

# GIS 中地理本体的建立与比较

安 杨<sup>1, 2</sup> 边馥苓<sup>2</sup> 关侏红<sup>3</sup>

(1 武汉大学计算机学院, 武汉市珞喻路 129 号, 430079)

(2 武汉大学空间信息与数字工程研究中心, 武汉市珞喻路 129 号, 430079)

(3 同济大学电子与信息工程学院, 上海市四平路 1239 号, 200092)

**摘 要:**重点研究了地理本体的建模和地理本体比较方法, 对地理本体比较方法的核心思想从结构和语义两个方面对地理本体中的概念进行了比较, 并计算了其相似度, 以判断各种地理概念之间的关系。

**关键词:**地理本体; 语义相似度

中图法分类号: P208

## 1 地理本体的建模

地理本体提供地理信息领域中所公认的概念并对其进行明确的形式化定义, 以解决由于地理认知不同所引起的问题及其逻辑描述的互译问题, 能够应用于地理信息集成、共享等方面。

本文在对 OGC 的 GML3.0 规范进行详细分析的基础上, 定义了自己的地理本体。地理空间由不同种类的地理实体构成, 一个本体概念(类)由一系列的性质和行为相同的地理实体组成, 而每个地理实体又可以映射为计算机世界中的一个对象, 每个对象都具有几何(空间)属性、时间信息以及与其他对象的拓扑、语义关系等。地理本体结构是按层次方式组织的, 概念(类)之间存在同义和蕴含关系。同义关系表示两个概念(类)语义相等; 蕴含关系表示两个概念(类)是继承关系。基于这一认识, 在地理信息领域最重要的概念是要素、坐标系和空间关系。因此, 建立地理本体的中心就是 GeoFeature、GeoRelation 和 GeoCoordinationSystem 本体。地理本体模型如图 1 所示。

## 2 地理概念的比较

地理本体的建立为地理信息的共享提供了前

提和保障。地理本体允许用户在其基础上定义自己的应用本体。由于空间信息数据繁多、结构复杂, 用户在地理本体上进行扩展, 各自定义的本体间可能存在较大的差异, 因此, 本体中地理概念的比较是基于本体进行空间信息共享所必须解决的问题。

为了解决空间信息集成中的两种异构问题, 考虑从结构和语义两方面对概念进行比较。从结构方面对概念进行比较主要是对概念的名称进行比较。而从语义方面对概念进行比较主要是确定在地理本体中概念之间的语义关系。

为了衡量相关元素的关系, 用一个相似度系数, 即相似度来描述两个描述间各相关元素的相似程度。相似度系数用一个小数来表示, 其取值范围是[0, 1], 0 表示两个元素完全不同, 1 表示完全相同。相似度需要通过结构和语义两方面来确定。定义结构相似度为  $SynSim$ , 语义相似度为  $SemSim$ , 则所比较的两个概念的相似度  $SimDegree$  定义如下:

$$SimDegree = SimW * SynSim + (1 - SimW) * SemSim$$

其中  $SimW$  是用来确定语义相似度和结构相似度在相似度计算中所占的权重, 取值范围为 0~1 的实数, 由用户给出。当用户倾向于结构相似度的比较时,  $SimW$  的值就大; 当用户倾向于语

