

基于概念格属性约简的本体合并方法

胡蕊

(大连海事大学 信息科学技术学院, 辽宁 大连 110310)

摘要: 提出一种基于概念格属性约简的本体合并方法。该方法先对源本体中的概念进行外延和内涵的分析, 以此为形式背景构造概念格。对所构造的概念格进行属性约简, 删除本体合并中的冗余信息, 约简绝对不必要属性, 以最少信息精确地反映概念格及本体的本质和结构。最后通过实例分析此方法的可行性。

关键词: 概念格; 属性约简; 本体合并

中图法分类号 TP393 **文献标识码** A

Ontology Merging Method Based on Attributes Reduction of Concept Lattice

HU Rui

(Faculty of Information Science and Technology, Dalian Maritime University, Dalian 110310, China)

Abstract: This paper presents an ontology merging method based on concept lattice's attribute reduction, which first conducts an analysis on the extension and connotation of the concept in the original sources. With this background, this paper constructs the concept lattice, then conducts attribute reduction on the concept lattice, removes redundant information on the ontology merging, and reduces any unnecessary attribute. All this is done to use the minimum amount of information that accurately reflects the nature and structure of the concept lattice and the body. Finally, an example analysis is used to prove the feasibility of this method.

Keywords: concept lattice; attribute reduction; ontology merging

1 引言

由于本体在知识共享和知识重用中的核心作用, 本体在知识工程和语义 Web 领域得到广泛应用。目前已建好的本体一般覆盖同一领域的不同方面, 内容上常交叉重复, 而用户在解决某一问题时往往需要一个特定领域的多方面知识, 这就需要将分布的、异构的资源进行合并, 因此本体合并也是必然的发展趋势。

本体合并是指将相同或相似领域内已存在的本体合并在一起, 消除重叠的和不协调的部分。本体合并的输入信息为同一领域的多个不同的本体, 它们作为合并的源本体。而输出信息则是合并后的全局本体即目标本体。目前基于本体合并的方法较少, 本文提出一种基于概念格属性约简的本体合并方法, 在合并过程中利用概念格属性约简知识, 删除冗余信息, 约简绝对不必要属性, 以最少信息精确地体现了概念格及本体的结构与本质。

概念格属性约简是在保持形式背景上所有概念的外延集不变的前提下, 寻找极小属性子集, 该属性子集依然能够完全确定形式背景上的原有概念, 并保持它们之间原有的层次结构关系。

2 基于概念格属性约简的本体合并方法

2.1 本体与概念格及属性约简的基本概念

2.1.1 形式背景

概念格分析的数据一般用形式背景来表示, 形式背景的定义如下。

称 (U, A, I) 为一个形式背景, 其中

$U = \{x_1, \dots, x_n\}$ 为对象集, $A = \{a_1, \dots, a_m\}$ 为属性集, I 为 U 和 A 之间的二元关系, 即 $I \subseteq U \times A$ 。若 $(x, a) \in I$, 则称 x 具有属性 a , 记为 xIa 。

2.1.2 概念格

概念格, 又称为 Galois 格, 是德国数学家 Wille R. 于 1982 年首次提出的。概念格是根据数据集中对象与属性之间的二元关系建立的

一种概念层次结构，体现了概念之间的泛化和特化关系。作为数据分析和知识处理的有力工具，概念格理论已被广泛应用于知识工程、数据挖掘、信息检索、软件工程等领域。

许多定义共享了关于本体最核心的部分：本体概念以及本体概念之间的 is—a 关系。概念格体现的是一个层次结构，其中每一个节点均表示为一个概念，每个概念都还有其各自的属性和从上层概念中继承来的属性。概念格中，概念之间存在一种偏序关系，从某种程度上来说，本体概念间的 is—a 层次关系和概念格中的偏序关系非常相似。

2.1.3 约简

对于形式背景 (U, A, I) ，如果存在属性集 $D \subseteq A$ ，使得 $L(U, D, I_D) \cong L(U, A, I)$ ，则称 D 是 (U, A, I) 的协调集。若进一步 $\forall d \in D$ ， $L(U, D - \{d\}, I_{D - \{d\}})$ 与 $L(U, A, I)$ 不同构，则称 D 是 (U, A, I) 的约简。 (U, A, I) 的所有约简的交集称为 (U, A, I) 的核心。

设形式背景 (U, A, I) 的所有约简为 $\{D_i \mid D_i \text{ 是约简}, i \in \tau\}$ (τ 为一个指标集)。可将属性集 A 分为以下 3 部分。

