

形式概念分析

理论与应用

*Formal Concept Analysis:
Theory and Application*

徐伟华 李金海 魏 玲 张 涛 著



科学出版社

形式概念分析理论与应用

徐伟华 李金海 魏 玲 张 涛 著

科 学 出 版 社

北 京

内 容 简 介

本书以形式概念分析的理论与实践应用为主线,系统深入地介绍了基于形式概念分析的众多理论与应用的最新研究成果。全书共13章,着重阐述了基于直观图的形式背景、属性偏序概念图生成理论、形式模糊背景的非矩阵知识约简方法、区间集概念格、形式背景对特殊序结构的刻画和表示、概念格的决策逻辑、直觉模糊广义协调决策形式背景中的数据分析、模糊数据集中基于形式概念分析的双向学习、概念格与多粒度粗糙集的规则比较、形式概念分析的属性拓扑及其基本应用和三元概念分析。

全书注重系统性、严谨性、理论性和可读性,可以作为高等院校应用数学、计算机、信息科学、系统工程等专业高年级本科生及研究生的教学用书,也适合作为相关专业科研工作者的参考书。

图书在版编目(CIP)数据

形式概念分析理论与应用 / 徐伟华等著. —北京: 科学出版社, 2016.7
ISBN 978-7-03-049317-0

I. ①形… II. ①徐… III. ①数据处理 IV. ①TP274

中国版本图书馆 CIP 数据核字 (2016) 第 153604 号

责任编辑: 杨 岭 黄明冀 / 责任校对: 杨悦蕾

责任印制: 余少力 / 封面设计: 墨创文化

科学出版社出版

北京东黄城根北街16号

邮政编码: 100717

<http://www.sciencep.com>

成都锦瑞印刷有限责任公司印刷

科学出版社发行 各地新华书店经销

*

2016年8月第 一 版 开本: 787×1092 1/16

2016年8月第一次印刷 印张: 17 1/2

字数: 415 千字

定价: 120.00 元

(如有印装质量问题, 我社负责调换)

编委会成员

顾 问：张文修

主 编：徐伟华 李金海 魏 玲 张 涛

编 委：（以姓氏笔画为序）

万 青 马建敏 米据生 祁建军

李少雄 李文涛 李克文 李金海

张晓燕 张 涛 邵明文 庞晋中

洪文学 姚红娟 钱 婷 徐伟华

郭兰坤 景 嫒 解 滨 裴 铎

翟岩慧 魏昕宇 魏 玲

序

形式概念分析是由德国数学家 Wille R. 教授于 1982 年提出的, 它处理的是一类特殊的关系数据库。通过对象集与属性集之间的某种关联, 建立由对象集与属性集对形成的形式概念, 而把关系数据库称为形式背景。由于分析数据的核心工具是概念格, 国内又称它为概念格理论(或简称概念格)。

形式概念分析早期研究的学者并不多, 基本上限于 Wille R. 教授本人及其学生与同事。随着概念格理论结果的丰富和有关算法及程序的成功开发, 逐渐引起了其他领域感兴趣的学者的关注。我于 20 世纪 90 年代初期在香港与梁怡教授和梁广锡教授曾合作研究不确定性推理, 提出了包含度理论, 已注意到概念格在知识获取中的应用, 并考虑了多值关系数据库的形式背景, 这些内容已收集在我与梁怡教授合作的《不确定性推理原理》一书中。由于当时正关注着遗传算法收敛性研究, 因此未能继续研究下去。

1998 年左右, 在学生的推动下, 我开始了粗糙集的研究。当时, 我即将从研究生院院长的位置上退下来, 有较多时间阅读文献和从事研究工作。于是以后的几年时间里, 我招收了比较多的博士生, 并开始了与加拿大里贾拿大学姚一豫教授的合作。当我们团队在粗糙集方面研究得如火如荼的时候, 姚一豫教授发给了我们他尚未发表的文章, 文章从粗糙集的观点来研究概念格。受姚一豫教授研究工作的启发, 我们在研究粗糙集的同时也开始重新研究概念格。

