语义支持的地理要素属性相似性计算模型

谭永滨[1,2,3],唐瑶[1],李小龙[1,2],刘波[1,2],危小建[1,2]

(1. 东华理工大学 测绘工程学院,南昌 330013;2. 流域生态与地理环境监测国家测绘地理信息局重点实验室,南昌 330013;
3. 东华理工大学 江西省数字国土重点实验室,南昌 330013)

摘要:现势性是发挥数据应用价值的关键。鉴于在地理要素数据匹配过程中,常常因为分类标准不同或地图综合处理等原因而产生要素属性间的语义异质性,该文试图通过引入本体语义技术解决语义异质性问题,利用属性枚举的方法表达地理要素类别语义,构建地理要素分类本体结构;同时将地物属性信息划分为要素类别信息、地物名称信息以及辅助信息 3 个部分,提出了语义支持的地理要素属性相似性模型。该模型从概念语义内涵的角度计算地理要素类别的相似性;从地物名称字面出发计算地物名称相似度;根据不同的辅助属性类型计算辅助属性相似度。实验结果表明,同尺度多时态与跨尺度相同时间版本环境下,所提出的模型可有效合理地评价候选地物间的属性相似性,较好地为地理数据匹配提供支持。

关键词:语义分析;相似性;要素属性匹配;本体;数据匹配

doi:10.3969/j.issn.1000-3177.2017.01.022

中图分类号:P208 **文献标识码**:A **文章编号**:1000-3177(2017)149-0126-08

Semantic-based Geographic Feature Property Similarity Measurement Model

TAN Yong-bin[1,2,3], TANG Yao[1], LI Xiao-long[1,2], LIU Bo[1,2], WEI Xiao-jian[1,2]

(1. *School of Geomatics, East China University of Technology, Nanchang* 330013, *China*;
2. *Key Laboratory of Watershed Ecology and Geographical Environment Monitoring, National Administration of Surveying, Mapping and Geoinformation, Nanchang* 330013, *China*;
3. *Key Laboratory for Digital Land and Resources of Jiangxi Province, East China University of Technology, Nanchang* 330013, *China*)

Abstract: Data currency plays an important role in the data employment. During the geographic feature matching process, there was some semantic heterogeneity among geographic data of different scales in different time produced by different classification standard or map generalization processes. In this paper, ontology semantic technology was introduced to solve the semantic heterogeneity. A geographic class was represented as ontology properties, and then the geographic category ontology was constructed. Each geographic feature was divided as three parts, which were category information, object name and assistant information, respectively. Finally, a geographic feature properties similarity model was proposed with the geographic category ontology to match geographic objects as the full perspective of attribute information, which could provide a foundational base for geographic updating.

Key words: semantic analysis; similarity; feature property matching; ontology; data matching

收稿日期:2016-09-09 **修订日期**:2016-10-26

基金项目:东华理工大学博士启动基金项目(DHBK2015310);东华理工大学江西省数字国土重点实验室开放研究基金资助项目(DLLJ201619);江西省自然科学基金(20151BAB213030)。

作者简介:谭永滨(1985-),男,博士,主要从事地理本体、概念语义研究。

E-mail: ybtan@ecut.edu.cn

通信作者:唐瑶(1986-),女,硕士,主要从事语义分析、地理建模与分析研究。

E-mail: yaotang@ecut.edu.cn

引用格式:谭永滨,唐瑶,李小龙,刘波,危小建. 语义支持的地理要素属性相似性计算模型[J]. 遥感信息,2017,32(1):126-133

0 引 言

地理空间数据信息的现势性是地理信息科学(Geographic Information Science,GIS)的灵魂,远远高于几何精确性[1]。描述不同时期的地物目标变化(空间及属性变化)的增量数据是实现多比例尺地理数据级联更新、快速增量制图等应用的重要数据源[2]。提取增量数据的关键在于度量不同地理空间数据中的地理要素的相似性,识别相同(似)的地理要素并形成地理匹配要素集合。根据地理要素的特点,其相似性包括空间相似性与属性相似性两方面。空间相似性是指根据特定内容和比例尺对空间的匹配和排序[3],通常是指两个空间对象在某个特定粒度下表现为相似的现象。而属性相似性是指不同时期间地理要素所代表的地物目标的属性特征间的相似度。一般地理增量数据提取过程,先利用空间相似性算法查找到空间上相关的地理对象,形成候选相似地理数据集;接着使用属性相似性算法确定候选集合中对象与待匹配对象间的相似度,以最终确定两时期地理数据中的相同(似)对象。因此如何准确度量空间与属性相似性是提取增量数据的关键。

