

引用格式:贾小斌,艾廷华,彭子凤. 基于特征匹配的土地利用数据语义邻近度计算[J]. 资源科学, 2017, 39(3): 408-417. [Jia X B, Ai T H, Peng Z F. Computation of semantic proximity in land-use data based on feature-matching[J]. *Resources Science*, 2017, 39(3): 408-417]. DOI: 10.18402/resci.2017.03.03

基于特征匹配的土地利用数据语义邻近度计算

贾小斌¹, 艾廷华², 彭子凤³

(1. 中电科新型智慧城市研究院有限公司, 深圳 518026;

2. 武汉大学资源与环境科学学院, 武汉 430079;

3. 深圳市规划国土房产信息中心, 深圳 518034)

摘要: 量化语义关系的判断是地理数据处理的关键,也是地理本体建模与表达的难点,土地利用数据作为典型的专题性地理数据,其语义特征上的邻近关系如何,是土地资源统计、整理、规划、管理和开发需要关注的问题。针对土地利用数据的语义特征,从《土地利用现状分类》的依据出发,结合其权属特征,建立了土地利用数据语义特征的多细节层次表达模型;随后应用特征匹配的方法,通过计算其在权属、覆盖成因、覆盖类型、用途、利用状态、植被类型、附着物性质、利用方式等方面的匹配关系得到土地利用数据语义邻近度的具体度量值,实现土地利用数据从语义建模到邻近度计算的整个过程。在试验中则以具有不同语义特征的土地利用数据实体为例进行语义邻近关系计算,并将试验结果与实际经验判断进行比较,证明该模型具有较强的实用性,计算结果符合人类的认知。

关键词: 土地利用数据;语义特征;邻近度;特征匹配

DOI: 10.18402/resci.2017.03.03

1 引言

土地利用数据是一种在语义特征上具有类型多层次划分的专题地图数据,其在语义特征上的邻近关系如何,是土地资源统计、整理、规划、管理和开发需要关心的问题。当前,国内外针对土地利用数据语义邻近关系计算方面的专题研究不多,在相关研究中涉及到的主要有:Oosterom建立了多边形群层次剖分GAP树结构中的语义层次树,但对层次树中任意两结点间的语义距离如何计算没有给出相应方法^[1]。艾廷华等针对土地利用数据综合中的语义邻近关系进行了定性讨论,将语义关系做出了邻近与非邻近的二值化判断^[2]。艾廷华等提出了一种基于有序量的语义邻近度计算方法,能够对土地利用数据进行有效的定量计算,但只是在“等级量”上确定了其间的邻近级别,在“差异量”层次上的量化描述还有待进一步加强^[3]。刘耀林等中提出了

基于本体方法的土地利用数据语义邻近度计算模型,可以很好表达土地利用数据的语义特征,但对语义邻近关系上综合度量方法论述不够充分^[4]。此外,李霖等学者针对其他类型的空间地理数据也进行了基于本体方法的语义邻近关系计算研究^[5-10]。

相对于空间距离,语义距离的计算难以找到有效的量化模型,涉及的因素多,属于多维参量计算问题,各因子难以量化,不同因子集成中难以匹配融合,一直是该领域的难点问题。本研究针对土地利用数据这种语义特征较强的专题地图数据提出一种全新的高效的语义邻近度计算方法,实现土地利用数据从语义建模到邻近关系计算的全过程。具体区分为以下二个步骤:①从《土地利用现状分类》^[11]中对土地利用类型的概念定义出发,对土地利用数据的语义特征进行形式化表达,构建出其在语义特征上的多细节层次表达模型;②运用特征

收稿日期:2016-03-16; 修订日期:2016-04-30

基金项目:数字制图与国土信息应用工程国家测绘地理信息局重点实验室开放基金(GCWD201403)。

作者简介:贾小斌,男,河南南阳人,博士,研究方向为土地资源数据整理。E-mail: jiaxiaobin_123@126.com

2017年3月

匹配的方法,通过计算土地利用数据在其权属、覆盖成因、覆盖类型、用途、利用状态、植被类型、附着物性质、利用方式等特征上的匹配关系进而得到其语义邻近关系的具体度量值。

2 土地利用数据语义的多细节层次表达

土地利用数据语义特征主要体现在其用地类型和权属2个方面。用地类型(土地利用现状分类)是以服务国土资源管理为主,主要依据土地的自然属性、覆盖特征、利用方式、土地用途、经营特点等因素进行综合性分类的结果^[1]。图1为土地利用现状分类情况,采用2个层次的分类体系,其中一级类有8类,二级类有38类。(土地)权属是指土地产权

的归属,是存在于土地之中的排他性权利。文中对权属采用二值化处理,只区分为相同和不同2种情形,不作深入讨论。综合土地利用数据的用地类型和权属特征,建立如图2所示的土地利用数据语义特征多细节层次表达模型,区分3个层次11个语义特征项,其中语义特征项用地类型和权属位于第1层,覆盖成因、覆盖类型、用途、利用状态位于第2层,茎的性质、生长习性、植被用途、附着物性质、利用方式位于第3层。

定义三元组 $I_i: \langle Name_i, \{v_{i1}, v_{i2}, \dots, v_{ij}\}, R_i \rangle$ 表示土地利用数据语义特征多细节层次表达模型中第 i 个语义特征项, $Name_i$ 表示该项名称, $\{v_{i1}, v_{i2}, \dots, v_{ij}\}$

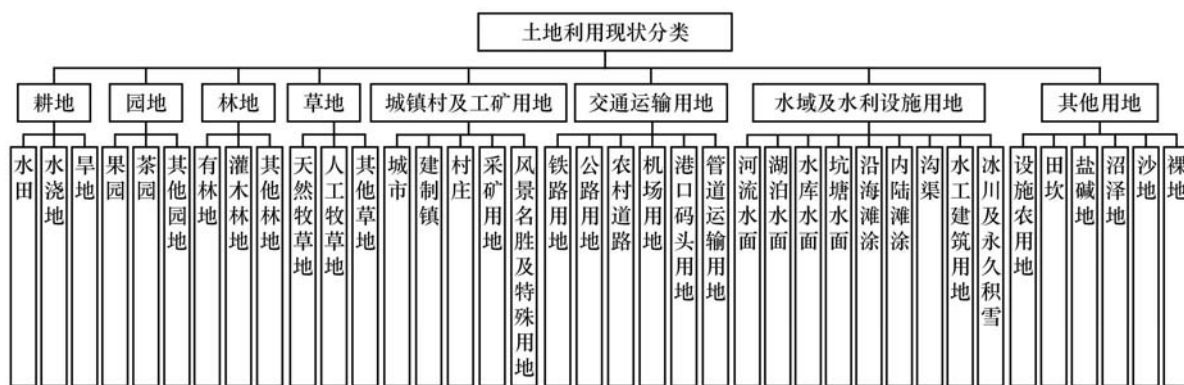


图1 土地利用现状分类(农村土地调查时)

Figure 1 Land use status category (for rural land survey)