在 21 世纪前十年中, 魏玲博士、吴伟志博士、米据生博士、祁建军博士、马建敏博士、杨宏志博士、仇国芳博士、徐伟华博士、邵明文博士、张红英博士、李同军博士、王虹博士、李鸿儒博士、王霞博士、宋笑雪博士、李金海博士、范世青硕士等在概念格方面做出了出色的工作, 他们的研究论文都发表在国内外重要期刊上。他们的研究成果可以归结为以下几个方面: ①给出了形式背景的属性约简定义、概念格属性约简理论与方法和核心属性的判定定理; ②给出了决策形式背景的定义、决策规则获取方法和决策条件属性约简的理论与方法; ③给出了概念格之间的广义偏序关系, 以及广义协调形式背景决策条件属性约简的理论与方法; ④给出了偏序关系下, 协调与不协调决策形式背景的知识获取与属性约简方法; ⑤给出了对象概念格与属性概念格的属性约简理论、方法与属性特征; ⑥给出了基于概念格的信息粒认知模型及认知学习的双向逼近迭代算法; ⑦在模糊概念格的基础上, 给出了三种变精度概念格及其性质与近似推理; ⑧给出了 4 类模糊概念格及其相互关系, 以及基于 4 种不同模糊概念格的模糊推理; ⑨充分利用了依赖空间方法在概念格研究中的作用, 给出了一个形式背景所有概念的格关系; ⑩给出了多粒度形式背景, 研究了多粒度形式背景下的信息粒度关系, 为粒计算提供了一种新的数学模型。

十多年来，概念格研究开始转向计算机科学、信息科学等领域，这开启了概念格与其他理论交叉融合研究的浪潮，属于概念格蓬勃发展的阶段。现在，概念格已被认为是“概念数据分析与知识处理”的重要数学工具，应用相当广泛，并且吸引了诸多学者的研究兴趣。

令人可喜的是，我们这个团队的博士已成为各个高等院校的学术骨干，他们大多已是教授并招收硕士生和博士生，他们和他们的学生坚持进行着对概念格的研究，并提出了一些新颖的思路。比如，把认知计算与概念格结合起来研究认知概念学习；将粒计算和概念格交叉研究概念格的粒计算方法；将包含度理论与概念格融合给出近似推理的更多方法；将区间集和形式背景结合起来，提出区间集概念格；把概念格和决策逻辑统一起来研究，给出概念格的决策逻辑；形式背景推广到区间值或三维形式背景的三元概念格；从属性偏序的角度提出属性偏序概念图，以及利用概念格的属性拓扑工具在可视化方面的成功应用等。这些新的研究思路将为概念格的发展注入新鲜的活力，值得期待。

《形式概念分析理论与应用》的编著者均是从事概念格一线科研工作的年轻学者，他们将自己最新的一些科研成果总结成书，我非常高兴为该书作序，并希望他们今后取得更多更优秀的成果。

张文修

2016年4月于西安

前 言

大数据时代已经来临,生物计算、物联网+、社交网络、云计算等新兴服务促使人类面临前所未有的挑战和机遇。一些具有海量、高维、分布式、动态等特征的大规模复杂数据不断涌现,人们迫切要去分析处理这些复杂数据,从中找到有价值的信息。然而,直接处理这些数据的面临着严重的计算问题和维数灾难问题。千古留传的“瞎子摸象”的故事就反映出在缺少视觉信息条件下对不完全信息进行融合会产生认知偏差。因此,如何有效、快速地处理这些复杂数据,并提取出隐含其中的、潜在的有用知识,一直是智能信息处理领域的一个研究热点。

形式概念分析(formal concept analysis, FCA)是德国数学 Wille R. 教授于 1982 年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、关系等用形式化的语境表述出来,然后根据语境,利用伽罗瓦连接构造出概念格(concept lattice),即本体,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,自动产生本体。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,在机器学习、数据挖掘、知识工程、管理科学、软件工程、信息检索等领域得到了广泛的应用。

概念这一术语源于哲学,它是外延和内涵的统一体。概念之间的泛化和例化关系使所有的概念构成一个完备格,又称为概念格。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。自 2003 年起,每年都召开一次“形式概念分析国际会议”(ICFCA)与“概念格及其应用国际会议”(CLA)。后来,越来越多的学者投入到形式概念分析理论与应用这一领域的研究中,并取得了丰硕的成果。1999 年, Ganter 出版了学术著作 *Formal Concept Analysis: Mathematical Foundations*。这一著作的出版大大推动了形式概念分析这一领域的发展。2007 年,国内学者马垣对 Ganter 这一力作进行了翻译并于科学出版社出版。同时,马老师于 2011 年又出版了《形式概念分析及其新进展》,整理了我国部分学者关于形式概念分析的最新成果。

