

《智能信息处理》课程作业

形式概念分析及其应用

田坤

作业	分数[20]
得分	

2021 年 11 月 29 日

形式概念分析及其应用

田坤

(大连海事大学 信息科学技术学院, 大连 116026)

摘要 形式概念分析是由 Wille 教授于 1982 年首先提出的, 它提供了一种支持数据分析的有效工具。形式概念包括外延和内涵两部分, 它本质上描述了形式对象与形式属性之间的关系。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。从形式背景中生成概念格的过程实质上是一种概念聚类过程。概念格可以用于完成许多机器学习任务, 并且概念格在信息检索、数字图书馆、软件工程和知识发现等方面也有很广泛的应用。本文介绍了形式概念分析的相关概念, 并对形式概念分析的各方面应用进行阐述, 并在此基础上对未来做出展望。

关键词 形式概念分析; 形式背景; 概念格; 形式概念应用

Formal Concept Analysis and Its Application

Tian Kun

(School of Information Science and Technology, Dalian Maritime University, Dalian 116026)

Abstract Formal concept analysis was first proposed by Professor Wille in 1982. It provides an effective tool to support data analysis. The concept of form includes denotation and connotation, which essentially describes the relationship between formal objects and formal attributes. Concept lattice vividly and concisely reflects the generalization and specialization relationship between these concepts through Hasse diagram. The process of generating concept lattice from formal context is essentially a concept clustering process. Concept lattice can be used to complete many machine learning tasks, and it is also widely used in information retrieval, digital library, software engineering and knowledge discovery. This article introduces the related concepts of formal concept analysis, expounds various aspects of the application of formal concept analysis, and looks forward to the future on this basis.

Key words Formal concept analysis; Formal background; Application of formal concepts

1 引言

在计算机与网络信息技术飞速发展的今天, 各个领域的信息与数据急剧增加, 并且由于人类的参与使数据与信息中的不确定性更加显著, 信息与数据中的关系更加复杂。如何从大量杂乱无章和强干扰的数据中挖掘潜在的、新颖的、正确的、有利的价值知识, 这给智能信息处理提出了严峻的挑战, 形式概念分析 (Formal Concept Analysis, FCA) 对于处理复杂的信息不失为一种有效的方法。早在 1940 年 Birkhoff 就已为该方法提供较好的数学理论基础 [1]; 之后, Ganter 等人将其作为一个较好

的数据分析方法, 深化、完善该理论基础, 并将它们扩展到各种现实应用中 [2]。形式概念分析提供了一种较好的层次化 (形式) 对象的分析方法, 它能够识别那些具有共同 (形式) 属性的一组 (形式) 对象的组合 [2]。一方面, 形式概念分析已具备较完善的理论基础, 在应用到其他领域中时, 具有形式化方面的表达能力, 从而具有较强的理论和技术说服力; 另一方面, 形式对象和形式属性这种二元关系经常出现在软件世界中, 对这种二元关系处理的方法也推动形式概念分析技术在其他领域应用的不断发展。由此, 形式概念分析开始从应用角度走向理论形式化研究, 而理论研究将进一步促进形式概念分析在各个领域中的应用。

2 形式概念

2.1 概念的含义及表示

在英文中，概念（concept）一词，其词根词缀为共同（con-）以及拿（-cept-），译为概念、观念、思想，常常指被广泛接受的思想或概念。概念（concept）一词源于拉丁语 conceptus,其过去分词为 concipere。该词的词根词缀意思为共同（com-）以及截取（capere）。而在中文中，概念又由两个字概和念构成。概，是古代一种量具用词，表示用作对古代量具‘斛’的满量状态做出校准。原定义为：量米粟时，使用木板在斗斛上刮平，使其处于一定范围以内，不至于过满，表示对事物做出限定，使其不超出范围。现代字义为：处于一定范围内，如：大概、概念、概括。念，则是令心，心之力，心通思，自然地思维，常思。

