

引文格式: LUO Guowei, ZHANG Xingchang, QI Lixin, et al. The Fast Positioning and Optimal Combination Method of Change Vector Object[J]. Acta Geodaetica et Cartographica Sinica, 2014, 43(12): 1285-1292. (罗国玮, 张新长, 齐立新, 等. 矢量数据变化对象的快速定位与最优组合匹配方法[J]. 测绘学报, 2014, 43(12): 1285-1292.) DOI: 10.13485/j.cnki.11-2089.2014.0191

矢量数据变化对象的快速定位与最优组合匹配方法

罗国玮^{1,2}, 张新长^{1,3}, 齐立新⁴, 郭泰圣¹

1. 中山大学 地理科学与规划学院, 广东 广州 510275; 2. 广西师范学院, 广西 南宁 530001; 3. 广东省城市化与地理环境空间模拟重点实验室, 广东 广州 510275; 4. 广东省国土资源测绘院, 广东 广州 510500

The Fast Positioning and Optimal Combination Matching Method of Change Vector Object

LUO Guowei^{1,2}, ZHANG Xingchang^{1,3}, QI Lixin⁴, GUO Taisheng¹

1. School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China; 2. Guangxi Teachers Education University, Nanning 530001, China; 3. Guangdong Key Laboratory for Urbanization and Geo-simulation, Guangzhou 510275, China; 4. Institute of Surveying and Mapping of Land and Resources Department of Guangdong Province, Guangzhou 510500, China

Abstract: The change-only information is important to the recording of object life cycle, the establishment of spatial-temporal database and the updating of GIS database. To solve the problem of low efficiency of traditional method in change detection when the data volume is large, we proposed a highly efficient method of change detection based on the grid-partitioning of data and the comparison of synthesis of spatial and attribute information. This method only detects the changed grid to reduce the detection region. In order to solve the matching problem of old features and new features, we propose a method named optimal combination-matching method. The method selects the optimally matched features through the comparison of the characteristic of spatial information and semantic information. The method's high efficiency and accuracy in change detection of large volume of spatial data and matching of changed features is validated by experiment.

Key words: change-only information; grid-partitioning; semantic tree; optimal combination-matching

摘要: 要素变化信息对地物生命周期的记录、时空数据库的构建、GIS 数据库的更新有重要意义。针对大数据量的变化信息发现, 本文采用基于格网划分的方法, 通过对空间特征与属性特征汇总信息的对比, 只对发生变化的格网进行检测, 缩小了检测范围与空间查询区域。为解决要素变化前后的匹配问题, 提出一种最优组合匹配法, 通过对组合对象空间特征及语义特征的综合比较, 从候选要素中选取最佳匹配对象。试验证明, 该方法能够高效准确地实现大数据量的矢量数据变化信息的探测, 并能很好地解决非一对一的要素匹配问题。

关键词: 变化信息; 格网划分; 语义树; 最优组合匹配

中图分类号: P208

文献标识码: A

文章编号: 1001-1595(2014)12-1285-08

基金项目: 国家自然科学基金重点项目(41431178); 国家 863 计划(2013AA122302); 高等学校博士点专项基金(20120171110030)

1 引言

GIS 矢量数据的要素变化信息对于记录要素生命周期、时空数据库的建立、GIS 客户数据库更

新^[1]信息提取等方面有重要的意义。GIS 矢量数据变化信息检测与匹配是为了发现新旧数据中的变化信息, 并建立变化要素在新旧数据中的匹配关系。国内外学者从空间信息、属性信息对新旧

要素的变化特征进行了研究,目前,对矢量要素变化检测方法的研究多集中在数据更新信息识别的方面,文献[2—3]运用基于目标匹配的方法提取变化信息。文献[4]采用了基于事件的变化检测,变化信息必须由事件驱动,对数据的完整性要求高且不具备通用性。在空间数据匹配研究中,研究主要以几何特征(形状^[5-6]、距离^[7]、方向^[8]等)、拓扑特征^[9-10]、属性特征作为匹配依据,在一对一的要素匹配中取得了良好的匹配效果。文献[11—12]运用了概率理论的匹配模型,通过融合多种匹配指标,计算候选目标概率大小来确定匹配目标,但匹配计算量大且没有完全解决多对多的匹配问题。为解决多对多的要素匹配问题,文献[13]提出了增量凸包法,文献[14]采用了组合匹配的方法。另外还有一些基于机器学习、神经网络^[15]等方面的匹配方法,由于其需要学习样本的参与,使用起来比较复杂。以上这些匹配方法主要是针对制图因素而非变化信息的匹配,文献[2—3]中所进行的变化对象匹配没有给出多对多要素匹配的具体方案。在变化信息的匹配中,由于对象本身发生了几何、拓扑、属性信息的改变,采用常规的特征匹配指标很难确定同名实体。

