

DOI: 10.11992/tis.201809021

网络出版地址: <http://kns.net/kcms/detail/23.1538.tp.20190520.1330.004.html>

# 概念格在不完备形式背景中的知识获取模型

王雯<sup>1,2</sup>, 康向平<sup>3,4</sup>, 武燕<sup>2</sup>

(1. 太原工业学院 自动化系, 山西 太原 030008; 2. 太原理工大学 信息工程学院, 山西 太原 030024; 3. 同济大学 嵌入式系统与服务计算教育部重点实验室, 上海 201804; 4. 同济大学 计算机科学与技术系, 上海 201804)

**摘要:** 为了使概念格模型具有更强的数据处理能力, 消除不完备信息带来的影响, 针对经典概念格的局限性, 本文将粗糙集中的粒化思维融入到概念格中。首先探讨了概念格视角下的信息粒化方法, 然后提出了基于等价类和基于极大相容类的知识获取方法, 最后给出了实例分析。这些方法一方面有助于概念格与粗糙集的融合, 另一方面也为探索不完备形式背景的分析处理机制提供了有益思路。

**关键词:** 概念格; 粗糙集; 不完备形式背景; 等价类; 极大相容类; 信息粒化; 数据处理

**中图分类号:** TP18 **文献标志码:** A **文章编号:** 1673-4785(2019)05-1048-08

**中文引用格式:** 王雯, 康向平, 武燕. 概念格在不完备形式背景中的知识获取模型 [J]. 智能系统学报, 2019, 14(5): 1048-1055.

**英文引用格式:** WANG Wen, KANG Xiangping, WU Yan. Knowledge acquisition model of concept lattice in an incomplete formal context[J]. CAAI transactions on intelligent systems, 2019, 14(5): 1048-1055.

## Knowledge acquisition model of concept lattice in an incomplete formal context

WANG Wen<sup>1,2</sup>, KANG Xiangping<sup>3,4</sup>, WU Yan<sup>2</sup>

(1. Department of Automation, Taiyuan Industrial College, Taiyuan 030008, China; 2. College of Information Engineering, Taiyuan University of Technology, Taiyuan 030024, China; 3. Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai 201804, China; 4. Department of Computer Science and Technology, Tongji University, Shanghai 201804, China)

**Abstract:** To improve the data processing ability of the concept lattice model and eliminate the impact of incomplete information, in this paper, considering the limitations of classical concept lattice in actual applications, the granulation idea in rough sets is integrated into concept lattice. First, we explore information granulation methods from the perspective of concept lattice and then propose a knowledge acquisition method based on the equivalence class and maximal tolerance class, and then, carry out the case analysis. These methods are helpful for the integration of concept lattice and rough sets. Moreover, they also provide some useful ideas for exploring the analysis and processing mechanism of incomplete formal contexts.

**Keywords:** concept lattice; rough set; incomplete formal context; equivalence class; maximal tolerance class; information granulation; data processing

通过模拟人类的概念认知思维, Wille<sup>[1]</sup> 教授于 1982 年在格论基础上发展了面向概念建模的概念格理论。作为格论的重要应用分支, 概念格具有坚实的数学理论基础, 其中概念、格代数结

构、对偶伽罗瓦连接等是核心要素。近年来, 伴随着概念格自身理论发展以及与粗糙集<sup>[2-3]</sup>、模糊集等的深度融合, 其理论体系日趋成熟, 应用范围也得到了极大的拓展。

粗糙集无需借助先验知识, 且符合人类的近似思维, 易于理解, 因此受到了国内外学者的广泛关注<sup>[4]</sup>。近年来, 对于不完备信息的处理, 粗糙

收稿日期: 2018-09-13. 网络出版日期: 2019-05-21.

基金项目: 国家自然科学基金项目 (61603278).

通信作者: 王雯. E-mail: [wangwen80971@163.com](mailto:wangwen80971@163.com).

集已经取得了大量的研究成果。把缺失值理解为任何可能值, Kryszkiewicz<sup>[5-6]</sup>提出了基于容差关系的拓展粗糙集模型,其中容差关系满足自反性和对称性,随后,将缺失值视为不存在或是不允许比较的未知值, Stefanowski<sup>[7]</sup>提出了基于不对称相似关系(满足自反性和传递性)的粗糙集拓展模型,结合上述2种模型的优点,王国胤<sup>[4]</sup>进一步提出了基于限制容差关系(满足自反性和对称性)的扩充模型,该模型相对以往模型更加符合实际, Leung等<sup>[8]</sup>将极大相容类视为一个粒,探讨了不完备信息系统中的知识获取,张文修<sup>[9]</sup>将不完备信息系统理解为一个集值信息系统,并探讨了面向集值的相容关系。

目前,关于不完备信息系统已有大量的研究成果,然而对于不完备形式背景的分析处理尚处于起步阶段<sup>[10-16]</sup>。从研究对象来讲,形式背景本质上是一种特殊的信息系统,它们之间存在着天然联系,具有较强的相容性,这也意味着面向不完备信息系统的知识获取方法对于不完备形式背景的处理具有一定的借鉴作用。事实上,将不完备信息系统中的方法推广到不完备形式背景中,也是一项非常有意义的研究工作,其不仅能为不完备形式背景的分析处理提供必要的支撑,而且也有助于概念格与粗糙集的理论融合。

考虑到不完备形式背景的普遍性以及经典概念格的局限性,在概念格框架体系中,本文融入了粗糙集中的粒化思维,探讨了概念格视角下的信息粒化,提出了基于等价类和基于极大相容类的知识获取方法。这些方法一方面有助于概念格与粗糙集的融合,另一方面也为探索不完备形式背景的分析处理机制提供了有益思路。