由于地理要素数据采集的时期不同,所采用的分类标准也不一致。因此属性相似性度量存在两方面语义异质性的问题:1)新旧数据中同一类别地理要素因不同分类标准而产生语义异质性;2)同一地物因数据表达尺度的不同造成数据表现也不相同,进而影响地理要素的类别属性,即不同尺度下相同的地理要素属于不同的类别(可能属于上下义关系的类别)。本体技术起源于哲学领域,后被引入地理信息领域[4],是解决语义异质性的重要技术方法,广泛应用于语义互操作[5-7]、概念的形式化表达[8-10]等方向。信息本体包含4个基本特征:概念化、明确、形式化与共享[11],其主要目的是清晰明确地表达概念中隐含的语义信息。如何清晰准确地表达概念的语义信息是实现对象关联的基础。常用的语义表达方法有多种,如比喻法、代数法、属性枚举法等。属性枚举法通过列举出概念的本体属性[8]表达概念语义。本文采用属性枚举法表达地理要素类别的语义,同时在文献[8,12]提取的基础地理要素本体属性表的基础增加新的本体属性:新旧分类代码信息、比例尺,共同构建多尺度地理要素分类本体。

本文通过引入本体技术,分析地理要素的属性信息的语义内涵,构建多尺度地理要素分类本体;将属性信息分为要素类别、要素名称与辅助信息3部分,结合语义信息及基本相似度算法实现地理要素属性的相似性度量。

1 多尺度地理要素分类本体

根据基础地理要素分类的特点及应用情况,本文将基础地理要素类别本体形式化定义为:$O=(C, P_C, R, H_C, T_C)$。其中,C 表示地理要素分类集合;P_C 表示地理要素分类 c 的本体属性向量集合;R 表示要素分类间关系集合;H_C 包含具有分类关系的要素分类,包括了父、子要素分类间的上下位关系(kind-of)以及要素分类与实例间的关系(is-a)的集合;T_C 表示其他类型的要素分类间关系,如同义关系等。要素分类的本体属性集合 P_C 定义为:$P_C=(C,A)$,其中 C 为地理要素分类集合;A 为要素类别对应的本体属性向量。由于每个本体属性对于要素分类的重要性各不相同,本文在本体属性集合向量增加权重分量以表达特定本体属性的重要性。本体属性集合向量结构如下:$A=\{(a_1,v_1,\omega_1),(a_2,v_2,\omega_2),\cdots(a_n,v_n,\omega_n)\}$,其中 a 为本体属性分量;v 为相应的本体属性值;ω 为本体属性的权重,满足所有本体属性的 ω 值之和为1。现行的基础地理要素分类标准《GB/T 13923—2006》按照应用领域关系将地理要素类别划分为大类、中类、小类和子类。本文借助本体构建工具 Protégé4.1,在王红等[12]构建的基础地理要素分类本体库基础上建立本文的形式化本体结构,部分本体结构如图1所示。基础地理类别"常年河"的语义信息可的形式化表达为:

$O_{常年河}=$({"常年河"},{〖"常年河",{("物质性","水",1/7),("流通",{"水","石","土"},1/7),("成因","天然",1/7),("时间状态","连续",1/7),("旧分类代码","21010",1/7),("新分类代码","210100",1/7),("比例尺","比例尺等级",1/7)}〗},{"kind-of"},{〖"kind-of",("常年河","地面河流"),("常年河","地下河段"),("常年河","消失河段")〗},{})。

