

属性拓扑理论及其应用

张 涛 著



科学出版社

属性拓扑理论及其应用

张 涛 著



科学出版社

北 京

内 容 简 介

本书以属性拓扑理论及其应用为主线,系统地介绍了属性拓扑基本理论及其应用的最新研究成果.全书分为基础知识、概念计算、关联分析、记忆模型4篇,共13章.基础知识篇着重阐述了属性拓扑基本理论及其基本性质;概念计算篇着重阐述基于属性拓扑的全局形式概念搜索、基于拓扑分解的并行概念计算、增量式概念认知学习以及概念树与概念格的相互转化;关联分析篇着重阐述属性拓扑与频繁关联分析、偏序关联分析和粒关联分析的关系与规则挖掘方法;记忆模型篇定义了记忆属性拓扑的记忆模型,并讨论了基于属性拓扑的记忆激活与遗忘机制.

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、信息科学、计算机、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辅导工具书.

图书在版编目(CIP)数据

属性拓扑理论及其应用/张涛著. —北京:科学出版社,2017.5

ISBN 978-7-03-052760-8

I. ①属… II. ①张… III. ①拓扑 IV. ①O189

中国版本图书馆CIP数据核字(2017)第102605号

责任编辑:周 涵/责任校对:彭 涛

责任印制:张 伟/封面设计:陈 敬

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>

北京九州迅驰传媒文化有限公司 印刷

科学出版社发行 各地新华书店经销

*

2017年5月第 一 版 开本:720×1000 B5

2017年5月第一次印刷 印张:13 3/4 插页:2

字数:278 000

定价:78.00 元

(如有印装质量问题,我社负责调换)

撰写人员名单

(按姓氏笔画排序)

白冬辉 任宏雷 刘梦奇
李 慧 李和合 杨 爽
张 涛 曹海兰 路 静
魏昕宇

前 言

2011年,受洪文学教授邀请,马垣先生到燕山大学讲学,其主要内容是形式概念分析的基础知识.其实,作为一个刚刚进入这个领域的门外汉,我有幸聆听了马先生的全部课程.

在课下的讨论中,马先生提到形式概念分析面临的两大应用障碍:过于抽象和运算复杂度高.恰巧在此之前我做了几年的信息可视化,主要目的是将抽象的信息以直观的形式表现给用户,从而降低用户理解的难度.于是我就有了将概念计算过程可视化的念头.自从有了这个念头,听课过程中便开起了小差.马先生在黑板上一丝不苟地严谨推理,而我却在偷偷地尝试将概念计算过程可视化.

在经历多次失败之后,终于在复杂网络分析中找到了突破点.如果将形式背景中的属性表示成结点,将对象对属性的耦合程度表示成结点的关联,形式背景就可以描述成一个属性之间相互关联的图结构.如果将属性替换为计算机,这样的结构就如同计算机的网络拓扑描述,因此将这个描述模型命名为属性拓扑.

有了属性拓扑的基本表示之后,开始研究如何利用属性拓扑进行概念计算.对最初的设想经历了一系列的修正后,最终将属性拓扑看作一个通信网络,拓扑中的耦合关系看作通信协议,这样概念计算问题就自然转化成了网络中的点到点通信问题和路径搜索问题.于是,基于串行思想的形式概念分析算法逐渐成型,该算法解决了概念计算的可视化问题.

但运算速度问题始终未能解决,几次试图减少运算复杂度未果的情况下转战并行计算.这时候,属性拓扑的网络模型再次发挥作用.反向利用拓扑中的耦合关系,将一个完整的属性拓扑分解为若干个概念上耦合的子拓扑,从而完成全并行的概念计算.经过实验验证,这样的解法尤其适合于大背景下的概念提取.于是,通过并行机制提高概念计算的速度,也算是完成了一次“曲线救国”.

在解决了串行和并行的概念计算问题之后,我们又研究了增量式数据对属性拓扑的影响以及增量式概念计算过程,同样得到了令人振奋的结果.至此,最初利用属性拓扑进行概念计算的初衷得以全部实现.

在进行概念计算的研究过程中,发现借助属性拓扑的图论本质会为其带来更大的应用空间.利用图和树两种数据结构间的转化关系,我们证明了属性拓扑与属性偏序图间的可转化关系,为属性拓扑的偏序关系发现铺平了道路;证明了属性拓扑与频繁模式树的可转化关系,并设计了关联关系发现算法;研究了属性拓扑对属性

聚类特性的描述,实现了“拓扑粒”的描述与计算.这些扩展为属性拓扑今后在数据分析领域的发展奠定了基础.

