

《智能信息处理》课程作业

## 三支概念分析的理论研究现状及其技术展望

刘行顺

作业	分数[20]
得分	

2021 年 11 月 29 日

# 三支概念分析的理论研究现状及其技术展望

刘行顺<sup>1)</sup>

(大连海事大学信息科学技术学院 辽宁大连)

**摘 要** 概念格是形式概念分析理论中的核心数据分析工具，三支决策是三分而治思想的理论化、方法化与策略化，两者的结合已带来诸多新问题、新思想与新方法。本文简述了形式概念分析的形成意义以及发展进程，并在此基础上详细介绍了三支概念分析的相关技术。最后本文给出了三支概念分析未来优缺点的展望。

**关键词** 形式概念分析；三支概念分析；概念格；基本理论；技术展望

## The current situation of theoretical research and its technical prospect of three conceptual analyses

Liu Xingshun<sup>1)</sup>

**Abstract** Concept grid is the core data analysis tool in the theory of formal concept analysis, and the three decisions are the theoretical, methodological and strategic of the three-part and ruling thought, and the combination of the two has brought many new problems, new ideas and new methods. This paper briefly describes the formation significance of formal conceptual analysis and the development process, and on this basis, introduces in detail the three related techniques of conceptual analysis. Finally, this paper gives a three-to-concept analysis of future advantages and disadvantages of the outlook.

**Key words** Formal concept analysis; three conceptual analyses; concept grid; basic theory; technology outlook

## 引言

随着信息技术的飞速发展，知识的获取显得愈发重要。形式概念分析(Formal Concept Analysis)[1]是知识表示与处理的一种有效的数学工具，自1982年 Wille[2]提出这一概念以来，人们围绕概念格的构造、知识发现、规则提取与数学分析以及其它方面的应用等等展开了深入地研究，取得了一系列研究成果。形式概念分析理论(FCA)是数据分析和知识获取的有效数学工具之一，对哲学中的“概念”这一人类思维的基本单元进行了形式化表达，进一步又形成概念格这一核心数据结构，进而进行知识表示。FCA的基本概念是形式背景、形式概念和相应的概念格，它们由对象集和属性集之间的二元关系决定。概念格作为形式概念分析的核心结构，反映了概念之间的层次结构，通过数据中的形式概念和层次结构来反映数据中隐藏的信息和知识。近年来，FCA受到越来越多的关注。

三支概念分析(Three-way Concept Analysis, 3WCA)[3]是 QI 等结合三支决策(Three-way

Decisions, 3WD)[4]与形式概念分析(Formal Concept Analysis, FCA)[5]的思想提出的一种新理论。3WCA的研究对象主要是由对象(属性)导出的三支概念。3WCA 的提出，既为三支决策理论提供了一种新的模型，也丰富了形式概念分析理论。

三支决策是 Yao 于 2012 年提出的一种决策方法，其核心思想是按照评价标准，将给定区域划分成正域、负域、边界域，并采取相应策略进行处理。随后，三支决策广泛的应用于工程、管理等领域。

## 1 FCA 基本理论

**定义 1.**<sup>[6]</sup>称三元组  $(G, M, I)$  是一个形式背景，其中  $G = \{x_1, x_2, \dots, x_n\}$ ，每一个  $x_i (i \leq n)$  被称为一个对象； $M = \{a_1, a_2, \dots, a_l\}$ ，每一个  $a_i$  是一个属性；对于  $x \in G$  与  $a \in M$ ，如果  $x I a$ ，记  $(x, a) \in I$ ，并且称对象  $x$  拥有属性  $a$ ，或者属性  $a$  被对象  $x$  拥有。

Wille 和 Ganter 在形式背景上定义了一对算子，并由该诱导算子生成形式概念。

**定义 2.**<sup>[6]</sup>设  $(G, M, I)$  是一个形式背景，

$X \subseteq G, A \subseteq M$ , 一对算子定义如下:  $*$ :  
 $P(G) \rightarrow P(M), X^* = \{a \in M \mid \forall x \in X, (x, a) \in I\}$ ;  
 $*: P(M) \rightarrow P(G), A^* = \{x \in G \mid \forall a \in A, (x, a) \in I\}$ .

定义 3.<sup>[6]</sup> 设  $(G, M, I)$  是一个形式背景,  $X \subseteq G, A \subseteq M$ . 若  $X^* = A$  且  $A^* = X$ , 则称  $(X, A)$  为形式概念, 简称概念. 其中  $X$  为  $(X, A)$  的外延,  $A$  为  $(X, A)$  的内涵.

由定义 3 可知, 对于一个形式概念  $(X, A)$  来说,  $*$  算子描述了对象与属性间“共同拥有”的含义.

