

文章编号: 1671-8860(2006)07-0649-04

文献标志码: A

不同空间数据格式的数据量实验研究

张立亭^{1, 2, 3}祝国瑞^{1, 2}周世健³陈竹安³

(1 武汉大学资源与环境科学学院, 武汉市珞喻路 129 号, 430079)

(2 武汉大学教育部地理信息系统重点实验室, 武汉市珞喻路 129 号, 430079)

(3 东华理工学院测量系, 抚州市学府路 14 号, 344000)

摘要: 针对一些通用的空间数据的数据交换格式进行了研究, 并就不同数据格式的数据量进行了对比实验。结果表明, 在栅格数据中, .bmp 格式的数据量最大, .jpg 格式和 .gif 格式的数据量较小; 矢量格式的数据量的大小与空间数据的属性有关, 在几种常用的矢量格式中, .e00 格式的数据量最大。

关键词: 数据格式; 数据量; 栅格格式; 矢量格式

中图法分类号: P208

在网络数据共享方面, 尽管不少学者研究了良好的共享系统平台^[1-3], 甚至一些国家也出台了空间数据交换标准^[4], 但很少有人针对适合网络传输特点的空间数据格式开展研究。空间数据共享效果与网上传输速度有关, 而网上传输速度除了与带宽有关外, 还与传输的数据量有着密切的关系。本文针对一些通用的数据交换格式进行了研究。

1 不同空间数据格式的数据实验

在众多的数据格式中, 由于数据结构不同, 其数据量也必然不同。目前, 通用 GIS 软件空间数据矢量交换格式有 ESRI E00 格式(.e00)、ESRI Shape 格式(.shp)、MapInfo 交换格式(.mif/.tab)、AutoCAD 的 .dxg 格式(.dwg)、地球空间数据交换格式(.vct)。通用空间栅格数据格式有.tif、.gif、.jpg、.bmp 等格式^[5]。

空间数据有栅格表示法和矢量表示法两种表示方法。为了更清楚地研究空间数据中不同数据格式与数据量的关系, 本文对深圳市龙港区土地利用图(如图 1)按矢量格式和栅格格式进行了分层, 并且用不同的数据格式分别进行存储, 比较其数据量。图 1 中, 不同的灰度表示不同用途的土

地类型。分层时, 根据不同的灰度按照土地利用分类分别提取耕地、园地、林地、疏林地、交通用地、城镇居民点用地、农村居民点用地、灌木林、水域和裸岩石砾地等, 如图 2 所示。

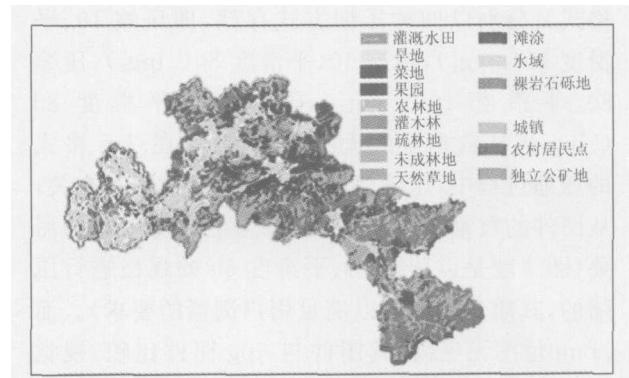


图 1 龙港区土地利用图

Fig. 1 Land Using Map of Longgong District



图 2 土地利用部分分层图

Fig. 2 Parts Delamination Map of Land Using

收稿日期: 2006-04-26。

项目来源: 国家自然科学基金资助项目(40574008); 江西省教育厅科技资助项目(赣教技[2005]208); 江西省数字国土重点实验室开放研究基金资助项目(DLLJ200509)。

