《智能信息处理》课程考试

基于形式概念分析的本体构建方法

马盼盼

考核	到课[10]	作业[20]	考试[70]	课程成绩[100]
得分				

2020 年 12 月 11 日

基于形式概念分析的本体构建方法

马盼盼

(大连海事大学信息科学技术学院，辽宁大连 110291)

摘 要：形式概念分析的思想主要来源于哲学，而概念格作为形式概念分析的核心数据结构，在规则提取与数据分析方面有着广泛的应用。形式背景是形式概念分析的一个重要元素。利用形式背景生成概念格，再运用概念格的构造算法可自动产生本体。本文采用形式概念分析的方法来构建本体，简述了有关形式概念分析及本体的概念，介绍了本体构造的过程。通过概念格图形的形式来展现本体的研究领域中概念及概念之间的关系，寻找所有隐含概念及概念间的关系，从而清楚的表达出本体的结构。

关键词：形式概念分析；概念格；本体

Ontology construction method based on formal concept analysis

MA PANPAN

(College of Information Science and Technology，Dalian Maritime University，Dalian 110291，China)

Abstract：The idea of formal concept analysis is mainly derived from philosophy. As the core data structure of formal concept analysis, concept lattice has a wide range of applications in rule extraction and data analysis. Formal background is an important element of formal conceptual analysis. Using the formal background to generate the concept lattice, and then using the concept lattice construction algorithm can automatically generate the ontology. This paper uses the method of formal concept analysis to construct the ontology, briefly describes the concept of formal concept analysis and ontology, and introduces the process of ontology construction. Through the form of concept lattice graphics, the concept and the relationship between concepts in the research field of the ontology are displayed, and all implicit concepts and the relations between the concepts are found, so as to clearly express the structure of the ontology.

Key words：formal concept analysis; concept lattice; ontology

0 引言

本体的本质是共享概念模型的明确的形式化的规范说明，其目标[1]是通过概念模型对信息作完全的形式化描述，提供对该领域知识的共同理解，并从不同层次的形式化模式上给出这些信息的互相关系，使计算机可以理解并处理网上的信息。因此，构建本体就成为本体应用的关键的问题，目前该领域的研究还处于探索阶段，没有形成成熟、统一的方法。本文探讨了基于形式概念分析(Formal Concept Analysis，FCA)的理论来构建本体。形式概念分析能从形式背景中发现概念结构，生成概念格，具有明确的层次关系和丰富的语义信息。概念格的 Hasse 图能清晰地表达概念之间的层次关系，即本体的层次，从而清楚地表达出本体的结构。

本文首先介绍了有关形式概念分析及本体的基本概念，又介绍了用形式概念分析来进行本体构建的方法，最后介绍了概念格的相关算法。

1 形式概念分析与本体简介

1.1 形式概念分析

形式概念分析是 20 世纪 90 年代 Wille 提出的一种从形式背景进行数据分析和规则提取的强有力工具[1]，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的语境表述出来，然后根据语境，构造出概念格 (concept lattice)，从而清楚地表达出概念及概念间

关系的结构。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，识别出领域内的对象、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自动产生概念格。形式概念分析强调以人的认知为中心，提供了一种与传统的、统计的数据分析和知识表示完全不同的方法，成为了人工智能学科的重要研究对象，在机器学习、数据挖掘、本体研究、软件工程、知识发现以及 Web 语义检索等领域得到了广泛的应用[2]。现实世界是由各种各样的对象组成的，每个对象都有自己的一组属性或者特征。概念就是指对象、属性以及它们之间的关系，概念反应了对象的特有属性，分为两部分：一部分是对象，一部分是属性集。因此，概念也可以表示为（对象，属性集）的二元组形式。背景是概念的集合，也就是对象集合及其具有的属性的集合。任何一个概念都是从背景中提取出来的一个子集，通常以对象-属性集的二维表表示一个背景，用 1 表示某个对象具有某个属性，而用 0 表示某个对象不具有某个属性。形式概念分析是做为一种数学理论被提出的，是人们组织和分析数据的一种方法，将数据及其结构、本质以及依赖关系进行形象化的一种描述。那么，对现实世界中的概念和背景在形式概念分析时就会形成形式概念和形式背景。

定义 1.1[3] 形式概念：设形式对象集 G，形式属性集 M，二元关系 I　G×M。若 X　G 并且 Y　M，X={ x|x∈G，　y∈Y，xIy }，Y={ y|y∈G，　x∈x，xIy }，则二元组(X,Y)称为形式概念其中 X 称为形式概念的外延，表示属于这个形式概念的对象的集合；Y 称为形式概念的内涵，属于这个形式概念的属性的集合。

