

# 不完备形式背景上的知识表示

智慧来

(河南理工大学计算机科学与技术学院 焦作 454000)

**摘要** 不完备形式背景中包含有不确定性信息,其上的知识表示与完备形式背景上的知识表示既有区别又有联系。为了研究两者的内在联系,定义了偏小近似形式背景与偏大近似形式背景,以及偏小近似概念格与偏大近似概念格,提出了偏大近似概念格上粗糙概念的识别方法,研究了偏小近似概念格与偏大近似概念格之间的蕴含关系。结论表明,可以用偏大近似概念格来作为不完备形式背景的知识表示工具。

**关键词** 不完备形式背景,粗糙概念,精确概念,概念格

**中图分类号** TP181 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.1.061

## Knowledge Representation on Incomplete Formal Context

ZHI Hui-lai

(School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454000, China)

**Abstract** Incomplete formal context contains uncertainty information, and thus knowledge representation on incomplete formal context and the one on complete formal context have distinction, and also have connection. In order to study their internal relationship, upper and lower approximate formal context, as well as upper and lower approximate concept lattice were defined respectively. Then a recognition method of rough concept was put forward. Moreover, the relationship between upper approximate concept lattice and lower approximate concept lattice was also studied. Conclusion shows that upper approximate concept lattice can be used as a knowledge representation tool for incomplete formal context.

**Keywords** Incomplete formal context, Rough concept, Exact concept, Concept lattice

## 1 引言

形式概念分析是研究知识表示和推理的理论,是在格论和序论上发展起来的,具有严格的数学模型,已经在数据挖掘领域中有许多成功的应用<sup>[1,2]</sup>。

通常研究的形式背景是完备的,即对象和属性之间的关系是已知的,即:若用函数  $I(g, m)$  表示对象  $g$  和属性  $m$  之间的关系,则  $I(g, m)$  的取值是确定的。根据  $I(g, m)$  取值范围的不同,可以得到不同类型的形式背景,其中包括布尔值形式背景<sup>[3]</sup>、区间值形式背景<sup>[4]</sup>、模糊形式背景<sup>[5]</sup>、混合值形式背景<sup>[6]</sup>等。

然而遗憾的是,受到人的认知能力以及客观事物复杂性的制约,人们时常不能准确判断对象和属性之间的关系,使得获取的形式背景存在数据缺失,也就是无法确定  $I(g, m)$  的取值,从而得到的形式背景是不完备的。

粗糙集理论和形式概念分析具有相同的研究对象,粗糙集理论中的信息系统亦即形式概念分析理论中的形式背景。在粗糙集理论中,对于一个不完备的信息系统,通常的做法是用 1 或 0 补信息系统中的缺失信息,从而使其转化为一个完备的信息系统,然后再定义信息系统中的上、下近似等概念<sup>[7,8]</sup>。

对于不完备信息系统,李金海研究了近似概念构造、规则提取与属性约简<sup>[9]</sup>。康向平等在不完备形式背景中引入了相

容概念和相容规则,重点讨论了基于相容概念的相容规则集合获取<sup>[10]</sup>。在上述研究中,无论是由近似概念构成的近似概念格,还是由相容概念构成的相容概念格,其中都包含有确定的信息和不确定的信息。本文将借用粗糙集理论中的相关概念,提出并定义偏小近似形式背景、偏大近似形式背景,并构造二者上的概念格,进而研究二者之间的内在联系,为不完备形式背景中的信息处理提供理论和技术上的支持。

## 2 基本概念

在形式概念分析中,数据是用形式背景来表示的,下面给出它的形式化定义。

**定义 1** 一个形式背景是一个三元组  $K=(G, M, I)$ ,  $G$  和  $M$  是两个集合,  $I \subseteq G \times M$  是一个二元关系,我们将  $G$  的元素称为对象,  $M$  的元素称为属性,  $I$  表示对象和属性之间的关系,  $gIm$  或  $(g, m) \in I$  表示对象  $g$  具有属性  $m$ 。

**定义 2** 在形式背景  $K=(G, M, I)$  中,  $A \subseteq G$ , 定义  $f(A) = \{m \in M \mid gIm, \forall g \in A\}$  ( $A$  中对象共同属性的集合);相应地,  $B \subseteq M$ , 定义  $g(B) = \{g \in G \mid gIm, \forall m \in B\}$  (具有  $B$  中所有属性的对象的集合)。

**定义 3** 在形式背景  $K=(G, M, I)$  中,  $A \subseteq G, B \subseteq M$ , 如果有  $f(A)=B, g(B)=A$ , 则称二元组  $C=(A, B)$  是一个概念,  $A$  是概念  $C$  的外延,  $B$  是概念  $C$  的内涵, 并记  $K$  的全体概念为  $B(K)$ 。

