

# 基于粒计算的概念格拓展模型

康向平<sup>1,2</sup>, 苗夺谦<sup>1,2</sup>

(1. 同济大学 电子与信息工程学院, 上海 201804;

2. 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)

**摘要:** 将粒计算融入到概念格研究中, 结合相似度模型和概念格结构信息, 提出一种基于粒计算的概念格拓展模型, 其有助于扩展经典概念的内涵和外延, 也有助于压缩概念的规模. 该模型是概念格和粒计算融合研究的一次有益探索和尝试, 同时对概念格拓展也不失为一种有效手段.

**关键词:** 概念格; 粒计算; 关系粒; 粒化概念格

**中图分类号:** TP182

**文献标志码:** A

## Expanded Concept Lattice Based on Granular Computing

KANG Xiangping<sup>1,2</sup>, MIAO Duoqian<sup>1,2</sup>

(1. College of Electronics and Information Engineering, Tongji University, Shanghai 201804, China; 2. Key Laboratory of Embedded System and Service Computing of the Ministry of Education, Tongji University, Shanghai 201804, China)

**Abstract:** By introducing granular computing into concept lattice, and integrating similarity measure model and structure information of concept lattice, the paper proposes an expanded concept lattice model based on granular computing. It can help to expand intent and extent of classical concept, and also can effectively reduce the scale of concepts to some extent. The model is not only a useful attempt and exploration for the fusion of these two theories, but also an effective means of expanding the classical concept lattice.

**Key words:** concept lattice; granular computing; relation granules; granulation concept lattice

德国 Wille 教授于 1982 年提出了形式概念分析<sup>[1]</sup>, 即概念格理论, 该理论是格序论的一个重要分支, 具有良好的数学性质, 主要用于概念的发现、排

序和显示. 在概念格构筑的体系中, 概念、完备格结构、伽罗瓦连接等是核心要素. 从内涵和外延 2 个不同角度对概念进行刻画和描述, 有助于深化人们对现实世界中概念的概要认识和准确理解; 完备格结构作为一种基于序关系建立起来的概念层次模型, 直观地反映了概念之间的泛化/特化关系; 伽罗瓦连接则是概念格体系的基石. 概念格作为一种强有力的数据分析工具, 受到了学者的广泛关注, 其在基础理论<sup>[2-7]</sup>、格生成算法<sup>[8-9]</sup>、规则获取<sup>[10-14]</sup>、属性约简<sup>[15-16]</sup>等方面, 以及在不同领域的应用<sup>[17-18]</sup>取得了大量的研究成果. 近年来, 许多学者系统研究了概念格与模糊集<sup>[19-21]</sup>、粗糙集<sup>[22-28]</sup>、神经网络<sup>[29]</sup>、概率论<sup>[30]</sup>、粒计算<sup>[31-33]</sup>等知识发现方法之间的联系, 旨在探索复杂数据处理的新途径. 这种多技术之间的渗透和交叉, 有助于推动概念格向深层次和多元化方向发展. Burusco<sup>[21]</sup>将模糊集引入概念格, 用于度量对象和属性之间的不确定性关系, 有助于解决复杂问题中的模糊信息和不确定信息; Dias 等<sup>[29]</sup>探讨了概念格和神经网络融合理论; Jiang 等<sup>[30]</sup>将概率论引入到概念格中, 提出了一种新的数据挖掘方法; Yao<sup>[28]</sup>将粗糙集中的上下近似思想引入到概念格中, 探讨了概念格中的近似算子; Kang 等<sup>[32]</sup>将粒计算引入概念格, 提出了概念格在不同粒度下的知识获取模型, 该模型为概念格构建和规则获取提供了统一的基于模糊粒化基的求解方法, 不仅可以有效简化复杂的概念格结构和压缩庞大的概念规模, 而且就决策规则集的完备性和无冗余性做了一些有益的尝试. 关于概念格的最新相关研究成果, 详见文献<sup>[34]</sup>.

在经典概念格中, 对于概念的刻画是精确的. 然

收稿日期: 2015-08-11

基金项目: 国家自然科学基金(61273304, 61202170); 中国博士后科学基金(2014M560352); 教育部高等学校博士学科点专项科研基金(20130072130004)

第一作者: 康向平(1982—), 男, 工学博士, 主要研究方向为概念格、粗糙集、粒计算等. E-mail: tongji\_kangxp@sina.com

通讯作者: 苗夺谦(1964—), 男, 教授、博士生导师, 工学博士, 主要研究方向为智能信息处理、人工智能与模式识别、数据挖掘、粗糙集、粒计算等. E-mail: miaoduoqian@163.com

而,这种精确性在实际应用中可能存在一定的局限性,不利于人们从看似不相关的数据中发现一些有价值的信息.在此情况下,就需要考虑进一步拓展概念的内涵和外延,将概念“可能”包含的一些对象也纳入到其外延中.例如,在地震预测中,若一些测量结果符合地震的“大部分”特征而非“全部”特征时,专家一般提前判定地震可能会发生,而不是等到“全部”特征符合后再做出判定.此外,在一些大规模复杂的数据集中,概念的这种精确性通常会使得知识获取过程淹没在一些不必要的细节中,并最终导致海量概念和复杂的格结构,不利于更好地从整体认识和理解问题.针对上述问题,本文尝试引入粒计算.粒计算以粒、粒度、粒推理等为核心要素,通过模拟人类思维方式在大规模复杂数据建模与分析中独特的优势,可以降低知识获取的“分辨率”,扩大知识度量的“尺度”,有助于人们对问题的宏观认识和理解.

目前,关于概念格与粒计算的融合研究还尚不多见<sup>[31-33]</sup>,据此,本文将粒计算融入到概念格研究中,提出了粒化概念格和粒化概念.粒计算的引入不仅有助于扩展经典概念的内涵和外延,而且也有助于压缩概念的规模并简化复杂的格结构.

### 1 概念格简介

形式背景是一个三元组  $(G, M, I)$ ,其中  $G$  和  $M$  是非空有限集,  $I \subseteq G \times M$  是一个从集合  $G$  到集合  $M$  的二元关系.表 1 是一个从超市采集的数据集,其中  $1, 2, \dots, 9$  表示顾客编号;  $a, b, \dots, e$  表示商品编号,若某位消费者购买了某种商品,则用“ $\times$ ”进行标记.