针对大数据量的变化检测,为提高变化信息检测的效率,本文采用基于格网划分的变化信息快速定位方法。根据要素变化前后的空间关系及语义关系,对常规的匹配方法进行了优化,提出了一种最优组合匹配方法,目的是提高空间数据变化前后对象的匹配精度,匹配结果能更好地为地物变化过程的跟踪记录及客户 GIS 数据库增量信息获取提供数据服务。

2 变化类型与匹配关系

在 GIS 矢量数据中,要素的变化类型及匹配关系如表 1 所示。在现实中,地理要素会出现多种变化,如对比旧数据,新数据中的同一要素发生了位移与几何形变两种变化,要想检测出这些复合的要素变化非常困难,但是可以通过对变化前后要素的特征分析对比,确定它们的匹配关系。

3 基于格网划分的变化信息快速定位

传统的矢量数据的要素变化检测方法是通过对要素进行逐一比较来发现变化信息,需要进行大量的空间查询与属性对比。本文提出一种基于格网划分的变化信息快速定位方法,可以对新旧

数据中的空间信息和属性信息变化进行快速而准确的定位。具体过程如下:

(1) 通过检测确保新旧数据是同一比例尺及采用同一坐标系,分别对新旧数据增加特征点坐标属性字段(Center_X、Center_Y)及存储要素属性汇总信息的属性字段(TotalStr)。

表 1 要素变化类型及匹配关系

Tab. 1 Feature change of type and matching relation

要素类型	变化类型	匹配关系
点	新增	1:0
	删除	0:1
	位移、属性变化	1:1
线	新增	1:0
	删除	0:1
	位移、扩张、缩小、属性变化	1:1
	合并	1:n
	分裂	m:1
	聚合	m:n
面	新增	1:0
	删除	0:1
	位移、扩张、缩小、属性变化	1:1
	合并	1:n
	分裂	m:1
	聚合	m:n

(2) 分别对新旧数据进行全局查询,计算要素特征点坐标、要素属性汇总信息、并确定新旧数据对比的范围($X_{\min}, X_{\max}, Y_{\min}, Y_{\max}$)。特征点坐标代表要素所处的位置,点要素直接取其坐标,线要素取其中点,面要素取其质心。要素属性汇总信息是将要素属性字段的字段值按字段名的字符串匹配排序进行拼接,每个字段值之间用特殊符号分隔,如式(1)所示

$$\text{Totalstr. value} = \text{Fields}(0). \text{value} + " | " + \text{Fields}(1). \text{value} + " | " + \dots + " | " \text{Fields}(n). \text{value} + " | " \quad (1)$$

(3) 将变化检测范围按统一的宽度和高度划分为 $m \times n$ 个规则的格网^[16], m 与 n 按式(2)计算,格网的宽度根据检测范围和要素总数由系统自动确定。格网的宽度记为 G_{width} , 高度记为 G_{height} 。根据格网的总数定义数组变量来存储格网中要素的汇总信息

$$\left. \begin{aligned} m &= (X_{\max} - X_{\min}) / G_{\text{width}} \\ n &= (Y_{\max} - Y_{\min}) / G_{\text{height}} \end{aligned} \right\} \quad (2)$$

(4) 分别对新旧数据按特征点坐标(Center_X、Center_Y)进行排序查询,并按特征点坐标将要素匹配到相应的格网。格网编号与特征点坐标

匹配方法按式(3)计算,其中 INT()为向下取整。按式(4)对格网内要素的特征点坐标(Center_X、Center_Y)、几何值信息、属性汇总信息按排序结果依次累加到相应的网格变量。线要素和面要素的几何值信息为要素的弧段长度,点要素不需要累加几何值信息^[17]

$$\left. \begin{aligned} p &= \text{INT}((\text{Center_X} - X_{\min}) / G_{\text{width}}) \\ q &= \text{INT}((\text{Center_Y} - Y_{\min}) / G_{\text{height}}) \end{aligned} \right\} \quad (3)$$

$$\left. \begin{aligned} \text{gcenx}(p, q) &= \sum_{i=1}^k \text{Center_X}_i \\ \text{gcenty}(p, q) &= \sum_{i=1}^k \text{Center_Y}_i \\ \text{glen}(p, q) &= \sum_{i=1}^k \text{length}_i \\ \text{gstr}(p, q) &= \sum_{i=1}^k \text{TotalStr}_i \\ \text{gcount}(p, q) &= k \end{aligned} \right\} \quad (4)$$

gcenx()、gcenty()分别表示网格中要素特征点 X 坐标及 Y 坐标的和;glen()表示网格中要素的弧段长度和;gstr()表示网格中要素属性值字符串拼接;gcount()为行号网格中要素个数。

