

一种矢量数据的双层次多尺度表达模型与检索技术

程昌秀 陆 锋

(中国科学院地理科学与资源研究所资源与环境信息系统国家重点实验室,北京 100101)

摘 要 空间数据的多尺度表达是当代 GIS 研究的热点问题之一。该文针对矢量数据快速可视化的需求,结合制图综合领域的相关理论,提出了一种矢量数据双层次多尺度表达模型,用来将矢量数据抽象为空间要素和要素内的点坐标两个层次进行表达。其中空间要素层次的表达以空间要素为最小研究单元,通过建立多尺度索引来描述空间要素因尺度改变而引起的数量或性质变化;要素点坐标层次的表达则是以要素内坐标点为最小研究单元,通过尺度层次标记的方式来表达空间要素内的点坐标随尺度变化的渐变过程。该模型在开源数据库管理系统 PostgreSQL 支持下,扩展了相应的索引与函数,实现了矢量数据的双层次多尺度表达模型,同时设计了相应的检索算法,并以某城市 1:10 000 土地利用数据为例,对上述模型与检索算法进行了验证。实验结果表明,在基本不影响可视化效果的前提下,该矢量数据多尺度模型能极大地提高海量矢量数据的可视化与传输的效率。

关键词 多尺度表达 多尺度数据库 多尺度索引

中图法分类号: P208;TP391.3 文献标识码: A 文章编号: 1006-8961(2009)06-1012-06

A Bi-hierarchical Multi-scale Model and Indexing Approach for Large Vector Data Representation

CHENG Chang-xiu, LU Feng

(Institute of Geographical Sciences and Natural Resources Research, Beijing 100101)

Abstract Multi-representation of spatial data is one of the hot topics in modern GIS. This paper proposed a bi-hierarchical multi-scale model to represent large vector data with two hierarchies, namely spatial elements and vertex coordinates, so as to improve the visualization efficiency. Firstly spatial elements were taken as the minimal units and organized with a multi-scale index to describe the changes of quantity or property of elements brought with scale changes. Then vertex coordinates were regarded as the minimal units and tagged with scale hierarchy labels to describe the geometric changes arising from different scales. Under the support of PostgreSQL, this paper implemented the bi-hierarchical multi-scale model by extending the index and functions, and developed the searching algorithms. Finally, a 1:10,000 scale land use dataset was taken as an example to validate the model and searching algorithms. The results showed that the visualization and transmission efficiency of large vector data could achieve great improvement with the proposed model and searching algorithms.

Keywords multi-representations, multi-scale database, multi-scale index

1 引言

现代信息社会的目标之一是在任何时间、任何

地点,为任何人提供灵活的信息访问服务^[1]。地理信息系统(GIS)作为访问空间数据的先进工具,应该为不同客户端的运行环境提供不同详细程度(尺度)的空间数据,特别是当用户访问海量空间数据

基金项目:中国科学院知识创新工程重点方向性项目(KZCX2-YW-308);国家高技术研究发展计划(863)项目(2007AA120401)

收稿日期:2008-12-03;改回日期:2009-03-22

第一作者简介:程昌秀(1973~),女,副研究员。2001年于中国农业大学获土壤学专业博士学位。主要从事多重表达模型、空间数据库关键技术的研究,发表论文30余篇。E-mail:chengex@lreis.ac.cn

时,灵活的信息服务功能显得尤为重要。

近年来,随着小波压缩、影像金字塔模型^[2]等理论的发展与成熟,栅格数据的多尺度表达与可视化问题已基本得以解决;而矢量数据则由于各空间要素间存在着复杂的联动、关联及演化关系,致使矢量数据的多尺度表达与管理问题一直未得到很好的解决,也一直是业界的研究热点之一。

矢量数据多的尺度表达模型主要用于描述空间要素从精细到粗略的渐变过程。与制图综合不同的是,多尺度表达模型是以地理空间数据为研究对象,从可视化角度来研究空间信息的变化情况,其受比例尺、制图区域地理特征以及制图符号等因素的影响较小。此外,多尺度表达模型是用层次化的数据结构来记录空间要素随尺度变化的综合过程,而制图综合则是以物理文件的形式来记录对应某尺度综合后的地图数据。