表 1 一个典型的形式背景

Tab.1 A classical formal context					
顾客编号	商品编号				
	a	b	c	d	e
1	$\times$		$\times$	$\times$	$\times$
2		$\times$		$\times$	$\times$
3		$\times$	$\times$		
4	$\times$		$\times$		$\times$
5			$\times$	$\times$	
6		$\times$	$\times$		$\times$
7		$\times$		$\times$	
8	$\times$		$\times$	$\times$	$\times$
9		$\times$	$\times$		

定义 1 设  $K=(G, M, I)$  是一个形式背景,  $A \subseteq G, B \subseteq M$ . 若  $A'=B$  且  $B'=A$ , 则称  $(A, B)$  是一个概念, 此时称  $A$  和  $B$  分别为概念的外延和内涵. 其中, 算子  $A'$  被描述为

$$A' = \{m \in M \mid (g, m) \in I, \forall g \in A\}$$

相应地,算子  $B'$  被定义为

$$B' = \{g \in G \mid (g, m) \in I, \forall m \in B\}$$

对于任意概念  $(A_1, B_1)$  和  $(A_2, B_2)$ , 它们之间的偏序关系定义为

$$(A_1, B_1) \leq (A_2, B_2) \Leftrightarrow A_1 \subseteq A_2 \Leftrightarrow B_2 \subseteq B_1$$

事实上,  $(B(K), \leq)$  是一个完备格, 其中  $B(K)$  是一个由  $K$  中所有概念组成的集合. 图 1 即是表 1 中的概念格.

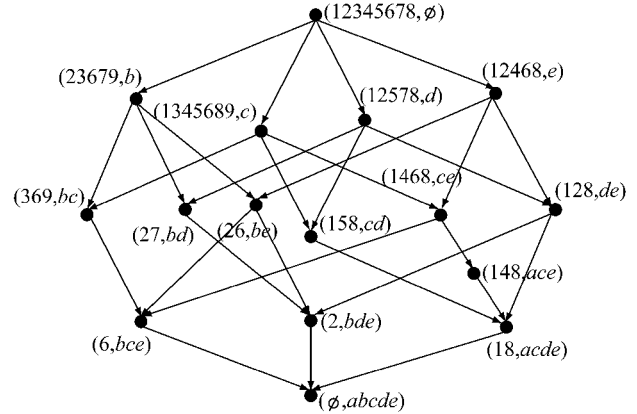


图 1 由表 1 生成的概念格

Fig.1 Concept lattice with respect to Tab.1

引理 1 在形式背景  $(G, M, I)$  中, 设  $A, A_1, A_2 \subseteq G, B, B_1, B_2 \subseteq M$ , 有下列结论成立: ①  $A_1 \subseteq A_2 \Rightarrow A'_2 \subseteq A'_1$ ; ②  $B_1 \subseteq B_2 \Rightarrow B'_2 \subseteq B'_1$ ; ③  $A \subseteq A'', B \subseteq B''$ ; ④  $A' = A'', B' = B''$ .

定义 2 设  $K=(G, M, I)$  是一个形式背景, 对于任意  $g \in G$ , 称  $\gamma g$  为对象概念, 即

$$\gamma g = (g'', g')$$

对于任意  $m \in M$ , 称  $\mu m$  为属性概念, 即

$$\mu m = (m', m'')$$

$\gamma(G)$  表示由所有对象概念组成的集合,  $\mu(M)$  表示由所有属性概念组成的集合. 对象概念和属性概念作为概念格中的基础概念, 扮演着重要角色. 通常, 图 1 也可以简化表示为图 2, 其中“ $m$ ”代表  $\mu m$ , “ $g$ ”代表  $\gamma g$ .

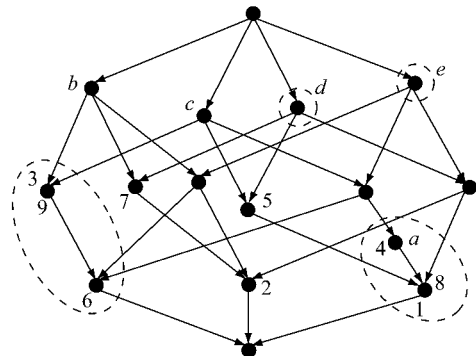


图 2 图 1 的简化概念格

Fig.2 Simplified concept lattice with respect to Fig.1

**引理 2** 在形式背景  $K=(G, M, I)$  中, 设  $B \subseteq M$ , 则

$$B' = \bigcap \{L \in \text{Int}(K) \mid B \subseteq L\}$$

其中,  $\text{Int}(K)$  是一个由所有概念内涵组成的集合.

关于概念格的更详细介绍, 请查阅文献 [2].

## 2 粒化形式背景

概念格的粒化本质上是形式背景的粒化, 据此, 本文探讨了形式背景中的对象粒和关系粒, 并在此基础上构建了粒化形式背景.

### 2.1 相似度模型和对象粒

在传统的相似度度量模型中, 特征模型是最常见的度量模型之一. 特征模型主要包含两大类, 一类是基于公共特征集构造的对称性特征模型, 例如:

$$\rho(g_i, g_j) = \begin{cases} 1, & g'_i = g'_j \\ \frac{1}{|M|} (|g'_i \cap g'_j| + \sigma |(M \setminus g'_i) \cap (M \setminus g'_j)|), & g'_i \neq g'_j \end{cases}$$

其中, 参数  $\sigma$  体现了公共特征集  $g_i \cap g_j$  和公共非特征集  $M \setminus g'_i \cap M \setminus g'_j$  对相似度的影响是不对称的.