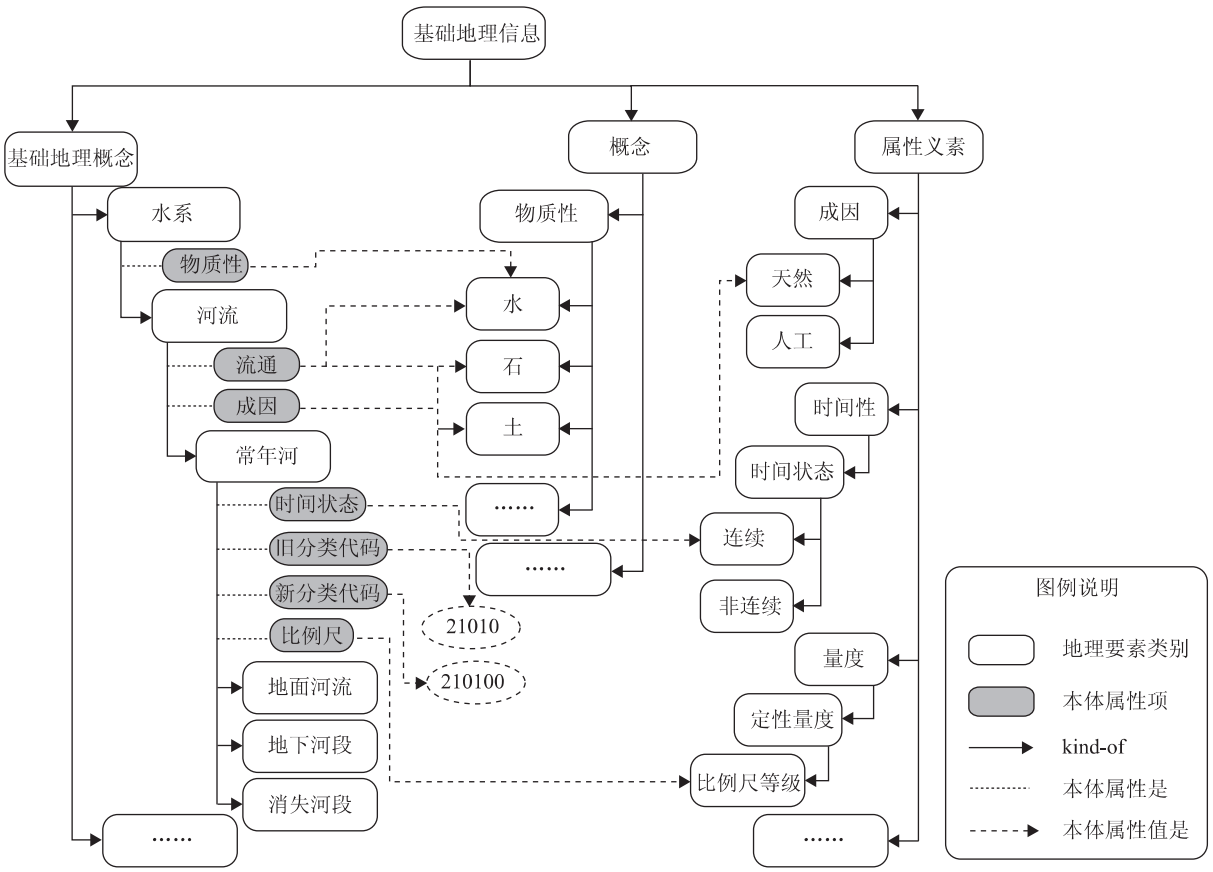


图 1　基础地理信息分类本体结构(部分)

2　基本相似度算法

目前在人工智能领域,度量概念相似性的算法众多,大致可分为。相似性作为描述两实体间相关关系的度量,在知识与行为理论、数据挖掘、信息获取、数据集成等领域研究的发挥着重要的作用。计算机科学家将语义相似性定义为层次结构中两实体的语义距离[13];认知心理学家和语言学家认为语义相似性是关于实体特征描述的相似性[14]。目前实体匹配的方法与模型的研究多种多样,大致可分为基于字符串、基于外部特征、基于信息内容、基于语义距离与基于本体属性等 5 类相似性算法。

2.1　基于字符串的相似度模型

该模型能够简便地计算概念的相似性且允许单独计算任意两个概念的相似性。但该模型仅适用于比较具有近似名称的概念,若比较同名异义、异名同义的概念则会获得错误的相似性结果,如“时令河”与“时令湖”两概念,在字面字符串上具有较高的相似度,但是实际上却是分别表达不同的含义。因此基于字符串的模型通常用于辅助其他模型计算相似度。

2.2　基于外部特征的相似性模型

该模型是通过抽取概念对象的外部显著特征属性,以集合的形式表达概念对象。比较概念间的各类特征属性项,形成共同特征属性集与差异特征属性集,利用差异模型或比率模型计算概念间的相似度。Tversky[15]利用该思路计算几何图形之间的相似度。该模型从概念对象的外在特征出发计算相似度,可在一定程度上避免同义词与一词多义等情况的影响,改善相似度结果的准确性。但并未比较概念的本质属性的相似性,容易在某些情况下容易被外在特征相同但本质并不完全相同的概念影响相似度结果。

2.3　基于信息内容的相似性模型

该模型基于信息论出现的,主要是利用概念间相同信息在语料库出现的概率计算相似度。该模型的基本原理是:概念间共享的信息量越大,语义相似度也相应越高;反之越小。对于概念树而言,子概念节点是父节点的细化,因此父节点的信息量较低,而子节点较多。目前的算法大多结合概念间共同父节点的信息量实现的相似度计算,但不同的算法由于利用的节点不同会产生不同的相似度结果[16-18]。

2.4　基于语义距离的相似性模型

该模型从本体语义关系的角度计算相似性,可消除同义词与一词多义等情况产生的影响。当前常

用的方法是结合已建成的树状或图状的本体层次结构,利用概念在本体结构中的位置,计算概念间的路径距离作为相似度结果。该模型可在分类体系构建完成的前提下能够快速地计算出概念间的相似度,且已有一些学者通过引入层次深度、亚层次密度[19]等因素提高结果的精确性。