定义 1.2 形式背景：三元组 K=(G,M,I)被称为形式背景，其中 G 为形式对象的集合，M 为形式属性的集合，I 是 G 和 M 之间的二元关系，I　G×M。若 g 是 G 中的一个形式 对象，m 是 M 中一个形式属性，那么用（g，m）∈I 表达 g 与 m 之间的关系，读作“形式对象 g 具有形式属性 m”[1]。

定义 1.3 概念格：对于形式背景 H=(G,M,I)存在唯一的一个偏序集与之对应，并且该偏序集的子集的上确 界与下确界都存在，这个偏序集产生的格结构称为概念格。

1.2 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。

本体的目标是获取、描述和表示相关领域的知识，提供对该领域知识的共同理解，确定该领域内共同认可的词汇，并且在不同层次的形式化模式上给出这些词语和词语之间相互关系的明确定义，本体论是反省客观存在的概念模型。用一个公式表示：本体=概念+关系+函数+公理+实例。

2 本体构建方法

本体表示的是现实世界的模型，因此建立的本体必须能够客观反映现实。因此本体的构建应该是一个反复迭代的过程，这个过程将贯穿于本体的整个生命周期[4]。

首先要明确构建的本体将覆盖的专业领域、应将本体的目的、作用以及它的系统开发，维护和应用对象，这些对于领域本体的建立过程中有着很大的关系，所以应当在开发本体前注意，对于特定的专业领域的一些特殊的表达法和特定的详细内容的注释，应当明确。

1）在领域本体创建的初始阶段，尽可能列举出系统想要陈述的或要向用户解释的所有概念。这上面的概念和术语是需要声明或解释的。而不必在意所要表达的概念之间的意思是否重叠，也不要考虑这些概念到底用何种方式（类、属性还是实例）来表达。

对领域文档的预处理，依据“对象-属性”关系，生成领域形式背景，依据单值背景，构造概念格（Hasse 图），将概念格进行转换

处理，得到领域本体概念层次模型

2）设计元本体，定义领域中概念与概念的关系。尽量做到领域无关性，并且包含的元概念数量尽可能少。概念的定义可以使用元本体中定义的元概念，也可重用已有本体。

3）定义类和类层次，领域中的概念主要用类来描述。而对类层次的定义有以下 3 种方法：自上向下法、自下向上法、混合法

4）定义类的属性和属性约束，仅仅定义类还不足以描述整个领域，还要进一步描述类的内部结构。本体中用属性来描述类的内部结构。属性类也有属性，即属性约束。

5）对领域本体编码

6）对领域本体模型进行扩充。对领域本体原型的扩充是在领域专家的参与下完成的。包括：对领域本体属性、实例、公理、非分类关系的扩充。

7)对领域本体的复用。通过本体的映射，发现领域本体间是否存在着关系。若存在，就可以在本体映射的基础上，进行同域本体的合并或者结盟，形成新的领域本体；对概念格的复用，概念格是在构建领域本体的过程中形成的中间产物，在二次开发过程中，可直接应用已有的领域概念格，将之与新的开发过程中的概念格进行合并；对形式背景和领域关键概念的复用

8）本体的检验评价

本体形式化以后，是否满足了我们刚开始提出的需求、是否满足本体的建立准则、本体中的术语是否被清晰定义、本体中的概念及其关系是否完整等问题都需要我们在本体建立过程后进行检验和评估[5]。

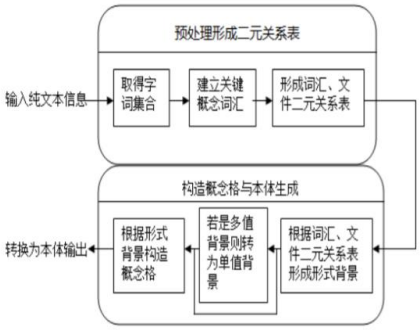

图 1 Hasse 图

3 概念格相关算法

3.1 Uschold 和 King 方法

建立在企业本体基础之上，是相关企业间术语和定义的集合，该方法只提供开发本体的指导方针，目前企业本体在爱丁堡大学人工智能研究所及他的合作伙伴，具体的步骤：

（1）确定本体应用的目的和范围：根据所研究的领域和任务，建立相应的领域本体或过程本体，领域越大，所建立本体越大，因此需限制研究的范围。

（2）本体分析：定义本体所有术语的意义及其之间的关系，该步骤需领域专家的参与，对该领域越了解，所建的本体就越完善。

（3）本体表示：一般用语义模型表示本体。

（4）本体评价：建立本体的评价标准是清晰性、一致性、完善性、可扩展性。清晰性就是本体中的术语应该无二义性；一致性指的是本体中的关系逻辑上应当一致；完整性指的是本体中的概念以及关系应当是完整的，应该包括该领域的所有概念，但是很难达到，需要不断完善；可扩展性本体应用能够扩展，在该领域不断发展时能加入新的概念。

