

一种基于概念格胶合的模糊本体合并方法^{*}

李兆海, 李冠宇

(大连海事大学 信息科学技术学院 计算机科学与技术系, 辽宁 大连 116026)

摘要: 为了解决本体异构问题, 实现本体资源的共享和重用, 需要进行本体集成。本体合并是集成的一种方式。本体对概念外延和内涵揭示得不够, 对概念的表达深度不够。形式概念分析可以很好地表示和处理概念以及概念之间的关系, 但是在语言表达和推理等方面还很不足, 因此需要将它们结合起来表达和处理知识, 提高语义表达能力。提出了一种基于模糊概念格胶合的模糊本体合并方法, 为模糊本体集成提供了一种新的方法。

关键词: 形式概念分析; 模糊概念格; 概念格胶合; 模糊本体; 模糊本体合并

中图分类号: TP391.1 **文献标志码:** A **文章编号:** 1001-3695(2012)07-2517-03

doi:10.3969/j.issn.1001-3695.2012.07.030

Concept lattice gluing based fuzzy ontology merging method

LI Zhao-hai, LI Guan-yu

(Dept. of Computer Science & Technology, College of Information Science & Technology, Dalian Maritime University, Dalian Liaoning 116026, China)

Abstract: In order to solve the problem of heterogeneity and achieve the goal of sharing and reusing resource, ontology should be integrated from different sources. Ontology merging is a way of ontology integration. Ontology has its disadvantages for the expression of concept because it's not well enough to express the extent and intent of concept. Formal concept analysis could solve this problem, but it has difficulty in coding and reasoning. It may be a better way to combine formal concept analysis with ontology. This paper proposed a novel method of fuzzy ontology merge based on fuzzy concept gluing. The processing of fuzzy ontology merging transformed to the processing of fuzzy concept lattice gluing, shows a new way of fuzzy ontology integration in the semantic Web.

Key words: formal concept analysis(FCA); fuzzy concept lattice; concept lattice gluing; fuzzy ontology; fuzzy ontology merging

语义网是对当前万维网的一个扩展, 用机器可理解的方式来描述信息。在语义网中, 通过本体(ontology)来对信息进行描述。本体是概念的明确的形式化表达, 描述了概念以及概念之间的关系。然而客观世界中的信息往往是不精确的, 如“年轻人很喜欢吃很辣的食物”, 这里的“年轻人”、“很喜欢”“很辣”都是一些模糊的词语, 以往的精确本体不能够处理这些不精确信息, 需要引入模糊本体来进行描述。由于本体构建语言、构建方法及表达方式等方面的不同导致了本体的异构, 因此需要对本体进行集成, 即本体集成(ontology integration)。本体合并(ontology merging)是本体集成的一种最有效的方式, 是把两个或者多个本体合并生成一个新的本体, 以达到知识共享和重用为目的。

本体仍然存在许多不足, 一是本体的粒度不易控制, 二是本体所表达概念的深度不够, 对概念的内涵和外延方面揭示得不够, 而且本体的形式化程度、一般化程度都不高^[1]。因此, 需要引入形式概念分析来解决上述问题。形式概念分析(FCA)是对概念的哲学理解的数学化表达, 是数据的内在结构、关联和相互依存关系的可视化方法。但形式概念分析在编码、推理能力等方面仍有很多不足, 因此需要将两者结合, 才能更好地进行知识表达和处理。特别的, 本体的合并、集成问题

可以转换成概念格的胶合、分解问题来处理。因此, 本文把概念格的胶合应用到模糊本体合并中来, 提出了一个基于概念格胶合的模糊本体合并模型。

1 背景知识

1.1 模糊本体

由于精确本体模型不能表示不精确信息, 如书“很贵”、食物“很辣”等, 因此需要引入模糊本体来表示模糊信息。Straccia^[2]最先提出模糊本体的思想, 引入模糊逻辑到本体模型中, 可以表示模糊概念以及概念之间的模糊关系。目前有很多方法可以产生模糊本体, 如基于模糊理论和形式概念分析的FOGA^[3]方法, 通过构建模糊形式背景产生模糊概念格, 概念聚类产生层次关系; 最后生成模糊本体。Lv等人^[4]结合模糊集理论从关系数据库模式产生模糊本体模型。Zhang等人^[5]提出一种从模糊UML模型中产生模糊本体的方法。

