

文章编号:1007-130X(2024)04-0725-09

三支概念背景下属性粒化效率的度量*

张晓燕, 王佳一

(西南大学人工智能学院, 重庆 400715)

摘要:三支概念分析是三支决策与形式概念分析结合的产物,该理论相对于形式概念分析最大的进步是可以同时研究形式背景中“共同具有”和“共同不具有”的信息。属性粒化是一种基于粒度树与剪枝将属性分解为子属性,形成新属性集合的理论。而由于同一粒度树上剪枝众多,如何选择剪枝,确定最优的粒化方向以进行进一步运算成为保证属性粒化效率的关键问题。通过理论推导,证明了原三支概念与属性粒化得到的新三支概念存在紧密的内在联系,以此作为度量属性粒化效率的基础。首先,基于属性粒化层次的关系,将属性粒化层次分为存在偏序关系的属性粒化层次和不存在偏序关系的属性粒化层次。进一步,给出细化系数的含义,并分别阐述了细化系数在 2 种属性粒化层次中的度量作用,从而达到度量不同属性粒化效率的目的。

关键词:三支概念;属性粒化;多粒度;细化系数;粒度树

中图分类号:TP18

文献标志码:A

doi:10.3969/j.issn.1007-130X.2024.04.017

Efficiency measurement of attribute granulation under the background of three-way concept

ZHANG Xiao-yan, WANG Jia-yi

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

Abstract: The three-way concept analysis is a combination of three-way decision and formal concept analysis. The greatest progress of this theory compared with formal concept analysis is that it can simultaneously study the information that is “commonly shared” and “not commonly shared” in the formal context. Attribute granulation is a theory based on the decomposition of attributes into sub-attributes using a granularity tree and pruning, forming a new set of attributes. However, due to the numerous prunings on the same granularity tree, the key issue to ensure the efficiency of attribute granulation is how to choose the pruning and determine the optimal direction for further operations to achieve optimal granulation results. In this paper, through theoretical derivations, it is proved that there is a close internal relationship between the original three-way concepts and the new three-way concepts obtained from attribute granulation, which can be used as the basis for measuring the efficiency of attribute granulation. Firstly, based on the relationship of attribute granulation levels, the attribute granulation levels are divided into attribute granulation levels with partial order relationships and attribute granulation levels without partial order relationships. Furthermore, the definition of refinement coefficients is given, and the measurement roles of refinement coefficients in the two types of attribute granulation levels are respectively explained, so as to achieve the purpose of measuring the efficiency of different attribute granulation directions.

Key words: three-way concept; granularity of attributes; multi-granularity; coefficient of elaboration; granularity tree

* 收稿日期:2022-12-08;修回日期:2023-05-19

基金项目:国家自然科学基金(61976245)

通信地址:400715 重庆市西南大学人工智能学院

Address: College of Artificial Intelligence, Southwest University, Chongqing 400715, P. R. China

1 引言

形式概念分析^[1]是一种以形式背景为研究对象的理论,是一种以概念格相关理论为核心概念的数学工具,常用于进行数据分析。概念格展示了概念之间的偏序关系,刻画对象和属性的内在关系^[2]。然而,基本的概念格因其只研究了“共同(被)具有”关系的问题,忽略了“共同不(被)具有”关系的问题而存在局限性^[3,4],且这种局限性也导致该理论在实际应用中受到限制^[5]。

2014年, Qi^[6]提出了新的理论以弥补经典概念格理论的局限性,即三支概念分析。三支概念分析提出了对象诱导的三支概念格与属性诱导的三支概念格,2种三支概念格均同时研究了“共同(被)具有”和“共同不(被)具有”的问题。其所获得的三支概念格相较于以往经典概念格研究范围更加全面,也会使概念识别在实际应用过程中更加精确。

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^[7]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,仿照对象诱导概念格与属性诱导概念格提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,并分析4种概念之间的异同。苏新等^[8]进行了基于对象和基于属性的三支概念格合并方法比较。Wei等^[9]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。

除对三支概念本身的扩展和推广,学科交叉融合研究也为丰富三支概念体系做出了极大的贡献。Li等^[10]将多粒度与三支概念融合,提出了基于多粒度的三支认知概念学习。龙柄翰等^[11]提出模糊三支概念分析,将模糊集理论与三支概念分析相结合,考虑模糊背景中“共同具有的程度”与“共同不具有的程度”2方面不确定的共性信息极大地提高了三支概念在生产力方面的意义。

多粒度研究方面,知识粒化^[12]、属性粒化^[13,14]和属性聚类^[15]等概念的提出将多粒度与概念认知联系起来^[16],为了缓解庞大的概念个数,在约束宽松的情况下减少工作量提供了解决方案,并且为获取数据的多层次概念知识表示与处理方法上提供了新的方法。多粒度方面的研究包括研究对象的

粗化与细化,在属性粒化研究中,其主要研究方向包括属性的吸收,即粗化,以及属性的分解,即细化。在属性的粗化与细化之中,概念也会随之转化,可以获得在不同粒度上的概念分析。Belohlavek等^[17]提出了给予属性粒化的形式概念分析方法,该理论中进行了属性粒化,生成不同粒度层次的形式背景的工具主要是粒度树与剪枝。