随着智能信息技术突飞猛进的发展,2000 年左右,形式概念分析理论在我国受到了

极大的关注，并不断涌现出众多的优秀学者和有价值的成果。特别是以西安交通大学张文修教授为代表的研究团队，在 21 世纪前十年对形式概念分析的理论，尤其是对应用进行了深入的研究和探讨。而且，近期又有许多学者对概念格的研究提出了一些非常新颖且应用价值极大的思路。比如，把认知计算与概念格结合起来得到新课题“认知概念学习”；还有将粒计算和概念格交叉研究得到“概念格的粒计算方法”；另外包含度与概念格的融合也取得一些新进展。这些成果大都发表在 *IEEE Transactions on Cybernetics*, *Information Sciences*, *Fuzzy Sets and Systems*, *Knowledge-Based Systems*, 《中国科学》等国际重要期刊上。

同时，像本书中某些研究结果所显示的那样，形式概念分析理论与方法对研究信息系统中若干问题，如知识的获取，规则的协调性及矛盾规则的排除，都是一种非常有效的方法。可见，形式概念分析理论不仅是研究不确定性推理的一般理论，而且是研究不确定性现象的思想论、哲学论。

基于上述出发点，全书以形式概念分析的理论与实践应用为主线，系统深入地介绍了基于形式概念分析的众多理论与应用的最新研究成果。全书除预备知识之外共计 12 章，着重阐述了基于直观图的形式背景、属性偏序概念图生成理论、形式模糊背景的非矩阵知识约简方法、区间集概念格、形式背景对特殊序结构的刻画和表示、概念格的决策逻辑、直觉模糊广义协调决策形式背景中的数据分析、模糊数据集中基于形式概念分析的双向学习、概念格与多粒度粗糙集的规则比较、形式概念分析的属性拓扑及其基本应用和三元概念分析。本书是由国内形式概念分析及其粒计算研究领域的 20 多位一线科研工作者和学者合作撰写的。全书由徐伟华、李金海、魏玲、张涛共同编写并统稿。具体章节撰写分工如下：

第 1, 11 章由昆明理工大学李金海(副教授)撰写；

第 2 章由西北大学魏玲(教授)、西安工程大学万青(博士)撰写；

第 3 章由燕山大学兼东北大学秦皇岛分校洪文学(教授)、燕山大学李少雄(博士生)撰写；

第 4 章由中国石油大学(华东)邵明文(教授)、李克文(教授)撰写；

第 5 章由长安大学马建敏(副教授)、姚红娟(硕士)、景嫻(硕士)、西安交通大学张文修(教授)撰写；

第 6 章由河北师范大学解滨(教授)和米据生(教授)、承德医学院裴铎(讲师)撰写；

第 7 章由湖南师范大学郭兰坤(副教授)撰写；

第 8 章由山西大学翟岩慧(博士)撰写；

第 9, 10 章由重庆理工大学徐伟华(教授)、张晓燕(副教授)、李文涛(博士生)、庞晋中(硕士)撰写；

第 12 章由燕山大学张涛(副教授)、魏昕宇(硕士)撰写；

第 13 章由西北大学魏玲(教授)、西安工程大学万青(博士)、西北大学钱婷(博士生)、西安电子科技大学祁建军(教授)撰写。

本书研究内容为基于形式概念分析解决复杂系统的知识发现、不确定决策分析、规则获取等问题提供了有利工具。书中汇集了当前形式概念分析、粒计算研究团队的最新

研究成果，其写作目的是让更多的学生、老师、学者了解形式概念分析理论、方法及应用，共同促进该领域的发展，并希望能为形式概念分析的研究发展做出更大的贡献。

本书在编写过程中参考了许多同类著作，吸收了许多观点已在参考文献中列出，在此向文献作者表示由衷的感谢。特别感谢西安交通大学张文修教授及各章节负责人、评审人对本书提出的宝贵意见。

本书的出版得到了国家自然科学基金项目 61105041, 61472463, 61402064, 61170107, 61573127, 61300121, 10901025, 11371014, 11401195, 61074130, 61273019, 81273740, 81373767, 61305057, 61562050, 61363056, 61303107 的资助，在此一并表示感谢。

