Definability of concept in incomplete information systems

Shunliang Huang*

Business School, Shandong University of Technology, No.88, Gongqingtuan Rd., Zibo 255000, China

Received 1 August 2014, www.cmnt.lv

Abstract

An incomplete information table (a set) can be expressed as a family of complete information tables (sets). The family of complete information sets maybe constructs an interval set. A concept in incomplete information situation, called partially known concept, is said to be definable if its extension can be expressed as an interval set. The new definition of definability proposed in this paper, named interval definable, is different from its usual meaning in the rough sets theory where a concept is definable means that its extension is a definable set, which is the union of some equivalence classes. The new definition of definability not only provides a new interpretation of interval sets, but also endows more general meaning and deeper understanding of definability.

Keywords: incomplete information, definability, concept representation, interval sets, uncertainty

1 Introduction

Definability, in rough sets theory [1, 2], is an important notion. In many studies, the notion of definability is introduced in two ways through equivalence classes and approximations [3], respectively. In general, definability and approximation can be defined each other. A set is said to be definable if it is the union of some equivalence classes. Alternatively, some authors considered the definability of a set based on its approximations. A set is said to be definable if its lower and upper approximations are equal [2, 4-7]. The two definitions of definability are equivalent. Yao [3] has investigated explicitly the notion of definability in the framework of rough sets theory and given a new interpretation. In the above mentioned studies, for an incomplete information table, the subsets of the incomplete information table are usually indefinable. In order to describe these subsets, many new theories are introduced such as rough sets theory, interval sets theory [8-10], etc. Interval sets, as a new set theory introduced, is used to describe partially known concept. In practice, partially known concept can be described by incomplete information table. By Lipski's model [11, 12], an incomplete information table can be expressed as a family of complete tables. This means that a subset of an incomplete information table can be expressed as a family of sets, which are the subsets of those complete tables, respectively. A family of sets maybe constructs an interval set. From this point of view, the notion of definability can be defined in a different way, namely, if an incomplete information table can be expressed as an interval set, then the information table is definable. As a more primitive notion, this paper not only provides a different viewpoint with the interpretation of interval sets, but also endows more general meaning and deeper understanding the notion of definability.

The rest of this paper is organized as follows. In section 2, definability of concept in complete information is examined. The notion of meaning triangle [13] is introduced to interpret the relation between the intension and the extension of a concept. A logic language, which is used by Pawlak [2], is adopted for the representation of the intension of a concept. In section 3, a new definition of definability is given. A sub-language of the above mentioned logic language is defined to describe the incomplete information. Based on the logic language, a constructive method of interval sets is introduced, and it is proved that an incomplete information table can be expressed as an interval set, namely, is definable in sense of new definition of definability. Finally, some concluding remarks are given.

2 Definability of concept in complete information

In the classical view, every concept is understood as a unit of thought that consists of two parts, namely, the intension and the extension of the concept [14, 15]. The intension of a concept specifies the necessary and sufficient conditions for a thing being a member of a specific set. The extension of a concept is a list naming every object that is a member of a specific set. The name, intension, and extension of a concept, as three vertexes, construct a triangle just like meaning triangle proposed by Ogden and Richards [13]. The classical view of concepts enables us to study concepts in a logic setting in terms of intension and also in a set-theoretic setting in terms of extension [16]. In a logic language, the intension of a concept can be denoted by a logic formula, and the extension is a set corresponding with the logic formula. By logic language, we can describe an information table and study the representation, interpretation and processing of information.

* ***Corresponding author*** e-mail: huangshunliang@163.com

Definition 1 Let U be a finite and nonempty set, called the universe, an information table T can be expressed as a quadruple:

$$T=(U,AT,\{V_a \mid a\in AT\},\{f_a \mid a\in AT\}), \quad (1)$$

where AT is a finite set of attributes, V_a is a nonempty set of values for an attribute $a\in AT$, and $f_a:U\rightarrow V_a$ is an information or a description function. It is assumed that the mapping f_a is single-valued. In this case, an information table is called complete information table.

One can define a logic language to express the intension of a concept as a equation of the language. We adopt the decision logic language L used and studied by Pawlak [2]. Equations of L are recursively constructed based on a set of atomic equations corresponding to some basic concepts.