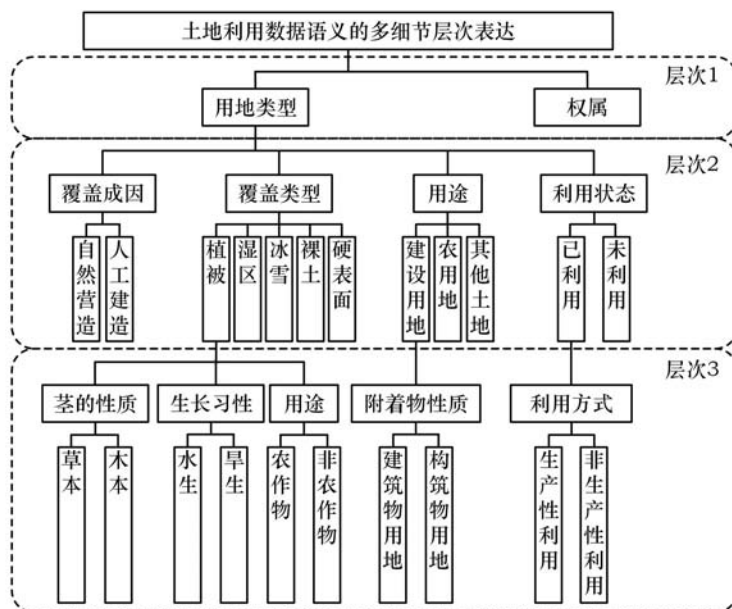


图2 土地利用数据语义的多细节层次表达

Figure 2 The hierarchical representation of level of details of land-use data

表示该项值域, j 表示值域中元素值个数, R_i 表示该项层次的级别, $R_i \in \{1, 2, 3\}$ 。如<覆盖成因, {自然营造, 人工建造}, 2>。则土地利用数据语义特征的多细节层次表达模型可表示为集合 $\{I_1, I_2, \dots, I_i, i=11\} = \{<权属, \{\dots\}, 1>, <用地类型, \{\dots\}, 1>, <覆盖成因, \{\dots\}, 2>, <覆盖类型, \{\dots\}, 2>, <用途, \{\dots\}, 2>, <利用状态, \{\dots\}, 2>, <茎的性质, \{\dots\}, 3>, <生长习性, \{\dots\}, 3>, <植被用途, \{\dots\}, 3>, <附着物性质, \{\dots\}, 3>, <利用方式, \{\dots\}, 3>\}$ 。对任意土地利用数据实体对象 E_i , 定义函数 $V_{ir}(E_i, I)$ 代表其在语义特征项 $I: <Name, \{v_1, v_2, \dots, v_i\}, R>$ 的取值, 若实体 E_i 在语义特征项 I 上有语义特征, 则 $V_{ir}(E_i, I) = I.v$; 反之, $V_{ir}(E_i, I) = 0$ 。定义实体 E_i 的语义特征表达函数为 $G_{sf}(E_i)$, 则其语义特征可表达为: $G_{sf}(E_i) = \{I_1.v, I_2.v, \dots, I_n.v, n=11\}$ 。如区域中权属为 OS , 用地类型为水田的实体 E , 其语义特征可表达为 $G_{sf}(E) = \{OS, 水田, 人工建造, 植被, 农用地, 已利用, 草本, 水生, 农作物, 0, 生产性利用\}$ 。

3 基于特征匹配的土地利用数据语义邻近关系计算

语义邻近度是指对象在语义关系描述上的邻近程度, 其计算模型有基于特征匹配的模型、基于语义关系的模型以及二者的混合模型^[12-14, 16]。这里应用基于特征匹配的模型来进行土地利用数据语义邻近度计算。

3.1 基于特征匹配的邻近度计算模型

基于特征匹配的模型是依据概念或对象的特征(如性质、作用和属性)来判断语义邻近度, 共同的属性特征会增加两个概念的邻近度, 反之则降低。如图3所示, 若对象 a 、 b 分别有特征集合 A 和 B , $A \cap B$ 表示 a 与 b 的相同特征的集合, $A - B$ 表示属

于 a 但不属于 b 的特征集合, $B - A$ 表示属于 b 但不属于 a 的特征集合, 则基于特征匹配的语义邻近度计算的比率模型可表示为公式(1)^[13]: 其中, 参数 α, β 为修正系数, G 为密度函数。邻近度 $S(a, b)$ 在 0~1 之间, 值为 1 表示对象 a 和 b 所有特征相同, 反之, 值为 0 则表示对象 a 和 b 无相同特征。

$$S(a, b) = \frac{G(A \cap B)}{G(A \cap B) + \alpha G(A - B) + \beta G(B - A)} \quad (1)$$

$$\alpha \geq 0, \beta \geq 0$$

3.2 土地利用数据语义邻近度计算

运用公式(1)中模型对土地利用数据进行语义邻近关系计算, 需首先在同一语义特征项上进行匹配, 随后在此基础上确定土地利用数据语义邻近关系的整体度量值。

(1) 单一特征项上的语义匹配关系。土地利用数据在单一语义特征项上的匹配关系可区分为三种情形: 即取值相同或不同、没有取值。为了方便计算, 这里分别用具体值来代替。定义土地利用数据实体 E_1 和 E_2 在语义特征项 $I: <Name, \{v_1, v_2, \dots, v_i\}, R>$ 上匹配关系函数 $G_{ir}(E_1, E_2, I)$, 则 $G_{ir}(E_1, E_2, I)$ 可表达为公式(2)所示。若实体 E_1 和 E_2 在语义特征项 I 上任意一个没有取值, 则 $G_{ir}(E_1, E_2, I) = 0$; 若 E_1 和 E_2 有相同取值, 则 $G_{ir}(E_1, E_2, I) = 1$; 若 E_1 和 E_2 取值不同, 则 $0 < G_{ir}(E_1, E_2, I) < 1$ 。

$$G_{ir}(E_1, E_2, I) = \begin{cases} 1 & V_{ir}(E_1, I) = V_{ir}(E_2, I) \neq 0 \\ k & 0 < k < 1, V_{ir}(E_1, I) \neq V_{ir}(E_2, I) \\ 0 & V_{ir}(E_1, I) \times V_{ir}(E_2, I) = 0 \end{cases} \quad (2)$$

(2) 权属和用地类型上的定性分析。土地利用数据语义特征主要体现在其用地类型和权属两个方面。若土地利用数据实体在权属特征与用地类型上相同, 则语义邻近度取值为 1。(土地) 权属是指土地产权的归属, 是存在于土地之中的排他性权利。若权属不同, 则进行语义邻近关系计算没有任何意义, 令其语义邻近度取值为 0。若权属相同用地类型不同, 其语义邻近度取值在 0~1 之间。对任意土地利用数据实体 E_1 和 E_2 , 定义函数 $G_{smb}(E_1, E_2)$

