概念格理论研究及其发展

商迪

作业	分数
得分	

2020 年 11 月 13 日

第43卷 第10期 计 算 机 学 报 Vol. 25 No. 6
2020 年 10 月 CHINESE JOURNAL OF COMPUTRS Dec. 2020

概念格理论研究及其发展

商迪

（大连海事大学 计算机技术 辽宁省大连市 中国 116026）

摘 要：概念格是根据对象和属性之间的二元关系建立起来的概念层次结构，又由于概念格能生成 Hasse 示图，这使它能生动、简明地体现出各概念之间的关系．由于其可视性强，又与其他的理论不断结合，而被广泛地应用于软件工程、知识发现等领域。随着理论的发展，近年来概念格在概念分析展示、构造关联等方面的优势越来越明显，与其它学科理论(认知计算、粒计算等)的融合发展也越来越广。本文从概念格理论的起源定义、研究内容、发展等方面总结了概念格的研究进展，还提出了未来的一些发展方向。

关键词：概念格；知识发现；形式概念分析；

中图法分类号：TP311.20 DOI 号： 10.3969/j.issn.1001-3695.2020.01.030

Research and development of conceptual Lattice theory

Di Shang

Department of computer, Dalian Maritime University, City Dalian

Abstract：Concept lattice is a concept hierarchy established according to the binary relationship between objects and attributes, and because concept lattice can generate Hasse diagrams, it can vividly and concisely reflect the relationship between concepts. Due to its strong visibility and constant combination with other theories, it has been widely used in software engineering, knowledge discovery and other fields. With the development of theory, in recent years, concept lattice has more and more obvious advantages in concept analysis, demonstration, construction association and other aspects, and its integration development with other disciplines (cognitive computing, particle computing, etc.) is also more and more extensive. This paper summarizes the research progress of concept lattices from the aspects of origin definition, research content and development of concept lattices, and puts forward some development directions in the future.

Key words：Concept lattice; Knowledge discovery; Formal concept analysis;

0 引言

概念格，也称为 Cralois 格，又叫做形式概念分析，由 Wille R 于 1982 年首先提出，概念格的每个节点是一个形式概念，由两部分组成:外延，即概念所覆盖的实例；内涵，即概念的描述，该概念覆盖实例的共同特征。另外，概念格通过 Hasse 图生动和简洁地体现了这些概念之间的泛化和特化关系。因此，概念格被认为是进行数据分析的有力工具。概念格是形式概念分析理论中的核心数据结构,它利用二元关系建立一种概念间的层次关系,概念格是应用数学的分支，它来源于哲学相关领域内对概念的理解,随着研究的深入，很多学者逐渐认识到概念格自身结构的巨大优势，研究从开始的单纯理论扩展发展到理论与实际应用相结合，并且融合交叉多个相关理论，形式概念分析越来越多地被应用到数据挖掘,信息检索,软件工程等方面，成为许多专家学者关注的热点。

1 概念格基础理论

1.1 基本定理

形式背景（formal context）可以表示为三元组 T=(O，D，R)，其中 O 是事例（对

象）集合，D 是描述符（属性）集合，R 是 O 和 D 之间的一个二元关系，则存在唯一的一个偏序集与之对应，并且这个偏序集产生一种格结构，这种由背景（O，D，R）所诱导的格 L 称为概念格。格 L 中的每个节点是一个序偶(称为概念)，记为(X,Y)，其中 X∈P(O)称为概念的外延；其中 Y∈P(D)称为概念的内涵；每一个序偶关于关系 R 是完备的，即有性质：

1)X={x∈O|∀y∈Y,xRy}

2)Y={y∈D|∀x∈X,xRy}

在概念格节点间能够建立起一种偏序关系。具体地，给定概念 H1=(X1,Y1) 和 H2=(X2,Y2) ，则 H1<H2<==>Y1⊂Y2,领先次序意味着 H1 是 H2 的父节点或称直接泛化。根据偏序关系可生成格的 Hasse 图：如果 H1<H2 并且不存在另一个元素 H3 使得 H1<H3<H2，则从 H1 到 H2 就存在一条边。

概念格将每一个节点表示为一个形式概念，每个形式概念包含概念的外延(extent)和内涵(intent)两部分内容。外延表示此概念所包含的所有对象的集合、即此概念所涵盖的实例，内涵则表示概念中所有对象的共有特征。对于给定的形式背景 K=(G.M.1)(其中 C 为对象集合，M 为属性集合，I 是 G 与 M 之间的一个二元关系)，存在惟一一个偏序集合与之相对应。由偏序集构成一种格结构，并且此偏序集满足自反性、反对称性和传递性。若 g∈C.m∈M.glm 表示对象 g 具有 m 属性。格中的每个节点称之为概念，记作 C(X,Y),X∈G 是概念 C(X,Y)的外延。

