

基于范畴论的形式化地理本体对齐和集成研究

胡玲¹ 李霖² 王红³

(内江师范学院计算机科学学院 内江 641100)¹ (武汉大学资源与环境学院 武汉 430079)²
(湖北大学资源环境学院 武汉 430062)³

摘要 本体对齐可解决本体间的异构,而范畴论可屏蔽本体间的异构性并为本体集成提供统一的方法框架。将范畴论引入本体对齐研究领域,在范畴论的基础上,结合地理本体特征,对态射进行了重新定义,构造了一个更复杂的范畴。其中范畴中的对象为地理本体,而态射不仅能表达相等关系,也可以描述地理本体间的多种语义关系。在态射定义的基础上,根据态射的特性定义了本体对齐和本体对齐的复合,最后对本体集成进行了阐述。

关键词 范畴,态射,本体对齐,复合,外推

中图分类号 TP301.2 **文献标识码** A

Formalizing Geo-ontology Alignment and Integration Based on Category Theory

HU Ling¹ LI Lin² WANG Hong³

(College of Computer Science, Neijiang Normal University, Neijiang 641100, China)¹

(School of Resources and Environment Science, Wuhan University, Wuhan 430079, China)²

(Faculty of Resources and Environment Science, Hubei University, Wuhan 430062, China)³

Abstract Ontology alignment can solve the heterogeneity between ontology, while category theory can shield the heterogeneity between ontologies and provide uniform methodological framework for ontology integration. This paper applied category theory into the field of ontology alignment. Based on the features of geo-ontology it redefined morphism and intended to construct a more complex category, where objects are geo-ontologies, but morphism can not only express equal relationship but also describe multiple semantic relationships between geo-ontologies. Based on the definition and the features of morphism, this paper also defined ontology alignment and composing alignment, and elaborated ontology integration.

Keywords Category theory, Morphism, Ontology alignment, Composing, Pushout

1 引言

使用地理数据的行业的业务需要受各部门的体制、技术和其他因素的影响,不同的部门建立了各自的地理信息系统,然而其缺乏统一的组织和模式管理,导致数据难以共享。地理信息系统融合的关键是语义异构,语义异构主要是指不同的地理信息源的概念和类别具有不同的含义。为了解决这一问题,研究人员提出了本体的概念:两个系统通过共享的术语集——本体来达到对术语集的统一理解,从而实现系统之间语义层次上的集成与互操作,但所有的系统均基于同一本体系统来进行开发是不可能。因此,为了真正实现系统之间的语义互操作,还需要解决不同本体之间的集成。目前,许多本体已经被开发且本体之间包含交叉重叠的信息,如何合理利用已存在的本体是亟待解决的问题。例如某公司既需要使用统一规范的本体,同时又要把它和特定的本体一起使用,或在具体应用时可能需要同时使用不同领域的本体,或同时使用同一领域但不同角度的本体;而本体构建者则希望在已存在

的本体的基础上通过扩展已存在的本体得到新的本体或合并已存在的领域本体等。无论是上述哪一种情况,了解不同本体内术语之间的关系都很重要。当定义不同本体内术语之间的关系时,就需要建立两个本体的对齐^[1]。文献[2,3]中也提出本体对齐可用于解决异构问题。近几年来,本体对齐一直是一个重要的研究课题,不同的研究人员开发了相同领域的知识本体,而这些本体通常是异质的。因此,需要一种本体对齐来整合或完善本体知识体系^[4]。本体对齐和本体合并是本体工程中重要的步骤之一,且有待进一步研究^[5]。

范畴论的历史始于1945年艾林伯格(Eilenberg)和麦克莱恩(MacLane)共同署名的著名论文“自然等价的一般理论”中,并逐渐扩展到对各类数学结构之间的关系的调查研究。1963年,F. William Lawvere的论文开创了范畴逻辑和模型理论的研究领域,使得范畴论在数学语义领域内得到广泛应用。范畴论作为数学的一种分支,是一种高度概括和抽象的结构,且提供了本体集成的一系列形式化方法(如余极限、极限等),为语义的集成提供了一个统一的方法框架,它正在被