(5) 对新旧数据中编号相同网格中的特征点 X 坐标、特征点 Y 坐标、弧段长度、属性值拼接字符串进行对比

$$\left. \begin{aligned} c_x(p, q) &= |N(\text{gcenx}(p, q)) - O(\text{gcenx}(p, q))| / O(\text{gcenx}(p, q)) \\ c_y(p, q) &= |N(\text{gcenty}(p, q)) - O(\text{gcenty}(p, q))| / O(\text{gcenty}(p, q)) \\ c_l(p, q) &= |N(\text{glen}(p, q)) - O(\text{glen}(p, q))| / O(\text{glen}(p, q)) \end{aligned} \right\} \quad (5)$$

按式(5)计算网格中几何特征的变化率, $c_x()$ 为网格中要素特征点 X 坐标和的变化率; $c_y()$ 为网格中要素特征点 y 坐标和的变化率; $c_l()$ 为网格中弧段长度的变化率; $N()$ 与 $O()$ 分别代表新旧数据。当新旧数据中的对应网格满足以下条件之一, 该网格中的要素存在变化情况, 需要对网格内的要素逐一作变化检测(如图 1 所示):

(1) 要素特征点 X 坐标和的变化率 $c_x()$ 或要素特征点 Y 坐标和的变化率 $c_y()$ 大于给定阈值, 阈值由测量点坐标误差许可范围确定。

(2) 要素弧段长度和变化率 $c_l()$ 大于给定阈值, 阈值由边长测量误差许可范围确定。

(3) 要素数量不同。

(4) 要素属性值拼接字符串 gstr 不同。

变化要素的发现方法是在目标数据中搜索与源数据的面积、弧段长度、方向、重心等空间特征与语义特征相同(差异小于阈值)的对象, 当搜索结果为空时说明要素发生了变化。在作新旧要素逐一变化对比时, 需要通过大量的空间查询, 当数据范围较大时, 花费时间较多。由于本文已对查询空间进行了格网划分, 且要素属性中记录了该要素的重心坐标, 进行要素空间查询时通过属性过滤($\text{Center_X} \geq \text{GridX}_{\text{左}} \text{ AND } \text{Center_X} \leq \text{GridX}_{\text{右}} \text{ AND } \text{Center_Y} \geq \text{GridY}_{\text{下}} \text{ AND } \text{Center_Y} \leq \text{GridY}_{\text{上}}$), 只对要素所在格网内的要素进行, 大大缩小了查询的范围, 提高了查询的效率。

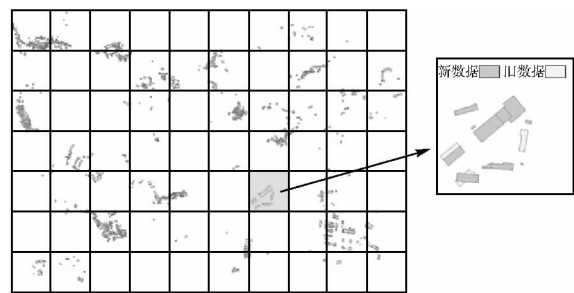


图 1 格网划分

Fig. 1 Grid partition

4 变化要素最优匹配

具有匹配关系说明新要素对旧要素有继承性, 继承性可以通过相似度来量化。变化要素的匹配关系需要综合考虑空间相似度及语义相似度。

4.1 空间相似度检测

空间相似度检测的目的是确定候选匹配要素, 并与语义相似度结合来确定要素的匹配关系。下面分别介绍点、线、面的空间相似度检测方法。

4.1.1 点要素

点要素的候选匹配要素通过缓冲区来搜索。以源数据中的点要素进行半径为 r 的缓冲区搜索, r 的值根据不同的地物取最大可能位移距离。目标数据中落入缓冲区的点作为候选匹配要素, 候选要素与源要素的空间相似度按式(6)计算, $d(A, B)$ 为源数据中点要素 A 与目标要素中点要素 B 的空间相似度, $\text{dis}(A, B)$ 为 A 与 B 的欧氏距离

$$d(A, B) = 1 - \frac{\text{dis}(A, B)}{r} \quad (6)$$

4.1.2 线要素与面要素

目前, 面要素及线要素的空间特征匹配方法

有很多,面要素可通过对象的面积、位置、形状、方向来匹配,线要素可通过长度、位置、方向、弯曲度等特征来匹配。由于是对变化了的对象进行匹配,对象的单个空间特征都可能发生了改变,且变化前后的对象数目存在非一对一的关系,采用单个空间特征匹配难以保证准确性^[3]。重叠面积能够综合反映对象变化前后的位置与几何相似性且计算方法简单而快速,并有利于进行对象间多对多的匹配;拓扑差异能够反映对象变化前后的拓扑关系相似性,因此本文采用了重叠面积比与拓扑差异来计算空间相似度:

(1) 通过面积重叠度确定候选要素。线要素与面要素的匹配,设源数据中的要素(线要素的缓冲区)为 A ,搜索目标数据中与 A 相交的要素 B_i ,按式(7),当 A 与 B_i (线要素的缓冲区)的相交面积与 A 与 B_i 中面积比较小的要素面积的比值 $S(A, B)$ 大于给定阈值 ϕ , B_i 作为与 A 匹配的候选要素。为顾及非一对一的要素匹配,需要对候选要素进行反向搜索,搜索源数据中与目标数据中候选要素 B_i 相交的要素 A_i 。当 B_i 与 A_i (线要素的缓冲区)的相交面积与 A_i 与 B_i 中面积比较小的要素面积的比值 $S(A, B)$ 大于阈值, A_i 作为与 B_i 匹配的候选要素。只要搜索到新的候选匹配要素,就以此候选要素作为源要素,反向进行搜索,经反复迭代,直到没有搜索到新的候选匹配要素为止。阈值 ϕ 的设定是为避免将过多的将匹配可能性低的要素纳入候选匹配对象,所取值根据地物类型来确定。

$$S(A, B) = \frac{\text{area}(A \cap B)}{\min(\text{area}(A), \text{area}(B))} \quad (7)$$

经过反复双向搜索,会得到集合 $A(a_1, a_2, \dots, a_m)$ 与集合 $B(b_1, b_2, \dots, b_n)$ 的候选匹配要素。

(2) 拓扑相似度计算。在多要素的匹配中,变化前后要素组合之间的拓扑关系是空间相似性的重要因子,在其他因素相等的情况下,变化前后组合拓扑信息差异小匹配的可能性大,例如相接的两个对象比相离的两个对象合并为一个对象的可能性大。本文定义候选要素之间的拓扑距离,相接或相交为 1,相离为 2,单要素距离为 0。源候选要素之间与目标候选要素之间的拓扑差异为两组要素拓扑距离的和。拓扑相似度采用式(8)计算,其中 $td(A, B)$ 为源要素与目标要素的拓扑差异; λ 为可调节参数,本文经多次试验, λ 取 1 时相拓扑似度计算结果较理想

$$st(A, B) = \frac{\lambda}{\lambda + td(A, B)} \quad (8)$$

4.2 语义相似度检测

源要素与目标要素间的语义相似度反映了要素变化前后的语义继承关系。语义相似度通过要素属性值的相似度来确定,按文献[18]提出的语义相似度评价模型计算。因对象的属性值有可能发生变化,属性项中的特征码及类别等变化可能性小的字段应加大权重。

空间矢量数据的属性字段类型大多为文本或数字,这两类字段能较好地反映对象的语义特征。对一般字符型的属性字段,本文采用字符串的编辑距离^[19]来计算相似度。字符串编辑距离的相似度计算方法有时无法反映概念间的语义距离^[20],字符串编辑距离很远但是表达的语义信息很相似的情况会经常出现,如“沟渠”与“河流”。针对要素类别的字段,本文采用了基于语义树的计算方法,通过事先利用专家知识构建好的领域本体^[21]语义树,来计算两个概念间的语义距离。如图 2 所示,要计算概念 A 与概念 B 之间的语义相似度参考按文献[22]提出的语义相似度计算模型,按式(9)、式(10)进行计算。

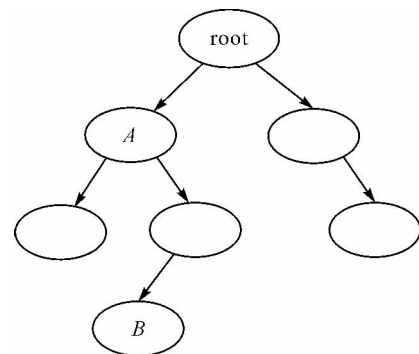


图 2 层次语义树

Fig. 2 Hierarchical semantic tree

$$\text{sim}(A, B) = -\log \frac{\text{comdis}(A, B)}{2^{\text{maxdepth}}} \quad (9)$$

$$\text{comdis}(A, B) = \text{dis}(A, \text{com}(A, B)) + \text{dis}(B, \text{com}(A, B)) \quad (10)$$

式(9)中 maxdepth 为语义树的最大深度;式(10)中 $\text{com}(A, B)$ 为节点 A 与节点 B 在语义树中的最近共同祖先;式(10)中 $\text{dis}(A, \text{com}(A, B))$ 与 $\text{dis}(B, \text{com}(A, B))$ 分别计算 A 、 B 到最近共同祖先的有向边距离。

对于数值型字段,本文认为数值在一定变化

阈值内存在相似性,变化范围超出阈值时相似度为 0。数值型字段相似度计算如式(11)所示。 v 为变化前的数值, v' 为变化后的数值, ϵ 为变化阈值,且 $\epsilon < v$

$$z(v, v') = \begin{cases} 1 - |(v' - v)|/v & |v - v'| < \epsilon, \epsilon < v \\ 0 & |v - v'| \geq \epsilon, \epsilon < v \end{cases} \quad (11)$$

4.3 匹配策略

本文采用空间相似度与语义相似度结合起来计算对象的相似度,计算模型如式(12)所示

$$\text{sim}(A, B) = r(A, B)\omega_1 + \text{st}(A, B)\omega_2 + \text{sem}(A, B)\omega_3 \quad (12)$$

