## 基于形式概念分析的排序算法

郭佳伟

（大连海事大学 辽宁大连 计算机技术系）

摘要：

近年来,本体作为一种有效的、表现概念层次结构和语义的模型,被越来越多的领域所应用.应该说,形式概念分析的出现能很好地解决目前计算机应用领域中存在的一些困难,如人机交互或机器与机器之间的通信、自动推理、知识表示和重用等.但是,在能很好地应用本体之前,我们面临一个新的难题:概念格的构建.本文对现有的领域概念格构建方法做了具体事例的介绍 ,最后对形式概念分析用于领域本体构建方法做了分析、比较和总结。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语境表示出来，然后根据语境，构造出概念格，即本体，从而清楚的表达出本体的结构。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CLCA，自动产生本体。

关键字：形式概念分析；形式背景；本体；概念格

**Sorting algorithm based on formal concept analysis**

GUO JiaWei

（Dalian Maritime University Dalian, Liaoning Department of Computer Technology）

Abstract: In recent years, ontology as a kind of effective, performance concept hierarchy and semantic model, was applied by more and more fields. It should be said that the emergence of formal concept analysis can well solve some difficulties that exist in the field of computer application, such as human-computer interaction or communication between the machine and machine, automatic reasoning, knowledge representation, and reuse and so on. However, before a good application ontology, we face a new problem: the construction of concept lattice. In this paper, the existing method of building concept lattice to the specific examples, finally, the formal concept analysis is used in the domain ontology building method is analyzed, compared and summarized. Based on the mathematics of formal concept analysis, the composition of the ontology concepts, attributes, and relationships with formal context, and then according to the context, constructs the concept lattice, the ontology, so as to clearly express the structure of ontology. The ontology construction process is half automation, in the formation of the concept phase, which require the participation of domain experts, identified in the field of object, attribute, build their relationship, after the concept generation, can construct the context, and then using the generating algorithm of concept lattice CLCA, automatically generate ontology.

Keywords: Formal concept analysis. In the form of background; Ontology; Concept lattic

# 1 概述

形式概念分析作为一种数学理论于1982年由德国学者Wille首先提出，概念格结构模型是其核心数据结构。概念格本质上描述对象与属性之间的联系，表明概念之间的泛化、特化关系，其相应的Hasse图实现了数据的可视化。目前形式概念分析已被广泛研究并应用到机器学习、软件工程和信息检索等领域。基于形式概念分析的粗糙集模型、基于概念格的多示例集成学习模型、基于概念格的不同粒度下的领域本体模型及形式概念分析在不同粒度下知识获取模型，这些模型不仅在理论上拓展形式概念分析方法，而且对形式概念分析的应用起到积极的推动作用。

# 2.相关定义

## 2.1本体

给出构成相关领域词汇的基本术语和关系，以及利用这些术语和关系的构成的规定这些词汇外延的规则的定义，本体是概念模型的明确的规范说明，是共享概念模型的形式化规范说明，其中基本构建元语：类，关系，函数，公理，实例。概念可以指任何事物；关系表示概念间的相互作用；函数是一种特殊的关系；公理表示永真断言；实例表示元素。

本体，从而清楚地表达出本体的结构。在形式概念分析中，概念的外延被理解为属于这个概念的对象的集合，而内涵则被认为是所有这些对象所共有的特征或属性集，这实现了对概念的哲学理解的形式化。所有的概念连同它们之间的泛化/例化关系构成一个概念格。

## 2.2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。其本质上描述了对象和特征之间的联系，表明了概念之间的泛化与例化关系。

# 3．排序算法形式概念分析

排序算法是数据结构学科经典的内容，其中内部排序现有的算法有很多种，且各有什么特点。本文主要是对内排序的冒泡排序、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、堆排序九种算法为对象集，把算法的平均时间复杂度，最好时间复杂度，最差时间复杂度，稳定性，辅助空间为属性集构建形式背景。



表1 形式背景

此形式背景为多值的，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，其余的认为时间复杂度低。存储空间为O(1)的算法认为其存储空间小，其余的认为存储空间大。并且把时间复杂度低，算法稳定，存储空间小用1表示，把时间复杂度高，算法不稳定，存储空间大用0表示，简化后的形式背景如表2。



表2 简化的形式背景

可以依据基于约简意义的概念格构造方法对其进行简约。把属性值相同的对象进行合并。通过观察，可以看出冒泡排序和插入排序可以合并，交换排序和选择排序可以合并，基数排序和归并排序可以合并。



表3 约简后的形式背景

把对象的属性值为1的作为默认值。把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。则有对象集{a,b,c,d,e,f}，属性集{1,2,3,4,5}。

