

《智能信息处理》课程作业

基于形式概念分析的研究

王壮

作业	分数[20]
得分	

2021 年 11 月 28 日

基于形式概念分析的研究

王壮

(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

Research Based on Formal Concept Analysis

Wangzhuang

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

1 引言

形式概念分析 (Formal Concept Analysis, 简称 FCA) 由 Wille R 于 1982 年提出，其目的是对格理论进行重构，以促进格理论研究者和格理论的潜在使用者之间更好的交流[1]。形式概念分析提供了一种较好的层次化对象的分析方法，他能够识别那些具有共同属性的一组对象的组合。一方面，形式概念分析已具备较完善的理论基础，在应用软件工程，特别是软件维护活动中时，具有形式化方面的表达能力，从而具有较强的理论和技术说服力；另一方

面，形式对象和形式属性这种二元关系经常出现在软件世界中，对这种二元关系处理的方法也推动形式概念分析技术在软件领域应用的不断发展。概念格主要用于认知计算、机器学习、模式识别、专家系统、决策分析、网页搜索等领域。已逐步发展成一门成熟的知识发现体系，并演变出很多不同的研究范式[4]。近年来，概念格应用研究出现一些新领域，比如认知概念学习，规则提取，三支决策等等。本文主要是对形式概念分析的各个方面进行详细介绍。

2 形式概念

2.1 理论概述

信息技术得迅速发展, 在商业管理、政府部门、科学研究和工业 数据处理等领域都收集并且存储了大量的数据, 信息爆炸或者信息 泛滥使当今数字化社会面临着巨大的挑战, 有人说我们是数据的富翁, 知识的穷人。面对如此的尴尬的局面, 迫使人类不得不思考如何 才能解决这个问题。形式概念分析 是如何从大量的数据信息中自动 地抽取有价值的知识和信息, 因而成为人工智能中非常活跃的研究领域。

2.2 形式概念分析产生的背景

人类在认知过程中, 把所有感 觉的事物的共同特点抽出来, 加以概括, 就称为概念。在哲学中, 概 念被定义为思维的基本形式之一, 它反映了可观事物的一般的、本质 的特性。概念被理解为由外延和内 涵两个部分所组成的思想单元。对概念的这种理解源于古希腊哲学, 发展于十七世纪的近代学院派, 进一步发展成德国标准, 最终成为了世界标准。由于概念是思维的基本 单元, 因此基于对概念的研究成为人工智能的一个重要组成部分。在知识表示、知识管理、机器学习、 专家系统等不同的领域, 研究者们 从不同的角度和观点来分析概念, 形成了对概念的不同形式化描述 方法。形式概念分析是一种对数据进 行分析的工具或者方法, 特别是可以对给定的信息进行分析和处理。而数据应该是从人类有意义的 可以理解的思维单位----概念中抽 取而形成的形式化的单元。形式化 表明的是所有处理的数据是形式 化的数学实体, 不需要和人类思维 中的概念完全相同, 同时它也指出 形式概念分析处理的基本数据形式是形式背景, 形式背景是人类背 景知识中的一小部分。形式概念分析是由德国的 Darmstadt 技术大学数学系于 70 年 代未发展起来的。当时 Darmstadt 的 研究小组致力于系统研究并发展一种基于 Birkhoff 格理论的应用软 件。自 1981 年在关于有序集合的 Banff 会议的专题演讲上对形式概 念分析做了首次描述以后, 数以百 计的

相关论文开始出版和发表, 甚至包括一篇格理论的数学基础的书。Darmstadt 研究小组此后一直从事多达上百个基于格理论的应用软件研究, 因此研究先前的部分成员已经建立了一个很小的原型系统, 并且进一步使其逐步实用化。Darmstadt 小组的研究显然在形式概念分析的诞生和应用过程中起到了重大的作用。他们首先把多种可能性系统的详细阐述为一种数据分析方法, 并且进行试验, 进一步开发许多应用。这个工作成功的决定性因素存在于背景的形式化和把“概念”作为外延和内涵实体的解释。在这里“概念”已经被形式化, 而且对它的理解深深植根于哲学, 这个在文献中有详细的描述。这种思想的传统甚至在标准 DIN 2330 和 DIN 2331(DIN 表 示 “Deutsche Industrienorm”, 对于由 German National Bureau of Standard 发布的标准是很典型的) 中发现了表述, Darmstadt 小组在开发初期, 多次讨论了这两个标准。

3 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。[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]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 1 所示为偏序集、 格和完备格之间的关系。