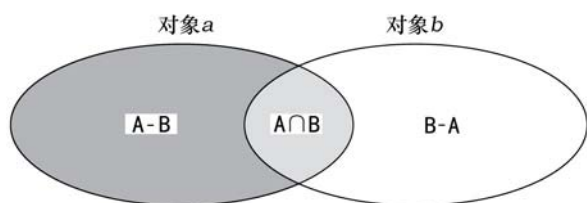


图3 对象 a 和 b 相似性度量的维恩图表达

Figure 3 Computation the similarity of object a and b represented by Venn diagram

2017年3月

表示其语义邻近度, $V_{os}(E_1)$ 和 $V_{LUT}(E_1)$ 分别表示实体 E_1 在特征项权属和用地类型上的取值, 则 $G_{SND}(E_1, E_2)$ 可表达为公式(3)。

$$G_{SND}(E_1, E_2) = \begin{cases} 0 & V_{os}(E_1) \neq V_{os}(E_2) \\ 0 < k < 1 & V_{os}(E_1) = V_{os}(E_2) \\ k & V_{LUT}(E_1) \neq V_{LUT}(E_2) \\ 1 & V_{os}(E_1) = V_{os}(E_2) \\ & V_{LUT}(E_1) = V_{LUT}(E_2) \end{cases} \quad (3)$$

(3) 定量计算。若土地利用数据实体 E_1 和 E_2 权属相同但用地类型不同, 将其语义特征分别用 $G_{SF}(E_1)$ 和 $G_{SF}(E_2)$ 表示, 依据特征匹配的比率模型(公式(1)), 其语义邻近度 $G_{SND}(E_1, E_2)$ 可表达为公式(4):

$$G_{SND}(E_1, E_2) = G(G_{SF}(E_1) \cap G_{SF}(E_2)) / [G(G_{SF}(E_1) \cap G_{SF}(E_2)) + \alpha G(G_{SF}(E_1) - G_{SF}(E_2)) + \beta G(G_{SF}(E_2) - G_{SF}(E_1))] \quad (4)$$

为简化表达, 定义函数 $G_{SED}(E_1, E_2)$ 、 $G_{SUD}(E_1, E_2)$ 和 $G_{SUD}(E_2, E_1)$, 其中 $G_{SED}(E_1, E_2)$ 代表实体 E_1 和 E_2 语义特征相同部分的度量值, 即 $G_{SED}(E_1, E_2) = G(G_{SF}(E_1) \cap G_{SF}(E_2))$; 函数 $G_{SUD}(E_1, E_2)$ 代表实体 E_1 和 E_2 语义特征不同的部分中属于 E_1 而不属于 E_2 的度量值, 即 $G_{SUD}(E_1, E_2) = G(G_{SF}(E_1) - G_{SF}(E_2))$; 则函数 $G_{SUD}(E_2, E_1)$ 代表实体 E_2 和 E_1 语义特征不同的部分中属于 E_2 而不属于 E_1 的度量值, 即 $G_{SUD}(E_2, E_1) = G(G_{SF}(E_2) - G_{SF}(E_1))$ 。则实体 E_1 和 E_2 的语义邻近度 $G_{SND}(E_1, E_2)$ 可表达为公式(5):

$$G_{SND}(E_1, E_2) = \frac{G_{SED}(E_1, E_2)}{G_{SED}(E_1, E_2) + \alpha G_{SUD}(E_1, E_2) + \beta G_{SUD}(E_2, E_1)} \quad (5)$$

实体不同级别的语义特征对其邻近关系影响的权重不同, 一般情况下, 权重与级别高低成反比。定义函数 $G_w(I_i \times R_i)$ 代表语义特征项 I_i 所在级别 R_i 对实体语义邻近关系影响的权重, 则 $G_{SED}(E_1, E_2)$ 、 $G_{SUD}(E_1, E_2)$ 和 $G_{SUD}(E_2, E_1)$ 可分别表示为公式(6)-公式(8):

$$G_{SED}(E_1, E_2) = G(G_{SF}(E_1) \cap G_{SF}(E_2)) = \sum_{i=3}^{11} G_w(I_i \times R_i), \quad G_{IR}(E_1, E_2, I_i) = 1 \quad (6)$$

$$G_{SUD}(E_1, E_2) = G(G_{SF}(E_1) - G_{SF}(E_2)) = \sum_{i=3}^{11} G_w(I_i \times R_i), \quad 0 < G_{IR}(E_1, E_2, I_i) < 1 \quad (7)$$

$$G_{SUD}(E_2, E_1) = G(G_{SF}(E_2) - G_{SF}(E_1)) = \sum_{i=3}^{11} G_w(I_i \times R_i), \quad 0 < G_{IR}(E_1, E_2, I_i) < 1 \quad (8)$$

3.3 土地利用数据语义邻近度计算步骤

对任意土地利用数据实体 E_1 和 E_2 , 其语义邻近度 $G_{SND}(E_1, E_2)$ 的计算步骤如下:

(1) 进行语义特征的多细节层次表达, 即: $G_{SF}(E_1) = E_1 \times \{I_1.v, I_2.v, \dots, I_n.v, n=11\}$, $G_{SF}(E_2) = E_2 \times \{I_1.v, I_2.v, \dots, I_n.v, n=11\}$ 。

(2) 权属和用地类型上的定性分析, 若 $G_{os}(E_1) \neq G_{os}(E_2)$, 则: $G_{SND}(E_1, E_2) = 0$; 若 $G_{os}(E_1) = G_{os}(E_2)$ 且 $G_{LUT}(E_1) = G_{LUT}(E_2)$, 则 $G_{SND}(E_1, E_2) = 1$ 。

(3) 定量计算, 当 $G_{os}(E_1) = G_{os}(E_2)$ 且 $G_{LUT}(E_1) \neq G_{LUT}(E_2)$, 用公式(6)-公式(8)计算实体 E_1 和 E_2 语义特征中相同部分和不同部分的度量值, 令 $G_{SED}(E_1, E_2) = m$ 、 $G_{SUD}(E_1, E_2) = n$ 、 $G_{SUD}(E_2, E_1) = k$, 则 $G_{SND}(E_1, E_2) = m / (m + \alpha \times n + \beta \times k)$ 。

4 试验

以农村土地调查数据为例, 若有权属特征为 OS 的土地利用数据实体集合 $\{E_1, E_2, E_3, \dots, E_i, i=38\}$, 其用地类型分别对应图1中的38个二级类, 用上述语义邻近度计算方法进行语义邻近关系的计算。

(1) 对该38个实体进行语义特征的多细节层次表达, 如 $G_{SF}(E_1) = \{OS, \text{水田}, \text{人工建造}, \text{植被}, \text{农用地}, \text{已利用}, \text{草本}, \text{水生}, \text{农作物}, 0, \text{生产性利用}\}$; $G_{SF}(E_2) = \{OS, \text{水浇地}, \text{人工建造}, \text{植被}, \text{农用地}, \text{已利用}, \text{草本}, \text{旱生}, \text{农作物}, 0, \text{生产性利用}\}$ 。其他实体语义特征的多细节层次表达见表1。具有相同二级类的土地利用实体在同一个语义特征项的取值并不一定具有唯一性, 为了方便后续相似度的计算, 这里针对本文中的38个特定实体, 限定其在各个语义特征项上的具有单一取值。