通常, 人们会从多个角度去思考和分析问题, 虽然看待问题的角度不同, 但在本质是相同的. 例如, 当对人的品行进行评价时, 虽然“他们都是品德高尚的人”和“他们都不具有卑劣、庸俗、自私等特征”是两种不同的评价, 但它们却从不同方面反映了相类似的信息. 类似地, 在上述模型中, 对象  $g_i$  和对象  $g_j$  的“公共特征集(即  $g_i \cap g_j$ )”和“公共非特征集(即  $M \setminus g'_i \cap M \setminus g'_j$ )”是评价  $g_i$  和  $g_j$  相似程度的 2 个不同方面, 它们是构建模型的 2 种重要因素. 换句话说, 上述模型是一种融合了“公共特征集”和“公共非特征集”的对称相似度模型. 例如, 在表 1 中, 当  $\sigma=1$  时, 基于定义 3 中的相似度模型, 可以得到一个如表 2 所示的模糊相似矩阵.

表 2 模糊相似矩阵

Tab.2 Fuzzy similarity matrix

	1	2	3	4	5	6	7	8	9
1	1.0	0.4	0.2	0.8	0.6	0.4	0.2	1.0	0.2
2	0.4	1.0	0.4	0.2	0.4	0.6	0.8	0.4	0.4
3	0.2	0.4	1.0	0.4	0.6	0.8	0.6	0.2	1.0
4	0.8	0.2	0.4	1.0	0.4	0.6	0.0	0.8	0.4
5	0.6	0.4	0.6	0.4	1.0	0.4	0.6	0.6	0.6
6	0.4	0.6	0.8	0.6	0.4	1.0	0.4	0.4	0.8
7	0.2	0.8	0.6	0.0	0.6	0.4	1.0	0.2	0.6
8	1.0	0.4	0.2	0.8	0.6	0.4	0.2	1.0	0.2
9	0.2	0.4	1.0	0.4	0.6	0.8	0.6	0.2	1.0

**定理 1** 在形式背景  $(G, M, I)$  中, 对于任意对象  $g_i, g_j \in G$ , 有下述性质成立:

$$\rho_1(g_i, g_j) = \frac{|B_i \cap B_j|}{|B_i \cup B_j|}$$

其中  $B_i$  和  $B_j$  分别是对象  $g_i$  和  $g_j$  的特征集. 另一类是基于差异特征集构造的不对称性特征模型, 例如:

$$\rho_2(g_i, g_j) = \frac{|B_i \cap B_j|}{|B_i \cap B_j| + \alpha |B_i - B_j| + \beta |B_j - B_i|},$$

$$0 \leq \alpha, \beta \leq 1, \alpha + \beta = 1$$

其中,  $B_i$  和  $B_j$  分别是对象  $g_i$  和  $g_j$  的特征集. 该模型同时引入了共同特征集和差异特征集, 相对于差异特征集, 共同特征对于相似度的影响要更大一些. 此外, 参数  $\alpha$  和  $\beta$  体现了 2 个不同的差异特征集  $B_i - B_j$  和  $B_j - B_i$  对于相似度的影响是不对称的.

本文研究的重点是对称相似度模型, 详见定义 3.

**定义 3** 在形式背景  $(G, M, I)$  中, 设  $0 \leq \sigma \leq 1$ , 对于任意  $g_i, g_j \in G$ , 定义

- (1)  $0 \leq \rho(g_i, g_j) \leq 1$ .
- (2) 若  $g'_i = g'_j$ , 那么  $\rho(g_i, g_j) = 1$ .
- (3) 若  $g'_j \subseteq x' \subseteq g'_i$  且  $g'_j \subseteq y' \subseteq g'_i$ , 则  $\rho(g_i, g_j) \leq \rho(x, y)$ .
- (4) 若  $g'_j \subseteq z' \subseteq g'_i$ , 则  $\rho(g_i, g_j) \leq \rho(g_i, z)$  且  $\rho(g_i, g_j) \leq \rho(g_j, z)$ .

**证明** 由  $\rho$  的定义即得性质(1)和(2)成立.

(3) 当  $g'_j \subseteq x' \subseteq g'_i, g'_j \subseteq y' \subseteq g'_i$  时, 有  $M \setminus g'_i \subseteq M \setminus x' \subseteq M \setminus g'_j$  且  $M \setminus g'_i \subseteq M \setminus y' \subseteq M \setminus g'_j$  进而

$M \setminus g'_i \cap M \setminus g'_j \subseteq M \setminus x' \cap M \setminus y'$  且  $g'_i \cap g'_j \subseteq x' \cap y'$  显然, 结论  $\rho(g_i, g_j) \leq \rho(x, y)$  成立.

(4) 由  $g'_j \subseteq z' \subseteq g'_i$  可知  $M \setminus g'_i \subseteq M \setminus z' \subseteq M \setminus g'_j$ , 进而有

$g'_i \cap g'_j \subseteq g'_i \cap z'$  且  $M \setminus g'_i \cap M \setminus g'_j \subseteq M \setminus g'_i \cap M \setminus z'$  显然, 有结论  $\rho(g_i, g_j) \leq \rho(g_i, z)$  成立. 类似地, 有  $\rho(g_i, g_j) \leq \rho(g_j, z)$  成立. 证毕.

由上述定理可知, 本文定义的相似度模型是合理的. 事实上, 依据该模型可以导出一个模糊相似关系矩阵

$$R = (\rho(g_i, g_j))_{|G| \times |G|}$$

进而可以借助传递闭包算法探讨对象集  $G$  的粒化. 首先基于模糊传递闭包算法生成  $R$  的模糊等价关系矩阵  $R^+$ . 早期传递闭包算法往往伴随着较高的时间复杂度为  $O(n^3 \log n)$ , 为此许多学者对算法进行了改进, 例如, Lee<sup>[35]</sup> 提出的算法的时间复杂度为  $O(n^2)$ . 然后在模糊等价关系矩阵  $R^+$  中引入参数

$0 \leq \lambda \leq 1$ , 求解  $R^+$  的  $\lambda$ -截等价关系矩阵  $R_\lambda^+$ . 本质上,  $R_\lambda^+$  是对象集  $G$  中的一个等价关系,  $G_\lambda = G/R_\lambda^+$  是  $G$  的一个划分. 其中,  $X \in G/R_\lambda^+$  是一个等价类, 也称为一个对象粒. 在下文中, 由  $g \in G$  决定的对象粒记为

$$[g] = \{h \in G \mid (g, h) \in R_\lambda^+\}$$

$\lambda$  的值越小, 相应  $G$  中的对象粒就越大. 一个从表 2 导出的模糊等价矩阵如表 3 所示. 在表 3 中, 当  $\lambda = 0.8$  时, 有

$$G_\lambda = \{[1], [2], [3], [5]\}$$

其中,  $[1] = [4] = [8] = \{1, 4, 8\}$ ,  $[2] = [7] = \{2, 7\}$ ,  $[3] = [6] = [9] = \{3, 6, 9\}$ ,  $[5] = \{5\}$ .