通过对自动制图综合技术中的选取(Selection)、合并(Merging)、聚合(Aggregation)、融合(Amalgamation)、压缩(Collapsing)、简化(Simplification)、光滑(Smoothing)、移位(Displacement)、夸张(Exaggeration)、重分类(Reclassification)等综合算子的分析,矢量数据随尺度变化的渐变过程可以大致归结为以下两个层次:

(1)空间要素的渐变 其是以空间要素为最小研究单元,用于描述空间要素因尺度改变而引起的数量或性质的变化。如选取、合并、聚合、融合等,其描述的是空间要素由多到少的变化过程,而压缩则包含了空间要素性质的变化,重分类既包含了空间要素数量的变化,又包含了性质的变化。

(2)空间要素内点坐标的渐变 其是以空间要素内的坐标点为最小研究单元,在空间要素的数量、性质不发生变化的情况下,因尺度变化而引起的要素内的点坐标数目或位置的变化,如线简化、光滑、移位等。

这两个层次的渐变相辅相成、交替出现,它们描述了空间要素随尺度变化的全过程。因此,本文拟从上述两个层次出发来设计实现矢量数据的多尺度表达模型。

2 双层次多尺度表达模型设计思想

目前,业界常见的矢量数据多尺度表达模型的实现方案主要有如下2种:

(1)多尺度矢量数据索引技术 该技术是针对大尺度的矢量数据,在不改变已有矢量数据的基础上,通过生成一种基于尺度的索引结构来描述空间要素随尺度变化的过程或结果。查询时,则根据视窗显示尺度和空间范围,快速检索出某尺度下某区域内的矢量数据。描述空间要素渐变过程的代表性多尺度索引技术包括 Reactive-tree^[3-4]、GAP-tree^[5-6]、Multiple R-tree^[7],而描述空间要素内的坐标渐变过程的代表性索引有 z-value^[8-9]。

(2)多尺度矢量存储结构 该结构不仅改变了矢量数据的存储模型,并加入了尺度维信息。查询时,可通过数据片断的重组或运算来得到某个尺度的矢量数据表达。描述空间要素间渐变过程的代表性多尺度存储结构有 Map Cube Model^[10]、Scale-based Model^[11];描述要素内坐标渐变过程的代表性多尺度存储结构有 Strip trees^[12]、BLG-tree^[3]、PR-file^[13]等。

一般而言,多尺度矢量数据索引技术更擅长描述空间要素间的渐变过程,而多尺度矢量存储结构则更擅长描述要素内的渐变过程。因此,近年来业界开始出现联合使用多尺度索引和多尺度存储结构的技术,如 Multi-scale Hilbert R-tree^[14-15]、Reactive-tree/GAP-tree/BLG-tree 的集成^[16]。本文也采用多尺度索引与多尺度存储结构混合的技术方案,即将 Reactive-tree/GAP-tree 合二为一,提出了一种扩展的 GAP-trees 多尺度索引结构,用于表达空间要素的渐变过程。

虽然 BLG-tree 是较成熟的多尺度矢量存储结构,但由于其树型的存储结构与传统的、现行的 Geometry 存储结构(如 WKB、WKT 等)不兼容,因此本文利用 Geometry 中坐标的尺度标记 M 值来记录各坐标点所在的尺度层次,以实现多尺度存储结构与传统的、现行的 Geometry 模型的兼容。

3 多尺度模型生成与检索技术

基于上述设计思想,本文设计实现了一种矢量数据的双层次多尺度表达模型与检索技术。

3.1 多尺度模型的产生式规则

为了实现多尺度索引与数据结构的自动生成,本文参考制图综合领域的综合规则研发了适合多尺度表达模型自动生成的产生式规则,以验证基于扩展 GAP-trees 的多尺度索引技术和基于尺度层次标