表 1 土地利用数据语义特征的多细节层次表达

Table 1 Semantic characteristic of land-use data expression

土地利用数据语义特征的多细节层次表达	
E_1	{OS,水田,人工建造,植被,农用地,已利用,草本,水生,农作物,0,生产性利用}
E_2	{OS,水浇地,人工建造,植被,农用地,已利用,草本,旱生,农作物,0,生产性利用}
E_3	{OS,旱地,人工建造,植被,农用地,已利用,草本,旱生,农作物,0,生产性利用}
E_4	{OS,果园,人工建造,植被,农用地,已利用,木本,旱生,农作物,0,生产性利用}
E_5	{OS,茶园,人工建造,植被,农用地,已利用,木本,旱生,农作物,0,生产性利用}
E_6	{OS,其他园地,人工建造,植被,农用地,已利用,木本,旱生,农作物,0,生产性利用}
E_7	{OS,有林地,自然营造,植被,农用地,已利用,木本,旱生,非农作物,0,生产性利用}
E_8	{OS,灌木林地,自然营造,植被,农用地,已利用,木本,旱生,非农作物,0,生产性利用}
E_9	{OS,其他林地,人工建造,植被,农用地,已利用,木本,旱生,非农作物,0,生产性利用}
E_{10}	{OS,天然牧草地,自然营造,植被,农用地,已利用,草本,旱生,非农作物,0,生产性利用}
E_{11}	{OS,人工牧草地,人工建造,植被,农用地,已利用,草本,旱生,非农作物,0,生产性利用}
E_{12}	{OS,其他草地,自然营造,植被,其他土地,未利用,草本,旱生,非农作物,0,0}
E_{13}	{OS,城市,人工建造,硬表面,建设用地,已利用,0,0,0,建筑物用地,非生产性利用}
E_{14}	{OS,建制镇,人工建造,硬表面,建设用地,已利用,0,0,0,建筑物用地,非生产性利用}
E_{15}	{OS,村庄,人工建造,硬表面,建设用地,已利用,0,0,0,建筑物用地,非生产性利用}
E_{16}	{OS,采矿用地,人工建造,硬表面,建设用地,已利用,0,0,0,构筑物用地,生产性利用}
E_{17}	{OS,风景名胜及特殊用地,人工建造,硬表面,建设用地,已利用,0,0,0,建筑物用地,非生产性利用}
E_{18}	{OS,铁路用地,人工建造,硬表面,建设用地,已利用,0,0,0,构筑物用地,非生产性利用}
E_{19}	{OS,公路用地,人工建造,硬表面,建设用地,已利用,0,0,0,构筑物用地,非生产性利用}
E_{20}	{OS,农村道路,人工建造,硬表面,农用地,已利用,0,0,0,0,非生产性利用}
E_{21}	{OS,机场用地,人工建造,硬表面,建设用地,已利用,0,0,0,构筑物用地,非生产性利用}
E_{22}	{OS,港口码头用地,人工建造,硬表面,建设用地,已利用,0,0,0,构筑物用地,非生产性利用}
E_{23}	{OS,管道运输用地,人工建造,硬表面,建设用地,已利用,0,0,0,构筑物用地,非生产性利用}
E_{24}	{OS,河流水面,自然营造,湿区,其他土地,未利用,0,0,0,0,0}
E_{25}	{OS,湖泊水面,自然营造,湿区,其他土地,未利用,0,0,0,0,0}
E_{26}	{OS,水库水面,人工建造,湿区,建设用地,已利用,0,0,0,构筑物用地,非生产性利用}
E_{27}	{OS,坑塘水面,人工建造,湿区,农用地,已利用,0,0,0,0,非生产性利用}
E_{28}	{OS,沿海滩涂,自然营造,湿区,其他土地,未利用,0,0,0,0,0}
E_{29}	{OS,内陆滩涂,自然营造,湿区,其他土地,未利用,0,0,0,0,0}
E_{30}	{OS,沟渠,人工建造,硬表面,农用地,已利用,0,0,0,0,非生产性利用}
E_{31}	{OS,水工建筑用地,人工建造,硬表面,建设用地,已利用,0,0,0,构筑物用地,非生产性利用}
E_{32}	{OS,冰川及永久积雪,自然营造,冰雪,其他土地,未利用,0,0,0,0,0}
E_{33}	{OS,设施农用地,人工建造,硬表面,农用地,已利用,0,0,0,0,生产性利用}
E_{34}	{OS,田坎,人工建造,植被,农用地,已利用,草本,旱生,非农作物,0,非生产性利用}
E_{35}	{OS,盐碱地,自然营造,植被,其他土地,未利用,草本,旱生,非农作物,0,0}
E_{36}	{OS,沼泽地,自然营造,植被,其他土地,未利用,草本,水生,非农作物,0,0}
E_{37}	{OS,沙地,自然营造,裸土,其他土地,未利用,0,0,0,0,0}
E_{38}	{OS,裸地,自然营造,裸土,其他土地,未利用,0,0,0,0,0}

(2)计算该38个实体相互间的语义邻近度。在实体语义邻近度计算时,土地的自然属性和社会经济属性上的语义特征是用地类型在不同语义层次上的反映,若将用地类型加入到语义邻近度的计算中,有重复之嫌,因此计算时将其排除。设置公式

(5)中参数 $\alpha=\beta=1$,令公式(6)、公式(7)和公式(8)中的权重函数 $G_w(I_i \times R_i) = I_i \times R_i$, $3 \leq i \leq 11$,分别求得 $G_{SED}(E_i, E_j)$ 、 $G_{SUD}(E_i, E_j)$ 、 $G_{SUD}(E_j, E_i)$ 以及 $G_{SND}(E_i, E_j)$ 。如对实体 E_1 和 E_2 ,其在语义特征项

2017年3月

{覆盖成因,覆盖类型,用途,利用状态,植被茎的性质,植被用途,利用方式}上有相同语义特征,则 $G_{SED}(E_1, E_2)=2+2+2+2+3+3+3=17$;其在语义特征项{植被生长习性}上有不同语义特征,则: $G_{SUD}(E_1, E_2)$