本书注重系统性、严谨性、理论性和可读性，可以作为高等院校应用数学、计算机、信息科学、系统工程等专业高年级本科生及研究生的教学用书，也可作为相关专业科研工作者的参考辅导工具书。

由于作者水平有限，加上时间仓促，书中难免存在不足之处，敬请读者批评指正。

目 录

第 1 章 预备知识	1
1.1 格论基础	1
1.2 形式背景与概念格	2
1.2.1 形式背景与形式概念	2
1.2.2 概念格	3
1.3 面向对象概念格和面向属性概念格	3
1.3.1 各种算子及其性质	3
1.3.2 面向对象概念与面向属性概念	4
参考文献	4
第 2 章 基于直观图的形式背景	5
2.1 直观图的基本定义	5
2.2 基于直观图的粒转化	7
2.2.1 带标记的对象类与对象概念的转化	7
2.2.2 带标记的属性类与属性概念的转化	9
2.3 基于直观图的近似概念获取	10
2.3.1 k -阶关系的定义及性质	10
2.3.2 近似概念的定义及获取	12
2.4 基于直观图的不可约元判定及属性特征分析	15
2.4.1 基于对象直观图的不可约元判定	15
2.4.2 基于属性直观图的不可约元判定	18
2.4.3 基于属性直观图的属性特征分析	19
2.5 总结与展望	22
参考文献	22
第 3 章 属性偏序概念图	23
3.1 认知机理模型	23
3.1.1 理想的知识分类模型	23
3.1.2 认知机理模型	24
3.2 属性特征与对象特征的定义	25
3.2.1 形式背景中属性特征的定义	26
3.2.2 形式背景中对象特征的定义	27
3.3 属性偏序概念图及其生成	27

3.3.1	属性偏序概念图和对象偏序概念图	28
3.3.2	基于特征属性计算的偏序概念图生成	28
3.3.3	基于论域划分原理的属性偏序概念图生成	32
3.4	概念格与属性偏序概念图异同比较	36
3.4.1	5个著名形式背景的比较	36
3.4.2	相同与不同的本质	43
3.5	属性偏序概念图应用	47
3.5.1	工程问题上的规则提取	47
3.5.2	医学问题上的知识发现	49
3.6	总结与展望	51
	参考文献	52
第4章	模糊形式背景的非矩阵知识约简方法	53
4.1	模糊形式背景和模糊-经典概念格	54
4.2	模糊子背景和相应的概念格	56
4.3	模糊形式背景中的交不可约元	57
4.4	模糊形式背景中的基于模糊-经典概念格的属性特征	60
4.5	模糊形式背景下基于模糊-经典概念格的属性约简	64
4.6	总结与展望	65
	参考文献	66
第5章	区间集概念格	69
5.1	区间集及其代数结构	69
5.1.1	区间集	69
5.1.2	区间集的运算	69
5.2	区间集概念格及其区间集约简方法	71
5.2.1	区间集概念格	71
5.2.2	区间集概念格的区间集协调集判定方法	74
5.2.3	区间集概念格的区间集约简方法	77
5.3	区间集概念格的近似表示	80
5.3.1	对偶概念格	80
5.3.2	对偶区间集概念格	81
5.3.3	区间集概念格的近似表示	82
5.4	区间集概念信息粒	85
5.4.1	区间集概念认知系统	85
5.4.2	区间集概念信息粒的转化	86
5.5	总结与展望	87
	参考文献	87
第6章	Axiality-概念格及其属性约简方法	89
6.1	基于不可约元的 irr-型属性约简方法	89

