

技术方法

基于缓冲区融合的临近节点检查方法研究

焦英华

(山东省国土测绘院, 山东 济南 250013)

摘要:临近节点是影响矢量数据质量的关键问题。本文针对数据维度提升带来的高维信息这一特征,将要素节点升级为节点缓冲区,设计实现了基于要素节点缓冲区融合的临近节点检查方法,通过多源数据合并、要素转节点、相同节点融合、缓冲以及缓冲区融合等步骤获得要素节点缓冲区融合面,最后通过面积容差判定出临近节点所在范围。实验证明,算法通用性强,准确度高,效率稳定,定位准确,是一种有效的多源矢量数据临近节点检查方法。

关键词:多源数据融合;临近节点;拓扑检查;地理要素

中图分类号:P208

文献标识码:A

doi:10.12128/j.issn.1672-6979.2022.02.010

引文格式:焦英华.基于缓冲区融合的临近节点检查方法研究[J].山东国土资源,2022,38(2):67-69. JIAO Ying-hua. Study on Adjacent Nodes Inspection Based on Buffer Fusion Method[J].Shandong Land and Resources,2022,38(2):67-69.

0 引言

矢量数据^[1-4]是地理要素按几何对象类型分层存储的数据集,包含空间信息以及属性信息,是实现地理信息系统功能的数据基础,在农村土地承包经营权确权登记发证、第三次全国国土调查、房地一体确权登记等工作的汇交数据成果中,矢量数据是最基础也是最核心的组成部分。矢量数据具有的空间分析功能使得其广泛应用于建设规划、资源管理、环境分析等各个行业领域,数据质量的好坏直接影响着数据在各个领域的应用^[5]。通过多源数据融合,可以有效提升数据维度,维度提升可以增加对象的属性信息,对于表述某些空间关系更为有利。矢量数据对精度、图形拓扑关系和属性逻辑关系的准确性要求极高,测绘生产内业在采集、编辑、整理的各个过程中不可避免的产生拓扑错误,而临近节点是其中最容易出现错误^[6]。实现高精度的矢量数据临近节点检查,是保证数据质量的关键问题。

本文研究了基于缓冲区融合的矢量数据临近节点检查方法,并通过实证分析对算法精度、效率和稳定性进行了验证。

1 算法实现

本文选用微软公司的 Microsoft Visual Studio 2010^[7](以下简称 VS2010)作为开发平台,C#^[8]作为开发语言,ArcObjects^[9](以下简称 AO)作为主要开发工具包,实现基于缓冲区融合的矢量数据临近节点检查。

首先将所有待检查数据进行合并,然后将要素的节点全部取出,融合完相同的节点后,生成缓冲区并且按照是否相交进行融合,读取融合缓冲区的面积众数作为判定阈值,大于该阈值的融合缓冲区即为临近节点所在位置,算法流程如图1所示。

(1)数据合并。临近节点检查数据来源不同,如房地一体临近节点检查中,需将房屋、宗地、阳台、楼梯等数据合并到一起,综合检查其临近节点,因此首先将不同来源数据合并到一起,如果数据的空间参考不一致,需要先进行坐标转换再将数据合并。

(2)要素转节点。使用 AO 中的 ESRI.ArcGIS.DataManagementTools.FeatureVerticesToPoints 类,将合并的要素转化为节点集合。

(3)相似节点融合。要素转节点后,邻接位置存

收稿日期:2021-09-13;修订日期:2021-09-28;编辑:曹丽丽

基金项目:省级自然资源监测监管大数据应用服务平台建设,项目编号:2019JZZY020103

作者简介:焦英华(1978—),男,山东临邑人,高级工程师,主要从事地理信息系统研发、空间数据挖掘等工作;E-mail:wah4480@163.com

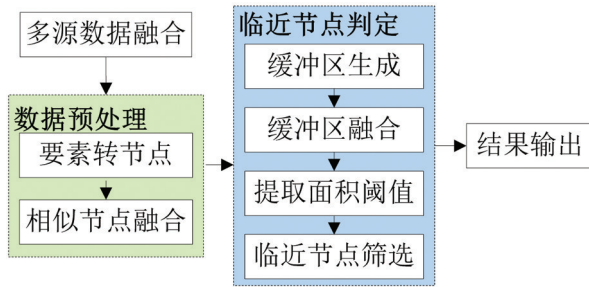


图 1 算法流程图

在重复节点,其坐标值在一定容差范围内,但并不一定完全一致,因此,使用 ESRI. ArcGIS. DataManagementTools.Dissolve 类,将生成的拓扑容差范围内相接触的点融合。避免不同数据源位置精度不一致造成的同一节点坐标偏差,同时降低数据量,提高计算效率,为下一步缓冲区融合做准备。

