

引文格式: 丁建勋,刘亚楠,张新长,等. OpenStreetMap 数据下的空间数据更新方法[J].测绘通报,2016(6):94-97.DOI: 10.13474/j.cnki.11-2246.2016.0199.

OpenStreetMap 数据下的空间数据更新方法

丁建勋¹,刘亚楠¹,张新长²,杨文杰³

(1. 珠海市测绘院,广东 珠海 519015; 2. 中山大学地理科学与规划学院,广东 广州 510275;

3. 广东省国土资源技术中心,广东 广州 510075)

Research on Updating Spatial Data Combined with Open Street Map

DING Jianxun, LIU Yanan, ZHANG Xinchang, YANG Wenjie

摘要: 针对利用专业测绘数据对空间数据库更新中存在的现势性问题,提出了利用 OpenStreetMaps 数据对空间数据库进行更新的方法。阐述了基于 OpenStreetMap 数据的空间数据更新方法的关键理论与技术,特别是 OpenStreetMap 数据与专业测绘数据的变化要素检测算法:基于格网的匹配检测算法和基于几何量算的变化要素检测。在此基础上,利用珠海市的专业测绘数据与相应区域的 OpenStreetMap 数据作为试验数据,编程实现了基于 OpenStreetMap 的空间数据更新。

关键词: OpenStreetMap; 空间数据; 变化要素检测; 数据更新

中图分类号: P208

文献标识码: B

文章编号: 0494-0911(2016)06-0094-04

随着网络技术与 GIS 技术的快速发展,用户对地理数据的现势性要求越来越高^[1]。如何保持空间数据的现势性已经成为目前 GIS 界面临的重要课题,新兴的自发地理信息(volunteered geographic information, VGI) 为这个重要的课题提供了一个全新的解决思路。

VGI 是地理信息数据获取的一种全新方式,该数据由用户进行编辑提供,用户既是数据的使用者,又是数据的提供者,即充分发挥公众的参与性,填补了当前地理数据采集的不足,同时其现实性高、细节丰富、获取免费。目前比较成功的 VGI 平台有 OpenStreetMap(OSM)、Google Map Marker、维基地图^[2],本文主要结合 OSM 数据进行更新方法的研究。

一、关键技术分析

1. OSM 数据预处理

为了使 OSM 数据与专业测绘数据的匹配最优化,在进行更新操作之前需要对 OSM 数据进行预处理,通过分析 OSM 数据可以发现其具有以下特点:

- 1) OSM 数据中道路数据比较丰富,其他类型的数据信息量并不丰富。
- 2) 道路数据分为不同的等级,包括一级道路、二级道路、居民地道路、高速公路、人行道等。
- 3) OSM 数据中道路线不仅用单线表示,部分道路线还会用双线表示。
- 4) OSM 数据中通常一个要素代表一条道路,

与专业测绘数据中差别比较大。

鉴于以上的数据特点,OSM 数据的预处理主要包括以下 4 个部分:①根据需要提取 OSM 中的有效数据;②提取双线道路数据的中心线;③减少 OSM 数据和专业测绘数据之间的几何误差;④对 OSM 数据节点分割,使 OSM 数据和专业测绘数据的匹配尽可能达到 1:1^[3]。

2. 变化要素检测

OSM 数据与专业测绘数据的变化要素检测是整个更新流程的核心部分,变化要素检测算法要求能精确、快速地查找出 OSM 数据相对于数据库中专业测绘数据的变化部分。

本节基于自发地理信息数据更新的特点提出了适用于自发地理信息数据更新的变化要素检测方法,分别介绍了格网的划分方法、利用线要素经过的格网面积计算重叠度,以及根据重叠度的数值判断要素的变化情况^[4]。

(1) 格网的划分

首先需要建立完全覆盖 OSM 数据与专业测绘数据的格网,建立格网的步骤如下:

- 1) 对比要进行变化要素检测的 OSM 数据与专业测绘数据的边界范围,选取范围的标准如下

$$\left. \begin{aligned} X_{\max} &= \max(X_{c-\max}, X_{p-\max}) \\ X_{\min} &= \min(X_{c-\min}, X_{p-\min}) \\ Y_{\max} &= \max(Y_{c-\max}, Y_{p-\max}) \\ Y_{\min} &= \min(Y_{c-\min}, Y_{p-\min}) \end{aligned} \right\} \quad (1)$$

收稿日期: 2016-03-07

基金项目: 国家自然科学基金重点项目(41431178)