## 1 概念格基本知识

通常,称  $(G, M, I)$  是一个形式背景,若  $G$  是对象集,  $M$  是属性集,  $I$  表示  $G$  和  $M$  之间的二元关系。在此意义下,  $(x, m) \in I$  或  $m(x) = 1$  表示对象  $x \in G$  拥有属性  $m \in M$ , 相反,  $(x, m) \notin I$  或  $m(x) = 0$  表示对象  $x$  不拥有属性  $m$ 。

**定义1** 在  $K = (G, M, I)$  中,对于任意  $X \in 2^G$  和  $B \in 2^M$ , 相应的经典内涵算子与外延算子定义为

$$X^* = \{m \in M \mid (x, m) \in I, \forall x \in X\} \quad (1)$$

$$B^* = \{x \in G \mid (x, m) \in I, \forall m \in B\} \quad (2)$$

式中  $2^X$  是集合  $X$  的幂集。若  $B = X^*$  且  $X = B^*$ , 称  $(X, B)$  是一个概念,其中  $X$  和  $B$  分别称为概念的外延和内涵。在此意义下,任意概念  $(X_1, B_1)$  和  $(X_2, B_2)$  之间的序关系定义为

$$(X_1, B_1) \leq (X_2, B_2) \Leftrightarrow X_1 \subseteq X_2 \Leftrightarrow B_2 \subseteq B_1$$

基于上述认识和理解,一个格代数结构  $(\mathcal{B}(K), \leq)$  可以从  $K$  中推导出来,其中  $\mathcal{B}(K)$  是  $K$  中所有概念的集合。关于概念格详细资料请查阅文献<sup>[17]</sup>。

对象和属性之间的关系如果存在缺失,则称  $(G, M, I)$  是一个不完备形式背景。一个典型的不完备形式背景如表1所示,其中,若  $x \in G$  和  $m \in M$  之间的关系是缺失的(即无法确定是  $(x, m) \in I$  还是  $(x, m) \notin I$ ),则记为  $m(x) = *$ 。

表1 一个典型的不完备形式背景  
Table 1 A typical incomplete formal context

	$a$	$b$	$c$	$d$	$e$
1	1	1	0	0	1
2	1	*	0	0	0
3	0	0	0	1	1
4	0	*	0	1	1
5	0	1	1	1	0
6	0	1	1	*	0
7	1	0	1	*	1
8	1	0	1	0	1
9	1	*	1	1	1

## 2 概念格与粒化分析

等价类、极大相容类等是粗糙集中的基本粒。在粒内部,不同对象往往拥有相同的特征和相近的特征值,这就意味着粒内部不同对象之间的特征值是可以相互借鉴的。据此,本文尝试在概念格理论框架内探讨基于二元关系的信息粒化。

二元关系与形式背景存在着天然联系,它们之间可以相互表示。通常,二元关系有2种不同的类型,即内部二元关系和外部二元关系。其中,内部二元关系存在于单个集合的内部,例如,偏序集  $(V, \leq)$  中的序关系“ $\leq$ ”即是一种内部二元关系;外部二元关系是指存在于集合之间的关系,例如,  $(V, W, \leq)$  中的序关系“ $\leq$ ”即是一种外部二元关系。

**推论1** 内部二元关系是一种特殊的外部二元关系。例如,虽然  $(V, \leq)$  与  $(V, V, \leq)$  在形式上存在一定差异,但本质上却是相同的。

**推论2** 内部二元关系与外部二元关系均可以表示为形式背景。在下文中,无论是内部还是外部二元关系均统一表示为形式背景。

在粒化思维下,对于数据集的分析和处理,人们通常注重的是集合,而非单个元素;注重的是集合之间的关系,而非单个元素之间的关系。

**定义2** 设 $\mathcal{R}$ 是非空集合 $X$ 中的一个二元关系。若 $X \subseteq X$ 和 $Y \subseteq X$ 之间满足下述条件:

对于任意 $x \in X$ 和任意 $y \in Y$ ,均有 $(x,y) \in \mathcal{R}$ ,即 $X \times Y \subseteq \mathcal{R}$ ,称 $X$ 和 $Y$ 是关联的。设 $X$ 和 $Y$ 是关联的,即 $X \times Y \subseteq \mathcal{R}$ 。若不存在 $X_1 \times Y_1 \subseteq \mathcal{R}$ 满足 $X \times Y \subset X_1 \times Y_1$ ,称 $X$ 和 $Y$ 是极大关联的。

**定义3** 称 $K=(X,X,\mathcal{R})$ 是一个关系背景,其中 $\mathcal{R}$ 是非空集合 $X$ 中的一个二元关系。

**定理1** 在关系背景 $K=(X,X,\mathcal{R})$ 中,下列论述是等价的:

- 1)  $(X,Y) \in \mathcal{B}(K)$ ;
- 2)  $X$ 和 $Y$ 是极大关联的。

**证明** 当 $(X,Y) \in \mathcal{B}(K)$ 时,由定义1可知,对于任意 $x \in X$ 和任意 $y \in Y$ ,有 $(x,y) \in \mathcal{R}$ 成立,这就意味着 $X$ 和 $Y$ 是关联的。假设 $X$ 和 $Y$ 不是极大关联的,即存在 $X_1 \times Y_1 \subseteq \mathcal{R}$ 满足 $X \times Y \subset X_1 \times Y_1$ 。在此情形下,必然有 $Y \subset X^*$ 且 $X \subseteq Y^*$ 成立,或 $Y \subseteq X^*$ 且 $X \subset Y^*$ 成立,而非 $Y = X^*$ 且 $X = Y^*$ 成立,由此即得 $(X,Y) \notin \mathcal{B}(K)$ ,显然,该结论与已知条件 $(X,Y) \in \mathcal{B}(K)$ 是相互矛盾的。故当 $(X,Y) \in \mathcal{B}(K)$ 时, $X$ 和 $Y$ 是极大关联的。证毕。

