

基于概念格的知识推理研究分析

周子涵

(大连海事大学 信息科学技术学院 大连 181272)

摘 要: 本文主要讨论了形式概念、形式背景的相关内容;对于概念格的构建和简约进行了简单说明;最后,根据一个球类运动的实际情况建立一个形式背景和概念格,讨论了采用概念格的方式进行信息知识处理的优点。

关键词: 形式概念;形式背景;概念格;约简;知识推理

Research and Analysis of Knowledge Reasoning Based on Concept Lattice

Zihanzhou

(School of Information Science and Technology, Dalian Maritime University, Dalian 181272)

Abstract In this paper, we mainly discuss the related concepts of formal concept, formal context and briefly describe the construction and simplification of concept lattice. Finally, we construct a formal context and concept lattice according to the actual situation of a ball game, and discuss the advantages such information knowledge processing.

Key words formal concept; formal context; concept lattice; reduction; knowledge reasoning

1 形式概念

1.1 概念定义

形式概念分析(Formal Concept Analysis, FCA)是一种从形式背景进行数据分析和规则提取的强有力工具,形式概念分析建立在数学基础之上,对组成本体的概念、属性以及关系等用形式化的语境表述出来,然后根据语境,构造出概念格(concept lattice),即本体,从而清楚地表达出本体的结构。这种本体构建的过程是半自动化的,在概念的形成阶段,需要领域专家的参与,识别出领域内的对象、属性,构建其间的关系,在概念生成之后,可以构造语境,然后利用概念格的生成算法 CLCA,自动产生本体。

概念格是 FCA 的核心数据结构。概念格的每

个节点是一个概念,由外延和内涵组成。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。概念格结构模型是形式概念分析理论中的核心数据结构。其本质上描述了对对象和特征之间的联系,表明了概念之间的泛化与例化关系。这种概念格构建的过程是半自动化的。

对于人类可以认知的事物我们都可以建立一个对应的对象和属性集合,这样的展现就是概念。关于概念的表达方法目前有三种形式:表达式法、二维表法、图示法,具体表现形式如图 1 所示。

①表达式法 概念 = (对象集, 属性集)

②二维表法

	属性1	属性2	...	属性m
对象1				
...				
对象n				

③图示法

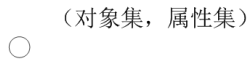


图 1 概念的三种表示

对于第二种方法, 每一行就是一个概念, 多个概念的结合就是一个背景, 关于背景的描述在下文有说明。

1.2 自然概念定义

自然概念就是人们头脑中认为的自然事物然后用概念的形式表现出来。这部分的自然概念只是概念的一部分是其子集, 不是一个抽象的概念而是一个具体的事物展现。表现形式和概念的表现方式一样如图 2 所示。

①表达式法 电脑 = ({电脑}, {屏幕, 键盘})
电视 = ({电视}, {屏幕, 遥控器})

②二维表法

	屏幕	键盘	遥控器
电脑	√	√	
电视	√		√

③图示法 ({电脑}, {屏幕, 键盘}) ({电视}, {屏幕, 遥控器})

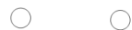


图 2 自然概念的三种表示

1.3 形式概念定义

形式概念 (Formal Concept Analysis, FCA) 由 R.Wille 于 1982 年提出是应用数学的一个分支、是信息处理的一种理论、是知识处理的一种理论。形式概念通俗的语言描述是: 形式对象集和形式属性集的总和, 定义如下。

定义 2^[2]. 设 A 是对象集合 G 的一个子集, 定义 $f(A) := \{m \in M | \forall g \in A, gIm\}$ (A 中对象共同属性的集合)。 gIm 表示 $g \in G$ 与 $m \in M$ 之间存在关系 I , 读作: 对象 g 具有属性 M 。

相应地, 设 B 是属性集合 M 的一个子集, 我们定义 $g(A) := \{g \in G | \forall m \in B, gIm\}$ (具有 B 中

所有属性的对象的集合)。

定义 3^[2]. 形式背景 (G, M, I) 上的一个形式概念 (Formal Concept) 是二元组 (A, B) , 其中 $A \subseteq G, B \subseteq M$, 且满足 $f(A) = B$ 和 $g(B) = A$ 。我们称 A 是形式概念 (A, B) 的外延, B 是形式概念 (A, B) 的内涵。

$\xi(G, M, I)$ 表示形式背景 (G, M, I) 的所有形式概念的集合。

相应的概念三种表示方法, 如图 3 所示。

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, 从而可以获得一些新的结论。如图 4 所示。

①表达式法 形式概念 = (形式对象集, 形式属性集)

