

《智能信息处理》课程作业

形式概念分析的概念及实际应用

刘鑫

作业	分数[20]
得分	

2021 年 11 月 23 日

形式概念分析的概念及实际应用

刘鑫

(大连海事大学 信息科学技术学院, 大连 116026)

摘 要 形式概念分析一般也被称作概念格, 又叫 Cralois 格, 它是由德国的 WhileR 教授于 1982 年提出, 它提供了一种支持数据分析的有效工具, 且被主要用于认知计算、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。本篇介绍了有关形式概念分析的基本概念, 包括形式概念、形式背景、概念格及概念格的生成方法。接着列举了一些近年来形式概念分析的相关应用, 并对这些应用进行了简要的介绍。

关键词 形式概念; 形式背景; 概念格

Concept and practical application of formal concept analysis

LIU Xin

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

Abstract Formal conceptual analysis, also known as concept grid, also known as Cralois, was proposed in 1982 by Professor While R of Germany, which provides an effective tool to support data analysis and is mainly used in cognitive computing, machine learning, pattern recognition, expert systems, decision analysis, web search and other fields. In recent years, some new fields have appeared in the study of concept grid application, such as cognitive concept learning, rule extraction, three decision-making, and so on. This paper introduces the basic concepts of formal conceptual analysis, including formal concepts, formal background, concept grid and concept grid generation methods. Then some applications of formal conceptual analysis in recent years are listed, and these applications are briefly introduced.

Key words formal concept; formal context; concept lattice

1 引言

形式概念分析^[1], 也称为概念格, 是由 Wille R 于 1982 年提出的, 其基本思想是基于对象与属性之间的关系, 根据这一关系来建立一种概念层次结构, 其中每个概念都是对象与属性的统一体。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。作为数据分析和知识处理的形式化工具, 形式概念分析已经被广泛地应用于软件工程、知识工程等领域。

形式概念分析的使用就是把人类已经知道的知识或者常识通过此技术手段发现和挖掘新知识, 从而对原先已经认识的世界知识进行利用然后改造世界。由此可见, 形式概念分析的引入, 对知识

的整合作出了巨大的贡献。本文先介绍了形式概念形式概念、形式背景、概念格的定义, 以及概念格的生成方式, 然后列举了几个形式概念分析的应用。从本文中, 可以感受到概念格在知识分析过程中的突出表现。^[2]

2 形式概念

2.1 概念的含义

概念 (Concept) 是人类在认识过程中, 从感性认识上升到理性认识, 把所感知的事物的共同本质特点抽象出来, 加以概括, 是自我认知意识的一种表达, 形成概念式思维惯性。在人类所认知的思维体系中最基本的构筑单位。

心理学上认为, 概念是人脑对客观事物本质的反映, 这种反映是以词来标示和记载的。概念是思

维活动的结果和产物,同时又是思维活动借以进行的单元。表达概念的语言形式是词或词组。概念都有内涵和外延即其涵义和适用范围。概念随着社会历史和人类认识的发展而变化。

从广义上讲:事物能够改变模型的性质称为事物的概念。

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。

目前有三种概念表示方法形式,分别是表达式法,二维表法,以及图示法。如表 1 所示。

表 1 概念的表示形式

表示方法	具体表示								
表达式法	概念=(对象集, 属性集)								
图示法	(对象集, 属性集)								
二维表法	<table border="1"> <tr> <td></td><td>属性 1 ... 属性 n</td></tr> <tr> <td>对象 1</td><td></td></tr> <tr> <td>...</td><td></td></tr> <tr> <td>对象 n</td><td></td></tr> </table>		属性 1 ... 属性 n	对象 1		...		对象 n	
	属性 1 ... 属性 n								
对象 1									
...									
对象 n									

2.2 形式概念

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^[3] 在一个对象集合 O 和一个属性集合 A 之间的二元关系 $I \subseteq O \times A$ 。元组 $C=(O,A,I)$ 被称为一个形式背景。对于一个对象集合 $O \subseteq O$, 集合的公共属性 $\sigma(O)$ 被定义为:
 $\sigma(O)=\{a \in A | (o,a) \in I, \forall o \in O\}$ 。

(1) 同样的, 对于一个属性集合 $A \subseteq A$, 集合的公共对象 $\tau(A)$ 被定义为: $\tau(A)=\{o \in O | (o,a) \in I, \forall a \in A\}$ 。

(2) 当且仅当 $A=\sigma(O)$ 以及 $O=\tau(A)$ 时可以将一个元组 $c=(O,A)$ 称为是一个概念。对于一个概念 $c=(O,A)$, O 被称为是概念 c 的外延, A 被称为是概念 c 的内涵。

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

2.3 形式背景

定义 2: 一个形式背景 $K=(G,M,I)$,由集合 G , M 以及他们之间的关系组成 I , G 的元素称为对象。 M 的元素称为属性。为了表示一个对象 o 和一个属性 m 在关系 I 中,可以写成 oIm 或 $(o,m) \in I$,读成“对象 o 有属性 m ”。形式背景实例如表 2 所示:

表 2 形式背景定义实例

	A	B	C	D
1	X	X		X
2	X		X	
3		X	X	
4	X	X		X
5	X			

根据定义 2, 一个形式背景能够用一个三线表来表示,表的每一行是一个对象,每一列是一个属性。数字表示对象,字母表示属性,若 g 行 m 列的交叉处是 X ,则表示对象 g 具有属性 m 。

在一些实际应用过程中,一些复杂的形式背景可能会出现冗余,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 概念格定义