构建模糊本体的工具可以使用Bobillo等人^[6]开发的Protégé插件。该插件可以可视化地创建和编辑模糊本体, 表示成OWL2的形式, 并可以转换成一些本体推理工具支持的语言进行推理。

定义1 模糊本体可以表示成六元组的形式 $O_f = (C_f,$

收稿日期: 2011-11-30; 修回日期: 2012-01-04 基金项目: 国家自然科学基金资助项目(60972090, 61073057)

作者简介: 李兆海(1988-), 男, 云南临沧人, 硕士研究生, CCF会员, 主要研究方向智能信息处理(zhaohai_seaboy@126.com); 李冠宇(1963-), 男, 辽宁丹东人, 教授, 博士, 主要研究方向智能信息处理。

CP_f, R, RP, I, A_f)。其中:

a) C_f 是模糊概念集,概念也可以叫做类。

b) CP_f 是模糊概念的属性集合,集合中的每个元素可以看做是一个三元组 $CP_f = (c_f, v_f, l_f)$ 。这里 c_f 是模糊本体中的一个概念, v_f 是属性的值, l_f 表示限定词。比如“空气很湿润”,概念“空气”的属性值是“湿润”,限定词是“很”。

c) R 是概念间的模糊关系的集合。

d) RP 是概念间关系的属性集合。

e) I 是模糊概念的实例集合。

f) A_f 是公理的集合,或者模糊规则的集合。

1.2 模糊本体合并

为了解决本体异构,实现本体资源的共享和重用,需要进行本体集成^[7]。本体合并是本体集成的一种方式。本体合并是把两个或多个本体合并成一个本体的过程。目前已有很多精确的本体合并或者集成的方法,映射方法主要有基于语法的方法,如基于编辑距离的方法、基于概念结构的方法、基于机器学习的方法。

目前国内对于模糊本体的合并或者集成的研究还很少, Nguyen 等人^[8,9]提出了一种基于共识(consensus)的模糊本体集成方法,分别在概念层和关系层上进行集成,并提出冲突解决办法。但是文中给出的算法没有经过实验验证,无法保证其正确性和适用性。

1.3 模糊形式概念分析

模糊形式概念分析,首先根据领域知识构建模糊形式背景,然后根据概念格生成算法将此模糊形式背景生成模糊概念格,这一过程称为模糊形式概念分析(fuzzy formal concept analysis)。

定义2 一个模糊形式背景可以表示成三元组的形式 $K = (O, M, I = \varphi(O \times M))$ 。其中: O 是对象集; M 是属性集; I 是 $O \times M$ 上的模糊子集。每一个关系 $(o, m) \in I$ 有一个 $[0, 1]$ 上的隶属度值 $\mu(o, m)$ 。

定义3 模糊形式背景 K 中的每一个对象 o 都可以表示为 $\Phi(o) = \{p_1(\mu_1), p_2(\mu_2), \dots, p_n(\mu_n)\}$ 。其中: μ_i 是属性 p_i 属于对象 o 的隶属度, $\Phi(o)$ 称为对象 o 的模糊表示。

定义4 给定一个模糊形式背景 $K = (O, M, I)$, 定义 $A' = \{m \in M \mid \forall o \in A: \mu(o, m) \geq \gamma\}$, 其中 $A \subseteq O$; 定义 $B' = \{o \in B \mid \forall m \in M: \mu(o, m) \geq \gamma\}$, 其中 $B \subseteq M$, 则称 $(\varphi(A), B)$ 为模糊形式背景 K 上隶属度为 γ 的一个形式概念, 其中 $A \subseteq O, B \subseteq M, A' = B, B' = A$ 。对于每一个对象 $o \in \varphi(A)$, 均有一个隶属度 $\mu(o)$ 定义为

$$\mu_o = \min_{m \in B} \mu(o, m)$$

其中: $\mu(o, m)$ 是对象 o 和属性值 m 的隶属度值。如果 B 为空集, 则 $\mu(o) = 1$, 那么 A 和 B 分别称为形式概念 $(\varphi(A), B)$ 的外延和内涵。

定义5 给定形式概念 $C_1 = (A_1, B_1)$ 和 $C_2 = (A_2, B_2)$, 如果满足 $A_1 \subseteq A_2$, 且满足偏序关系 $(A_1, B_1) \leq (A_2, B_2)$, 给定模糊形式背景 K 中的所有模糊概念, 通过此偏序关系构建出来的完全格称做模糊概念格。

