, 1)$ *with certainty 1, supported by* x_1, x_2

$r_5^1 : (a_4^1, 2) \Longrightarrow (d, 2)$ *with certainty 1, supported by* x_9

$r_6^1 : (a_4^1, 1) \Longrightarrow (d, 2)$ *with certainty 1, supported by* x_{10}

Possible rules*:*

$r_7^1 : (a_1^1, 4) \longrightarrow (d, 1)$ *with certainty 2/3, supported by* x_5, x_6

$r_7^{1\prime} : (a_1^1, 4) \longrightarrow (d, 2)$ *with certainty 1/3, supported by* x_7

$r_8^1 : (a_1^1, 5) \longrightarrow (d, 1)$ *with certainty 2/3, supported by* x_5, x_6

$r_8^{1\prime} : (a_1^1, 5) \longrightarrow (d, 2)$ *with certainty 1/3, supported by* x_8

We can see that, based on the same reduct $\{a_1, a_4\}$, the set of decision rules derived from $(U, C^2 \cup \{d\})$ are more general than that from $(U, C^1 \cup \{d\})$.

5. Conclusion

In traditional rough-set-data-analysis, each object under each attribute in information tables can only take on one value. However, in many real-life applications, objects are usually measured at different scales under the same attribute, i.e. an object can take on as many values as there are scales under the same attribute, on the other hand, the precise values of some attributes for some objects may be missing. Such a system is called an incomplete multi-scale information table. We have developed in this paper a general framework for the study of knowledge acquisition in incomplete multi-scale decision tables from the perspective of granular computing. We have analyzed information granules with reference to different levels of scales in incomplete multi-scale information tables. We have also defined lower and upper approximations at different levels of scales in incomplete multi-scale information tables and examined their properties. We have further discussed optimal scale selection with various requirements in incomplete multi-scale decision tables and have presented relationships of different notions of optimal scales. With reference to different levels of scales and by using rough set approach, we have explored knowledge acquisition in the sense of rule induction in consistent and inconsistent incomplete multi-scale decision tables. For further study, new approaches to granular representation and new models for knowledge acquisition in complicated multi-scale information tables need to be formulated.

Acknowledgement

This work was supported by grants from the National Natural Science Foundation of China (Nos. 61573321, 61272021, 61202206, 61173181, and 61322211), and the Zhejiang Provincial Natural Science Foundation of China (Nos. LZ12F03002 and LY14F030001).

References

[1] C.L.P. Chen, C.Y. Zhang, Data-intensive applications, challenges, techniques and technologies: A survey on Big Data, Information Sciences 275 (2014) 314–347.

[2] G. Chiaselotti, D. Ciucci, T. Gentile, Simple graphs in granular computing, Information Sciences 340-341 (2016) 279–304.

[3] P.G. Clark, J. Grzymala-Busse, W. Rzasa, Mining incomplete data with singleton, subset and concept probabilistic approximations, Information Sciences 280 (2014) 368–384.

[4] P.G. Clark, J. Grzymala-Busse, W. Rzasa, Consistency of incomplete data, Information Sciences 322 (2015) 197–222.

[5] S. Demri, E. Orlowska, Incomplete Information: Structure, Inferenece, Complexity, Springer Verlag, Heidelberg, 2002.

[6] S. Greco, B. Matarazzo, R. Slowinski, Handling missing values in rough set analysis of multi-attribute and multi-criteria decision problems, in: N. Zhong, A. Skowron, S. Ohsuda (Eds.), New Directions in Rough Sets, Data Mining, and Granular-soft Computing, Lecture Notes in Artificial Intelligence 1711, Springer-Verlag, Berlin, 1999, pp. 146–157.

[7] J. Grzymala-Busse, A rough set approach to data with missing attribute values, in: G.Y. Wang, J.F. Peters, A. Skowron, Y.Y. Yao (Eds.), The 1st International Conference of Rough Sets and Knowledge Technology, Lecture Notes in Artificial Intelligence 4062, Springer, Berlin, 2006, pp. 58–67.

[8] S.M. Gu, W.Z. Wu, On knowledge acquisition in multi-scale decision systems, International Journal of Machine Learning and Cybernetics 4 (2013) 477–486.

[9] A.E. Hassanien, Rough set approach for attribute reduction and rule generation: A case of patients with suspected breast cancer, Journal of the American Society for Information Science and Technology 55 (2004) 954–962.

[10] T.P. Hong, L.H. Tseng, S.L. Wang, Learning rules from incomplete training examples by rough sets, Expert Systems with Applications 22 (2002) 285–293.