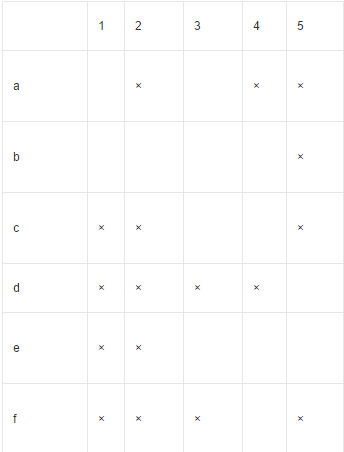


表4 单值形式背景

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（继承关系）的单值形式背景，也就是基于属性个数比较的排序。例如，表4中对象b有1个属性，属性个数最少，将其放在第一位。对象e有两个属性，放在第二位。对象a,c有3个属性，所以将其放在e后面2位。对象d,f有四个属性，放在最后两位。这样由表4得到的带有父子关系的单值形式背景如表5所示。

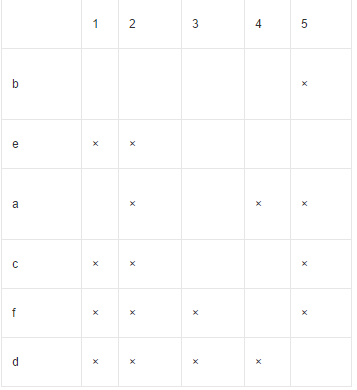


表5

# 4．分析与总结

利用此表通过Chein算法生成概念格，其概念格如图1所示。

图1 排序算法的概念格

通过该概念格，可以发现一些规律：

(1)If1 Then 2

1代表算法的平均时间复杂度，2代表算法的最好时间复杂度。也就是说算法的平均时间复杂度低的时候，它的最好时间复杂度一般也是低的。这和实际意义是相符的，因为算法的最好时间复杂度一般都小于或等于算法的平均时间复杂度，当它的平均时间复杂低的时候，它的最好时间复杂度必然也是低的。

(2)If3 Then1, 2

3代表最坏时间复杂度，该规则说明当算法的最坏时间复杂低的时候，它的最好时间复杂度和平均时间复杂度也是低的。这是因为一个算法的平均时间复杂度和最好时间复杂度都是小于或等于它的最坏时间复杂度的，当最坏时间复杂度低的时候，它的平均时间复杂度和最好时间复杂度也必然是低的。

(3)If4 Then 2

4代表算法的稳定性。该规则表明，当算法是稳定的时候，它的最好时间复杂度是低的。

# 结束语

对基于形式概念分析理论的知识获取模型进行研究，所获研究成果不仅从理论上丰富和发展了形式概念分析，而且由于其广泛的应用背景，这些结果同样具有重要的应用价值。排序算法是数据结构的重要内容，也对编程人员有着重要作用，选择一个合适的算法能提高程序的运行效率。形式概念分析是获得飞速发展的一种分析数据的工具。本文将形式概念分析引入到排序算法的研究，构建9种算法的概念格，从而全面的分析它们的特征和关系，能帮助大家更好的理解和运用这些算法。

# 参考文献

1 王西锋;张晓孪;;形式概念分析在基于事例推理中的应用[J];宝鸡文理学院学报(自然科学版);2006年04期   
2 秦昆;王新洲;张鹏林;傅晓强;;图像数据挖掘软件原型系统的设计与开发[J];测绘信息与工程;2005年06期   
3 秦昆;李振宇;杜鹢;基于概念分析的空间数据挖掘研究进展[J];地球信息科学学报;2009年01期   
4 张文元;秦昆;张成才;杨丽娜;马茹菲;;基于知识的遥感图像地物提取方法研究[J];地理空间信息;2007年01期   
5 贾泽露;刘耀林;;可视化空间数据挖掘研究综述[J];地理空间信息;2009年06期   
6 周涛;张艳宁;袁和金;陆惠玲;;降序加权join半概念格快速挖掘算法[J];计算机工程与应用;2006年29期   
7 王贤敏;牛瑞卿;;三峡库区岩性植被关联规则挖掘[J];计算机工程与应用;2008年31期   
8 王贤敏;牛瑞卿;;三峡库区岩性空间分布规律挖掘[J];计算机科学;2009年06期   
9 吴涛;秦昆;;图像纹理特征数据挖掘的理论与方法探讨[J];计算机时代;2006年08期   
10 杜琳;陈云亮;朱静;;图像数据挖掘研究综述[J];计算机应用与软件;2011年02期