

# 属性聚类下三支概念的对比

张晓燕\*, 王佳一

(西南大学 人工智能学院, 重庆 400715)

(\* 通信作者电子邮箱 zxy19790915@163.com)

**摘要:** 三支概念分析是人工智能领域一个非常重要的研究方向, 该理论最大的优势是可以同时研究形式背景中对象“共同具有”和“共同不具有”的属性。众所周知, 经过属性聚类生成的新形式背景与原形式背景具有较强的联系, 同时原三支概念与经过属性聚类得到的新三支概念也存在紧密的内在联系。为此, 进行属性聚类下三支概念的对比研究和分析。首先基于属性聚类提出悲观属性聚类、乐观属性聚类以及一般属性聚类的概念, 并研究了这三种属性聚类的关系; 然后, 通过对比聚类过程与三支概念形成的过程, 研究了原三支概念与新三支概念的区别, 分别从面向对象和面向属性的角度提出两个最低约束指数, 探索了属性聚类对三支概念格的影响, 进一步丰富了三支概念分析理论, 为可视化数据处理领域提供了可行的思路。

**关键词:** 三支概念; 属性聚类; 多粒度; 最低约束指数; 等价类

**中图分类号:** TP18 **文献标志码:** A

## Comparison of three-way concepts under attribute clustering

ZHANG Xiaoyan\*, WANG Jiayi

(College of Artificial Intelligence, Southwest University, Chongqing 400715, China)

**Abstract:** Three-way concept analysis is a very important topic in the field of artificial intelligence. The biggest advantage of this theory is that it can study “attributes that are commonly possessed” and “attributes that are commonly not possessed” of the objects in the formal context at the same time. It is well known that the new formal context generated by attribute clustering has a strong connection with the original formal context, and there is a close internal connection between the original three-way concepts and the new three-way concepts obtained by attribute clustering. Therefore, the comparative study and analysis of three-way concepts under attribute clustering were carried out. Firstly, the concepts of pessimistic, optimistic and general attribute clusterings were proposed on the basis of attribute clustering, and the relationship among these three concepts was studied. Moreover, the difference between the original three-way concepts and the new ones was studied by comparing the clustering process with the formal process of three-way concepts. Furthermore, two minimum constraint indexes were put forward from the perspective of object-oriented and attribute-oriented respectively, and the influence of attribute clustering on three-way concept lattice was explored. The above results further enrich the theory of three-way concept analysis and provide feasible ideas for the field of visual data processing.

**Key words:** three-way concept; attribute clustering; multi-granularity; minimum constraint index; equivalence class

## 0 引言

形式概念分析<sup>[1]</sup>作为一种数据分析和知识处理的数学工具, 它的主要思想是以形式背景为研究对象, 以形式概念与概念格为核心概念, 通过概念之间的偏序关系刻画对象和属性的内在关系<sup>[2]</sup>。因此, 该理论是一种非常有效的可视化数据处理方法。然而, 传统的形式概念分析理论只研究了数据“共同(被)具有”关系的问题, 而忽略了“共同不(被)具有”关系的问题<sup>[3,4]</sup>, 导致现实应用中该理论具有一定的局限性<sup>[5]</sup>。

2014年, Qi等<sup>[6]</sup>提出了新的形式概念分析理论, 即三支概念分析。三支概念分析是将三支决策理论与概念格相融合, 提出了对象诱导的三支概念格与属性诱导的三支概念格。这两种三支概念格均同时研究了“共同(被)具有”和“共

同不(被)具有”的关系。它们获得的概念更加精确, 构建的三支概念格相较于以往经典概念格, 在实际应用过程中概念识别也会更加精确。

在三支概念正式提出后, 许多学者对此进行了扩展和深入研究。Qian等<sup>[7]</sup>利用形式背景的叠置与并置, 提出了三支概念格的构造方法, 并仿照对象诱导概念格与属性诱导概念格给出了对象诱导的三支面向对象概念格和属性诱导的三支面向属性概念格的定义, 分析了四种概念之间的异同。苏新等<sup>[8]</sup>比较研究了基于对象和基于属性的三支概念格合并方法。Wei等<sup>[9]</sup>立足于三支概念格, 在三支协调的意义下研究了决策背景的规则获取问题, 并与强协调决策背景所获得的一般决策规则进行了详细的比较研究。

除了对三支概念本身的研究外, 学科交叉融合也极大扩

收稿日期: 2022-03-30; 修回日期: 2022-06-20; 录用日期: 2022-06-22。 基金项目: 国家自然科学基金资助项目(61976245)。

作者简介: 张晓燕(1979—), 女, 山西怀仁人, 副教授, 博士, CCF会员, 主要研究方向: 不确定性推理、人工智能的数学基础、粒计算与知识发现; 王佳一(1999—), 女, 河北正定人, 硕士研究生, 主要研究方向: 形式概念分析、认知计算。

展了三支概念的前景。Li等<sup>[10]</sup>将多粒度与三支概念结合,研究了基于多粒度的三支认知概念学习模型。龙柄翰等<sup>[11]</sup>建立了模糊三支概念分析,将模糊集理论与三支概念分析相融合,重点考虑模糊背景中“共同具有的程度”与“共同不具有的程度”两方面不确定的共性信息,为三支概念在模糊领域的实际应用奠定了较好的基础。

在多粒度研究方面,对象粒化、属性粒化以及关系粒化等概念<sup>[12-13]</sup>的提出将多粒度与概念认知联系起来。这些研究缓解了庞大的概念个数,在约束宽松的情况下减小了时间空间复杂度,并且为获取数据的多层次概念知识表示与处理方法提供了新的方法<sup>[14]</sup>。多粒度方面的研究主要包括对象(属性)的粗化与细化研究。属性粒化也称为属性的吸收和分解,在粒化过程中概念也会随之转化,从而可以获得在不同粒度空间的形式概念。另外,Belohlavek等<sup>[15]</sup>通过粒度树与剪枝研究了给予属性粒化的形式概念分析方法,从而生成不同粒度层次的形式背景。然而,剪枝的构造过程中要求不同粒度层次的属性在某种意义上存在偏序关系。这一条件较为苛刻,甚至难以实现。基于此种局限,Liu等<sup>[16]</sup>推广了该方法,提出了属性聚类的概念,即设定一个等价类条件,再以此为基础进行属性吸收,从而构建多粒度层次。