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

1.1 棚格格式

棚格数据以.jpg、.gif、.tif、.png 和.bmp 等格式存储, 得出的数据量如表 1 所示。

表 1 不同棚格格式的数据量

Tab. 1 Data Quantity of Different Raster Formats

| 类型 | .jpg/Kb | | .gif/Kb | | .png/Mb | | .tif/Mb | | .bmp/Mb | |
|-------|-------------------|-------------------|-------------------|-------------------|---------|------|---------|------|---------|--|
| | .jpg ₁ | .jpg ₂ | .jpg ₃ | .jpg ₄ | /Mb | /Mb | /Mb | /Mb | /Mb | |
| 土地利用 | 702 | 560 | 212 | 189 | 865 | 3.41 | 4.73 | 17.1 | | |
| 耕地 | 312 | 274 | 140 | 134 | 226 | 1.26 | 1.60 | 17.1 | | |
| 园地 | 418 | 362 | 162 | 155 | 306 | 1.79 | 2.20 | 17.1 | | |
| 林地 | 365 | 304 | 148 | 135 | 325 | 1.46 | 2.01 | 17.1 | | |
| 灌木林 | 296 | 254 | 132 | 125 | 390 | 1.31 | 1.85 | 17.1 | | |
| 疏林地 | 299 | 258 | 135 | 129 | 203 | 1.14 | 1.50 | 17.1 | | |
| 交通 | 300 | 256 | 138 | 129 | 182 | 1.01 | 1.29 | 17.1 | | |
| 城镇居民点 | 426 | 345 | 162 | 146 | 355 | 1.46 | 1.90 | 17.1 | | |
| 农村居民点 | 242 | 214 | 124 | 120 | 137 | 0.84 | 1.13 | 17.1 | | |
| 独立工矿 | 247 | 216 | 124 | 119 | 143 | 0.91 | 1.21 | 17.1 | | |
| 水域 | 291 | 250 | 133 | 124 | 165 | 1.01 | 1.30 | 17.1 | | |
| 裸岩石砾地 | 224 | 199 | 119 | 115 | 121 | 0.79 | 1.06 | 17.1 | | |

表 1 中,.jpg、.bmp、.gif、.png、.tif 等格式的数据量都是在.cdr 的基础上转存的棚格数据量。.jpg 格式和.bmp 格式都是 RGB 色彩(24 位元), 转存图件大小都是以宽 3 000 像素、高 2 000 像素规格存储的, 解析度为 300 dpi。其中,.jpg 格式又分别以四种压缩方式存储, 即压缩 10、平滑度 10(.jpg₁), 压缩 10、平滑度 80(.jpg₂), 压缩 80、平滑度 10(.jpg₃), 压缩 80、平滑度 80(.jpg₄)。从数据量的大小可以看出, 在.jpg 格式的压缩过程中, 压缩值和平滑度都与数据量有关; 从图件的质量来说, 完全能够满足用户用图的需要(图 2 就是以压缩 80、平滑度 80 的规格进行压缩的, 其质量完全可以满足用户浏览的要求)。而.bmp 位图无压缩, 其图件与.jpg 图件比较, 视觉上的差别很难觉察;.gif、.png、.tif 格式存储的图都有压缩,.gif 格式的数据量较小, 是因为.gif 格式的图件在存储过程中有颜色信息损失。

1.2 矢量格式

矢量数据以.cdr、.dwg、.tab、.shp、.e00 等格式存储, 得出的数据量如表 2 所示。

当然, 分层的目的是为了在传输过程中可以根据用户对土地空间数据需求的重点和兴趣不同减少冗余数据的传输, 但如果用户同时需要几种数据时, 可以进行图形叠加。应该注意的是, 在利用棚格数据图进行叠加分析时, 分层前的相对位置必须能很好地复原, 这就要求在将矢量数据转存为棚格数据时, 将图面设置成同样的大小, 且图面的西南角坐标相同, 这样才便于棚格数据的叠加。

表 2 不同矢量格式的数据量

Tab. 2 Data Quantity of Different Vector Formats

| 类型 | .cdr/Kb | .dwg/Mb | .tab/Mb | .shp/Mb | .e00/Mb |
|-------|---------|---------|---------|---------|---------|
| 土地利用 | 865 | 13.1 | 13.4 | 24.1 | 92.9 |
| 耕地 | 47.5 | 0.67 | 0.67 | 1.22 | 3.39 |
| 园地 | 83 | 1.09 | 1.08 | 2.04 | 5.61 |
| 林地 | 102 | 1.36 | 1.33 | 2.55 | 7.9 |
| 灌木林 | 56.9 | 0.81 | 0.83 | 1.49 | 4.22 |
| 疏林地 | 53 | 0.78 | 0.78 | 1.45 | 4.01 |
| 交通 | 119 | 1.19 | 1.19 | 2.20 | 10.2 |
| 城镇居民点 | 426 | 7.49 | 7.72 | 13.5 | 43.5 |
| 农村居民点 | 36.6 | 0.54 | 0.54 | 0.99 | 2.71 |
| 独立工矿 | 35.7 | 0.56 | 0.55 | 1.02 | 2.81 |
| 水域 | 118 | 1.18 | 1.03 | 1.88 | 5.37 |
| 裸岩石砾地 | 24.2 | 0.43 | 0.43 | 0.79 | 2.15 |

表 2 中列出了 5 种较为通用的矢量格式的数据量, 这些数据量的大小也存在较大的区别, 如.cdr 格式的数据量明显较其他格式的数据量小, 因为.cdr 格式不带属性格式, 而其余的三种矢量数据都带有属性格式。从数据结构来说, 其复杂程度远远高于.cdr 格式。因此, 如果用户不需要属性数据时, 也可以用.cdr 格式传输。同时,.e00 格式和.shp 格式的数据量明显大于.tab 格式和.dwg 格式的数据量, 这主要是因其数据结构不同而造成的^[6]。