**定义4** 设 $\mathcal{R}$ 是非空集合 $X$ 中的一个二元关系。若 $\mathcal{D}$ 是满足下述条件的 $X$ 中的最大子集,对于任意 $x,y \in \mathcal{D}$ ,均有 $(x,y) \in \mathcal{R}$ ,即 $X \times X \subseteq \mathcal{R}$ ;则称 $\mathcal{D}$ 是 $X$ 中的一个信息粒。

**定理2** 在关系背景 $K=(X,X,\mathcal{R})$ 中,下述结论成立:

$(\mathcal{D},\mathcal{D}) \in \mathcal{B}(K)$ ,当且仅当 $\mathcal{D}$ 是一个信息粒

**证明** 当 $(\mathcal{D},\mathcal{D}) \in \mathcal{B}(K)$ 成立时,由定理1得 $\mathcal{D}$ 和 $\mathcal{D}$ 是极大关联的,这也意味着 $\mathcal{D}$ 将满足下述条件:

对于任意 $x,y \in \mathcal{D}$ ,均有 $(x,y) \in \mathcal{R}$ ,即 $\mathcal{D} \times \mathcal{D} \subseteq \mathcal{R}$

假设 $\mathcal{D}_1$ 是满足上述条件的最大集合,其中 $\mathcal{D} \subset \mathcal{D}_1$ 。在此情形下, $\mathcal{D}_1 \times \mathcal{D}_1 \subseteq \mathcal{R}$ 是成立的,进而判定 $(\mathcal{D},\mathcal{D}) \notin \mathcal{B}(K)$ 。显然,该结论与 $(\mathcal{D},\mathcal{D}) \in \mathcal{B}(K)$ 是矛盾的。故当 $(\mathcal{D},\mathcal{D}) \in \mathcal{B}(K)$ 成立时, $\mathcal{D}$ 是满足上述条件的最大集合,即 $\mathcal{D}$ 是一个信息粒。

反过来,当 $\mathcal{D}$ 是信息粒时,由定义4即得 $\mathcal{D} \times \mathcal{D} \subseteq \mathcal{R}$ ,且不存在 $\mathcal{D}_1 \times \mathcal{D}_1 \subseteq \mathcal{R}$ 满足 $\mathcal{D} \subset \mathcal{D}_1$ 。在此情形下,由定义2可知 $\mathcal{D}$ 和 $\mathcal{D}$ 是极大关联的,进而由定理1得 $(\mathcal{D},\mathcal{D}) \in \mathcal{B}(K)$ 。故当 $\mathcal{D}$ 是一个信息粒,有 $(\mathcal{D},\mathcal{D}) \in \mathcal{B}(K)$ 成立。证毕。

在定理2中,对于任意 $(\mathcal{D},\mathcal{D}) \in \mathcal{B}(K)$ ,当 $\mathcal{R}$ 是

一个满足自反性和对称性的相容关系时,粒 $\mathcal{D}$ 本质上是一个相容类;当 $\mathcal{R}$ 是一个满足自反性、对称性和传递性的等价关系时,粒 $\mathcal{D}$ 是一个等价类。

### 3 基于等价类的数据分析模型

**定义5** 设 $(G,M,I)$ 是一个不完备形式背景, $0 \leq \delta \leq 1, 1 \leq \kappa \leq 2$ 。对于任意 $x \in G$ 和 $y \in G$ ,它们之间的相似度定义为

$$\text{Sim}(x,y) = \frac{|\text{Sam}_-(x,y)| + \delta \times |\text{Lac}_-(x,y)|}{|\text{Sam}_-(x,y)| + \delta \times |\text{Lac}_-(x,y)| + \kappa \times |\text{Dif}_-(x,y)|} \quad (3)$$

式中:

$$\text{Sam}_-(x,y) = \{a \in M | m(x) = m(y) = 1 \text{ or } m(x) = m(y) = 0\} \quad (4)$$

$$\text{Lac}_-(x,y) = \{m \in M | m(x) = * \text{ or } m(y) = *\} \quad (5)$$

$$\text{Dif}_-(x,y) = \{m \in M | m(x) = 0 \text{ and } m(y) = 1, \text{ or } m(x) = 1 \text{ and } m(y) = 0\} \quad (6)$$

参数 $\delta < 1$ 意味着 $\text{Lac}_-$ 对 $\text{Sim}$ 的影响力要小于 $\text{Sam}_-$ 和 $\text{Dif}_-$ ;参数 $1 < \kappa \leq 2$ 意味着 $\text{Dif}_-$ 对 $\text{Sim}$ 的影响力大于 $\text{Sam}_-$ 和 $\text{Lac}_-$ 。从相似度 $\text{Sim}$ 出发,首先会生成一个模糊相似关系矩阵 $\tilde{R}$ ,进而借助传递闭包算法将 $\tilde{R}$ 转化为模糊等价关系矩阵 $\bar{R}$ 。在此基础上,用户设置阈值参数 $0 \leq \lambda \leq 1$ ,可最终得到一个 $\lambda$ -阶等价关系矩阵 $R_\lambda$ 。 $R_\lambda$ 本质上是一个等价关系。此时,用户可依据定理2从 $R_\lambda$ 推导出若干个等价类,其中 $x \in G$ 决定的等价类记为 $[x]_\lambda$ 。