6.1.1	不可约元	89
6.1.2	协调决策形式背景的属性约简	90
6.1.3	不协调决策形式背景的属性约简	91
6.2	同态映射环境下的属性约简方法	93
6.2.1	协调决策形式背景的同态约简	93
6.2.2	同态约简的辨识矩阵	96
6.3	Axialities 导出概念格的属性约简方法	97
6.3.1	伽罗瓦连接与幂集上的二元关系	97
6.3.2	Axiality-概念格	99
6.3.3	Axiality-概念格的属性约简	101
6.3.4	Axiality-概念格的属性特征	105
6.4	总结与展望	108
	参考文献	109
第 7 章	形式概念分析对特殊序结构的表示	110
7.1	代数格与代数 domain	111
7.2	逼近概念对代数格的表示	112
7.2.1	闭包算子与代数格	112
7.2.2	并半格与代数格	113
7.2.3	逼近概念与代数格	114
7.3	相容 F -扩张形式背景和关联规则系统	114
7.3.1	相容 F -扩张形式背景和 F -概念	114
7.3.2	F -扩张形式背景诱导的关联规则系统	116
7.3.3	形式关联规则系统	117
7.3.4	形式关联规则系统诱导的 F -扩张形式背景	119
7.4	形式关联规则系统对代数 domain 的表示	120
7.4.1	形式关联规则系统范畴	120
7.4.2	Scott 连续映射和 A -联络之间的一一对应	121
7.4.3	形式关联规则系统和代数 domain 的范畴等价	123
7.5	总结与展望	126
	参考文献	127
第 8 章	概念格的决策逻辑	128
8.1	决策蕴涵的语义特征	129
8.1.1	决策蕴涵及其语义	129
8.1.2	一致闭包和模型判定	130
8.1.3	完备性判定	131
8.2	决策蕴涵的语构特征	133
8.3	决策蕴涵规范基	135
8.3.1	决策前提和决策蕴涵规范基	135

8.3.2 决策背景的决策蕴涵规范基	137
8.3.3 决策蕴涵规范基的生成算法	140
8.3.4 实验比较	142
8.4 总结与展望	144
参考文献	145
第9章 直觉模糊广义协调决策形式背景中的数据分析	147
9.1 蕴含映射与广义协调的基础定义和性质	147
9.2 蕴含映射下的属性约简	149
9.3 蕴含映射下基于辨识矩阵的属性约简	155
9.4 直觉模糊广义协调决策形式背景下 的概念格属性约简	156
9.5 数据分析方法与案例研究	159
9.6 总结与展望	164
参考文献	164
第10章 模糊数据集中基于形式概念分析的双向学习法	166
10.1 模糊数据集中的概念格	166
10.2 模糊数据集中的双向学习系统	168
10.3 模糊数据集中的双向学习机理	171
10.4 模糊数据集中的双向学习算法	173
10.4.1 模糊数据集中双向学习算法	173
10.4.2 时间复杂度分析	175
10.5 案例分析和试验评估	176
10.5.1 案例分析	176
10.5.2 实验评估	180
10.6 总结与展望	186
参考文献	187
第11章 概念格与多粒度粗糙集的规则比较	188
11.1 多粒度粗糙集与规则	188
11.1.1 多粒度粗糙集	188
11.1.2 粒规则与析取规则	190
11.2 决策信息系统转化为决策形式背景	190
11.3 多粒度粗糙集与概念格的规则比较	193
11.3.1 多粒度粗糙集与概念格规则之间的异同	193
11.3.2 多粒度粗糙集与概念格中的规则支持度和置信度	200
11.4 总结与展望	204
参考文献	204
第12章 属性拓扑及其基本应用	206
12.1 形式背景中属性耦合关系	206

12.1.1	二元属性间的基本耦合	206
12.1.2	形式背景下的属性分类	207
12.2	形式背景的属性拓扑表示	209
12.2.1	形式背景的预处理	209
12.2.3	属性拓扑的生成	210
12.3	属性拓扑的性质	213
12.3.1	属性拓扑的基本性质	213
12.3.2	基于属性拓扑的属性分类及相关性质	215
12.4	基于属性拓扑的形式概念计算算法	216
12.4.1	属性拓扑的分解定理	217
12.4.2	基于属性子拓扑的形式概念计算算法	221
12.4.3	利用属性拓扑的全路径搜索方法计算形式概念	226
12.4.4	基于原始背景的形式概念复原	228
12.5	属性拓扑在博客数据分析中的应用	228
12.5.1	博客数据分析的现实意义	228
12.5.2	实验数据与数据预处理	229
12.5.3	形式概念计算	232
12.6	总结与展望	238
	参考文献	239
第 13 章	三元概念分析	241
13.1	三元概念分析的基本概念	241
13.2	三元背景的蕴含及规则	247
13.3	三元因子分析	250
13.4	模糊三元概念分析	254
	参考文献	257
附录		259
索引		261

第1章 预备知识

为了避免后续章节在基础知识方面的重复叙述,本章对形式概念分析中的基本概念作统一介绍,希望达到统一全书基本符号的目的,使其更加紧凑与严谨。更多基本概念的介绍参见文献[1]~[6]。