点要素直接按公式进行源要素与目标要素一对一相似度计算,线要素与面要素变化匹配涉及非一对一的匹配, A, B 分别代表新旧数据候选匹配对象的集合。 $r(A, B)$ 对点要素为距离相似度,对线与面要素为面积重叠度; $\text{st}(A, B)$ 代表候选匹配要素的拓扑相似度; $\text{sem}(A, B)$ 为语义相似度; $\omega_1, \omega_2, \omega_3$ 为可调节权重值, $\omega_1 + \omega_2 + \omega_3 = 1$,点要素取 $\omega_2 = 0$ 。针对变化信息可能存在一对多、多对一或多对多匹配,本文提出了最优组合匹配法。从候选匹配要素中选取相似度最大的候选匹配组合,如选取的候选匹配组合的相似度大于阈值,表明这对组合为最终的匹配结果。具体算法:设 $A(a_1, a_2, \dots, a_m), B(b_1, b_2, \dots, b_n)$ 为候选匹配要素的集合,根据排列组合公式集合 A 与集合 B 分别有 $2^m - 1, 2^n - 1$ 项候选组合。集合中组合的生成方法是依次将十进制数 1 至 $2^m - 1$ 转换成二进制数,判断该二进制数的值,从右往左判断不为 0 的项为集合中选取的要素的下标,如 110 表示要生成的组合为 (a_3, a_2) 。源要素组合与目标要素组合的面积重叠度按式(13)计算

$$r(A, B) = \frac{\text{area}(P_A \cap P_B)}{\text{area}(P_A \cup P_B)} \quad (13)$$

式中, P_A, P_B 对于面要素分别代表源要素组合与目标要素组合,对于线要素代表源要素缓冲区组合与目标要素缓冲区组合。 P_A, P_B 为组合中各元素的并集。

源要素组合与目标要素组合的拓扑相似度按式(8)计算,如组合中的要素个数大于 2,该组合的拓扑距离取组合内两两要素之间拓扑距离的最大值。

源要素组合与目标要素组合的语义相似度计算方法如下:如两个组合的元素数量均为 1,按本文 4.2 节的方法求取集合的语义相似度。如组合

中元素不为 1,按以下方法计算:对可累加数值型字段分别对两个组合中该字段的值的求和,对非累加数值型字段取两个组合中该字段的平均值,再用式(11)计算相似度;对字符型字段,根据字段性质采用字符编辑距离法或层次语义树法求出两两相应字段值的相似度,再计算字段相似度的平均值;求出两组合各属性字段的相似性之后,再按语义相似度评价模型^[18]计算源要素组合与目标要素组合的语义相似度。

当候选要素较多时,生成的组合较多,计算量也相应增大。为减少计算时间,本文对不存在相交或包含关系的组合对不再进行计算,直接将相似度赋为 0。在匹配过程中,需要进行大量的空间查询操作,对点要素及面要素的匹配,将空间搜索范围限制在与源要素所在格网及相邻的格网内,以提高匹配的速度。

完成源候选要素与目标候选要素生成的所有组合的相似度计算之后,进行匹配要素的最终确定。方法如下:①将相似度大于阈值的组合对存入队列;②从队列中选取相似度最大的组合对作为相匹配要素记录存入匹配列表,同时将这对组合从队列中删除,并将队列中含有最大相似度组合对中的要素的其他组合对从队列中删除;③重复步骤 2,直到队列中的所有组合对被删除。匹配列表为匹配的最终结果,根据列表中每一条匹配记录的要素数目,可识别匹配关系 $1:0, 0:1, 1:1, 1:n, m:1, m:n$ 。

5 试验案例

5.1 变化信息快速定位试验

本文采用 Visual Studio 2010 结合 ArcGIS Engine 10 开发了 GIS 矢量数据变化检测与匹配原型系统,分别选取不同时期的大比例尺地形图中的多组点、线、面要素运用全局逐个检测法和格网法进行了试验,要素的变化形式包括新增、消失、位移、几何形变、合并、分裂、聚合、属性变化等,如图 3 所示。表 2 列出了格网划分法与全局逐个检测的效率对比信息。试验证明,虽然在网格信息初始化、要素特征点计算等方面花费一些时间,但格网划分法只对变化了的区域进行检测,可以大大减少检测的时间,检测要素越多加速效果越明显,且该方法的效率不会受到变化要素分布区域的影响,只与要素的变化率有关,要素变化率越低,加速比越高。本文还采用了不同的格网

