

《智能信息处理》课程作业

## 基于排序算法特征与关系的形式概念分析

方婉青

作业	分数[20]
得分	

2021 年 11 月 29 日

# 基于排序算法特征与关系的形式概念分析

方婉青

(大连海事大学 信息科学与技术学院 大连 116026)

**摘要** 形式概念分析 (Formal Concept Analysis) 作为一种应用在数据挖掘、信息检索和机器学习等诸多领域的强有力的工具,能够清楚的表达出本体的结构,自引入以来得到了广泛的关注。本文简要介绍了形式概念分析的基本概念、过程和基本方法,以排序算法特征与关系为形式背景进行形式概念分析。

**关键词** 形式概念分析, 概念格, 排序算法

## Formal Concept Analysis Base on the Features and Relationships of Sort Algorithm

Wanqing Fang

( Department of Information Science Technology, Dalian Maritime University, Dalian 116026)

**Abstract** Formal concept analysis, as a powerful tool for data mining, information retrieval and machine learning, has been widely concerned since its introduction. This paper briefly introduces the basic concepts, processes and methods of formal conceptual analysis, and carries on the formal conceptual analysis with the features and relationships of sort algorithms as the background of formal concept analysis.

**Keywords** Formal concept analysis, Concept lattice, Sort algorithm

### 1 绪论

形式概念分析<sup>[1]</sup>,建立在数学基础之上,对概念、属性以及关系等组成概念格的部分用形式化的语境表述出来,然后根据这个语境,构造出概念格,从而清楚地表示出概念格的结构。形式概念分析是与传统数据分析和知识表示完全不同的一种方法,目前是人工智能的重要研究对象,广泛应用于机器学习、数据挖掘、信息检索等众多领域。形式概念分析的逻辑如图一所示。

排序算法是是数据结构最基础的算法之一,现有的排序算法有很多种,归纳出它们具有的特点有着重要的意义。本文通过分析冒泡排序、交换排序、选择排序、插入排序、希尔排序、基数排序快速排序、归并排序、堆排序这九种排序算法,将它们的平均时间复杂度、最好时间复杂度、最差时间复杂度、稳定性、辅助空间为属性集来构建其

形式背景并建立对应的概念格。

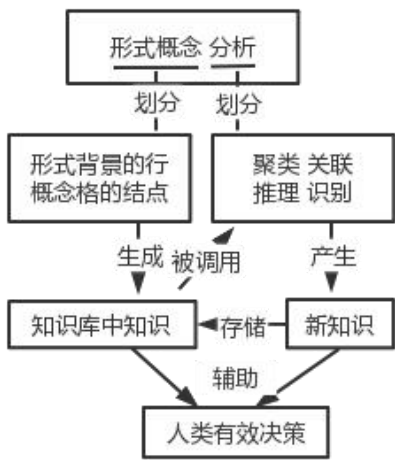


图1 形式概念分析逻辑

## 2 形式背景

形式背景是领域本体的抽象描述，形式概念分析首先要建立形式背景。假定  $G$  为所领域本体中的全部对象， $M$  为刻画这些对象的所有属性的集合， $I \subseteq G \times M$  为  $G$  和  $M$  中元素之间的二元关系，这样得到的一个三元组  $(G, M, I)$  即为该领域本体的形式背景。该三元组可以表示为二维表。表 1 为形式背景的示例，其中  $G = \{g_1, g_2, g_3, g_4\}$  是对象集合， $M = \{a, b, c, d\}$  是属性集合，二元关系  $I$  为确定性关系。

表 1 形式背景

	a	b	c	d	e
1	1	1	1	1	0
2	1	1	1	0	0
3	0	1	1	0	0
4	0	1	0	0	1

表 2 排序算法特征与关系的形式背景

	平均时间 复杂度	最好时间 复杂度	最坏 时间 复杂度	稳 定 性	辅助 空间
冒泡 排序	$O(n)$	$O(n)$	$O(n)$	是	$O(1)$
交换 排序	$O(n)$	$O(n)$	$O(n)$	否	$O(1)$
选择 排序	$O(n)$	$O(n)$	$O(n)$	否	$O(1)$
插入 排序	$O(n)$	$O(n)$	$O(n)$	是	$O(1)$
希尔 排序	$O(n \log n)$	$O(n)$	$O(n)$	否	$O(1)$
基数 排序	$O(d(r+n))$	$O(d(n+rd))$	$O(d(r+n))$	是	$O(n+rd)$
快速 排序	$O(n \log n)$	$O(n \log n)$	$O(n)$	否	$O(n \log n)$
归并 排序	$O(n \log n)$	$O(n \log n)$	$O(n \log n)$	是	$O(n)$
堆排 序	$O(n \log n)$	$O(n \log n)$	$O(n \log n)$	否	$O(1)$