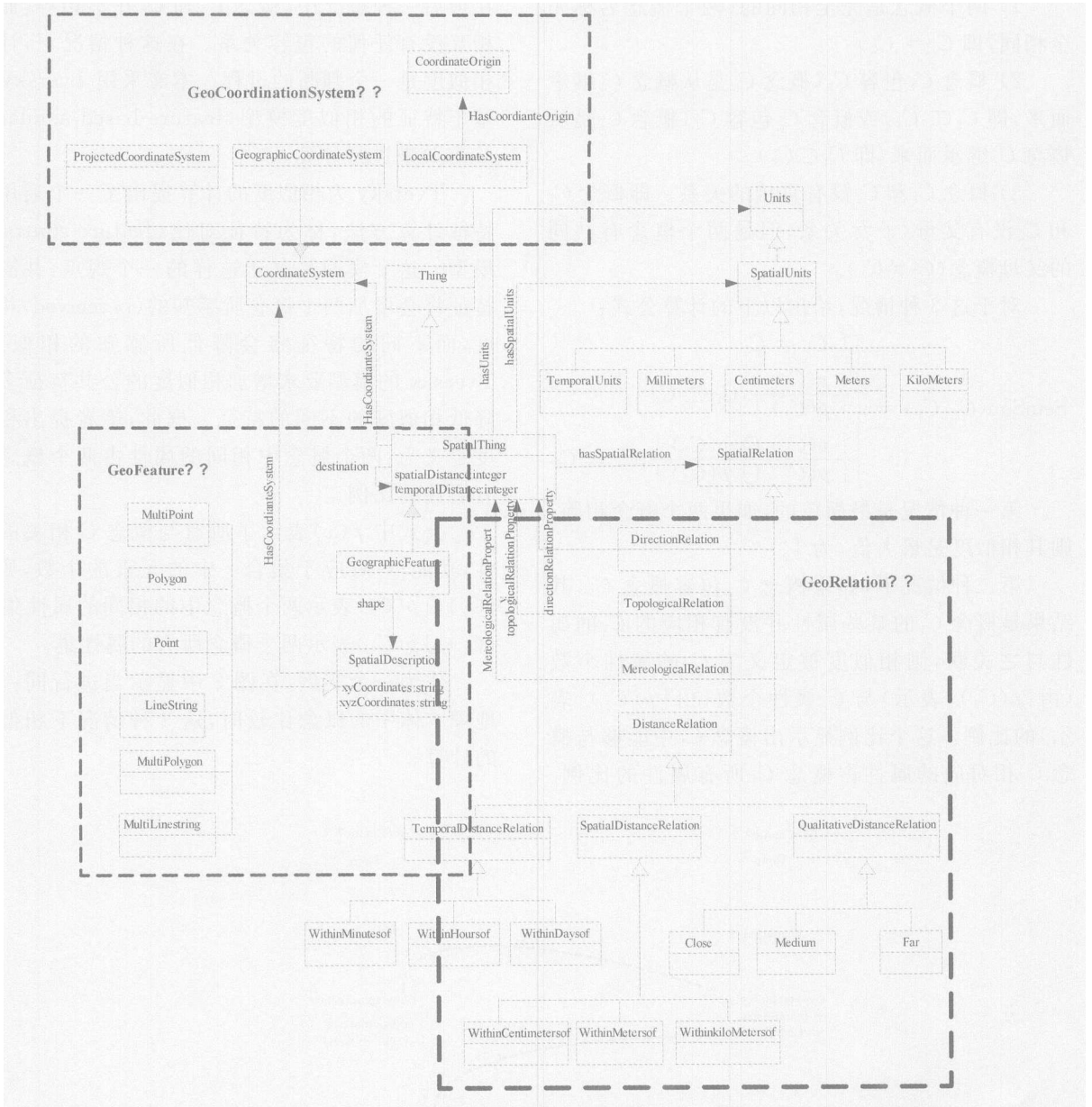


图 1 地理本体的模型

Fig. 1 Model of Geo Ontology

义相似度的比较时,  $SimW$  的值就小。例如, 当  $SimW = 0$ , 就是只考虑语义相似度的比较。由于本文主要是研究语义异构, 因此,  $SimW$  的值取 0。

### 3 语义相似度的计算

确定本体中的概念之间的语义关系, 用产生式规则推理概念之间的关系。整个推理过程以以下 3 个公理为基础。

公理 1 本体概念之间的关系是由概念之间的蕴含或同义关系组成, 概念的语义由其在属性集  $(a_1, a_2, \dots, a_m)$  的特征子集  $\{X_1, \dots, X_n\}$  描述。

公理 2 如果两个概念(类)  $C_s$  和  $C_t$  的属性是

完全可比, 即两个类有共同的属性集且在所有属性上均有真值, 值域是相关的, 则两个类  $C_s$  和  $C_t$  是同义关系。记为: 如果  $\{X_{s_1}, \dots, X_{s_k}\} = \{X_{t_1}, \dots, X_{t_k}\}$ , 则  $C_s \equiv_s C_t$ 。其中  $\equiv_s$  表示语义相等。

公理 3 如果两个类  $C_s$  和  $C_t$  的属性集部分可比, 即一个类  $C_s$  的一个实例在另一个类  $C_t$  的所有属性上均有真值, 且值域是相关的, 反之不成立, 则  $C_s$  语义包含  $C_t$ 。记为如果  $\{X_{s_1}, \dots, X_{s_k}\} \in \{X_{t_1}, \dots, X_{t_l}\}$ , 则  $C_s \Rightarrow_s C_t$ 。其中  $\Rightarrow_s$  表示语义包含。