在研究增量式概念计算过程中,发现该过程与人脑认知过程的相似性,于是开始从记忆模型角度对属性拓扑进行重新解释.在学习了心理学知识之后,利用属性拓扑完成了记忆、唤醒与遗忘的解释,构造出了以属性拓扑为基础的记忆模型.这使得属性拓扑在原有的描述基础上加入了时间轴,可以分析不同时间下概念认知与关联认知的变化过程,符合当前认知计算与类脑计算的研究方向.

时至今日,属性拓扑已经在概念计算、关联分析、记忆模型三个大方向上初露头角.三个方向相互关联、相互促进、相互补充,构成了本书的主体内容.全书由张涛负责统稿,具体章节撰写分工如下:

第1章由张涛、李和合撰写;第2章由白冬辉、张涛、李和合撰写;第3章由李慧、曹海兰撰写;第4章由白冬辉撰写;第5章由曹海兰撰写;第6章由李慧、魏昕宇、李和合撰写;第7章至第10章由魏昕宇撰写;第11章至第13章由杨爽撰写.书稿的校对工作由李和合和刘梦奇负责.

回首属性拓扑的研究之路,过程绝非一帆风顺.在每次遇到迷茫的时候总能有幸得到我的导师洪文学先生的耐心指导,让我得以坚持至今,感谢之情非言语所能表达;在研究过程中刘文远教授的悉心点拨为属性拓扑的应用指明了道路;还需要感谢的是一路陪我走来的研究生团队,在属性拓扑还不被认可的时候,可以顶住压力随我一起潜心研究,并为本书的写作贡献了重要力量.尤其是作为早期成员的路静与任宏雷,为属性拓扑的基础研究做出了重要贡献.

在撰写本书的过程中,我们参考了许多同类著作,吸收了许多观点,在此由衷地表示感谢,并在参考文献中列出.特别感谢马垣教授对该研究提出的宝贵意见和建议.

本书的出版得到了国家自然科学基金项目(项目编号:61603327)、河北省自然科学基金项目(项目编号:F2015203013)、河北省青年拔尖人才支持计划的资助,在此一并表示感谢.

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.

张 涛
燕山大学
2017年3月

目 录

第一篇 基础知识

第 1 章 预备知识	3
1.1 形式概念分析	3
1.1.1 形式背景	3
1.1.2 形式概念	5
1.1.3 形式背景的子背景与形式概念	6
1.1.4 概念格与 Hasse 图	6
1.2 图论	8
1.2.1 图的定义与术语	8
1.2.2 图的存储结构	11
1.3 关联规则	14
1.3.1 关联规则基础	14
1.3.2 关联规则分类	16
1.4 本章小结	18
参考文献	18
第 2 章 属性拓扑的基本理论	21
2.1 形式背景预处理	21
2.2 属性拓扑的定义	22
2.3 属性拓扑的属性分类	27
2.3.1 顶层属性和伴生属性	27
2.3.2 父属性和子属性	28
2.3.3 全局属性、空属性与对等属性	28
2.4 属性拓扑的基础运算	29
2.4.1 增加属性	29
2.4.2 删除属性	31
2.4.3 合并属性	32
2.4.4 交换属性	33
2.4.5 子图合并	33

2.5 属性拓扑的转置: 对象拓扑	34
2.6 决策连续形式背景的离散化	35
2.6.1 数据空间的色度学可视化	35
2.6.2 可视化空间离散化	36
2.6.3 形式背景生成	38
2.7 本章小结	38
参考文献	39

第二篇 概念计算

第 3 章 基于属性拓扑的全局形式概念搜索	43
3.1 算法基础	43
3.1.1 属性拓扑性质	43
3.1.2 属性拓扑的有序化处理	45
3.1.3 结点排序	47
3.1.4 路径的表示方法	47
3.2 结点搜索	50
3.2.1 结点搜索过程	50
3.2.2 结点搜索过程的数据更新	52
3.3 结点回溯	54
3.4 算法总流程	56
3.5 本章小结	57
参考文献	57
第 4 章 基于拓扑分解的并行概念计算	58
4.1 并行概念计算现状	58
4.2 属性拓扑的自下而上分解	59
4.2.1 基于 Upper-set 和 Level 的属性排序	59
4.2.2 属性拓扑的自下而上分解	62
4.3 BDAT 的子属性拓扑约简	64
4.3.1 BDAT 与概念之间的关联	65
4.3.2 BDAT 子拓扑的约简	67
4.4 基于 BDAT 的并行概念计算算法	69
4.5 实验结果与分析	71
4.6 本章小结	75
参考文献	75