作者简介: 丁建勋(1973—),男,高级工程师,主要研究方向为城市测量新技术应用与管理。E-mail: djxr@126.com

式中 c 表示专业测绘数据; p 表示 OSM 数据。

2) 选定格网的范围后,对数据范围区域进行划分。考虑到区域连续性、一致性和完整性,根据 OSM 数据和专业测绘数据的大小将整个区域等分成若干相邻的等正方形^[5],等分成的小正方形的数量可用下式表示

$$\left. \begin{aligned} \text{num}_X &= \frac{(X_{\max} - X_{\min})}{\text{width}} - 1 \\ \text{num}_Y &= \frac{(Y_{\max} - Y_{\min})}{\text{height}} - 1 \end{aligned} \right\} \quad (2)$$

式中 X_{\max} 、 X_{\min} 分别表示数据范围的 X 坐标的最大值和最小值; Y_{\max} 、 Y_{\min} 表示数据范围 Y 坐标的最大值和最小值; num_X 表示 X 轴方向格网的列数; num_Y 表示 Y 轴方向格网的行数; width 表示格网的宽度; height 表示格网的高度。

(2) 线目标格网索引几何确定

本文采用的线要素经过格网的集合与线要素的缓冲区具有图形上的相似性,可以将格网面要素的集合当作线要素的变异缓冲区,而线要素经过的格网的集合是与线要素相交的格网的集合,不需要重新建立线要素的缓冲区^[6]。

设线要素的可用节点的几何表示为 $\{ \text{node}_1, \text{node}_2, \dots, \text{node}_n \}$, 与线要素相交的格网索引可表示为 $\{ \text{Grid}_1, \text{Grid}_2, \dots, \text{Grid}_m \}$, 如图 1 所示。

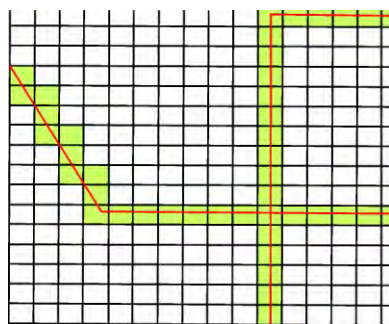


图 1 线要素相交格网集合

利用相交格网的方法比利用缓冲区的方法更适用于自发地理信息与专业测绘数据的变化检测匹配,该方法能够过滤变化较小的要素,因为当格网的边长设置一个合理的值时,即使要素发生较小的变化,相交的格网仍然相同或仅有很小部分的差别。可以发现,由于道路数据发生了微小的改变,但 OSM 数据的相交格网仅发生了一个格网的变化,深色框中的格网消失,新增了浅色框中的格网,如图 2 所示。

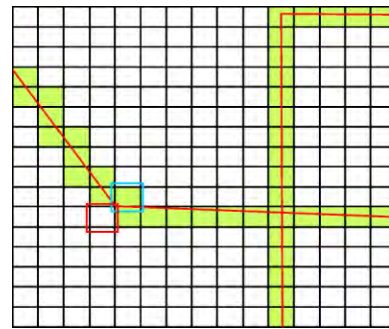


图 2 较小变化的线要素相交格网集合

(3) 重叠度计算

变化区域的专业测绘数据要素集合为 $A\{a_1, a_2, \dots, a_n\}$, 变化区域的 OSM 数据要素集合为 $B\{b_1, b_2, \dots, b_n\}$ 。本文利用双向匹配检测法对变化要素进行匹配检测,首先利用专业测绘数据对 OSM 数据进行搜索,检测出专业测绘数据相对于 OSM 数据中删除的变化类型要素;然后利用 OSM 数据对专业测绘数据进行搜索,检测出 OSM 数据相对于专业测绘数据要素新增、移动、合并、分割等变化类型^[7],见表 1。

表 1 搜索类型表

搜索方式	检测变化类型
OSM 数据匹配专业测绘数据	新增、位移、扩张、缩小、合并、分裂、属性变化
专业测绘数据匹配 OSM 数据	删除

本文将要素的变化类型归为两类:新增和删除。

进行搜索匹配前,需建立一个临时库并新建一个要素类,变化检测中变化的要素将存储在临时库的变化要素类中,此变化要素类保留了专业测绘数据和 OSM 数据中的属性字段,同时新建了两个属性字段,分别用来存放变化要素在专业测绘数据和 OSM 数据中的 ID 及要素变化类型。

利用专业测绘数据线要素的相交格网集合 $\{ \text{Grid}_1, \text{Grid}_2, \dots, \text{Grid}_m \}$ 与 OSM 数据线要素的相交格网集合 $\{ \text{grid}_1, \text{grid}_2, \dots, \text{grid}_n \}$ 作为操作对象。