到稿日期:2014-02-28 返修日期:2014-06-06 本文受国家自然科学基金(60975033),河南理工大学博士基金(B2011-102)资助。

智慧来(1981-),男,博士,讲师,主要研究方向为形式概念格分析、知识表示和推理, E-mail: zhihuilai@126.com。

**定义 4** 在形式背景  $K=(G, M, I)$  中, 概念  $C_1=(A_1, B_1), C_2=(A_2, B_2) \in B(K)$ , 定义  $C_1 \leq C_2 \Leftrightarrow A_1 \subseteq A_2 \Leftrightarrow B_1 \supseteq B_2$ , 并称  $C_2$  是  $C_1$  的父概念,  $C_1$  是  $C_2$  的子概念,  $\leq$  称为概念的层次序,  $B(K)$  中的所有概念用这种序组成的集合称为概念格, 记做  $L(G, M, I)$ . 说明: 为了简便起见, 对于  $\{g\} \subseteq G$ , 用  $f(g)$  代替  $f(\{g\})$ ; 相应地, 对于  $\{m\} \subseteq M$ , 用  $g(m)$  代替  $g(\{m\})$ .

说明: 为了简便起见, 对于  $\{g\} \subseteq G$ , 用  $f(g)$  代替  $f(\{g\})$ ; 相应地, 对于  $\{m\} \subseteq M$ , 用  $g(m)$  代替  $g(\{m\})$ .

**定义 5** 对于一个形式背景  $(G, M, I)$ , 如果任意两个满足  $f(g)=f(h)$  的元素  $g, h \in G$  都有  $g=h$ , 而且对偶地任意两个满足  $g(m)=g(n)$  的元素  $m, n (m, n \in M)$  都有  $m=n$ , 则称这个形式背景是净化的.

**定义 6**<sup>[11]</sup> 对于一个净化的形式背景  $(G, M, I)$ , 如果满足对任意一个元素  $g \in G$  都不存在  $A \subseteq G$  使得  $\bigcap_{a \in A} f(a) = f(g)$  成立, 而且对偶地满足对任意一个元素  $m (m \in M)$  都不存在  $B \subseteq G$  使得  $\bigcap_{b \in B} g(b) = g(m)$  成立, 则称这个形式背景为纯化的.

### 3 完备形式背景及其上、下近似形式背景

在实际应用中, 例如在三支决策<sup>[12,13]</sup>中, 对象和属性之间的关系有时是难以确定的, 也就是说对于给定的一个对象  $g \in G$ , 一个属性  $m \in M, (g, m) \in I$  是否成立难以确定. 包含不确定性信息的形式背景, 称为不完备形式背景.

**定义 7** 一个不完备形式背景是一个三元组  $IK=(G, M, I), G$  和  $M$  是两个集合,  $I \subseteq G \times M$  是一个二元关系, 我们将  $G$  的元素称为对象,  $M$  的元素称为属性,  $I$  表示对象和属性之间的关系, 当  $I(g, m)=1$  时表示对象  $g$  具有属性  $m$ , 当  $I(g, m)=0$  时表示对象  $g$  不具有属性  $m$ , 当  $I(g, m)=*$  时表示对象  $g$  是否具有属性  $m$  尚不确定.

表 1 是一个不完备形式背景的例子. 对于不完备形式背景中的不确定性信息, 可以有两种处理方式: 第一种方式是把  $I(g, m)=*$  处理为  $I(g, m)=0$ ; 第二种方式是把  $I(g, m)=*$  处理为  $I(g, m)=1$ .

表 1 不完备形式背景  $IK_1$

	a	b	c	d	e
1	1	0	*	1	0
2	0	1	0	*	0
3	1	0	1	0	1
4	1	1	0	0	1

借用粗糙集中的概念, 在下文中把第一种处理方式得到的形式背景称为偏小近似形式背景, 并记做  $IK^U$ ; 把第二种处理方式得到的形式背景称为偏大近似形式背景, 并记做  $IK^L$ .

**定义 8** 给定一个不完备形式背景  $IK$ , 称其偏小近似形式背景  $IK^U$  的概念格  $L(K^U)$  为偏小近似概念格; 称其偏大近似形式背景  $IK^L$  的概念格  $L(K^L)$  为偏大近似概念格.

### 4 不完备形式背景上的粗糙概念

**定义 9** 对于概念格中的任意一个概念  $(A, B)$ , 若在不完备形式背景  $IK$  中存在  $a \in A, b \in B$  且  $I(a, b)=*$ , 则称这个概念为粗糙概念(Rough Concept), 否则称这个概念为精确概念(Exact Concept).