第 5 章 增量式概念认知学习	77
5.1 引言	77
5.2 增量式概念学习的形式背景处理	78
5.3 基于 RDFFCS 的增量式概念认知学习	79
5.3.1 新增对象的分类	79
5.3.2 对象拓扑的坍塌	81
5.3.3 递归式深度优先形式概念搜索	84
5.3.4 基于 RDFFCS 的增量式概念更新	87
5.4 基于概念树的增量式概念认知学习	91
5.4.1 路径更新对概念的影响	91
5.4.2 概念更新对概念树的影响	95
5.4.3 基于概念树的概念更新算法	97
5.4.4 算法流程	99
5.5 本章小结	101
参考文献	101
第 6 章 属性拓扑与概念格的双向转化	103
6.1 属性拓扑到概念格的转化	103
6.1.1 概念树的生成	103
6.1.2 属性拓扑到概念格转化关系	104
6.2 概念格到属性拓扑的转化	110
6.3 本章小结	112
参考文献	113

第三篇 关联分析

第 7 章 属性拓扑与频繁关联挖掘	117
7.1 属性拓扑与频繁模式树的二元关系描述	117
7.1.1 形式背景视角下的频繁模式树	117
7.1.2 属性拓扑和 FP-tree 的异同点	120
7.1.3 三种二元关系转化	120
7.2 属性拓扑到频繁模式树转化算法	124
7.3 频繁模式树到属性拓扑转化算法	129
7.4 本章小结	132
参考文献	132

第 8 章 属性拓扑与频繁关联规则	134
8.1 频繁净化形式背景	134
8.2 二元频繁模式挖掘	135
8.3 三元及以上频繁模式挖掘	136
8.3.1 频繁约简属性拓扑	136
8.3.2 BFSX 算法	137
8.3.3 BFSW 算法	138
8.4 算法总体流程	139
8.5 本章小结	141
参考文献	141
第 9 章 属性拓扑与偏序关联规则挖掘	142
9.1 属性偏序二元关系描述	142
9.2 双向转化的数学基础	144
9.3 属性拓扑到属性偏序的二元关系转化	145
9.3.1 属性偏序到属性拓扑的二元关系转化	145
9.3.2 属性偏序到属性拓扑的二元关系转化	148
9.4 从属性拓扑到属性偏序转换算法	148
9.4.1 算法描述	148
9.4.2 算法示例	149
9.5 属性偏序到属性拓扑二元关系转化	151
9.5.1 算法描述	151
9.5.2 算法示例	153
9.6 本章小结	156
参考文献	156
第 10 章 属性拓扑粒度关联规则挖掘	157
10.1 拓扑粒的基本概念	157
10.2 全网络边介数计算	158
10.2.1 单根结点属性权值计算	158
10.2.2 单根结点拓扑边介数计算	159
10.2.3 全网络边介数计算	161
10.3 基于拓扑分裂的属性拓扑粒结构划分算法	161
10.4 属性拓扑粒划分实验	162
10.4.1 属性拓扑的退化	162
10.4.2 粒结构划分	162
10.5 本章小结	164

参考文献	164
------------	-----

第四篇 记 忆 模 型

第 11 章 属性拓扑的记忆模型	167
11.1 引言	167
11.2 当前的记忆模型	168
11.2.1 记忆信息的三级加工模型	168
11.2.2 人脑记忆粒化模型	168
11.2.3 记忆机制的 Object-Attribute-Relation(OAR) 模型	169
11.2.4 遗忘曲线图	170
11.3 属性拓扑的记忆特性分析	170
11.3.1 认知的本质	171
11.3.2 记忆特性与属性拓扑	171
11.4 属性拓扑的记忆模型	172
11.4.1 人脑记忆特性的进一步挖掘	172
11.4.2 属性拓扑的激活模型	174
11.4.3 属性拓扑的遗忘模型	174
11.4.4 几种记忆模型的对比分析	176
11.5 本章小结	177
参考文献	177
第 12 章 属性拓扑的记忆激活机制	180
12.1 属性拓扑激活	180
12.2 属性拓扑激活的认知分析	183
12.2.1 激活集的认知分析	183
12.2.2 新增集的认知分析	184
12.2.3 激活子网络结构的认知分析	186
12.2.4 属性拓扑的激活算法及其认知分析	188
12.3 实验	190
12.4 本章小结	193
参考文献	193
第 13 章 属性拓扑的记忆遗忘机制	195
13.1 属性拓扑的遗忘模型	195
13.1.1 属性拓扑的属性分类阶段	195
13.1.2 属性拓扑的遗忘阶段	197

