

《智能信息处理》课程作业

## 形式概念分析

洪川宇

作业	分数[20]
得分	

2020 年 11 月 12 日

# 形式概念分析

洪川宇

(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，它反映了概念的哲学理解，形式概念分析的核心部分为概念格，在数据分析中，概念格是一种有效的工具。形式概念分析的每个节点都由两部分组成：外延和内涵。此外，形式概念分析通过哈斯图生动和简洁地体现了这些概念之间的关系。当今时代随着科技的不断进步，概念格的理论越来越多地被应用到各种领域，例如数据挖掘、信息检索、软件工程，在处理大规模数据的过程中发挥的重大的作用。本文主要介绍了形式概念分析的主要定义，以及概念格的构建方法及概念格的应用。

**关键词：**形式概念分析；形式背景；概念格；形式概念

## Review of formal concept analysis

HongChuanyu

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

**Abstract:** Formal concept analysis was proposed by Professor Wille in Germany in the early 1980s, which reflects the philosophical understanding of concepts. The core part of formal concept analysis is concept lattice, which is an effective tool in data analysis. Each node of formal concept analysis consists of two parts: denotation and connotation. In addition, the analysis of formal concepts embodies the relationship between these concepts vividly and concisely through Haas. Nowadays, with the continuous progress of science and technology, the theory of concept lattice is more and more applied to various fields, such as data mining, information retrieval, software engineering, which plays an important role in the process of processing large-scale data.This paper mainly introduces the definition of formal concept analysis, the construction method of concept lattice and the application of concept lattice.

**Keywords:** Formal concept analysis; Formal background; Concept lattice; Formal concept

根据百度百科显示，概念是人类在认识过程中，从感性认识上升到理性认识，把所感知的事物的共同本质特点抽象出来，加以概括，是自我认知意识的一种表达，形成概念式思维惯性。在人类所认知的思维体系中最基本的构筑单位。

概念是反应对象的特有属性的思维形式，是从对象的属性中抽出特有属性概括而成，并且每一个概念都有内涵和外延，其中概念的外延是指抽象出的对象，概念的内涵是指抽象出的特有属性。

目前来说，人类可以认知的事物都可以由概念表示，总的来说概念的表示法有三种方法，分别是

表达式法，二维表法和图示法，如图 1 所示。

①表达式法	概念 = (对象集, 属性集)																				
②二维表法	<table><tr><td></td><td>属性1</td><td>属性2</td><td>...</td><td>属性m</td></tr><tr><td>对象1</td><td></td><td></td><td></td><td></td></tr><tr><td>...</td><td></td><td></td><td></td><td></td></tr><tr><td>对象n</td><td></td><td></td><td></td><td></td></tr></table>		属性1	属性2	...	属性m	对象1					...					对象n				
	属性1	属性2	...	属性m																	
对象1																					
...																					
对象n																					
③图示法	<div>(对象集, 属性集)</div> <div>○</div>																				

图 1 概念的三种表示

对于以上三种概念的表示方法，可以这样理解。

现将要把对象人和狗，分别用以上三种方法抽象成概念。人的属性可抽象成语言和哺乳，狗的属性可抽象成狗吠和哺乳，由此根据概念的三种表示方法，可将人和狗抽象成以下三种形式。

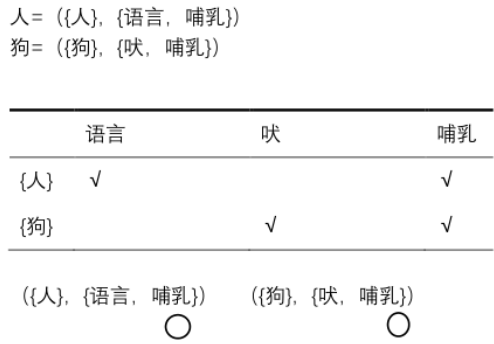


图2 三种不同的表示方法

### 1.2 自然概念定义

自然概念的定义和概念的定义类似，自然概念是人们认为很自然的事物的属性集。自然概念是概念的一个子集，自然概念不是一个抽象的概念，而是某个事物的具体体现。通俗的说自然概念就是人们头脑中认为的自然事物，然后人们对它加以理解，最后用概念的形式表现出来。