宽度做了试验,结果显示格网宽度在一定范围内 变化对加速效果影响不大。

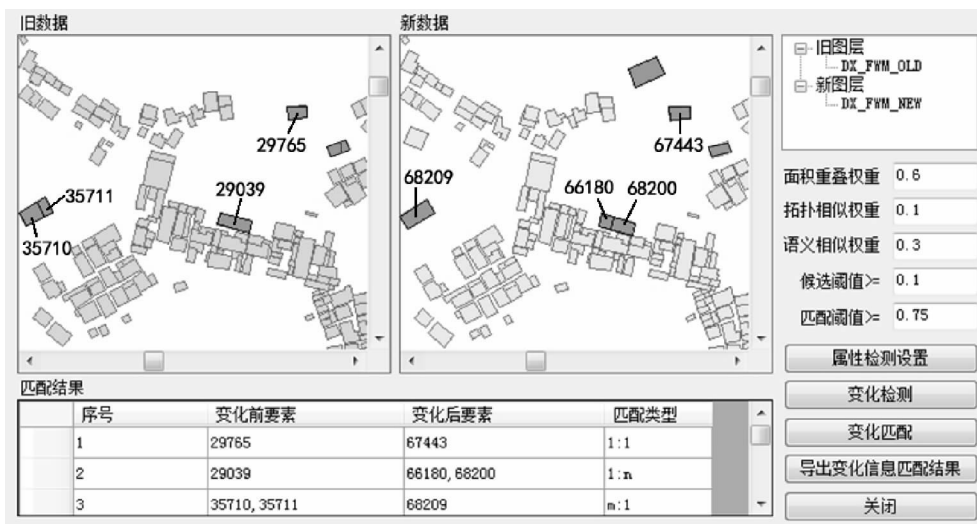


图 3 GIS 数据变化检测与匹配原型系统

Fig. 3 GIS data change detection and matching prototype system

表 2 格网划分法与全局逐个检测的检测效率对比

Tab. 2 Comparison of efficiency grid partition method and traversal method

类型	检测要素	变化要素	变化率/(%)	全局逐个 检测时间/s	格网法时间/s	加速比
点	3376	566	16.77	85	12	7.08
点	6482	1094	16.88	281	25	11.24
点	8815	1301	14.76	484	36	13.44
线	2136	409	19.15	67	10	6.70
线	4533	449	9.91	187	15	12.47
线	7653	809	10.57	425	27	15.74
面	3342	443	13.26	141	19	7.42
面	5264	733	13.92	335	30	11.17
面	9922	1354	13.65	1339	57	23.49

在实际运用中,可能出现要素特征点正好落在格网边界上,解决的方法是统一将格网左右边界上的特征点统一归左边网格,上下格网边界上的特征点统一归上边网格。

5.2 最优组合匹配试验

本文利用变化对象快速定位试验中检测出的要素进行匹配试验。运用式(13)计算匹配相似度时,权重 ω_1 、 ω_2 、 ω_3 根据数据源的不同,采用对已

匹配好的样本进行特征分析结合专家经验来设定,试验中点要素分别取(0.7,0,0.3),线要素与面要素分别取(0.6,0.1,0.3)。试验结果显示(表 3),对比常规的组合方法^[13-14],最优组合匹配法的匹配精度有所提高。应用实例如表 4 所示,最优组合匹配法在一对多、多对一及多对多的匹配中具有明显的优势。但在要素发生了较大位移时,系统没有将其纳入候选要素,会导致误匹配。







表 3 匹配试验结果对比

Tab. 3 Comparison of the matching experiment

类型	变化数	常规组合匹配法			最优组合匹配法		
		正确匹配数	错误匹配数	准确率/(%)	正确匹配数	错误匹配数	准确率/(%)
点	566	524	42	92.58	539	27	95.25
线	409	367	42	89.73	393	16	96.09
面	433	393	40	90.76	418	15	96.54

表 4 应用实例

Tab. 4 Examples of application

变化前要素	变化后要素	匹配结果对比
		常规组合匹配法:A、B 与 C 匹配 最优组合匹配法:考虑了语义和拓扑相似度,A 与 C 综合相似度最高,A 与 C 匹配
		常规组合匹配法:A 与 B、C、D 匹配 最优组合匹配法:考虑了语义和拓扑相似度,A 与 B、C 匹配
		常规匹配法:A、B 与 C、D、E 匹配 最优组合匹配法:考虑了语义和拓扑相似度,A、B 与 C、D 匹配

6 结 论

GIS 矢量数据的变化信息的发现与匹配能为地物生命周期的记录、时空数据获取及客户空间数据更新信息的发布提供服务。本文对空间矢量数据的变化对象的快速定位与匹配进行了研究,经过试验,得出了以下结论:①基于格网划分的变化信息快速定位法大大缩小了检测范围,极大地提高了变化信息的检测速度,且不受变化对象分布范围的影响,能够准确高效地检测出矢量数据空间信息及属性信息的变化;②最优组合匹配法通过对变化对象空间信息与语义信息的综合对比,从候选要素中找出最优组合,提高了匹配的准确度,能更好地解决多对多的变化要素匹配问题。本文方法的不足之处在于匹配的准确度对相似度因子权重及匹配阈值的设置有较大的依赖性。这是下一步研究需要解决的问题。

参考文献:

[1] JIAN C L, HUANG M L. Research on Geographic Information Database Incremental Updating Method[C]//Proceedings of International Conference on Audio Language and Image Processing. Shanghai: [s. n.], 2010: 985-989.

