

《智能信息处理》课程作业

形式概念分析基本概念及应用

王煜坤

作业	分数[20]
得分	

2020 年 11 月

形式概念分析基本概念及应用

王煜坤

(大连海事大学 信息科学技术学院, 大连 116026)

摘要 形式概念分析, 也称为概念格, 又叫做 Cralois 格, 由 Wille R 于 1982 年首先提出, 它提供了一种支持数据分析的有效工具。本文介绍了有关形式概念分析的基本概念, 包括形式概念、形式背景、概念格及概念格的生成方法。接着列举了一些近年来形式概念分析的相关应用, 并对这些应用进行了简要的介绍。

关键词 形式概念; 形式背景; 概念格

Basic concepts and applications of formal concept analysis

WANG Yu-kun

(School of Information Science and Technology, Dalian Maritime University, Dalian 116026)

Abstract Formal concept analysis, also known as concept lattice or Cralois lattice, was first proposed by Wille R in 1982. It provides an effective tool to support data analysis. This paper introduces the basic concepts of formal concept analysis, including formal concept, formal background, concept lattice and the generation method of concept lattice. Then it lists some applications of formal concept analysis in recent years, and briefly introduces these applications.

Key words formal concept; formal context; concept lattice

1 引言

形式概念分析最早由德国教授 Wille R 提出, 以形式背景和形式概念为基础, 开创了概念格理论与应用研究的先河。随后形式概念分析作为一种数据和知识处理的有效工具, 已被广泛应用于机器学习、模式识别、数据挖掘、计算机推理、知识发现等领域。

形式概念分析的使用就是把人类已经知道的知识或者常识通过此技术手段发现和挖掘新知识, 从而对原先已经认识的世界知识进行利用然后改造世界。由此可见, 形式概念分析的引入, 对知识的整合作出了巨大的贡献。本文先介绍了形式概念形式概念、形式背景、概念格的定义, 以及概念格的生成方式, 然后列举了几个形式概念分析的应用。从本文中, 可以感受到概念格在知识分析过程中的突出表现。^[1]

2 形式概念

2.1 概念

定义 1. 概念 (Idea; Notion; Concept) 是反应对象特有属性的思维方式, 是从对象的属性中抽出特有属性概括而成的。它的表达语言形式是词或词组。概念包括内涵和外延, 外延由内涵所决定。

心理学上认为, 概念是人脑对客观事物本质的反映, 这种反映是以词来标示和记载的。概念是思维活动的结果和产物, 同时又是思维活动借以进行的单元。概念随着社会历史和人类认识的发展而变化。^[2]

关于概念的表示方法目前有三种形式: 表达式法、二维表法、图示法, 具体表现形式如表 1 所示。

表 1 概念的三种表示

概念的表示方法	具体表示	
表达式法	概念= (对象集, 属性集)	
二维表法		属性 1 ... 属性 n
	对象 1	
	...	
	对象 n	
图示法	○ (对象集, 属性集)	

2.2 自然概念

自然概念就是人们把认为很自然的事物用概念的形式表达出来。例如，人，狗等。自然概念包含于下述的形式概念，形式概念是一个更大的概念集合。

2.3 形式概念

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. 定义 4 (形式概念) 形式概念 C 是一个二元组 (X, Y) ，其中 $X \subseteq O, Y \subseteq A$ ， O 是对象集， A 是属性集，满足以下两点要求：1) $Y = f(X)$ ；2) $X = g(Y)$ 。 X 为概念的外延，记为 $Ex(C)$ ， Y 为概念的内涵，记为 $In(C)$ 。

形式概念的作用就是构建自然概念的层次连通结构。为了解释自然概念和形式概念的区别，这里给出如下例子进行对比。如表 2, 3 所示。

表 2 自然概念

	书写	擦除
铅笔	√	
橡皮		√

表 3 形式概念

	书写	擦除
铅笔, 橡皮		
铅笔	√	
橡皮		√
∅	√	√

从以上两图表中可以看出，形式概念存在一些自然概念中不存在的情况，自然概念是形式概念的子集。且形式概念和自然概念所对应的对象集元素数量有以下对应关系：

$$N=2^n \quad (2.3)$$

其中 N 是形式概念元素数目， n 是自然概念元素数目。

3 形式背景

3.1 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足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^[3]. (形式背景) 形式背景 K 是一个三元组 (O, A, I) ，是形式概念分析的输入。其中， O 是对象集， A 是属性集， $I \subseteq O \times A$ 是 O 和 A 之间的二元关系，对于一个对象 $o \in O$ ，属性 $a \in A$ ，那么 oIa 就表示对象 o 具有属性 a ， $\forall o \in O, a \in A$ ，有 $oIa \rightarrow (o, a) \in I$ 。