13.2 遗忘过程中的再刺激·····	198
13.2.1 基本概念·····	199
13.2.2 属性结点的直接刺激·····	200
13.2.3 属性结点的间接刺激·····	201
13.2.4 属性结点的再刺激分析·····	202
13.3 属性拓扑的遗忘算法·····	202
13.4 实验·····	203
13.5 本章小结·····	208
参考文献·····	208

彩图

第一篇

基础知识

第1章 预备知识

属性拓扑是一种基于图论的表示方法,最早来源于对形式概念分析中概念计算的研究,进而在关联规则等方面获得了应用.因此,作为预备知识,本章介绍形式概念分析、图论以及关联规则的部分基本内容.

1.1 形式概念分析

1.1.1 形式背景

形式概念分析是以数学化的概念和概念层次为基础的应用数学领域,它激发了人们对于概念数据分析和知识处理的数学思考.形式背景是形式概念分析的基本表示方法^[1],其定义如下.

定义 1-1^[2] 形式背景可以用三元组 $K = (G, M, I)$ 表示,其中 G 表示所有对象的集合, M 表示所有属性的集合, $I \subseteq G \times M$ 表示对象与属性之间的关系, $G \times M$ 表示集合 G 与集合 M 的笛卡儿积.

定义 1-2^[2] 设 $K = (G, M, I)$ 是一个形式背景,若 $A \subseteq G, B \subseteq M$, 令

$$f(A) = \{m \in M | \forall g \in A, (g, m) \in I\} \quad (1-1)$$

及

$$g(B) = \{g \in G | \forall m \in B, (g, m) \in I\} \quad (1-2)$$

性质 1-1 如果 $K = (G, M, I)$ 是一个形式背景, $A, A_1, A_2 \subseteq G$ 是对象的集合, $B, B_1, B_2 \subseteq M$ 是属性的集合,则下面的一些性质^[1]:

$$A_1 \subseteq A_2 \Rightarrow f(A_2) \subseteq f(A_1)$$

$$B_1 \subseteq B_2 \Rightarrow g(B_2) \subseteq g(B_1)$$

$$A \subseteq g(f(A)), B \subseteq f(g(B))$$

$$f(A) = f(g(f(A))), g(B) = g(f(g(B)))$$

$$A \subseteq g(B) \Leftrightarrow B \subseteq f(A) \Leftrightarrow A \times B \subseteq I$$

表 1-1 为一个形式背景,显然,该形式背景的对象集为 $G = \{1, 2, 3, 4, 5, 6, 7, 8\}$, 属性集为 $M = \{a, b, c, d, e, f, g, h, i\}$. 表 1-1 的符号意义见表 1-2.

表 1-1 生物和水的形式背景

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>	<i>i</i>
1	×	×					×		
2	×	×					×	×	
3	×	×	×				×	×	
4	×		×				×	×	×
5	×	×		×		×			
6	×	×	×	×		×			
7	×		×	×	×				
8	×		×	×		×			

表 1-2 符号表

属性	符号	对象	符号
需要水	<i>a</i>	蚂蝗	1
水里生活	<i>b</i>	娃娃鱼	2
陆地生活	<i>c</i>	蛙	3
有叶绿素	<i>d</i>	狗	4
双子叶	<i>e</i>	水草	5
单子叶	<i>f</i>	芦苇	6
能运动	<i>g</i>	豆	7
有四肢	<i>h</i>	玉米	8
哺乳	<i>i</i>	—	—