1.1 格论基础

定义 1.1 设 S 为一个集合,若它满足

- (1)自反性: $\forall a \in S, a < a$;
- (2)反对称性: $\forall a, b \in S, a < b, b < a \Rightarrow a = b$;
- (3)传递性: $\forall a, b, c \in S, a < b, b < c \Rightarrow a < c$ 。

则称 $(S, <)$ 为偏序集。

定义 1.2 设 $(S, <)$ 为偏序集,若任意 $a, b \in S, \{a, b\}$ 有上确界元和下确界元,则称 $(S, <)$ 是一个格。

定义 1.3 设 $(S, <)$ 为格,若 S 的任一子集有上确界元和下确界元,则称 $(S, <)$ 是完备格。

需要指出的是,一个有限格必是完备的。

定义 1.4 设 2^S 为 S 的幂集,若映射 $\rho: 2^S \rightarrow 2^S$ 满足

- (1)保序性: $X \subseteq Y \Rightarrow \rho(X) \subseteq \rho(Y)$;
- (2)增值性: $X \subseteq \rho(X)$;
- (3)幂等性: $\rho(\rho(X)) = \rho(X)$ 。

则称 ρ 是 S 上的闭包算子。

定义 1.5 设 $\psi: S \rightarrow T, \varphi: T \rightarrow S$ 为偏序集 $(S, <)$ 和 $(T, \leq)$ 之间的一对映射,若它们满足

- (1) $a < b \Rightarrow \psi(b) \leq \psi(a)$;
- (2) $x \leq y \Rightarrow \varphi(y) < \varphi(x)$;
- (3) $a < \varphi\psi(a), x \leq \psi\varphi(x)$ 。

则称序对 (ψ, φ) 是 S 和 T 之间的伽罗瓦(Galois)连接。

准确地讲,定义 1.5 实际上给出的是反序伽罗瓦连接。

1.2 形式背景与概念格

1.2.1 形式背景与形式概念

定义 1.6 三元组 (U, A, I) 称为形式背景, 其中 $U = \{x_1, x_2, \dots, x_m\}$ 是非空有限对象集, $A = \{a_1, a_2, \dots, a_n\}$ 是非空有限属性集, I 是笛卡儿积 $U \times A$ 上的二元关系。约定, $(x, a) \in I$ 表示对象 x 拥有属性 a , $(x, a) \notin I$ 表示对象 x 不拥有属性 a 。

例 1.1 表 1.1 描述了一个形式背景, $U = \{x_1, x_2, x_3, x_4, x_5, x_6\}$, $A = \{a_1, a_2, a_3, a_4\}$, 其中数字 1 表示对象拥有属性, 数字 0 表示对象不拥有属性。

表 1.1 形式背景 (U, A, I)

U	a_1	a_2	a_3	a_4
x_1	1	0	1	0
x_2	1	1	0	0
x_3	1	0	0	1
x_4	0	1	0	1
x_5	0	1	1	0
x_6	1	1	1	0

为了从形式背景 (U, A, I) 中诱导出概念, 需进一步给出如下算子: 任意 $X \subseteq U$, $B \subseteq A$,

$$f(X) = \{a \in A \mid \forall x \in X, (x, a) \in I\} \quad (1.1)$$

$$g(B) = \{x \in U \mid \forall a \in B, (x, a) \in I\} \quad (1.2)$$

上述算子的直观意义非常明确, $f(X)$ 表示 X 中所有对象共同拥有的属性组成的集合, $g(B)$ 表示拥有 B 中所有属性的对象组成的集合。

为了方便, 通常记 f 和 g 的复合算子为 fg , 且 g 和 f 的复合算子为 gf 。易证, 算子 f 和 g 具有如下性质。

性质 1.1 设 (U, A, I) 为形式背景, 任意 $X, X_1, X_2 \subseteq U$, $B, B_1, B_2 \subseteq A$, 则

