

形式概念分析在本体构建中的应用

翟正国

(大连海事大学, 信息科学与技术学院, 1120161212)

摘要: 本体作为共享概念的形式化说明, 用于表达数据源的语义、识别和建立概念间的语义关联、达成语义一致, 提供了语义异构问题的解决途径。传统的关系数据库由于具备技术成熟、应用广泛、数据管理能力强、数据安全程度高等特点, 依然是数据存储的主要手段。然而, 关系数据只表达了“属性/值”二元关系, 不具备语义描述的能力, 如何在关系数据中直接抽取数据的语义, 构建本体, 提炼丰富的语义信息, 弥补关系数据表的缺陷, 是一个重要的研究方向。文中探讨了运用形式概念分析的技术, 结合给出的数据, 构建本体的方法。在研究本体构建的基本方法基础上, 论述了形式概念分析方法, 并结合给定数据信息, 演示了运用形式概念分析的方法构建本体的过程。

Using Formal Concept Analysis for Ontology Building

Zhai Zhengguo

(Dalian Maritime University, College of Information Science and Engineering, 1120161212)

Abstract: Ontology by way of formal specification of share concepts, provides a method to resolve the semantic heterogeneous, since it can be used to express the semantics of data sources, identify and construct the semantic relationship of the concepts, compass the consistency. Traditional relational databases is still the main means to store data for its characteristic of mature technique, widely used applications, powerful data management and security. However, relational data only express the “attribute/value” relation, it does not have the semantic describing ability. Therefore, it is an important research orientation that how to take out the semantics in the relational data directly, refine the rich semantic information, construct the ontology, make up the defect of relational data schema. This paper discusses the technique of using formal concept analysis in combination with data model and the data information for ontology construction. On the basis of analyzing present status about building ontology, introduces the formal concept analysis method concerned. Finally, under given data model and data information, demonstrates how to construct an ontology.

Key words: ontology; formal concept analysis; concept lattice

0 引言

本体最早是一个哲学上的概念。从哲学的范畴来说, 本体是客观存在一个系统的解释或说明, 关注的是客观的抽象本质。本体定义: 本体是共享概念模型的明确的形式化规范说明。本体体现的是共同认可的知识, 反映的是相关领域中公认的概念集, 它所针对的是团体而不是个体。本体的目标是捕获相关领域的知识, 提供对该领域知识的共同理解, 确定该领域内共同认可的概念, 并从不同层次的形式化模式上给出这些概念(术语)和概念之间相互关系的明确定义。形式概念分析(formal concept analysis)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), 即本体, 从而清楚地表达出本体的结构。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 CLCA, 自动产生本体。

1 本体的构建

在决定构建本体之前, 必须决定构建什么类型的本体, 怎么构建本体。从描述范围来看, 本体包括领域本体和公共本体^[1]。领域本体

和特定的应用相关,描述了现实世界内小范围的一个模型;相反,公共本体包含公共的概念和关系,可用于不同的应用之中。公共本体作为本体构建的基石,便于扩展、添加新的概念和关系。更进一步,要清楚是采用人工构建还是自动构建的方式。目前已经有一些自动从文本信息构建本体的文献发表^{[3][4]}。其基本思想是:从文本中分离出本体概念;使用自然语言处理的办法,分析句子中两个概念的关系。事实上,在分离概念的时候,需要人工专家的参与,因而也只是实现了半自动化。还要决定是用自底向上的方式构建本体,还是使用自顶向下的方式。本体的构建需要3个阶段^{[5][6]}:

- (1)本体获取,即知识获取的过程。这需要识别目标和范围,发现领域内关键的概念及关系,给出定义描述并用准确的词汇表达出来。
- (2)本体译码,决定领域知识在概念模型中的结构。即把上一阶段获取的概念用形式化的语言明确表示出来。先要确立本体中用到的基本词汇,比如类(classes)、属性(properties)、字面值(facets)等,然后选择合适的知识模型来表达这些词汇。
- (3)本体集成,重用原有本体,加快本体开发。这一步比较困难,因为针对不同的本体甚至是结构完全不同、概念抽取方法迥异的本体,很难识别出公共的概念。如何提供指导或者工具辅助本体集成,是开发本体遇到的一个极大挑战。