[11] Q.H. Hu, J.F. Liu, D.R. Yu, Mixed feature selection based on granulation and approximation, Knowledge-Based Systems 21 (2008) 294–304.

[12] M. Inuiguchi, S. Hirano, S. Tsumoto (Eds.), Rough Set Theory and Granular Computing, Springer, Berlin, 2002.

[13] J. Kerr-Wilson, W. Pedrycz, Design of rule-based models through information granulation, Expert Systems with Applications 46 (2016) 274–285.

[14] M. Kryszkiewicz, Rough set approach to incomplete information systems, Information Sciences 112 (1998) 39–49.

[15] M. Kryszkiewicz, Rules in incomplete information systems, Information Sciences 113 (1999) 271–292.

[16] Y. Leung, D.Y. Li, Maximal consistent block technique for rule acquisition in incomplete information systems, Information Sciences 153 (2003) 85–106.

[17] Y. Leung, W.Z. Wu, W.X. Zhang, Knowledge acquisition in incomplete information systems: A rough set approach, European Journal of Operational Research 168 (2006) 164–180.

[18] Y. Leung, J.S. Zhang, Z.B. Xu, Clustering by scale-space filtering, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (2000) 1396–1410.

[19] S.Y. Li, T.R. Li, Incremental update of approximations in dominance-based rough sets approach under the variation of attribute values, Information Sciences 294 (2015) 348–361.

[20] H.X. Li, M.H. Wang, X.Z. Zhou, J.B. Zhao, An interval set model for learning rules from incomplete information table, International Journal of Approximate Reasoning 53 (2012) 24–37.

[21] J.Y. Liang, Z.B. Xu, The algorithm on knowledge reduction in incomplete information systems, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10 (2002) 95–103.

[22] G.P. Lin, J.Y. Liang, Y.H. Qian, An information fusion approach by combining multigranulation rough sets and evidence theory, Information Sciences 314 (2015) 184–199.
[23] T.Y. Lin, Y.Y. Yao, L.A. Zadeh (Eds.), Data Mining, Rough Sets and Granular Computing, Physica-Verlag, Heidelberg, 2002.
[24] P.J. Lingras, Y.Y. Yao, Data mining using extensions of the rough set model, Journal of the American Society for Information Science 49 (1998) 415–422.
[25] W. Lipski, On databases with incomplete information, Journal of the ACM 26 (1981) 41–70.
[26] J.S. Mi, W.Z. Wu, W.X. Zhang, Approaches to knowledge reduction based on variable precision rough sets model, Information Sciences 159 (2004) 255–272.
[27] E. Orlowska (Ed.), Incomplete Information: Rough Set Analysis, Physica-Verlag, Heidelberg, 1998.
[28] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Boston, 1991.
[29] Z. Pawlak, Drawing conclusions from data—the rough set way, International Journal of Intelligent Systems 16 (2001) 3–11.
[30] W. Pedrycz (Ed.), Granular Computing: An Emerging Paradigm, Physica-Verlag, Heidelberg, 2001.
[31] W. Pedrycz, Granular Computing: Analysis and Design of Intelligent Systems, CRC Press/Francis Taylor, Boca Raton, 2013.
[32] W. Pedrycz, A. Skowron, V. Kreinovich (Eds.), Handbook of Granular Computing, Wiley, 2008.
[33] W. Pedrycz, G. Succi, A. Sillitti, J. Iljazi, Data description: A general framework of information granules, Knowledge-Based Systems 80 (2015) 98–108.
[34] Y.H. Qian, J.Y. Liang, C.Y. Dang, Knowledge structure, knowledge granulation and knowledge distance in a knowledge base, International Journal of Approximate Reasoning 50 (2009) 174–188.
[35] Y.H. Qian, J.Y. Liang, C.Y. Dang, Incomplete multigranulation rough set, IEEE Trasactions on Systems, Man and Cybernetics—Part A 40 (2010) 420–431.
[36] Y.H. Qian, J.Y. Liang, D.Y. Li, F. Wang, N.N. Ma, Approximation reduction in inconsistent incomplete decision tables, Knowledge-Based Systems 23 (2010) 427–433.
[37] Y.H. Qian, J.Y. Liang, Y.Y. Yao, C.Y. Dang, MGRS: A multi-granulation rough set, Information Sciences 180 (2010) 949–970.
[38] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, 1976.
[39] A. Skowron, The relationship between rough set theory and evidence theory, Bulletin of the Polish Academy of Sciences: Mathematics 37 (1989) 87–90.
[40] A. Skowron, The rough sets theory and evidence theory, Fundamenta Informaticae 13 (1990) 245–262.
[41] A. Skowron, J. Grzymala-Busse, From rough set theory to evidence theory, in: R.R. Yager, M. Fedrizzi, J. Kacprzyk (Eds.), Advance in the Dempster-Shafer Theory of Evidence, Wiley, New York, 1994, pp. 193–236.
[42] A. Skowron, C. Rauszer, The discernibility matrices and functions in information systems, in: R. Slowinski (Ed.), Intelligent Decision Support—Handbook of Applications and Advances of the Rough Sets Theory, Kluwer Academic Publishers, Boston, 1992, pp. 331–362.
[43] R. Slowinski, J. Stefanowski, Rough classification in incomplete information systems, Mathematical and Computer Modelling 12 (1989) 1347–1357.
[44] R. Ureña, F. Chiclana, J.A. Morente-Molinera, E. Herrera-Viedma, Managing incomplete preference relations in decision making: A review and future trends, Information Sciences 302 (2015) 14–32.
[45] W.Z. Wu, Attribute reduction based on evidence theory in incomplete decision systems, Information Sciences 178 (2008) 1355–1371.
[46] W.Z. Wu, Y. Leung, Theory and applications of granular labelled partitions in multi-scale decision tables, Information Sciences 181 (2011) 3878–3897.
[47] W.Z. Wu, Y. Leung, Optimal scale selection for multi-scale decision tables, International Journal of Approximate Reasoning 54 (2013) 1107–1129.
[48] W.Z. Wu, Y. Leung, J.S. Mi, Granular computing and knowledge reduction in formal contexts, IEEE Transactions on Knowledge and Data Engineering 21 (2009) 1461–1474.
[49] W.Z. Wu, Y. Leung, W.X. Zhang, Connections between rough set theory and Dempster-Shafer theory of evidence, International Journal of General Systems 31 (2002) 405–430.
[50] X.P. Yang, Z.J. Lu, T.J. Li, Decision-theoretic rough sets in incomplete information system, Fundamenta Informaticae 126 (2013) 353–375.
[51] X.B. Yang, X.N. Song, Z.H. Chen, J.Y. Yang, On multigranulation rough sets in incomplete information system, International Journal of Machine Learning and Cybernetics 3 (2012) 223–232.