记形式背景  $(G, M, I)$  上所有概念构成的集合为  $L(G, M, I)$ .

对于  $(X_1, A_1), (X_2, A_2) \in L(G, M, I)$ , 定义其偏序关系为:  $(X_1, A_1) \leq (X_2, A_2) \Leftrightarrow X_1 \subseteq X_2 \Leftrightarrow A_1 \supseteq A_2$

其中, 下确界和上确界分别为:

$(X_1, A_1) \wedge (X_2, A_2) = (X_1, X_2, (A_1 \cup A_2)^*)$ ,  
 $(X_1, A_1) \vee (X_2, A_2) = ((X_1 \cup X_2)^*, A_1 \cap A_2)$ ,  
 则  $L(G, M, I)$  在如上定义的偏序关系“ $\leq$ ”下是一个完备格, 称为概念格. 表 1 是一个形式背景  $(G, M, I)$ , 其中, 对象集  $G = \{1, 2, 3, 4\}$ , 属性集  $M = \{a, b, c, d, e\}$ . 如果对象与属性有关, 在对象行与属性列的交叉处用 1 标记, 否则用 0 标记. 形式背景  $(G, M, I)$  的补背景  $(G, M, I^c)$  如表 2 所示

表 1 形式背景  $(G, M, I)$

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

表 2 补背景  $(G, M, I^c)$

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

根据概念格定义可得形式背景  $(G, M, I)$  及其补背景对应的概念格, 分别如图 1 和图 2 所示:

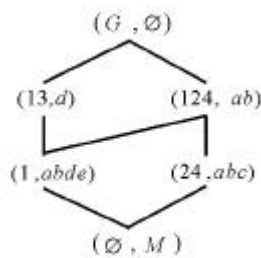


图 1 形式背景  $(G, M, I)$  的概念格

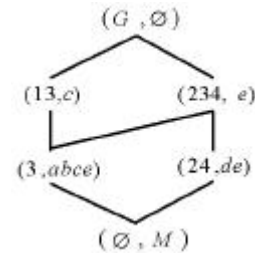


图 2 补背景  $(G, M, I^c)$  的概念格

## 2 三支概念分析理论

### 2.1 3WCA 基本概念

3WCA 的基本概念是三支概念以及三支概念格。按照从对象出发, 将属性集三分的理念, 可定义对象导出三支概念和对象导出三支概念格; 对偶地, 按照从属性出发, 将对象集三分的角度, 又可定义属性导出三支概念和属性导出三支概念格。为了定义“共同不具有”这个意义, Qi 等给出了负算子的概念, 而将 Wille 提出的经典算子相应的称为正算子。

针对一个形式背景  $(G, M, I)$ , 负算子  $*$  定义为: 对于任意的  $X \subseteq G, B \subseteq M$ ,

$$X^* = \{a \mid a \in M, \forall x \in X, xI^c a\}$$

$$B^* = \{x \mid x \in G, \forall a \in B, xI^c a\}$$

$X^*$  表示  $X$  中所有对象共同不具有的属性集合,  $B^*$  表示共同不具有  $B$  中所有属性的对象集合。如果一个二元对  $(X, B)$  是一个负概念。这恰与哲学中的“负概念”的语义一致<sup>[7]</sup>。而负概念也就是  $(G, M, I)$  补背景上的该概念。

结合正、负两种算子, 即可同时反映两种共性, 于是得到对象导出三支算子 (OE 算子) 与属性导出三支算子 (AE 算子)<sup>[3, 8]</sup>.

### 2.2 导出三支概念格

定义 4. 设  $(G, M, I)$  是一个形式背景。对于任意的对象子集  $A, B \subseteq M$ , 一对 OE 算子

$$(O, E1)^<: P(G) \rightarrow DP(M)$$

$$(O, E2)^>: DP(M) \rightarrow P(G)$$

定义如下:

$$X^< = (X^*, X^*), (A, B)^> = A^* \cap B^*$$

$X, (A, B)$  称为  $(G, M, I)$  的一个对象导出三支概念, 简称 OE 概念, 当且仅当  $X^< = (A, B)$  与  $(A, B)^> = X$  同时成立,  $X$  成为 OE 概念  $(X, (A, B))$  的外延,  $(A, B)$  称为内涵。

从 OE 概念出发, 我们可使所有 OE 概念组

成的集合  $OEL(G, M, I)$  形成一个格，称之为对象导出三支概念格(简称 OE 概念格)，其中的偏序关系为：

$$(X, (A, B)) \leq (Y, (C, D)) \Leftrightarrow X \subseteq Y \Leftrightarrow (C, D) \subseteq (A, B),$$