$$= G_{SUD}(E_2, E_1)=3; G_{SND}(E_1, E_2)=17/(17+3+3)\approx 0.74。$$

其他语义邻近度计算结果见表2。

(3)结果分析。在实体权属相同情况下,实体在用地类型上的语义关系越近,其语义邻近度取值

表2 土地利用数据语义邻近度计算结果

Table 2 Semantic neighboring degree among land-use data

G_{SND}	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9	E_{10}	E_{11}	E_{12}	E_{13}	E_{14}	E_{15}	E_{16}	E_{17}	E_{18}	E_{19}
E_1	1	0.74	0.74	0.54	0.54	0.54	0.29	0.29	0.38	0.43	0.54	0.14	0.10	0.10	0.10	0.18	0.10	0.10	0.10
E_2	0.74	1	1.00	0.74	0.74	0.74	0.43	0.43	0.54	0.60	0.74	0.25	0.10	0.10	0.10	0.18	0.10	0.10	0.10
E_3	0.74	1.00	1	0.74	0.74	0.74	0.43	0.43	0.54	0.60	0.74	0.25	0.10	0.10	0.10	0.18	0.10	0.10	0.10
E_4	0.54	0.74	0.74	1	1.00	1.00	0.60	0.60	0.74	0.43	0.54	0.14	0.10	0.10	0.10	0.18	0.10	0.10	0.10
E_5	0.54	0.74	0.74	1.00	1	1.00	0.60	0.60	0.74	0.43	0.54	0.14	0.10	0.10	0.10	0.18	0.10	0.10	0.10
E_6	0.54	0.74	0.74	1.00	1.00	1	0.60	0.60	0.74	0.43	0.54	0.14	0.10	0.10	0.10	0.18	0.10	0.10	0.10
E_7	0.29	0.43	0.43	0.60	0.60	0.60	1	1.00	0.82	0.74	0.60	0.33	0.05	0.05	0.05	0.12	0.05	0.05	0.05
E_8	0.29	0.43	0.43	0.60	0.60	0.60	1.00	1	0.82	0.74	0.60	0.33	0.05	0.05	0.05	0.12	0.05	0.05	0.05
E_9	0.38	0.54	0.54	0.74	0.74	0.74	0.82	0.82	1	0.60	0.74	0.25	0.10	0.10	0.10	0.18	0.10	0.10	0.10
E_{10}	0.43	0.60	0.60	0.43	0.43	0.43	0.74	0.74	0.60	1	0.82	0.48	0.05	0.05	0.05	0.12	0.05	0.05	0.05
E_{11}	0.54	0.74	0.74	0.54	0.54	0.54	0.60	0.60	0.74	0.82	1	0.38	0.10	0.10	0.10	0.18	0.10	0.10	0.10
E_{12}	0.14	0.25	0.25	0.14	0.14	0.14	0.33	0.33	0.25	0.48	0.38	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
E_{13}	0.10	0.10	0.10	0.10	0.10	0.10	0.05	0.05	0.10	0.05	0.10	0.00	1	1.00	1.00	0.40	1.00	0.65	0.65
E_{14}	0.10	0.10	0.10	0.10	0.10	0.10	0.05	0.05	0.10	0.05	0.10	0.00	1.00	1	1.00	0.40	1.00	0.65	0.65
E_{15}	0.10	0.10	0.10	0.10	0.10	0.10	0.05	0.05	0.10	0.05	0.10	0.00	1.00	1.00	1	0.40	1.00	0.65	0.65
E_{16}	0.18	0.18	0.18	0.18	0.18	0.18	0.12	0.12	0.18	0.12	0.18	0.00	0.40	0.40	0.40	1	0.40	0.65	0.65
E_{17}	0.10	0.10	0.10	0.10	0.10	0.10	0.05	0.05	0.10	0.05	0.10	0.00	1.00	1.00	1.00	0.40	1	0.65	0.65
E_{18}	0.10	0.10	0.10	0.10	0.10	0.10	0.05	0.05	0.10	0.05	0.10	0.00	0.65	0.65	0.65	0.65	0.65	1	1.00
E_{19}	0.10	0.10	0.10	0.10	0.10	0.10	0.05	0.05	0.10	0.05	0.10	0.00	0.65	0.65	0.65	0.65	0.65	1.00	1
E_{20}	0.18	0.18	0.18	0.18	0.18	0.18	0.11	0.11	0.18	0.11	0.18	0.00	0.47	0.47	0.47	0.27	0.47	0.47	0.47
E_{21}	0.10	0.10	0.10	0.10	0.10	0.10	0.05	0.05	0.10	0.05	0.10	0.00	0.65	0.65	0.65	0.65	0.65	1.00	1.00
E_{22}	0.10	0.10	0.10	0.10	0.10	0.10	0.05	0.05	0.10	0.05	0.10	0.00	0.65	0.65	0.65	0.65	0.65	1.00	1.00
E_{23}	0.10	0.10	0.10	0.10	0.10	0.10	0.05	0.05	0.10	0.05	0.10	0.00	0.65	0.65	0.65	0.65	0.65	1.00	1.00
E_{24}	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.05	0.00	0.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00
E_{25}	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.05	0.00	0.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00
E_{26}	0.10	0.10	0.10	0.10	0.10	0.10	0.05	0.05	0.10	0.05	0.10	0.00	0.47	0.47	0.47	0.47	0.47	0.75	0.75
E_{27}	0.18	0.18	0.18	0.18	0.18	0.18	0.11	0.11	0.18	0.11	0.18	0.00	0.33	0.33	0.33	0.17	0.33	0.33	0.33
E_{28}	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.05	0.00	0.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00
E_{29}	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.05	0.00	0.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00
E_{30}	0.18	0.18	0.18	0.18	0.18	0.18	0.11	0.11	0.18	0.11	0.18	0.00	0.47	0.47	0.47	0.27	0.47	0.47	0.47
E_{31}	0.10	0.10	0.10	0.10	0.10	0.10	0.05	0.05	0.10	0.05	0.10	0.00	0.65	0.65	0.65	0.65	0.65	1.00	1.00
E_{32}	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.05	0.00	0.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00
E_{33}	0.29	0.29	0.29	0.29	0.29	0.29	0.21	0.21	0.29	0.21	0.29	0.00	0.27	0.27	0.27	0.47	0.27	0.27	0.27
E_{34}	0.38	0.54	0.54	0.38	0.38	0.38	0.43	0.43	0.54	0.60	0.74	0.38	0.18	0.18	0.18	0.10	0.18	0.18	0.18
E_{35}	0.14	0.25	0.25	0.14	0.14	0.14	0.33	0.33	0.25	0.48	0.38	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
E_{36}	0.25	0.14	0.14	0.05	0.05	0.05	0.21	0.21	0.14	0.33	0.25	0.70	0.00	0.00	0.00	0.00	0.00	0.00	0.00
E_{37}	0.00	0.00	0.00	0.00	0.00	0	0.05	0.05	0.00	0.05	0.00	0.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00
E_{38}	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.05	0.00	0.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00

(续表2)