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 4 形式背景定义

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

形式背景区别于形式概念，它存在其特定的背景。可以理解为将形式概念至于特定背景的二维表。

3.2 形式背景上的计算

形式背景上的计算包括：对象/属性约简，关联规则抽取。下面分别举例介绍这两种计算。

3.2.1 对象/属性约简

由于形式概念有这样一条性质：内涵决定外延，即属性决定概念。因此，我们可以将属性相同的概念进行合并，即属性约简。具体过程如下，表 5 经过属性约简后可以转变为表 6。

表 5 约简前的形式背景

	设备	计算机	复印机	打印机	专用装置
咨询	√	√	√	√	√
方案			√	√	
安装	√	√	√	√	√
指导		√	√	√	√

表 6 约简后的形式背景

	设备	计算机	复印机	打印机	专用装置
咨询/	√	√	√	√	√
安装					
方案			√	√	
指导		√	√	√	√

3.2.2 关联规则抽取

观察形式背景的属性，我们可以发现，当一些属性出现时，另一些属性一定会出现。我们将其分别定义为条件属性和决策属性。例如下表 7 中，我们可以看出：如果 c 属性存在，C 属性页一定存在；如果 d 属性存在，D 属性页一定存在

表 7 关联规则抽取

	a	b	c	d	e	f	C	D
1	0	1	1	0	1	1	1	0
2	0	0	1	1	0	0	1	1
3	1	0	0	1	0	1	0	1
4	1	1	1	1	1	0	1	1

4 概念格及概念格的生成

4.1 概念格

在了解概念格之前，先要明确格和概念格的定义。

定义 5 格是任两个元素的上确界和下确界都存在的偏序集

定义 6 完备格是任何子集的上确界和下确界都存在的偏序集

有了这两个概念，概念格的定义就可以作出清晰的阐述。

定义 7 概念格是元素为概念的完备格

概念格的表示举例，如下图 1

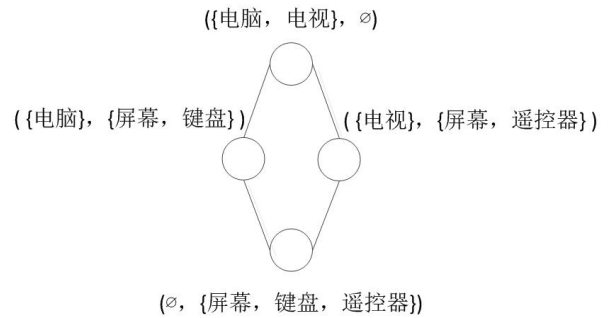


图 1 概念格的表示

4.2 概念格的生成

概念格的生成包括如下图 2 中的七个步骤

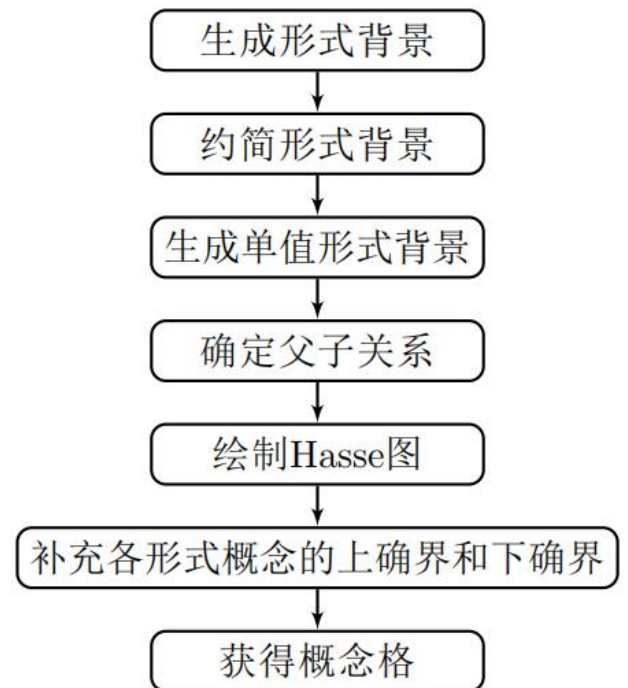


图 2 概念格生成步骤

5 概念格的应用

5.1 学习图像特征

论文^[4]提出了一种新的方法，可以从多变量数据中学习图的表示。在文章中，每个节点代表一组数据点，边表示簇之间的子集-超集关系，可以是相互重叠的。其方法的关键是使用形式化概念分析，以基于代数闭性。实验证明，该方法可以有效地与基线方法相比，提取聚类的层次结构。

5.2 公路交通气象服务形式背景分析

公路交通出行受到多个基础气象要素的影响和制约,研究不同气象要素对公路交通出行的综合影响,对更好地开展气象服务意义重大。国家级气象服务部门综合研究分析天气、能见度、风力等多个高影响天气要素对公路交通出行的影响,建立了公路通行的气象条件模型。基于此模型,结合形式概念分析模型设计,建立了公路交通气象形式概念分析形式背景,揭示气象要素与公路通行条件的相关性,为公路交通气象服务知识库建模和应用提供知识基础。[5]