到稿日期:2011-10-20 返修日期:2012-02-10 本文受国家自然科学基金(40871178),四川省教育厅重点项目(11ZA030)资助。

胡玲(1969—),女,硕士,副教授,主要研究方向为计算机应用、地理信息系统理论,E-mail:hhuling@163.com;李霖(1960—),男,教授,博士生导师,主要研究方向为数字地理制图;王红(1975—),女,博士后,副教授,主要研究方向为地理信息系统理论与技术。

广泛用于计算语义模型及其应用领域^[6]。文献[7]中认为范畴论引入了一种面向对象的范式,这种范式使我们只需要了解接口,不需要了解内部的细节问题(如本体语言的异构),即这种特性可屏蔽本体间的异构性。利用范畴论原理,使我们得到的结果是在正确的数学理论框架支撑下的形式化结构^[8]。本文结合地理信息本身的特点,以数学中的范畴论为理论基础,通过利用范畴论中态射、(余)锥、(余)极限等概念来屏蔽本体的各种异构,通过范畴本体对齐建立地理本体语义关系,最终实现本体集成。

2 地理本体

2.1 地理本体的逻辑模型

针对地理信息本身的特点,借鉴以前的定义来定义地理信息本体如下^[9]。

定义 1 Geo-ontology = $\langle C, R, H, P, P^R, I \rangle$, 其中, C 概念(Concept)表示一组共享某些相同属性的对象的集合; R 概念之间的语义关系(Relationship)可理解为概念之间的横向关系,例如等价关系; H 概念之间的层次关系(Hierarchy)主要是指父/子类关系,可理解为概念之间的纵向关系,层次关系也是一种语义关系; P 属性(Property)分为对象属性和数据属性,对象属性表示个体之间的关系,数据属性表示个体到数值的关系; P^R 对属性的限制主要是对属性取值的类型、范围及多少等的限制; I 实例(Individual)。

2.2 地理本体语义

地理本体通过本体属性表达语义,本体属性集描述了地理本体概念的内涵,反映了概念本质,可将地理概念用一组结构化的属性组取代文字描述。

定义 2(地理概念的语义)^[10,11] $G_c = (c, \{A\})$, 其中 c 为概念术语, A 表示概念的属性,即为本体属性。地理本体语义的本质就是指地理概念的语义。

2.3 地理概念的语义关系

根据现有的地理信息概念定义,可以分析出一些基础地理信息概念之间的语义关系,其主要包括上下位关系(is-a)、部分整体关系(part-of)及附属关系(appertaining-to) 3 大类^[9]。

3 地理本体对齐

3.1 改进的地理本体范畴

范畴论的两大要素是对象和态射,但最基本和最核心的概念是态射。所谓“态”就是保持结构(关系)之态,因此对态射而言,重要的是使对象的原有结构(关系)保持不变,通过它,才会形成更高层次的结构。然而态射函数(即结构保持函数)只能表示同义关系(即两个概念语义相等)^[7,12,13]。而在实际的应用中,当两个不同的本体中概念可能是相等(相似)关系,也可能是蕴含关系或其他关系时,用范畴本体中的态射就无法表达,也就无法利用范畴论提供的形式化方法实现本体的集成。因此在范畴论的基础上,结合地理本体特征,对态射进行了重新定义,构造了一个更复杂的范畴,其中范畴中的对象为地理本体,而态射不仅能表达相等关系,也可以表示其他语义关系。

定义 3(态射) 设 $A, B \in O$ (O 为地理本体集合), $A = \langle C_1, R_1, H_1, P_1, P^{R_1}, I_1 \rangle$, $B = \langle C_2, R_2, H_2, P_2, P^{R_2}, I_2 \rangle$, 定义每一个态射 $f: A \rightarrow B$ 由三元组 $\langle e_A, e_B, R \rangle$ 集构成, 其中 e_A 和 e_B 分别表示地理本体 A 和 B 的概念、关系、属性和实例等