上，下确界为：对于任意的  $(X, (A, B)), (Y, (C, D)) \in OEL(G, M, I)$ ,

$$\begin{aligned} (X, (A, B)) \vee (Y, (C, D)) &= (X, Y)^{<>}, (A, B) \cap (C, D), \\ (X, (A, B)) \wedge (Y, (C, D)) &= (X \cap Y, ((A, B) \cup (C, D))^{><}) \end{aligned}$$

对偶的，我们也定义了一对 AE 算子，并形成 AE 概念以及 AE 概念格。

图 3 是例 1 的对象导出三支概念格我们以对象导出三支概念  $(13, (d, c))$  为例解释其语义<sup>[9]</sup>。从 3WD 角度：对对象子集  $\{1, 3\}$  出发，可以将属性全集三分为  $\{c\}$ ， $\{d\}$  与  $\{a, b, e\}$ ；从 FCA 角度：对象 1 与 3 共同具有的属性是  $d$ ，共同不具有的属性是

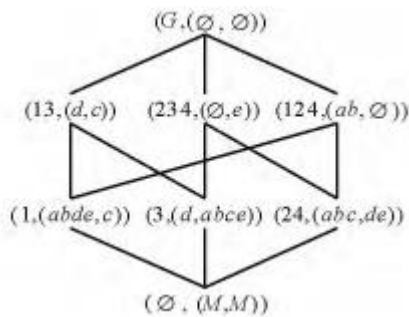


图 3 表 1 的对象导出三支概念格

$c$ ；共同具有  $d$ 、同时共同不具有  $c$  的对象也恰为 1 与 3；而  $\{c\} \cup \{d\}$  的补集中的属性则是对象 1 与 3 具有差异的属性。所以，该三支概念清晰、完整、准确的反映了一个对象子集的共性(包括正面和负面的两种共性)与差异性，以及对于属性集的三分。图 4 是例 1 的属性导出三支概念格，其中概念的解释与对象导出三支概念类似，只不过是属性角度出发，考虑一个属性子集被对象集共同具有以及共同不具有的情况<sup>[9]</sup>。

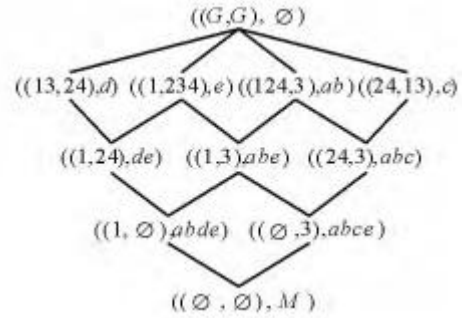


图 4 表 1 的属性导出三支概念格

从图 2 与 3 可以看出，这两个三支概念格明显比图 1 的经典概念格复杂，这是信息反映更为精细所付出的代价，自然也直接导致 3WCA 应用受限。

### 3 三支概念分析展望

#### 3.1 不完备背景的深入研究

相对于完备性的基础信息，不完备的情况在我们的生活和科研领域更为多见。利用部分已知概念这一特殊的三支概念，以及区间集所起的三支作用，可以先从理论上寻求表述和分析不完备背景的方法，探讨不同类型的部分已知概念族的结构，进而应用于实际问题当中<sup>[9]</sup>。

#### 3.2 决策形式背景的规则获取

FCA 中基于决策形式背景进行的规则提取，往往需要设定决策形式背景的协调性。类似地，对于 3WCA 而言，协调性的设定也是必须的，而且会更为复杂一些，但是其含有的信息也更丰富。由于牵涉到了决策形式背景的补背景及其上的三支规则，也就是负规则的引入，这使得基于决策形式背景的三支角度的规则获取变得复杂，而更有意义<sup>[9]</sup>。但是这部分工作还没有很好很深刻的成果。

#### 3.3 与三支决策理论的深入结合

三分而治再分析效用，是三支决策理论发展到目前为止较为完全的思想框架<sup>[10]</sup>，而三支概念分析仅利用 3WCA 中的导出算子考虑了三分，对于治略与效用尚未考虑。即使是三分，目前的研究也多是对完备背景的对象集和属性集进行精确的三分，而对于模糊背景或者不完备背景来说，三分方法的研究是一个非常有意义的问题，而且才刚刚起步。

另外，三分以后，处于边界域中的个体其信息是不够明确的，如何对它们进行更进一步的持续深入的分析，也是不容忽视的问题。Gao 等<sup>[11]</sup>在

对三支决策的可行策略进行研究的过程中,通过分析不同域中对象的相似性和差异性,提出了一种动态迁移方案,为边界域中对象给出了一种在不同域之间进行转移的方案。这一思想为三支概念中处于边界域的对象(属性)的进一步分析提供一个新的思路。