记的多尺度矢量数据表达模型的研究成果。规则参数的设置界面如图 1 所示。根据这些规则,即可生

成面向空间要素的多尺度索引结构和面向要素点坐标的多尺度存储结构。



图 1 产生式规则参数的设置界面

Fig. 1 The main form of setting some parameters for product rules

3.2 基于扩展 GAP-trees 的多尺度索引结构

基于扩展 GAP-trees 的多尺度索引结构在逻辑上,表现为如图 2 所示的若干树结构。图中每个节点的数据结构如下:

```

Struct
| Integer ID; /* 空间要素的编码 */
Numeric Scale; /* 空间要素所在的比例尺 */
NodeList * Children; /* 空间要素的子节点列表的指针 */
};

```

图 2 中的每一棵树不仅表示了空间要素随尺度变化的渐变过程,还表明了每个空间要素的生命周期。对于零散的节点(如图 2 中的 18),其生命周期为全过程,即它将在比例尺 0 到比例尺 3 内一直存在;对于没有父节点的节点(如图 2 中的 22, 23 等),其生命周期处于比例尺 0 到该节点所在比例尺的值之间;对于有父节点的节点(如图 2 中的 3, 21 等),其生命周期处于其父节点所在比例尺的值到该节点所在比例尺的值之间,如空间要素 3 的生命周期是比例尺 1 到比例尺 3。

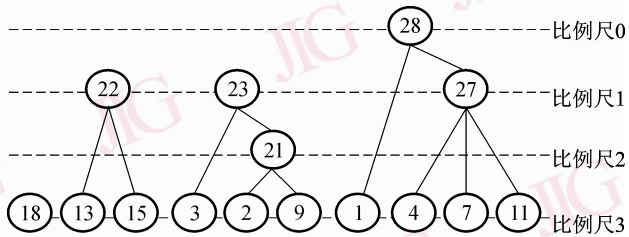


图 2 多尺度索引树示例

Fig. 2 Demo of multi-scale index tree

系统通过此树结构和各节点的生命周期来描述空间要素的渐变过程,例如当显示比例尺在“比例尺 1”附近,且处于由小到大的渐变中,则空间要素 23 将消失,取而代之的是要素 3 和空间要素 21。

3.3 基于尺度层次标记的多尺度存储结构

对于那些生命周期跨多个尺度的空间要素(如图 2 中的要素 18、要素 3 等),系统将基于 Douglas-Peucker 算法对将出现在不同尺度的点进行级别标定,并将标定值记录在 Geometry 中的每个点的尺度层次标记的尺度值 M 中(如图 3 所示)。其中每个

点的尺度值 M 也标定了节点在尺度空间内的生存周期。若当前显示尺度为 1, 则图 3 中所有小于等于 2 的坐标都应该被选出。

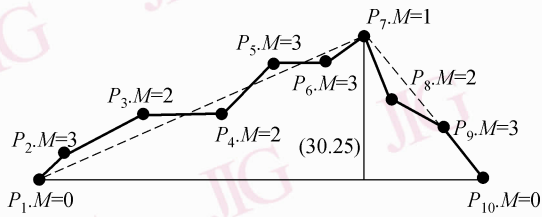


图 3 多尺度存储结构示例

Fig. 3 Demo of multi-scale storage structure

由于本文的重点是讨论矢量数据的多尺度表达模型与检索技术, 而非讨论 Douglas-Peucker 算法自身的缺陷或改进, 因此, 本文只是沿用 Douglas-Peucker 算法。如果需解决自相交问题, 可参考文献 [13] 的研究成果。

3.4 多尺度的检索机制

基于上述多尺度索引机制和多尺度存储结构, 在可视化过程中的数据检索流程如下:

(1) 根据前端传来的任一显示比例尺, 在既定的比例尺序列 (如图 2 所示的比例尺 0 ~ 比例尺 3) 中, 首先找到一个最近的大于等于此显示比例尺的

比例尺值作为要获取的数据的尺度 (如比例尺 1), 以下简称“当前比例尺”;