要素; R 表示 e_A 和 e_B 的语义关系(如同义关系、蕴含关系、交叠关系等)。

3.2 本体对齐

许多文献中都给出了本体对齐的定义,如文献[14]中认为本体对齐是不同的本体之间存在的语义关系的明确说明。文献[15]中认为本体对齐是描述两个不同本体中实体的相应关系的映射的集合,本体对齐并不是关系实体间关系的集合,而是映射的集合。本体对齐将两个本体正确地连接起来,可以用于多个本体的联合查询等任务。文献[16]中认为本体对齐是在两个源本体之间寻找相似性的过程,该操作的结果是对两个本体之间的相似点进行详细的说明。文献[17]中认为本体对齐实际上就是建立两个本体之间二元关系集的过程。文献[18]中定义本体对齐就是找到两个实体之间的关系(相等关系和蕴含关系等)。文献[19]中指出语义的互操作是通过找到不同本体的实体之间的关系来实现本体间的一致性,这个过程称为本体对齐。

本文结合文献中的定义和地理本体的特征,将范畴论引入本体对齐,定义本体对齐为态射集,其中本体作为范畴论的对象。

定义 4(地理本体对齐) 设 A, B 是地理本体, A, B 的本体对齐是关于 A 和 B 的概念、关系、属性和实例等要素之间的态射集,用 (O, P_1, P_2) 表示。其中 O 是地理本体, $P_1: O \rightarrow A$, $P_2: O \rightarrow B$ 是态射,态射不仅能表示同义关系,也可以表示多种语义关系。

根据第 2.2 节中地理本体语义的定义,可推导出地理本体语义对齐的建立应遵循以下 4 个规则。

规则 1 如果两个地理概念 C_1, C_2 有完全相同的本体属性集 $\{A\}$, 同时所有属性上均有真值,且值域是相关的,则认为这两个概念是同义关系,用符号“ $\equiv$ ”表示,即:

if $G_{C1} \equiv G_{C2}$, 则 $C_1 \equiv C_2$

规则 2 如果两个概念 C_1, C_2 本体属性集不同,即使概念名相同,实质上也是不同的地理概念。用符号“ $\perp$ ”表示,即: if $G_{C1} \perp G_{C2}$, 则 $C_1 \perp C_2$ 。

规则 3 如果两个概念 C_1, C_2 的本体属性集部分可比,即概念 C_1 在概念 C_2 的所有本体属性上均有真值,且值域是相关的,反之不成立,则 C_1 语义蕴含 C_2 (或 C_2 被包含于 C_1), 用符号“ $\supset$ 或 $\subset$ ”表示,即: if $G_{C1} \in G_{C2}$, 则 $C_1 \supset C_2$ (或 $C_2 \subset C_1$)。

规则 4 如果两个概念 C_1, C_2 的本体属性集部分相同,即概念 C_1 在概念 C_2 的部分属性上有真值,且值域相关,反之亦然,则这两个概念为交叉关系,用符号“ $\bowtie$ ”表示。

3.3 本体对齐的复合

复合是对齐重用的核心操作:如果在本体 A 和 B 之间、 B 和 C 之间存在对齐关系,利用复合就可以得到本体 A 和 C 之间的对齐关系^[3,12],即利用已存在的对齐关系推导出新的对齐关系。根据范畴论的定义,态射具有复合性,因此本体对齐的复合实质上是态射的复合。

定义 5(本体对齐的复合) 设 $A, B, C \in O$ (O 为地理本体集), 态射 $f: A \rightarrow B, g: B \rightarrow C$, 其中 $\langle e_1, e_2, R_1 \rangle \in f, \langle e_2, e_3, R_2 \rangle \in g$, 则存在复合态射 $g \circ f$ 是一个三元组 $\langle e_1, e_3, R, n \rangle \in g \circ f, R = \xi(R_1, R_2)$ 由函数 ξ 表示, $\xi: R \times R \rightarrow R$ 是一个关系集,假设有以下 5 种关系,那么得到的复合关系由表 1 描述。该复合表由关系代数得到并且可以由基本关系扩展到关系的聚合^[3,12]。