当 $\delta = \kappa = 1$ 时,用户从表1能推导出如表2、表3所示的模糊相似关系矩阵和模糊等价关系矩阵。当 $\lambda = 1$ 时, $\{1\}$ 、 $\{2\}$ 、 $\{3,4\}$ 、 $\{5,6\}$ 、 $\{7,8\}$ 、 $\{9\}$ 是等价类;当 $\lambda = 0.8$ 时, $\{1,2\}$ 、 $\{3,4\}$ 、 $\{5,6\}$ 、 $\{7,8,9\}$ 是等价类;当 $\lambda = 0.6$ 时, $\{1,2,\dots,9\}$ 是唯一的等价类。

表2 一个模糊相似关系矩阵  
Table 2 A fuzzy similarity relation matrix

	1	2	3	4	5	6	7	8	9
1	1.0	0.8	0.4	0.6	0.2	0.4	0.6	0.6	0.6
2	0.8	1.0	0.4	0.4	0.4	0.6	0.6	0.6	0.4
3	0.4	0.4	1.0	1.0	0.4	0.4	0.6	0.4	0.6
4	0.6	0.4	1.0	1.0	0.6	0.6	0.6	0.4	0.6
5	0.2	0.4	0.4	0.6	1.0	1.0	0.4	0.2	0.6
6	0.4	0.6	0.4	0.6	1.0	1.0	0.4	0.4	0.6
7	0.6	0.6	0.6	0.6	0.4	0.4	1.0	1.0	1.0
8	0.6	0.6	0.4	0.4	0.2	0.4	1.0	1.0	0.8
9	0.6	0.4	0.6	0.6	0.6	0.6	1.0	0.8	1.0

表3 一个模糊等价关系矩阵  
Table 3 A fuzzy equivalence relation matrix

	1	2	3	4	5	6	7	8	9
1	1	0.8	0.6	0.6	0.6	0.6	0.6	0.6	0.6
2	0.8	1	0.6	0.6	0.6	0.6	0.6	0.6	0.6
3	0.6	0.6	1	1.0	0.6	0.6	0.6	0.6	0.6
4	0.6	0.6	1.0	1	0.6	0.6	0.6	0.6	0.6
5	0.6	0.6	0.6	0.6	1	1.0	0.6	0.6	0.6
6	0.6	0.6	0.6	0.6	1.0	1	0.6	0.6	0.6
7	0.6	0.6	0.6	0.6	0.6	0.6	1	1.0	1.0
8	0.6	0.6	0.6	0.6	0.6	0.6	1.0	1	1
9	0.6	0.6	0.6	0.6	0.6	0.6	1.0	1	1

为避免单一粒度认知的片面性,对于任意缺失值的估计,用户可以设置多个阈值参数 $\lambda_1, \lambda_2, \dots, \lambda_N, 0 \leq \lambda_i \leq 1$ ,进而结合多个粒度下的分类结果去分析和求解问题。

**定义6** 设 $R_{\lambda_i}$ 是 $\lambda_i$ -阶模糊等价关系矩阵,称 $\lambda_i$ 是边界值,若不存在 $\lambda_k$ 满足 $\lambda_i < \lambda_k$ 且 $R_{\lambda_i} = R_{\lambda_k}$ 。

**准则1** 在不完备形式背景 $(G, M, I)$ 中,设 $m(x) = *$ , $\lambda_i$ 是边界值, $i = 1, 2, \dots, N$ 。

若 $\alpha \geq \beta$ ,则 $m(x) = 1$ ;若 $\alpha < \beta$ ,则 $m(x) = 0$ 。

其中

$$(\alpha\beta) = \begin{pmatrix} \lambda_1 & \lambda_2 & \dots & \lambda_N \end{pmatrix} \times \begin{pmatrix} \rho_{\lambda_1}(x, m, 1) & \phi_{\lambda_1}(x, m, 0) \\ \vdots & \vdots \\ \rho_{\lambda_N}(x, m, 1) & \phi_{\lambda_N}(x, m, 0) \end{pmatrix}$$

$$\rho_{\lambda_i}(x, m, 1) = \frac{|\{y \in [x]_{\lambda_i} \mid m(y) = 1\}|}{|[x]_{\lambda_i}|}$$

$$\phi_{\lambda_i}(x, m, 0) = \frac{|\{y \in [x]_{\lambda_i} \mid m(y) = 0\}|}{|[x]_{\lambda_i}|}$$

基于上述判定准则,用户可以对不完备形式背景进行预处理,从而得到一个完备的形式背景。接上例,对于 $d(7) = *$ ,依据准则1及下述计算结果,即得 $d(7) = 0$ 。

$$\begin{pmatrix} 1 & 0.8 & 0.6 \end{pmatrix} \times \begin{pmatrix} 0.00 & 0.50 \\ 0.33 & 0.33 \\ 0.44 & 0.33 \end{pmatrix} = \begin{pmatrix} 0.53 & 0.97 \end{pmatrix}$$

其中 $(\lambda_1 \lambda_2 \lambda_3) = (1 \ 0.8 \ 0.6)$

类似地,用户也可以判定其它缺失值,相应的判定结果如表4所示。在此基础上,复用经典概念格生成算法,用户可以从表4生成一个格代数结构,如图1所示。

表4 表1的一个完备化形式背景  
Table 4 A complete formal context from table 1

	a	b	c	d	e
1	1	1	0	0	1
2	1	1	0	0	0
3	0	0	0	1	1
4	0	0	0	1	1
5	0	1	1	1	0
6	0	1	1	1	0
7	1	0	1	0	1
8	1	0	1	0	1
9	1	0	1	1	1