形式背景一般都需要从数据源中提取，不是直接存在的，这就需要对数据源进行科学的分析，采取合理的策略和算法来提取形式背景。

针对九种排序算法特征与关系，提取相关形式背景，结果如表 2 所示。

这个形式背景是多值的，为了简便，把平均时间复杂度、最好时间复杂度、最差时间复杂度为  $O(n)$  的算法认为其时间复杂度高，其余的时间复杂度低。存储空间为  $O(1)$  的认为其存储空间小，其余的存储空间大，并且把时间复杂度低，算法稳定，存储空间小用 1 表示，否则用 0 表示。简化后的形式背景如表 3 所示。

表 3 排序算法特征与关系的简化形式背景

	平均时 间复杂 度	最好时 间复杂 度	最坏时间 复杂度	稳定 性	辅助 空间
冒泡 排序	0	0	0	1	1
交换 排序	0	0	0	0	1
选择 排序	0	0	0	0	1
插入 排序	0	0	0	1	1
希尔 排序	1	0	0	0	1
基数 排序	1	1	1	1	0
快速 排序	1	1	0	0	0
归并 排序	1	1	1	1	0
堆排 序	1	1	1	0	1

## 3 概念格

概念格的构造是形式概念分析的首要任务，构造概念格的过程实际上是概念类聚的过程<sup>[1]</sup>。因此，在概念格中，建格算法具有很重要的地位。对于同一批数据，所生成

的格是唯一的，它不受到数据或属性的排列次序的影响。概念格的建格算法可以分为两类：批处理算法和增量算法。

本文根据排序算法特征与关系的简化形式背景产生对应概念格。首先，根据所给形式背景约减生成单值形式背景，再确定单值形式背景中的父子关系，根据父子继承关系绘制 Hasse 图，最后补充各形式概念的上确界和下确界，形成概念格。

3.1 形式背景的约简

形式背景的约简是形式概念分析中的重要组成部分。通过形式背景约简后，获得的知识更加简洁、清晰，并且反映了属性之间的依赖关系。形式背景的约减包括聚类（行约减）和关联（列约减）。通过观察可以得出，冒泡排序和插入排序、交换排序和选择排序、基数排序与归并排序有相同属性的行，故将其合并。表中并没有相同的列，所以不进行列约减。最后得到约减后的形式背景如表 4。

表 4 排序算法特征与关系的简约形式背景					
	平均 时间 复杂 度	最好 时间 复杂 度	最坏 时间 复杂 度	稳定 性	辅助 空间
冒泡排序/ 插入排序	0	0	0	1	1
交换排序/ 选择排序	0	0	0	0	1
希尔排序	1	0	0	0	1
基数排序/ 归并排序	1	1	1	1	0
快速排序	1	1	0	0	0
堆排序	1	1	1	0	1

3.2 生成单值形式背景

把对象的属性值为 1 的作为默认值，简约后的形式背景变为单值的形式背景。并用 1,2,3,4,5 依次代表各属性，用 a,b,c,d,e,f 依次

代表个对象。那么对象集为{a,b,c,d,e,f}，属性集为{1,2,3,4,5}，单值形式背景如表。单值的形式背景即根据前一步约减后的形式背景，把值为“1”的位置改为“×”，去掉其他位置的“0”以表示该形式对象有此属性。最后得出结果见表 5。

表 5 排序算法特征与关系的单值形式背景					
	1	2	3	4	5
a				×	×
b					×
c	×				×
d	×	×	×	×	
e	×	×			
f	×	×	×		×