在三支概念格背景下,对属性粒化的研究尚且不足,且现在无法通过有效手段度量属性粒化效率,严重降低了概念区分与细化的速度,需要大量冗余的计算^[18]。针对此种情况,本文提出以三支概念格为背景的属性粒化效率度量方法,尝试探讨对解决此类问题的方法。

2 基础知识

2.1 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 或者 1 这 2 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, 对于 $X \subseteq G, A \subseteq M$, 现定义如下 2 个分别作用于属性子集和对象子集的导出算子 X^* 和 A^* :

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

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

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^* 。

设有形式背景 $K = (G, M, I)$, 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^[20]。

本节内容详见文献^[2]中有关于概念格的理论叙述。

2.2 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”2种关系。在应用中, 只研究单方面往往会使研究结果具有片面性, 从而在利用概念寻找对象和属性的过程中出现错误, 而从正反 2 方

面研究则会使研究结果更加精准,提高寻找结果的正确率。

为了研究“共同不具有”的关系,在这里定义 2.1 节导出算子的负算子。对于子集 $X \subseteq G, A \subseteq M$, 有:

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

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

显然,在负算子作用下,“共同不具有”这一关系可以被研究。但是,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。特别需要说明的一点是,单个对象子集在运算后会得到 2 个属性子集,即“共同具有的属性”和“共同不具有的属性”。同理,由于出发点不同,单个属性子集在运算后会得到 2 个对象子集,从而可以得出 2 种概念。

对于 $X \subseteq G$ 和 $A, B \subseteq M$:

$$X^0 = (X^*, \bar{X}^*)$$

$$(A, B)^0 = A^* \cap \bar{B}^*$$

若 $X^0 = (A, B)$ 且 $(A, B)^0 = X$, 则 $(X, (A, B))$ 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 $A \subseteq M$:

$$A^\wedge = (A^*, \bar{A}^*)$$

$$(X, Y)^\wedge = X^* \cap \bar{Y}^*$$

若 $A^\wedge = (X, Y)$ 且 $(X, Y)^\wedge = A$, 则 $((X, Y), A)$ 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^[21]。

本节内容详见文献[8]中有关三支概念格定义的叙述。

2.3 属性粒化

属性粒化是一种由旧的属性集构建新的属性集的方式,其构建方法是根据不同属性的内在联系,通过粒度树与剪枝的构造得到不同的属性集。相对于纯代数构造方法,通过粒度树这一工具的属性粒化显然要求属性之间存在某种关联,至少是人为规定的联系。

对于形式背景 $K = (G, M, I)$, 任意属性 $m \in M$, 属性 m 的粒度树 T_m 是满足以下全部条件的含有根节点的树:

该树中的任意一个节点都代表一个属性,且都有确定且唯一的名称来标记。树的根节点为 m 。

对于每一个节点,此处记为 z , 均有一对象集合 $z^* \in G$, 该对象集由具有属性 z 的集合构成。

若 z_1, z_2, \dots, z_n 是节点 z 的 n 个子节点, 则

$\{z_1^*, z_2^*, \dots, z_n^*\}$ 是 $\{z\}$ 的一个划分。

剪枝 C 是粒度树 T_m 上满足如下条件的节点集:对于每一个叶节点 u 而言,在从 u 到根节点 m 的路径上存在唯一的节点存在于该剪枝。给定形式背景 $K = (G, M, I)$, 如表 1 所示,其中, $G = \{a, b, c, d, e\}$ 为对象集, $M = \{\text{Color, Black, Red, Blue, Light blue, Dark blue}\}$ 为属性集, I 为 G 和 M 之间的二元关系。

Table 1 Formal background $K = (G, M, I)$

表 1 形式背景 $K = (G, M, I)$

对象	属性					
	Color	Black	Red	Blue	Light blue	Dark blue
a	1	1	0	0	0	0
b	1	0	1	0	0	0
c	1	0	0	1	0	1
d	1	0	0	1	1	0
e	1	0	0	1	1	0

表 1 中基于属性集的粒度树如图 1 所示。

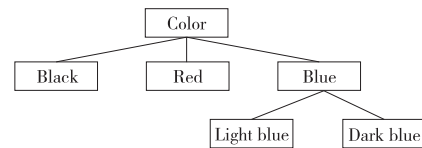


Figure 1 Granularity tree corresponding to attribute set in table 1

图 1 表 1 中基于属性集的粒度树

$\{\text{Red, Light blue, Dark blue}\}$ 不是剪枝,因为其到根节点 Color 的路径上还有 Black 这一子节点。 $\{\text{Blue, Red, Light blue, Dark blue}\}$ 是剪枝。

在形式背景 $K = (G, M, I)$ 的粒度树中任取一剪枝,该剪枝代表了特定的属性粒化层次。任意属性 m , 都是其粒度树 T_m 上的一个剪枝。不同属性粒度树上的剪枝可以导出形式背景 (G, M_C, I_C) , 其中 $M_C = \bigcup_{m \in M} C_m, I_C$ 为 G 与 M_C 的子关系。对于任意 2 个属性粒化层次 C_1 和 C_2 , 若存在偏序关系 $C_1 \leq C_2$, 前者是后者的细化, 则对于 $\forall m \in M, C_{1m} \leq C_{2m}$ 。由此,可在不同的粒化层次上进行概念分析,其概念生成算子如下所示:

对于 $X \subseteq G, A \subseteq M_C$, 有:

$$X^{*C} = \{m : m \in M_C, \forall x \in X, I_C(x, m) = 1\}$$

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

得到粒度树后,可以通过从粒度树上获取不同的剪枝得到不同的粒度层次,有些粒度层次之间存在偏序关系,而有些粒度层次之间不存在偏序关系。本文将对这 2 种情况分别进行讨论。

本节内容详见文献[15]中关于粒度树、剪枝及属性粒化的叙述。

2.4 获取概念格

引理 1 设 $K = (G, M, I)$ 是一个形式背景, 对于 $X, X_0 \subseteq G$ 和 $A \subseteq M$, 下面性质成立:

(1) 若 (X, X^*) 不是一个概念, 那么一定存在 $X_0 \subseteq G$ 使得 $X^* = X_0^*$, 且 (X_0, X^*) 是一个概念。

(2) 若 $X^* = A$ 和 $X \subseteq A^*$, 那么 (A^*, A) 是一个概念。

引理 1 是 Next Clousure 算法的基础, 利用此概念可以获得所有概念, 所有的概念组成概念格 $L(K)$ 。Next Clousure 算法如算法 1 所示。

算法 1 Next Clousure 算法

输入: 形式背景 $K = (G, M, I)$ 。

输出: 全部概念集合 C 。

```

步骤 1 初始化  $C = \emptyset$ ;
        重命名全部对象为  $1:i$ ;
步骤 2 令  $C = C \cup \{\emptyset^{**}, \emptyset^*\}$ ; /* 为对象子集和属性
        子集导出算子, ** 是* 的二次运算, 即复合运算 */
         $D = \text{sort}(G, \text{DESC})$ ; /* 将对象集元素进行分类
        并按降序排序 */
步骤 3  $S = \emptyset^*$ ;
        while  $S \neq \emptyset^{**}$  do
             $J = D - S$ ;
            foreach  $i$  in  $J$  do
                 $F = (S \cap \{1, \dots, i-1\} \cup \{i\})^*$ ;
                 $T = F^*$ ;
                 $U = T \cap \{1, \dots, i-1\}$ ;
                if  $i \in T - S$  and  $S \cap \{1, \dots, i-1\} = U$  then
                     $C = C \cup \{T, F\}$ ;
                     $S = T$ ;
                    break;
                endif
            endfor
        endwhile
步骤 4 return  $C$ 

```

利用 Next Clousure 算法同样可以得到反概念。反概念是相对概念而言的, 反概念可以看作是原形式背景的补背景下的概念。对于一个形式背景, 所有反概念组成该形式背景下的反概念格。

3 存在偏序关系的不同粒度层次

3.1 属性粒化前后三支概念

属性粒化前后, 形式概念背景中的属性集也随

之进行了粗细转化, 在由粗属性集到细属性集的转化中, 对于相同对象集, 新属性可能部分出现在与之对应的属性集中, 也有可能全部不出现在对应的属性集中。

现实应用中, 属性粒化常用于更加精确地定义概念。例如, 在将“不同种类牛仔裤”进行区分定义时, 某属性可以由“蓝色”到“深蓝色”及“浅蓝色”, 再到具体色号层层细化, 使定义更加细化。而在属性粒化由粗到细的过程中, 随着属性集向更细的方向转化, 必定会有更多概念的形成。以下将对此结论进行理论证明。

为了方便说明, 对于剪枝后的形式背景 $K_C = (G, M_C, I_C)$ 形成的概念格记为 $L(K_C)$ 。对于概念格中的概念数量, 由于概念格本质上是集合, 将其元素的数目记为 $|L(K_C)|$ 。相似地, 由这种概念格形成的面向对象三支概念格记为 $OEL(K_C)$, 面向属性三支概念格记为 $AEL(K_C)$ 。

定理 1 在属性粒化的过程中, 随着属性粒度层次的细化, 属性集中元素的增多, 概念格中概念的数量不会减少。

证明 在属性粒化过程中, 对于任意属性分解过程 $a \rightarrow \{a_1, a_2, \dots, a_l\}$, 若该属性分解导致概念格中元素减少, 则为在原概念格中存在概念的内涵 B , $a \in B$ 的基础上, 新概念格中不存在概念的内涵 B_C , 令 $B_C \cap \{a_1, a_2, \dots, a_l\} \neq \emptyset$ 。

对于新属性集中的元素 $\forall a_p (1 \leq p \leq l) \in \{a_1, a_2, \dots, a_l\}$, 设 $A = (M \cap M_C) \cup \{a_p\}$, 则因为 $(A^*C)^*C = A^*C^*C$, 且 $(A^*C^*C)^*C = A^*C^*C^*C = A^*C$, 因此 (A^*C, A^*C^*C) 是一个概念格, 因为 $A \subseteq A^*C^*C$, 所以 $a_p \in A^*C^*C$, 则在新的概念格中存在概念内涵 $A^*C^*C \cap \{a_1, a_2, \dots, a_l\} \neq \emptyset$ 。以上假设不成立, 即任意属性分解均不会造成概念格中元素减少, 原定理得证。