概念格是形式概念分析理论中用于数据分析与处理的核心工具,也是一种挖掘数据关联的有效方法。概念格的每个节点是一个形式概念,由两部分组成,分别是外延和内涵,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。从数据集中(概念格中称为形式背景)中生成概念格的过程实质上是一种概念聚类过程;然而，概念格可以用于许多机器学习的任务。目前，已经有了一些建造概念格的算法，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得到应用。

3.2 概念哥构造

假设给定形式背景 (context)为三元组 $T=(O, D, R)$,其中 O 是事例集合 , 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合与之对应 ,并且这个偏序集合产生一种格结构 ,这种由背景 (O, D, R) 所诱导的格 L 就称为一个概念格。格 L 中的每个节点是一个序偶(称为概念),记为 (X, Y) ,其中 $X \in P(O)$ 称为概念的外延; $Y \in P(D)$ 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:

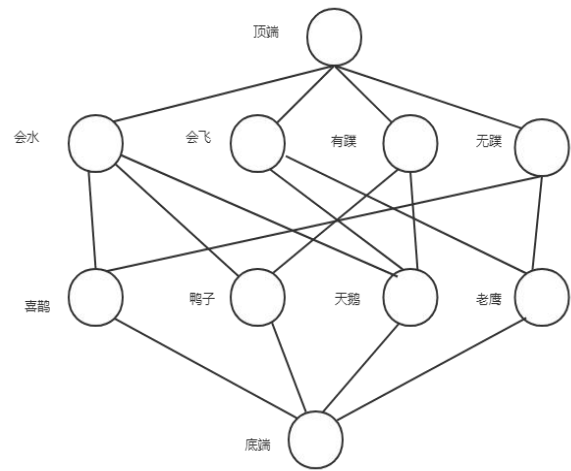
- 1) $X = \{x \in O | y \in Y, x R y\}$;
- 2) $Y = \{y \in D | x \in X, x R y\}$ 。

在概念格节点间能够建立起一种偏序关系。 给定 $H1=(X1, Y1)$ 和 $H2=(X2, Y2)$, 则 $H1 < H2$ $Y1 \leq Y2$,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。表 1 表示了一个形式背景。 其中 $O = \{1, 2, 3, 4\}$, $D = \{a1, a2, a3, b1, b2, b3, c1, c2, d1, d2, d3, d4\}$, R 描述了 O 中元素拥有的 D 中的属性值集。表 3 是形式背景的例子，图 1 是相应的概念格的 Hasse 图。

表 3 形式背景引例

	会水	会飞	有蹼	无蹼
喜鹊		X		X
鸭子	X		X	
天鹅	X	X	X	
老鹰		X		X

图 1 概念格表示



4 概念格的应用

4.1 数字图书馆及文献检索

Neuss 和 Kent 使用概念格进行 Internet 上文档元信息的自动分类和分析。他们讨论的两个著名的资源发现服务 WHOIS++ 和 HARVEST,都可以包含在一个基于概念格的更为一般和具有良好结构的方法中。Cole 和 Eklund 将概念格方法应用于分析和可视化 具有 1962 个属性和 4 000 个处方摘要的医药数据库。他们演示了使用概念缩放来制作嵌套的 Hasse 图 ,对于概念的清晰表达具有很大好处。 作者还 将概念格方法用于分析一组 Email 文档 ,以获得用户感兴趣的信息。 Eklund 和 Martin 展示了概念层次进行 Web 文档索引和导航的能力。他们描述了一个基于概念格的工具 WebKB。 该工具提供了一种形式语言 ,用来描述概念的语义及概念间的层次关系。Kent 和 Bowman 建造了基于概念格的用于 数字图书馆的系统 Nebula 及相应接口。作者认为该系统满足 Richard Furuta 在 Digital Library '95 上提出的数字图书馆系统应该满足的 4 个条件。

4.2 形式概念分析方法在 Web 上的应用

随着我国 Web 应用领域的进一步扩大,使得 人们对于 Web 的质量要求也变得越来越高。在信息 网络技术迅速发展的今天,也导致 Web 的应用以及构造成为了进行软件测评研究的一项重要内容。而在进行 Web 的具体构建过程中还存在有差异性、

分布性以及平台性等诸多特性，其对于 Web 应用软件的进一步发展也有着非常重要的意义[4]。但是 Web 跟互联网技术都具备有开发周期短以及更新速度过快等特点，这也就对其调试以及应用测试等问题提出了更高的挑战。借助于形式概念分析法的应用，能够有效解决 Web 开发过程中的更新速度快这一难题，并能够突破 Web 网页在差异性和分布性上面存在的差异性。此外传统的 Web 网页在应用过程中虽然取得了一定的效果，但是与实际的生活以及应用需求还存在着比较大的差距，并难以充分满足人们的实际需求，比如在传统的自动化测试过程中就无法很好的解决填充表单这一问题。随着互联网技术应用范围的不断变大，使得 Web 网页也开始融入到人们的日常生活之中，并在电子商务、电子教育以及安全性测试等多个环节得到了广泛的应用。通过形式概念分析法能够很好的避免原本测试模式下的局限性从而实现技术以及用户需求的有效统一。

5 结语

通过本篇可以大致了解形式概念分析的基本概念以及基本方法。并介绍了若干的应用领域。可见，形式概念分析对领域认识起着至关重要的作用。在认识一个领域的知识集的时候，系统地整合了各个概念，并以更加规范化的方式呈现出来，方便之后进行知识挖掘。

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] [2]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.
- [3] 许佳卿, 彭鑫, 赵文耘. 一种基于模糊形式概念分析的程序聚类方法[J]. 计算机研究与发展, 2009 (9):11.
- [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).