(i) $a=v$ is an atomic equation, where $a\in AT$, $v\in V_a$.

(ii) If p and q are equations in L, respectively, then $\neg p, p\wedge q, p\vee q$ are also equations in L.

Given an information table $T=(U,AT,V_a,f_a)$, $f_a(x)=v$, written as $p=(a=v)$, is an equation. If object $x\in U$ satisfies p, then it can be denoted by $x\models p$. For two equations p and q, the semantics of the logic language can be defined as follows:

(i) $x\models p$, iif $f_a(x)=v$; (ii) $x\models p\wedge q$, iff $x\models p$ and $x\models q$; (iii) $x\models p\vee q$, iff $x\models p$ or $x\models q$; (iv) $x\models\neg p$, iff not $x\models p$; (v) $x\models p\rightarrow q$, iff $x\models\neg p$ and $x\models q$; (vi) $x\models p\leftrightarrow q$, iff $x\models p\rightarrow q$ and $x\models q\rightarrow p$.

Definition 2: let p be an equation of L. With respect to an information table $T=(U,AT,V_a,f_a)$, the set of all the objects that satisfy the equation p, denoted by $m(p)$, is called meaning set of the logic equation.

The meaning set of an equation is a subset of U, and can be denoted as follows:

$$m(p)=\{x\in U \mid x\models p\}. \quad (2)$$

The meaning set $m(p)$ includes all objects expressed by the formula p. From the point of view of concept, p represents the intension of a concept, and $m(p)$ represents the extension of a concept. Thus, a pair $(p,m(p))$ represents a concept. Obviously, the following proposition holds [16].

Proposition 1: Let p and q be two atomic formulas of the logic language L in a complete information table, the logic connectives and set-theoretic operators can be expressed by each other as follows:

$$\begin{aligned}
&(\text{i})m(\neg p)=U-m(p);\\
&(\text{ii})m(p\wedge q)=m(p)\cap m(q);\\
&(\text{iii})m(p\vee q)=m(p)\cup m(q);\\
&(\text{iv})m(p\rightarrow q)=m(\neg p)\cup m(q);\\
&(\text{v})m(p\leftrightarrow q)=m(p\rightarrow q)\cap m(q\rightarrow p).
\end{aligned} \quad (3)$$

With the introduction of language L, we can discuss the definability of sets with a formal description of concepts.

Definition 3 Let $T=(U,AT,V_a,f_a)$ be an information table. A subset of $X\subseteq U$, representing the extension of a concept, is called a definable set if and only if there is a formula p of L such that:

$$X=m(p). \quad (4)$$

Otherwise, it is indefinable.

A definable set is something that can be described precisely by using the properties of definable set.

Definition 4 Let $T=(U,AT,V_a,f_a)$ be an information table. 2^U is the power set of U. p is a formula of L. The set, consisted of all definable sets, is called the systems of definable sets. It can be denoted by:

$$DEF(U,L)=\{X\in 2^U \mid X=m(p), p\in L\}. \quad (5)$$

A concept, the pair $(p,m(p))$ in an information table T, where $p\in L$, is definable if its extension $m(p)$ is definable. Table 1 is an example of a complete information table.

TABLE 1 An complete information table

	a_1	a_2
x_1	0	0
x_2	0	0
x_3	1	1
x_4	0	1

The example of logic equations and their meaning sets can be computed as follows:

$$m(a_1=0)=\{x_1,x_2,x_4\},$$

$$m(a_2=0)=\{x_1,x_2\};$$

$$m((a_1=0)\wedge(a_2=0))=m(a_1=0)\cap m(a_2=0)=\{x_1,x_2\}.$$

3 Definability of concept in incomplete information

3.1 INCOMPLETE INFORMATION TABLE

For various reasons, the information contained in an information table is usually incomplete. There are many situations about incomplete information, but in our discussion, incompleteness means that instead of having a single value of an attribute, we have a subset of the attribute domain, which represents our knowledge that the actual value, though unknown, is one of the values in this subset [11]. This extends the idea of Codd's null value [17], corresponding to the case where this subset is the whole attribute domain. Mathematically, an incomplete information table can be defined as follows.