由上述形式背景可以看出，形式背景的性质明显成立：

$$f(\{1, 2, 3\}) = \{a, b, g\} \subseteq f(\{2, 3\}) = \{a, b, g, h\}$$

$$g(\{a, b, c\}) = \{3, 6\} \subseteq g(\{a, b\}) = \{1, 2, 3, 5, 6\}$$

$$\{4, 5\} \subseteq g(f(\{4, 5\})) = \{1, 2, 3, 4, 5, 6, 7, 8\}$$

$$\{c, d\} \subseteq f(g(\{c, d\})) = \{a, c, d\}$$

$$f(\{4, 7\}) = f(g(f(\{4, 7\}))) = \{a, c\}$$

$$g(\{b, h\}) = g(f(g(\{b, h\}))) = \{2, 3\}$$

$$\{5, 6, 8\} \subseteq g(\{a, d\}) = \{5, 6, 7, 8\}$$

$$\Leftrightarrow \{a, d\} \subseteq f(\{5, 6, 8\}) = \{a, d, f\}$$

$$\Leftrightarrow \{5, 6, 8\} \times \{a, d\} \subseteq I$$

1.1.2 形式概念

设 $K = (G, M, I)$ 是一个形式背景, 并设 $A \subseteq G, B \subseteq M$, 如果 A, B 满足 $f(A) = B, g(B) = A$, 则称二元组 (A, B) 是形式背景 K 中的一个概念, 并将 A 称为概念 (A, B) 的外延, B 称为概念 (A, B) 的内涵. 由此可见概念是外延和内涵的统一. 通常用 $\mathfrak{B}(G, M, I)$ 或 $\mathfrak{B}(K)$ 表示形式背景 $K = (G, M, I)$ 上的所有概念集合.

定义 1-3 形式背景的全局概念是指以该背景的所有属性为内涵, 对应的对象集为外延构成的概念 (称为全属性全局概念) 或者以所有的对象为外延, 对应的属性集为内涵构成的概念 (称为全对象全局概念). 形式背景 $K = (G, M, I)$ 下的全局概念只有两个, 即 $(G, f(G))$ 和 $(g(M), M)$.

性质 1-2 由性质 1-1 可知, $(g(f(A)), f(A))$ 一定是概念. 又因为如果 (A', B') 是概念, 且 $A \subseteq A'$, 则 $A \subseteq g(f(A)) \subseteq g(f(A'))$, 所以可知 $g(f(A))$ 是包含对象集 A 的最小外延, 于是 $A \subseteq G$ 是外延, 当且仅当 $A = g(f(A))$. 同理可知, $B \subseteq M$ 是内涵, 当且仅当 $B = f(g(B))$.

性质 1-3 两个外延的并集并不一定还是外延, 同样两个内涵的并集也不一定是内涵. 但是, 任意数量的外延的交集一定是还是外延, 任意数量的内涵的交集也一定还是内涵. 即: 若 T 是索引集, 而且对每个 $t \in T, A_t \subseteq G$ 都是对象的集合, 则必有

$$f\left(\bigcup_{t \in T} A_t\right) = \bigcap_{t \in T} f(A_t)$$

由形式概念的定义可知, 表 1-1 所示形式背景中,

$$f(\{1, 2, 3, 4, 5, 6, 7, 8\}) = \{a \in M \mid \forall g \in \{1, 2, 3, 4, 5, 6, 7, 8\}, (g, a) \in I\}$$

并且

$$g(a) = \{g \in G \mid (g, a) \in I\}$$

因此, 二元组 $(12345678, a)$ 为一个形式概念, 其中 $\{1, 2, 3, 4, 5, 6, 7, 8\}$ 为形式概念的外延, $\{a\}$ 为形式概念的内涵. 又因为这个形式概念的外延集合 $\{1, 2, 3, 4, 5, 6, 7, 8\} = G$, 于是, 概念 $(12345678, a)$ 为一个全局概念, 并且是全对象全局概念. 显然可以直接看出该形式背景的全属性全局概念为 $(\emptyset, abcdefghi)$.

依照形式概念的定义, 可以计算出表 1-1 背景下的所有形式概念: $(12345678, a)$, $(\emptyset, abcdefghi)$, $(1234, ag)$, $(34678, ac)$, $(12356, ab)$, $(5678, ad)$, $(234, agh)$, $(123, abg)$, $(36, abc)$, $(678, acd)$, $(568, adf)$, $(56, abdf)$, $(68, acdf)$, $(23, abgh)$, $(34, acgh)$, $(3, abcdgh)$, $(4, acghi)$, $(6, abcdf)$, $(7, acde)$.

这些概念中, 做任意外延的交集, 如概念 $(678, acd)$ 和 $(568, adf)$, $\{6, 7, 8\} \cap \{5, 6, 8\} = \{6, 8\}$, 而对象集合 $\{6, 8\}$ 仍然是外延, 对应的概念为 $(68, acdf)$; 做任意