假定, 概念  $C_1$  和概念  $C_2$  都属于本体  $O$ , 在比较的过程中, 根据以上 3 个公理, 存在以下 3 种不同的情况。

1) 两个概念是完全相同的(两个概念名称完全相同,即  $C_1 = C_2$ )。

2) 概念  $C_1$  包容  $C_2$  (概念  $C_2$  是从概念  $C_1$  继承而来,即  $C_2 \subset C_1$ ) 或概念  $C_2$  包容  $C_1$  (概念  $C_1$  是从概念  $C_2$  继承而来,即  $C_1 \subset C_2$ )。

3) 概念  $C_1$  和  $C_2$  没有直接的关系。即概念  $C_1$  和  $C_2$  没有父母/子女关系,但是两个概念有共同的父母概念( $C_1 \neq C_2$ )。

对于这3种情况,给出以下的计算公式:

$$\text{SemSim}(C_1, C_2) = \begin{cases} 1, C_1 = C_2 \\ \frac{|p(C_2)|}{|p(C_1)|}, C_1 \supset C_2 \\ \frac{|p(C_1) \cap p(C_2)|}{|p(C_1) \cup p(C_2)|}, C_1 \neq C_2 \end{cases}$$

第一种情况是最简单的,如果两个概念相等,则其相似度是最大值,为1。

第二种情况下,如果概念  $C_2$  包容概念  $C_1$ ,其结果是概念  $C_1$  的某些属性并没有相应的  $C_2$  的属性与之关联,则相似度被定义为  $C_2$  的属性个数(由  $|p(C_2)|$  表示)与  $C_1$  属性个数(由  $|p(C_1)|$  表示)的比例。这个比例显示出概念  $C_1$  中能够与概念  $C_2$  相对应的属性占概念  $C_1$  所有属性的比例。

在最后一种情况中,概念  $C_1$  和  $C_2$  并不相等,而且相互没有任何的包容关系。在这种情况下,计算相似度是一个判断的过程。本文采用 Tversky 的基于特征的相似度模型(feature-based similarity model)来进行计算。

Tversky 为相似度的计算提出了一个通用的特征计算方法,称为特征对比(feature-contrast)模型。这个模型是基于这样的一个观点:共同的特征将会增加两个概念所感知的(perceived)相似度,而不同的特征将会降低所感知的相似度。Tversky 的模型要求增加相似度的公共特征多于降低相似度的不同的特征。据此,笔者提出相似度定义为:两个概念中相同的属性占两个概念所有属性的比例。

公式中  $p(x)$  表示了所有与概念  $C$  相关的属性,功能  $|s|$  对应于集合  $s$  中的元素的个数,则  $p(C_1) \cap p(C_2)$  表示两个概念中的相同的属性集,  $p(C_1) \cup p(C_2)$  表示两个概念所有的属性集。

作为一个实例,在图2中显示当进行同一个地理本体中的概念比较时,这3种情况下相似度的计算。

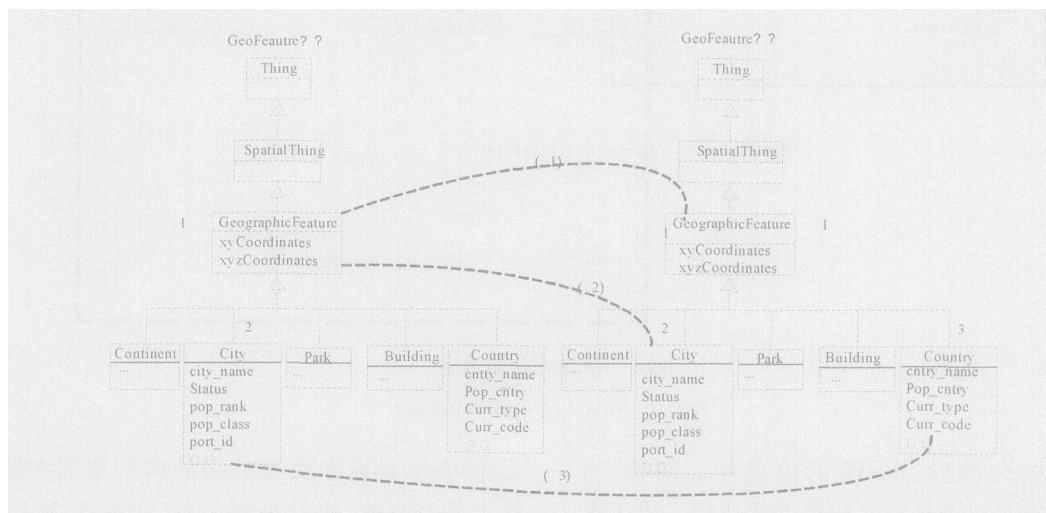


图2 在同一个本体中概念的比较

Fig. 2 Concept Comparison in Same Ontology

在例子中,概念所在的本体是笔者所定义的地理本体中的 GeoFeature 本体,该本体在图2中只表示了例子中所需的概念。表1显示了图2中地理概念的比较情况。

表1 本体比较的3种情况

Tab. 1 Comparison in Same Ontology

$C_1$	$C_2$
(1) Geographic-Feature(1)	Geographic Feature(1)
(2) Geographic-Feature(1)	City(2)
(3) City(2)	Country(3)