然而,在多粒度层次方面,对三支概念的研究尚且不足。同时,目前的研究无法通过粒度转化直接得到多粒度概念格和多粒度三支概念格,这对于多个粒度层次下的分析十分不利,且需要大量冗余的计算<sup>[17]</sup>。针对该情况,本文提出以属性聚类为基础的三支概念构建,并重点研究不同概念格下的三支概念转化,进一步丰富三支形式概念分析理论。

## 1 基础知识

### 1.1 概念格

设有形式背景  $K = (G, M, I)$ , 其中:  $G$  为对象集,  $M$  为属性集,  $I$  为  $G$  和  $M$  之间的二元关系。在经典形式背景中,  $I$  的取值只有 0 或 1 两种可能。对于  $x \in G, m \in M$ , 当  $I(x, m) = 1$  时, 表示对象  $x$  和属性  $m$  存在关系  $I$ ; 当  $I(x, m) = 0$  时, 表示对象  $x$  和属性  $m$  不存在关系  $I$ 。

为研究对象子集和属性子集之间的关系, 现给出两个导出算子, 分别作用于属性子集和对象子集, 对于  $X \subseteq G, A \subseteq M$ , 有:

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

$$A^* = \{x: x \in G, \forall m \in A, I(x, m) = 1\}$$

特别地, 当对象子集或属性子集中仅有一个元素时, 记  $\{x\}^*$  为  $x^*$ , 记  $\{m\}^*$  为  $m^*$ 。

对于形式背景  $K = (G, M, I)$ , 若对于  $X \subseteq G, A \subseteq M$ , 有  $X^* = A$  且  $A^* = X$ , 则  $(X, A)$  称为一个形式概念, 其中概念的外延为  $X$ , 概念的内涵为  $A$ 。形式背景  $K = (G, M, I)$  下的所有形式概念的集合为  $L(K)$ ,  $L(K)$  即为概念格<sup>[18]</sup>。

概念的上下近似算子定义如下:

$$X^d = \{m \in M: m^* \subseteq X\}$$

$$A^d = \{x \in G: x^* \subseteq A\}$$

$$X^u = \{m \in M: m^* \cap X \neq \emptyset\}$$

$$A^u = \{x \in G: x^* \cap A \neq \emptyset\}$$

由上可知, 上下近似算子将本身较为严格的概念格约束放宽, 使概念格更具实际意义。根据上下近似算子, 可以定义新的算子形成面向对象和面向属性的概念格。

对于形式背景  $K = (G, M, I)$ , 若对于  $X \subseteq G, A \subseteq M$ , 有  $X^u = A$  且  $A^d = X$ , 则  $(X, A)$  称为面向属性的概念。在形式背景下, 所有面向属性的概念称为面向属性概念格, 即  $L_p(K)$ 。进一步, 若有  $X^d = A$  且  $A^u = X$ , 则  $(X, A)$  可被称为面向对象的概念。所有面向对象的概念称为面向对象概念格, 即  $L_o(K)$ 。

### 1.2 三支概念格

如果说概念格研究的是共同具有的关系, 那么三支概念格就是同时研究共同具有和共同不具有两个关系。在应用中, 单方面的研究往往具有片面性, 从正反两方面研究能使研究结果更精准, 提高概念分析的正确率。

所谓“反方向”, 其实就是原形式背景的补背景。这里记  $K^c = (G, M, I^c)$  为原形式背景  $K = (G, M, I)$  的补背景。原形式背景的补背景主要有两种情况, 即 I 型混合背景和 II 型混合背景。I 型混合背景即形式背景与补背景的并置, 记为  $K|K^c = \{G, M \cup \bar{M}, I \cup I^c\}$ , 其中  $\bar{M}$  记为  $\bar{M} = \{\bar{a}_i: \bar{a}_i \in M\}$ , 对应的概念格为  $L(K|K^c)$ ; II 型混合背景即形式背景与补背景的叠置, 记为  $K/K^c = \{G \cup \bar{G}, M, I \cup I^c\}$ , 对应的概念格为  $L(K/K^c)$ 。

为了研究共同不具有的关系, 下面给出负算子<sup>[19]</sup>的定义。对于子集  $X \subseteq G, A \subseteq M$ , 有:

$$X^{\bar{}} = \{m: m \in M, \forall x \in X, I(x, m) = 0\}$$

$$A^{\bar{}} = \{x: x \in G, \forall m \in A, I(x, m) = 0\}$$

显然, 通过负算子可以表示并研究形式背景的“共同不具有”这一关系。但如果需要同时研究具有和不具有的关系, 还需要定义一对算子。需要特别说明的是, 单个对象子集在运算后会得到两个属性子集, 即共同具有的属性 and 共同不具有的属性。同理, 单个属性子集在运算后也会得到两个对象子集。因此, 由于出发点的不同, 可以得到下面两种三支概念。

对于  $X \subseteq G$  和  $A, B \subseteq M$ :  $X^0 = (X^*, X^{\bar{}})$ ,  $(A, B)^0 = A^* \cap B^{\bar{}}$ 。当  $X^0 = (A, B)$  且  $(A, B)^0 = X$ , 则称  $(X, (A, B))$  为对象诱导的三支概念, 简称 OE 概念, 其中  $X$  为 OE 概念的外延,  $(A, B)$  为 OE 概念的内涵。