类似地, 有如下定理:

定理 2 在属性粒化的过程中, 随着属性粒度层次的细化, 属性集中元素增多, 面向对象三支概念格中概念的数量不会减少。

定理 3 在属性粒化的过程中, 随着属性粒度层次的细化, 属性集中元素的增多, 面向属性三支概念格中概念的数量不会减少。

3.2 细化系数

为了研究属性聚类对三支概念的影响, 需要在同一对象集下, 进一步建立属性聚类前后三支概念的联系。对于属性集 $M = \{m_1, m_2, \dots, m_{|M|}\}$, 聚

类后表示为 $M_R = \{([m]_R)_1, ([m]_R)_2, \dots, ([m]_R)_k\} (k \leq |M|)$, 其中, R 为聚类关系, $\forall 1 \leq p, q \leq k$, 且 $p \neq q$, 根据等价类与集合的性质, $([m]_R)_p \cap ([m]_R)_q = \emptyset$ 。定义一般属性聚类中的算子与负算子, 对于 $X \subseteq G$ 和 $A \subseteq M_R$: 如果在定义事物的时候, 选择描述事物的关键词越多, 分类越细致, 对于事物的定义就会越精确。相似地, 概念格在不同的属性层次移动, 由粗到细的时候, 概念也会由粗到细, 得到更加精确的概念。概念格中概念的数量也会随之变化。根据定理 1, 较粗属性层次对应的概念格中的概念数量会大于或等于较细属性层次对应的概念格中概念的数量, 因此可以通过比较层次移动时概念格中概念数量的变化来比较属性粒化的效果。

定义 1 对于属性粒化层次 C_1 和 C_2 , 若存在偏序关系 $C_1 \leq C_2$, 即前者是后者的细化, 在度量细化时, 定义细化系数 e_{LC} 如式(1)所示:

$$e_{LC} = \frac{|L(K_{C_1})|}{|L(K_{C_2})|} \quad (1)$$

细化系数 e_{LC} 越大, 表明新的属性层次产生了越多的新概念, 此次属性粒化越有效。

相似地, 根据定理 2 和定理 3, 可以定义出与以上细化系数相同的面向对象细化系数和面向属性细化系数。

定义 2 对于属性粒化层次 C_1 和 C_2 , 若存在偏序关系 $C_1 \leq C_2$, 即前者是后者的细化, 在度量细化时, 定义面向对象细化系数 e_{OELC} 如式(2)所示:

$$e_{OELC} = \frac{|OEL(K_{C_1})|}{|OEL(K_{C_2})|} \quad (2)$$

定义 3 对于属性粒化层次 C_1 和 C_2 , 若存在偏序关系 $C_1 \leq C_2$, 即前者是后者的细化, 在度量细化时, 定义面向属性细化系数 e_{AELC} 如式(3)所示:

$$e_{AELC} = \frac{|AEL(K_{C_1})|}{|AEL(K_{C_2})|} \quad (3)$$

以下用示例说明 2 种细化系数的用法。

例 1 现有一购物网站搜索场景利用如表 2 所示形式背景 $K = (G, M, I)$ 表示, 其中, $M = \{m_1, m_2, m_3\}$, $M_{C_1} = \{m_1, m_2, m_{31}, m_{32}\}$, $M_{C_2} = \{m_1, m_2, m_{31}, m_{32}, m_{311}, m_{312}\}$ 。对象代表 4 种不同的鞋品, 属性代表商品的特征, m_1, m_2 和 m_3 分别代表网面材质、黑色和安踏品牌, 且粒度树如图 2 所示。对于存在偏序关系 $C_2 < C_1$ 的属性粒化层次, 及生成 2 个新的形式背景 $K_{C_1} = (G, M_{C_1}, I_{C_1})$ 和

$K_{C_2} = (G, M_{C_2}, I_{C_2})$, m_{31} 和 m_{32} 分别代表安踏品牌休闲系列和安踏品牌运动系列, m_{311} 和 m_{312} 分别代表安踏品牌运动系列男款和安踏品牌运动系列女款。

Table 2 Formal background $K = (G, M, I)$ for example 1

表 2 例 1 对应的形式背景 $K = (G, M, I)$

对象	m_1	m_2	m_3	m_{31}	m_{32}	m_{311}	m_{312}
a	1	0	1	1	0	1	0
b	0	1	1	1	0	0	1
c	1	1	1	0	1	0	0
d	1	0	0	0	0	0	0

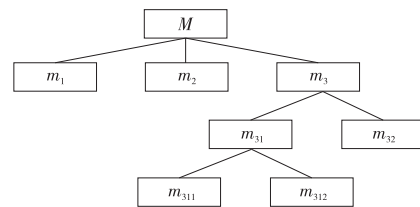


Figure 2 Granularity tree of table 2

图 2 表 2 中的粒度树

计算可得:

$$|OEL(K)| = 8, |AEL(K)| = 7, |OEL(K_{C_1})| = 11, e_{OELC_1} = 11/8, |AEL(K_{C_1})| = 10, e_{AELC_1} = 10/7。$$

同理, $|OEL(K_{C_2})| = 12, e_{OELC_2} = 12/11 > 1; |AEL(K_{C_2})| = 15, e_{AELC_2} = 3/2 > 1$ 。该属性粒化在 C_2 上存在意义。