(2) 在索引树中, 找出如下 4 类节点作为此尺度下应该显示的空间要素: ①既无父节点也无子节点的零散的节点; ②节点的尺度等于当前比例尺的节点; ③节点的尺度小于当前比例尺, 并且其父节点的尺度大于当前比例尺的节点; ④节点的尺度小于当前比例尺的所有根节点;

(3) 对于第 2 步中选出的每个空间对象, 仅选出的值 M 小于等于当前比例尺编号的点来表示当前尺度空间对象的形态。

4 应用实例与分析

本文以某市 1:10 000 土地利用数据为例对上述模型进行了验证。测试数据包含了 186 329 个多边形 (约 240 M)、11 668 862 个节点。后台数据库选用运行在 Dell2600 服务器上的 PostgreSQL + PostGIS。客户端软件 GeoStreamer 是一个用 Java 开发的空问数据浏览软件 (如图 4 所示)。

按系统要求规则定义了相应的参数后, 本文先后选定了 0.000 836, 0.001 638, 0.003 212, 0.006 295,

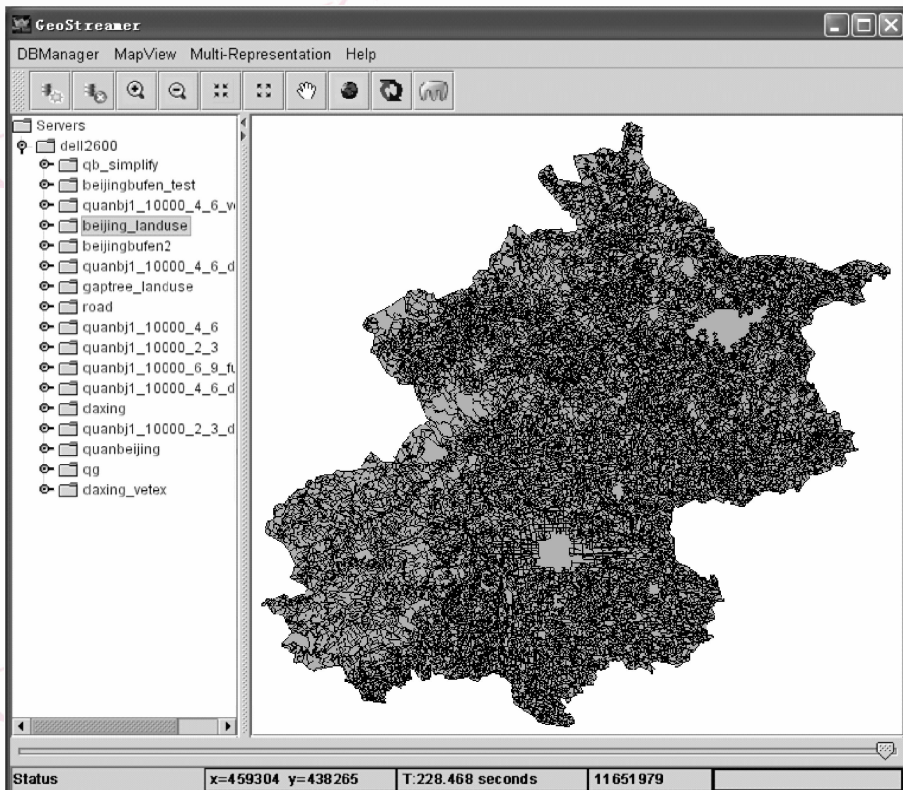


图 4 GeoStreamer 主界面

Fig. 4 The main form of GeoStreamer

0.012 335 和 0.024 182 作为关键尺度的比例尺。通过设定数据综合规则,系统可自动生成上述 6 个关键尺度的多尺度索引与数据存储结构。根据已生成的土地利用多尺度数据索引和存储结构,当用户请求数据时,系统即可根据 3.4 节给出的检索机制

找出相应尺度的数据,展示给用户。

为了对比本文所提出的多尺度双层次模型与检索方法与常规无尺度方法的区别,在同样的运行环境下,本文提出的技术方案和常规无尺度方案在不同显示尺度下的可视化效果和系统耗时的结果如图 5 所示。