2 形式概念分析

形式概念分析^[7]是应用数学和格论的一个分支,它建立在概念和概念层次的数学化基础之上。一个概念就是最大限度地收集对集合中共同特点有帮助的元素,并且运用形式概念分析的方法,可以发现、构造和展示由属性(Attributes)和对象(Objects)构成的概念(Concept)及其之间的关系。因而,形式概念分析的方法已经运用在软件开发等众多环节之中。

2.1 相关概念及定义

定义1:一个形式化的语境(Context) $k=(G, M, I)$, 包括两个稽核(G 和 M)和一个二元关系。在这个语境中, G 中的元素称为对象, M 中

的元素称为属性。一般用 $g I m$, 或者 $(g, m) \in I$ 来表达对象 g 和属性 m 的关系,读作“对象 g 具有属性 m ”。根据定义1,可以通过矩阵来表示语境。每行的开头是对象名,每列的开头是属性名。行 g 和列 m 的交叉表示对象 g 具有属性 m , 见表1。

表1 语境的矩阵表示

属性 \ 对象	m	m	m	m
g	I			
g		I		
g			I	

定义2:对一个对象集 A , 定义 $A' = \{m \in M \mid g I m, \text{ 对所有的 } g \in A\}$ (即 A 中所有的对象共有的属性集合)。相应地, 对一个属性集 B , 定义 $B' = \{g \in G \mid g I m, \text{ 对所有的 } m \in B\}$ (即包含 B 中所有的属性的集合)。

定义3:语境 (G, M, I) 中的形式概念(Formal Concept)是个集合对 (A, B) , 其中 AG, BM , 并且 $A' = B, B' = A$ 。 A, B 分别称作概念 (A, B) 的外延(Extent)和内涵(Intent)。 $\beta(G, M, I)$ 表示语境 (G, M, I) 中的所有概念集。

定义4:如果 $(A_1, B_1), (A_2, B_2)$ 都是语境中的概念, 并且 $A_1 A_2$, 那么 (A_1, B_1) 被称作 (A_2, B_2) 的子概念(Sub concept), (A_2, B_2) 则是 (A_1, B_1) 的超概念(Super concept), 记为 $(A_1, B_1) \leq (A_2, B_2)$ 。“ \leq ”反映了概念间的层次关系。由层次关系搭构的所有 (G, M, I) 的概念记作 $\beta(G, M, I)$, 被叫作概念格^[9](Concept Lattice)。

线路图^[10](Line Diagram)是概念格的图示化表示。线路图包含语境中对象、属性之间的关系,是语境的一个等价表示形式。在特定的语境中,包含类的继承。通过查看线路图,能够容易地发现属性之间的依赖和关联。

2.2 概念格构造算法(CLCA)

概念格构造的算法如下,这里使用一种自底向上的计算方式计算出概念格。

变量说明:

N :语境中单个对象具备属性的最大数目,即是格中的继承层次数目,但不包括上下界所占的层次。

A:集合,其中元素为语境中单个属性或多个属性的集合。

Node:本体节点集。Node(i, j)表示第 i 层上的第 j 个节点。

算法:

```
(1) 计算出 N;
(2) for(int i=N; i>=1; i--)
{
    计算出 A; //A 中元素是属性集, 这些元素中包含
    的属性个数为 i;
    for(j=1; j<size f(A); j++)
    {
        a=A'(j);
        b=A(j);
        if(a'==b&&a==b')
        {
            Node(i, j).attribute list=b;
            Node(i, j).object list=a;
            判断该 attribute list(即 b)是否包含于下
            层节点的 attribute list 中, 确定节点间
            父子关系;
        }
    }
}
```