4 不存在偏序关系的不同粒度层次

在实际应用中, 经常使用属性粒化对现有概念进行细化, 而属性细化方向的选择则与属性粒化的效率息息相关, 细化方向选择的不同会导致属性粒化效率的差别, 例如, 在试图细化“牛仔裤”这个概念的时候, 存在多种细化方向, 显然, 将“蓝色”细化为“深蓝色”和“浅蓝色”相较于将“红色”细化为“浅红色”和“深红色”更加有效。

例 2 现有一购物网站搜索场景利用如表 3 所示的形式背景 $K = (G, M, I)$ 表示, 其中, $M = \{m_1, m_2, m_3\}$, $M_{C_1} = \{m_1, m_2, m_{31}, m_{32}\}$, $M_{C_3} = \{m_{11}, m_{12}, m_2, m_3\}$ 。对象代表 4 种不同的鞋品, 属性代表商品的特征, m_1, m_2 和 m_3 分别代表网面材质、黑色和安踏品牌, 且粒度树如图 3 所示。不存在偏序关系的属性粒化层次及生成的新形式背景 $K_{C_1} = (G, M_{C_1}, I_{C_1})$ 和 $K_{C_3} = (G, M_{C_3}, I_{C_3})$, m_{11} 和 m_{12} 分别代表全网面鞋和半网面鞋, m_{31} 和

m_{32} 分别代表安踏品牌休闲系列与安踏品牌运动系列。

Table 3 Formal background $K=(G,M,I)$ for example 2
表 3 例 2 对应的形式背景 $K=(G,M,I)$

对象	m_1	m_{11}	m_{12}	m_2	m_3	m_{31}	m_{32}
a	1	1	0	0	1	1	0
b	0	0	0	1	1	1	0
c	1	0	1	1	1	0	1
d	1	1	0	0	0	0	1

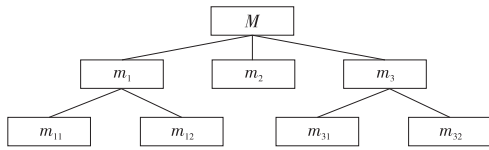


Figure 3 Granularity tree of table 3

图 3 表 3 中的粒度树

计算可得:

$$|OEL(K_{C_3})| = 10, e_{OELC_3} = 5/4 < e_{OELC_1}$$

$$|AEL(K_{C_3})| = 10, e_{AELC_3} = 10/7 = e_{AELC_1}$$

因此,属性粒化方向 C_1 优于 C_3 。

在具体的现实操作中,细化方向的选择往往更加复杂,且相较于以上实例更加抽象和不直观,因此通过细化算子来度量不存在偏序关系的不同粒度层次的优劣,以此来选择粒度层次进行进一步细化,从而避免冗余的计算和资源的浪费,是十分有必要的。

5 算法

5.1 理论依据

以下一组定理是获取面向对象三支概念格与面向属性三支概念格的理论依据。

定理 4 设 $K=(G,M,I)$ 是一个形式背景,若 (X,A) 是一个概念且 (X,B) 是一个反概念,那么 $(X,(A,B))$ 一定是一个 OE 概念,记由此得到的概念集合为 O_1 。

定理 5 设 $K=(G,M,I)$ 是一个形式背景,若 (X,A) 是一个概念且 (Y,B) 是一个反概念,且 Y 为 X 的最小上界, X 不属于 $FL(K)$ 的外延集,那么 $(X,(A,B))$ 一定是一个 OE 概念,记由此得到的概念集合为 O_2 。

定理 6 设 $K=(G,M,I)$ 是一个形式背景,若 (X,A) 是一个概念且 (Y,B) 是一个反概念,且 X 为 Y 的最小上界, Y 不属于 $L(K)$ 的内涵集,那么 $(Y,(A,B))$ 一定是一个 OE 概念,记由此得到

的概念集合为 O_3 。

定理 7 设 $K=(G,M,I)$ 是一个形式背景,若 (X,A) 是一个概念且 (Y,B) 是一个反概念,且 X,Y 均为 $X \cap Y$ 的最小上界, X 不属于 $FL(K)$ 的外延集, Y 不属于 $L(K)$ 的内涵集,那么 $(X \cap Y,(A,B))$ 一定是一个 OE 概念,记由此得到的概念集合为 O_4 。

定理 8 设 $K=(G,M,I)$ 是一个形式背景,则 $OEOL(K) = O_1 \cup O_2 \cup O_3 \cup O_4$ 。

定理 9 设 $K=(G,M,I)$ 是一个形式背景,若 (X,A) 是一个概念且 (Y,A) 是一个反概念,那么 $((X,Y),A)$ 一定是一个 AE 概念,记由此得到的概念集合为 A_1 。

定理 10 设 $K=(G,M,I)$ 是一个形式背景,若 (X,A) 是一个概念且 (Y,B) 是一个反概念,且 B 为 A 的最小上界, A 不属于 $FL(K)$ 的内涵集,那么 $((X,Y),A)$ 一定是一个 AE 概念,记由此得到的概念集合为 A_2 。