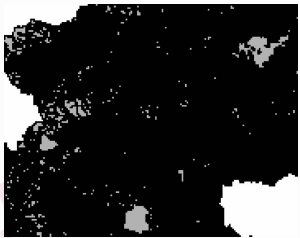
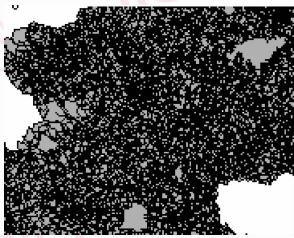
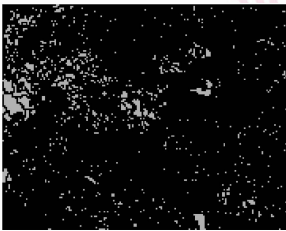
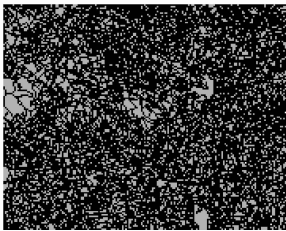
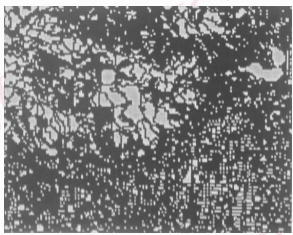
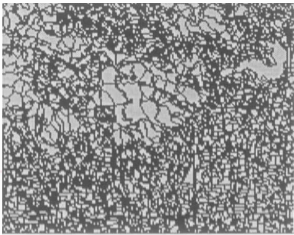
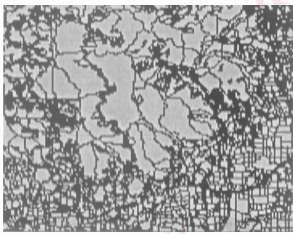
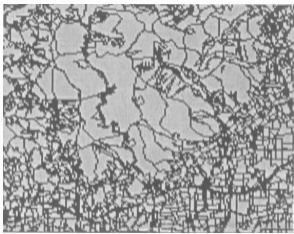
显示比例尺		常规无尺度方法	多尺度双层次模型
0.000 836	截屏图		
	系统耗时 (s)	205.800	4.613
0.001 638	截屏图		
	系统耗时 (s)	117.855	6.932
0.003 212	截屏图		
	系统耗时 (s)	30.701	2.893
0.006 295	截屏图		
	系统耗时 (s)	8.457	1.089
0.012 335	截屏图		
	系统耗时 (s)	2.434	0.440

图 5 多尺度双层次模型与常规无尺度方法的可视化效果及耗时比较

Fig. 5 Difference between bi-hierarchical multi-scale model approach and normal method

由图5可见,在可视化效果基本相同的情况下,采用本文所提出的双层次多尺度矢量数据表达模型可以极大地减少数据传输和显示的系统耗时,而且显示比例尺越小,表现越优异。

5 结 论

本文以多尺度索引和多尺度数据存储结构为基础,提出了一种矢量数据双层次多尺度表达模型,并发展了相应的数据检索技术。矢量数据的双层次多尺度表达模型分别从空间要素、要素几何两个层次来描述空间要素随尺度变化的全过程。其中的多尺度索引结构,通过树结构描述了空间要素因尺度变化而引起的数量或性质的渐变,而多尺度存储结构则通过尺度层次标记描述了要素内的点坐标因尺度变化而引起的数目或位置的渐变。它们分别通过尺度属性和级别属性定义了对象的生存周期。对于生命周期跨多个尺度的空间要素,其在不同尺度的空间形态可以通过其多尺度空间数据结构来进一步描述。当用户请求浏览数据时,系统就根据当前的显示尺度仅选出相应的数据传给用户,在保证数据渐进可视化效果的前提下,本文所提出的矢量数据的多尺度模型能极大地提高海量矢量数据的可视化与传输的效率。

参考文献 (References)