**定义 10**<sup>[14]</sup> 在概念格  $L(G, M, I)$  中, 如果一个概念具有

形式  $(g(f(a)), f(a))$  且  $a \in G$ , 则称  $(g(f(a)), f(a))$  是一个对象概念; 如果一个概念具有形式  $(g(b), f(g(b)))$  且  $b \in M$ , 则称  $(g(b), f(g(b)))$  是一个属性概念.

**定理 1** 在纯化形式背景中, 对象概念是上确界不可约元, 上确界不可约元一定是对象概念.

证明: 首先证明在纯化形式背景中对象概念一定是上确界不可约元, 然后证明上确界不可约元一定是对象概念.

(1) 证明纯化形式背景中对象概念一定是上确界不可约元. 反证法, 假设对象概念  $(A, B)$  不是上确界不可约元, 即对象概念有两个或两个以上的下近邻, 其中的两个下近邻记做  $(A_i, B_i), i$  为指标集,  $i \in T$ . 由于  $(A, B)$  是对象概念, 因此存在对象  $g \in A$  使得  $f(g)=B$ . 根据概念格基本定理, 有  $B = \bigcap_{t \in T} B_t$ , 又因为  $f(A_t)=B_t (t \in T)$ , 则有  $f(g) = \bigcap_{t \in T} f(A_t)$  ( $t \in T$ ), 因此  $g$  是可约简的, 与纯化形式背景的前提矛盾, 故定理成立.

(2) 证明纯化形式背景中上确界不可约元一定是对象概念. 反证法, 令概念  $(A, B)$  是一个上确界不可约元, 假设  $(A, B)$  不是对象概念, 则对于其概念外延中的对象  $g_1, g_2, \dots, g_{|A|}$  不存在一个  $1, 2, \dots, |A|$  的排列  $p_1, p_2, \dots, p_{|A|}$  使得  $f(g_{p_1}) \supseteq f(g_{p_2}) \supseteq \dots \supseteq f(g_{p_{|A|}})$  成立, 这意味着至少存在两个对象  $g_i, g_j$ , 使得  $f(g_i) \supseteq f(g_j)$  和  $f(g_j) \supseteq f(g_i)$  都不成立, 则必有  $f(g_i) \cap f(g_j) = f(A)$  成立, 这意味着  $(A, B)$  至少有两个子概念, 其中一个概念的外延中包含  $g_i$ , 另一个概念的外延中包含  $g_j$ , 而这显然与上确界不可约元只有一个下近邻矛盾, 故定理成立.

**定理 2** 在纯化形式背景中, 属性概念是下确界不可约元, 下确界不可约元一定是属性概念.

证明: 由定理 1, 根据对偶原理可知本定理成立.

根据定理 1 和定理 2, 可以在概念格中识别所有的对象概念和属性概念. 在概念格  $L(K^U)$  中标注对象概念和属性概念, 如图 1 所示.

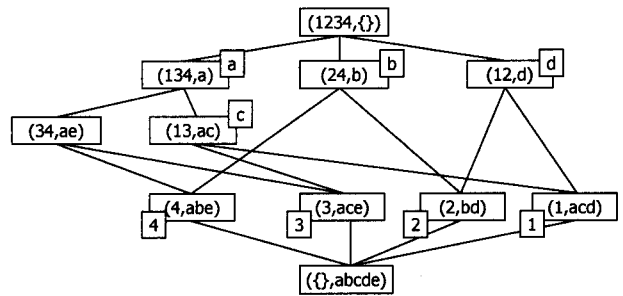


图 1 标注了对象概念和属性概念的  $L(K^U)$

**定义 11** 在概念格  $L(G, M, I)$  中, 对于给定的对象  $a \subseteq G$  与属性  $b \subseteq M$ , 记  $G(a, b)$  为:

$$G(a, b) = \{(A, B) \mid (A, B) \text{ 是一个概念, 且 } (g(f(a)), f(a)) \leq (A, B) \leq (g(b), f(g(b)))\}$$

则称  $G(a, b)$  为一个由对象  $a$  和属性  $b$  决定的信息粒, 简称信息粒.

**定理 3** 对于一个不完备形式背景  $IK$ , 若有  $I(g, m)=*$ , 则在偏大近似概念格  $L(K^L)$  中, 信息粒  $G(g, m)$  包含的概念是粗糙概念.

证明: 对任意的一个概念  $(A, B) \in G(g, m)$ , 根据概念之间的继承关系可知  $A$  包含  $g$  并且  $B$  包含  $m$ , 进而根据定义 9 可知, 概念  $(A, B)$  是一个粗糙概念.