定理 11 设 $K=(G,M,I)$ 是一个形式背景,若 (X,A) 是一个概念且 (Y,B) 是一个反概念,且 A 为 B 的最小上界, B 不属于 $L(K)$ 的内涵集,那么 $((X,Y),B)$ 一定是一个 AE 概念,记由此得到的概念集合为 A_3 。

定理 12 设 $K=(G,M,I)$ 是一个形式背景,若 (X,A) 是一个概念且 (Y,B) 是一个反概念,且 A,B 均为 $A \cap B$ 的最小上界, A 不属于 $FL(K)$ 的内涵集, B 不属于 $L(K)$ 的内涵集,那么 $((X,Y),A \cap B)$ 一定是一个 AE 概念,记由此得到的概念集合为 A_4 。

定理 13 设 $K=(G,M,I)$ 是一个形式背景,则 $AEPL(K) = A_1 \cup A_2 \cup A_3 \cup A_4$ 。

5.2 获取 e_{OELC} 算法

根据以上理论可以获取 e_{OELC} , 如算法 2 所示。

算法 2 获取 e_{OELC}

输入:概念格 $L(K),L(K_C)$ 与反概念格 $FL(K),FL(K_C)$ 。

输出:参数 e_{OELC} 。

步骤 1 初始化 $O_1 = \emptyset, O_2 = \emptyset, o_1 = 0, o_2 = 0$;

步骤 2 令 $L_{pG} \cap F_{qG} = W$;

if $W = L_{pG}$ 且 $W = F_{qG}$

then $O_1 = O_1 \cup (W, (L_{pM}, F_{qM}))$;

/* L_{pG} 和 L_{pM} 分别表示在 $L(K)$ 中概念的外延和内涵, F_{qG} 和 F_{qM} 分别表示在 $FL(K)$ 中反概念的外延和内涵 */

else if $W = L_{pG}, W \neq F_{qG}$ 且 F_{qG} 为 W 最小上界

```

    then  $O_1 = O_1 \cup (W, (L_{pM}, F_{qM}));$ 
else if  $W \neq L_{pG}, W = F_{qG}$  且  $L_{pG}$  为  $W$  最小上界
    then  $W_1 = W_1 \cup (W, (L_{pM}, F_{qM}));$ 
else if  $W \neq L_{pG}, W \neq F_{qG}$  且  $L_{pG}$  与  $F_{qG}$  均为  $W$  最小上界
    then  $O_1 = O_1 \cup (W, (L_{pM}, F_{qM}));$ 
end if
 $o_1$  为  $O_1$  中元素的个数;
步骤 3 令  $L_{pGC} \cap F_{qGC} = V;$ 
if  $V = L_{pGC}$  且  $V = F_{qGC}$ 
    then  $O_2 = O_2 \cup (V, (L_{pMC}, F_{qMC}));$ 
    /*  $L_{pGC}$  和  $L_{pMC}$  分别表示在  $L(K_C)$  中概念的外延和内涵,  $F_{qGC}$  和  $F_{qMC}$  分别表示在  $FL(K_C)$  中反概念的外延和内涵 */
else if  $V = L_{pGC}, V \neq F_{qGC}$  且  $F_{qGC}$  为  $V$  最小上界
    then  $O_2 = O_2 \cup (V, (L_{pMC}, F_{qMC}));$ 
else if  $V \neq L_{pGC}, V = F_{qGC}$  且  $L_{pGC}$  为  $V$  最小上界
    then  $O_2 = O_2 \cup (V, (L_{pMC}, F_{qMC}));$ 
else if  $V \neq L_{pGC}, V \neq F_{qGC}$  且  $L_{pGC}$  与  $F_{qGC}$  均为  $V$  最小上界
    then  $O_2 = O_2 \cup (V, (L_{pMC}, F_{qMC}));$ 
end if
步骤 4  $e_{OELC} = o_2/o_1;$ 
步骤 5 输出  $e_{OELC}$ 

```

例 3 现某购物网站需要对搜索机制改良,对 10 件短袖的标签进行细分,根据最新网络关键词捕捉报告,近期“全棉”与“含棉混合面料”的区别引起消费者广泛讨论,故生成如图 4 所表示的粒度树。

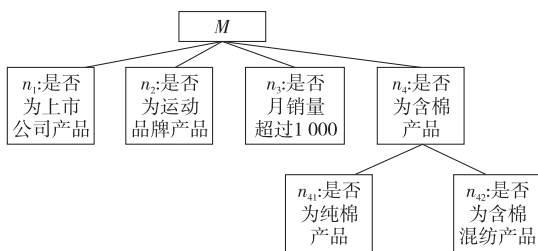


Figure 4 Granularity tree of commodity research
图 4 商品调研粒度树

根据粒度树与“商品-标签”对应关系,有如表 4 所示形式背景 $K = (G, M, I)$ 和 $K_C = (G, M_C, I)$, 其中, $G = \{\alpha, \beta, \dots, \tau\}, M = \{n_1, n_2, n_3, n_4\}, M_C = \{n_1, n_2, n_3, n_{41}, n_{42}\}$ 。