表 2 给定的对象—属性集

属性 对象	A	B	C	D	E	F	G	H
a					X			X
b					X			X
c			X					X
d	X	X						
e			X					X
f	X							
g	X							
h			X					X

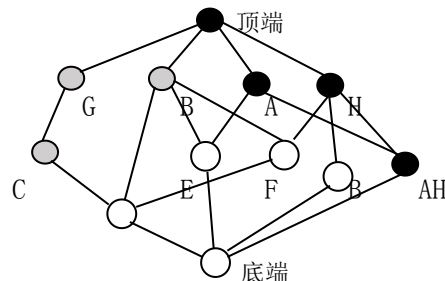
表 2 中:“×”表示连接对象和属性的符号,

参 考 文 献

[1] 蒋平. 形式概念分析在软件工程中的应用

比如说(调试, 测试)可以组成一个对象-属性集, 表示了其在软件工程中操作的优先级别和对象属性。同时, 对于其他部分的开发工作, 也可以通过对其具体的特征集合进行收集。

由以上数据信息得出的概念格如下,



图一 概念格

从图中对象的数量和节点的颜色可以看出, 系统对于每个对象的事件分配是不同的。颜色最暗的表示最需要紧急处理的事件或者过程, 稍微暗些的表示可以通过其它环节进行预先或者推迟处理的过程, 最后则是按照系统正常开发次序进行完成的工作。

3 结束语

在介绍了 FCA 的方法之后, 在演示给定数据库模式和数据信息的基础上, 运用 FCA 来构建本体——概念格。从概念格的线路图上, 可以清楚地浏览概念之间的层次关系。

要构建本体, 应先分离出概念, 建立概念之间的关系, 这里离不开领域专家的参与; 然后可以构建数据库模式(或者建立 XML 文件, 再转换成关系数据表); 其次还要做一个多值语境到单值语境的转换, 这一步必须对数据信息的数值做分类处理, 可以是人工的, 也可以借助工具; 最后运用 CLCA 算法, 自底向上地自动构建概念格。

可以看出, 虽然本体的开发过程依然离不开人的因素, 但概念格作为本体的构建方式, 清楚表达了概念以及概念之间的关系, 而且容易为人们所理解。CLCA 算法也给本体构建提供了指引。

[J]. 计算机技术与发展, 2008.

[2] 胡可云, 陆玉昌, 石纯一. 概念格及其应用扩展[J]. 清华大学学报: 自然科学版, 2000, 40(9): 77~81.

[3] 胡可云, 陆玉昌, 石纯一. 概念格及其应

- 用扩展[J]. 清华大学学报: 自然科学版, 2000, 40(9): 77~81.
- [4] 谢润. 概念格建格算法研究[D]. 西南交通大学, 2006.
- [5] 孙小兵. 形式概念分析在软件维护中的应用综述[D]. 扬州大学, 2010.
- [6] 刘树鹏, 李冠宇. 基于形式概念分析的本体合并方法[J]. 计算机工程与设计, 2011(04): 09-121.
- [7] 乌弘毅, 黄应辉. 模糊改您格构建的 BORDAT 方法 [J]. 计算机技术与发展, 2010(10): 15-21.
- [8] 蒋平. 概念格构建系统的设计与时间[J]. 计算机技术与发展, 2010(10): 01-05.
- [9] StummeG, Maedche A. FCA-merge:bottom-up merging of ontologies[C]//17th Artificial Intelligence(IJCAI' 01).Germany:Springer, 2001 :225-230.
- [10] Uschold M.Ontologies:Principles, Methods and Applications[J] .The Knowledge Engineering Review , 1996, 11(2):93-120.
- [11] Tilley T, Cole R, Becker P, Survey of Formal Concept Analysis Support for Software Engineering Activities[C]In:Stumme G.Proceedings of the First International Conference on Formal Concept Analysis - ICFCA' 03. [s. l.] :[s. n.] , 2003:119 -124 .
-