- 1) 绝对必要属性 (核心属性) $b: b \in \bigcap_{i \in \tau} D_i$ 。
- 2) 相对必要属性 $c: c \in \bigcup_{i \in \tau} D_i - \bigcap_{i \in \tau} D_i$ 。
- 3) 绝对不必要属性 $d: d \in A - \bigcup_{i \in \tau} D_i$ 。

其中，非核心中的属性，称之为不必要属性 $e: e \in A - \bigcap_{i \in \tau} D_i$ 。它要么是相对必要属性，要么是绝对不必要属性。对于绝对必要属性 b ，相对必要属性 c ，绝对不必要属性 d ，显然 $b^* \neq c^*$ ， $c^* \neq d^*$ ， $b^* \neq d^*$ 。

在定义中，根据属性与约简的关系，属性被分为 3 类：绝对必要属性 (核心属性)、相对必要属性、绝对不必要属性。不同类型的属性在概念格约简中所起的作用是不同的，且不同类型的属性有不同的特征。

设 (U, A, I) 是形式背景， $\forall a \in A$ ，记 $G_a = \{g \mid g \in A, g^* \supset a^*\}$ 。下列命题成立：

- 1) a 是核心属性 $\Leftrightarrow (a^{**} - \{a\})^* \neq a^*$ ；
- 2) a 是相对必要属性 $\Leftrightarrow (a^{**} - \{a\})^* = a^*$ ，且 $G_a^* \neq a^*$ ；
- 3) a 是绝对不必要属性 $\Leftrightarrow (a^{**} - \{a\})^* = a^*$ ，且 $G_a^* = a^*$ 。

2.1.4 本体

本体 X 如图(1)所示， X_1 拥有属性 a_1 ， X_2 和 X_3 都是 X_1 的子概念，因此 X_2 和 X_3 也都拥有属性 a_1 。而 X_2 和 X_3 又分别拥有各自的属性 a_2 和 a_3 。以该本体 X 中的概念作为形式背景下的对象，以概念中所拥有的属性作为形式背景下中对象所拥有的属性，采用经典的 Godin 算法来构建相应的概念格。

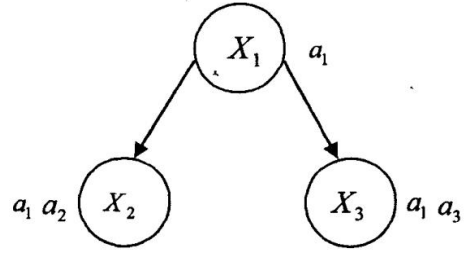


图 1 本体 X 的层次结构

图 2 表示的是本体 X 所对应的概念格的 Hasse 图。从图中看出，概念格也是层次结构，每一个节点均表示为一个概念，下层概念继承上层概念的属性。因此概念格中的概念关系也属于一个偏序关系，与本体概念中的 is—a 关系极其相似。

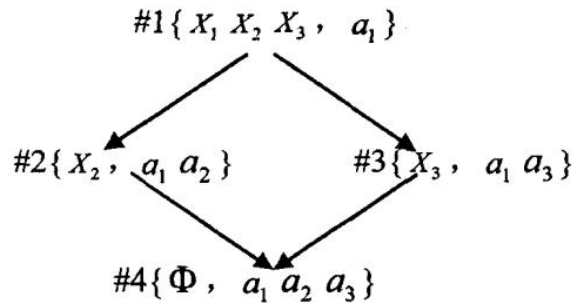


图 2 本体 X 所对应的概念格的 Hasse 图

2.2 本体合并算法

根据概念格的合并理论，分析两个形式背景中会出现以下三种情况：

- ①如果两个概念 m, n ，存在概念 m is—a 概念 n ，则在合并后的形式背景中只出现 n 。即：如果 $(g, m) \in I$ 并且 m is—a n 意味着 $(g, n) \in I$ ；
- ②如果两个形式背景中存在两个名称不同但意义相同的本体概念如 a, b ，那么可以在合并后的形式背景中以 “ $a \setminus b$ ” 来表示这个概念，意思就是 a 或者 b ；
- ③如果两个形式背景中存在两个名称、含义均不同，但它们从领域文档中的分布一样如 p, q ，那么在整理后的形式背景中以 $p(q)$ 来表示，意思是 p 且 q 。本文采用合并的两个源本体中，只要其中一个本体的某一概念拥有该属性，则合