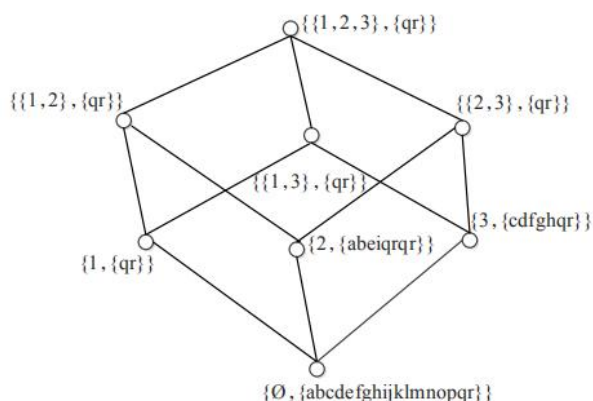


图3 公路通行气象条件模型概念格

根据形式背景分析后形成的概念格(如图3),表示了公路交通气象服务中不同公路通行条件与高影响气象要素之间的关联关系,实现了基于形式概念分析的公路交通气象服务知识表示。在分析过程中,使用一阶谓词逻辑表达概念内涵,使用合取范式表示概念外延,为气象服务领域知识获取和本体知识库构建开展了有益的探讨。

5.3 形式概念分析在软件工程中的应用

5.3.1 形式概念分析方法在需求分析中的应用

形式概念这种分析方法是在近年来才兴起的,在这种分析方法出现之后就被广泛性的使用到了各个领域当中。而需求分析则是软件在不同需求当中应用性的分析,这也就表明在分析的过程当中能够具体分析其所使用的各种使用的环境,之后则能够对不同类别软件的应用信息进行有效的整合与管理。所以从大的方向来看,形式概念这种分析方法能够促进在各种实际应用当中软件使用的灵活性和有效性。具体到其实际使用的要求来看,这种分析方法的语境和属性都能够在需求分析当中起到非常重要的作用,因而也就能够为相关和同类别的软件提供一些基础和借鉴。

5.3.2 形式概念分析方法在结构设计中的应用

设计和研发作为一种高科技领域范畴的工作,主要是对于数据结构进行改革和提升。所以在基本需求分析和管理的基礎之上对其数据进行合理化的组织和分析,这样才能够总结出较为科学和有效的方法。一般来说假设这种方式能够较为合理的表明概念分析方法当中对于概念的构造原理,之后则可以通过有关的分析器来对每个项目进行分析和使用。因而在形式概念分析的过程当中需要具体分析变量之间的关系,这样才能够分析不同项目特征所能形成的相关概念,这样才能够形成系统化的概念格,也才能够更加有效的促进形式概念这种分析方法在结构设计中的有效使用。

5.3.3 形式概念分析方法在系统设计中的应用在对于软件工程系统设计当中,这一阶段最主要的任务和工作就是需要构造出系统的概念格,之后则能够在此基础上出现较多概念格的构造算法。而这种算法是能够更加科学有效的计算出格的概念和不同层次的关系,这样才能够具体分析出不同项目所组成集合的核心内容。在寻找到概念分析的核心之后就可以较为准确的获得相关概念间的关系。通过对于不同集合关系进行分析和整合就可以获得系统化的概念格。在之后相关专家就可以通过一系列的计算来对概念格的数值范围进行分析和计算,这样才能够促进软件应用的最大化。[6]

6 总结

从本文中,我们概括性地了解了形式概念分析相关的一些概念和基本方法,并展示了它的一些应用场景。可见,形式概念分析对领域认识起着至关重要的作用。在认识一个领域的知识集的时候,系统地整合了各个概念,并以更加规范化的方式呈现出来,方便之后进行知识挖掘。

参考文献

- [1] 徐春明,林强,王璨,杨楠,薄瑜. 高校大学生分类管理中基于等价类的属性约简算法应用[J]. 计算机科学与应用, 2020, 10(4): 665-675. <https://doi.org/10.12677/CSA.2020.104069>
- [2] 周超,任志宇,毋文超. 基于形式概念分析的语义角色挖掘算法[J]. 计算机科学, 2018.

- [3] 朱佳, 王向前, 张宝隆, 等. 基于形式概念分析的煤矿事故本体构建[J]. 工矿自动化, 2018, 44(05):30-34.
- [4] Yoneda, Yuka & Sugiyama, Mahito & Washio, Takashi. (2018). Learning Graph Representation via Formal Concept Analysis.
- [5] 渠寒花, 惠建忠, 何险峰, 等. 气象服务形式概念分析模型研究[J]. 计算机工程与应用, 2018, 54(9):257-264. DOI:10.3778/j.issn.1002-8331.1611-0029.
- [6] 胡鑫. 形式概念分析在软件工程中的应用[J]. 电脑迷, 2016, (5):39. DOI:10.3969/j.issn.1672-528X.2016.05.034.