表1 复合关系表

$R_2 \backslash R_1$	$\equiv$	$\supset$	$\subset$	$\bowtie$	$\perp$
$\equiv$ (相等)	$\equiv$	$\supset$	$\subset$	$\bowtie$	$\perp$
$\supset$ (蕴含)	$\supset$	$\supset$	$\supset \equiv \bowtie$	$\supset$	$\supset \perp$
$\subset$ (蕴含于)	$\subset$	$\supset \equiv \bowtie$	$\subset$	$\subset \perp$	$\perp$
$\bowtie$ (交叉)	$\bowtie$	$\supset \perp$	$\subset \perp$	$\supset \equiv \perp$	$\supset \perp$
$\perp$ (不等)	$\perp$	$\perp$	$\subset \perp$	$\subset \perp$	$\supset \equiv \perp$

4 地理本体集成

范畴论提供了本体集成的一系列形式化方法(如余极限、极限等),为语义的集成提供了一个统一的方法框架。本文利用余极限的形式化方法来说明地理本体的集成。外推(push-out)是余极限的特殊形式,其代表了建构的过程,这个过程将各个内在互连的系统建构成一个更为庞大的系统。作为一种形式化的结构,外推确保了合并后的结果是完整的、唯一的和最小的^[20]。两个本体的集成如图1所示,其中若A和B表示地理本体,O则表示Q和P的语义交集,即具有语义关系的集合,S表示Q和P的语义并集^[13],S为求解的结果(即集成本体)。因此,利用外推实现本体集成应首先确定地理本体的关系。多个本体的集成实质上就是通过连续的外推来实现的,其中箭头表示态射,A,B,C为源本体,O和O₁为本体对齐,D和F为利用外推得到的集成本体,如图2所示。不论是两个本体或多个本体的集成都首先要找到源本体间的对齐关系。因此研究本体对齐的目的正如文献[15]中指出的,本体对齐的结果能进一步支持态射的可视化,可将源本体转换为另一个本体或作为不同本体之间的桥接本体。

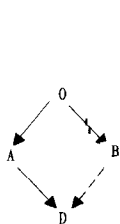


图1 外推

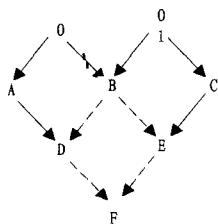
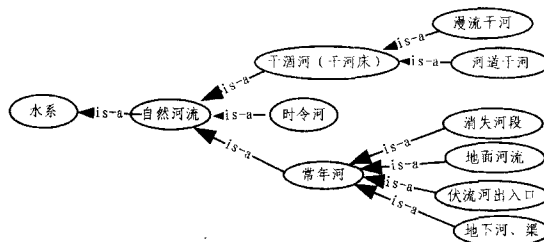
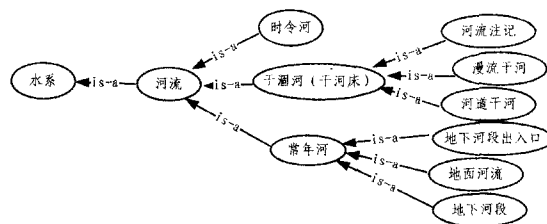
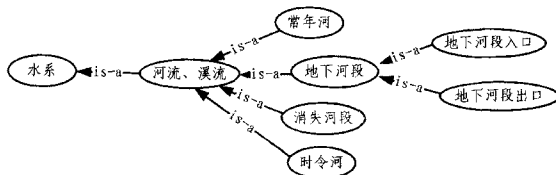


图2 外推的复合

5 案例分析

例1 利用斯坦福大学医学院医学信息中心开发的 on-

tology 编辑器和知识管理平台 Protégé 作为 ontology 建模工具开发平台,根据不同的分类标准建立水系“河流”,其中图3是采用《GB14804-93 地形图要素分类与代码》为标准建立的本体树的部分截图,图4是采用《GBT 13923-2006 1:500~1:2000 基础地理信息要素分类与代码》为标准建立的本体树的部分截图,图5是采用其他标准建立的本体树的部分截图。



由于采用了不同的分类标准,不同的地理本体系统间存在语义异构,因此要建立不同本体间的对齐关系,步骤如下:

(1)根据第2.2节的定义分别提取两个本体的地理概念的本体属性,如提取的地下河段、地下河段出入口、时令河等概念的部分本体属性如表2所列。

表2 部分本体属性和值域

概念名称	物质性	成因	空间形状	空间方向	空间位置		空间关系	地域+空间关系	时间性		功能性
					位置	地域性			周期性	生命周期	
地下河段	水	天然	槽形	出入口	陆地以下	岩溶地区	内	岩溶地区内	无	常态期	流通水
地下河段出入口	水	天然	槽形			岩溶地区	内	岩溶地区内	无	常态期	连接/地下河段/流通水
时令河	水	天然	槽形						有:季节性	常态期	流通水
消失河段	水	天然	槽形			沼泽U/沙地	内	(沼泽U/沙地/)内	有:季节性	消亡期:消失	流通水

(2)根据第3.2节的4个规则为推理规则确定不同本体概念之间的语义关系,推导出本体之间的语义关系集R。

图1和图2的部分对齐关系(用二元关系表示)如下:

河流 $\equiv$ 河流,地下河段入口 $\subset$ 地下河段出入口

图2和图3的部分对齐关系如下:

河流 $\equiv$ 自然河流,地下河段出入口 $\perp$ 伏流河出入口

从以上对齐关系可以看出,本体1中的地下河段与本体2中的地下河段出入口存在对齐关系,而本体2中的地下河段出入口与本体3中的伏流河出入口存在对齐关系。如果能

通过相应的操作找到本体1的地下河段和本体3中的伏流河出入口之间的对齐关系,就可以通过已经存在的对齐关系推理出新的对齐关系,这个操作称为复合。

例2 例1中的地下河段入口、地下河段出入口、伏流河出入口的语义关系如图6所示。

从复合关系表1可以得出地下河段和伏流河出入口的关系 $\{\subset\} \circ \{\perp\} = \{\perp\}$ 。同理,可以推导出河流和自然河流之间的关系以及其他概念间的语义关系,利用对齐的复合特性和

(下转第266页)

- [19] Jalba A C, Wilkinson M H F, et al. Automatic diatom identification using contour analysis by morphological curvature scale spaces[J]. Machine Vision and Applications, 2010, 16(4): 217-227
- [20] Culverhouse P F, Simpson R G, Ellis R, et al. Automatic classification of field collected dinoflagellates by artificial neural network[J]. Marine Ecology Progress Series, 1996, 139: 281-287
- [21] Davis C S, Fredrik T, et al. A three-axis fast-tow digital Video Plankton Recorder for rapid surveys of plankton taxa and hydrography[J]. American Society of Limnology and Oceanography, Methods, 2005(3): 59-74
- [22] Tang X, Stewart W K, Huang H, et al. Automatic plankton image

recognition[J]. Artificial Intelligence Review, 1998, 12: 177-199

- [23] Tang X, Lin F, Remsen A. Binary plankton image classification[J]. IEEE Journal of Oceanic Engineering, 2006, 31(3): 728-735
- [24] 郑少平, 谢杰镇, 王博亮, 等. 基于形状特征描述子的赤潮优势种识别研究[J]. 厦门大学学报: 自然科学版, 2009, 48(5): 676-679
- [25] 王明时, 王学民, 张磊. 显微图像分析技术在赤潮生物识别中的应用[J]. 天津大学学报, 2004, 37(5): 392-395
- [26] Ruiz L A, Molto E, et al. Location and characterization of the stem-calyx area on oranges by computer vision[J]. Journal of Agricultural Engineering, 1996, 64: 165-172
- [27] Sarkar N, Wolfe R R. Feature extraction techniques for sorting tomatoes by computer vision[J]. Transactions of the ASAE, 1985, 28(3): 970-974

(上接第 244 页)

已存在的对齐关系就可以推导出更复杂的关系。



图 6 复合关系图

结束语 近年来, 范畴论被用于开发本体描述的新方法及解决视图更新问题的新途径^[21]。范畴论的高度概括、抽象和更强、更直观的表达力为本体集成提供了一种框架, 这种特性可屏蔽本体间的异构。本文将范畴论引入本体对齐的研究中, 在范畴论的基础上, 结合地理本体特征, 对态射进行了重新定义, 构造了一个更复杂的范畴。其中范畴中的对象为地理本体, 而态射不仅能表达相等关系, 也可以描述地理本体间的多种语义关系。在态射定义的基础上, 根据态射的特性定义了本体对齐和本体对齐的复合, 而研究本体对齐的最终目的是实现本体的集成, 因此对范畴论提供的本体集成的方法之一即外推进行了阐述。