例如, 给定一个模糊形式背景, 如表 1 所示, 有五个对象、七个属性值, 每行表示对象含有某属性的概率。

表 1 一个模糊形式背景

	a	b	c	d	e	f	g
1	0.9	0	0.6	0	0.9	0	0
2	0.7	0	0.6	0	0.9	0	0
3	0.6	0	0	0.9	0.8	0	0
4	0	0.8	0.6	0	0	0.6	0.6
5	0	0.6	0.7	0	0	0.6	0.5

根据渐进式算法构建出模糊概念格, 如图 1 所示。

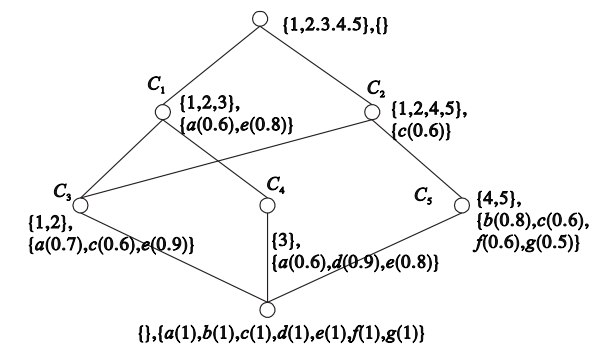


图 1 模糊概念格

1.4 概念格胶合

概念格胶合(concept lattice gluing)是把一些具有相同结构的部分“粘”在一起, 形成一个更大格的一种方法^[10]。

定义6 当下列条件成立时, 格 V 是它的两个子格 V_1 和 V_2 的理想过滤胶合:

- $V = V_1 \cap V_2$;
- V 中若有 $X \leq Y$, 则 $\{X, Y\} \subseteq V_1$ 或 $\{X, Y\} \subseteq V_2$;
- 可在 V 中找到元素 a 与 b , 使 $V_1 \cap V_2 = \{a\} \cup \{b\}$ 。

定义7 形式背景 $K_1 = (U_1, M_1, I_1)$ 与 $K_2 = (U_2, M_2, I_2)$ 的并集是 $K_1 \cup K_2 = (U_1 \cup U_2, M_1 \cup M_2, I_1 \cup I_2)$ 。如满足以下条件:

- $U_0 = U_1 \cap U_2$ 是 $K_1 \cup K_2$ 的外延;
- $M_0 = M_1 \cap M_2$ 是 $K_1 \cup K_2$ 的内涵;
- $I_0 = I_1 \cap I_2 = I_1 \cap U_0 \times M_0 = I_2 \cap U_0 \times M_0$;

就称 $K_1 \cup K_2$ 为形式背景 K_1 与 K_2 的胶合。

两个概念格的理想过滤胶合一定同构于形式背景胶合的概念格。实际上, 通常把两个没有共同元素的概念格做一些处理后才能把它们胶合在一起。首先修改形式背景 K_1 与 K_2 的概念, 使得两个背景中的概念一致; 然后再修改属性, 使得属性的概念一致。如果这些概念之间建立了同构映射, 则可以把这些概念中的属性和对象分别改成相同的名字。这样修改后的背景胶合的概念格就是它们概念格的理想过滤胶合。

概念格胶合也可以叫做概念格合并。目前国内外概念格合并大多数都是基于形式背景的合并再生成概念格, 很少在概念格基础上直接合并。

李云等人^[11]提出了概念格的纵向合并和横向合并的方法。概念格的横向合并指对象集相同、属性集不同的概念格之间的合并, 而纵向合并指属性集相同、对象集不同的概念格之间的合并。智慧来^[12]研究了概念格构造与应用中的关键技术, 统一了概念格的横向和纵向合并算法, 提出了统一的概念格合并理论。赵玉峰^[13]总结了经典形式背景横向合并的方法, 提出模糊形式背景横向合并生成模糊概念的方法。