表 3 模糊等价矩阵

Tab.3 Fuzzy equivalence matrix

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

### 2.2 关系粒和粒化形式背景

本节将重点探讨关系粒和粒化背景, 其中关系粒在粒化背景中扮演了重要角色.

**定义 4** 设  $K = (G, M, I)$  是一个形式背景, 对于任意  $g \in G, m \in M$ , 称  $[g] \times m \in G_\lambda \times M$  是一个关系粒; 若  $I_\lambda \subseteq G_\lambda \times M$ , 则称  $K_\lambda = (G_\lambda, M, I_\lambda)$  是  $K$  的一个粒化形式背景.

由上述定义可知, 任意关系粒  $[g] \times m$  是否满足  $([g], m) \in I_\lambda$ , 对于构造粒化背景至关重要, 需要给出科学合理的判定准则. 据此, 本文提出了两类不同的判定方法, 即一类是基于形式背景  $K$  的判定方法, 另一类是基于概念格  $(B(K), \leq)$  的判定方法.

**定义 5** 在形式背景  $(G, M, I)$  中, 对于任意  $[g] \times m \in G_\lambda \times M$ , 定义

$$\eta_\lambda(g, m) = \frac{1}{|[g]|} \times |[g] \cap m'|$$

其中  $\eta_\lambda$  是一个可信度函数.

显然,  $\eta_\lambda$  越大,  $([g], m) \in I_\lambda$  的可信度就越高;  $\eta_\lambda$  越小,  $([g], m) \notin I_\lambda$  的可信度就越高. 据此, 引入参数  $0 \leq \delta \leq 1$ , 并将其作为判定  $([g], m) \in I_\lambda$  是否可信的阈值. 依据实际需要, 用户可以灵活调整阈值参数.

**准则 1** 对于粒  $[g] \times m$ , 若  $\eta_\lambda(g, m) \geq \delta$ , 则  $([g], m) \in I_\lambda$ ; 若  $\eta_\lambda(g, m) < \delta$ , 则  $([g], m) \notin I_\lambda$ .

例如, 基于准则 1, 由表 1 可以导出如表 4 所示的粒化形式背景, 其中  $\lambda = 0.8$  且  $\delta = 0.6$ .

表 4 基于准则 1 导出的表 1 的粒化背景

Tab.4 Granulation context deduced from Tab.1

based on Criterion 1					
	a	b	c	d	e
[1]	×		×	×	×
[2]		×		×	
[3]		×	×		
[5]			×	×	

**引理 3** 设  $(G, M, I)$  是一个形式背景, 对于任意  $g \in G$  和  $m \in M, (g, m) \in I$  当且仅当  $\gamma g \leq \mu m$ .

依据上述引理, 本文将尝试融入概念格的结构信息, 探讨基于  $(B(K), \leq)$  的判定方法.

**准则 2\_a** 设  $[g] \times m$  是一个关系粒, 对于  $\forall g_1 \in [g]$ , 如果  $\gamma g_1 \leq \mu m$ , 那么  $([g], m) \in I_\lambda$ ; 如果存在  $g_1 \in [g]$  满足  $\gamma g_1 \not\leq \mu m$ , 那么  $([g], m) \notin I_\lambda$ .

上述准则等价于如下关系:

若  $\eta_\lambda(g, m) = 1$ , 则  $([g], m) \in I_\lambda$ ; 若  $\eta_\lambda(g, m) \neq 1$ , 则  $([g], m) \notin I_\lambda$ .

由上述讨论可知, 准则 2\_a 是准则 1 的一种特殊情况, 但其判定条件较准则 1 太过苛刻, 无助于用户从看似不相关的数据中发现有价值的信息. 据此, 可以借助下确界运算“ $\wedge$ ”将准则 2\_a 进一步弱化为准则 2\_b 和 2\_c.

**准则 2\_b** 设  $[g] \times m$  是一个关系粒, 若  $\wedge \gamma([g]) \leq \mu m$ , 则  $([g], m) \in I_\lambda$ ; 若  $\wedge \gamma([g]) \not\leq \mu m$ , 则  $([g], m) \notin I_\lambda$ .

**准则 2\_c** 设  $[g] \times m$  是一个关系粒, 若  $\wedge \gamma([g]) \not\leq \mu m$ , 则  $([g], m) \notin I_\lambda$ . 当  $\wedge \gamma([g]) \leq \mu m$  时, 若满足下述条件:

对于  $\forall H \subseteq [g]$ , 满足  $\wedge \gamma(H) \leq \mu m$  则  $([g], m) \in I_\lambda$ ; 若满足下述条件:

$\exists H \subseteq [g]$ , 满足  $\wedge \gamma(H) \not\leq \mu m$  则  $([g], m) \notin I_\lambda$ , 其中  $|H| = 2$ .

在上述基于概念格结构信息的判定准则中, 由于  $\wedge \gamma([g]) \leq \wedge \gamma(H) \leq \gamma g \leq \mu m$ , 其中  $H \subseteq [g]$ , 显然, 若作为判定条件,  $\wedge \gamma(H) \leq \mu m$  要弱于  $\gamma g \leq \mu m$ , 但强于  $\wedge \gamma([g]) \leq \mu m$ . 即准则 2\_c 判定条件的强弱介于准则 2\_b 和准则 2\_a 之间. 尤其是当  $|[g]|$  较大时, 准则 2\_b 中的判定条件  $\wedge \gamma([g]) \leq \mu m$  将弱化, 此时相应的判定结果有可能是合理的. 本质上, 定理 2 和定理 3 可以从可信度函数的角度客观地反映准则 2\_c 的合理性.