[2] WU Jianhua, FU Zhongliang. Methodology of Feature Change Detection and Matching in Data Updating[J]. Computer Applications, 2008, 28(6): 1612-1615. (吴建华, 付仲良. 数据更新中要素变化检测与匹配方法[J]. 计算机应用, 2008, 28(6): 1612-1615.)

[3] YING Shen, LI Lin, LIU Wanzeng, et al. Change-only Updating Based on Object Matching in Version Databases [J]. Geomatics and Information Science of Wuhan

University, 2009, 34(6): 752-755. (应申, 李霖, 刘万增, 等. 版本数据库中基于目标匹配的变化信息提取与数据更新[J]. 武汉大学学报: 信息科学版, 2009, 34(6): 752-755.)

[4] ZHANG Feng, LIU Nan, LIU Renyi, et al. Research of Cadastral Data Modeling and Database Updating Based on Spatio-temporal Process[J]. Acta Geodaetica et Cartographica Sinica, 2010, 39(3): 303-309. (张丰, 刘南, 刘仁义, 等. 面向对象的地籍时空过程表达与数据更新模型研究[J]. 测绘学报, 2010, 39(3): 303-309.)

[5] AI T H, CHENG X, LIU P, et al. A Shape Analysis and Template Matching of Building Features by the Fourier Transform Method [J]. Computers, Environment and Urban Systems, 2013, 41: 219-233.

[6] KIELER B, HUANG W, HAUNERT J H, et al. Matching River Datasets of Different Scales [M]. Advances in GIScience. Berlin: Springer, 2009: 135-154.

[7] DENG M, LI Z L, CHEN X Y. Extended Hausdorff Distance for Spatial Objects in GIS [J]. International Journal of Geographical Information Science, 2007, 21(4): 459-475.

[8] DENG M, LI Z L. A Statistical Model for Directional Relations between Spatial Objects[J]. Geoinformatica, 2008, 12: 193-217.

[9] PAN LI, WANG Hua. Automatic Recognition of Change Types of Residential Areas Using Topology Relations Model[J]. Geomatics and Information Science of Wuhan University, 2009, 34(3): 301-303. (潘励, 王华. 利用拓扑关系模型自动检测居民地的变化类型[J]. 武汉大学学报: 信息科学版, 2009, 34(3): 301-303.)

[10] DENG M, CHENG T, CHEN X Y, et al. Multi-level Topological Relations between Spatial Regions Based upon Topological Invariants[J]. Geoinformatica, 2007, 11: 239-267.

[11] TONG Xiaohua, DENG Susu, SHI Wenzhong. A Probabilistic Theory-based Matching Method [J]. Acta Geodaetica et Cartographica Sinica, 2007, 36(2): 210-217. (董小华, 邓慷慷, 史文中. 基于概率的地图实体匹配方法[J]. 测绘学报, 2007, 36(2): 210-217.)

[12] WALTER V, FRITSCH D. Matching Spatial Data Sets: A Statistical Approach[J]. International Journal of Geographical Information Systems, 1999, 13(5): 445-473.

[13] ZHANG Liping, GUO Qingsheng, SUN Yan. The Method of Matching Residential Features in Topographic Maps at Neighboring Scales [J]. Geomatics and Information Science of Wuhan University, 2008, 33(6): 604-607. (章莉萍, 郭庆胜, 孙艳. 相邻比例尺地形图之间居民地要素匹配方法研究[J]. 武汉大学学报: 信息科学版, 2008, 33(6): 604-607.)

[14] WU Jianhua. Entity Matching Methods Based on Combining Multi-similarity-characteristics Considering Environment Similarity[J]. Geography and Geo-Information Science, 2010, 26(4): 1-6. (吴建华. 顾及环境相似的多特征组合实体匹配方法[J]. 地理与地理信息科学, 2010, 26(4): 1-6.)