2.5 基于本体属性的相似性模型

该模型从概念的语义内涵出发,将每个概念表达为本体属性集合,并利用相关本体属性的相似性结合权重信息计算概念的相似性[20]。该模型可在不依赖于概念本体结构的基础上计算概念间的语义相似度,避免相似度计算结果因分类标准不同而产生结果差异的情况。

3 地物属性语义相似性模型

地理数据通常采用属性表结构来表达其所代表的地物的属性特征信息。通过分析发现,在地物众多的属性特征中,存在两类相对重要的属性信息:"分类代码"与"名称"。其中"分类代码"属性描述地物内在本质的类别信息;"名称"属性是直观地区分地物的描述信息。因此本文将地物属性的语义相似性划分为三部分:地物类别语义相似性、地物名称相似性和辅助属性语义相似性。地物属性语义相似性模型框架如图 2 所示。

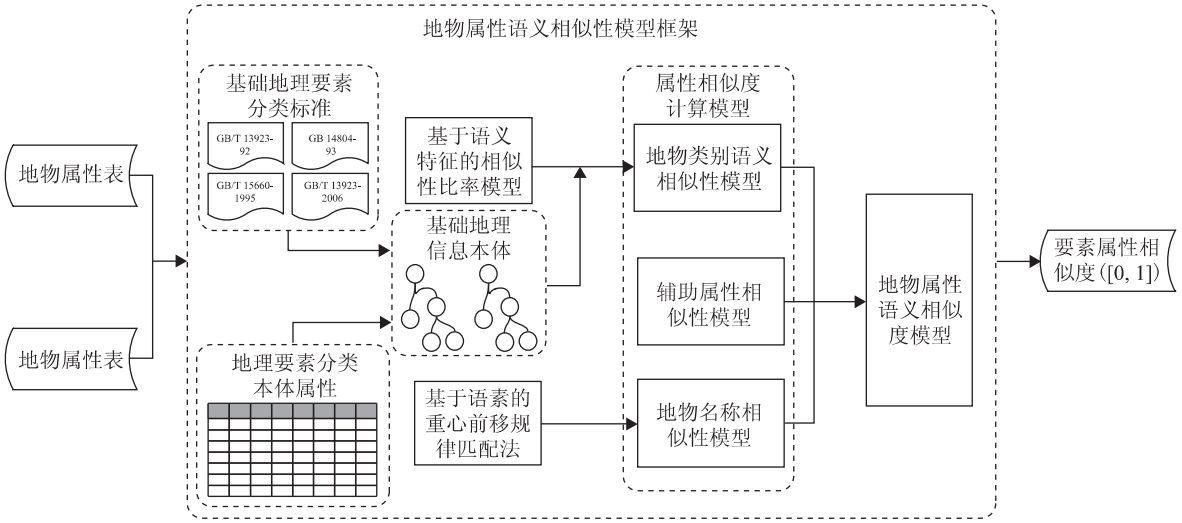


图 2 地物属性语义相似性模型框架

3.1 地理要素类别语义相似性

本文通过建立基础地理信息分类本体结构,结合分类概念的本体属性项,从语义内涵的角度采用本体属性支持的概念语义相似性模型[20]计算地理要素类别语义的相似度。在地理要素分类的本体属性描述中可能存在分号(";")、并集("∪")或交集("∩")的分隔符号的情况,如"主干道"与"岩石滩"的"地域性"本体属性值分别为{"城市;街区"}和{"海岸线∩干出线"}。本文根据不同的情况采取不同的策略,若是分号分隔符采取字符串分离策略,将一个属性值分割为两个子属性值,即{"城市;街区"}={"城市","街区"};若是并集分隔符,则认为属性值可只包含其中一项,即{"海岸线∪干出线"}={"海岸线","干出线"};若是交集分隔符,则认为属性值中必须包含所有值才算相同。

同时还规定:若待匹配的地理要素类别间关系为上下义关系,且父类别所属的地理要素的比例尺小于子类别的,则认为二者在语义上是一致的,即 $\mathrm{Sim}_{\mathrm{class}}(e_1,e_2)=1$。例如,计算 1:250 000 比例尺下的地理类别为"街道"的要素 e_1 与 1:50 000 比例尺下的地理类别为"次干道"要素 e_2 的类别语义相似度;按照《GB/T 13923—2006》规定,在 1:250 000 比例尺下地图中不采集"次干道"类别的要素,同时因为"街道"是"次干道"的父类别,所以可认为在此条件下 $\mathrm{Sim}_{\mathrm{class}}$("街道","次干道")= 1。