并后，目标本体中也一定有一概念拥有该属性的原则来进行形式背景的合并。

本文采用基于一一映射的概念格属性约简算法对概念格进行属性约简。算法 1 如下：

记形式背景 (U, A, I) 所有概念组成的集合为 $AL = \{(X, B) | (X, B) \in L(U, A, I)\}$ ，而所有概念的内涵组成的集合为

$AI = \{B | (X, B) \in AL\}$ ；对于任意属性集 $D \subseteq A$ ，记集合 $AID = \{B \cap D | B \in AI\}$ 。

算法 1 基于一一映射的概念格属性约简算法
输入 形式背景 (U, A, I)

输出 A 的一个相对约简

- (1) 令 $E=A$;
- (2) 若存在属性 $a \in E$ ，令 $D = E / \{a\}$ 使得映射 $f: EI \mapsto EID$ 为一一映射，转 (3)；否则转 (4)；
- (3) 选择满足 $f: EI \mapsto EID$ 为一一映射的属性集 $D = E / \{a\}$ ，执行 $E \leftarrow D$ ，并转 (2)；
- (4) 结束，输出结果 E

基于概念格属性约简的本体合并方法主要分为以下四个步骤：

(1) 对本体进行预处理：首先对两个源本体进行概念和属性的分析，并得出其中所有概念和属性的包含关系，同时可以得出相对应的形式背景。（下面例子中将表示出来）

(2) 根据 (1) 所得出的形式背景，对两个形式背景进行比较分析，可得出两个源本体合并以后所对应的总的形式背景。

(3) 构造概念格：通过 (2) 所得到概念和属性之间的配对关系所对应的形式背景，采用经典的 Godin 方法来构建概念格。

(4) 概念格属性约简：利用概念格属性约简以及算法 1 来约简属性，优化概念格中冗余的包含关系，对与用户需求无关紧要的绝对不必要属性进行删除，来优化概念格的结构。

3 实例与分析

下面以两个源本体合并为例来验证基于概念格的本体合并方法的可行性与有效性。

第一步：分别对两个本体 X_1 和 X_2 进行分析得出以下两个形式背景。

表 1 为本体 1 所对应的形式背景。其中形式背景中 1、2、3、5 均表示本体中的概念，而其中的 a、c、d、e、f、g、h、i 均表示本体的属性。若对象拥有某一属性，即用 1 表示，否则就用 0 表示。

表 1 本体 1 所对应的形式背景

	a	c	d	f	g	h	i
1	1	1	0	0	1	0	1
2	1	0	0	0	1	0	0
3	1	0	1	1	0	1	0
5	0	0	0	0	1	0	0

表 2 为本体 2 所对应的形式背景。其中形式背景中 3、4、5 均表示本体中的概念，而其中的 b、c、d、e、f、g、h 均表示本体的属性。若对象拥有某一属性，即用 1 表示，否则就用 0 表示。

表 2 本体 2 所对应的形式背景

	b	c	d	e	f	g	h
3	0	0	1	0	0	1	0
4	1	1	0	0	1	0	1
5	1	0	0	1	0	1	0

第二步：由上面两个源本体合并过程对其形式背景进行分析，两个形式背景会出现三种情况，出于对一般情况的分析与研究，充分的考虑到各对象所拥有的属性，因此可以得出合并后的形式背景：

表 3 表示的是两个源本体和合并后所对应的形式背景。而其中的 a、b、c、d、e、f、g、h、i 均表示两本体合并后的属性。若对象拥有某一属性，即用 1 表示，否则就用 0 表示。

表 3 两个源本体和合并后所对应的形式背景

	a	b	c	d	e	f	g	h	i
1	1	0	1	0	0	1	0	1	
2	1	0	1	0	0	0	1	0	1
3	1	0	0	1	0	0	1	0	1
4	0	1	1	0	0	1	0	1	0
5	0	1	0	0	1	0	1	0	0