Definition 5 Let $T=(U,AT,V_a,f_a)$ be an information table. If each object $x\in U$ is mapped into a nonempty subset of V_a such that $\varnothing\neq F_a(x)\subseteq V_a$, where $F_a:U\to 2^{V_a}-\{\varnothing\}$, then the information table is called incomplete information table.

An example of incomplete information table is given as follows:

TABLE 2 An incomplete information table

	a_1	a_2
x_1	{0}	{0}
x_2	{0}	{1}
x_3	{0,1}	{1}
x_4	{1,2}	{1}

3.2 REPRESENTATION OF INCOMPLETE INFORMATION

For an incomplete information table, Lipski proposed a reasonable semantic interpretation that an incomplete information table can be represented by multiple complete information tables, namely, an incomplete information table is equivalent to a family of complete information tables.

Definition 6: Let $T'=(U,AT,\{V_a\mid a\in AT\},\{f'_a\mid a\in AT\})$ be a complete information table and $T=(U,AT,\{V_a\mid a\in AT\},\{F_a\mid a\in AT\})$, be an incomplete information table. For each $x\in U$, $a\in AT$, if $f'_a(x)$ has single element and $f'_a(x)\subseteq F_a(x)$, then we say that T_1 is a completion of T. The set, consisted of all the completions of an incomplete information table T, is equivalent to the information table T, written as:

$$CT(T)=\{T'\mid T' \textit{ is a completion of } T\}. \tag{6}$$

All the completions of Table 2 are listed in Table 3.

TABLE 3 Completions of incomplete information TABLE 2

T1			T2			T3			T4		
	a_1	a_2		a_1	a_2		a_1	a_2	a_2	a_1	a_2
x_1	0	0	x_1	0	0	x_1	0	0	x_1	0	0
x_2	0	0	x_2	0	0	x_2	0	0	x_2	0	0
x_3	0	0	x_3	0	1	x_3	1	1	x_3	1	1
x_4	1	1	x_4	2	0	x_4	1	1	x_4	2	1

According to Lipski's model, we can solve many problems about incomplete information table by solving those problems in the family of complete information tables.

3.3 DEFINABILITY OF CONCEPT IN INCOMPLETE INFORMATION

Given an equation p of L, there is a semantic interpretation about p in each completion of an incomplete information table. In the family of semantic interpretations about p, there are two sets can well define the semantic of the given formula which is called lower bound and upper bound, respectively.

$$\begin{aligned}\|p\|_*&=\{x\in U\mid \forall T'\in CT(T), x\in\|p\|^{T'}\}=\bigcap_{T'\in CT(T)}\|p\|^{T'}\\ \|p\|^*&=\{x\in U\mid \exists T'\in CT(T), x\in\|p\|^{T'}\}=\bigcap_{T'\in CT(T)}\|p\|^{T'}\end{aligned}, \tag{7}$$

where $\|p\|^{T'}$ denotes the set of semantic of the formula p in complete information table T'. Apparently, for arbitrary formula p of L, $\|p\|_*\subseteq\|p\|^*$ always holds. According to the definition of interval sets [8], it can be proved that $\|p\|_*,\|p\|^*$ construct an interval set, written as:

$$[\|p\|_*,\|p\|^*]=\{A\subseteq U\mid \|p\|_*\subseteq A\subseteq\|p\|^*\}. \tag{8}$$

Formally, the semantic interval set of an atomic formula $a=v$ of L in an incomplete information table can be computed as follows:

$$\begin{aligned}\|a=v\|_*&=\{x\in U\mid F_a(x)=\{v\}\};\\ \|a=v\|^*&=\{x\in U\mid v\in F_a(x)\}\end{aligned}. \tag{9}$$

For complicate logic formula, semantic interval sets have the properties as follows:

Theorem 1 Let p and q be two atomic formulas of the logic language L in an incomplete information table, the following equations hold.

$$\begin{aligned}&\|p\wedge q\|_*=\|p\|_*\cap\|q\|_*;\\ &\|p\wedge q\|^*\subseteq\|p\|^*\cap\|q\|^*;\\ &\|p\vee q\|_*\supseteq\|p\|_*\cup\|q\|_*;\\ &\|p\vee q\|^*=\|p\|^*\cup\|q\|^*.\end{aligned} \tag{10}$$