3.3 确定父子关系

为了方便构造形式背景的概念格，将单值形式背景转换为带有父子关系（继承）的单值形式背景，也就是基于属性个数比较的排序。例如，表 5 中对象 b 有 1 个属性，属性个数最少，将其放在第一位。对象 a, c, e 有两个属性，放在第二位。对象 d, f 有四个属性，放在最后两位。这样由表 5 得到的带有父子关系的单值形式背景如表 6 所示。表 6 所示为基于属性个数的排序。

表 6 排序算法特征与关系基于属性个数的排序					
	1	2	3	4	5
b					×
e	×	×			
a				×	×
c	×				×
f	×	×	×		×
d	×	×	×	×	

3.4 绘制哈斯图

哈斯图在数学分支序理论中，是用来表示有限偏序集的一种数学图表，它是一种图形形式的对偏序集的传递简约。具体的说，对于偏序集合  $(S, \leq)$ ，把 S 的每个元素表示为平面上的顶点，并绘制从 x 到 y 向上的线段或弧线，只要 y 覆盖 x（就是说，只要  $x < y$  并且没有 z 使得  $x < z < y$ ）。这些弧线可以相互交叉但不能触及任何非其端

点的顶点。带有标注的顶点的这种图唯一确定这个集合的偏序。

哈斯图的作法是以“圆”表示元素；若  $x < y$ ，则  $y$  在  $x$  的上层；若  $y$  覆盖  $x$ ，则连线；不可比的元素在同层。应用哈斯图表示各结点所组成的偏序集及节点间的关系，由上到下表示的即为两节点间的父子关系，根据表 6 所绘哈斯图如图 2 所示。

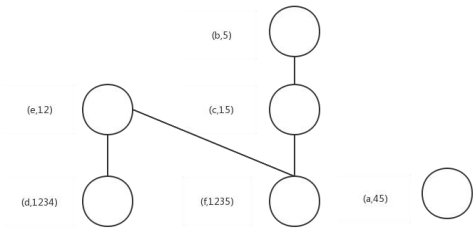


图 2 哈斯图

### 3.5 生成概念格

针对表 6 的约简形式背景，手动生成概念格。图 2 已经给出哈斯图，根据图 2 补出上下确界得到概念格。图 3 是得到的概念格。

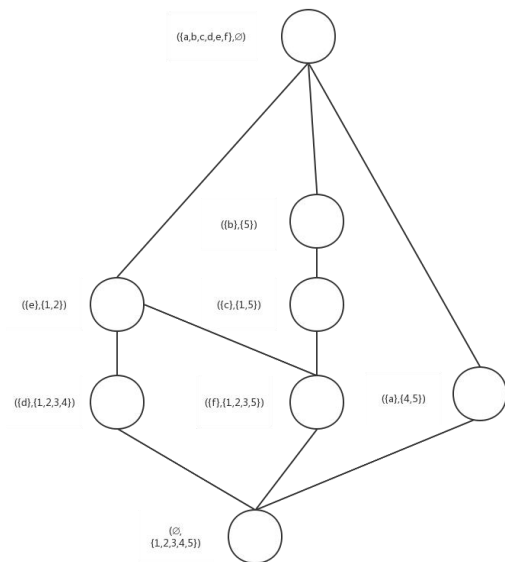


图 3 概念格

## 4 结 论

在使用形式概念分析这一工具的过程中，我们利用已有的排序算法知识，经过形

式化以后能够得到新的知识，利用这些新的知识，我们可以更好地理解已有的知识，学习到新的特征与关系。

随着对形式概念分析的研究的不断深入，之后我们仍然需要学习和把握国内与国外学术界的动态，更好地应用形式概念分析。

## 参 考 文 献

- [1] 智慧来,李逸楠.形式概念分析中的面向对象概念约简[J].海南热带海洋学院学报,2021,28(05):66-71.DOI:10.13307/j.issn.2096-3122.2021.05.09.
- [2] 苏新,陈永平,杨思春.基于对象和基于属性的三支概念格合并方法比较[J].计算机应用与软件,2021,38(11):282-287.
- [3] 张呈玲.概念格模型下属性约简的矩阵方法研究[D].闽南师范大学,2021.DOI:10.27726/d.cnki.gzzsf.2021.000240.
- [4] 张晨.C语言程序设计中选择结构排序算法研究[J].蚌埠学院学报,2021,10(05):54-56+97.DOI:10.13900/j.cnki.jbc.2021.05.012.