### 1.3 形式概念定义

形式概念(Formal Concept Analysis, FCA)由 Wille 博士于 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。形式概念是用来构建自然概念的层次联通结构的。形式概念的定义如下所示：

- 设形式对象集  $U \quad X \in U$
- 形式属性集  $A \quad B \in A$
- 二元关系  $R \in U \times A$
- 若  $X = \{x \mid x \in U, \forall a \in B, xRa\}$
- $B = \{a \mid a \in A, \forall x \in X, xRa\}$
- 则二元组  $(X, B)$  被称为形式概念
- $X$  中  $x$  每个  $x$  都有全部属性
- $B$  中  $a$  每个  $x$  都有的属性

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。下面继续以人狗的例子来说明形式概念。

形式概念

	语言	吠
人,狗		
人	√	√
狗		√
∅	√	√

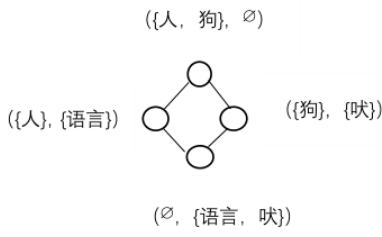


图3 形式概念的表示

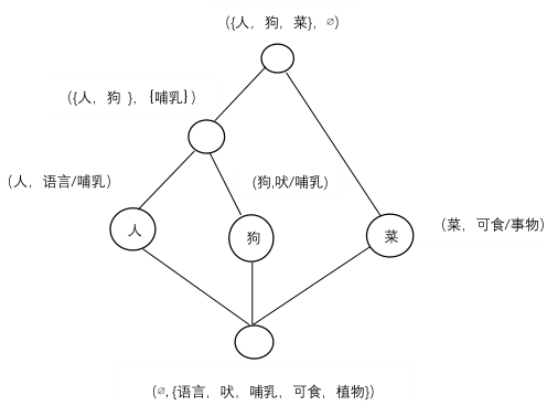


图4 形式概念的实例

如图4所示，为形式概念分析的实例，运用形式概念可以极大限度地对集合中具有某种关系或者含有某些共同属性的元素进行分类，发现由属性和对象构成的概念和概念之间的关系。例如人和狗都具有哺乳这一属性，则将人和狗、哺乳分别作为对象和属性提取出来，由此类推，则可以发现不同对象之间的联系和不同对象属性之间的联系。事实证明，应用形式概念，在信息检索，智能信息处理等方面起着巨大的作用。

### 1.4 形式背景

形式背景(formal context)顾名思义，形式指的是形式概念分析，背景指的是置于其中的环境，对形式背景进行约简可得约简背景。可以将形式背景定义成一个三元组，形式背景设为  $K$ ，则  $K = (G, M, I)$ ，

集合  $G$  为所有的对象，集合  $M$  为对象的属性， $I \subseteq G \times M$  为  $G$  和  $M$  中元素之间的二元关系集合。该三元组可以表示为二维表。在下面表 1 所示的形式背景中，关于对象集合  $G=\{g_1, g_2, g_3, g_4\}$ ，属性集合  $M=\{a, b, c, d\}$ ，二元关系  $I$  为确定性关系。<sup>[6]</sup>事实上，形式背景一般都不是直接存在的，需要从数据源中提取，从而就需要对数据源进行分析，采取不同的策略和算法来提取形式背景。

表 1 形式背景的示例

	a	b	c	d
$g_1$	√		√	√
$g_2$	√	√		
$g_3$	√	√	√	
$g_4$			√	

## 2 概念格

概念格，又称为 Galois 格，是 Wille 博士于 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。<sup>[3]</sup>

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。<sup>[3]</sup>

为使读者更好的了解定义概念格的，以下说明几个定义。在偏序集中，任意两元素的上下确界存在，则此偏序集为格。在格中，任意子集的上下确界存在，则此格为完备格，如图 5 所示为偏序集、格和完备格之间的关系。其中元素为概念的概念格称为概念格。如图 6 所示为概念格示意图，其具体的构成方法将在下文说明。