而在现代的定义中，概念具有两个基本特征，也就是概念的内涵和外延。因为概念的外延是从其内涵中演绎出来的，所以外延中的所有元素都具有共同的本质属性。概念的内涵就是指这个概念的含义，即该概念所反映的事物对象所特有的属性。概念的内涵为：任意标识或范畴位置，任意标识在范畴结构中的位置认知。或表达成：概念{S/合{正/反}}：合{正/反}是范畴的结构式，其中的正、反、合是三个固定的范畴结构位置，S是任意标识，S可以放到不同的位置上以获得不同的范畴意义而被定义认知为一个有某种抽象意义的概念。概念的外延就是指这个概念所反映的事物对象的范围，即具有概念所反映的属性的事物或对象。概念的外延为：内涵性质和外延种类。概念的外延包括：内涵性质{性质 \vee 范畴}，外延种类{个体 \vee 种类}。概念分为“性质、范畴、个体、种类”共四大基本类型。集合论中某集合的任何元素都可以而且只能属于这四类概念及其他的集合（不包括这个某集合本身）。

自然概念则是把人们认为很自然的事物用概念的形式表现出来。跟概念一样，自然概念也是由对象集以及其对应的属性集构成。自然概念是概念一个子集。自然概念的表示方法与概念的表示方法相同。

概念即对象集以及对象集所具有的属性集的总和。概念的表示方法有三种形式：表达式法、二维表法、图示法，如图1所示。其中，对象集为概

念的外延，而对象集所具有的属性集为概念的内涵。在使用二维表表示概念时，每一个对象及其属性（每一行）都为一个概念。

①表达式法 概念 = （对象集，属性集）

②二维表法

	属性1	属性2	...	属性m
对象1				
...				
对象n				

③图示法 (对象集，属性集)

○

图1 概念的三种表示方法

2.2 形式概念的定义

形式概念分析中的概念来源于哲学。在哲学中，概念被理解为由外延和内涵两部分组成的思想单元。形式概念分析(Formal Concept Analysis, FCA)这一概念是在1982年由 R.Wille 提出的。形式概念分析是应用数学的一个分支，是信息处理、知识处理的一种理论。形式概念即形式对象集以及形式对象集所具有的形式属性集的总和。其定义为：

定义 1^[1] 设 X 为形式对象集 U 的一个子集，B 为形式属性集 A 的一个子集，有二元关系 $R \subseteq U \times A$ ，若 $X = \{x|x \in U, \forall a \in B, xRa\}$ ， $B = \{a|a \in A, \forall x \in X, xRa\}$ ，则二元组 (X, B) 被称为形式概念。

举例如下：（{人，狗}，{哺乳，嗅觉}）。在该形式概念中，人和狗都具有嗅觉，且都是哺乳动物；反过来说，具有嗅觉的是人和狗，是哺乳动物的也是人和狗。

2.3 形式背景

单个的形式对象集以及形式对象集所具有的形式属性集构成了形式概念，形式概念的作用是构建自然概念的层次连通结构。而自然概念则是把人们认为很自然的事物用概念的形式表现出来。人脑对客观世界的认知可以构成许许多多的形式概念机，这就决定着单一的形式概念不足以描述自然界万事万物之间的关系，当不同的形势概念集相互联系时，人们就需要引入新的概念来描述它，将人脑中的形势概念集置于一张二维表中，就构成了形式背景（formal context），如图2所示。

	形式属性1	形式属性2	……	形式属性m
形式对象1	√			√
形式对象2		√		
……			……	
形式对象n	√		√	√

图2 形式背景

在实际应用过程中，一些复杂的形式背景可能会出现一些问题，比如对象或者属性的冗余。例如冗余信息可能表现在两个或多个不同的对象拥有相同的属性，人们在实际应用中把具有相同属性的不同对象合并来进行约简，称为“约简行”；同理，有时候不同的对象之间虽然具有不同的属性，但是属性与属性之间却存在着异构的现象，人们把这样的属性进行合并来进行约简，称为“约简列”；有时候多个属性之间能相互关联，某一个或多个决策属性可以由一个或者多个其他条件属性推出，人们通过“关联规则抽取”来进行约简。通过行或列的约简人们就可以进行信息的合并，从而消除形式背景之间的冗余。

3 概念格

3.1 概念格的概念

格（lattice）的意义是任两个元素的上确界和下确界都存在的偏序集。完备格为任一子集的上确界和下确界存在的偏序集，其特点是只有一个最高点，且只有一个最低点，且图中任何两点连通。概念格是元素为概念的完备格。概念格，也称为Cralois格，它提供了一种支持数据分析的有效工具。概念格的每个节点是一个形式概念，每一个形式概念都是由外延和内涵两部分组成。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。从形式背景中生成概念格的过程实质上是一种概念聚类过程。目前，概念格已被广泛应用于机器学习、模式识别、专家系统、数据挖掘、信息检索等领域^[2]。