Proof: we only prove $\|p\wedge q\|_*=\|p\|_*\cap\|q\|_*$. Suppose $p=(a_1=v_1)$ and $q=(a_2=v_2)$. First, we prove $\|p\wedge q\|_*\subseteq\|p\|_*\cap\|q\|_*$. Suppose $x\in\|p\wedge q\|_*$, this means x satisfies formulas p and q at the same time. According to Equation (9), $x\in\{x\in U\mid F_{a_1}(x)=\{v_1\}\}$ and $x\in\{x\in U\mid F_{a_2}(x)=\{v_2\}\}$ hold. Thus we have $x\in\|p\|_*$ and $x\in\|q\|_*$, namely $x\in\|p\|_*\cap\|q\|_*$. Therefore, $\|p\wedge q\|_*\subseteq\|p\|_*\cap\|q\|_*$. Secondly, we prove $\|p\wedge q\|_*\supseteq\|p\|_*\cap\|q\|_*$. Suppose $x\in\|p\|_*\cap\|q\|_*$, namely $x\in\|p\|_*$ and $x\in\|q\|_*$. According to Equation (9), we have $x\in\{x\in U\mid, F_{a_1}(x)=\{v_1\}\}$ and $x\in\{x\in U\mid F_{a_2}(x)=\{v_2\}\}$. This means $x\in\{x\in U\mid F_{a_1}(x)=\{v_1\}\wedge F_{a_2}(x)=\{v_2\}\}$, i.e.,

$x \in \|p \wedge q\|_*$. Then we have $\|p \wedge q\|_* \supseteq \|p\|_* \cap \|q\|_*$. Therefore, $\|p \wedge q\|_* = \|p\|_* \cap \|q\|_*$.

The properties above mentioned show that the operations of semantic interval sets may be used to interpret the logic operations. But the properties given in Theorem 1 cannot work directly for the task because there is inequality sign (inclusion). The existence of equality is due to that the atomic formulas involved in the complicated formula have the same attributes. For example, if $p=(a_1=0)$, $q=(a_1=1)$, $p \vee q = ((a_1=0) \vee (a_1=1))$ then in Table 2, the semantic set of these formulas can be computed as follows: $\|p\|_* = \|a_1=0\|_* = \{x_1, x_2\}, \|q\|_* = \|a_1=1\|_* = \varnothing$, $\|p \vee q\|_* = \|a_1=0 \vee a_1=1\|_* = \{x_1, x_2, x_3\}$. So, we can get $\|p \vee q\|_* \supseteq \|p\|_* \cup \|q\|_*$, where the equation $\|p\|_* \cup \|q\|_*$ can be interpreted as "known to be $a_1=0$ or known to be $a_1=1$", but the equation $\|p \vee q\|_*$ is interpreted as "known to be $a_1=0$ or 1". If we assume that p and q have no common attributes, then the formulas of Theorem 1 would be all equations. Therefore, we can define a type of logic language L_0, which is a subset of *L*, investigated by Lipski [11], as the description language of incomplete information.

(i) $a=v$ is an atomic equation, where $a \in AT$, $v \in V_a$.

(ii) If p and q are equations in L_0, respectively, so as $p \wedge q, p \vee q$ are also equations in L_0.

Definition 7: Let $T=(U, AT, \{V_a \mid a \in AT\}, \{F_a \mid a \in AT\})$ be an incomplete information table, p be a formula of L_0. For $X \subseteq U$, we say that X is interval-definable if there is an interval set $M(p)$ such that:

$$X \propto M(p), \tag{11}$$

where $M(p)=[m_*(p), m^*(p)]$, $\|p\|_* = m_*(p), \|p\|^* = m^*(p)$.

A concept in incomplete information is interval-definable if its extension is an interval-definable set, written as $(p, M(p))$. p is the description formula corresponding with the interval set $M(p)$. For a concept $(p, M(p))$, if $\|p\|_* = m_*(p) = \|p\|^* = m^*(p)$, then the concept is a precise concept, and the definability of the concept is same as usual definition. It means that the information about this concept is complete. All the interval-definable concepts construct a family, which is a family of interval-definable sets in an incomplete information table, or the universe U. We can give the following definition to describe this sets family.