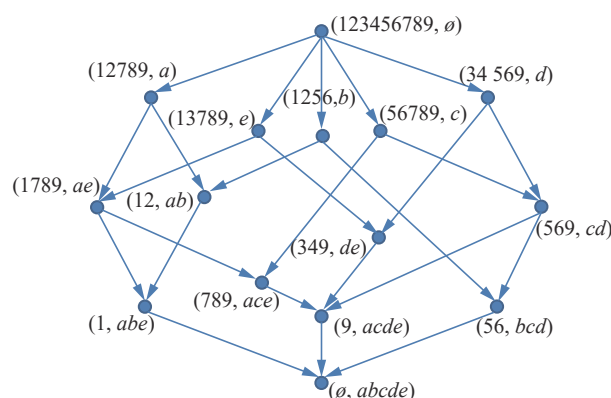


图1 基于准则1从表1导出的概念格结构  
Fig. 1 Concept lattice structure derived from table 1 based on criterion 1

## 4 基于极大相容类的数据分析模型

在不完备形式背景 $(G, M, I)$ 中,一个相容关系 $R_\sigma$ 可以描述为

$$R_\sigma = \{ (x, y) \in G \times G \mid \text{Sim}(x, y) \geq \sigma, 0 \leq \sigma \leq 1 \}$$

式中 $\text{Sim}$ 是定义5中构造的相似度模型。从定理2和 $R_\sigma$ 出发,用户即可推导出若干个极大相容类,并基于准则2对缺失值进行判定。

**准则2** 设 $m(x) = *$ , $x \in G$ 同时属于 $N$ 个极大相容类 $\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_N$ 。

若 $\hat{\alpha} \geq \hat{\beta}$ ,则 $m(x) = 1$ ;若 $\hat{\alpha} < \hat{\beta}$ ,则 $m(x) = 0$ 。

其中

$$(\hat{\alpha} \hat{\beta}) = (|\mathcal{D}_1| \mid |\mathcal{D}_2| \mid \dots \mid |\mathcal{D}_N|) \times \begin{pmatrix} \hat{\rho}_1(x, m, 1) & \hat{\phi}_1(x, m, 0) \\ \vdots & \vdots \\ \hat{\rho}_N(x, m, 1) & \hat{\phi}_N(x, m, 0) \end{pmatrix}$$



$$\hat{\rho}_i(x, m, 1) = \frac{|\{x \in \mathcal{D}_i | m(x) = 1\}| + 1}{|\{x \in \mathcal{D}_i | m(x) = 0\}| + 1}$$

$$\hat{\phi}_i(x, m, 0) = \frac{|\{x \in \mathcal{D}_i | m(x) = 0\}| + 1}{|\{x \in \mathcal{D}_i | m(x) = 1\}| + 1}$$

在表5中, 当  $\delta = \kappa = 1$ 、 $\sigma = 0.8$  时,  $\{1, 2\}$ 、 $\{2, 3, 4\}$ 、 $\{3, 4, 5\}$ 、 $\{5, 6, 7\}$ 、 $\{6, 7, 8\}$ 、 $\{7, 8, 9\}$ 、 $\{2, 7\}$  是极大相容类。基于准则2, 用户可以对不完备形式背景进行预处理, 进而得到一个完备的形式背景。例如, 对于  $d(7) = *$ , 基于准则2和下述计算结果

$$(3 \ 3 \ 3 \ 2) \times \begin{pmatrix} 1 & 1 \\ 1/3 & 3 \\ 1/3 & 3 \\ 2 & 1/2 \end{pmatrix} = (9 \ 22)$$

即可判定  $d(7) = 0$ , 其中,  $\mathcal{D}_1$ 、 $\mathcal{D}_2$ 、 $\mathcal{D}_3$ 、 $\mathcal{D}_4$  依次表示  $\{5, 6, 7\}$ 、 $\{6, 7, 8\}$ 、 $\{7, 8, 9\}$ 、 $\{2, 7\}$ 。

表5 一个不完备形式背景  
Table 5 A incomplete formal context

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>
1	0	0	0	1	1
2	1	0	0	1	1
3	1	0	0	1	0
4	1	*	0	1	0
5	1	0	1	1	0
6	1	0	1	0	0
7	1	*	1	*	1
8	1	0	1	0	1
9	1	1	1	0	1

类似地, 用户也可以判定其它缺失值, 相应的判定结果如表6所示。在此基础上, 复用经典概念格生成算法, 用户可以从表6生成一个格代数结构, 如图2所示。

表6 表5的一个完备化形式背景  
Table 6 Complete formal context generated from table 5

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>
1	0	0	0	1	1
2	1	0	0	1	1
3	1	0	0	1	0
4	1	0	0	1	0
5	1	0	1	1	0
6	1	0	1	0	0
7	1	0	1	0	1
8	1	0	1	0	1
9	1	1	1	0	1

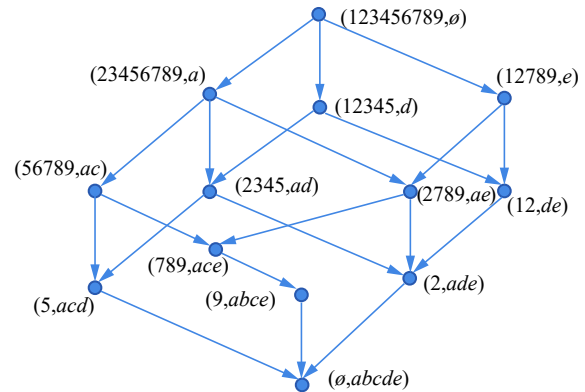


图2 基于准则2从表5导出的概念格结构  
Fig. 2 Concept lattice structure derived from Table 5 based on Criterion 2

基于准则1和准则2的模型本质上都属于间接处理模型, 即需要对原始数据集进行预处理, 给出缺失数据的估算值, 进而复用经典理论对完备形式背景进行分析处理。

在不完备形式背景  $(G, M, I)$  中, 一个相容关系  $R$  可以描述为

$$R = \{(x, y) \in G \times G | m(x) = m(y) \text{ or } m(x) = * \text{ or } m(y) = *, \forall m \in M\}$$