3.2 概念格的构建

概念格的表示形式是Hasse图，概念格的构建的基础是形式背景，形式背景描述了多个形式概念之间的关系，单个形式概念描述了形式对象以及形式对象所具有的形式属性之间的关系。所以，概念

格的构建必须明确不同形式对象以及不同形式对象所具有的形式属性。

概念格的构建包含以下几个步骤：生成形式背景，约简形式背景，生成单值形式背景，确定父子关系，绘制Hasse图，补充各形式概念的上确界和下确界，最后获得概念格。

在此我们举一个例子来说明概念格的构建。某公司正常运转需要一些设备，如计算机、打印机、空调、电灯以及专用设备。这些设备所对应的“对象-属性”二元关系如下所示。

计算机 = （咨询，方案，安装，指导，培训，配件，维修，服务）

打印机 = （咨询，安装，指导，配件，维修，服务）

空调 = （咨询，安装，指导，配件，维修，服务）

电灯 = （方案，安装，配件，维修）

专用设备 = （咨询，安装，指导，配件，维修）

这些形式概念组成的形式背景如图3所示。

	计 机 器	打 印 机	空 调	电 灯	专 用 设 备
咨询	Y	Y	Y		Y
方案	Y			Y	
安装	Y	Y	Y	Y	Y
指导	Y	Y	Y		Y
培训	Y				
配件	Y	Y	Y	Y	Y
维修	Y	Y	Y	Y	Y
服务	Y	Y	Y		

图3 形式背景

可以看到，在图3中存在不同行不同列的值相同的情况，这就产生了冗余。在形式背景部分已经提到了信息合并，消除冗余的方法，即约简行、约简列以及关联规则抽取。利用以上概念可以将具有相同属性的打印机、空调合并，将计算机、打印机、空调的共同属性指导以及咨询合并，把所有对象都有的配件、安装以及维修合并，得到约简后的形式背景，如图4所示。

	计算机	打印机/空 调	电 灯	专用设 备
方案	Y		Y	
指导/咨询	Y	Y		Y
培训	Y			

维修/配件/	Y	Y	Y	Y
安装				
服务	Y	Y		

图 4 最终的形式背景

之后，根据最终的形式背景可以通过基于父子关系的渐进式构建，生成 Hasse 图，得到概念格，如图 5 所示。

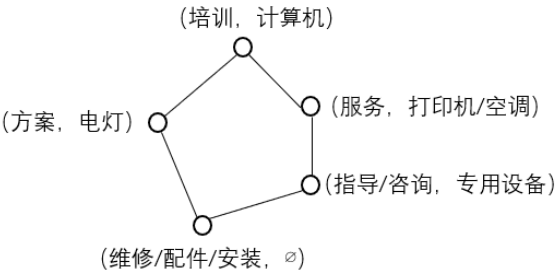


图 5 概念格

3.3 结论分析

根据以上例子可以得知，Hasse 图由于自身的特点可以清晰的表示概念格，Hasse 图的每一个结点都是对象，在图中省略了自反性、省略了箭头，使图像简洁，箭头默认朝下则是基于父子关系的渐进式构建。通过观察 Hasse 图，人们可以更好地了解不同形式对象之间属性的继承关系，使得人们可以快速得到更多有用的信息。对于人们已知的知识，经过概念格的形式进行推演，可以更好地了解不同形式概念之间的关系，帮助人们对系统有更好的认识，从而发现潜在的知识。

4 形式概念的应用

4.1 形式概念分析方法在需求分析中的应用

形式概念这种分析方法是在近年来才兴起的，在这种分析方法出现之后就被广泛性的使用到了各个领域当中。而需求分析则是软件在不同需求当中应用性的分析，这也就表明在分析的过程当中能够具体分析其所使用的各种使用的环境，之后则能够对不同类别软件的应用信息进行有效的整合与管理。所以从大的方向来看，形式概念这种分析方法能够促进在各种实际应用当中软件使用的灵活性和有效性。具体到其实际使用的要求来看，这种分析方法的语境和属性都能够在需求分析当中起到非常重要的作用，因而也就能够为相关和同类别的软件提供一些基础和借鉴。