Definition 8: Let $T=(U, AT, \{V_a \mid a \in AT\}, \{F_a \mid a \in AT\})$ be an incomplete information table. 2^U is the power set of U. p is a formula of L_0. The set, consisted of all interval-definable sets, is called the interval-definable concept system. It can be denoted by:

$$IDEF(U, L_0) = \{X \in 2^U \mid X \propto M(p), p \in L_0\}. \tag{12}$$

Two concepts included in the interval-definable concept system, which are inequality, include different information, or they have different uncertainty. In order to study the properties of interval-definable concepts, the knowledge ordering [9] is introduced to describe the difference of two concepts.

3.4 PROPERTIES OF INTERVAL-DEINABLE CONCEPTS

Let *A* and *B* be two interval-definable concepts in the system $IDEF(U, L_0)$. Assume that $[\varnothing, \varnothing] \in IDEF(U, L_0)$. Suppose that $A=[A_l, A_u]$ and $B=[B_l, B_u]$. We have more knowledge about the concept *A* than the concept *B*, as we are more sure about the former than the latter, if $A \subseteq B$. This suggests that the standard set inclusion provides a knowledge ordering on interval sets.

Definition 9 A knowledge ordering $\preceq_k$ [9] on interval sets can be defined by:

$$[B_l, B_u] \preceq_k [A_l, A_u] \Leftrightarrow [A_l, A_u] \subseteq [B_l, B_u] \Leftrightarrow B_l \subseteq A_l \subseteq A_u \subseteq B_u. \tag{13}$$

In some sense, the knowledge ordering reflects the fact that *A* is tighter than *B*. Generally, we use the standard set inclusion $\subseteq$ as a knowledge ordering.

The set intersection of two interval sets is an interval set, namely, for $A=[A_l, A_u]$ and $B=[B_l, B_u]$,

$$A \cap B = \begin{cases} [A_l \cup B_l, A_u \cap B_u], when\ A_l \cup B_l \subseteq A_u \cap B_u, \\ [\varnothing, \varnothing] \qquad ,otherwise. \end{cases} \tag{14}$$

The following properties hold for the relation $\subseteq$ on interval-definable concept system:

Theorem 2 For $A, B, C, D \in IDEF(U, L_0)$, the following equations hold.

$$A \subseteq B \Leftrightarrow A \cap B = A; A \subseteq B \Leftrightarrow A \cup B = B, \tag{15}$$

$$A \subseteq B\, and\, C \subseteq D \Rightarrow A \cap C \subseteq B \cap D, \tag{16}$$

$$A \cap B \subseteq A, A \cap B \subseteq B. \tag{17}$$

Proof We only prove $A \subseteq B \Leftrightarrow A \cap B = A$. Suppose $A=[A_l, A_u]$ and $B=[B_l, B_u]$. According to the definition of knowledge ordering, $[A_l, A_u] \subseteq [B_l, B_u] \Leftrightarrow B_l \subseteq A_l \subseteq A_u \subseteq B_u$. On the other hand, $A \cap B = A \Leftrightarrow [A_l \cup B_l, A_u \cap B_u] = [A_l, A_u]$. This means that

$A_l \cup B_l = A_l$, $A_u \cap B_u = A_u$. In the sense of general set-theoretic operations, this is equivalent to $B_l \subseteq A_l, A_u \subseteq B_u$, thus we have $B_l \subseteq A_l \subseteq A_u \subseteq B_u$. Therefore, $A \sqsubseteq B \Leftrightarrow A \cup B = A$ holds.

The relation $\sqsubseteq$ on $IDEF(U, L_0)$ is a reflexive and transitive relation. It is an ordering relation for defining the semi-lattice operation $\cap$. With respect to the knowledge ordering $\sqsubseteq$, we can say that the interval-definable concept system $IDEF(U, L_0)$ is a semi-lattice $IDEF(U, L_0), \cap)$, or $(IDEF(U, L_0), \sqsubseteq)$.