下文中, 由上述相容关系决定的极大相容类的集合记为  $\Delta$ 。

**准则3** 在任一极大相容类中, 不同对象在同一属性下应具有相同的属性值。在此情形下, 一个极大相容类  $\mathcal{D}$  在属性  $m \in M$  下的值域只可能有以下4种情况, 即  $\{1\}$ 、 $\{0\}$ 、 $\{1, *\}$ 、 $\{0, *\}$ , 而不可能是  $\{0, 1, *\}$ 。

当  $\mathcal{D}$  在  $m$  下的值域为  $\{1, *\}$  时, 若“\*”代表“0”, 则必然与准则3相违背, 此时, “\*”代表的只能是“1”; 同理, 若  $\mathcal{D}$  在  $m$  下的值域为  $\{0, *\}$ , 则“\*”代表的值只能是“0”。下文中,  $V_m(\mathcal{D}) = 1$  表示  $\mathcal{D}$  在  $m$  下的值域为  $\{1\}$  或  $\{1, *\}$ ;  $V_m(\mathcal{D}) = 0$  表示  $\mathcal{D}$  在  $m$  下的值域为  $\{0\}$  或  $\{0, *\}$ 。

**定义7** 设  $(G, M, I)$  是一个不完备形式背景, 对于任意集合  $X \in 2^G$ , 记

$$X^+ = \{m \in M | (m(x) = 1) \text{ or } (\exists \mathcal{D} \in \Delta, x \in \mathcal{D}, V_m(\mathcal{D}) = 1), \forall x \in X\}$$

对于任意集合  $B \in 2^M$ , 记

$$B^+ = \{x \in G | (m(x) = 1) \text{ or } (\exists \mathcal{D} \in \Delta, x \in \mathcal{D}, V_m(\mathcal{D}) = 1), \forall m \in B\}$$

与定义1中的经典算子相类似, 从上述算子出发同样可以导出一个格代数结构, 相关证明过程与经典算子类似, 在此不再详述。

例如, 在表5中, 从定理2和定义7出发, 可判定  $\{1\}$ 、 $\{2\}$ 、 $\{3, 4\}$ 、 $\{5\}$ 、 $\{6\}$ 、 $\{7, 8\}$ 、 $\{7, 9\}$  是极大相容类, 如表7所示, 进而基于定义7中的算子

生成图3所示的格代数结构。除(79, abce)之外,图3、图2中的其他结点均一致,这也从侧面反映了这样一个事实,即无论是基于准则2,还是基于准则3,所构造的模型可能会得到相似的结论。

表7 表5的一个粒化形式背景  
Table 7 A granular formal context from table 5

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>
{1}	0	0	0	1	1
{2}	1	0	0	1	1
{3, 4}	1	0	0	1	0
{5}	1	0	1	1	0
{6}	1	0	1	0	0
{7, 8}	1	0	1	0	1
{7, 9}	1	1	1	0	1

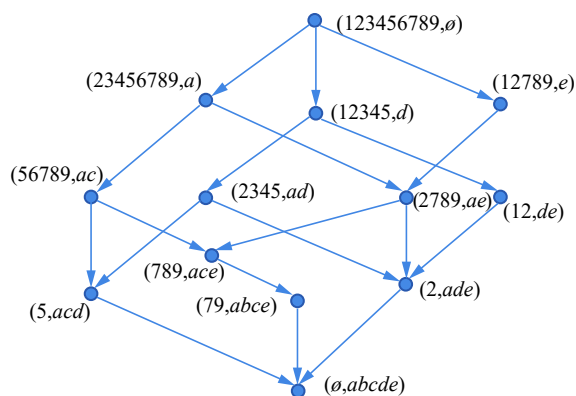


图3 基于准则3从表5导出的概念格结构

Fig. 3 Concept lattice structure derived from Table 5 based on Criterion 3

定义8 称  $\hat{K} = (\mathcal{A}, \mathcal{M}, \mathcal{J})$  是  $K = (G, \mathcal{M}, \mathcal{I})$  的粒化背景,若

$$(\mathcal{D}, m) \in \mathcal{J} \Leftrightarrow V_m(\mathcal{D}) = 1$$

在粒化背景  $\hat{K}$  中,对于任意  $\hat{X} \in 2^{\mathcal{A}}$ ,记

$$\hat{X}' = \{m \in \mathcal{M} \mid (\mathcal{D}, m) \in \mathcal{J}, \forall \mathcal{D} \in \hat{X}\}$$

对于任意  $B \in 2^{\mathcal{M}}$ ,记

$$B' = \{\mathcal{D} \in \mathcal{A} \mid (\mathcal{D}, m) \in \mathcal{J}, \forall m \in B\}$$

其中,  $(\mathcal{D}, m) \in \mathcal{J}$  亦可表示为  $m(\mathcal{D}) = 1$ 。

基于准则3的知识获取模型,其本质是将分析尺度放大,以极大相容类为研究对象,而非单个对象。在准则3下,任意  $\mathcal{D}$  在属性  $m$  的取值是确定的,即要么  $m(\mathcal{D}) = 1$  成立,要么  $m(\mathcal{D}) = 0$  成立,这就意味着,不完备形式背景可以转化为一个完备的粒化形式背景。事实上,在准则3下,无论是基于定义7,还是定义8,得到的格代数结构本质上是相同的,仅仅在外延表示形式上存在略微差异。