续表 2

G_{SND}	E_{20}	E_{21}	E_{22}	E_{23}	E_{24}	E_{25}	E_{26}	E_{27}	E_{28}	E_{29}	E_{30}	E_{31}	E_{32}	E_{33}	E_{34}	E_{35}	E_{36}	E_{37}	E_{38}
E_1	0.18	0.10	0.10	0.10	0.00	0.00	0.10	0.18	0.00	0.00	0.18	0.10	0.00	0.29	0.38	0.14	0.25	0.00	0.00
E_2	0.18	0.10	0.10	0.10	0.00	0.00	0.10	0.18	0.00	0.00	0.18	0.10	0.00	0.29	0.54	0.25	0.14	0.00	0.00
E_3	0.18	0.10	0.10	0.10	0.00	0.00	0.10	0.18	0.00	0.00	0.18	0.10	0.00	0.29	0.54	0.25	0.14	0.00	0.00
E_4	0.18	0.10	0.10	0.10	0.00	0.00	0.10	0.18	0.00	0.00	0.18	0.10	0.00	0.29	0.38	0.14	0.05	0.00	0.00
E_5	0.18	0.10	0.10	0.10	0.00	0.00	0.10	0.18	0.00	0.00	0.18	0.10	0.00	0.29	0.38	0.14	0.05	0.00	0.00
E_6	0.18	0.10	0.10	0.10	0.00	0.00	0.10	0.18	0.00	0.00	0.18	0.10	0.00	0.29	0.38	0.14	0.05	0.00	0.00
E_7	0.11	0.05	0.05	0.05	0.05	0.05	0.05	0.11	0.05	0.05	0.11	0.05	0.05	0.21	0.43	0.33	0.21	0.05	0.05
E_8	0.11	0.05	0.05	0.05	0.05	0.05	0.05	0.11	0.05	0.05	0.11	0.05	0.05	0.21	0.43	0.33	0.21	0.05	0.05
E_9	0.18	0.10	0.10	0.10	0.00	0.00	0.10	0.18	0.00	0.00	0.18	0.10	0.00	0.29	0.54	0.25	0.14	0.00	0.00
E_{10}	0.11	0.05	0.05	0.05	0.05	0.05	0.05	0.11	0.05	0.05	0.11	0.05	0.05	0.21	0.60	0.48	0.33	0.05	0.05
E_{11}	0.18	0.10	0.10	0.10	0.00	0.00	0.10	0.18	0.00	0.00	0.18	0.10	0.00	0.29	0.74	0.38	0.25	0.00	0.00
E_{12}	0.00	0.00	0.00	0.00	0.21	0.21	0.00	0.00	0.21	0.21	0.00	0.00	0.21	0.00	0.38	1.00	0.70	0.21	0.21
E_{13}	0.47	0.65	0.65	0.65	0.00	0.00	0.47	0.33	0.00	0.00	0.47	0.65	0.00	0.27	0.18	0.00	0.00	0.00	0.00
E_{14}	0.47	0.65	0.65	0.65	0.00	0.00	0.47	0.33	0.00	0.00	0.47	0.65	0.00	0.27	0.18	0.00	0.00	0.00	0.00
E_{15}	0.47	0.65	0.65	0.65	0.00	0.00	0.47	0.33	0.00	0.00	0.47	0.65	0.00	0.27	0.18	0.00	0.00	0.00	0.00
E_{16}	0.27	0.65	0.65	0.65	0.00	0.00	0.47	0.17	0.00	0.00	0.27	0.65	0.00	0.47	0.10	0.00	0.00	0.00	0.00
E_{17}	0.47	0.65	0.65	0.65	0.00	0.00	0.47	0.33	0.00	0.00	0.47	0.65	0.00	0.27	0.18	0.00	0.00	0.00	0.00
E_{18}	0.47	1.00	1.00	1.00	0.00	0.00	0.75	0.33	0.00	0.00	0.47	1.00	0.00	0.27	0.18	0.00	0.00	0.00	0.00
E_{19}	0.47	1.00	1.00	1.00	0.00	0.00	0.75	0.33	0.00	0.00	0.47	1.00	0.00	0.27	0.18	0.00	0.00	0.00	0.00
E_{20}	1	0.47	0.47	0.47	0.00	0.00	0.33	0.69	0.00	0.00	1.00	0.47	0.00	0.57	0.29	0.00	0.00	0.00	0.00
E_{21}	0.47	1	1.00	1.00	0.00	0.00	0.75	0.33	0.00	0.00	0.47	1.00	0.00	0.27	0.18	0.00	0.00	0.00	0.00
E_{22}	0.47	1.00	1	1.00	0.00	0.00	0.75	0.33	0.00	0.00	0.47	1.00	0.00	0.27	0.18	0.00	0.00	0.00	0.00
E_{23}	0.47	1.00	1.00	1	0.00	0.00	0.75	0.33	0.00	0.00	0.47	1.00	0.00	0.27	0.18	0.00	0.00	0.00	0.00
E_{24}	0.00	0.00	0.00	0.00	1	1.00	0.08	0.10	1.00	1.00	0.00	0.00	0.60	0.00	0.00	0.21	0.21	0.60	0.60
E_{25}	0.00	0.00	0.00	0.00	1.00	1	0.08	0.10	1.00	1.00	0.00	0.00	0.60	0.00	0.00	0.21	0.21	0.60	0.60
E_{26}	0.33	0.75	0.75	0.75	0.08	0.08	1	0.47	0.08	0.08	0.33	0.75	0.00	0.17	0.18	0.00	0.00	0.00	0.00
E_{27}	0.69	0.33	0.33	0.33	0.10	0.10	0.47	1	0.10	0.10	0.69	0.33	0.00	0.38	0.29	0.00	0.00	0.00	0.00
E_{28}	0.00	0.00	0.00	0.00	1.00	1.00	0.08	0.10	1	1.00	0.00	0.00	0.60	0.00	0.00	0.21	0.21	0.60	0.60
E_{29}	0.00	0.00	0.00	0.00	1.00	1.00	0.08	0.10	1.00	1	0.00	0.00	0.60	0.00	0.00	0.21	0.21	0.60	0.60
E_{30}	1.00	0.47	0.47	0.47	0.00	0.00	0.33	0.69	0.00	0.00	1	0.47	0.00	0.57	0.29	0.00	0.00	0.00	0.00
E_{31}	0.47	1.00	1.00	1.00	0.00	0.00	0.75	0.33	0.00	0.00	0.47	1	0.00	0.27	0.18	0.00	0.00	0.00	0.00
E_{32}	0.00	0.00	0.00	0.00	0.60	0.60	0.00	0.00	0.60	0.60	0.00	0.00	1	0.00	0.00	0.21	0.21	0.60	0.60
E_{33}	0.57	0.27	0.27	0.27	0.00	0.00	0.17	0.38	0.00	0.00	0.57	0.27	0.00	1	0.18	0.00	0.00	0.00	0.00
E_{34}	0.29	0.18	0.18	0.18	0.00	0.00	0.18	0.29	0.00	0.00	0.29	0.18	0.00	0.18	1	0.38	0.25	0.00	0.00
E_{35}	0.00	0.00	0.00	0.00	0.21	0.21	0.00	0.00	0.21	0.21	0.00	0.00	0.21	0.00	0.38	1	0.70	0.21	0.21
E_{36}	0.00	0.00	0.00	0.00	0.21	0.21	0.00	0.00	0.21	0.21	0.00	0.00	0.21	0.00	0.25	0.70	1	0.21	0.21
E_{37}	0.00	0.00	0.00	0.00	0.60	0.60	0.00	0.00	0.60	0.60	0.00	0.00	0.60	0.00	0.00	0.21	0.21	1	1.00
E_{38}	0.00	0.00	0.00	0.00	0.60	0.60	0.00	0.00	0.60	0.60	0.00	0.00	0.60	0.00	0.00	0.21	0.21	1.00	1