②二维表法

	形式属性1	...	形式属性2
形式对象1			
...			
形式对象n			

(形式对象集, 形式属性集)

③图示法



图 3 形式概念的三种表示

	屏幕	键盘	遥控器
电脑, 电视			
电脑	√	√	
电视	√		√
\emptyset	√	√	√

图 4 形式概念作用距离

图 4 的表现形式, 我们可以理解为: 如果某个事物具有电脑和电视的两种性, 也许就会有不一样的好处。例如, 目前的电视主要使用有线信号进行节目播放。如果一个电视机拥有两个系统, 一个系统是 windows 操作系统; 另一个系统就是普通的电视机系统, 这样就可以实现电脑电视的结合。通过切换信号源就可以进行系统的改变, 这样的产品既可以满足家中老年人对于简单的电视节目需求、也可以满足年轻人对智能设备的需求, 一个简单的设备满足了不同的客户群体。事实证明, 智能电视正在向这样的方向发展, 一种是电视系统; 一种是安

卓系统。虽然目前还不是 windows 操作系统, 这样的发展趋势已经展露。

1.4 形式背景

在这个客观世界中, 单独的事物并不能够描述一个具体的系统。同样地, 单独的形式概念并不是以描述一个形式概念集, 于是我们引入形式背景来解释这个问题。

定义 4^[2]. 一个形式背景 (Formal Context) 是一个三元组 $K = (G, M, I)$, 其中 G 是对象的集合, M 是属性的集合, I 是 G 和 M 之间的一个二元关系, 即, $I \subseteq G \times M$ 。

根据定义 4, 一个形式背景能够用一个矩形表来表示, 表的每一行是一个对象, 每一列是一个属性。若 g 行 m 列的交叉处是 X , 则表示对象 g 具有属性 m , 如表 1 所示。

表 1 定义 4 的表格显示

	属性 1	属性 2	属性 n
对象 1	X			
对象 2	X	X		
.....				
对象 m				X

形式背景就是信息世界, 相当于我们的客观世界。对于形式背景而言, 把形式概念置于其中的二维表就是形式背景, 如图 5 所示。

在一个复杂的形式概念集进行形式背景生成的时候会产生冗余信息。这样的冗余信息可能发生在不同的对象拥有相同的属性, 我们把这样的冗余问题进行约简, 此类问题称为“约简行”; 有时候不同的事物虽然有不同的属性, 但是属性和属性之间可能存在异构现象, 我们把这样的约简问题称为“约简列”; 有时候多个属性之间能相互关联, 某个属相可能由一个或者多个其他相关属性推出, 这样的约简问题我们称为“关联规则抽取”通过行和列的约简我们就可以进行信息的合并, Jiang 等人提出利用基于输入和输出参数的 Web 服务集的扩展概念格模型, 来有效地组织服务, 为 Web 服务的选择、优化和组合提供智能支持^[4]

2 概念格

2.1 概念格的

概念格, 又称为 Galois 格, 是 Wille 于 1982 年首先提出来的一种知识表示和处理模型。借助概

念格理论可以对具有对象和属性的数据库进行分析, 并通过概念的外延与内涵进行规则提取、决策优化等。目前, 概念格已被广泛应用于机器学习、模式识别、专家系统、数据挖掘、信息检索等领域。^[3]

概念格理论的主要思想是在形式背景中寻找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。^[3]

在形式背景部分我们已经提及过关于形式背景的约简, 其中包括: 约简行、约简列、关联规则抽取等方法。在概念格中, 进行的约简规则和形式背景一样, 方法类似、原理相同, 详细方法请查阅相关文献。之前我们提到过形式概念以二元表的形式可以组成形式背景, 也说过概念有三种表示方式: 表达式法, 二维表法, 图示法。所以我们可以使用图示法把形式概念进行表示, 这样我们就可以构建一个哈斯图, 如果哈斯图是完备格, 那么这个哈斯图就是概念格, 如图 6 所示。

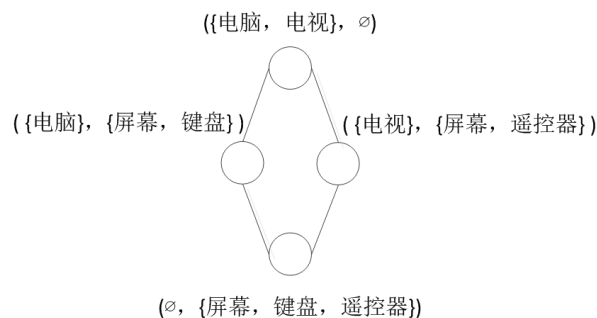


图 6 概念格的表示

3 概念格的应用

3.1 设计形式概念

概念格的构建以形式背景为基础, 形式背景由多个形式概念构成, 形式概念是一个由形式对象集、形式属性集、形式对象和形式属性之间关系构成的三元组。所以, 我们先假定实验需要的形式概念集。