2 实验结果分析

2.1 棚格格式的数据量分析

棚格数据除.bmp 格式外, 一般都存在压缩, 尽管压缩的方式各不相同^[7,8], 但在数据量的压缩方面, 确实具有很好的效果。本文结合表 1 中的数据分析比较不同格式的棚格数据的数据量。

图 3 为.jpg 格式的四种压缩方式及其数据量的关系图。图中显示,.jpg₁ 的数据量最大,.jpg₄ 的数据量最小,.jpg₂、.jpg₃ 的数据量处于其间。由此可见, 平滑度与数据量无关。

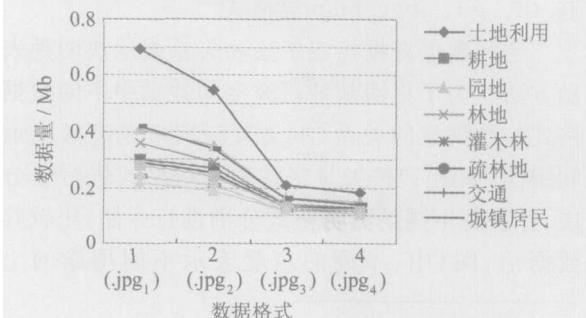


图 3 不同图层的.jpg 格式数据量图

Fig. 3 Data Quantity Chart of .jpg Format of Different Map Layers

如图4所示,. bmp 格式的数据量最大,其次是.tif 格式的数据量,. png 格式的数据量大于.gif 格式和.jpg 格式的数据量。其中,土地利用图层的数据量明显高于其他图层,园地、林地的数据量也高于其他种类。分析认为,土地利用图层的图斑面积最大,其次是园地、林地图层的图斑面积,. jpg 格式和.gif 格式的数据量相当。因此,以上几种栅格格式存储的数据量均与图斑面积大小有关,且.jpg 格式和.gif 格式的数据量较小。如果在没有压缩的情况下(如.bmp 格式),相同大小的图件栅格数据的数据量相同;如果存在压缩,数据量就不一定相同。

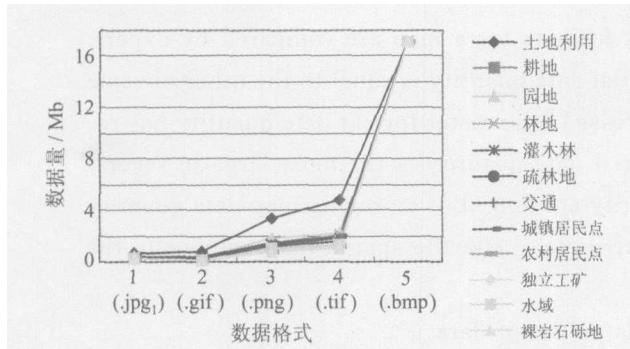


图4 不同栅格数据格式的数据量图

Fig. 4 Data Quantity Chart of Different Raster Formats

2.2 矢量格式的数据量分析

本文选取的矢量数据格式为.cdr,. dwg,. tab,. shp 和.e00 五种格式。这五种数据格式的数据量如图5所示。

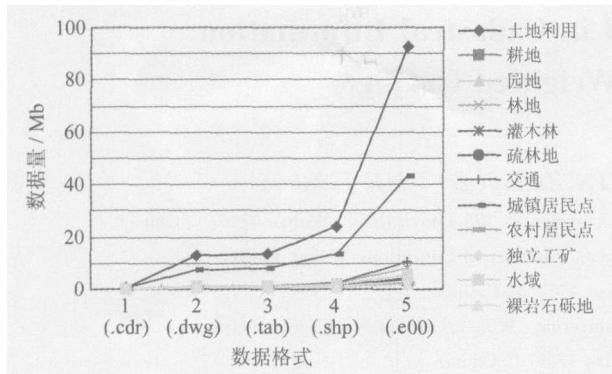


图5 数据矢量格式及其数据量关系图

Fig. 5 Relationship Chart of Different Vector Formats and Data Quantity

从图5可以很明显地看出,. e00 格式的数据量较其他四种格式的数据量大,. shp 格式的数据量略大于其他三种,数据量最小的是.cdr 格式。这是因为.cdr 格式不带属性格式,其矢量数据单独存放,而其余的三种矢量数据都带有属性格式,从数据结构来说,其复杂程度远远高于.cdr 格

式,. e00 格式的数据量明显大于.shp,. tab 和.dwg 格式,这主要是由于其数据结构不同造成的。因此,在用户需要矢量数据而不需要属性数据时,可以采用.cdr 格式传输。