2 本体和形式概念分析的关系

本体和形式概念分析都是知识建模强有力的工具,都有着独特的理论基础和基本方法。

本体和形式概念分析在知识建模中各有优势^[14],如表2所示。本体的形式化程度和一般化程度不够高,对外延与内涵揭示的深度不够高。形式概念分析通过概念的外延和内涵之间的联系生成概念格,通过概念格可以很清楚地知道概念之间的关系,挖掘出概念之间的隐含关系;但在编码、推理中还有很多缺陷。因此,应该将本体与形式概念分析融合在一起,以更好表达和处理知识。可以通过形式背景的胶合、分解或者概念格的胶合、分解来完成对本体的合并、集成等操作。

表2 本体和形式概念分析的比较

内容	本体	形式概念分析
建模对象	现实世界,可无数据	人工世界,有数据(形式背景)
完备性	不完备	比较完备
描述语言	RDF, RDFS, OWL 等	无
推理能力	很强	较弱
建模过程	自顶向下	自低向上

本体表示概念及概念间的关系,其中表达概念关系最核心的是 Kind-of, 可以证明它也是一个偏序关系,因此可以实现本体到概念格间的互相映射。概念格的节点、节点属性集、节点对象集、节点间的层次关系分别可以映射成本体中的类、类的属性集、实例和类间的关系。对本体的一些操作可以转换成对概念格这一精确数学结构进行处理,如本体的映射、合并的操作可以转换成对概念格的映射、分解、合并等操作。

3 基于概念格胶合的模糊本体合并模型

基于上述分析,本文提出一种基于概念格胶合的模糊本体合并方法,如图2所示。首先将源本体和目标本体分别转换成单值模糊形式背景;然后生成模糊概念格,在此基础上进行模糊概念格胶合;最后生成目标模糊本体。

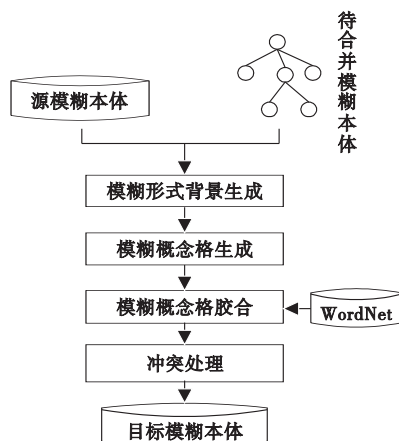


图2 模糊本体合并模型

首先将要合并的两个模糊本体转换成模糊形式背景,对应关系如表3所示。

然后根据概念格渐进式构造算法构造出模糊概念格,源模糊本体对应的概念格结构不变,与待合并模糊本体对应的概念格进行胶合。模糊概念格胶合算法 FCLG 算法如下:

FCLG Algorithm
 Input: sourceV, revisionV
 Output: targetV

```

1 Input sourceV and revisionV
2 while i < total concepts in revisionV do
3   while j < total concepts in sourceV do
4     s1 = sim(Ni, Cj);
5     if(s1 > 0)
6       merge(Ni, Cj);
7     elseif(hasRealitionInWordNet(Ni, Cj))
8       merge1(Ni, Cj);
9     elseif(sims(Ni, Cj) > λ)
10      merge(Ni, Cj);
11    end if
12    i++; j++;
13  end while
14 end while
  
```

表3 模糊本体和模糊形式背景对应关系

模糊本体	模糊形式背景
实例	对象(行)
属性	属性(列)
隶属度	属性值

对输入的待合并的模糊概念格中的每个节点的概念进行遍历,在源模糊本体对应的模糊概念格中从上到下进行查找,首先根据式(1)计算格相似度。

$$\text{sim}(C, N) = \frac{\sum_{k=1}^n (C_k \cap N) / (C_k \cup N)}{n} \quad (1)$$

其中: C 为源模糊概念格中的概念; N 为待合并概念格中的概念; n 指以概念 C 为根节点的子树的总节点数; C_k 为概念 C 为根节点子树中的第 k 个概念。

例如,一个节点 N 的外延是 $\{1, 2, 5\}$,内涵是 $\{a(0.5), c(0.8)\}$,源本体对应的模糊概念格如图1所示,现在比较节点 C_1 和 C_2 的相似度:

$$\begin{aligned} \text{sim}(C_1, N) &= \frac{0.5}{0.6+0.8+0.8} + \frac{0.5+0.6}{0.7+0.8+0.9} + \frac{0.5}{0.6+0.9+0.8+0.9} = 0.28 \\ \text{sim}(C_2, N) &= \frac{0.6}{0.6+0.8} + \frac{0.5+0.6}{0.7+0.8+0.9} + \frac{0.6}{0.5+0.8+0.8+0.6+0.5} = 0.36 \end{aligned}$$