3.2 地理要素名称属性相似性

地理要素的名称属性的相似性是判断地物是否相似的最为直观指标。由于地物的名称十分复杂,不可能通过建立本体结构实现其相似性比较,因此本文采用基于字面的相似度算法。通过分析地理要素的命名规律,发现地物名称的匹配与重心后移规律[21]的理念类似,但关注的语义重点位置不同。专业词汇关注要素类别是否相同,语义重点位于字符串的后部分;而地物名称关注于地物是否相同,语义重点集中于字符串的前部分。例如"关山大道"与"雄楚大道",对于要素分类相似性度量,关注点在于要素类别均为"大道",因此属于相同类别的可能性较大;而对于地物名称相似性度量,关注点在地物位

置分别为“关山”与“雄楚”，因此属于相同地物的可能性较小。

本文通过改进重心后移规律匹配法，将模型中的“匹配序”改进为“匹配逆序”，以符合地物名称的特点。基于重心前移的相似性模型如下所示：

$$\mathrm{Sim}(s_1,s_2)=\alpha\times\frac{1}{2}\times\left(\frac{k}{m}+\frac{k}{n}\right)+\beta\times\min\left(\frac{m}{k},\frac{k}{m}\right)\times\frac{1}{2}\times\left(\frac{\sum_{i=1}^{c}L_1(i)}{\sum_{t=1}^{m}t}+\frac{\sum_{i=1}^{c}L_2(i)}{\sum_{p=1}^{n}p}\right) \tag{1}$$

其中，α 表示 s_1，s_2 中含有相同语素个数的影响权重，β 表示相同语素在各个词中的位置关系的影响权重，满足 $\alpha+\beta=1$，通常推荐设置 $\alpha=0.6$，$\beta=0.4$[22]；m、n 分别是 s_1、s_2 的字符长度，c 是 s_1、s_2 匹配的字符数；$L_1(i)$、$L_2(i)$ 分别表示匹配字符 i 在 s_1、s_2 的匹配索引值。

3.3 辅助属性相似性

基础地理要素的属性数据不仅仅包含分类属性信息、要素名称信息，同时还存在描述辅助信息的属性字段。作为用户通过地理要素全面了解地物特征的字段，辅助属性的语义相似度计算也是不可或缺的。

1）基本属性相似度

本文将基本数据类型属性划分为4种类型：有序标称属性、无序标称属性、数值型属性和字符型属性。

（1）有序标称属性相似度

具有一定顺序的2个及2个以上的离散状态值的属性称为有序标称属性。针对此类属性，可先将值域里的所有值按照升序排列，然后为值域中每个属性值进行编号。例如，一般道路要素的属性表中都存在“道路等级”属性，即“一级道路、二级道路、……”，因此我们将其按重要性升序排列后，分别赋以编号。其相似度计算函数为：

$$\mathrm{Sim}_{\mathrm{nomi}}(p,q)=1-\frac{\mathrm{num}(p)-\mathrm{num}(q)}{n} \tag{2}$$

其中，num函数为获取当前属性值所对应的编号；n 为属性值域中值的数量。

（2）无序标称属性相似度

无序标称属性是指无程度的差别或次序关系的标称属性。如颜色属性就是一个无序标称属性，它可能存在的值有：红色、绿色、蓝色等，这些属性值之间相互独立，互不交叉。其相似度计算函数为：

$$\mathrm{Sim}_{\mathrm{bool}}(p,q)=\begin{cases}1 & p=q\\ 0 & p\neq q\end{cases} \tag{3}$$

（3）数值型属性相似度

数值型属性值是采用数字进行描述，允许数字存在阈值范围内的误差。其相似度计算函数为：

$$\mathrm{Sim}_{\mathrm{bool}}(p,q)=\begin{cases}1 & |p-q|\leqslant\lambda\\ 0 & |p-q|>\lambda\end{cases} \tag{4}$$

其中，λ 为允许的阈值范围上限，即若 p 与 q 的差值在阈值范围内，则认定 p 与 q 相同，相似度为1；反之认为不相同。

（4）字符型属性相似度

字符型属性相似度模型与地理要素名称相似性模型一致，采用重心后移规律匹配法计算相似度，计算方法见公式（1）。

2）辅助属性字段相似度

在单属性相似度计算结果的基础上，结合不同辅助属性的重要性，本文提出了基于权重的地理要素辅助属性相似性模型：

$$\mathrm{Sim}_{\mathrm{ma}}(p,q)=\sum_{i=1}^{n}\omega_i\mathrm{Sim}_{\mathrm{oa}}(p_i,q_i) \tag{5}$$

其中，$\mathrm{Sim}_{\mathrm{oa}}(p_i,q_i)$ 是第 i 组属性对的相似度计算结果，计算公式根据前文提供相应的属性类型计算模型；ω_i 为第 i 组属性在概念表达中的权重，同时满足 $\sum_{i=1}^{n}\omega_i=1$；n 为要素属性的匹配字段对的个数，即仅计算待计算的地理数据均包含的属性字段，若一方存在而另一方没有的属性字段不纳入计算范围。