(4)生成缓冲区。使用 AO 中的 ESRI.ArcGIS. AnalysisTools.Buffer 类,以检查临近节点的距离容差的一半作为缓冲半径,生成缓冲区面。

(5)缓冲区融合。使用 AO 中的 ESRI.ArcGIS. DataManagementTools.Dissolve 类,将生成的相接触的缓冲区融合。

(6)得到面积容差。读取生成的缓冲区融合数据,因临近节点为少数,所以单个节点生成的缓冲区面积应为众数,因此取面积众数作为面积判定阈值,大于该阈值的为临近节点所在缓冲区。其中,面积精度应与拓扑容差一致,如拓扑容差为 0.001,则面积取值应取三位有效数字。

(7)临近节点判定。大于获取的面积容差阈值的缓冲区融合图形为临近节点拓扑错误所在缓冲区范围。

(8)释放资源,输出检查结果。删除融合后要素节点、节点缓冲区等中间数据,保留临近节点判定结果。

2 实例分析

以某地籍子区房地一体数据作为测试数据,共计 9 类数据 923 个要素,数据存在 80 组临近节点。设计两组实验分别验证其精度稳定性和效率稳定性,实验中算法计算 CPU 为 i7-8700K 3.7GHz,临近节点检查阈值为 0.05 m。

2.1 准确性分析

本文将数据整体平移复制 1 到 10 倍,数据量从 9 类 923 个矢量图形到 9 类 9 230 个矢量图形时,随数据量增加,该算法对临近节点识别的准确性未有明显变化,算法运行结果如表 1 所示。

表 1 算法准确性分析表 (临近节点数量/组)

	1 倍	2 倍	3 倍	4 倍	5 倍	6 倍	7 倍	8 倍	9 倍	10 倍
测试 1	80	160	240	320	400	480	560	640	720	800
测试 2	80	160	240	320	400	480	560	640	720	800
测试 3	80	160	240	320	400	480	560	640	720	800
正确率	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

数据量倍增后,算法依旧检测出所有临近节点错误集合,具有良好的精度稳定性。

从实验结果可以看出,同源数据间临近节点、不同源数据间临近节点均被全部检测出。该算法无需进行数据融合、拓扑节点构建等预处理操作,即可对数据中存在的临近节点进行快速、准确检测,具有较好的通用性与稳定性。

2.2 效率性分析

本文将数据整体平移复制 1 到 10 倍,分别计算数据量从 9 类 923 个图斑到 9 类 9 230 个图斑时,该算法随数据量增加其运行效率保持稳定,算法运行时间如表 2 所示。

表 2 算法效率性分析表 (运行时间/s)

	1 倍	2 倍	3 倍	4 倍	5 倍	6 倍	7 倍	8 倍	9 倍	10 倍
测试 1	9.3	16.1	24.0	33.7	43.9	55.0	68.6	82.7	95.4	113.1
测试 2	9.6	16.9	24.3	33.8	43.7	57.1	67.4	79.4	94.0	111.7
测试 3	9.4	16.2	24.6	33.8	44.0	56.1	68.2	80.4	96.9	112.0
平均	9.4	16.4	24.3	33.8	43.9	56.1	68.1	80.8	95.4	112.3

数据量为 9 组 923 个时,算法运行时间约为 9.4 s,随数据量倍增,算法运行时间线性增长,在数据量为 9 组 9 230 个时,算法运行时间约 112.3 s,算法效率如图 2 所示。传统基于拓扑查询的临近节点检查,随数据量增加,不可避免地会出现效率严重降低现象,而该算法随数据量倍增未出现运行时间指数增长现象,具有良好的效率稳定性。

近年来,土地确权、三调、自然资源统一确权登记、房地一体等国家任务陆续开展,时间紧任务重,顺利完成数据汇交对实现成果信息化管理共享至关重要^[10-12]。研究通用、高效、准确的矢量数据质检方案是保证数据质量的重要方向。