**定理 2** 在  $K = (G, M, I)$  中, 设  $A \subseteq G, m \in M$ ,

有下述结论成立:

- (1) 若  $\wedge \gamma(A) \not\leq \mu m$ , 则  $(A \times m) \cap I = \emptyset$ .
- (2) 若存在  $g \in A$ , 满足  $(g, m) \in I$ , 则  $\wedge \gamma(A) \leq \mu m$ .

证明:当  $\wedge \gamma(A) \not\leq \mu m$  时, 假设存在  $g \in A$  满足  $(g, m) \in I$ , 则由引理 3 即得  $\gamma g \leq \mu m$ , 进而由  $\wedge \gamma(A) \leq \gamma g$  可导出  $\wedge \gamma(A) \leq \mu m$ , 与条件  $\wedge \gamma(A) \not\leq \mu m$  相矛盾! 故结论(1)成立.

当  $(g, m) \in I$  时, 由引理 3 可知  $\gamma g \leq \mu m$ , 进而由  $\wedge \gamma(A) \leq \gamma g$  即得结论(2)成立. 证毕.

**定理 3** 在定理 2 中, 有下述结论成立:

- (1) 若  $\wedge \gamma(A) \not\leq \mu m$ , 则  $\eta_\lambda(g, m) = 0$ .
- (2) 若  $\eta_\lambda(g, m) > 0$ , 则  $\wedge \gamma(A) \leq \mu m$ .
- (3) 设  $A, H \subseteq G$ , 若  $H \subseteq A$ , 则  $\wedge \gamma(A) \leq \wedge \gamma(H)$ .

证明:由定义 5 和定理 2 即得结论(1)和(2)成立. 结论(3)由上、下确界的定义即得, 在此不再详述. 证毕.

在基于准则 2\_a、2\_b 或 2\_c 的判定过程中, 用户可以借助概念格  $(B(K), \leq)$  的 Hasse 图, 即概念格的结构信息, 直观生动地判定任意关系粒  $[g] \times m$  是否满足  $([g], m) \in I_\lambda$ . 相对于准则 2\_a 和准则 2\_b, 用户选择准则 2\_c 得到的判定结果可能会更合理一些, 因为准则 2\_a 的判定条件太强, 不利于用户从看似不相关的数据中发现有价值的信息, 而准则 2\_b 的判定条件太弱, 可能无法为构造粒化背景提供更加科学合理的依据.

以准则 2\_c 为例, 当  $\lambda = 0.8$  且  $\delta = 0.6$  时, 基于图 2 所示的格结构信息, 用户可以从表 1 导出如表 5 所示的粒化背景. 例如, 在图 3 中, 对于关系粒  $[1] \times d$ , 由于  $\gamma 1 \wedge \gamma 8 = \gamma 1 = \gamma 8$ ,  $\gamma 1 \wedge \gamma 4 = \gamma 1$ ,  $\gamma 4 \wedge \gamma 8 = \gamma 8$ , 且  $\gamma 1 \wedge \gamma 4$ ,  $\gamma 1 \wedge \gamma 8$ ,  $\gamma 4 \wedge \gamma 8$  均小于  $\mu d$ , 由此可以判定  $([1], d) \in I_\lambda$ ; 对于关系粒  $[3] \times e$ , 由于  $\gamma 3 \wedge \gamma 9 = \gamma 3 = \gamma 9$  且  $\gamma 3 \wedge \gamma 9 \not\leq \mu e$ , 所以判定  $([3], e) \notin I_\lambda$ .

表 5 基于准则 2\_c 从表 1 导出的粒化背景

Tab. 5 Granulation context deduced from Tab. 1 based on Criterion 2\_c

	a	b	c	d	e
[1]	×		×	×	×
[2]		×		×	×
[3]		×	×		
[5]			×	×	

### 3 粒化概念和粒化概念格

在粒化背景  $K_\lambda = (G_\lambda, M, I_\lambda)$  中, 对于任意对象

粒集  $A_\lambda \subseteq G_\lambda$  和属性集  $B \subseteq M$ , 定义 1 中的算子可以形式化描述为

$$A'_\lambda = \{m \in M \mid ([g], m) \in I_\lambda, \forall [g] \in A_\lambda\}$$

$$B'_\lambda = \{g \in G \mid ([g], m) \in I_\lambda, \forall m \in B\}$$

若  $A'_\lambda = B$  且  $B'_\lambda = A_\lambda$ , 则称  $(A_\lambda, B)$  是一个  $\lambda$  粒化概念. 设  $(E_\lambda, B)$  和  $(F_\lambda, D)$  是  $\lambda$  粒化概念, 定义

$$(E_\lambda, B) \leq (F_\lambda, D) \Leftrightarrow E_\lambda \subseteq F_\lambda \Leftrightarrow B \supseteq D$$

称完备格  $(B(K_\lambda), \leq)$  是  $(B(K), \leq)$  的一个  $\lambda$  粒化概念格, 其中  $B(K_\lambda)$  是  $K_\lambda$  中所有  $\lambda$  粒化概念组成的集合. 例如, 由表 4 可以导出一个如图 3 所示的概念格结构, 该结构是图 1 所示概念格的一个粒化概念格.

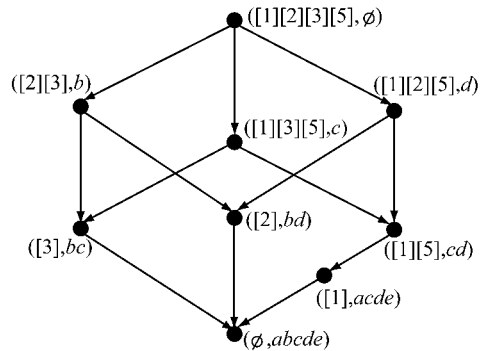


图 3 由表 4 生成的粒化概念格

Fig. 3 Granulation concept lattice with respect to Tab. 4

由上述讨论可知, 粒化形式背景中的算子是基于对象粒定义的, 虽然其与定义 1 中的算子形式上存在一定差异, 但本质上它们属于同一类算子.

### 4 约简、核和蕴涵规则

约简、核、蕴涵规则等是概念格中常见的问题, 本节将重点就上述问题给出一种基于粒化概念的求解方法.