第三步：根据表 3 中概念 1、2、3、4、5 以及所拥有的属性 {a、b、c、d、e、f、g、h、i} 应用经典的 Godin 算法构建概念格，如下：

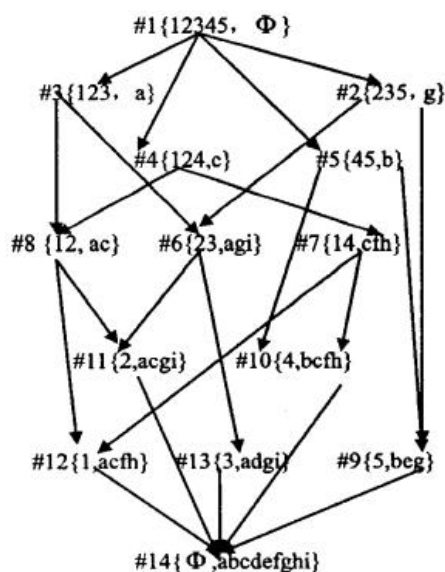


图3 本体合并后概念格的Hasse图

图3所得出的概念格结构比较粗糙, 还需要利用概念格的属性约简对该概念格进行优化, 对概念格中的冗余信息进行合并, 对其中的绝对不必要属性进行删除, 这样能以最少的信息精确地反映概念格以及本体的本质和结构。

第四步: 利用算法1 计算后得到表3 形式背景的属性约简集为{a、b、c、d、f、g}, 其对应的概念格的Hasse图如下图4。

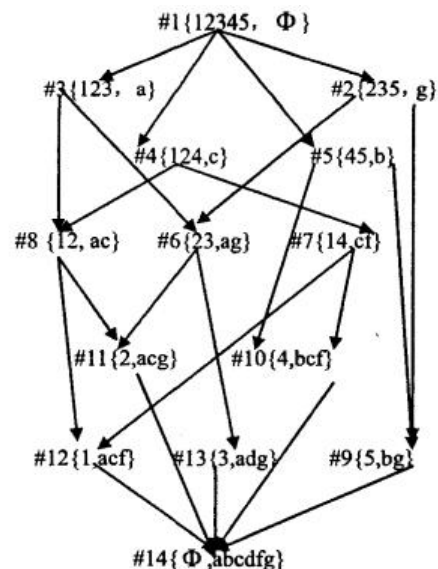


图4 属性约简后概念格的Hasse图

图4是对图3进行概念格的属性约简, 因此图3和图4中的概念格是同构, 其所对应的本体结构和语义是一致的, 拥有语义一致性和结构一致性。但是图4所对应的本体属性明显要比图3对应的本体的属性要少(少了属性e、h、

i), 以最少的信息精确的体现了本体的结构, 对于本体的结构也是一个优化。

4 结束语

文中利用概念格的属性约简的基本知识为已存在的源本体合并提供了一种方法和理论支持, 利于本体的应用与发展。通过属性约简, 对本体合并中的冗余信息进行合并, 对不必要属性进行删除, 优化了本体的结构, 并用实例证明该方法具有可行性和有效性。本文仅考虑两个源本体的情况。对于多个源本体的情况, 可转化为两个源本体的情况(两两组合即可, 只是需要反复进行)。

参考文献

- [1]张瑞玲, 白桂梅, 徐红升, 沈夏炯. 基于FCA的本体的构建与合并. 微电子学与计算机, 2008, (7): 40—43, 47.
- [2]张文修, 魏玲, 祁建军. 概念格的属性约简理论. 中国科学(E辑), 2005, 35(6): 628—639.
- [3]李云, 刘宗田, 陈岭等. 多概念格的横向合并算法. 电子学报, 2004, 32(11): 1849—1854.
- [4]李金海, 吕跃进. 一种新颖的概念格属性约简算法. 计算机工程与应用, 2008, 44(20): 148—151.
- [5]R.Wille. Restructuring lattice theory:an approach based on hierarchies of concepts[J]. in Ordered Sets, I.Rival, Ed.Reidel, Dordrecht, 1982, pp: 445—470.
- [6]Godin R, Mineau G, Missaoui R. Incremental structuring of knowledge bases[C]. Proc. of KRUSE. 1995, 95: 179—193.
- [7]李艳霞, 史一民, 李冠宇. 基于概念格的K—Means 算法研究[J]. 计算机工程与设计, 2011, 32(2): 656—658.
- [8]Skowron A. Extracting laws from decision tables: a rough set approach[J]. Computational Intelligence, 1995, 11(2): 371—388.
- [9]Berry M W, Drmac Z, Jessup E R. Matrices, vector spaces, and information retrieval[J]. SIAM review, 1999, 41(2): 335—362.
- [10]Carpineto C, Romano G. Information retrieval through hybrid navigation of lattice representations[J]. International Journal of Human-Computer Studies, 1996, 45(5): 553—578.