《智能信息处理》课程作业

基于形式概念分析的模糊三支半研究

曹凤龙

作业	分数
得分	

2021年 11月 29 日

基于形式概念分析的模糊三支半研究

曹凤龙

（大连海事大学 信息科学技术学院 辽宁 大连 116026）

摘 要：形式概念分析是由德国的Wille教授于20 世纪80年代初提出的，其核心数据结构概念格，也称 Galois 格，准确而简洁地描述了概念之间的层次关系，已成为一种重要的知识表示方法。概念格通过Hasse 图生动和简洁地体现了这些概念之间的泛化和特化关系。形式概念分析以数学为基础而建立，对本体中的概念、属性以及相互关系等用形式化的语言表述，然后根据语境构造出概念格，清晰的表达出本体的结构，目前形式概念分析已成为进行数据分析和规则提取的强有力的工具。本文主要介绍了形式概念分析的主要定义和相关概念以及模糊三支半的概念和定义，从格论角度出发讨论其偏序关系以及上下确界，并进行相关定理的证明。

关键词：形式概念分析；形式背景；概念格；模糊三支半；

Fuzzy three and a half studies based on formal conceptual analysis

CaoFengLong

(Dalian Maritime University,Computer Science and Technology,Liaoning,Dalian,116026,China)

Abstract: Form conceptual analysis was proposed by Professor Wille in Germany in the early 1980s, and its core data structure conceptual lattice, also known as the Galois lattice, accurately and succinctly describes the hierarchical relationship between concepts and has become an important method of knowledge representation.The conceptual lattice vividly and succinctly embodies the generalization and specialization relationship between these concepts through the Hasse graph.formal concept analysis is based on mathematics, and the formal language expresses the concepts, attributes and mutual relations in the ontology, and then constructs the concept lattice according to the context to clearly express the structure of the ontology. At present, formal concept analysis has become a powerful tool for data analysis and rule extraction.This paper mainly introduces the main definitions and related concepts of formal concept analysis and fuzzy three and a half concepts, discuss its partial order relationship and upper and lower confirmation bounds from the perspective of lattice theory, and make the proof of the correlation theorem.

Keywords: Form conceptual analysis; Formal background; Conceptual lattice; Fuzzy three and a half;

0 引言

在哲学中，概念被理解为由外延和内涵两个部分所组成的思想单元。基于形式概念分析的粗糙集模型、基于概念格的多示例集成学习模型、基于概念格的不同粒度下的领域本体模型及形式概念分析在不同粒度下知识获取模型，这些模型不仅在理论上拓展形式概念分析方法，而且对形式概念分析的应用起到积极的推动作用。

三支概念分析是三支决策理论与形式概念分析相结合产生的一种用于知识发现的理论．对三支半概念分析提出了一个统一的描述。基于三支决策中思想可知，三支概念分析要表达的是共同拥有的特征和共同不拥有的特征，每个三支概念的外延或者内涵是由正域和负域两部分构成．三支概念分析对于数据处理的优越性，使得这一理论得到了越来越多的关注，并且在人工智能、医疗决策、数据挖掘等领域中得到了广泛应用．

本文首先介绍了形式概念分析的基本概念，又介绍了形式背景，以及概念格的运算规则。最后介绍模糊三支半的定义和概念。为后续研究提供了思路。

1 基本概念

形式概念分析理论的主要思想源于哲学中对概念的定义。在哲学体系中概念是由外延和内涵两个部分组成的思想单元，外延被定义为属于这个概念的所有对象的集合，而内涵被定义为属于这个概念的所有对象所共同具有的属性的集合。概念格是形式概念分析的核心数据机构，其本质上描述了对象和特征之间的联系。一般认为外延是概念覆盖的实例，而内涵则是对于概念的描述，概念进一步可以通过Hasse图来实现可视化，通常Hasse 图的每一个节点就代表一个概念。形式概念分析提供了一种较好的层次化（形式）对象的分析方法，它能够识别那些具有共同（形式）属性的一组（形式）对象的组合。在应用形式概念分析方法的过程中，线路图的制定是非常重要的一个环节，其本身也是对于概念化的图形化表示。通过线路图能够对语境中所包含的对象和属性关系进行展示，在一些特定的语境下还包含有继承以及发展的关系，因此说形式概念分析其本质是一种准确性高以及使用范围广泛的分析模式。