Y∈M 是概念中对象的共有属性(内涵)。节点概念与节点概念之间存在着偏序关系，若 C1=(X1,Y1),C2=(X2,Y2),并且 X1<X2 <=>Y1<Y2，称 CI 为 C2 的父节点。

1.2 外延和内涵

概念格的每一个结点都是一个形式概念，并由内涵和外延两部分组成。外延可以理解为属于概念所有对象的集合，即概念所覆盖的实例。内涵是全部对象所拥有的共同特征或属性，即慨念的描述。

外延和内涵是概念的两个方面，他们相互制约又相互联系，并呈现反比关系，内涵越多，外延越少;反之，内涵越少,外延越多。这说明，人们在数据挖掘和分析时，可以利用这种反变规律，通过减少概念内涵来增加概念外延，或者增加概念内涵来减少概念外延。

由于概念格的结点反映出概念外延和内涵的有机统结点间关系体现出了典型的例化和泛化关系。概念格可以通过 Hasse 图简洁生动体现这概念之间的例化和泛化关系，表明一种概念的层次结构，从本质上体现出概念的共同属性(项目、特征)和实体(记录、对象、交易)之间的关系，因而概念格作为一种科学有力的数据挖掘方法，适合于作为规则发现的基础性数据结构。

1.3 优势和特征

概念格是一种核心数据结构，概念格模型根据属性和对象的二元关系进行概念层次结构的构建，并能直接反映出对象和属性的泛化和例化关系，以此更好地在概念层次上建立数据依赖及因果关系模型。概念格数据集的产生必须依托形式背景。常规的数据分析方法通常都存在一个缺陷，在模块化的区域内不能产生一个有效地识别类和对象的指导方案，这些数据分析方法往往会大规模减少给定数据信息，只为获得极少量的“重要参数”。而概念格技术依托数据集中对象和共同属性之间的二元关系，建立概念层次结构，结合格代数理论方法对数据进行分析处理，从而保证了数据信息的最大分解，并最大程度保留了数据之间的各类特殊关系。依托形式背景建立的概念格几乎包含了所有的数据细节，不会因人为因素减少和消除给定数据信息的复杂性，因此概念格在国际数据挖掘和分析领域运用广泛，备受关注。

结合概念格技术的概念定义、优劣对比.运用范围等信息，可以总结归纳出概念格的特征，具体表现在以下三方面:

①形式概念决定了形式背景下具有共同属性的一组最大的对象集合:

②概念格直观明了显示了一种对象和属性二元关系的层次分类;

③概念格中格与子格的关系充分显示出此具体属性是某种抽象属性的不同实例。

1.4 概念格的应用

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。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。在知识发现领域，概念格可以从关系数据中构造出来，然后从概念格上可以提取各种类型的知识，如蕴含规则、关联规则、分类规则等等;在软件工程领域，概念格可以从类库的规范说明上构造，从而对类库结构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虑掉无用的信息。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。

2 研究内容

概念格理论与方法是形式概念分析研究中的基本内容,该研究已取得一系列的重要成果,主要集中在概念格模型推广,概念格构造,概念格约简,基于概念格的规则提取,概念知识空间,概念格的粒计算方法等内容。

2.1 概念格的构造

概念格的构造过程实际上是概念聚类的过程，是应用形式概念分析的前提。通常，概念格的大小是在指数量级上的，而且要处理的数据又多数是海量的，概念格构造算法的研究始终是形式概念分析中的一个主要问题。概念格的构造算法可分为两类：批处理算法和增量算法。使用批处理算法构造概念格要完成两项任务：一是要生成所有的格节点，即所有概念的集合；二是要建立这些格节点间的直接前驱/直接后继关系。按照这两项任务完成次序的不同，我们可以将批处理算法分为任务分割生成模型和任务交叉生成模型。任务分割生成模型是首先生成全部的概念集合，然后再找出这些概念之间的直接前驱/直接后继关系；任务交叉生成模型是在生成概念的过程中同时确定概念之间的关系。增量算法的基本思想是将当前要插入的对象与格中所有的概念求交，根据交的结果进行不同的操作。典型的算法有：Godin 算法、Capineto 算法、AddAtom 算法和 AddIntent 算法。

2.2 概念格约简

概念格的约简能够有效地提高概念格的堆护效率，使形式背景中所蕴含的知识易于发现，简化知识的表示方式。约简概念格实际上是在保持对象集不变的条件下，如何求得最小的属性集的过程。目前对于概念格约简理论，国外研究的理论与方法大都是以从约简行与列的角度来约简属性和对象两个概念为基础，而国内的研究主要是以张文修等提出的理论为基础，给出概念格属性约简的判定定理、引入形式背景的可辨识属性矩阵，并依此为基础求得属性约简的方法，有的文献在此基础上引入了决策形式背景，进一步研究解决了决策形式背景下概念格属性约简方法。