**定义 6** 在  $K = (G, M, I)$  中, 设  $B \subseteq M$ , 如果  $m \in B$  且  $B' \neq (B - m)'$ , 则称  $m$  是  $B$  中的核属性, 设  $D \subseteq B \subseteq M$ , 若  $D$  是满足  $B' = D'$  的最小属性子集, 则称  $D$  是  $B$  的一个约简; 设  $B, D \subseteq M$ , 称  $B \rightarrow D$  是  $K$  中的蕴涵规则, 简称规则. 当  $B' \subseteq D'$  时, 称  $B \rightarrow D$  是一条确定规则; 当  $B' \not\subseteq D'$  时, 称  $B \rightarrow D$  是一条不确定规则.

**定理 4** 在  $K_\lambda = (G_\lambda, M, I_\lambda)$  中, 对于任意  $B \subseteq M$ , 记

$$B^+ = \bigcap \{L \in \text{Int}(K_\lambda) \mid B \subseteq L\}$$

若  $\text{Int}(K_\lambda)$  是一个由所有  $\lambda$  粒化概念内涵组成的集合, 则  $B^+ = B''$ .

证明:由引理 2 即得. 证毕.

**定理 5** 在  $K_\lambda = (G_\lambda, M, I_\lambda)$  中, 设  $m \in B \subseteq M$ , 若  $E_\lambda \neq F_\lambda$ , 则  $m$  是  $B$  中的核属性. 其中,  $(E_\lambda, B^+) \in B(K_\lambda)$  且  $(F_\lambda, (B-m)^+) \in B(K_\lambda)$ .

证明:由  $(E_\lambda, B^+) \in B(K_\lambda)$  可知  $E_\lambda = (B^+)'$ , 进而由定理 4 可知  $E_\lambda = B''' = B'$ . 同理可证  $F_\lambda = (B-m)'$ . 由于  $E_\lambda \neq F_\lambda$ , 所以  $B' \neq (B-m)'$ . 显然, 由定义 6 即得结论成立. 证毕.

**定理 6** 在  $K_\lambda = (G_\lambda, M, I_\lambda)$  中, 设  $D \subseteq B \subseteq M$ , 若  $D$  是满足  $E_\lambda = F_\lambda$  的最小集合, 则  $D$  是  $B$  的一个约简, 其中  $(E_\lambda, B^+) \in B(K_\lambda)$  且  $(F_\lambda, D^+) \in B(K_\lambda)$ .

证明:由  $(E_\lambda, B^+) \in B(K_\lambda)$  可知  $E_\lambda = (B^+)'$ , 进而由定理 4 可知  $E_\lambda = B''' = B'$ . 同理可证  $F_\lambda = D'$ . 由于  $E_\lambda = F_\lambda$ , 所以  $B' = D'$ , 又由于  $D$  是满足  $B' = D'$  的最小集合, 故由定义 6 即得结论成立. 证毕.

概念格非常适合发现一些有价值的规则型知识, 置信度和支持度是对规则的重要补充和定量描述. 通常, 在形式背景  $K = (G, M, I)$  中, 就某条规则  $B \rightarrow D$  而言, 其置信度是指在  $B$  出现的条件下,  $D$  出现的概率, 通常被形式化描述为

$$\text{Conf}(B \rightarrow D) = \frac{|B' \cap D'|}{|B'|}$$

如果  $\text{Conf}(B \rightarrow D) = 1$ , 则说明当  $B$  出现时,  $D$  一定会出现. 如果置信度太低, 则说明  $B$  和  $D$  之间的关系较弱,  $B \rightarrow D$  在实际应用中的价值不高. 支持度本质上揭示了  $B$  与  $D$  同时出现的概率, 通常被形式化描述为

$$\text{Sup}(B \rightarrow D) = \frac{|B' \cap D'|}{|G|}$$

如果支持度较高, 则说明  $B$  与  $D$  在一定程度上是相关的. 事实上, 在实际应用中, 用户往往会对象置信度和支持度提出一定的要求, 进而获取置信度和支持度较高的规则.

基于上述讨论, 关于规则的置信度和支持度, 本文给出了基于对象粒  $G_\lambda$  的定义.

**定义 7** 在  $K_\lambda = (G, M, I_\lambda)$  中, 对于任意规则  $B \rightarrow D$ , 其置信度和支持度分别定义为

$$\text{Conf}(B \rightarrow D) = \frac{|E_\lambda \cap F_\lambda|}{|E_\lambda|}$$

$$\text{Sup}(B \rightarrow D) = \frac{|E_\lambda \cap F_\lambda|}{|G_\lambda|}$$

其中,  $(E_\lambda, B^+) \in B(K_\lambda)$  且  $(F_\lambda, D^+) \in B(K_\lambda)$ .

例如, 在表 4 中, 对于规则  $c \rightarrow de$ , 由于  $([1][3][5], c)$  和  $([1], de)$  是  $\lambda$  粒化概念, 又  $\{c\}^+ = \{c\}$  且

$\{d, e\}^+ = \{a, c, d, e\}$ , 故有

$$\text{Conf}(c \rightarrow de) = \frac{|\{[1], [3], [5]\} \cap \{[1]\}|}{|\{[1], [3], [5]\}|} =$$

$$\frac{1}{3} = 0.33$$

$$\text{Sup}(c \rightarrow de) = \frac{|\{[1], [3], [5]\} \cap \{[1]\}|}{|G_\lambda|} =$$

$$\frac{1}{4} = 0.25$$

通常, 形式背景中存在着大量的冗余规则, 往往导致一些有价值的规则被淹没在海量的规则中. 一些常见的确定的冗余规则有: 若  $B \subseteq B_1$  且  $D_1 \subseteq D$ , 则  $B_1 \rightarrow D_1$  相对于确定规则  $B \rightarrow D$  是冗余的, 此时称  $B_1 \rightarrow D_1$  是一条相对冗余规则; 若  $D \subseteq B$ , 称  $B \rightarrow D$  是一条绝对冗余规则.

基于上述讨论, 在一个粒化形式背景中, 可以删除一些绝对冗余规则和相对冗余规则, 并最终得到一个规模较小的确定规则集, 称为源规则集. 例如, 表 1 和 4 中源规则集如表 6 所示.