现以此例,从准确性和效率性评估算法 2(与直接计算作比较)。

利用直接计算方法,求得 $OEL(K)$ 需要进行算子计算(包括正算子和负算子)544次,交集运算

Table 4 Formal background $K = (G, M, I)$ for example 3
表 4 例 3 对应的形式背景 $K = (G, M, I)$

对象	n_1	n_2	n_3	n_4	n_{41}	n_{42}
α	1	0	0	0	0	0
β	0	1	1	0	0	0
χ	1	1	1	0	0	0
δ	0	0	1	1	1	0
ϵ	0	1	1	1	0	1
φ	1	1	0	0	0	0
γ	0	0	0	1	1	0
η	1	0	0	1	0	1
ι	0	0	1	0	0	0
κ	1	1	1	1	1	0

256 次,比较运算 512 次,最后得出 $|OEL(K)| = 33$;求得 $OEL(K_C)$ 需要进行算子计算(包括正算子和负算子)2 112 次,交集运算 1 024 次,比较运算 2 048 次,最后得出 $|OEL(K_C)| = 52$ 。共计使用算子计算 2 656 次,交集运算 1 280 次,比较运算 2 560 次。最终计算得出 $e_{OELC} = 52/33$ 。

利用本文的改进算法 Next Closure 算法计算概念格与反概念格。计算 K 的概念格使用算子运算 32 次,比较运算 16 次;计算 K_C 的概念格使用算子运算 64 次,比较运算 32 次。利用 2 个计算结果进一步计算 e_{OELC} 的过程中,使用交集运算 1 280 次,比较运算 256 次,最终输出 $e_{OELC} = 52/33$ 。

在实验过程中,需要尤为注意的是,算子计算的复杂度与其他计算是不可同日而语的。以本数据集为例,一次算子计算包含 4~50 次的运算;而一次比较计算至多只涉及 5 次对应计算。因此,效率性的评估以算子计算的数目为主要评价指标。

综上,在计算 e_{OELC} 时,使用本文提出的算法相较直接计算方法大幅减少了正算子及负算子的计算量,极大降低了计算复杂度。2 次计算输出的 e_{OELC} 相等。算法 2 相比直接计算方法在保证准确性的前提下,极大提升了计算效率。

5.3 获取 e_{AELC} 算法

根据以上理论可以获取 e_{AELC} , 如算法 3 所示。

算法 3 获取 e_{AELC}

输入:概念格 $L(K), L(K_C)$, 与反概念格 $FL(K), FL(K_C)$ 。

输出:参数 e_{AELC} 。

步骤 1 初始化 $A_1 = \emptyset, A_2 = \emptyset, a_1 = 0, a_2 = 0$;

步骤 2 令 $L_{pM} \cap F_{qM} = N$;

if $N = L_{pM}$ 且 $N = F_{qM}$

then $A_1 = A_1 \cup ((L_{pG}, F_{qG}), N)$;
 / * L_{pG} 和 L_{pM} 分别表示在 $L(K)$ 中概念的外延和内涵, F_{qG} 和 F_{qM} 分别表示在 $FL(K)$ 中反概念的外延和内涵 * /
else if $N = L_{pM}, N \neq F_{qM}$ 且 F_{qM} 为 N 最小上界
then $A_1 = A_1 \cup ((L_{pG}, F_{qG}), N)$;
else if $N \neq L_{pM}, N = F_{qM}$ 且 L_{pM} 为 N 最小上界
then $A_1 = A_1 \cup ((L_{pG}, F_{qG}), N)$;
else if $N \neq L_{pM}, N \neq F_{qM}$ 且 L_{pM}, F_{qM} 均为 N 最小上界
then $A_1 = A_1 \cup ((L_{pG}, F_{qG}), N)$;
end if
 a_1 为 A_1 中元素的个数;

步骤 3 令 $L_{pMC} \cap F_{qMC} = Q$;

if $Q = L_{pMC}$ 且 $Q = F_{qMC}$
then $A_2 = A_2 \cup ((L_{pGC}, F_{qGC}), Q)$;
else if $Q = L_{pMC}, Q \neq F_{qMC}$ 且 F_{qMC} 为 Q 最小上界
then $A_2 = A_2 \cup ((L_{pGC}, F_{qGC}), Q)$;
else if $Q \neq L_{pMC}, Q = F_{qMC}$ 且 L_{pMC} 为 Q 最小上界
then $A_2 = A_2 \cup ((L_{pGC}, F_{qGC}), Q)$;
else if $Q \neq L_{pMC}, Q \neq F_{qMC}$ 且 L_{pMC}, F_{qMC} 均为 Q 最小上界
then $A_2 = A_2 \cup ((L_{pGC}, F_{qGC}), Q)$;
end if
 a_2 为 A_2 中元素的个数;

步骤 4 $e_{AELC} = a_2/a_1$;

步骤 5 输出 e_{AELC}

接例 3,从准确性和效率性评估算法 3(与直接计算作比较)。

利用直接计算方法,求得 $AEL(K)$ 需要进行算子计算 64 次,交集运算 16 次,比较运算 32 次,最后得出 $|AEL(K)|=12$;求得 $AEL(K_C)$ 需要进行算子计算(包括正算子和负算子)128 次,交集运算 32 次,比较运算 64 次,最后得出 $|AEL(K_C)|=17$ 。共计使用算子计算 192 次,交集运算 48 次,比较运算 96 次。最终计算得出 $e_{AELC} = 17/12$ 。