定义 12 若  $C=(A,B)$  是一个粗糙概念, 记  $r(C)$  为:

$$r(C) = (|\{a \in A \mid \exists b \in B, I(a,b) = * \}| + |\{b \in B \mid \exists a \in A, I(a,b) = * \}|) / (|A| + |B|)$$

则称  $r(C)$  为概念  $C$  的粗糙度。

例如, 在不完备形式背景  $IK_1$  中存在  $I(1,c) = *, I(2,d) = *$ , 则信息粒  $G(1,c)$  和  $G(2,d)$  中包含的概念  $(13,ac)$ 、 $(1,acd)$ 、 $(12,d)$ 、 $(2,bd)$  都是粗糙概念, 且有  $r((13,ac)) = r((1,acd)) = 1/2$ ,  $r((12,d)) = r((2,bd)) = 2/3$ 。

## 5 上、下近似形式背景之间的关系

定义 13<sup>[15]</sup> 在一个概念格中, 如果存在概念  $(A_1, B_1) > (A_2, B_2)$ , 并且不存在概念  $(A_3, B_3)$ , 使得  $(A_1, B_1) > (A_3, B_3) > (A_2, B_2)$ , 那么  $(A_1, B_1)$  和  $(A_2, B_2)$  称为父子概念对, 记做  $[(A_1, B_1), (A_2, B_2)]$ 。

定理 4 对于给定一个不完备形式背景, 其偏大近似概念格蕴含其偏小近似概念格。

证明: 对于一个不完备形式背景  $IK$ , 它的偏大近似形式背景  $IK^L$  可由偏小近似形式背景  $IK^U$  得到, 即将偏小近似形式背景  $IK^U$  中的关系  $\{I(g,m) = 0 \mid I(g,m) = *, I(g,m) \in IK\}$  修改为  $\{I(g,m) = 1 \mid I(g,m) = *, I(g,m) \in IK\}$ 。

将偏小近似形式背景  $IK^U$  调整为偏大近似形式背景  $IK^L$  之后, 构造偏大近似概念格  $L(IK^L)$  只需要对偏小近似概念格  $L(IK^U)$  进行更新。如果在  $L(IK^U)$  中存在父子概念对  $[(A_1, B_1), (A_2, B_2)]$ , 且满足  $A_1$  包含  $g$ ,  $B_2$  包含  $m$ , 那么需要生成概念  $(A_2 \cup g, B_1 \cup m)$ , 然后分下述 3 种情况进行更新。

情形 1: 若  $A_1 \subseteq A_2 \cup g, B_1 \cup m \supseteq B_2$ , 即  $(A_1, B_1)$  和  $(A_2, B_2)$  均蕴含在生成概念  $(A_2 \cup g, B_1 \cup m)$  中, 则由生成概念  $(A_2 \cup g, B_1 \cup m)$  代替父子概念对  $[(A_1, B_1), (A_2, B_2)]$ 。

情形 2: 若  $A_1 \subseteq A_2 \cup g$  成立,  $B_1 \cup m \supseteq B_2$  不成立, 即  $(A_1, B_1)$  蕴含在生成概念  $(A_2 \cup g, B_1 \cup m)$  中, 则由生成概念  $(A_2 \cup g, B_1 \cup m)$  代替父概念  $(A_1, B_1)$ ; 若  $A_1 \subseteq A_2 \cup g$  不成立,  $B_1 \cup m \supseteq B_2$  成立, 即  $(A_2, B_2)$  蕴含在生成概念  $(A_2 \cup g, B_1 \cup m)$  中, 则由生成概念  $(A_2 \cup g, B_1 \cup m)$  代替子概念  $(A_2, B_2)$ 。

情形 3: 若  $A_1 \subseteq A_2 \cup g$  不成立,  $B_1 \cup m \supseteq B_2$  不成立, 即  $(A_1, B_1)$  和  $(A_2, B_2)$  保持不变, 则不执行任何动作, 概念格中增加生成概念  $(A_2 \cup g, B_1 \cup m)$ 。