- [15] XU Junkui, WU Fang, WEI Huifeng. Areal Settlements Matching Algorithm Based on Artificial Neural Network Technique[J]. *Journal of Geomatics Science and Technology*, 2013, 30(3): 293-298. (许俊奎, 武芳, 魏慧峰. 神经网络在居民地面状匹配中的应用[J]. *测绘科学技术学报*, 2013, 30(3): 293-298.)
- [16] CHEN Yumin, GONG Jianya, SHI Wenzhong. A Distance-based Matching Algorithm for Multi-scale Road Networks [J]. *Acta Geodaetica et Cartographica Sinica*, 2007, 36(1): 84-90. (陈玉敏, 龚健雅, 史文中. 多尺度道路网的距离匹配算法研究[J]. *测绘学报*, 2007, 36(1): 84-90.)
- [17] GUO Taisheng, ZHANG Xinchang, LIANG Zhiyu. Research on Change Information Recognition Method of Vector Data Based on Neural Network Decision Tree [J]. *Acta Geodaetica et Cartographica Sinica*, 2013, 42(6): 937-944. (郭泰圣, 张新长, 梁志宇. 神经网络决策树的矢量数据变化信息快速识别方法[J]. *测绘学报*, 2013, 42(6): 937-944.)
- [18] COBB M A, CHUNG M J, FOLEY I H, et al. A Rule-based Approach for the Conation of Attributed Vector Data[J]. *GeoInformatica*, 1998, 2(1): 7-35.
- [19] DIAO Xingchun, TAN Mingchao, CAO Jianjun. New Method of Character String Similarity Computing Based on Fusing Multiple Edit Distances [J]. *Application Research of Computers*, 2010, 27(12): 4523-4525. (刁兴春, 谭明超, 曹建军. 一种融合多种编辑距离的字符串相似度计算方法[J]. *计算机应用研究*, 2010, 27(12): 4523-4525.)
- [20] HASTINGS J T. Automated Conflation of Digital Gazetteer Data[J]. *International Journal of Geographical Information Science*, 2008, 22(10): 1109-1127.
- [21] LOPEZ-PELLICER F J, LACASTA J, FLORCZYK A, et al. An Ontology for the Representation of Spatiotemporal Jurisdictional Domains in Information Retrieval Systems [J]. *International Journal of Geographical Information Science*, 2012, 26(4): 579-597.
- [22] LEACOCK C C. Milling in a Sparse Training Space for Word Sense Identification [EB/OL]. [2009-02-25]. http://arxiv.org/PS_cache/cmp-lg/pdf/9511/9511007v1.pdf.
(责任编辑: 丛树平)

收稿日期: 2014-01-17

修回日期: 2014-06-22

第一作者简介: 罗国玮(1979—), 男, 博士生, 高级工程师, 研究方向为空间数据更新与融合。

First author: LUO Guowei(1979—), male, PhD candidate, senior engineer, majors in spatial data updating and fusion.

E-mail: bestlgw@163.com

通信作者: 张新长

Corresponding author: ZHANG Xinchang

E-mail: eeszxc@mail.sysu.edu.cn

(上接第 1284 页)

- [13] CABELLO S, DE BERG M, VAN DIJK S, et al. Schematization of Road Networks[C]// *Proceedings of the Seventeenth Annual Symposium on Computational Geometry*. Massachusetts: [s. n.], 2001: 33-41.
- [14] WARE J M, ANAND S, TAYLOR G E, et al. Automated Production of Schematic Maps for Mobile Applications [J]. *Transactions in GIS*, 2006, 10(1): 25-42.
- [15] DONG Weihua, GUO Qingsheng, LIU Jiping, et al. Progressive Generalization Research of Schematic Road Network Maps [J]. *Geomatics and Information Science of Wuhan University*, 2007, 32(9): 829-832. (董卫华, 郭庆胜, 刘纪平, 等. 道路网示意性地图的渐进式综合研究[J]. *武汉大学学报: 信息科学版*, 2007, 32(9): 829-832.)
- [16] LI Z L, DONG W H. A Stroke-based Method for Automated Generation of Schematic Network Maps [J]. *International Journal of Geographical Information Science*, 2010, 24(11): 1631-1647.
- [17] HE Jin, FEI Lifan, HUANG Lina, et al. Study on the Method of Indirect Generalization for Contour Lines Based on the 3D Douglas-Peucker Algorithm [J]. *Acta Geodaetica et Cartographica Sinica*, 2013, 42(3): 467-473. (何津, 费立凡, 黄丽娜, 等. 三维 Douglas-Peucker 算法的等高线间接综合方法研究[J]. *测绘学报*, 2013, 42(3): 467-473.)
- [18] KLIPPEL A. Conceptualizing Wayfinding and Route Direction Elements[D]. Jacobs: Universitat Bremen, 2003.
- [19] THOMSON R C, RICHARDSON D E. The Good Continuity' Principle of Perceptual Organization Applied to the Generalization of Road Networks[C]// *Proceeding of the 19th International Cartographic Conference*. Ottawa: [s. n.], 1999: 1215-1225.
- [20] LI Z L. Algorithm Foundation of Multi-scale Spatial Representation[M]. London: CRC Press, 2006.
- [21] TOBLER W. Numeric Map Generalization [J]. *Cartographica*, 1966, 26(1): 9-25.
(责任编辑: 宋启凡)

收稿日期: 2014-05-09

修回日期: 2014-08-13

第一作者简介: 遆鹏(1977—), 男, 博士, 讲师, 研究方向为数字地图制图与空间数据可视化。

First author: TI Peng (1977—), male, PhD, lecturer, majors in digital cartography and geospatial visualization.

E-mail: peng_tj@hotmail.com

通信作者: 贾洪果

Corresponding author: JIA Hongguo

E-mail: lemon_gg@hotmail.com