对于  $X, Y \subseteq G$  和  $A \subseteq M$ :  $A^A = (A^*, A^{\bar{}})$ ,  $(X, Y)^A = X^* \cap Y^{\bar{}}$ 。当  $A^A = (X, Y)$  且  $(X, Y)^A = A$  时, 则称  $((X, Y), A)$  为属性诱导的三支概念, 简称 AE 概念, 其中  $(X, Y)$  为 AE 概念的外延,  $A$  为 AE 概念的内涵。

类似地, 对于上下近似算子, 也有相似的负算子定义:

$$X^{\bar{d}} = \{m \in M: m^{\bar{*}} \subseteq X\}$$

$$A^{\bar{d}} = \{x \in G: x^{\bar{*}} \subseteq A\}$$

$$X^{\bar{u}} = \{m \in M: m^{\bar{*}} \cap X \neq \emptyset\}$$

$$A^{\bar{u}} = \{x \in G: x^{\bar{*}} \cap A \neq \emptyset\}$$

于是, 根据负算子可以有面向对象三支概念和面向属性的三支概念。对于  $X \subseteq G$  和  $A, B \subseteq M$ , 定义:  $X^{0'} = (X^d, X^{\bar{d}})$ ,  $(A, B)^{0'} = A^d \cap B^{\bar{d}}$ 。当  $X^{0'} = (A, B)$  且  $(A, B)^{0'} = X$ , 则称  $(X, (A, B))$  为面向对象的三支概念, 简称 OEO 概念, 其中  $X$  为 OEO 概念的外延,  $(A, B)$  为 OEO 概念的内涵。在形式背景  $K = (G, M, I)$  中, 所有对象诱导的面向对象三支概念的集合被称为对象诱导的面向对象三支概念格, 记为  $OEOL(K)$ 。

对于  $X, Y \subseteq G$  和  $A \subseteq M$ , 定义:  $A^{A'} = (A^d, A^{\bar{d}})$ ,  $(X, Y)^{A'} = X^d \cap Y^{\bar{d}}$ 。当  $A^{A'} = (X, Y)$  且  $(X, Y)^{A'} = A$  时, 称  $((X, Y), A)$  为面

向属性的三支概念,简称 AEP 概念,其中  $(X, Y)$  为 AEP 概念的外延,  $A$  为 AEP 概念的内涵。在形式背景  $K = (G, M, I)$  中,所有属性诱导的面向属性三支概念的集合被称为属性诱导的面向属性三支概念格,记为  $AEPL(K)$ 。

无论是 OE 概念、AE 概念、OEO 概念还是 AEP 概念,均在定义中利用了负算子,即同时研究了共同具有和共同不具有的问题,它们的区别在于概念形成的过程和约束。

对于形式背景  $K = (G, M, I)$ , 给定任意  $x \in G$  和  $m \in M$ , 则有下列结论成立:

- 1)  $(m^*, (m^{**}, m^{***}), (m^{\bar{}} , (m^{\bar{**}} , m^{\bar{***}}))) \in OEL(K)$ ;
- 2)  $((x^{**}, x^{***}), x^*) , ((x^{\bar{**}} , x^{\bar{***}}), x^{\bar{}})) \in AEL(K)$ ;
- 3)  $(m^u , (m^{ud} , m^{u\bar{d}})) , (m^{\bar{u}} , (m^{\bar{u}d} , m^{\bar{u}\bar{d}})) \in OELO(K)$ ;
- 4)  $((x^{ud} , x^{u\bar{d}}), x^u) , ((x^{\bar{u}d} , x^{\bar{u}\bar{d}}), x^{\bar{u}}) \in AEPL(K)$ 。

### 1.3 属性聚类

在实际应用中,形式背景往往包括大量的属性和对象,但在某个具体的研究中,往往不需要研究全部的属性,因此属性约简或属性吸收就显得尤为重要。属性聚类作为一种较为通用的属性吸收方法,是基于粒度树和剪枝的属性粒化的推广。属性聚类的获取过程更简便,不要求聚类之前与之后的属性存在实际意义上的偏序关系,因此也具有更好的适用性。

属性聚类的特征是基于某种特定需求或者先验关系的。通过等价关系的作用,部分属性被聚合起来,形成新的属性(集)。需要特别说明的是,虽然新的属性集中元素均是以聚类之前的属性为元素的集合,但在新的形式背景下,每个集合被看作一个整体,被称为新的属性。

对于形式背景  $K = (G, M, I)$ ,  $R$  为先验关系或特定关系确立的等价关系,  $[m]_R$  为包含属性  $m$  的属性等价类,属性聚类后形成的新形式背景  $(G, M_R, I_R)$ <sup>[12]</sup> 如下:

$$M_R = \{[m]_R; m \in M\}$$

$$I_R = \{I_R(g, [m]_R) = 1; \exists n \in [m]_R, I(g, n) = 1\}$$

相较于旧的形式背景,新的形式背景的对象集相同,但属性集发生了变化,新的属性和旧的属性存在包含与被包含关系,属性聚类是一个由细粒度到粗粒度的过程。因此,在属性聚类下,可以在不同层次、不同粒度下进行研究。

## 2 属性聚类前后三支概念的变化

### 2.1 属性聚类的推广

在第 1 章属性聚类过程中定义对应关系时,采用了乐观定义方法,即在构成新属性的属性集中,只要有一个属性可以与对象  $x$  构成关系  $I$ , 那么新的属性就与对象  $x$  构成关系  $I_R$ 。这在应用中会存在一定的局限性。如在日常生活中,常常会遇到以下语境——“以下五条条件中符合三条的获得评比资格”或者“符合以下所有条件的获得评比资格”。这便是不同的属性聚类方式。