理应越大。如对实体 E_1 ,其用地类型为水田,与水田语义关系最近的应该是具有相同父类(耕地)的水浇地和旱地,计算结果显示:与实体 E_i 具有最大语义邻近度为实体 E_2 和 E_3 ,实体 E_1 和 E_2 、 E_1 和 E_3 语义邻近度相同,即 $G_{SND}(E_1, E_2)=G_{SND}(E_1, E_3)=0.74$,这是因为实体 E_1 和 E_2 、 E_3 在语义特征项{覆盖成因,

覆盖类型,用途,利用状态,植被茎的性质,植被用途,利用方式}上的取值相同,在语义特征项{植被生长习性}上取值不同,可见实体 E_1 与 E_2 、 E_3 的语义邻近度计算结果合理。对其他语义邻近度计算结果,这里不再一一举例分析,经笔者仔细审查,均与实际经验判断结果相符。

2017年3月

(4)应用实例。在土地利用数据整理过程中,进行尺度变换操作时,针对部分面积小于阈值的多边形实体,需按照与其位置邻近实体的邻近程度来确定瓜分面积多少,邻近程度表现为空间和语义关系的综合邻近程度^[2,17-26]。为了凸显本文的研究主题,这里在应用举例时,不考虑空间上的邻近关系,剖分位置也不详细讨论,具体参见文献^[2]。在图4a中,多边形实体 P_0 、 P_1 、 P_2 、 P_3 和 P_4 权属相同,其用地类型分别为水田、旱地、有林地、其他草地和设施农用地,其中实体 P_0 面积小于阈值。在土地利用数据整理过程中需将 P_0 剖分给 P_1 、 P_2 、 P_3 和 P_4 。由表2中可知实体 P_0 与 P_1 、 P_2 、 P_3 、 P_4 的语义邻近度分别为: $G_{SND}(P_0, P_1)=0.29$ 、 $G_{SND}(P_0, P_2)=0.43$ 、 $G_{SND}(P_0, P_3)=0.33$ 、 $G_{SND}(P_0, P_4)=0.21$ 。在图4b中,将实体 P_0 分割为 P_9 、 P_{10} 、 P_{11} 和 P_{12} ,其面积比例为: $G_{Area}(P_9):G_{Area}(P_{10}):G_{Area}(P_{11}):G_{Area}(P_{12})=G_{SND}(P_0, P_1):G_{SND}(P_0, P_2):G_{SND}(P_0, P_3):G_{SND}(P_0, P_4)=0.29:0.43:0.33:0.21$ 。在图4c中,依次将 P_9 与 P_1 、 P_{10} 与 P_2 、 P_{11} 与 P_3 、 P_{12} 与 P_4 合并。

5 结论

本文针对土地利用数据语义邻近关系提出了一种完整的计算方法,实现土地利用数据从语义建模到邻近度计算的全过程,通过实验结果与实际经验判断比较验证出该模型具有较强的实用性,计算结果符合人类的认知,并将语义邻近度的计算结果转化为空间可度量值应用到土地利用数据整理的尺度变换操作中以展示该研究的实用价值,该研究成果可应用在国土资源数据的统计、整理、规划、管理和开发中。主要创新在于:①提出一种新型土地

利用数据语义特征形式化表达方式,构建出土地利用数据语义特征的多细节层次表达模型;②将基于特征匹配的邻近度量模型应用到土地利用数据语义特征的邻近关系定量计算中。本研究的计算方法具有较强的可操作性,可推广应用到其他相关空间地理数据语义邻近关系计算中。

参考文献(References):

- [1] Oosterom P V. The GAP-tree, An Approach to On-the-Fly Map Generalization of an Area Partitioning[A]. Muller J C, Lagrange J P, Weibel R. GIS and Generalization: Methodology and Practice[C]. London: Taylor & Francis, 1995.
- [2] 艾廷华, 刘耀林. 土地利用数据综合中的聚合与融合[J]. 武汉大学学报(信息科学版), 2002, 27(5): 486-492. [Ai T H, Liu Y L. Aggregation and amalgamation in land-use data generalization [J]. *Geomatics and Information Science of Wuhan University*, 2002, 27(5): 486-492.]
- [3] 艾廷华, 杨帆, 李精忠. 第二次土地资源调查数据建库中的土地利用图综合缩编[J]. 武汉大学学报(信息科学版), 2010, 35(8): 887-891. [Ai T H, Yang F, Li J Z. Land-use data generalization for the database construction of the second land resource survey[J]. *Geomatics and Information Science of Wuhan University*, 2010, 35(8): 887-891.]
- [4] 刘耀林, 李红梅, 杨淳惠. 基于本体的土地利用数据综合研究[J]. 武汉大学学报(信息科学版), 2010, 35(8): 883-886. [Liu Y L, Li H M, Yang C H. Ontology based land use data generalization[J]. *Geomatics and Information Science of Wuhan University*, 2010, 35(8): 883-886.]
- [5] 李霖, 王红. 基于形式化本体的基础地理信息分类[J]. 武汉大学学报(信息科学版), 2006, 31(6): 523-526. [Li L, Wang H. Classification of fundamental geographic information based on

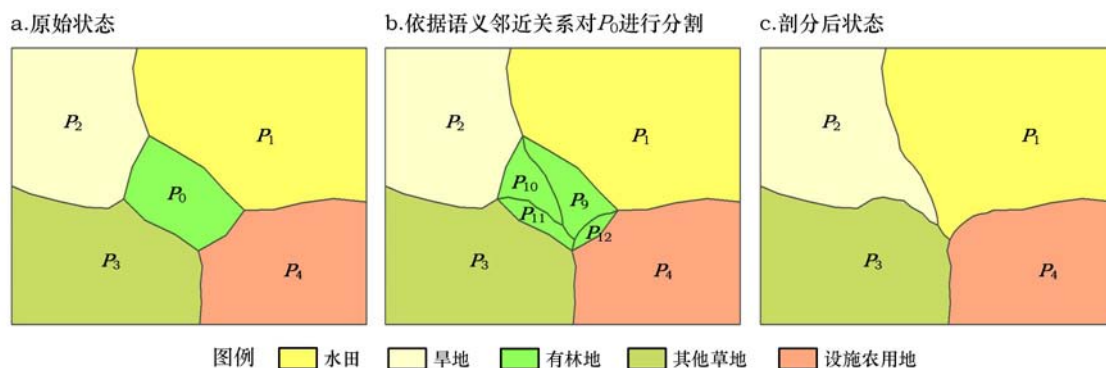


图4 土地利用数据语义邻近关系应用实例示意

Figure 4 An example for application of semantic neighboring degree in land-use data