3.4 地理要素属性语义相似度

本文根据地理要素类别、地物名称以及辅助属性的重要性，提出基于权重的地物属性语义相似度模型：

$$\mathrm{Sim}_{\mathrm{attr}}(p,q)=\alpha\times\mathrm{Sim}_{\mathrm{class}}(p,q)+\beta\times\mathrm{Sim}_{\mathrm{name}}(p,q)+\gamma\times\mathrm{Sim}_{\mathrm{ma}}(p,q) \tag{6}$$

其中，α，β，γ 分别为分类语义相似度、名称语义相似度和辅助属性语义相似度的权重，满足 $\alpha+\beta+\gamma=1$。$\mathrm{Sim}_{\mathrm{class}}$ 是地理信息分类语义相似性模型；$\mathrm{Sim}_{\mathrm{name}}$ 是地理要素名称相似性模型；$\mathrm{Sim}_{\mathrm{ma}}$ 是辅助属性相似性模型；相似度计算结果满足 $\mathrm{Sim}_{\mathrm{attr}}\in[0,1]$。

4 实验及结果分析

在地理数据匹配的实验中，本文从某市1999年的1∶50 000比例尺、2009年1∶50 000比例尺与2009年1∶250 000比例尺的道路数据中提取部分要素作为实验数据。实验数据的属性信息如

表 1 所示。针对原始数据,先利用空间相似度算法计算各地理要素间的空间相似度,将在空间位置、形态等空间特征上较为相似的要素作为候选匹配集合。

本文进行两组实验,分别验证同尺度多时态与跨尺度相同时间版本情况下,候选地理要素间的属性相似程度,并确定地物的最佳匹配组合,以进一步验证本文提出的模型正确性。实验中涉及的地理要素类别的本体属性信息如表 2 所示。表 2 仅保留实验地理要素类别均拥有的本体属性项。另外,本文假定类别语义相似度模型中各本体属性权重值均相同;根据文献[22]的建议,设定名称属性相似度模型中"匹配度"与"匹配序"的权重分别为 0.6 与 0.4;设定辅助属性相似度中各辅助属性的权重均相同;设定"地物属性语义相似度"中类别属性、名称属性与辅助属性的权重分别为 0.5、0.4 与 0.1。

表 1 实验数据属性信息表

要素编号 (ID)	类别代码 (GB)	名称 (NAME)	车道数 (LANE)	单/双行线 (SDTF)	路宽/m (WIDTH)	时间版本 (TEMPORAL)	比例尺度 (SCALE)
A	430501	沿江大道	6	双	22.5	2009 年	1∶50 000
B	42104	黄陂街	2	双	7.5	2000 年	1∶50 000
C	42105	花楼街	2	双	7.5	2000 年	1∶50 000
D	42102	中山大道	6	双	22.5	2000 年	1∶50 000
E	42102	沿河街道	4	双	15	2000 年	1∶50 000
F	430501	沿河大道	4	双	15	2009 年	1∶50 000
G	430500	中山大道	6	双	22.5	2009 年	1∶250 000
H	430500	沿江大道	6	双	22.5	2009 年	1∶250 000
I	430500	沿河大道	4	双	15	2009 年	1∶250 000

表 2 地物类别的本体属性表[8]

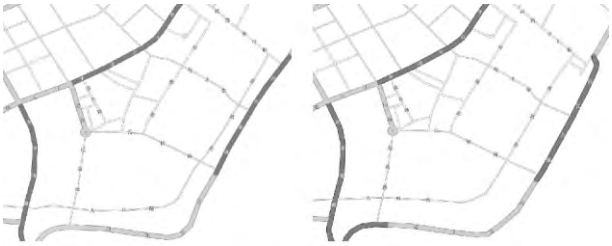
概念名称	显示比例尺	新标准分类代码	旧标准分类代码	成因	地域性	空间关系	功能	其他特性
街道	全比例尺	430500	42100	人工	城市∪街区	内	提供-道路运输服务∩(通行-机动车∪通行-非动车)	
主干道	全比例尺	430501	42102	人工	城市∪街区	内	提供-道路运输服务∩(通行-机动车∪通行-非动车)	主要
次干道	大比例尺、中比例尺	430502	42103	人工	城市∪街区	内	提供-道路运输服务∩(通行-机动车∪通行-非动车)	次要
支线	大比例尺、中比例尺	430503	42104	人工	街区	内	提供-道路运输服务∩(通行-机动车∪通行-非动车)	街道的最小单位
内部道路	大比例尺、中比例尺	430600	42105	人工	公园∪工矿区∪学校	内	提供-道路运输服务∩(通行-机动车∪通行-非动车)	