如果相似度大于阈值,则把这个概念合并到源模糊本体对应的概念格中,否则在 WordNet 中查找这两个概念的关系,如果是同义词或近义词,则合并;如果是上位关系,则插入到比较节点的父节点;如果是下位关系,则插入成子节点;如果是兄弟关系,则插入成兄弟节点。

如果在 WordNet 查找不到上述关系,则计算这两个概念结构的相似度,即:

$$\text{sims}(c_1, c_2) = \max(0, \frac{\min(|c_1|, |c_2|) - \text{ms}(c_1, c_2)}{\min(|c_1|, |c_2|)}) \quad (2)$$

其中: $|c_i|$ 指概念 c_i 的词的长度, $\text{ms}(c_i, c_j)$ 指词 c_i 到 c_j 转换所需的最少步数。

如果计算出的结构相似度大于阈值,则把这两个概念合并;否则舍弃该节点。

最后将生成的模糊概念格转换成模糊本体,模糊概念格中的概念对应模糊本体中的类;模糊概念格中的属性对应模糊本体中的属性;对象集对应类的实例;层次关系对应类的层次关系;隶属度对应模糊本体中的属性值。

4 结束语

语义网中用本体表示概念及概念间的关 (下转第 2533 页)

(上接第 2519 页)系,引入模糊本体可以处理不精确信息。由于本体的构建团队不同等造成本体异构,因此需要进行本体集成。本文通过实现本体到概念格间的互相映射,将本体合并、本体集问题转换成概念格胶合、分解的处理问题,并把概念格胶合应用于模糊本体合并,提出了一种基于模糊概念格胶合的模糊本体合并方法,为模糊本体集成提供了一种新的思路。

参考文献:

- [1] 徐宝祥,张云中. 信息系统建模理论发展研究[J]. 情报杂志, 2010, 29(5): 70-74.
- [2] STRACCIA U. Towards a fuzzy description logic for the semantic Web (preliminary report) [C]//Proc of the 2nd European Semantic Web Conference. 2005:167-181.
- [3] THO Q T, HUI S C, FONG A C M, *et al.* Automatic fuzzy ontology generation for semantic Web[J]. IEEE Trans on Knowledge and Data Engineering, 2006, 18(6): 842-856.
- [4] LV Yan-hui, MA Z M, ZHANG Fu. A fuzzy ontology generation framework from relational schema[C]//Proc of 5th International Conference on Fuzzy Systems and Knowledge Discovery. 2008:435-439.
- [5] ZHANG Fu, MA Z M, CHENG Jing-wei, *et al.* Fuzzy semantic Web ontology learning from fuzzy UML model [C]// Proc of the 18th ACM Conference on Information and Knowledge Management. 2009:1007-1016.
- [6] BOBILLO F, STRACCIA U. Fuzzy ontology representation using OWL 2 [J]. International Journal of Approximate Reasoning, 2011, 52(7): 1073-1094.
- [7] 于娟, 党延忠. 本体集成研究综述[J]. 计算机科学, 2008, 35(7): 9-13.
- [8] NGUYEN N T, TRUONG H B. A consensus-based method for fuzzy ontology integration [C]//Proc of the 2nd International Conference on Computational Collective Intelligence. 2010: 480-489.
- [9] DUONG T H, NGUYEN N T, KOZIERKIEWICZ-HETMANSKA A, *et al.* Fuzzy ontology integration using consensus to solve conflicts on concept level [C]//Proc of ACIIDS Posters. 2011:33-42.
- [10] 马垣, 曾子维, 迟呈英, 等. 形式概念及其新进展 [M]. 北京: 科学出版社, 2010: 170-181.
- [11] 李云, 刘宗田. 多概念格的横向合并算法 [J]. 电子学报, 2004, 32(11): 1849-1854.
- [12] 智慧来. 概念格构造与应用中的关键技术研究 [D]. 上海: 上海大学, 2010.
- [13] 赵玉锋. 形式背景横向合并概念生成 [D]. 西安: 西北大学, 2010.
- [14] 张云中. Ontology 和 FCA 在知识建模中的融合机理研究 [J]. 现代图书情报技术, 2010(3): 40-46.