1.1 形式概念定义

定义：对于形式背景 K，在 G 的幂集和 M 的幂集之间可以定义两个映射日 f 和 g 如下：

- ∀O⊆G: f(O)={d|∀x∈ O:(xId)}

- ∀D⊆M: g(D)={x|∀d∈ D:(xId)}

来自 P(G)×P（M）的二元组(O，D)如果满足两个条件：O=g(D)及 D=f(O)，则它被称为是形式背景K的一个形式概念，简称概念，记为 C=(O,D)，其中 D 和 O 分别被称为概念 C 的内涵和外延。K 的所有形式概念的集合被标记为 CS(K)。

我们可以把形式概念理解成为数学上的概念，因为形式概念等于对象集的属性集，其中对象集和属性集都是在数学上成立的。与概念的表示方法类似，形式概念也有三种表示方法，分别是表达式法，二维表法和图示法。形式概念的作用就是构建自然 概念的层次连通结构，为了更好的解释此作用，根据上述描述的内容，在自然概念的基础之上可以建立一个新的形式概念内容，从而对现实世界有一个好的理解。

1.2 形式背景

定义：一个形式背景K=(G,M,I)由集合 G、M以及它们之间的关系组成I, G的元素称为对象(Objects), M的元素称为属性(Attributes)。为了表示一个对象O和一个属性M在关系 I 中，可以写成 oIm 或(o,m)∈I，读成“对象o有属性m”。

1.3 概念格

概念格(concept lattice)是 FCA 的核心数据结构。概念格的每个节点是一个概念，每个概念由外延（Extent）和内涵(Intent)两部分组成。外延是概念所覆盖的实例，而内涵是概念的描述，是该概念所覆盖实例的共同特征。概念格可以通过其Hasse图生动简洁地体现概念之间的泛化和例化关系。因此，概念格被认为是一种支持数据 分析的有效工具。

	a	b	c	d
d1	1	1	1	1
d2	1	1	0	0
d3	0	1	1	1
d4	0	1	0	0
d5	0	1	1	0

表1 形式背景示例

所有的概念并构造出格结构以此刻画出数据集中对象与属性之间的关系。构造概念格是概念格应用的前提，但构造概念格已被证明是NP 问题；

因此，人们在构造概念格之前希望在保持格结构不变的情况下，尽可能的简化数据。目前，概念格约简的研究包括对象约简、属性约简、纵横向维护和内涵约简等。如图 1 所示为概念格示意图。

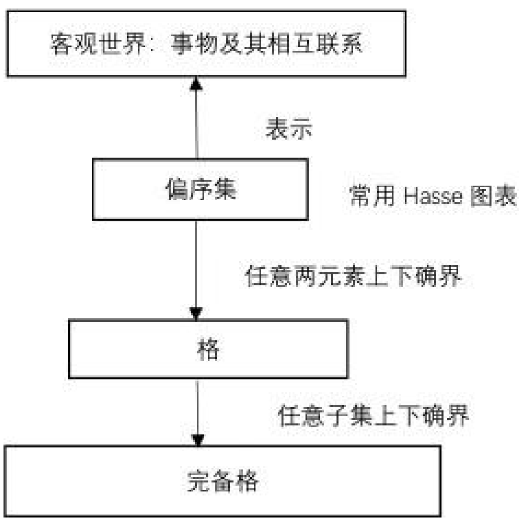

图1 概念格示意图

概念格理论的研究主要集中在以下几个方面：

(1) 概念格的建造

从数据集（在概念格中称为形式背景）中生成概念格的过程实质是一种概念聚类过程。对于同一批数据，所生成的格是唯一的。

(2) 概念格的约简

概念格的约简能够有效地提高概念格的维护效率。使形式背景中所蕴含的知识易于发现，简化知

识的表示方式。约简概念格实际上是在保持对象集不变的条件下，如何求得最小的属性集的过程。国内的研究主要是以张文修等提出的理论为基础。给出概念格属性约简的判定定理，引入形式背景的可辨识属性矩阵。并依此为基础求得属性约简的方法。