为解决此类问题,下面定义悲观属性聚类和一般属性聚类。

**定义 1** 对于形式背景  $K = (G, M, I)$ ,  $R$  为先验关系或特定关系确立的等价关系,  $[m]_R$  为包含属性  $m$  的属性等价类,悲观属性聚类后形成的新形式背景  $(G, M_R^b, I_R^b)$  如下:

$$M_R^b = \{[m]_R; m \in M\}$$

$$I_R^b = \{(g, [m]_R) \in G \times M_R^b; \forall n \in [m]_R, I(g, n) = 1\}$$

**定义 2** 对于形式背景  $K = (G, M, I)$ ,  $R$  为先验关系或特定关系确立的等价关系,  $[m]_R$  为包含属性  $m$  的属性等价类,  $N(m)$  为与约束和  $[m]_R$  有关的常数,原背景在一般属性聚类之后形成的新形式背景  $K_R^N = (G, M_R^N, I_R^N)$  如下:

$$M_R^N = \{[m]_R; m \in M\}$$

$$I_R^N = \left\{ (g, [m]_R) \in G \times M_R^N; \sum_{n \in [m]_R} I(g, n) \geq N(m) \right\}$$

显然,乐观属性聚类与悲观属性聚类均为一般属性聚类的特殊情况:当  $N(m) = 1$  时,为乐观属性聚类;当  $N(m) = |[m]_R|$  时,为悲观属性聚类。因此,在研究属性聚类性质的时候,只需要研究一般属性聚类的性质然后通过改变  $N(m)$  取值研究乐观属性聚类与悲观属性聚类的情况。

例 1 表 1 为某形式背景  $K = (G, M, I)$ , 现给出一组  $R$  等价类的选取方式以及两组不同的  $N(m)$  取值,构建两个形式背景  $K_R^{N_1} = \{G, M_R^{N_1}, I_R^{N_1}\}$  和  $K_R^{N_2} = \{G, M_R^{N_2}, I_R^{N_2}\}$  如表 2。

表 1 形式背景  $K = (G, M, I)$   
Tab. 1 Formal context  $K = (G, M, I)$

$G$	$a$	$b$	$c$	$d$	$e$	$f$	$g$	$h$	$i$	$j$	$k$
$\alpha$	1	1	0	0	0	0	1	0	1	1	0
$\beta$	0	1	1	1	1	1	1	0	0	1	1
$\gamma$	1	0	0	1	1	0	1	0	0	1	1
$\lambda$	0	0	0	0	1	1	1	1	1	0	0

现给出一组等价类  $R$  的构造方法,即归属于同一等价类的属性,在此定义下均为等价:  $[a]_R = \{a, b, c, d\}$ ,  $[e]_R = \{e, f, g, h\}$ ,  $[i]_R = \{i, j, k\}$ 。

给出两组  $([m]_R, N(m))$  的取值,用来比较研究  $N(m)$  取值的大小对属性聚类结果的影响:

$$N_1 = \{([a]_R, N_1(a) = 2), ([e]_R, N_1(e) = 2), ([i]_R, N_1(i) = 2)\}$$

$$N_2 = \{([a]_R, N_2(a) = 3), ([e]_R, N_2(e) = 3), ([i]_R, N_2(i) = 2)\}$$

根据此给出如表 2 所示的两个形式背景。

表 2 形式背景  $K_R^{N_1} = \{G, M_R^{N_1}, I_R^{N_1}\}$  与  $K_R^{N_2} = \{G, M_R^{N_2}, I_R^{N_2}\}$   
Tab. 2 Formal context  $K_R^{N_1} = \{G, M_R^{N_1}, I_R^{N_1}\}$  and  $K_R^{N_2} = \{G, M_R^{N_2}, I_R^{N_2}\}$

$G$	$K_R^{N_1} = \{G, M_R^{N_1}, I_R^{N_1}\}$			$K_R^{N_2} = \{G, M_R^{N_2}, I_R^{N_2}\}$		
	$[a]_R$	$[e]_R$	$[i]_R$	$[a]_R$	$[e]_R$	$[i]_R$
$\alpha$	1	0	1	0	0	1
$\beta$	1	1	1	1	1	1
$\gamma$	1	1	1	0	0	1
$\lambda$	0	1	0	0	1	0

### 2.2 属性聚类下三支概念格的表示

为了研究在属性聚类后三支概念的影响,建立在同一对象集下,属性聚类前后三支概念的联系,故在本文中利用等价类的方式表示属性集。对于属性集  $M = \{m_1, m_2, \dots, m_n\}$ ,

在聚类后以等价类的方式表示为:

$$M_R^N = \{([m]_R)_1, ([m]_R)_2, \dots, ([m]_R)_k\}; k \leq n$$

其中  $\forall 1 \leq i, j \leq k$ , 且  $i \neq j$ , 根据等价类与集合的性质,  $([m]_R)_i \cap ([m]_R)_j = \emptyset$ . 定义一般属性聚类中的算子与负算子, 对于  $X \subseteq G$  和  $A \subseteq M_R$ :

$$X^{*R} = \{[m]_R: [m]_R \in M_R, \forall x \in X, I_R^N(x, [m]_R) = 1\}$$

$$A^{*R} = \{x: x \in G, \forall [m]_R \in A, I_R^N(x, [m]_R) = 1\}$$

$$X^{\overline{*R}} = \{[m]_R: [m]_R \in M_R, \forall x \in X, I_R^N(x, [m]_R) = 0\}$$

$$A^{\overline{*R}} = \{x: x \in G, \forall [m]_R \in A, I_R^N(x, [m]_R) = 0\}$$

性质 1 在形式背景中  $K_R^N = (G, M_R^N, I_R^N)$ , 对于  $\forall X_1, X_2 \subseteq G, A_1, A_2 \subseteq M_R^N$ , 有以下结论成立:

- 1)  $X_1 \subseteq X_2 \Rightarrow X_2^{*R} \subseteq X_1^{*R}$ ;
- 2)  $A_1 \subseteq A_2 \Rightarrow A_2^{*R} \subseteq A_1^{*R}$ ;
- 3)  $X \subseteq X^{*R*R}$ ;
- 4)  $A \subseteq A^{*R*R}$ ;
- 5)  $X^{*R} = X^{*R*R*R}$ ;
- 6)  $A^{*R} = A^{*R*R*R}$ ;
- 7)  $(X_1 \cup X_2)^{*R} = X_1^{*R} \cap X_2^{*R}$ ;
- 8)  $(A_1 \cup A_2)^{*R} = A_1^{*R} \cap A_2^{*R}$ ;
- 9)  $X_1^{*R} \cup X_2^{*R} \subseteq (X_1 \cap X_2)^{*R}$ ;
- 10)  $A_1^{*R} \cup A_2^{*R} \subseteq (A_1 \cap A_2)^{*R}$ .