[52] J.T. Yao, Recent developments in granular computing: A bibliometrics study, in: Proceedings of IEEE International Conference on Granular Computing, Hangzhou, China, Aug 26–28, 2008, pp. 74–79.

[53] J.T. Yao, A.V. Vasilakos, W. Pedrycz, Granular computing: Perspectives and challenges, IEEE Transactions on Cybernetics 43 (2013) 1977–1989.

[54] Y.Y. Yao, Information granulation and rough set approximation, International Journal of Intelligent Systems 16 (2001) 87–104.

[55] Y.Y. Yao, Three-way decisions with probabilistic rough sets, Information Sciences 180 (2010) 341–353.

[56] Y.Y. Yao, Generalized rough set model, in: L. Polkowski, A. Skowron (Eds), Rough Sets in Knowledge Discovery 1. Methodology and Applications, Physica-Verlag, Heidelberg, 1998, pp. 286–318.

[57] Y.Y. Yao, P.J. Lingras, Interpretations of belief functions in the theory of rough sets, Information Sciences 104 (1998) 81–106.

[58] Y.Y. Yao, Y.H. She, Rough set models in multigranulation spaces, Information Sciences 327 (2016) 40–56.

[59] L.A. Zadeh, Fuzzy sets and information granularity, in: N. Gupta, R. Ragade, R.R. Yager (Eds.), Advances in Fuzzy Set Theory and Applications, North-Holland, Amsterdam, 1979, pp. 3–18.

[60] L.A. Zadeh, Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, Fuzzy Sets and Systems 90 (1997) 111–127.

[61] X.Y. Zhang, D.Q. Miao, Quantitative/qualitative region-change uncertainty/certainty in attribute reduction: Comparative region-change analyses based on granular computing, Information Sciences 334-335 (2016) 174–204.

[62] Y.Q. Zhang, Constructive granular systems with universal approximation and fast knowledge discovery, IEEE Transactions on Fuzzy Systems 13 (2005) 48–57.

[63] Y. Zhao, Y.Y. Yao, F. Luo, Data analysis based on discernibility and indiscernibility, Information Sciences 177 (2007) 4959–4976.