因此,这方面的研究工作还大有可为。

### 3.4 三支概念分析中的挑战性问题

#### 3.4.1 三支概念之间的相似性度量

三支概念之间的相似性度量是三支概念分析中的基本问题,这方面的研究将带来概念合并与格结构的进一步简化。但是,与经典概念之间的相似性度量<sup>[12]</sup>有所不同,三支概念一方面由于边界域带来的数据不一致性,使相似性度量方法的设计变得困难;另一方面像模糊值或缺失值环境下的不确定性因素,也会增加获取有效的相似性度量方法的难度<sup>[13]</sup>。

#### 3.4.2 多源数据的三支概念格构造

三支概念格目前主要针对单源数据进行构建,对于多源数据尚未见有关报道。类似地,多源数据的三支概念格一般也是基于信息融合思想实现的,但是它与普通的并行算法加速概念格构造有所不同,因为普通并行算法只加快建格速度,它并不影响最终计算结果。但是,基于信息融合的多源数据的三支概念格构造,它的最终结果会受信息融合函数选取的影响。也就是,选择不同的信息融合函数,得到的三支概念格结果可能是不同的<sup>[13]</sup>。另外,多源数据如果各个数据源之间存在属性重叠时,会给信息融合函数的设计带来一定的困难。当然,这些都是有待深入研究的细节问题。需要指出的是,它们有时还与问题背景有一定的关系,比如信息融合函数的权重选取通常与问题背景有关。

## 4 结语

文对概念格与三支决策相结合的研究历程、研究内容与研究展望作了扼要介绍,有关论述与讨论并未详细展开。

三支概念分析因为兼具三支决策理论与形式概念分析的优点,也因为其刻画语义的精确性与细致性,使得从认知角度获取新的研究增长点成为可能,从而得到了相关领域范围内学者,特别是国内学者的认可,使其在短短的五年内有了迅速的发展。需要指出的是,目前概念格与三支决策已开展的研究还不多,不管是三支概念分析,还是三支

概念学习,都处于初级研究阶段。比如,三支概念分析目前还没有完全摆脱 Wille 概念格的研究思路,三支概念学习也没有摆脱经典概念认知学习的研究方法,这两个主要方向都需要提出更多有价值且有难度的研究问题,特别是 Wille 概念格或经典概念认知学习没有触及到的问题,不断挖掘这类问题才能逐渐开辟一个新的有前景的交叉融合研究方向。

## 参 考 文 献

- [1] Ganter B, Wille R. Formal concept analysis: mathematical foundations[M]. Springer Science & Business Media, 2012.
- [2] Wille R. Restructuring lattice theory: an approach based on hierarchies of concepts[M]. Ordered sets, Springer, 1982, 445-470.
- [3] Qi J J, Wei L, Yao Y Y. Three-way formal concept analysis[J]. Lecture Notes in Computer Science, 2014, 8818: 732-741.
- [4] Yao Y Y. An outline of a theory of three-way decisions[C]. In: J. Yao, Y. Yang, R. Slowinski, S. Greco, H. Li, S. Mitra, L. Polkowski, editors. Rough Sets and Current Trends in Computing. Springer Berlin Heidelberg; volume 7413 of Lecture Notes in Computer Science; 2012, 1-17.
- [5] Wille R. Restructuring lattice theory: An approach based on hierarchies of concepts[M]. In: I. Rival, ed. Ordered Sets. Reidel: Dordrecht-Boston, 1982, 445-470.
- [6] Ganter B, Wille R. Formal concept analysis[M]. Mathematical Foundations, New York: Springer-verlag, 1999.
- [7] 金岳霖. 形式逻辑 [M]. 北京: 人民出版社, 1979.
- [8] WEI L, QIAN T. The three-way objectoriented concept lattice and the three-way property oriented concept lattice [C] // International Conference on Machine Learning & Cybernetics. 2015.
- [9] 魏玲, 高乐, 祁建军. 三支概念分析研究现状与展望[J]. 西北大学学报(自然科学版), 2019, 49(04): 527-537. DOI: 10.16152/j.cnki.xdxbzr.2019-04-005.
- [10] YAO Y Y. Three-way decision and granular computing [J]. International Journal of Approximate Reasoning. 2018, 103: 107-123.
- [11] GAO C, YAO Y Y. Actionable strategies in three-way decisions [J]. Knowledge-Based Systems, 2017, 133: 141-155.
- [12] WANG L D, LIU X D. A new model of evaluating concept similarity [J]. Knowledge-Based Systems, 2008, 21: 842-846.
- [13] 李金海, 邓硕. 概念格与三支决策及其研究展望[J]. 西北大学学报(自然科学版), 2017, 47(03): 321-329. DOI: 10.16152/j.cnki.xdxbzr.2017-03-002.