证明 这里只证明 1)、3)、5)、7)、9), 其余证明过程类似。

1) 对于  $X_1 \subseteq X_2$ , 假设存在  $[m]_R \in X_2^{*R}$  且  $[m]_R \notin X_1^{*R}$ , 即存在至少一个对象  $g \in X_1$  且  $g \notin X_2$ , 令  $I_R^N(g, [m]_R) = 0$ , 与  $X_1 \subseteq X_2$  相矛盾, 故假设不成立, 原命题得证。

3) 对于  $\forall g \in X$ , 对  $\forall [m]_R \in X^{*R}$ , 必有  $I_R^N(g, [m]_R) = 1$ , 且对于  $\forall [m]_R \in X^{*R}$ , 对  $\forall g \in X^{*R*R}$ , 必有  $I_R^N(g, [m]_R) = 1$ , 故  $X \subseteq X^{*R*R}$  成立。

5) 对于  $\forall [m]_R \in X^{*R}$  有  $\forall g \in X^{*R*R}, I_R^N(g, [m]_R) = 1 \Leftrightarrow \forall g \in X^{*R*R}$ , 则其中有  $\forall [m]_R \in X^{*R*R*R}, I_R^N(g, [m]_R) = 1$  成立, 即  $X^{*R} = X^{*R*R*R}$  成立。

7) 对于  $[m]_R \in (X_1 \cup X_2)^{*R}$ , 有  $\forall g_1 \in X_1, g_2 \in X_2$ , 令  $I_R^N(g_1, [m]_R) = 1$  且  $I_R^N(g_2, [m]_R) = 1 \Leftrightarrow \forall g_1 \in X_1, I_R^N(g_1, [m]_R) = 1$  且  $\forall g_2 \in X_2, I_R^N(g_2, [m]_R) = 1 \Leftrightarrow [m]_R \in X_1^{*R}$  且  $[m]_R \in X_2^{*R}$ , 即  $[m]_R \in X_1^{*R} \cap X_2^{*R}$ , 因此, 原命题得证。

9) 对于  $[m]_R \in X_1^{*R} \cup X_2^{*R}$ , 有  $[m]_R \in X_1^{*R}$  或  $[m]_R \in X_2^{*R} \Leftrightarrow \forall g_1 \in X_1, I_R^N(g_1, [m]_R) = 1$  或  $\forall g_2 \in X_2, I_R^N(g_2, [m]_R) = 1 \Leftrightarrow$  对于  $\forall g \in X_1 \cap X_2$ , 则必有  $I_R^N(g, [m]_R) = 1$ , 即对此有  $[m]_R \in (X_1 \cap X_2)^{*R}$ , 因此, 原命题得证。

定义 3 设有形式背景  $K_R^N = (G, M_R^N, I_R^N)$ , 若对于  $X \subseteq G$  和  $A \subseteq M_R$ , 有  $X^{*R} = A$  且  $A^{*R} = X$ , 则称  $(X, A)$  为一个  $R$  粒度一般属性聚类形式概念, 其中概念的外延为  $X$ , 概念的内涵为  $A$ . 形式背景  $K_R^N$  下所有形式概念的集合为  $L(K_R^N)$ .

由定义 3 可知,  $R$  粒度一般属性聚类形式概念也是概念, 符合基本概念的相关性质。

对于  $X \subseteq G$  和  $A, B \subseteq M_R^N$ , 记:  $X^{OR} = (X^{*R}, X^{\overline{*R}})$ ,  $(A, B)^{OR} = (A^{*R} \cap B^{\overline{*R}})$ , 则有下面定义:

定义 4 当存在一组  $X^{OR} = (A, B)$  且  $(A, B)^{OR} = X$ , 则  $(X, (A, B))$  为  $R$  粒度一般属性聚类对象诱导的三支概念, 简

称为 OER 概念, 其中  $X$  为 OER 概念的外延,  $(A, B)$  为 OER 概念的内涵. 记所有 OER 概念为 OER 概念格, 即  $OERL(K_R^N)$ .

对于  $X, Y \subseteq G$  和  $A \subseteq M_R$ :  $A^{AR} = (A^{*R}, A^{\overline{*R}})$ ,  $(X, Y)^{AR} = X^{*R} \cap Y^{\overline{*R}}$ .

定义 5 当存在一组  $A^{AR} = (X, Y)$  且  $(X, Y)^{AR} = A$  时,  $((X, Y), A)$  为  $R$  粒度一般属性聚类下属性诱导的三支概念, 简称为 AER 概念, 其中  $(X, Y)$  为 AER 概念的外延,  $A$  为 AER 概念的内涵. 记所有 AER 概念为 AER 概念格, 即  $AERL(K_R^N)$ .

于是可得 OER 概念与 AER 概念均为三支概念, 满足三支概念相关性质。

例 2 (续例 1) 对于表 2 给出的形式背景, 有以下  $OERL(K_R^N)$  及  $AERL(K_R^N)$ .