在实验中, 我们获取的对象包括 5 个, 即足球、篮球、排球、网球、羽毛球。根据标准的球类运动规则, 这些球类运动具有如下的“对象-属性”关系,

如图 7 所示。

足球 = ({足球}, {非回合制, 身体碰撞, 四人之上})
篮球 = ({篮球}, {非回合制, 身体碰撞, 四人之上, 室内比赛, 手控球})
排球 = ({排球}, {回合制, 非身体碰撞, 四人之上, 室内比赛, 手控球})
网球 = ({网球}, {回合制, 非身体碰撞, 手控球, 辅助机械})
羽毛球 = ({羽毛球}, {回合制, 非身体碰撞, 手控球, 辅助机械})

图 7 对象-属性二元关系

3.2 形式背景的构建

对于上面的对象进行 A~E 的排序, 非回合制为 a, 回合制为 b, 非身体碰撞为 c, 身体碰撞为 d, 四人之上为 e, 手控球为 f, 室内比赛为 g, 辅助机械为 h, 根据此符号表示, 可列写出形式背景如表 2 所示。

3.3 形式背景的约简

根据表格信息, 我们发现可以进行约简的行和列信息, 也就是可以对对象和属性进行约简, 整理结果如表 3 所示。

表 2 形式背景

	a	b	c	d	e	f	g	h
A	1			1	1			
B	1			1	1	1	1	
C		1	1		1	1	1	
D		1	1			1		1
E		1	1			1		1

表 3 形式背景的约简

	a,d	b,c	e	f	g	h
A	1		1			
B	1		1	1	1	
C		1	1	1	1	
D,E		1		1		1

根据属性个数重新排序, 如表 4 所示。

表 4 按照属性个数排序的形式背景

	a,d	b,c	e	f	g	h
A	1		1			
D,E		1		1		1
B	1		1	1	1	
C		1	1	1	1	

3.3 概念格的构建

根据上述的形式背景生成概念格, 如图 8 所示。在生成概念格的过程中要保证概念格是一个完备格, 也就是要保证任何子集的上确界和下确界都存在的偏序集。所以在构建概念格的过程中会存在其他结点。

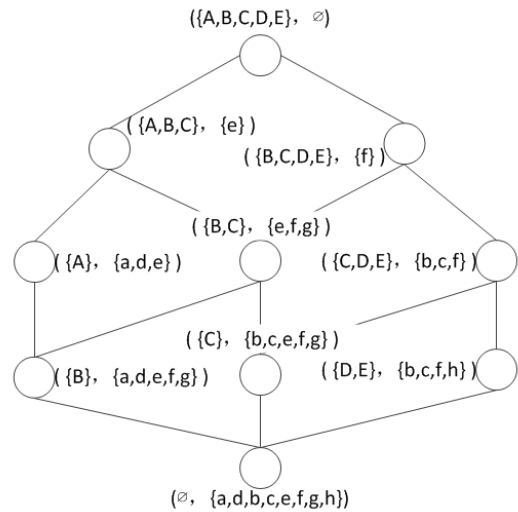


图 8 球类运动的概念格

3.4 结论分析

根据以上举例我们可以发现, 不同种类的球类运动可能具有相同的球类属性, 例如网球和羽毛球, 于是两个球类运动姑且可以认为是一种类型的; 同时, 不同的球类运动可能具有相同结果的属性和属性值, 例如回合制和非身体碰撞、非回合制和身体碰撞, 不难理解, 回合制比赛中双方一般在场地两侧互不直接接触, 所以也可以认为是属性的关联规则关系。

4 总结

对于一个系统我们可以找到此系统中的所有对象的论域, 进而根据这个论域建立起对应的背景, 此背景包括对象和属性, 这就是我们常识信息的内容。为了发现和挖掘新知识, 就需要进行扩展, 形式背景就是把对象重新排列组合, 形成新的对象对对应的新属性。

在使用形式概念分析过程中, 我们可以利用已知的知识或者常识, 经过形式化以后得到新的知识和常识, 从而帮助我们更好的理解研究的系统发现潜在的知识。

参 考 文 献

- [1] 概念, <http://dwz.cn/4nJWbo> 2016,10,17
- [2] 晏力, 刘鹏慧. 基于形式概念分析的属性约简[J]. 西华大学学报(自然科学版), 2012, 04: 37-41
- [3] 张云中. 基于形式概念分析的领域本体构建方法研究[D]. 吉林大学, 吉林, 2009