表 6 源规则集

Tab. 6 Source sets in Tab. 1 and Tab. 4	
$\lambda$	源规则集
1.0	$ad \rightarrow acde, cde \rightarrow acde, a \rightarrow ace$
0.8	$a \rightarrow acde, e \rightarrow acde$

## 5 结论及展望

把粒计算融入概念格研究中, 通过降低知识获取的“分辨率”, 扩大知识度量的“尺度”, 有助于人们从不相关的数据中发现一些潜在的有价值信息, 而且也能有效简化复杂的格结构和压缩概念的规模. 在粒化过程中, 提出了一种新的基于概念基的求解思路. 所谓概念基是指一个由低层概念构成的集合, 其在信息粒化的整个过程中扮演了重要的角色. 特别地, 在信息粒化过程中, 还提出了一种基于下确界运算的求解思路, 它与传统模型最大的区别是融合了概念格的结构信息.

总之, 本文把粒计算引入到概念格研究中, 对于概念格的拓展不失为一种有效的手段, 理论和实例证明了本文提出的结论的合理性和有效性. 虽然概念格是数据分析的有效工具, 但面对规模庞大的数据集以及复杂的知识发现任务, 在实际应用中仍存在着许许多多的问题. 本文的研究工作仅仅是一个尝试, 关于粒计算与概念格的融合理论还有待进一步深入.

## 参考文献:

- [1] Wille R. Restructuring lattice theory: an approach based on hierarchies of concepts [C]//Ordered Sets. Dordrecht: Reidel, 1982: 445-470.
- [2] Ganter B, Wille R. Formal concept analysis: mathematical foundations [M]. Berlin: Springer-Verlag, 1999.
- [3] Li J H, Mei C L, Lü Y J. Incomplete decision contexts: approximate concept construction, rule acquisition and knowledge reduction [J]. International Journal of Approximate Reasoning, 2013, 54(1): 149.
- [4] 薛金蓉, 安秋生, 郑军. 概念格的内涵缩减与数据库推理依赖[J]. 计算机研究与发展, 2014, 51(1): 96.  
XUE Jinrong, AN Qiusheng, ZHENG Jun. Intent reduction of concept lattice and database inference dependence [J]. Journal of Computer Research and Development, 2014, 51(1): 96.
- [5] 杨丽, 徐扬. 基于逻辑合取的语言真值概念格合并算法[J]. 电子学报, 2013, 41(11): 2149.  
YANG Li, XU Yang. Union algorithm of linguistic truth-valued concept lattice based on logical conjunction [J]. ACTA Electronica Sinica, 2013, 41(11): 2149.
- [6] 智慧来, 智东杰, 刘宗田. 概念格合并原理与算法[J]. 电子学报, 2010, 38(2): 455.  
ZHI Huilai, ZHI Dongjie, LIU Zongtian. Theory and algorithm of concept lattice union [J]. ACTA Electronica Sinica, 2010, 38(2): 455.
- [7] 张磊, 张宏莉, 殷丽华, 等. 概念格的属性渐减原理与算法研究[J]. 计算机研究与发展, 2013, 50(2): 248.  
ZHANG Lei, ZHANG Hongli, YIN Lihua, *et al.* Theory and algorithms of attribute decrement for concept lattice [J]. Journal of Computer Research and Development, 2013, 50(2): 248.
- [8] 谢志鹏, 刘宗田. 概念格的快速渐进式构造算法[J]. 计算机学报, 2002, 25(5): 490.  
XIE Zhipeng, LIU Zongtian. A fast incremental algorithm for building concept lattice [J]. Chinese Journal of Computers, 2002, 25(5): 490.
- [9] 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.  
LIU Zongtian, QIANG Yu, ZHOU Wen, *et al.* A fuzzy concept lattice model and its incremental construction algorithm [J]. Chinese Journal of Computers, 2007, 30(2): 184.
- [10] Zhai Y H, Li D Y, Qu K S. Fuzzy decision implications [J]. Knowledge-Based Systems, 2013, 37: 230.
- [11] Zhai Y H, Li D Y, Qu K S. Decision implication canonical basis: a logical perspective [J]. Journal of Computer and System Sciences, 2015, 81: 208.
- [12] 胡可云, 陆玉昌, 石纯一. 基于概念格的分类和关联规则的集成挖掘方法[J]. 软件学报, 2000, 11(11): 1478.  
HU Keyun, LU Yuchang, SHI Chunyi. An integrated mining approach for classification and association rule based on concept lattice [J]. Journal of Software, 2000, 11(11): 1478.
- [13] 梁吉业, 王俊红. 基于概念格的规则产生集挖掘算法[J]. 计算机研究与发展, 2004, 41(8): 1339.  
LIANG Jiye, WANG Junhong. An algorithm for extracting rule-generating sets based on concept lattice [J]. Journal of Computer Research and Development, 2004, 41(8): 1339.
- [14] 马垣, 张学东, 迟呈英. 紧致依赖与内涵亏值[J]. 软件学报, 2011, 22(5): 962.  
MA Yuan, ZHANG Xuedong, CHI Chengying. Compact dependencies and intent waned values [J]. Journal of Software, 2011, 22(5): 962.
- [15] Mi J S, Leung Y, Wu W Z. Approaches to attribute reduction in concept lattices induced by axialities [J]. Knowledge-Based Systems, 2010, 23(6): 504.
- [16] 张文修, 魏玲, 祁建军. 概念格的属性约简理论与方法[J]. 中国科学: E辑, 2005, 35(6): 628.  
ZHANG Wenxiu, WEI Ling, QI Jianjun. Attribute reduction theory and approach to concept lattice [J]. Science in China: Series E, 2005, 35(6): 628.
- [17] 李冲, 曹吉鸣, 马腾. 基于形式概念分析的项目成员综合相似度计算[J]. 同济大学学报: 自然科学版, 2014, 42(6): 983.  
LI Chong, CAO Jiming, MA Teng. Integrated similarity calculation of project members based on formal concept analysis [J]. Journal of Tongji University: Natural Science, 2014, 42(6): 983.
- [18] 李立峰, 张东晓. 概念格在二值命题逻辑命题集约简中的应用[J]. 电子学报, 2007, 35(8): 1538.  
LI Lifeng, ZHANG Dongxiao. The application of concept lattice theory in the reduction of the proposition set in two-valued propositional logic [J]. ACTA Electronica Sinica, 2007, 35(8): 1538.
- [19] Shao M W, Liu M, Zhang W X. Set approximations in fuzzy formal concept analysis [J]. Fuzzy Sets and Systems, 2007, 158(23): 2627.
- [20] Kang X P, Li D Y, Wang S G, *et al.* Formal concept analysis based on fuzzy granularity base for different granulations [J]. Fuzzy Sets and Systems, 2012, 203: 33.
- [21] Burusco A, Fuentes-Gonzales R. Construction of the L-fuzzy concept lattice [J]. Fuzzy Sets and Systems, 1998, 97: 109.
- [22] Wei L, Qi J J. Relation between concept lattice reduct and rough set reduct [J]. Knowledge-Based Systems, 2010, 23(8): 934.
- [23] Kang X P, Li D Y, Wang S G, *et al.* Rough set model based on formal concept analysis [J]. Information Sciences, 2013, 222: 611.
- [24] Chen J K, Li J J, Lin Y J, *et al.* Relations of reduction between covering generalized rough sets and concept lattices [J]. Information Sciences, 2015, 304: 16.
- [25] 仇国芳, 张志霞, 张炜. 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.  
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.
- [26] Lai H L, Zhang D X. Concept lattices of fuzzy contexts: formal concept analysis vs. rough set theory [J]. International Journal of Approximate Reasoning, 2009, 50: 695.
- [27] Wang L D, Liu X D. Concept analysis via rough set and AFS algebra [J]. Information Sciences, 2008, 178: 4125.
- [28] Yao Y Y. Concept lattices in rough set theory [C] // Proceedings of the 2004 IEEE Annual Meeting of the North