定理3 若  $(X, B) \in \mathcal{B}(K)$ ,  $(\hat{X}, B) \in \mathcal{B}(\hat{K})$ ,则

$$X = \bigcup_{\mathcal{D} \in \hat{X}} \mathcal{D}$$

证明 由定义7和定义8即得。证毕。

与准则1和准则2最大的不同是,基于准则3的知识获取模型无需对缺失信息进行预判定并给出估算值,而是直接在原始不完备形式背景上建立知识获取模型。

## 5 实例分析

表8是一个以医院病例为原型而得到的不完备形式背景,其中  $P_i (i=1, 2, \dots, 9)$  表示患者代码;  $(H, \text{yes})$ 、 $(H, \text{no})$ 、 $(M, \text{yes})$  等表示身体症状特征,其中  $H$ 、 $M$ 、 $T$ 、 $F$  依次表示 Headache、Muscle-pain、Temperature、Flu;若某患者具有某种症状,则在表8中用“1”来表示,反之用“0”来表示;若某诊断结论存在缺失,则用“\*”来表示。

表8 一个不完备形式背景  
Table 8 A incomplete formal context

患者	( <i>H</i> , yes)	( <i>H</i> , no)	( <i>M</i> , yes)	( <i>M</i> , no)	( <i>T</i> , high)	( <i>T</i> , very high)	( <i>T</i> , normal)	( <i>F</i> , yes)	( <i>F</i> , no)
$P_1$	0	1	1	0	1	0	0	1	0
$P_2$	1	0	0	1	1	0	0	1	0
$P_3$	1	0	1	0	1	1	0	1	0
$P_4$	0	1	1	0	0	0	1	0	1
$P_5$	*	0	0	1	1	0	0	0	1
$P_6$	0	1	1	0	*	1	0	1	0
$P_7$	1	0	0	1	1	0	*	1	0
$P_8$	1	0	0	1	1	0	0	*	*
$P_9$	0	1	1	0	*	0	1	0	1

当  $\delta = \kappa = 1$  时, 用户基于准则 1 得到的预处理结果, 如表 9 所示。其中, 当  $\lambda = 1$  时,  $\{1\}$ 、 $\{2, 5, 7, 8\}$ 、 $\{3\}$ 、 $\{6\}$ 、 $\{4, 9\}$  是等价类; 当  $\lambda = 0.89$  时,  $\{1,$

$6\}$ 、 $\{2, 5, 7, 8\}$ 、 $\{3\}$ 、 $\{4, 9\}$  是等价类; 当  $\lambda = 0.78$  时,  $\{1, 3, 6\}$ 、 $\{2, 5, 7, 8\}$ 、 $\{4, 9\}$  是等价类; 当  $\lambda = 0.67$  时,  $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$  是唯一等价类。

表 9 表 8 的一个完备化形式背景

Table 9 Complete formal context generated from table 8

患者	(H, yes)	(H, no)	(M, yes)	(M, no)	(T, high)	(T, very high)	(T, normal)	(F, yes)	(F, no)
$P_1$	0	1	1	0	1	0	0	1	0
$P_2$	1	0	0	1	1	0	0	1	0
$P_3$	1	0	1	0	1	1	0	1	0
$P_4$	0	1	1	0	0	0	1	0	1
$P_5$	1	0	0	1	1	0	0	0	1
$P_6$	0	1	1	0	1	1	0	1	0
$P_7$	1	0	0	1	1	0	0	1	0
$P_8$	1	0	0	1	1	0	0	1	0
$P_9$	0	1	1	0	0	0	1	0	1

当  $\delta = \kappa = 1$ 、 $\sigma = 0.78$  时, 用户基于准则 2 同样可以得到如表 9 所示的预处理结果, 其中,  $\{1, 6\}$ 、 $\{2, 5, 7, 8\}$ 、 $\{3, 6\}$ 、 $\{4, 9\}$  是极大相容类。

事实上, 表 9 中的完备化结果与下述事实性规则是吻合的, 这也在一定程度上反应了准则 1 和准则 2 的合理性和有效性。例如, 当患者体温是“very high”时, 则其体温一定也是“high”; 当患者体温“normal”时, 则其体温一定不是“high”和“very high”; 当患者“Flu, yes”时, 则一定不是“Flu, no”等。

事实性规则:  $(T, \text{very high}) \rightarrow (T, \text{high})$ 、 $(T, \text{normal}) \nrightarrow (T, \text{high})$ 、 $(T, \text{normal}) \nrightarrow (T, \text{very high})$ 、 $(H, \text{yes}) \nrightarrow (H, \text{no})$ 、 $(H, \text{no}) \nrightarrow (H, \text{yes})$ 、 $(F, \text{no}) \nrightarrow (F, \text{yes})$ 、 $(F, \text{yes}) \nrightarrow (F, \text{no})$ 。

在概念格经典理论中, 概念内涵和蕴涵规则都占有极其重要的地位。本文认为, 无论间接处理模型, 还是直接处理模型, 它们都应该得到相同的内涵集和蕴涵规则集 (内涵集和蕴涵规则集是相互唯一决定的)。据此, 本文提出了如下有效性判定准则:

有效性判定准则对于同一个不完备形式背景, 若多个不同模型得到的蕴涵规则集合或内涵集合是相同的, 则本文认为这些模型在一定程度上是有效的, 得到的结果在一定程度上也是可信的。

**定义 9** 设  $K = (G, M, I)$  为一个形式背景,  $B_1, B_2 \subseteq M$ 。若  $B_1' \subseteq B_2'$ , 则称  $B_1 \rightarrow B_2$  是  $K$  中的蕴涵规则。

准则 3 相对应的极大相容类有  $\{1\}$ 、 $\{2, 7, 8\}$ 、

$\{3\}$ 、 $\{4, 9\}$ 、 $\{5, 8\}$ 、 $\{6\}$ 。在此基础上, 基于定义 7 或定义 8, 用户可得到相应的格代数结构。经验证明, 基于准则 3 直接从表 8 得到的概念内涵集与基于准则 1 或准则 2 间接从表 8 中得到的概念内涵集是相同的。显然, 在这种情形下, 依据有效性判定准则给出的判定原理, 可以在一定程度认为基于准则 1、准则 2 或准则 3 构建的模型是合理的, 相应的求解结果也是可信的。