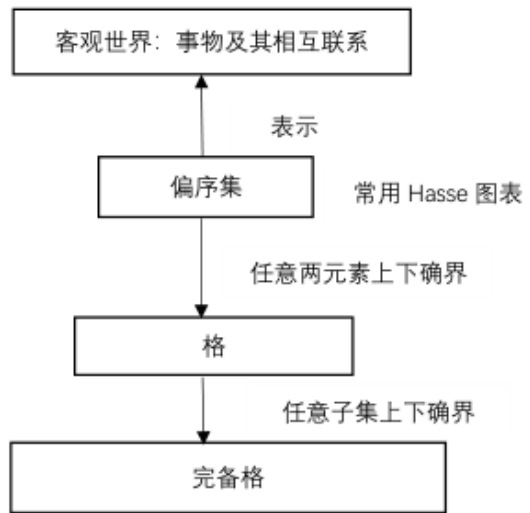


图 5 概念格的表示

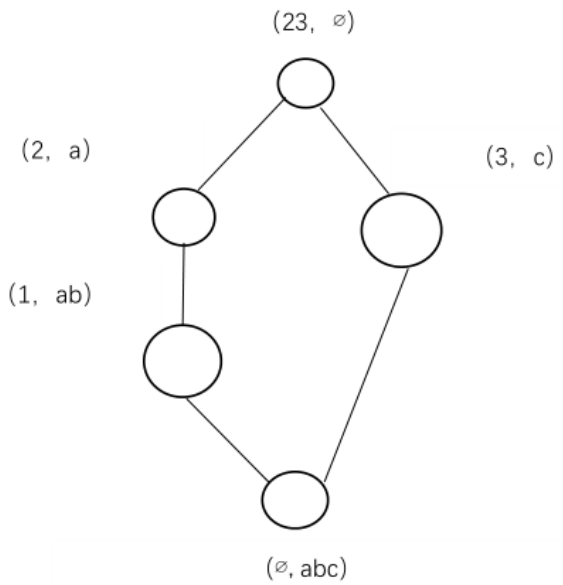


图 6 概念格的表示

## 3 概念格的生成

### 3.1 生成形式背景

在形式概念分析中，数据是用形式背景表示的。形式概念分析是 Wille 提出的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。形式概念分析强调以人的认知为中心，提供了一种与传统的、统计的数据分析和知识表示完全不同的方法，成为了人工智能学科的重要研究对象。生成形式背景分为以下四种：单值形式背景，单值粗糙形式背景，单值模糊形式背景，单值粗糙模糊形式背景。四种形式背景有着不同的生成方式，下面将举一个例子以下例子为某办公用品销售公司的业务活动的形式背景，如表 2 所示。

表 2 某办公用品销售公司的业务活动的形式背景

	设备	计算机	复印机	打字机	专用装置
咨询	✓	✓	✓	✓	✓
方案	✓	✓			
安装	✓	✓	✓	✓	✓
指导		✓	✓	✓	✓
培训		✓	✓		
配件	✓	✓	✓	✓	✓
维修	✓	✓	✓	✓	✓
服务		✓	✓	✓	

### 3.2 形式背景的约简

可以看到，在图 3 中存在不同行不同列的值相同的情况，这就产生了冗余。在形式背景部分已经提到了信息合并，消除冗余的方法，即约简行、约简列以及关联规则抽取。利用以上概念可以将具有相同属性的复印机、打字机合并，将复印机、打印机的共同属性指导以及咨询合并，把所有对象都有的配件、安装以及维修合并，得到约简后的形式背景，如表 3 所示。