将范畴论引入本体对齐的研究领域中的最大特点在于, 其表达能力和可扩展性较强, 不仅能表达多种语义关系, 而且能通过本体对齐的复合从已存在的对齐关系中推导出新的对齐关系, 可用于在分布式系统中多个本体的联合查询等任务。因此, 范畴论为地理信息本体语义集成提供了一种适合的形式化框架。

参考文献

- [1] Lambrix P, Tan H. A framework for aligning ontologies[M]. Principles and Practice of Semantic Web Reasoning, 2005: 17-31
- [2] Bouquet P, Euzenat J, Franconi E, et al. Specification of a common framework for characterizing alignment[J]. Kweb deliverable, D, 2004, 2
- [3] Euzenat J. Algebras of ontology alignment relations [C]// The 7th International Semantic Web-ISWC. 2008: 387-402
- [4] Shareha A, Rajeswari M, Ramachandram D. Multimodal Integration (Image and Text) Using Ontology Alignment[J]. American Journal of Applied Sciences, 2009, 6(6): 1217-1224
- [5] Lambrix P, Tan H. Ontology alignment and merging[M]. Anatomy Ontologies for Bioinformatics: Principles and Practice. Springer, 2008: 133-149
- [6] Cafezeiro I, Haeusler E. Semantic interoperability via category

theory[C]// ER'07 Tutorials, Posters, Panels and Industrial Contributions at the 26th International Conference on Conceptual modeling. New Zealand: Australian Computer Society, Inc., 2007, 83: 197-202

- [7] Hitzler P. Integrated view and comparison of alignment semantics[J]. Deliverable, Knowledge Web NoE, 2005, 2
- [8] Oyenat W, DeLoach S. Using Category Theory to Compose Multiagent Organization Design Models[R]. Manhattan. K. S. University, 2010: 1-34
- [9] 王红. 国家基础地理信息本体研究 [D]. 武汉: 武汉大学, 2009: 1-124
- [10] 李霖, 朱海红, 王红, 等. 基于形式本体的基础地理信息语义分析—以陆地水系要素类为例[J]. 测绘学报, 2008, 38(2): 230-242
- [11] 胡玲, 李霖, 王红. 基于范畴论的地理本体集成研究[J]. 计算机工程与设计, 2010, 31(16): 3732-3735
- [12] Zimmermann A, Krotzsch M, Euzenat J, et al. Formalizing ontology alignment and its operations with category theory[J]. Frontiers in Artificial Intelligence and Applications, 2006, 150: 277-289
- [13] Pierce B. Basic category theory for computer scientists [M]. Cambridge: The MIT Press, 1991: 1-20
- [14] Zimmermann A, Euzenat J. Three semantics for distributed systems and their relations with alignment composition[C]// The Semantic Web-ISWC 2006, 2006: 16-29
- [15] 于娟, 党延忠. 本体集成研究综述[J]. 计算机科学, 2008, 35(7): 9-13
- [16] de Bruijn J, Ehrig M, Feier C, et al. Ontology mediation, merging and aligning[M]// Davies J, Studer R, Warren P. Semantic Web Technologies: trends and research in ontology-based system. West Sussex(UK): John Wiley, 2006
- [17] Kalfoglou Y, Schorlemmer M. Ontology mapping: the state of the art[J]. The knowledge engineering review, 2003, 18(1): 1-31
- [18] Euzenat J, Valtchev P. Similarity-based ontology alignment in OWL-lite[C]// Ios Pr Inc. 2004: 333-337
- [19] Euzenat J, Barrasa J, Bouquet P, et al. D2. 2. 3: State of the art on ontology alignment[OL]. <http://www.starlab.vub.ac.be/research/projects/knowledgeweb/kweb-223.pdf>, 2004
- [20] Solomon A. Pushout: A Mathematical Model of Architectural Merger[J]. Lecture Notes in Computer Science, 2007, 4378: 389-399
- [21] Johnson M, Rosebrugh R. Ontology Engineering, Universal Algebra, and Category Theory[M]. Theory and Applications of Ontology, Computer Applications, 2010: 565-576