注:大比例尺为:大于 1∶5 000;中比例尺为:1∶5 000 ~ 1∶100 000;小比例尺为小于 1∶250 000;全比例尺:包含以上 3 种比例尺范围。

4.1 实验一:同尺度跨时间版本要素属性匹配实验

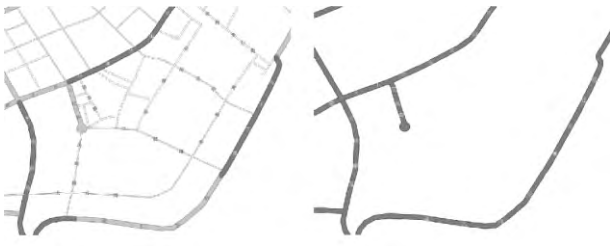
本实验以 1∶50 000 比例尺下,2000 年与 2009 年的道路数据中提取的部分候选匹配集为实验数据,要素的属性数据信息见表 1 中的要素 ABCDE,其中以要素 A 为更新数据,要素 BCDE 为候选集合中的数据。由于地理数据来源时期不同,因此存在新旧标准编码的情况,实验中涉及的 2000 年的要素类别代码为旧国家标准的 5 位编码,而 2009 年的要素类别代码为新国家标准的 6 位编码。道路数据图形如图 3 所示。

利用基础地理信息属性数据语义相似度模型,要素 BCDE 与要素 A 的属性相似度结果如表 3 所示。从表 3 的实验结果看到,候选匹配集合中与要素 A 的属性相似性,从高到低依次为:要素 E(沿江街道)、要素 D(中山大道)、要素 B(黄陂街)、要素 C(花楼街)"。结合表 1 与表 2 的信息发现,1)由于要素 A 与要素 E 的要素类别相同,要素分类语义相似



(a)2000年　(b)2009年

图3　实验道路数据



(a)比例尺为1:50 000　(b)比例尺为1:250 000

图4　实验道路数据

度高,而且二者在地物名称上也十分相近,虽然在辅助属性的相似度略低,但是二者的要素语义相似度仍然是最高;2)虽然要素D的分类代码(430501)与要素A的(42102)不相同,但是分析表2的地物要素类别的本体属性发现二者均表示相同的类别(即主干道),因此二者的要素类别是相同的;另外二者的名称差别较大,虽然辅助属性相似度也相等,但是重要性更高的地物名称语义相似性差距较大,因此要素A与要素D要素的语义相似度略低于其与要素E的;3)要素B与要素C在3类语义相似度上均存在较大差距,但是要素B的要素类别与要素A的更为相似(从表2中的"地域性"属性可以看到,"主干道"与"支线"均包含"街区"的属性值,而"内部道路"没有),因此要素B与要素A的属性语义相似度略高于要素C与要素A。实验结果符合实际更新过程。

表3　同尺度不同时间版本要素属性相似度结果

	要素A(沿江大道)			
	要素分类相似度	名称属性相似度	辅助属性相似度	要素属性相似度
要素E(沿江街道)	1	0.77	0.33	**0.841**
要素D(中山大道)	1	0.42	1	**0.768**
要素B(黄陂街)	0.7	0	0.33	**0.383**
要素C(花楼街)	0.6	0	0.33	**0.333**

4.2　实验二:同时期跨尺度要素属性匹配实验

本实验从2009年,1∶50 000与1∶250 000比例尺的道路数据中提取的部分候选匹配集为实验数据,要素的属性数据信息如表1所示中的要素FGHI,其中以要素F为数据尺度为1:50,000的更新数据,要素GHI分为1∶250 000的候选集合中的数据。道路数据图形如图4所示。

由于要素GHI的地图比例尺为1:250 000,属于"街道"的类别;要素F的地图比例尺为1:10 000,属于"次干道"的类别;"次干道"为"街道"的子类,因此本文认为此两类别在该情况下等效。根据基础地理信息属性数据语义相似度模型,要素GHI与要素A的属性相似度结果如表4所示。从表4的实验结果看到,候选匹配集合中与要素F的属性相似性,从高到低依次为:要素I(沿河大道)、要素H(沿江大道)、要素G(中山大道)。结合表1与表2的信息发现,要素F与要素I的分类属性虽然不相同(分别为"主干道"与"道路"),但依据上文的约定,二者的要素分类在语义层面是等效的,因此相似度等于1;同时由于二者的名称属性和辅助属性也都完全相同,因此相似度最高;同样,要素H、G与要素F的要素分类也属于等效要素类别,要素分类语义相似度等于1,但要素H的名称属性与要素F的更为相似,因此其属性语义相似度较高;最后要素G与要素F的属性语义相似度相对低。实验结果表明在数据更新中,应该优先选择要素I进行更新操作。