表 3 约简后的形式背景

	设备	计算机	复印机/打字机	专用装置
咨询/安装/配件/维修	✓	✓	✓	✓
方案	✓	✓		
指导		✓	✓	✓
培训		✓		
服务		✓	✓	

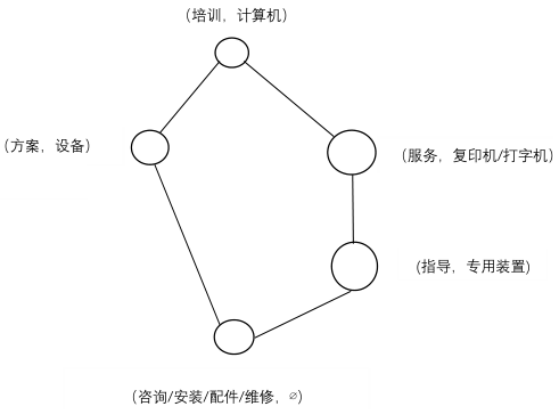
约简后的形式背景再经过基于属性个数的排序，得到最终的形式背景，如图 5 所示。通俗的来说培训包括一个属性，方案包括两个属性，由此类推，按照属性个数递增的顺序对约简后的形式背景进行重新排序，得到最终的形式背景，如表 4 所示。

表 4 最终的形式背景

	设备	计算机	复印机/打字机	专用装置
培训		✓		
方案	✓	✓		
服务		✓	✓	
指导		✓	✓	✓
咨询/安装/配件/维修	✓	✓	✓	✓

### 3.3 概念格的构建

之后，根据最终的形式背景可以通过基于父子关系的渐进式构建，生成 Hasse 图，得到概念格，如图 6 所示。



根据以上例子可以得知，哈斯图由于自身的特点可以清晰的表示概念格，哈斯图的每一个结点都是对象，在图中省略了自反性、省略了箭头，使图像简洁，箭头默认朝下则是基于父子关系的渐进式构建。通过观察哈斯图，人们可以更好地了解不同形式对象之间属性的继承关系，使得人们可以快速得到更多有用的信息。对于人们已知的知识，经过概念格的形式进行推演，可以更好地了解不同形式概念之间的关系，帮助人们对系统有更好的认识，从而发现潜在的知识。

## 4 总结

由 Wille R 于 1982 年首先提出的形式概念分析提供了一种支持数据分析的有效工具。形式概念包括外延和内涵两部分，它本质上描述了形式对象与形式属性之间的关系。形式背景则以二维表的形式描述了不同形式对象以及他们的形式属性之间的关系。在此基础上的概念格则是以哈斯图的形式，描述了不同形式对象之间的父子关系。概念格结构

模型来源于形式概念分析理论,是形式概念分析理论中的核心数据分析工具。

概念格体现了一种概念层次结构,实现了对数据的可视化。因此,概念格受到了人们的广泛关注。概念格理论经过几十年的发展,如今已被广泛运用于软件工程、知识工程、人工智能等领域。本文从概念一词的含义到形式概念的定义及表示,再到形式背景,再到格的解释以及概念格的概念以及应用等方面总结了概念格的研究进展。当然,概念格仍是一个高速发展的领域,对于粗糙模糊概念格的研究、基于概念格的数据挖掘模型的实现等等都是以后的重点研究方向。

### 参 考 文 献

- [1] Wille R. Restructuring lattice theory: an approach based on hierarchies of concepts[C]//International Conference on Formal Concept Analysis. Springer, Berlin, Heidelberg, 2009: 314-339.
- [2] 降惠.概念格理论研究进展与发展综述[J].办公自动化,2019,24(09):18-21+28.
- [3] Haiyun Xu,Chao Wang,Kun Dong,Zenghui Yue.Identification and Prediction of Interdisciplinary Research Topics: A Study Based on the Concept Lattice Theory[J].Journal of Data and Information Science,2019,4(01):60-88.
- [4] 于雪苹.基于概念格理论的研究及其应用分析[J].计算机学报, 2018.
- [5] 陈朝晖. 基于概念格的数据挖掘研究及应用[D]. 西安电子科技大学, 2014.
- [6] 梁高明, 张忠磊. 基于概念格的数据挖掘方法研究[J]. 科技信息, 2007, 2007(8):55-56.