首先利用专业测绘数据对 OSM 数据进行搜索匹配,即遍历专业测绘数据中的线要素,判断 OSM 数据中是否有要素的相交格网集合与专业测绘数据中的相交格网集合相交。如果专业测绘数据的某个要素的相交格网集合与 OSM 数据中的所有要素的相交格网集合均没有相交,则判定参考数据中的该要素已经在匹配数据中删除,将该要素存入临时库中并将该要素的操作类型赋值为“删除”。

遍历 OSM 数据中的线要素,获得线要素的相交格网集合,判断专业测绘数据中是否有要素的格网集合与其相交,如果专业测绘数据中不存在,则此要素为新增要素,将该要素存入临时库并将其操作类型赋值为“新增”;如果专业测绘数据中存在要素的格网集合与 OSM 数据中的要素的格网集合相交的要素,则需要计算数据之间要素的重叠度,重叠度的计算公式如下

$$S = \frac{\text{Area}[\text{Grid}(A_i) \cap \text{Grid}(B_i)]}{\text{Area}[\text{Grid}(A_i) \cup \text{Grid}(B_i)]} \quad (3)$$

式中, S 表示重叠度; A_i 表示专业测绘数据中的要素; B_i 表示 OSM 数据中的要素,此要素的格网集合与 A_i 的格网集合相交; $\text{Grid}(A_i)$ 表示专业测绘数据中要素 A_i 的相交格网集合; $\text{Grid}(B_i)$ 表示 OSM 数据中的要素 B_i 的相交格网索引; $\text{Area}[\text{Grid}(A_i) \cap \text{Grid}(B_i)]$ 表示相交格网集合 $\text{Grid}(A_i)$ 与 $\text{Grid}(B_i)$ 的交集面积; $\text{Area}[\text{Grid}(A_i) \cup \text{Grid}(B_i)]$ 表示相交格网集合 $\text{Grid}(A_i)$ 与 $\text{Grid}(B_i)$ 的并集面积。

在理想情况下,若专业测绘数据中的要素和 OSM 数据中的要素相比没有发生变化或发生微小变化时,重叠度 S 应等于 1^[8]。如果专业测绘数据中的要素相对于 OSM 数据中的要素中发生了小范围的移动,则 S 的范围在 0~1 之间(不包括 0 和 1),但在实际工作中,由于数据精度等的原因,通常会有上下阈值。

重叠度 S 根据数据的特点可以分为以下情况:

- 1) 重叠度 $S=0$,说明 OSM 数据中的要素为新增的要素。
- 2) 重叠度 $0 < S < Q_1$,其中 Q_1 表示下阈值,此时说明 OSM 数据中的要素与专业测绘数据中的要素并不是匹配要素,只是发生了相交或两条相邻较近的要素。
- 3) 重叠度 $Q_1 < S < Q_2$,其中 Q_2 表示上阈值,说明 OSM 数据中的要素与专业测绘数据中的要素是匹配要素,并且发生了变化。
- 4) 重叠度 $Q_2 < S < 1$,说明两者数据间的要素是匹配要素,但要素间只发生了较小的变化,不需要进行更新。
- 5) 重叠度 $S=1$,说明 OSM 数据中的要素与专业测绘数据中的要素完全匹配,并没有发生任何变化。

当检测到要素发生变化时,将专业测绘数据中的要素存入临时库,将操作类型赋值为“删除”,然后将 OSM 数据中的要素存入临时库,在操作类型中

标注为“新增”。

3. 空间冲突检测

在利用 OSM 数据对空间数据库进行更新时,需要对更新后的道路线数据进行空间冲突的检测与处理,使最终的数据结果符合现实地理实体的表达。

(1) 冲突约束规则表达式

本文在研究空间冲突检测方法的基础上,结合空间完整性约束表达式,增加几何约束规则和属性约束规则,用约束集合的形式表示冲突规则的约束^[9]。

$$\text{rule} = \{ \text{ID}, \text{FeatUreClass}_1, \text{FeatureClass}_2, \text{TopoRule}, \text{GeoRule}, \text{AttriRule} \} \quad (4)$$

式中, ID 表示冲突规则的编号; FeatUreClass_1 和 FeatureClass_2 表示规则约束的要素类; TopoRule 表示拓扑冲突约束规则; GeoRule 表示几何度量冲突约束规则; AttriRule 表示属性冲突约束规则。

(2) 冲突约束规则

根据数据间的特点,制定了道路网常用的冲突规则约束。设更新后的道路线要素类为 A ,冲突检查要素类为 B , A 中的要素集合可表示为 $\{a_1, a_2, \dots, a_n\}$, B 的要素集合可表示为 $\{b_1, b_2, \dots, b_m\}$ 。