综合上述 3 种情形可知, 偏小近似概念格中的概念都蕴含在偏大近似概念格中, 故定理成立。

推论 1 在偏大近似概念格  $L(IK^L)$  中, 若一个概念是精确概念而不是粗糙概念, 则这个概念一定存在于偏小近似概念格  $L(IK^U)$  中。

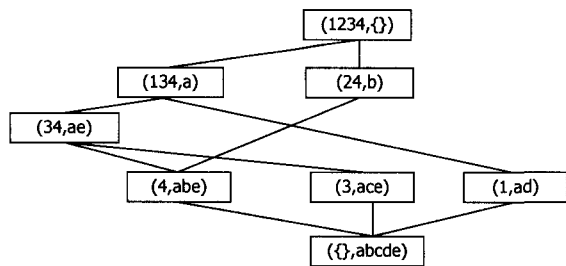


图 2 偏小近似概念格  $L(IK_1^U)$

例如, 偏小近似概念格  $L(IK_1^U)$  如图 2 所示, 它有 8 个概

念, 即:  $(1234, \{\}), (134, a), (24, b), (34, ae), (4, abe), (3, ace), (1, ad), (\{\}, abcde)$ 。其中  $(1234, \{\}), (134, a), (24, b), (34, ae), (4, abe), (3, ace), (\{\}, abcde)$  等 7 个概念存在于偏大近似概念格  $L(IK_1^L)$  中, 且这些概念都是精确概念。另外, 概念  $(1, ad)$  蕴含在偏大近似概念格  $L(IK_1^L)$  中的  $(1, acd)$  中。

由上述讨论看到, 偏大近似概念格中包含了偏小近似概念格中的所有信息, 因此, 可以用偏大近似概念格作为不完备形式背景的知识表示工具。

结束语 本文提出并定义偏小近似形式背景、偏大近似形式背景, 并研究二者之间的内在联系, 得到的结论是偏大近似形式背景能够蕴含偏小近似形式背景的所有信息, 可以用偏大近似概念格作为不完备形式背景上的知识表示工具。需要进一步研究的问题还有很多, 例如: 对于不完备形式背景上的关联规则, 如何衡量其不确定性; 偏小近似形式背景上的关联规则与偏大近似形式背景上的关联规则之间有什么联系; 如何在不完备形式背景上进行对象检索等问题。

## 参考文献

- [1] Jonas P, Ignatov Dmitry I, Kuznetsov Sergei O. Formal concept analysis in knowledge processing: A survey on applications[J]. Expert Systems with Applications, 2013, 40(16): 6538-6560
- [2] Jonas P, Kuznetsov Sergei O, Ignatov Dmitry I. Formal Concept Analysis in knowledge processing: A survey on models and techniques[J]. Expert Systems with Applications, 2013, 40(16): 6601-6623
- [3] Ganter B, Wille R. Formal Concept Analysis: mathematical foundation[M]. New York: Springer-Verlag, 1999
- [4] Jaoua A, Elloumi S. Galois connection, formal concepts and Galois lattice in real relation application in a real classifier[J]. Journal of Systems & Software, 2002, 60: 149-163
- [5] 刘宗田, 强宇, 周文, 等. 一种模糊概念格模型及其渐进式构造算法[J]. 计算机学报, 2007, 30(2): 184-188
- [6] 智慧来. 面向异构数据分析的形式概念分析扩展模型[J]. 电子学报, 2013, 41(12): 2451-2455
- [7] Gong Zeng-tai. Rough set theory for the incomplete interval valued fuzzy information systems[J]. Journal of Intelligent and Fuzzy Systems, 2014, 26(2): 889-900
- [8] Yang Xi-bei, Song Xiao-ning, Chen Ze-hua, et al. On multigranulation rough sets in incomplete information system[J]. International Journal of Machine Learning and Cybernetics, 2012, 3(3): 223-232
- [9] 李金海. 面向规则提取的概念格约简方法及其实现算法[D]. 西安: 西安交通大学, 2012
- [10] 康向平, 李德玉, 曲开社. 不完备形式背景中的知识获取方法[J]. 计算机工程与应用, 2010, 46(36): 16-19
- [11] 智慧来, 智东杰. 纯化形式背景及其性质研究[J]. 计算机工程与应用, 2011, 47(35): 61-62
- [12] 贾修一, 商琳. 一种求三支决策阈值的模拟退火算法[J]. 小型微型计算机系统, 2013, 34(11): 2603-2606
- [13] 刘盾, 李天瑞, 李华雄. 粗糙集理论: 基于三支决策视角[J]. 南京大学学报: 自然科学版, 2013, 49(5): 574-581
- [14] 智慧来, 智东杰. 形式概念分析中的对象概念与属性概念[J]. 计算机工程与应用, 2013, 49(18): 112-115
- [15] 智慧来, 智东杰. 关系粒度的概念格增量维护与关联规则更新[J]. 计算机科学, 2013, 40(4): 256-258