在第一种情况中,两个概念均和 GeographicFeature 相关。这是最简单的情况。

在第二种情况中,第一个概念与 GeographicFeature 相关,而另一个概念与 City 相关。由于概念 GeographicFeature 包含了概念 City,概念 GeographicFeature 的属性集是概念 City 属性集 {xyCoordinates, xyzCoordinates, city\_name, Status, pop\_rank, pop\_class, port\_id} 的子集,因此,两个概念的属性集只有部分相同,则两个概念的

相似度利用各自输出属性个数的比率来计算为:

$$|p(C_1)| / |p(C_2)| = 2/7 = 0.29$$

在第三种情况中, 第一个概念与概念 City 相关, 而第二个概念与 Country 相关。概念 City 的属性集是{xyCoordinates、xyzCoordinates、city\_name、Status、pop\_rank、pop\_class、port\_id}, 概念 Country 的属性集是{xyCoordinates、xyzCoordinates、entry\_name、Pop\_entry、Curr\_type、Curr\_code}。由于这两个概念没有父母/子女关系, 相似度计算如下:

$$S_1 = p(\text{City}) = \{\text{xyCoordinates、xyzCoordinates、city\_name、Status、pop\_rank、pop\_class、port\_id}\}$$

$$S_2 = p(\text{Country}) = \{\text{xyCoordinates、xyzCoordinates、entry\_name、Pop\_entry、Curr\_type、Curr\_code}\}$$

$$S_3 = p(\text{City}) \cap p(\text{Country}) = \{\text{xyCoordinates、xyzCoordinates}\}$$

$$S_4 = p(\text{City}) \cup p(\text{Country}) = \{\text{xyCoordinates、xyzCoordinates、city\_name、Status、pop\_rank、pop\_class、port\_id、entry\_name、Pop\_entry、Curr\_type、Curr\_code}\}$$

$$\text{SemSim}(\text{City}, \text{Country}) = |S_3| / |S_4| = 2/11 =$$

0.18

计算相似度的结果表示概念 City 和 Country 的相似度很低。

对地理概念进行相似度的计算后, 其结果可体现地理实体之间的关系, 为地理信息的共享和集成提供基础和依据。

### 参 考 文 献

- [1] 关侏红, 虞为, 安杨. GML 模式匹配算法. 武汉大学学报·信息科学版, 2004, 29(2): 169-175
- [2] Guan Jihong, Zhou S G, Chen J P, et al. Ontology-based GML Schema Matching for Information Integration [C]. ICMLC'03, Xi'an, 2003
- [3] Weinstein P C, Birmingham W. Comparing Concepts in Differentiated Ontologies [C]. The 12th Workshop on Knowledge Acquisition, Modeling and Management (KAW'99). Banff, Alberta, Canada, 1999
- [4] 安杨, 边馥苓, 关侏红. 基于 Ontology 的网络地理服务描述与发现. 武汉大学学报·信息科学版, 2004, 29(12): 1063-1066

第一作者简介: 安杨, 副教授, 博士。研究方向: 语义网, 网络地理服务, 地理信息系统。

E-mail: yangan@whu.edu.cn

## Design and Comparison of Geo Ontology in GIS

AN Yang<sup>1,2</sup> BIAN Fuling<sup>1</sup> GUAN Jihong<sup>3</sup>

(1 School of Computer Science, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

(2 Research Center of Spatial Information and Digital Engineering, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

(3 School of Electronics and Information Engineering, Tongji University, 1239 Siping Road, Shanghai 200092, China)

**Abstract:** Ontology is introduced in GIS domain to resolve all kinds of heterogonous issue of GIS. Geo ontology model is designed, and Geo concepts comparing method is proposed, in which concept in Geo Ontology is compared from structure and semantics level to adjust the relation of geo concepts.

**Key words:** geo ontology; semantic similarity degree

**About the first author:** AN Yang, associate professor, Ph D, majors in semantic web, Geo web services, GIS.

E-mail: yangan@whu.edu.cn