- American Fuzzy Information Processing Society (NAFIPS04). Piscataway: IEEE, 2004; 796.
- [29] Dias S M, Nogueira B M, Zarate L E. Adaptation of FCANN method to extract and represent comprehensible knowledge from neural networks [J]. *Studies in Computational Intelligence*, 2008, 134:163.
- [30] Jiang L, Deogun J. SPICE: a new frame work for data mining based on probability logic and formal concept analysis [J]. *Fundamenta Informaticae*, 2007, 78: 467.
- [31] 仇国芳, 马建敏, 杨宏志, 等. 概念粒计算系统的数学模型 [J]. *中国科学: E辑*, 2009, 39 (12): 1239.  
QIU Guofang, MA Jianmin, YANG Hongzhi, *et al.* Mathematical model of concept granule computing system [J]. *Science in China: Series E*, 2009, 39(12): 1239.
- [32] Kang X P, Li D Y, Wang S G. Research on domain ontology in different granulations based on concept lattice [J]. *Knowledge-Based Systems*, 2012, 27: 152.
- [33] Wu W Z, Leung Y, Mi J S. Granular computing and knowledge reduction in formal contexts [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2009, 21(10): 1461.
- [34] Sarmah A K, Hazarika S M, Sinha S K. Formal concept analysis: current trends and directions [J]. *Artificial Intelligence Review*, 2015, 44(1): 47.
- [35] Lee H S. An optimal algorithm for computing the max-min transitive closure of a fuzzy similarity matrix [J]. *Fuzzy Sets and Systems*, 2001, 123: 129.

~~~~~

(上接第 1095 页)

- [3] White R J. Exploration of a strategy for reducing gear noise in planetary transmission and evaluation of laser vibrometry as a means for measuring transmission error[D]. Cleveland: Case Western Reserve University, 2006.
- [4] Koffel G, Ville F, Changenet C, *et al.* Investigations on the power losses and thermal effects in gear transmission [J]. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, 2009, 223(3):469.
- [5] Inalpolat M, Kahraman A. A theoretical and experimental investigation of modulation sidebands of planetary gear sets [J]. *Journal of the Sound and Vibration*, 2009, 323 (3/5): 677.
- [6] 肖正明, 秦大同, 武文辉, 等. 土压平衡盾构机行星减速器动态性能测试与分析 [J]. *振动与冲击*, 2012, 31(9):154.  
XIAO Zhengming, QIN Datong, WU Wenhui, *et al.* Testing and analysis on dynamic performance of planetary gearbox in EPB shield tunneling machine [J]. *Journal of Vibration and Shock*, 2012, 31(9): 154.
- [7] Hohn B R, Michaelis K, Otto H P. Influence of immersion depth of dip lubricated gears on power loss, bulk temperature and scuffing load carrying capacity [J]. *International Journal of Mechanics and Materials in Design*, 2008, 4(2): 145.
- [8] 马璇, 李建华, 陈国定. 减速器齿轮传动系统的稳态热分析及试验研究 [J]. *西北工业大学学报*, 2002, 20(1):32.  
MA Xuan, LI Jianhua, CHEN Guoding. Experimentally supported stable thermal analysis of gearbox transmission system [J]. *Journal of Northwestern Polytechnical University*, 2002, 20(1): 32.
- [9] Barthod M, Hayne B, Tébec J L, *et al.* Experimental study of dynamic and noise produced by a gearing excited by a multi-harmonic excitation [J]. *Applied Acoustics*, 2007, 68(9): 982.
- [10] Marques P M T, Camacho R, Martins R C, *et al.* Efficiency of a planetary multiplier gearbox: influence of operating conditions and gear oil formulation [J]. *Tribology International*, 2015, 92: 272.
- [11] Song C, Zhu C, Liu H, *et al.* Dynamic analysis and experimental study of a marine gearbox with crossed beveloid gears [J]. *Mechanism and Machine Theory*, 2015, 92: 17.
- [12] Helsen J, Marrant B, Vanhollebeke F, *et al.* Assessment of excitation mechanisms and structural flexibility influence in excitation propagation in multi-megawatt wind turbine gearboxes: experiments and flexible multibody model optimization [J]. *Mechanical Systems and Signal Processing*, 2013, 40(1): 114.