规则一:当更新后的道路要素类中两个道路线要素的交点多于阈值时,判定为发生冲突的要素。

规则二:当更新后的道路线要素类中两个道路线的相邻交点间的距离小于阈值时,则判定为发生冲突的要素。

规则三:当冲突检测要素类为居民地面数据时,若存在道路线要素穿过居民地要素,则判定为发生冲突的要素。

规则四:当冲突检测要素类为河流线数据时,若存在道路线要素与河流线要素的夹角小于某个角度,则判定为发生冲突的要素^[10]。

二、试验分析

本文选取珠海市某区域作为研究对象,试验数据分别为珠海小比例尺专业测绘数据和从网上获取的 OSM 数据,如图 3 所示。

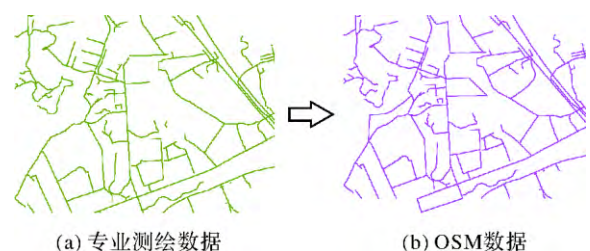


图 3

可以发现,OSM 数据相对于专业测绘数据有要素发生了变化:①道路线消失;②新增了若干条道路线;③部分 OSM 道路线形状和专业测绘数据大致相同,但发生了少量偏移;④部分 OSM 数据相对于专业测绘数据形状大致相同,部分细节有所差别。其中①、②属于道路线新增和删除,③、④属于 OSM 中“伪更新”的要素数据;①、②属于需要更新的变化要素,③、④属于不需要更新的变化要素。

1. 基于格网的变化要素检测结果

试验中首先利用基于缓冲区的变化要素检测对要素进行变化要素的检测匹配,经过处理的 OSM 数据仍然在细节方面与专业测绘数据有很大差别。而利用基于格网的变化要素检测,通过设置合适的格网大小,能去除变化检测中较小变化的影响,去除结果如图 4 所示。

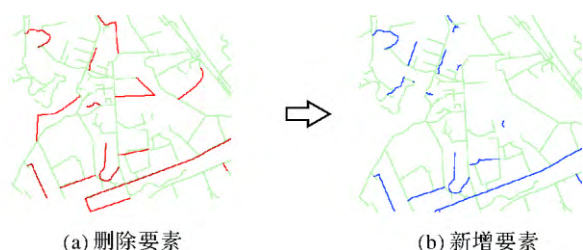


图 4 基于格网的变化要素检测

2. 基于几何信息量的变化要素检测结果

基于格网的初次匹配检测能去除一部分细小变化影响,但不能去除所有的“伪更新”要素,通过几何信息量算对初次匹配检测的结果进行再次匹配检测,检测结果如图 5 所示。



图 5 去除“伪更新”要素后的新增要素

三、结束语

本文在基于自发地理信息数据的空间数据更新方面进行了相关的研究工作,提出了一些实现方法,设计了技术路线,并在自发地理信息数据和专业测绘数据变化检测匹配中提出了相应的方法,作了一些利用自发地理信息数据作为更新数据源更新空间数据库的尝试。然而,基于自发地理信息数据的空间数据更新仍然处于刚刚起步的阶段,现阶段的研究相对较少,因此,本文提出的方法难免出现一些问题,需要深入研究和进一步完善。

参考文献:

- [1] 赵仁亮,陈军.数字环境下的数据缩编更新方法[J].地理信息世界,2008(3):17-21.
- [2] 李德仁,钱新林.浅论自发地理信息的数据管理[J].武汉大学学报(信息科学版),2010,35(4):379-383.
- [3] HAKLAY M. How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets [J]. Environment and Planning B Planning and Design, 2010 (37): 682-703.
- [4] 田文文.基于自发地理信息的空间数据变化发现与更新方法研究[D].武汉:武汉大学,2013.
- [5] 陈玉敏,龚健雅,史文中.多尺度道路网的距离匹配算法[J].测绘学报,2007,36(1):84-90.
- [6] 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.
- [7] 张新长,郭泰圣,唐铁.一种自适应的矢量数据增量更新方法研究[J].测绘学报,2012,41(4):613-619.
- [8] 陈舒燕.基于 OpenStreetMap 的出行可达性分析与实现[D].上海:上海师范大学,2010.
- [9] 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.
- [10] 宋振.线目标空间冲突自动检测研究现状[J].测绘,2009(6):246-248.