其中元素为概念的格称为 概念格。如图 2 所示为概念格示意图，其具体的构成方法将在下文说明。

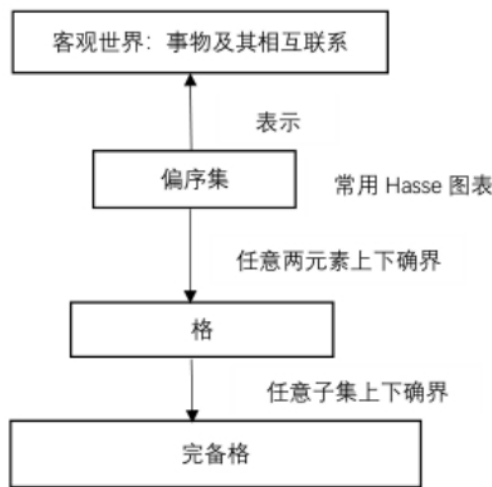


图 1 概念格的表示

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。概念格的生成包括如下图 3 中的七个步骤：

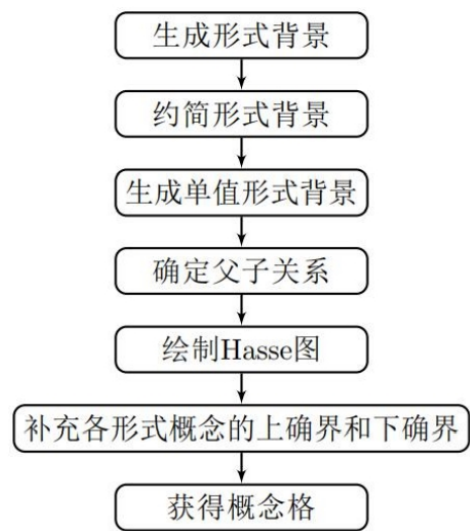


图 3 概念格的产生步骤

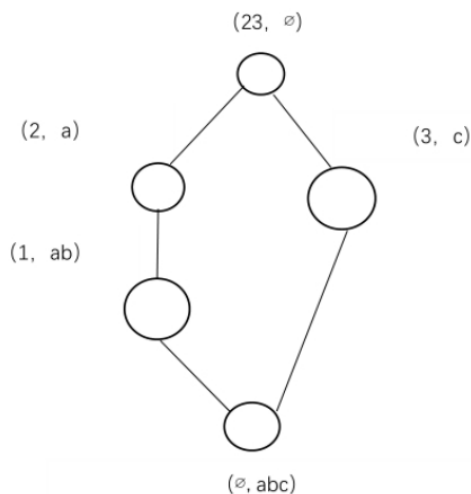


图 2 概念格的表示

4 概念格的生成

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，

5 概念格的应用

2002 年，陈世权和程里 春又提出了描述概念内涵和外延模糊关系的模糊概念格，建立了概念的 内涵与外延的模糊映射，用模糊集描述模糊概念思维，并进一步 探讨了模糊属性映射和模糊对象 映射的隶属函数的一些基本数学性质，证明全体模糊概念构成一个完备格，相关的研究工作对模糊概念刻画、模糊推理、模糊形式语言 具有重要的理论和应用价值。以上面的都是针对单个概念格 进行的，随着网络技术尤其是互联网的飞速发展，数据的分布式存储和并行处理的需求越来越迫切，对多概念格的合并算法也有了一些研究。概念格的研究并不仅仅限于理论方面，在应用上也有一定的发展，比如在软件工程、信息检索、知识发现、数据挖掘、规则提取、聚类分析等领域。在知识发现领域，概念格可以从关系数据中提取各种 类型的知识，比如关联规则、蕴涵规则、分类规则等等；在软件工程 领域，概念格可以从类库以及类库的规范说明上构造，从而对类库的可视化以及类库的重构和优化提供支持；在知识工程领域，概念格可

以用于知识库的重新结构化；在信息检索方面，概念格可以实现对信息的有机组织并过滤掉无用的信息，比如和粗糙集理论相结合对信息系统进行属性约简，来简化信息表。目前也有人在应用概念格进行搜索引擎方面的研究。而且有人指出概念格将会在生物和生命科学领域有重大应用

6 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] 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.
- [2] 李金海, 魏玲, 张卓,等. 概念格理论与方法及其研究展望[J]. 模式识别与人工智能, 2020(7).
- [3] 徐伟华等.形式概念分析的理论与应用[M].北京:科学出版社,2016
- [4] 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
- [5] 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.