对于形式背景  $K_R^{N_1} = \{G, M_R^{N_1}, I_R^{N_1}\}$ ,  $OERL(K_R^{N_1})$  包括:

- $(\{\alpha\}, (\{[a]_R, [i]_R\}, \{[e]_R\}))$
- $(\{\beta, \gamma\}, (\{[a]_R, [i]_R, [e]_R\}, \emptyset))$
- $(\{\lambda\}, (\{[i]_R\}, \{[a]_R, [e]_R\}))$
- $(\{\alpha, \beta, \gamma\}, (\{[a]_R, [i]_R\}, \emptyset))$
- $(\{\beta, \gamma, \lambda\}, (\{[e]_R\}, \emptyset))$
- $(G, (\emptyset, \emptyset))$

$AERL(K_R^{N_1})$  包括:

- $(\{\{\alpha, \beta, \gamma\}, \{\lambda\}\}, \{[a]_R, [i]_R\})$
- $(\{\{\beta, \gamma, \lambda\}, \{\alpha\}\}, \{[e]_R\})$
- $(\{\{\beta, \gamma\}, \emptyset\}, \{[a]_R, [i]_R, [e]_R\})$

对于形式背景  $K_R^{N_2} = \{G, M_R^{N_2}, I_R^{N_2}\}$ ,  $OERL(K_R^{N_2})$  包括:

- $(\{\alpha, \gamma\}, (\{[i]_R\}, \{[e]_R, [a]_R\}))$
- $(\{\beta\}, (M_R^{N_2}, \emptyset))$
- $(\{\lambda\}, (\{[e]_R\}, \{[a]_R, [i]_R\}))$
- $(\{\beta, \gamma\}, (\{[e]_R\}, \emptyset))$
- $(\{\alpha, \beta, \gamma\}, (\{[i]_R\}, \emptyset))$
- $(G, (\emptyset, \emptyset))$

$AERL(K_R^{N_2})$  包括:

- $(\{\{\beta\}, \{\alpha, \gamma, \lambda\}\}, \{[a]_R\})$
- $(\{\{\beta, \lambda\}, \{\alpha, \gamma\}\}, \{[e]_R\})$
- $(\{\{\alpha, \beta, \lambda\}, \{\lambda\}\}, \{[i]_R\})$
- $(\{\{\beta\}, \{\alpha, \gamma\}\}, \{[a]_R, [e]_R\})$
- $(\{\{\beta\}, \emptyset\}, M_R^{N_2})$

### 2.3 属性聚类前后三支概念的比较

在多粒度形式概念的构造中, 往往采用分别构建的方法, 这就造成了大量冗余计算以及资源的浪费. 事实上, 在构建多粒度过程中, 不同粒度的形式背景存在关联. 这种关联可以作为推导路径, 辅助构建多粒度形式概念。

在  $R$  水平一般属性聚类之后, 新的属性由原属性集组成, 新的对应关系可由原对应关系表示. 任何原属性集中的属性, 都能在新属性集中找到对应小属性集合包含, 且该属性集合的存在是唯一的。

三支概念的一种较为简单的获取方式是利用 I 型混合背景与 II 型混合背景, 这种方法比直接利用定义求解更加迅速直观. 在经过  $R$  水平一般属性聚类后, 经  $I_R^N$  原则计算, 新形成的属性要么与某对象存在关系  $I_R^N$ , 要么不存在关系  $I_R^N$ . 以 OE 概念与 OER 概念  $(X, (A, B))$  为例, 对于确定的对象集

$X$ , 在其聚类前的属性集合对  $(A, B)$ , 与其聚类后的属性集合对 (如果存在), 记为  $(A_R, B_R)$ 。任取  $A_R$  中的元素  $[m]_R$ , 必有超过  $N(m)$  个元素与对象集  $X$  中任意元素存在关系  $I$ 。且根据等价类的提取原则,  $A_R$  中各个  $[m]_R$  交集为空。即有:

$$\frac{\left| \left( \bigcup_{[m]_R \in A_R} [m]_R \right) \cap A \right|}{\left| \bigcup_{[m]_R \in A_R} [m]_R \right|} \geq \frac{\sum_{[m]_R \in A_R} N(m)}{\sum_{[m]_R \in A_R} |[m]_R|}$$

也可被记为:

$$\left| \left( \bigcup_{[m]_R \in A_R} [m]_R \right) \cap A \right| \geq \sum_{[m]_R \in A_R} N(m)$$

同理, 任取  $B_R$  中的元素  $[m]_R$ , 必有超过  $(|[m]_R| - N(m))$  个元素与对象集  $X$  中任意元素不存在关系  $I$ 。且根据等价类的提取原则,  $B_R$  中各个  $[m]_R$  交集为空。即有:

$$\frac{\left| \left( \bigcup_{[m]_R \in B_R} [m]_R \right) \cap B \right|}{\left| \bigcup_{[m]_R \in B_R} [m]_R \right|} \geq \frac{\sum_{[m]_R \in B_R} (|[m]_R| - N(m))}{\sum_{[m]_R \in B_R} |[m]_R|}$$

也可被记为:

$$\left| \left( \bigcup_{[m]_R \in B_R} [m]_R \right) \cap B \right| \geq \sum_{[m]_R \in B_R} (|[m]_R| - N(m))$$

对此可以看出, 经过属性聚类的三支概念事实上是对本身严格的约束条件予以放松处理。但是, 该放松处理并不是无限放松, 而是有一定限度的。接下来, 给出最低约束指数的定义来度量这种放松限度。

**定义 6** 对于形式概念  $K = (G, M, I)$  及其经过属性约简后的形式背景  $K_R^N = (G, M_R^N, I_R^N)$ , 记新形式背景下  $R$  水平一般属性聚类下对象诱导的三支概念格为  $OERL(K_R^N)$ , 其中, 概念格中的概念数目为  $|OERL(K_R^N)|$ 。定义在  $R$  等价类属性约简过程中, 对三支概念的最低 OER 约束指数为:

$$e_{OER} = \frac{1}{|OERL(K_R^N)|} \sum_{OERL(K_R^N)} \frac{\sum_{[m]_R \in A_R} N(m)}{\sum_{[m]_R \in A_R} |[m]_R|}$$

根据其性质, 也可记为:

$$e_{OER} = 1 - \left[ \frac{1}{|OERL(K_R^N)|} \sum_{OERL(K_R^N)} \frac{\sum_{[m]_R \in A_R} (|[m]_R| - N(m))}{\sum_{[m]_R \in A_R} |[m]_R|} \right]$$