总体来看, 准则 1 和准则 2 需要引入阈值参数, 属于间接处理模型; 而准则 3 则无需引入阈值, 属于直接处理模型。在实际应用中, 如果数据缺失量大, 本文倾向于选用直接处理模型, 因为先选择间接模型可能会导致原始数据集失真; 如果数据缺失量少, 则无论是直接模型还是间接模型, 都可以选用。

## 6 结束语

为消除不完备信息带来的影响, 使概念格模型具有更强的数据处理能力与更广的应用领域, 本文尝试将经典形式背景中的知识获取方法进一步拓展到不完备形式背景中, 依次探讨了基于等价类的分析模型和基于极大相容类的分析模型。在实际应用中, 用户既可以选择基于准则 1 或准则 2 的间接处理模型, 也可以选择基于准则 3 的直接处理模型。此外, 对于模型的机制有效性和结果可信性, 本文也尝试性的进行了探讨, 并提出了一些初步的验证方法。相信本文所做的工作能为下一步相关研究提供一些有益的思路。

## 参考文献:

- [1] WILLE R. Restructuring lattice theory: an approach based on hierarchies of concepts[M]. RIVAL I. Ordered Sets. Dordrecht: Reidel, 1982: 445–470.
- [2] 张文修, 姚一豫, 梁怡. 粗糙集与概念格 [M]. 西安: 西安交通大学出版社, 2006.
- [3] 仇国芳, 张志霞, 张伟. 基于粗糙集方法的概念格理论研究综述 [J]. 模糊系统与数学, 2014, 28(1): 168–177.  
QIU Guofang, ZHANG Zhixia, ZHANG Wei. A survey for study on concept lattice theory via rough set[J]. Fuzzy systems and mathematics, 2014, 28(1): 168–177.
- [4] 王国胤. Rough 集理论在不完备信息系统中的扩充 [J]. 计算机研究与发展, 2002, 39(10): 1238–1243.  
WANG Guoyin. Extension of rough set under incomplete information systems[J]. Journal of computer research and development, 2002, 39(10): 1238–1243.
- [5] KRYSZKIEWICZ M. Rough set approach to incomplete information systems[J]. Information sciences, 1998, 112(1/2/3/4): 39–49.
- [6] KRYSZKIEWICZ M. Rules in incomplete information systems[J]. Information sciences, 1999, 113(3/4): 271–292.
- [7] STEFANOWSKI J, TSOUKIAS A. Incomplete information tables and rough classification[J]. Computational intelligence, 2001, 17(3): 546–566.
- [8] LEUNG Y, LI Deyu. Maximal consistent block technique for rule acquisition in incomplete information systems[J]. Information sciences, 2003, 153: 85–106.
- [9] ZHANG Wenxiu, MI Jusheng. Incomplete information system and its optimal selections[J]. Computers & mathematics with applications, 2004, 48(5/6): 691–698.
- [10] YAO Yiyu. Interval sets and three-way concept analysis in incomplete contexts[J]. International journal of machine learning and cybernetics, 2017, 8(1): 3–20.
- [11] LI Jinhai, MEI Changlin, LV Yuejin. Incomplete decision contexts: approximate concept construction, rule acquisition and knowledge reduction[J]. International journal of approximate reasoning, 2013, 54(1): 149–165.
- [12] LI Meizheng, WANG Guoyi. Approximate concept construction with three-way decisions and attribute reduction in incomplete contexts[J]. Knowledge-based systems, 2016, 91: 165–178.
- [13] 张慧雯, 刘文奇, 李金海. 完备形式背景下近似概念格的公理化方法 [J]. 计算机科学, 2015, 42(6): 67–70, 92.  
ZHANG Huiwen, LIU Wenqi, LI Jinhai. Axiomatic characterizations of approximate concept lattices in incomplete contexts[J]. Computer science, 2015, 42(6): 67–70, 92.
- [14] 智慧来. 不完备形式背景上的知识表示 [J]. 计算机科学, 2015, 42(1): 276–278.  
ZHI Huilai. Knowledge representation on incomplete formal context[J]. Computer science, 2015, 42(1): 276–278.
- [15] 李磊军, 李美争, 解滨, 等. 三支决策视角下概念格的分析 and 比较 [J]. 模式识别与人工智能, 2016, 29(10): 951–960.  
LI Leijun, LI Meizheng, XIE Bin, et al. Analysis and comparison of concept lattices from the perspective of three-way decisions[J]. Pattern recognition and artificial intelligence, 2016, 29(10): 951–960.
- [16] 王振, 魏玲. 基于单边区间集概念格的不完备形式背景的属性约简 [J]. 计算机科学, 2018, 45(1): 73–78.  
WANG Zhen, WEI Ling. Attribute reduction of partially-known formal concept lattices for incomplete contexts[J]. Computer science, 2018, 45(1): 73–78.
- [17] GANTER B, WILLE R. Formal concept analysis: mathematical foundations[M]. Berlin: Springer-Verlag, 1999.

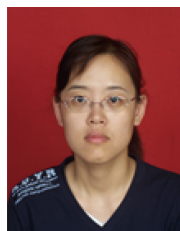
## 作者简介:



王雯, 女, 1984 年生, 硕士研究生, 主要研究方向为智能控制理论及应用、粒计算。



康向平, 男, 1982 年生, 博士研究生, 主要研究方向为粗糙集、概念格、粒计算。



武燕, 女, 1982 年生, 硕士研究生, 主要研究方向为智能信息处理。