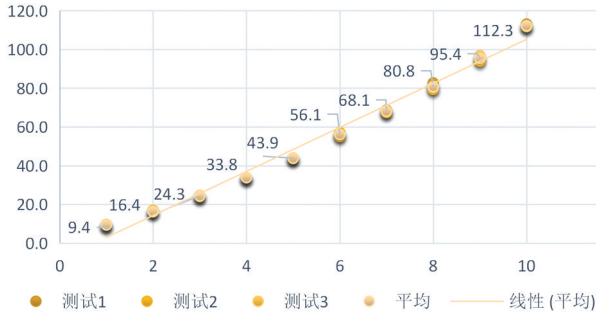


图 2 算法效率折线图

3 结论

本文设计实现了基于要素节点缓冲区融合的矢量数据临近节点检查算法。

(1) 基于维度提升带来高维信息这一特征,将临近节点关系描述转换为缓冲区面关系描述,检查算法描述简单,同时检查结果具有较高的准确性。

(2) 检查效率线性增长,随数据量倍增,本算法检查效率成平缓线性增长趋势,具有较好的效率稳定性。

(3) 通用性强,既能够检查要素内部又能有检查要素间临近节点,且之前无需做拓扑关系构建等处理,自动合并多源数据检查数据临近节点。

(4) 检查结果直观定位准确,检查结果维度同样提升为缓冲区簇形式,非常规的两两点集合,错误结果定位更直观准确。

参考文献:

- [1] 张标,陈楠,高瑛,等.矢量数据质检工具的设计与实现[J].北京测绘,2020,34(10):149-152.
- [2] 王月.农村土地承包经营权确权登记数据库质量控制与评价[D].西安:长安大学,2017:44-46.
- [3] 李琪.第三次国土调查数据分析与共享服务平台设计与研究[J].中国信息化,2020,315(7):103-105.
- [4] 叶爱东.农村房地一体权籍数据库质量控制与检查方法研究[J].国土资源信息化,2019(6):49-53.
- [5] 龚健雅.空间数据质量控制与评价方法及实现技术[J].武汉大学学报(信息科学版),2004,29(8):686-690.
- [6] 朱风云,周小红,汤小平.土地利用更新调查内业数据处理方法//第三届长三角科技论坛(测绘分论坛)暨2006江苏省测绘学术年会论文集[C].南京:江苏省测绘学会,2006:52-54.
- [7] 王玮.应用 Microsoft Visual Studio 2010 开发项目的优势[J].现代阅读(教育版),2011(23):221-222.
- [8] 李凯,王劲松.Visual C# 与 Matlab 混合编程原理及其在数据图形化中的应用[J].数字技术与应用,2014(9):55-56.
- [9] 尚晓光,侯克鹏.基于 ArcObjects 组件与 Microsoft.NET 框架的营口某铁矿开采沉陷预计系统[J].金属矿山,2018(6):129-134.
- [10] 关梅,杨志军,刘政,等.以第三次国土调查成果为森林资源“一张图”细化调查基础底图数据融合技术路线的探讨[J].山东国土资源,2020,36(8):68-71.
- [11] 安国强,仲阳,吴建廷,等.国土资源大数据上云高效运行体系及关键技术研究:以山东省为例[J].山东国土资源,2020,36(7):74-80.
- [12] 陈金林,叶爱东.ArcGIS 在农村集体土地所有权数据库质量检查中的应用[J].遥感信息,2014(5):114-117.

Study on Adjacent Nodes Inspection Based on Buffer Fusion Method

JIAO Yinghua

(Shandong Institute of Land Surveying and Mapping, Shandong Jinan 250013, China)

Abstract: The adjacent node is the key problem that affects the quality of vector data. Aiming at the characteristics of high dimensional information brought by data dimension promotion, the feature node is upgraded to node buffer, and the adjacent node inspection method based on feature node buffer fusion is designed and implemented. The feature node buffer fusion surface is obtained through the steps of multi-source data merging, feature to node, same node fusion, buffer and buffer fusion, and finally the range of adjacent nodes is determined through area tolerance. Experiments show that the algorithm has strong universality, high accuracy, stable efficiency and accurate positioning. It is an effective method for detecting adjacent nodes of multi-source vector data.

Key words: Multi-source data fusion; adjacent nodes; topology check; geographical elements