显然,  $e_{OER}$  的大小与等价类  $R$  以及  $N(m)$  有关。因此, 在进行属性约简时, 如果等价类  $R$  已确定, 可以通过调整  $N(m)$  的设定提高  $e_{OER}$ ; 如果  $N(m)$  已经确定, 可以通过调整等价类  $R$  提高  $e_{OER}$ 。

于是, 可知以下性质成立:

性质 2 对于 AE 概念和 AER 概念  $((X, Y), A)$ , 以及指定的对象集合对  $(X, Y)$ , 记由  $I$  和  $I_R^N$  确定的属性集分别为  $A$  与  $A_R$ 。对于任取  $[m]_R \in A_R$ , 必有  $[m]_R \in X^{*R}$  且  $[m]_R \in Y^{\overline{R}}$ , 则对于  $[m]_R$  必须同时满足:

- 1) 对  $[m]_R \in A_R, \forall x \in X, \sum_{n \in [m]_R} I(x, n) \geq N(m)$ ;
- 2) 对  $[m]_R \in A_R, \forall y \in Y, \sum_{n \in [m]_R} I(y, n) \leq |[m]_R| - N(m)$ 。

进一步, 可以有如下成立:

$$1) \sum_{x \in X} \sum_{[m]_R \in A_R} \sum_{n \in [m]_R} I(x, n) \geq |X| \sum_{[m]_R \in A_R} N(m);$$

$$2) \sum_{y \in Y} \sum_{[m]_R \in A_R} \sum_{n \in [m]_R} I(y, n) \geq |Y| \sum_{[m]_R \in A_R} (|[m]_R| - N(m))。$$

即:

$$1) \frac{\sum_{x \in X} \sum_{[m]_R \in A_R} \sum_{n \in [m]_R} I(x, n)}{|X| \sum_{[m]_R \in A_R} |[m]_R|} \geq \frac{\sum_{[m]_R \in A_R} N(m)}{\sum_{[m]_R \in A_R} |[m]_R|};$$

$$2) \frac{\sum_{y \in Y} \sum_{[m]_R \in A_R} \sum_{n \in [m]_R} I(y, n)}{|Y| \sum_{[m]_R \in A_R} |[m]_R|} \geq \frac{\sum_{[m]_R \in A_R} (|[m]_R| - N(m))}{\sum_{[m]_R \in A_R} |[m]_R|}。$$

**定义 7** 对于形式概念  $K = (G, M, I)$  及其经过属性约简后的形式背景  $K_R^N = (G, M_R^N, I_R^N)$ , 记新形式背景下  $R$  水平一般属性聚类下属性诱导的三支概念格为  $AERL(K_R^N)$ , 其中, 概念格中的概念数目为  $|AERL(K_R^N)|$ 。定义在  $R$  等价类属性约简过程中, 对三支概念的最低 AER 约束指数为:

$$e_{AER} = \frac{1}{|AERL(K_R^N)|} \sum_{AERL(K_R^N)} \frac{\sum_{[m]_R \in A_R} N(m)}{\sum_{[m]_R \in A_R} |[m]_R|}$$

根据其性质, 也可记为:

$$e_{AER} = 1 - \left[ \frac{1}{|AERL(K_R^N)|} \sum_{AERL(K_R^N)} \frac{\sum_{[m]_R \in A_R} (|[m]_R| - N(m))}{\sum_{[m]_R \in A_R} |[m]_R|} \right]$$

显然, 类似于最低 OER 约束指数,  $e_{AER}$  的大小依然与等价类  $R$  以及  $N(m)$  有关。因此, 在进行属性约简时候, 如果等价类  $R$  已确定, 可以通过调整  $N(m)$  的设定来提高  $e_{AER}$ ; 如果  $N(m)$  已经确定, 可以通过调整等价类  $R$  来提高  $e_{AER}$ 。

例 3 (续例 2) 针对例 2 给出的四个概念格, 计算相应的  $e_{OER}$  及  $e_{AER}$ 。

对  $K_R^{N_1} = \{G, M_R^{N_1}, I_R^{N_1}\}$ :

$$e_{OER} = \frac{1}{|OERL(K_R^{N_1})|} \sum_{OERL(K_R^{N_1})} \frac{\sum_{[m]_R \in A_R} N(m)}{\sum_{[m]_R \in A_R} |[m]_R|} = \frac{1}{5} \times \left( \frac{4}{7} + \frac{6}{11} + \frac{4}{7} + \frac{4}{7} + \frac{2}{4} \right) \approx 0.55$$

$$e_{AER}^{N_1} = \frac{1}{3} \times \left( \frac{4}{7} + \frac{2}{4} + \frac{6}{11} \right) \approx 0.54$$

对  $K_R^{N_2} = \{G, M_R^{N_2}, I_R^{N_2}\}$ ,

$$e_{OER}^{N_2} = \frac{1}{5} \times \left( \frac{2}{3} + \frac{8}{11} + \frac{3}{4} + \frac{3}{4} + \frac{2}{3} \right) \approx 0.71$$

$$e_{AER}^{N_2} = \frac{1}{5} \times \left( \frac{3}{4} + \frac{3}{4} + \frac{2}{3} + \frac{6}{8} + \frac{8}{11} \right) \approx 0.73$$

经比可以发现,  $e_{OER}^{N_2} \geq e_{OER}^{N_1}$  且  $e_{AER}^{N_2} \geq e_{AER}^{N_1}$ 。因此, 无论对于哪种三支概念来说, 在聚类过程中,  $K_R^{N_2}$  的约束都强于  $K_R^{N_1}$ 。

### 3 结语

本文是对多粒度属性聚类和三支概念的融合。首先, 将属性聚类推广为一般属性聚类, 乐观属性聚类与悲观属性聚类三种情况; 然后通过对属性聚类前后三支概念的构成, 定义了度量属性聚类对三支概念的放松影响的两个指数; 接着研究发现属性聚类过程中等价类的选取以及约束条件的选择是影响指数大小的两个因素, 进一步完善了三支概念分析理论。