4.2 形式概念分析方法在结构设计中的应用

软件设计和研发作为一种高科技领域范畴的

工作，主要是对于数据结构进行改革和提升。所以在基本需求分析和管理的的基础之上对其数据进行合理化的组织和分析，这样才能够总结出较为科学和有效的方法。一般来说假设这种方式能够较为合理的表明概念分析方法当中对于概念的构造原理，之后则可以通过有关的分析器来对每个项目进行分析和使用。因而在形式概念分析的过程当中需要具体分析变量之间的关系，这样才能够分析不同项目特征所能够形成的相关概念，这样才能够形成系统化的概念格，也才能够更加有效的促进形式概念这种分析方法在结构设计中的有效使用。

4.3 形式概念分析方法在系统设计中的应用

在对于软件工程系统设计当中，这一阶段最主要的任务和工作就是需要构造出系统的概念格，之后则能够在此基础上出现较多概念格的构造算法。而这种算法是能够更加科学有效的计算出格的概念和不同层次的关系，这样才能够具体分析出不同项目所组成集合的核心内容。在寻找到概念分析法的核心之后就可以较为准确的获得相关概念间的关系。通过对于不同集合关系进行分析和整合就可以获得系统化的概念格。在之后相关专家就可以通过一系列的计算来对概念格的数值范围进行分析和计算，这样才能够促进软件应用的最大化。

5 研究趋势与展望

形式概念分析在软件维护领域的应用在近十几年来取得长足的发展，尽管如此，我们认为如下一些方面仍值得国内外学者关注：

（1）应用领域的深入。形式概念分析技术在程序理解和重构方面已取得大量的研究成果，而形式概念分析在修改影响分析、调试以及测试方面的研究工作目前还处于起步阶段，甚至在修改后的系统验证方面还没有找到相关研究文献，因而形式概念分析技术在这些方面的研究还有待深入。

（2）应用层次之间的可跟踪化。虽然有部分研究人员将形式概念分析应用于需求层次或者设计层次，但各个层次使用形式概念分析技术所提取出来的概念是不同的，但这些概念之间从纵向的角度看是存在一定的可跟踪性，但目前还没有相关的从需求层次到设计层次以及规约层次之间的概念跟踪性研究。

（3）应用工具的支撑。尽管形式概念分析技术已有一些工具支撑，但这些工具主要还是用于概

念格的构造,当这些工具应用到某个具体的领域中还有一些不适应,需进一步修改和扩展,而这方面的工具目前还相当缺乏,所以开发适合某个具体领域形式概念分析的工具很有必要。

(4) 形式概念分析支持软件维护的统一化。形式概念分析可用于各种软件维护活动中,形式概念分析具有较完善的理论基础,研究如何利用该技术统一软件维护的各种活动:从修改(软件)理解到修改实施,修改后测试以及最后的系统验证,这样的研究将会很有意义。

6 总结

从本文中,我们概括性地了解了形式概念分析相关的一些概念和基本方法,并展示了它的一些应用场景。由 Wille R 于 1982 年首先提出的形式概念分析提供了一种支持数据分析的有效工具。形式概念包括外延和内涵两部分,它本质上描述了形式对象与形式属性之间的关系。形式背景则以二维表的形式描述了不同形式对象以及他们的形式属性之间的关系。在此基础上的概念格则是以 Hasse 图的形式,描述了不同形式对象之间的父子关系。概念格结构模型来源于形式概念分析理论,是形式概念分析理论中的核心数据分析工具。

概念格体现了一种概念层次结构,实现了对数据的可视化。因此,概念格受到了人们的广泛关注。概念格理论经过几十年的发展,如今已被广泛运用于软件工程、知识工程、人工智能等领域。本文从概念一词的含义到形式概念的定义及表示,再到形式背景,再到格的解释以及概念格的概念以及应用等方面总结了概念格的研究进展。当然,概念格仍是一个高速发展的领域,对于粗糙模糊概念格的研究、基于概念格的数据挖掘模型的实现等等都是以后的重点研究方向。

参 考 文 献

- [1] 胡可云,陆玉昌,石纯一.概念格及其应用进展[J].清华大学学报(自然科学版),2000,40(9):77-81.
- [2] 张云中.基于形式概念分析的领域本体构建方法研究[D].吉林大学,吉林,2009.
- [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] 朵琳,杨丙.一种基于用户兴趣概念格的推荐评分预测方法[J].小

型微型计算机系统,2020(10).

- [5] 周超,任志宇,毋文超.基于形式概念分析的语义角色挖掘算法[J].计算机科学,2018.

- [6] 朱佳,王向前,张宝隆,等.基于形式概念分析的煤矿事故本体构建[J].工矿自动化,2018,44(05):30-34.