- formal ontology[J]. *Geomatics and Information Science of Wuhan University*, 2006, 31(6): 523-526.]
- [6] 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. [Li L, Zhu H H, Wang H, *et al.* Semantic analyses of the fundamental geographic information based on formal ontology-exemplifying hydrological category[J]. *Acta Geodaetica Cartographica Sinica*, 2008, 37(2): 230-235.]
- [7] 谭永滨, 李霖, 王伟. 基于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. [Tan Y B, Li L, Wang W. Semantic similarity measurement model between fundamental geographic information concepts based on ontological property[J]. *Acta Geodaetica et Cartographica Sinica*, 2013, 42(5): 782-789.]
- [8] 郑茂辉, 冯学智, 蒋莹滢, 等. 基于描述逻辑本体的GIS多重表达[J]. *测绘学报*, 2006, 35(3): 261-266. [Zheng M H, Feng X Z, Jiang Y Y, *et al.* A formal approach for multiple representations in GIS based on DL ontologies[J]. *Acta Geodaetica et Cartographica Sinica*, 2006, 35(3): 261-266.]
- [9] 李德仁, 崔巍. 地理本体与空间信息多级网格[J]. *测绘学报*, 2006, 35(2): 143-148. [Li D R, Cui W. Geographic ontology and SIMG[J]. *Acta Geodaetica Cartographica Sinica*, 2006, 35(2): 143-148.]
- [10] 安杨, 边馥苓, 关佑红. GIS中地理本体的建立与比较[J]. *武汉大学学报(信息科学版)*, 2006, 31(12): 1108-1111. [An Y, Bian F L, Guan J H. Design and comparison of Geo ontology in GIS[J]. *Geomatics and Information Science of Wuhan University*, 2006, 31(12): 1108-1111.]
- [11] 国土资源部. 土地利用现状分类[EB/OL]. (2007-08-10)[2016-03-16]. <http://www.doc88.com/p-9049335535584.html>. [Ministry of Land and Resources. Classification of Land-use Status [EB/OL]. (2007-08-10)[2016-03-16]. <http://www.doc88.com/p-9049335535584.html>.]
- [12] Rips L, Shoben E, Smith E. Semantic distance and the verification of semantic relations[J]. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12(1): 1-20.
- [13] versky A. Features of similarity[J]. *Psychological Review*, 1977, 84(4): 327-352.
- [14] Rodríguez M, Egenhofer M, Rugg R. Assessing Semantic Similarities among Geospatial Feature Class Definitions[A]. Vckovski A, Brassel K, Schek H. *Interoperating Geographic Information Systems*[M]. Berlin: Springer, 1999.
- [15] Feng C C, Flewelling D M. Assessment of semantic similarity between land use/land cover classification systems[J]. *Computers, Environment and Urban Systems*, 2004, 28(3): 229-246.
- [16] Schwering A, Raubal M. Measuring Semantic Similarity Between Geospatial Conceptual Regions[A]. Rodríguez M, Cruz I, Levashkin S, *et al.* *GeoSpatial Semantics*[M]. Berlin: Springer, 2005.
- [17] Ahlqvist O. Using Semantic Similarity Metrics to Uncover Category and Land Cover Change[A]. Rodríguez M, Cruz I, Levashkin S, *et al.* *GeoSpatial Semantics*[M]. Berlin: Springer, 2005.
- [18] Ahlqvist O, Ban H. *Categorical Measurement Semantics: A New Second Space for Geography*[Z]. Oxford: Blackwell Publishing Ltd, 2007.
- [19] Jansen L J M, Groom G, Carrai G. Land-cover harmonisation and semantic similarity: Some methodological issues[J]. *Journal of Land Use Science*, 2008, 3(2-3): 131-160.
- [20] Ahlqvist O, Shortridge A. Spatial and semantic dimensions of landscape heterogeneity[J]. *Landscape Ecology*, 2010, 25(4): 573-590.
- [21] 贾小斌. 地图综合的智能体技术-以土地利用图斑综合为例[D]. 武汉: 武汉大学, 2013. [Jia X B. *Agent Technology Applied in Map Generalization: Case Study for Land-use Patch* [D]. Wuhan: Wuhan University, 2013.]
- [22] 陈先伟. 土地利用数据库综合的结构化模型和算法研究[D]. 武汉: 武汉大学, 2005. [Chen X W. *Research on Structured Model and Algorithms in Land Use Database Generalization* [D]. Wuhan: Wuhan University, 2005.]
- [23] 翁杰, 郭庆胜, 王晓妍, 等. 一种改进的图斑合并算法[J]. *武汉大学学报(信息科学版)*, 2012, 37(9): 1116-1119. [Weng J, Guo Q S, Wang X Y, *et al.* An improved algorithm for combination of land-use data[J]. *Geomatics and Information Science of Wuhan University*, 2012, 37(9): 1116-1119.]
- [24] 贾小斌, 艾廷华. 土地利用数据综合中图斑多边形决策推理[J]. *农业工程学报*, 2013, 29(9): 214-222. [Jia X B, Ai T H. Decision-making reasoning in land-use patch generalization[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2013, 29(9): 214-222.]
- [25] 王晓妍, 郭庆胜, 翁杰, 等. 零散多边形综合质量评价研究[J]. *武汉大学学报(信息科学版)*, 2012, 37(9): 1112-1115. [Wang X Y, Guo Q S, Weng J, *et al.* Quality assessment of scattered polygon generalization[J]. *Geomatics and Information Science of Wuhan University*, 2012, 37(9): 1112-1115.]
- [26] 郭庆胜, 王晓妍, 刘纪平. 图斑群合并的渐进式方法研究[J]. *武汉大学学报(信息科学版)*, 2012, 37(2): 220-223. [Guo Q S, Wang X Y, Liu J P. Progressive combination of polygon groups [J]. *Geomatics and Information Science of Wuhan University*, 2012, 37(2): 220-223.]

Computation of semantic proximity in land-use data based on feature-matching

JIA Xiaobin¹, AI Tinghua², PENG Zifeng³

(1. The Smart City Research Institute of China Electronics Technology Group Corporation, Shenzhen 518026, China;

2. School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China;

3. Shenzhen Municipal Planning and Land Real Estate Information Center, Shenzhen 518034, China)

Abstract: This research aimed to compute the semantic proximity in land-use data. First, this research established a model to express the layer of details of the semantic characteristics in land-use data, started from the basic situation of land-use category and combined with characteristics of ownership. Second, feature-matching was used to gain concrete metric values by computing the matching relationship among ownership, reason of coverage, land cover type, usage, using state, nature of the stem in vegetation, habits of growth in vegetation, usage of vegetation and land use patterns. This research achieves the whole process from semantic modeling to the semantic proximity in land-use data. Third, this experiment takes 38 entities in land-use data as an example to compute semantic proximity among those with the same characteristic in ownership but different characteristic in land use type, and compares the result of the experiment with the judgment of the practical experience to discover that the result of calculations conforms to human cognition and the methods used in calculating the semantic similarity have strong practicability. Finally, the computation results of semantic proximity were transformed into a spatial measurable value to divide the land-use patch into this have the adjacent relationship in spatial position based on the semantic proximity which the area of is smaller than the threshold when the scale is change smaller in the processing of land-use data in order to demonstrate the practical value of the research findings. The results of this research can be directly applied to statistics, processing, planning, management and development of land resources, and in computing semantic proximity in land-use data to calculate semantic adjacent relationships in other related spatial geographic data.

Key words: land-use data; semantic feature; proximity; feature matching