## 参考文献 (References)

- [1] GANTER B, WILLE R. Formal Concept Analysis: Mathematical Foundations[M]. Berlin: Springer, 1999: 183-214.
- [2] 李金海,魏玲,张卓,等. 概念格理论与方法及其研究展望[J]. 模式识别与人工智能, 2020, 33(7): 619-642. (LI J H, WEI L, ZHANG Z, et al. Concept lattice theory and method and its research prospect [J]. Pattern Recognition and Artificial Intelligence, 2020, 33(7): 619-642.)
- [3] 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): 1339-1344. (LIANG J Y, WANG J H. An algorithm for extracting rule-generating sets based on concept lattice[J]. Journal of Computer Research and Development, 2004, 41(8): 1339-1344.)
- [4] 胡学钢,薛峰,张玉红,等. 基于概念格的决策表属性约简方法[J]. 模式识别与人工智能, 2009, 22(4): 624-629. (HU X G, XUE F, ZHANG Y H, et al. Attribute reduction methods of decision table based on concept lattice[J]. Pattern Recognition and Artificial Intelligence, 2009, 22(4): 624-629.)
- [5] 马垣,马文胜. 概念格多属性渐减式构造[J]. 软件学报, 2015, 26(12): 3162-3173. (MA Y, MA W S. Construction of multi-attributes decrement for concept lattice [J]. Journal of Software, 2015, 26(12): 3162-3173.)
- [6] QI J J, WEI L, YAO Y Y. Three-way formal concept analysis[C]// Proceedings of the 2014 International Conference on Rough Sets and Knowledge Technology, LNCS 8818. Cham: Springer, 2014: 732-741.
- [7] QIAN T, WEI L, QI J J. Constructing three-way concept lattices based on apposition and sub-position of formal contexts [J]. Knowledge-Based Systems, 2017, 116: 39-48.
- [8] 苏新,陈永平,杨思春. 基于对象和基于属性的三支概念格合并方法比较[J]. 计算机应用与软件, 2021, 38(11): 282-287. (SU X, CHEN Y P, YANG S C. The comparison of the merging methods of three concept lattices based on object and attribute[J]. Computer Applications and Software, 2021, 38(11): 282-287.)
- [9] WEI L, LIU L, QI J J, et al. Rules acquisition of formal decision contexts based on three-way concept lattices [J]. Information Sciences, 2020, 516: 529-544.
- [10] LI J H, HUANG C C, QI J J, et al. Three-way cognitive concept learning via multi-granularity [J]. Information Sciences, 2017, 378: 244-263.
- [11] 龙柄翰,徐伟华. 模糊三支概念分析与模糊三支概念格[J]. 南京大学学报(自然科学), 2019, 55(4): 537-545. (LONG B H, XU W H. Fuzzy three-way concept analysis and fuzzy three-way concept lattice [J]. Journal of Nanjing University (Natural Science), 2019, 55(4): 537-545.)
- [12] 贺晓丽,折延宏. 基于属性粒化的近似概念分析及规则提取[J]. 山东大学学报(理学版), 2020, 55(5): 13-21. (HE X L, SHE Y H. Approximate concept analysis and rule acquisition based on attribute granularity[J]. Journal of Shandong University (Natural Science), 2020, 55(5): 13-21.)
- [13] 许佳卿,彭鑫,赵文耘. 一种基于模糊形式概念分析的程序聚类方法[J]. 计算机研究与发展, 2009, 46(9): 1556-1566. (XU J Q, PENG X, ZHAO W Y. Program clustering for comprehension based on fuzzy formal concept analysis [J]. Journal of Computer Research and Development, 2009, 46(9): 1556-1566.)
- [14] 李敏,杨亚锋,李丽红. 基于聚类与粒度度量的高效决策表约简[J]. 电脑知识与技术, 2019, 15(26): 246-248. (LI M, YANG Y F, LI L H. Efficient decision table reduction based on clustering and granulation measurement[J]. Computer Knowledge and Technology, 2019, 15(26): 246-248.)
- [15] BELOHLAVEK R, DE BAETS B, KONECNY J. Granularity of attributes in formal concept analysis [J]. Information Sciences, 2014, 260: 149-170.
- [16] LIU Z, LI B, PEI Z, et al. Formal concept analysis via multi-granulation attributes [C]// Proceedings of the 12th International Conference on Intelligent Systems and Knowledge Engineering, Nanjing: Conference Publishing Services, 2017: 1-6.
- [17] 刘美玉,祁建军,刘伟. 三支概念格中的关联规则提取算法[J]. 西安交通大学学报, 2021, 55(9): 189-196. (LIU M Y, QI J J, LIU W. Extracting association rules in three-way concept lattices [J]. Journal of Xi'an Jiaotong University, 2021, 55(9): 189-196.)
- [18] 万青,马盈仓,李金海. 基于直观图的三支概念获取及属性特征分析[J]. 计算机科学与探索, 2022, 16(12): 2879-2889. (WAN Q, MA Y C, LI J H. Three-way concept acquisition and attribute characteristic analysis based on pictorial diagrams [J]. Journal of Frontiers of Computer Science and Technology, 2022, 16(12): 2879-2889.)
- [19] 常丽娜,魏玲. 基于OE-近似概念格的不完备决策背景的规则提取[J]. 山东大学学报(理学版), 2021, 56(11): 31-37. (CHANG L N, WEI L. Rules acquisition based on OE-approximate concept lattice in incomplete formal decision contexts [J]. Journal of Shandong University (Natural Science), 2021, 56(11): 31-37.)

This work is partially supported by National Natural Science Foundation of China (61976245).

**ZHANG Xiaoyan**, born in 1979, Ph. D., associate professor. Her research interests include uncertainty reasoning, mathematical basis of artificial intelligence, granular computing and knowledge discovery.

**WANG Jiayi**, born in 1999, M. S. candidate. Her research interests include formal concept analysis, cognitive computing.