- (1) $X_1 \subseteq X_2 \Rightarrow f(X_2) \subseteq f(X_1)$;
- (2) $B_1 \subseteq B_2 \Rightarrow g(B_2) \subseteq g(B_1)$;
- (3) $X \subseteq gf(X)$;
- (4) $B \subseteq fg(B)$;
- (5) $f(X) = fgf(X)$;
- (6) $g(B) = gfg(B)$;
- (7) $f(X_1 \cup X_2) = f(X_1) \cap f(X_2)$;
- (8) $g(B_1 \cup B_2) = g(B_1) \cap g(B_2)$;
- (9) $f(X_1 \cap X_2) \supseteq f(X_1) \cup f(X_2)$;
- (10) $g(B_1 \cap B_2) \supseteq g(B_1) \cup g(B_2)$ 。

定义 1.7 设 (U, A, I) 为形式背景, $X \subseteq U$, $B \subseteq A$ 。若 $f(X) = B$ 且 $g(B) = X$,

则称序对 (X, B) 为形式概念(或简称概念)。同时, 称 X 为概念 (X, B) 的外延, B 为概念 (X, B) 的内涵。

1.2.2 概念格

概念是人类进行认知的基本单元, 不仅如此, 概念与概念之间还存在特化-泛化关系, 具体如下:

设 (X_1, B_1) 和 (X_2, B_2) 是形式背景 (U, A, I) 的两个概念, 若 $X_1 \subseteq X_2$ 或 $B_2 \subseteq B_1$, 则称 (X_1, B_1) 是 (X_2, B_2) 的特化概念, 或 (X_2, B_2) 是 (X_1, B_1) 的泛化概念。记概念之间的这种特化-泛化关系为 “ $\leq$ ”。

形式背景 (U, A, I) 的所有概念连同特化-泛化关系 $\leq$ 构成一个完备格, 记为 $L(U, A, I)$, 其中上确界元和下确界元分别定义为

$$(X_1, B_1) \vee (X_2, B_2) = (gf(X_1 \cup X_2), B_1 \cap B_2) \quad (1.3)$$

$$(X_1, B_1) \wedge (X_2, B_2) = (X_1 \cap X_2, fg(B_1 \cup B_2)) \quad (1.4)$$

上述完备格 $L(U, A, I)$ 称为概念格, 它是形式概念分析用于分析数据的核心工具, 其理论基础是序理论和格理论。

1.3 面向对象概念格和面向属性概念格

1.3.1 各种算子及其性质

设 (U, A, I) 为形式背景, $X \subseteq U, B \subseteq A$, 定义

$$xI = \{a \in A \mid (x, a) \in I\}$$

$$Ia = \{x \in U \mid (x, a) \in I\}$$

$$X^\square = \{a \in A \mid Ia \subseteq X\}$$

$$B^\square = \{x \in U \mid xI \subseteq B\}$$

$$X^\diamond = \{a \in A \mid Ia \cap X \neq \emptyset\}$$

$$B^\diamond = \{x \in U \mid xI \cap B \neq \emptyset\}$$

上述算子的语义解释如下: xI 表示对象 x 拥有的属性组成的集合, Ia 表示拥有属性 a 的对象组成的集合, $X^\square$ 表示对象集 X 所覆盖的属性组成的集合, $B^\square$ 表示拥有属性不超过 B 的对象组成的集合, $X^\diamond$ 表示与对象集 X 交叉的属性组成的集合, $B^\diamond$ 表示至少拥有 B 中一个属性的对象组成的集合。

性质 1.2 设 (U, A, I) 为形式背景, 任意 $X, X_1, X_2 \subseteq U, B, B_1, B_2 \subseteq A$, 则

$$(1) X_1 \subseteq X_2 \Rightarrow X_1^\square \subseteq X_2^\square, X_1^\diamond \subseteq X_2^\diamond;$$

$$(2) B_1 \subseteq B_2 \Rightarrow B_1^\square \subseteq B_2^\square, B_1^\diamond \subseteq B_2^\diamond;$$

$$(3) X^{\square\diamond} \subseteq X \subseteq X^{\diamond\square};$$

$$(4) B^{\square\diamond} \subseteq B \subseteq B^{\diamond\square};$$

$$(5) X^\square = X^{\square\square\square}, X^\diamond = X^{\diamond\diamond\diamond};$$

$$(6) B^\square = B^{\square\square\square}, B^\diamond = B^{\diamond\diamond\diamond};$$

$$(7) (X_1 \cap X_2)^\square = X_1^\square \cap X_2^\square, (B_1 \cap B_2)^\square = B_1^\square \cap B_2^\square;$$