4 Concluding remarks

The notion of definability in complete information table is examined. For studying the incomplete information table, a new definition of definability, named interval-definable, is proposed. The foundation of this notion is that an incomplete information table can be represented by an interval set and expressed as a family of complete information tables. Based on the definition of interval-definable, the interval-definable concept system is defined. By introducing a knowledge ordering, the properties of interval-definable concept system are investigated. The conclusion indicates that the interval-definable concept system is a semi-lattice with respect to the operation $\cap$, or relation $\sqsubseteq$. By studying definability in incomplete information table and interval sets, we hope that we can gain more insights into the representation and processing of imprecise or partially known concepts, and into the approximations of indefinable or complex concepts. More development and applications of interval sets in incomplete information will be studied in the near future.

Acknowledgements

This project is supported by the Humanities and Social Science Youth Foundation of the Ministry of Education (No.11YJCZH070), the Natural Science Foundation of Shandong Province (No.ZR2011GM006), and Shandong Province Higher School Humanities and Social Science Plan (No.J12WF86).

References

[1] Pawlak Z 1982 Rough sets *Int J of Computer and Inf Sci* **11**(5) 341-56
[2] Pawlak Z 1991 Rough sets-Theoretical Aspects of Reasoning about Data *Kluwer Academic Publishers:* Dordrecht
[3] Yao Y Y 2007 A note on definability and approximations *LNCS Trans. on Rough Sets VII* **4400** 274-82
[4] Buszkowski W 1998 Approximation spaces and definability for incomplete information systems *Rough Sets and Current Trends in Computing* LNAI **1424** 115-22
[5] Grzymala-Busse J W 2005 Incomplete data and generalization of indiscernibility relation, definability, and approximations *The 10th Int Conf on Rough Sets, Fuzzy sets, Data Mining, and Granular Computing* **1** Springer-Verlag: Heidelberg 244-53
[6] Järvinen J, Kortelainen J 2004 A note on definability in rough sets theory *Current Issues in Data and Knowledge Engineering, Akademicka Oficyna Wydawnicza EXIT* Warsaw 272-7
[7] Wasilewska A 1987 Definable sets in knowledge representation systems *Bulletin of the Polish Academy of Sci: Math* **35** 629-35
[8] Yao Y Y 1993 Interval-set algebra for qualitative knowledge representation *ICCI '93 Proceedings of the Fifth International Conference on Computing and Information* 370-74
[9] Yao Y Y 2009 Interval sets and interval-set algebra *Proceedings of The 8th IEEE International Conference on Cognitive Informatics* 307-14
[10] Miao D, Wang G, Yao Y, Liang J, Wu W, Zhang Y 2012 The Cloud Model and Granular Computing *Science Press*
[11] Lipski W J R 1979 On semantic issues connected with incomplete information databases *ACM Transactions on Database Systems* **4**(3) 262-96
[12] Lipski W J R 1981 On database with incomplete information *J of the ACM* **28**(1) 41-70
[13] Ogden C K, Richards I A 1946 The Meaning of Meaning: A Study of the Influence of Language Upon Thought and of the Science of Symbolism (8th ed) *Harcourt Brace Jovanovich:* New York
[14] Smith E E 1989 Concepts and induction in Posner M I (Ed.) Foundations of Cognitive Science *Cambridge MA MIT* 501-26
[15] Sowa J F 1984 Conceptual Structures: Information Processing in Mind and Machine *Addison-Wesley Longman Publishin Co Inc Boston MA USA*
[16] Yao Y Y 2009 Interpreting concept learning in cognitive informatics and granular computing *IEEE Transactions on Systems, Man, and Cybernetics Part B: Cybernetics* **39**(4) 855-66
[17] Codd E F 1975 Understanding relations (installment #7) *FDT Bulletin of ACM-SIGMOD* **7**(3-4) 23-8

Authors

Shunliang Huang, born in November, 1977, Shandong, China

Current position, grades: Doctor of Engineering, Associate professor in Shandong University of Technology.
University studies: Control Theory and Control Engineering in Shandong University.
Scientific interest: rough sets theory, data mining.
Publications: 22 papers.