- 1 Spaccapietra S, Vangenot C, Parent C, *et al.* MurMur: A research agenda on multiple representations [A]. In: Proceedings of Database Applications in Non-Traditional Environments [C], Kyoto, Japan, 1999: 373-384.
- 2 Yang Jing-yu, Zhang Yong-sheng, Yu Mei-jiao, *et al.* Image pyramid model dynamically reconstruction based on multi-resolution analysis of wavelet transform [J]. Science of Surveying and Mapping, 2007, **32**(5): 50-51, 62. [杨靖宇,张永生,于美娇等. 基于小波变换多分辨率分析特性的遥感影像动态重构的金字塔模型[J]. 测绘科学, 2007, **32**(5): 50-51, 62.]
- 3 Van Oosterom Peter. The Reactive-tree: a storage structure for a seamless, scaleless geographic database [A]. In: Proceedings of Auto-Carto [C], Baltimore, Maryland, USA, 1991, **10**: 393-407.
- 4 Li Jun, Jing Ning, Sun Mao-yin. A mechanism of implementing visualization with level of detail at multi-scale [J]. Journal of

- Software, 2002, **13**(10): 2037-2043. [李军,景宁,孙茂印. 多比例尺下细节层次可视化的实现机制[J]. 软件学报. 2002, **13**(10): 2037-2043.]
- 5 Van Oosterom P, Schenkelaars V. The development of an interactive multi-scale GIS [J]. International Journal of Geographical Information Systems, 1995, **9**(5): 489-507.
- 6 Tian Peng, Zheng Kou-gen, Zhang Yin, *et al.* Polygon generalization based on C-tree in scaleless GIS [J]. Journal of Image and Graphics, 2001, **6A**(8): 765-770. [田鹏,郑扣根,张引等. 基于C-Tree的无级比例尺GIS多边形综合技术[J]. 中国图象图形学报, 2001, **6A**(8): 765-770.]
- 7 Kwon J, Yoon Y. An access method for integrating multi-scale geometric data [A]. In: Proceedings of the 6th East-European Conference on Advances in Databases and Information Systems [C], Bratislava, Slovakia, 2002: 204-217.
- 8 Zhou X, Prasher S, Masaru K. Database support for spatial generalisation for WWW and mobile applications [A]. In: Proceedings of the IEEE Computer Society, 3rd International Conference on Web Information Systems Engineering [C], Washington, DC, USA, 2002: 239-246.
- 9 Lu Sang-lu, Zhou Xiao-fang, Chen Gui-hai, *et al.* Spatial information retrieval and database generalization [J]. Journal of Software, 2002, **13**(8): 1534-1539. [陆桑璐,周晓方,陈贵海等. 空间信息检索及其数据库概化技术[J]. 软件学报, 2002, **13**(8): 1534-1539.]
- 10 Timpf S. Map cube model-a model for multi-scale data [A]. In: Proceedings of 8th International Symposium on Spatial Data Handling [C], Vancouver, Canada, 1998: 190-201.
- 11 Wei Z, Cho S, Kim J, *et al.* A New Spatial Data Representation in GIS, URL [EB/OL]. <http://www.Geocomputation.org/2000/GC019/Gc019.htm>. 2002.
- 12 Ballard D H. Strip trees: a hierarchical representation for curves [J]. Communications of the ACM, 1981, **24**(5): 310-321.
- 13 Becker B, Six H W, Widmayer P. Spatial priority search: an access technique for scaleless maps [A]. In: Proceedings of ACM Special Interest Group on Management Of Data [C], New York, USA, 1991: 128-137.
- 14 Chan E P F, Chow K K W. On multi-scale display of geometric objects [J]. Data & Knowledge Engineering, 2002, **40**(1): 91-119.
- 15 Ye Chang-chun, Zhou Xing-ming. A method for organizing map dataset to support multi-scale [J]. Chinese Journal of Computers, 2004, **27**(7): 964-970. [叶常春,周兴铭. 一种支持多比例尺表示的地图数据组织方法[J]. 计算机学报, 2004, **27**(7): 964-970.]
- 16 Van Oosterom P. Variable-scale topological data structures suitable for progressive transfer: The GAP-face tree and GAP-edge forest [J]. Cartography and Geographic Information Science, 2005, **32**(4): 331-346.