从土地的利用类型与数据量的关系来说,图5中的土地利用图、城镇居民点图层的数据量较其他用地类型的数据量大,原因是这两种图中的属性数据比其他类型的要复杂。至于图中图面的形状与数据量没有太多的关系,如同是以线状形式为主的交通用地层与水域用地层,交通用地除了要记录线路名称外,线路的等级、路面宽度、路面材料、路基宽度等也是必须标明的,故其属性当然比水域用地的属性复杂,数据量也较水域用地的数据量大。可以认为,矢量图的数据量的大小与空间数据的属性有关。

3 结语

通过以上的实验认为,在数据共享时,如果用户只浏览图件而不需要交互操作时,尽量采用.jpg 格式或.gif 格式传输图件;当所存的图件需要为用户提供交互式操作时,可以用.tab 格式或.shp 格式传输图件。当然,也可以用国内通用的 GIS 软件的数据格式存储图件。

参 考 文 献

- [1] Hall W, Davis H, Hutchings G. Rethinking Hypermedia: the Microcosm Approach [M]. Boston: Kluwer Academic Publishers, 1996
- [2] 韩海洋, 黄建明, 龚健雅, 等. Internet 下多数据源、超媒体空间信息的分布式调度与管理 [OL]. <http://www.e2map.com>, 2003
- [3] 王艳东, 龚健雅. 一种 GIS 互操作模型研究 [J]. 武汉大学学报·信息科学版, 2005, 30(3): 218-221
- [4] 中国可持续发展信息共享系统的研究开发项目管理办公室. 中国可持续发展信息共享数据库数据字典内容标准 [OL]. <http://www.sdinfo.net.cn/ngcc/sdinfo/document/stddic.htm>, 2003
- [5] 中国可持续发展信息共享系统的研究开发项目管理办公室. 中国可持续发展信息共享系统的研究开发集中共享数据上交技术规定 [OL]. <http://www.acca21.org.cn/projet/xinxigongxiang/sjsj.htm>, 2003
- [6] ESRI. ESRI Shapefile Technical Description [OL]. <http://www.gischina.com/main/doc/simchin/gisforum/format/shapefile.pdf>, 1998
- [7] 高俊, 夏运均, 游雄, 等. 虚拟现实在地形环境仿真中的应用 [M]. 北京:解放军出版社, 1999: 66-70

- [8] Gregory K W. The JPEG Still Picture Compression Standard[OL]. <http://www.gischina.com/main/doc/simchin/gisforum/format/jpg.pdf>, 1991

第一作者简介: 张立亭, 教授, 博士生。现主要从事 GIS 和 LIS 的理论与应用研究。

E-mail: ltzhang@ecit.edu.cn

Data Quantity Experiments of Different Spatial Data Formats

ZHANG Liting^{1,2,3} ZHUGuorui^{1,2} ZHOU Shijian³ CHEN Zhu'an³

(1) School of Resource and Environment Science, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

(2) Key Laboratory of Geographic Information System, Ministry of Education, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

(3) Survey Department of East China Institute of Technology, 14 Xuefu Road, Fuzhou 344000, China)

Abstract: The spatial data quantity of different formats for a map are compared by experiments. Storaging map by raster format, its spatial data quantity is equal to the other of same size map if not be compressed. If only be compressed, the raster format data quantity has relation to chart speckle area size. .bmp file spatial data quantity is the most large in several raster formats, .jpg file and .gif file data quantity are less. But storing map data quantity by vector format has relation to spatial data attribute. .e00 file spatial data amount is the most large in several vector formats.

Key words: data format; data amount; raster data; vector data

About the first author: ZHANG Liting, professor, Ph.D candidate, engaged in theory and applications of GIS and LIS.

E-mail: ltzhang@ecit.edu.cn

(上接第 642 页)

Spatio-Temporal Data Model in Cadastral Information System Based on Weighted Graph

LEI Qihong^{1,2} LIU Yaolin^{1,2} YIN Zhangcai³ HE Jianhua^{1,2}

(1) School of Resource and Environment Science, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

(2) Key Laboratory of Geographic Information System, Ministry of Education, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

(3) School of Resources and Environmental Engineering, Wuhan University of Technology, 122 Luoshi Road, Wuhan 430070, China)

Abstract: A new spatio-temporal data model based on graph theory is presented, the organizing strategy of spatio-temporal data based on weighted graph is studied, and a formalized method to describe spatio-temporal semantics is provided. The application foreground of transformation between different spatio-temporal data models is analyzed with the model.

Key words: spatio-temporal data model; change process of land parcel; weighted graph

About the first author: LEI Qihong, Ph.D candidate, majors in land resource management, land sustainable use and land information system.

E-mail: leiqihong@163.com

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>