表4　跨尺度相同时间版本要素属性相似度结果

	要素F(沿河大道)			
	要素分类相似度	名称属性相似度	辅助属性相似度	要素属性相似度
要素I(沿河大道)	1	1	1	**1**
要素H(沿江大道)	1	0.73	0.33	**0.825**
要素G(中山大道)	1	0.42	0.33	**0.701**

5　结束语

本文提出定量计算基础地理要素的语义相似度模型,以明确地理要素间属性数据的匹配程度,可应用于要素属性数据关联与变化检测过程中。相似性模型结合本体理论,利用基础地理要素本体属性集合,从要素分类、要素名称与辅助属性3个角度定量计算相似性。在本文结尾通过对交通道路网地理要素进行实例分析,证明本文提出的地理要素属性语义相似性模型计算结果符合人类认识心理且可行合

理。本文当前研究的关注点在于结合本体技术,提出要素属性相似性模型框架。未来研究工作重点将是通过引入层次分析法合理配置模型中不同指标的权重值以取代经验值,以及需进一步细化要素名称和辅助属性信息,有针对性地计算各种属性类型的相似度。

参考文献:

[1] 李德仁. 利用遥感影像进行变化检测[J]. 武汉大学学报:信息科学版,2003,28(S1):7-12.

[2] 李德仁,眭海刚,单杰. 论地理国情监测的技术支撑[J]. 武汉大学学报:信息科学版,2012,37(5):505-512,502.

[3] HOLT A,BENWELL G L. Using spatial similarity for exploratory spatial data analysis:some directions[C]. University of Otago,1997.

[4] EGENHOFER M J,MARK D M. Naive geography:spatial information theory:a theoretical basis for GIS[M]. Springer Berlin Heidelberg,1995:1-15.

[5] EUZENAT J,SHVAIKO P. Ontology matching[M]. Heidelberg:Springer-Verlag New York,2007.

[6] BITTNER T, DONNELLY M, SMITH B. A spatio-temporal ontology for geographic information integration[J]. International Journal of Geographical Information Science 2009,23(6):765-798.

[7] BUCCELLA A,CECHICH A,FILLOTTRANI P. Ontology-driven geographic information integration:a survey of current approaches[J]. Computers & Geosciences,2009,35(4):710-723.

[8] 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,2008,37(2):230-235,242.

[9] 谭永滨,朱海红,李霖,等. 数字城市框架下地理信息服务语义分析[J]. 地理与地理信息科学,2012,28(2):5-8,19.

[10] STOCK K,REITSMA F,OU Y,et al. To ontologise or not to ontologise:an information model for a geospatial knowledge infrastructure[J]. Computers & Geosciences,2012,45(4):98-108.

[11] STUDER R,BENJAMINS V R,FENSEL D. Knowledge engineering:principles and methods[J]. Data & Knowledge Engineering,1998,25(1/2):161-197.

[12] 王红,李霖,朱海红. 国家基础地理信息本体关键问题研究[M]. 北京:科学出版社:2011.

[13] BRIGHT M W,HURSON A R,PAKZAD S. Automated resolution of semantic heterogeneity in multidatabases[J]. Acm Transactions on Database Systems,1994,19(19):212-253.

[14] MCRAE K, BOISVERT S. Automatic semantic similarity priming[J]. Journal of Experimental Psychology Learning Memory & Cognition,1998,24(3):558-572.

[15] TVERSKY,A. Features of similarity[J]. Psychological Review 1977,84(4):327-352.

[16] RESNI P. Using information content to evaluate semantic similarity in a taxonomy:international joint conference on artificial intelligence[C]. San Francisco:Morgan Kaufmann Publishers Inc. 1995:448-453.

[17] LORD P W,STEVENS R D,BRASS A,et al. Investigating semantic similarity measures across the gene ontology:the relationship between sequence and annotation[J]. Bioinformatics,2003,19(10):1275-1283.

[18] LIN D. An information-theoretic definition of similarity:fifteenth international conference on machine learning[C]. San Francisco:Morgan Kaufmann Publishers Inc. 1998:296-304.

[19] LIU H,BAO H,XU D. Concept vector for semantic similarity and relatedness based on WordNet structure[J]. Journal of Systems & Software,2012,85(2):370-381.

[20] 谭永滨,李霖,王伟,等. 本体属性的基础地理信息概念语义相似性计算模型[J]. 测绘学报,2013,42(5):782-789.

[21] 吴志强. 经济信息检索后控制词表的研制[D]. 南京:南京农业大学,1999.

[22] 王源,吴晓滨,涂从文,等. 后控规范的计算机处理[J]. 现代图书情报技术,1993,24(2):4-7.