（5）本体的建立：对所有本体按照以上标准进行检验，符合要求的文件的形式存放，否则转到（2）。

3.2 Gruninger 和 Fox 方法

（1）定义直接可能的应用和所有解决方案，并且提供潜在非形式化的对象和关系的语义表示。

（2）能力问题作为约束的条件，能够解决什么样的问题和这个问题应当怎样去解决，这里的问题用属于表示，答案用公理和形式化定义回答，由于在没有形式化 Ontology 之前进行的，所以叫非形式化的能力问题。

（3）术语的规范化：从非形式化能力问题中提取非形式化的术语，然后用 Ontology 形式化语言进行定义。

（4）形式化的能力问题：一旦能力问题脱离了非形式化，Ontology 术语已经定义，则能力问题自然形式化了。

（5）形式化公理：术语定义所遵循的公

理用一阶谓词逻辑表示，其中包括定义和语义或者解释。

（6）说明问题的解决方案必须是完全的。

3.3 Berneras 方法

这种方法开发本体由应用开发控制。所以每一个应用都有相应的表示该应用所需的 Ontology。这些本体既能重用其他的 Ontology，也能被后继应用集成，应用于电子网络的开发。开发过程如下：

（1）应用的说明：提供应用的上下文和应用模型所需要的组件。

（2）相关本体论范畴的初步设计：首先搜索已经存在的 Ontologies，然后进行抽象、提炼、扩充。

（3）Ontology 的构造：最小关联原则用来确保模型既能相互依赖，有尽可能一致，一直得到最大同构。

3.4 Methontology 方法

这种方法由马德里大学工艺分校开发人工智能图书馆使用。它分为三个阶段：

（1）管理阶段：这一阶段的系统规划规则包括任务的进展情况、需要的资源、如何保证质量等问题。

（2）开发阶段：规范说明 y 概念化 y 形式化 y 执行 y 维护。

（3）维护阶段：包括知识的获取、系统的集成、评价、文档的说明与解释、配置的管理与改善。

3.5 基于 Sensus 方法

这个 Ontology 用于自然语言程序，有 ISI 自然语言组企图为机器翻译提供广泛的概念结构，共有 5 万多个概念，为了能在 Sensus 基础上构造特定领域的 Ontology，必须把不相关的术语从 Sensus 中剪出掉，具体过程如下：

（1）定义/叶子术语；

（2）把叶子属于手工的和 Senses 术语相连；

（3）找出叶子节点到 Sensus 根的路；

（4）增加和域相关并且没有出现的概念；

（5）用启发式思维找出全部的特定的域的术语：对于某些有两条以上路经过的结点必须是一颗子树的父节点，那么这颗子树以上的所有结点都和该域相关，是要增加的术语。对于高层结点通常有多条路经过，则很难判断。

4 结语

本体作为一种新的知识组织方式， 力图去解决知识的共享和重利用问题， 在知识越来越丰富的今天， 受到了越来越多的关注， 在许多方面有着广泛的应用前景，许多研究也都相继开展起来。然而， 我们也看到， 基于本体知识库系统理论及应用还处于初步阶段， 其理论和方法还有待于进一步完善。

参考文献：

[1] [德]B.甘特尔,R.威尔.形式概念分析[M].马垣,张学东等译.北京：科学出版社,2007.

[2] R. Wille, "Methods of conceptual knowledge processing", Formal Concept Analysis 4th International Conference ICFCA 2006, vol. 3874, pp. 1-29, February 13-17, 2006.

[3] L. Yang, K. Cormican and M. Yu, "Ontology-based systems engineering: a state-of-the-art review", Comput. Ind., vol. 111, pp. 148-171, 2019

[4] M. N. Asim, M. Wasim, M. U. G. Khan, W. Mahmood and H. M. Abbasi, "A survey of ontology learning techniques and applications", Database, vol. 2018, pp. 1-24, 2018.

[5] 马良荔,孙煜飞,柳青.语义 Web 中的本体匹配研究[J].计算机应用研究,2017,05:1-3.