2.3 概念格规则提取

概念格上的规则提取具有广泛的应用前景。规则挖掘是近年来数据挖掘的研究课题，每个概念格节点本质上就是一个最大项目集，为关联规则挖掘提供了平台，体现了概念之间的包含与分类关系，更加易于理解和表示。由于规则本身是由内涵间的关系来描述的，而表现的却是外延之间的包含与被包含关系，正是由于概念节点统一了内涵与外延之间的关系，基于概念格的分类规则的提取在知识发现等方面有着广泛的应用。目前，对于概念格上分类规则的研究主要集中在优化概念格的构建和求解算法上。

2.4 概念知识空间

知识空间理论已成功应用于教育领域，它为构建知识评价体系提供了一种科学的途径。形式概念分析作为知识发现的一种有力工具已被广泛用于信息检索、知识评价等领域。

知识空间理论(简称"KST")是由美国数学心理学家于 1983 年首先提出的一种数学理论。其通过分析学生对不同水平的一系列有关问题解答情况来确定学生在不同知识中的认知水平。基于教育学和心理学等理

论,KST 建立了一套数学理论来反映教育规律，为教育评价提供了一种有效的科学方法，也是-种测试学生知识水平和构建学生知识结构的理论。1990 年 Koppen 和 Doignon 研究了基于专家问询生成知识空间的方法。1993 年 Dowling 在有限知识空间中提出了一种构建知识基的方法，同时研究了根据有限状态族生成知识空间的另种不同方法。

2.5 概念格的粒计算方法

形式概念分析的粒计算方法，是指借助粒计算处理问题的基本思想，分析形式背景中对象与属性之间的依赖关系。粒计算解决问题的基本思想通常包括粒的形成、粒的转化粒的语义解释等。

不同于概念格改胜算法的研究思路,形式概念分析的粒计算方法着重概念个数 ILI 的优化，即概念结果的优化。需要指出的是，一方面,形式概念分析的粒计算方法改变了传统的计算所有概念及其层次结构的研究思路;另一方面，它以具体问题背景为导向优化概念所得结果。简而言之，传统的研究思路是先构建形式背景的所有概念及层次结构,再来发掘它的具体应用；形式概念分析的粒计算方法反过来考虑这件事，即它根据具体问题背景先确定需要哪些概念及结构，甚至有时候这部分概念不包含在传统研究的范围之内，再快速有效找出这些目标概念及结构。

3 概念格的发展前景

自概念格提出以来，就受到了科研工作者的高度重视，这也使它也在众多的领域得到了快速的发展和极其广泛的应用。然而概念格构造理论的研究领域仍就是一个年轻且发展快速的领域，它具有极大的发展空间和应用的潜力。因为概念格所具有的良好的数学结构特性，不仅使得其易于发现数据中对象之间、属性之间的相互依赖关系，而且概念格还有利于从信息中挖掘出知识，从概念格中可以得到对我们有用的规则。也因此使得它在智能数据处理、知识发现、数据挖掘等方面得到了广泛的应用。随着网络技术的普及和发展，概念格在电子商务、Web 服务器管理、个性化导航、入侵检测、搜索引擎等方面也得到了广泛的应用。

现有的规则提取算法，虽然可以从概念格中提取大量的规则，但是往往提取出来的很多规则都是用户不感兴趣的，我们应当研究基于用户兴趣的形式背景知识的约束概念格，使概念格具有指向性和针对性;在未来的发展中，我们应当努力拓广概念格的结点结构，用更简洁的形式表示更丰富的知识，发现处理概念之间关系的更科学计算机方法，进而更好地进行规则提取和知识发现。还应该努力将概念格理论与知识发现、背景知识及实际应用相结合，产生更科学高效的知识发现和规则提取方法，更好地指导人们实践。

4 结语

在知识发现领域，概念格可以从关系数据中构造出来，然后从概念格上可以提取各种类型的知识，如蕴含规则、关联规则、分类规则等等；在软件工程领域，概念格可以从类库的规范说明上构造，从而对类库结构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虑掉无用的信息。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。

如今，概念格以其独特的优势正在赢得越来越多的研究者关注，并在各个应用领域获得了广泛应用。然而这仍是一个高速发展的领域。进步的研究方向包括:高效的建格算法及剪枝算法;从格上产生有用的规则;其他数据挖掘(KDD)任务;多方法的融合等等。

5 参考文献

[1] 降惠.概念格理论研究进展与发展综述[J].办公自动化,2019,24(09):20-23+30.

[2] 王凯，李绍稳，张友华等.概念格理论与方法研究[J].农业网络信息， 2010(03):14-18.

[3] 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).

[4] 张云中, 柳迪, 张原铭. 基于形式概念分析的知识发现研究态势[J].情报科学,2018,036(009):153-158.

[5] 毛华, 康然.概念格理论发展分析[J]. 河北大学学报(自然

科学版), 2015, 35(6):667-672.

[6] 郭伦众.概念格的性质及生成算法的研究[D].2016.

[7] LI Jinjin,SUN Wen.知识空间,形式背景和知识基[J]. 西北大学学报(自然科学版), 2019(4).