利用本文改进算法 Next Clousure 算法计算概念格与反概念格。计算 K 的概念格使用算子运算 32 次,比较运算 16 次;计算 K_C 的概念格使用算子运算 64 次,比较运算 32 次。利用 2 个计算结果进一步计算 e_{AELC} 过程中,使用交集运算 1 280 次,比较运算 256 次,最终输出 $e_{AELC} = 17/12$ 。2 次计算输出的 e_{AELC} 相等。

综上,在计算 e_{AELC} 时,使用本文提出算法相比

直接计算方法减少了正算子及负算子的计算量。算法 3 相比直接计算方法在保证准确性的前提下,极大地提升了计算效率。

6 结束语

本文对概念格、三支概念格以及属性粒化的定义及性质进行了概述,并且研究了属性粒化前后三支概念的关系,定义细化系数,以此度量属性粒化的效率。本文研究结果仍存在不能精确选择、不能进行精确性度量等问题。下一步将针对目前发现的问题进行深入的研究。

参考文献:

- [1] Gantter B, Wille R. Formal concept analysis: Mathematical foundations[M]. New York: Springer Verlag, 1999.
- [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 Jin-hai, Wei Ling, Zhang Zhuo, 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 Ji-ye, Wang Jun-hong. 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 Xue-gang, Xue Feng, Zhang Yu-hong, 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 Yuan, Ma Weng-sheng. Construction of multi-attribute decrement for concept lattice[J]. Journal of Software, 2015, 26(12): 3162-3173.
- [6] Qi J J. Three-way formal concept analysis[C]//Proc of the 9th International Conference on Rough Sets and Knowledge Technology, 2014: 732-741.
- [7] Qian T, Wei L, Qi J J. Constructing three-way concept lattices based on apposition and subposition 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 Xin, Chen Yong-ping, Yang Si-chun. 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(自然科学版), 2019, 55(4): 537-545.
Long Bing-han, Xu Wei-hua. Fuzzy three-way concept analysis and fuzzy three-way concept lattice [J]. Journal of Nanjing University (Natural Science), 2019, 55(4): 537-545.
- [12] 何流, 张贤勇, 唐孝. 基于知识粒化的粗糙集近似精度的泛化改进[J]. 系统工程理论与实践, 2021, 41(5): 1343-1352.
He Liu, Zhang Xian-yong, Tang Xiao. Knowledge granulation-based generalized improvement of approximate accuracy regarding rough sets [J]. Systems Engineering—Theory & Practice, 2021, 41(5): 1343-1352.
- [13] 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 Xiao-li, She Yan-hong. Approximate concept and rule acquisition based on attribute granularity[J]. Journal of Shandong University (Natural Science), 2020, 55(5): 13-21.
- [14] 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 Jia-qing, Peng Xin, Zhao Wen-yun. Program clustering for comprehension based on fuzzy formal concept analysis [J]. Journal of Computer Research and Development, 2009, 46(9): 1556-1566.
- [15] 折延宏, 胡梦婷, 贺晓丽, 等. 两种多粒度形式概念分析模型比较研究[J]. 计算机工程与应用, 2020, 56(10): 51-55.
She Yan-hong, Hu Meng-ting, He Xiao-li, et al. Comparative study of two multi granularitition formal concept analysis models [J]. Computer Engineering and Applications, 2020, 56(10): 51-55.
- [16] 李敏, 杨亚锋, 李丽红. 基于聚类与粒化度量的高效决策表约简[J]. 电脑知识与技术, 2019, 15(26): 246-248.
Li Min, Yang Ya-feng, Li Li-hong. Efficient decision table reduction based on clustering and granulation measurement [J]. Computer Knowledge and Technology 2019, 15 (26): 246-248.
- [17] Belohlavek R, De B B, Konecny J. Granularity of attributes in formal concept analysis [J]. Information Sciences, 2014, 260: 149-170.
- [18] Liu Z C, Li B, Pei Z, et al. Formal concept analysis via multi-granulation attributes[C]//Proc of 2017 12th International Conference on Intelligent Systems and Knowledge Engineering, 2017: 1-6.
- [19] 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 Mei-yu, Qi Jian-jun, Liu Wei. Extracting association rules in three-way lattices [J]. Journal of Xi'an Jiaotong University, 2021, 55(9): 189-196.
- [20] 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 Qing, Ma Ying-cang, Li Jin-hai. 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.
- [21] 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 Li-na, Wei Ling. 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.

作者简介:



张晓燕(1979-),女,山西怀仁人,博士,副教授,CCF会员(K3429M),研究方向为不确定性推理、人工智能的数学基础、粒计算与知识发现。**E-mail:** zxy19790915@163.com

ZHANG Xiao-yan, born in 1979, PhD, associate professor, CCF member(K3429M), her research interests include uncertainty reasoning, mathematical basis of artificial intelligence, granular computing, and knowledge discovery.



王佳一(1999-),女,河北保定人,硕士生,研究方向为形式概念分析和认知计算。**E-mail:** 1335083303@qq.com

WANG Jia-yi, born in 1999, MS candidate, her research interests include formal concept analysis and cognitive computing.