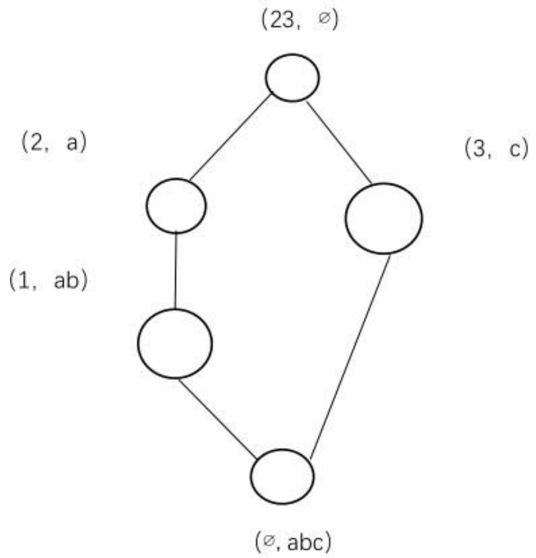

图2 概念格示意图

（3）概念格的运算

定义 ：如果形式背景 $K_1=(U_1, A_1, I_1)$和$K_2=(U_2, A_2, I_2)$满足 $U_1 \subseteq U$，$U_2 \subseteq U$，$A_1 \subseteq A$，$A_2 \subseteq A$，则称K_1 和 K_2 是同域形式背景，$L(K_1)$和$L(K_2)$是同域概念格，如果$U_2 \cap U_2=\varnothing$，则称K_2 和 K_2、$L(K_2)$和 $L(K_2)$分别是外延独立的，简称独立的。

定义 ： 如果 $K_1=(U_1, A_1, I_1)$和 $K_2=(U_2, A_2, I_2)$是同域且独立的，则 $K_1+K_2=(U_1 \cup U_2, A_1 \cup A_2, I_1 \cup I_2)$。

定义 ： 对于 $C_1=(O_1, D_1)$和 $C_2=(O_2, D_2)$，如果$D_1=D_2$，则称 C_1 内涵等于C_2，简称$C_1=C_2$。

定义 ： 对于 $C_1=(O_1, D_1)$和 $C_2=(O_2, D_2)$，如果 $D_1 \subset D_2$，则称 C_1 内涵小于 C_2，简称 C_1 大于 C_2，或称C_2 小于 C_1。

定义 ： 对于 $C_1=(O_1, D_1)$，$C_2=(O_2, D_2)$和$C_3=(O_3, D_3)$，定义 C_1+C_2 等于 C_3，如果 $O_3=O_1 \cap O_2$，$D_3=D_1 \cap D_2$。

2 模糊三支半概念

2.1模糊三支半算子

定义 ：设 $K\sim=(G, M, I\sim)$为一个模糊形式背景，给定阈值α，$X \subseteq G$，$A \subseteq M$.

定义在X上的模糊三支半算子$X\rightarrow=(X^{-*}, X^{-*\bullet})$，式中：$X^{-*}=\{(a, \mu(X, a))|a \in M, \forall x \in X, \mu(x, a) \geq \alpha\}$； $X^{-*\bullet}=\{(a, \mu(X, a))|a \in M, \forall x \in X, \mu(x, a)<\alpha\}$. 结合经典半概念与经典形式概念的相关性可知：半概念中定义的算子与形式概念中的算子是一致的，这恰好验证了每一个经典形式概念都是经典半概念，但反之不成立的结论．本研究定义的模糊三支半算子也是在模糊三支算子定义中考虑一半.

2.2 三支半概念定义

定义 ：设 $K^{\sim}=(G, M, I^{\sim})$为一个模糊形式背景，给定阈值$\alpha$，$X \subseteq G$，$A, B \subseteq M$，$\phi(A)$ 为 A 上的模糊 集，若 $X\rightarrow=(\phi(A), \phi(B))$，则称（$X$，($\phi(A)$， $\phi(B)$)）为面向对象的模糊三支半概念，简称模糊 OE-半概念。对比模糊 OE-概念和模糊OE-半概念，可以发 现模糊 OE-概念一定是模糊OE-半概念，反之则不成立这也验证了经典形式概念分析与经典半概念之间的规律。

2.3 模糊OE-半概念的性质

定义 ：设 $K^{\sim}=(G, M, I^{\sim})$为一个模糊形式背景，$OEFSL(G, M, I^{\sim})$为在 $K^{\sim}=(G, M, I^{\sim})$下生成的所有 模糊OE-半概念的集合。对于任意的$(X, (\phi(A), \phi(B)))$， $(Y, (\phi(C), \phi(D))) \in OEFSL(G, M, I^{\sim})$，定义二者的偏序关系为 $(X, (\phi(A), \phi(B))) \leqslant (Y, (\phi(C), \phi(D))) \Leftrightarrow X \subseteq Y, (\phi(C), \phi(D)) \subseteq (\phi(A), \phi(B))$，

式中：$(X, (\phi(A), \phi(B)))$为$(Y(\phi(C)\phi(D)))$的子概念；$(Y(\phi(C)\phi(D)))$为$(X(\phi(A), \phi(B)))$的超概念。定理 1 设 $K^{\sim}=(G, M, I^{\sim})$为一个模糊形式背景，$OEFSL(G, M, I^{\sim})$为在$K^{\sim}=(G, M, I^{\sim})$下生成的所有模糊OE-半概念的集合。对于任意的$(X, (\phi(A)\phi(B)))(Y(\phi(C)\phi(D))) \in OEFSL(G, M, I^{\sim})$ 上确界和下确界分别为

$(X(\phi(A)\phi(B))) \wedge (Y(\phi(C)\phi(D)))=$
$(X \cap Y((X \cap Y)^{-*}(X \cap Y)^{-*\bullet}))$；
$(X(\phi(A)\phi(B))) \vee (Y(\phi(C)\phi(D)))=$
$(X \cup Y((X \cup Y)^{-*}(X \cup Y)^{-*\bullet}))$

$OEFSL(G, M, I^{\sim})$在上述定义中给出的偏序关系下是一个完备格，称为模糊OE-半概念格。

3 总结

在实际数据处理和信息提取中，面对的信息背景并非是经典的 0-1背景，也存在只须单向讨论共同拥有和共同不拥有的属性集情况．提出面向对象的模糊三支半概念。该概念能够有效处理实际应用中模糊形式背景下的数据，单向考虑更符合实际意义。与模糊三支概念分析相比，面向对象的模糊三

支半概念的求解更简单，计算量小事实上，
面向对象的模糊三支半概念在数据挖掘、人工智能等很多领域都具有广泛的应用。面向属性的模糊三支半概念也可以得到相应结论.

4 参考文献

[1] 汪文威， 祁建军. 三支概念的构建算法. 西安电子科技大学学报， 2017, 44 (1) :71-76.

[2]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.

[3]胡可云， 陆玉昌， 石纯一． 概念格及其应用进展[J]． 清华大学学报:

[4]李金海, 李玉斐, 米允龙, 吴伟志. 多粒度形式概念分析的介粒度标记方法[J]. 计算机研究与发展,2020,57(02):447-458.

[5]毕强，滕广青. 国外形式概念分析与概念格理论应用研究的前沿进展及热点分析[J]． 现代图书情报技术, 2010,(11):17-23.

[6]张文修，魏玲，祁建军. 概念格的属性约简理论与方法 [M]. 中国科学 E 辑：信息科学，2005

[7] VOMBROCK B，WILLE R． Semiconcept and protoconcept algebras：the basic theorems[C]// Proc of the Third International Conference on Formal Concept Analysis． Berlin：Springer-Verlag，2005：34-48．

[8] VOMBROCK B．Complete subalgebras of semiconcept algebras and protoconcept algebras[C]// Proc of the Third International Conference on Formal Concept Analysis． Berlin：Springer-Verlag，2005：329-343．

[9] WANG G，MAO H． Approximation operators based on preconcepts[J]．Open Mathematics，2020，18(1)：400- 416．

[10] JULIA K． Semiconcept graphs with variables[J]．Con- ceptual Structures： Integration and Interfaces， 2002， 2393：369-381．

[11] MAO H． Approximation operators for semiconcept[J]． Journal of Intelligence and Fuzzy Systems，2019，36(4)： 333-343．