

基于小波分析的线状要素压缩优化的综合性研究

王玉海¹ 朱长青²

(1 信息工程大学理学院,郑州市陇海中路 66 号,450052)
(2 信息工程大学测绘学院,郑州市陇海中路 66 号,450052)

摘要:利用小波分析理论、Douglas 算法、曲率分析算法,对线状要素的压缩优化进行了深入的研究,并在建立线状要素压缩优化的自适应模型的基础上,对实际数据进行了试验。结果表明,此方法不仅能够保持高的压缩比,而且能够很好地保持原数据的特征。

关键词:小波分析;Douglas 算法;曲率分析算法;线状要素

中图法分类号:P208; P283.1

线状要素的压缩优化在地理环境仿真、制图综合、GIS 等研究中具有重要作用^[1]。线状要素的压缩方法基本都是通过减少数字化线划要素上点的数量达到简化图形和数据压缩的目的,同时又设法保持线划的基本特征。其传统的常用算法可归结为三种:垂距限差法、角度限差法、曲率分析法^[2]。其中,垂距限差法和角度限差法对起始点的选择比较敏感,即对同一组数据,只因起始点的不同,所得到的压缩结果也不同。曲率分析法是根据每个数据点的曲率大小或曲率差分大小来决定该点的取舍,而曲率本身是连续函数的概念,对于离散的点集来说,要考察它的曲率,只能得到其近似曲率,而近似曲率只能根据该点和它周围点来计算得到,所以近似曲率受周围点的位置影响非常大,即如果该点的周围点集过于密集或过于稀疏,则得到的曲率误差非常大,对压缩的结果就会产生很大的影响。在测绘领域,笔者曾应用二进制小波和多进制小波理论来解决矢量地图数据压缩这一难题,取得了理想的效果^[3-5],但同时也发现了一些问题。本文在小波分析的理论基础上,结合 Douglas 算法和曲率分析算法,并根据线状要素主要是等高线数据的自身特点,建立了线状要素压缩和优化的自适应模型。

1 线状要素压缩优化的自适应模型

1.1 基于小波分析的线状要素的压缩优化模型

利用正交小波的分解公式对一线状要素 $f(x)$ (其数据为 $L_0 = \{(x_m, y_m), m = 0, 1, 2, \dots, n\}$) 进行小波分解,如图 1 所示,得到低分辨率部分 $f_1(x)$ 和细节部分(高频部分) $g_1(x)$ 。可以看出, $f_1(x)$ 保留了原曲线的总体轮廓特征,可以作为 $f(x)$ 的近似逼近,传统的方法就是将 $f_1(x)$ 作为 $f(x)$ 的压缩数据,其弊端为:所有的点包括特征点都产生了位移,误差较大,地形地貌遭到破坏。而高频部分 $g_1(x)$ 包含了 $f(x)$ 的主要特征信息:在高频部分的数据中,绝对值较大的部分

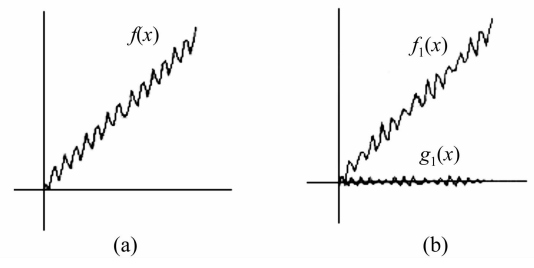


图 1 曲线分解图
Fig. 1 Decomposition of a Curve

对应着原数据中奇异性较大的数据点,即原曲线中曲率较大的部分,也就是线状要素中特征点所在的位置;相对应地, $g_1(x)$ 中接近于零的部分对应着原数据中奇异性非常小的数据点,即原曲线中曲率较小、比较平缓的部分,也就是线状要素中一些冗余数据点所在的位置,可以将这些冗余点去掉,使原曲线得到压缩优化的目的。

基于上述分析,首先建立一个基本的基于小波分析的线状要素的压缩优化模型:① 选取三系数的小波基,对原数据 L_0 进行正交小波分解,得到低分辨率部分数据 $f_1(x)$ 和 高频部分数据 $g_1(x)$;② 选取阈值 ω ,如果 $|g_1(x)| < \omega$,则数据点为冗余点;否则为特征点,并将特征点插入到 $f_1(x)$ (可以根据小波基的支集长度得到特征点在 $f_1(x)$ 中的正确位置);③ 将插入了特征点的低分辨率部分数据 $f_1(x)$ 作为一次小波变换后的压缩数据 L_1 。

1.2 模型中阈值的选取

在对 $g_1(x)$ 的分析中,主要是选取一阈值 ω ,小于 ω 的数据对应着原曲线中的平滑数据;大于 ω 的数据则对应着原曲线中的特征数据。由于低分辨率部分数据 $f_1(x)$ 本身就是原数据 $f(x)$ 的近似表示,且数据量只有原数据的 50%,所以插入了特征点的 $f_1(x)$ 即 L_1 与原数据 $f(x)$ 相比较,其数据量明显减少,特征得到了很好的保留,且更加突出,使原数据得到了压缩优化的目的。

对于复杂程度不同的地形地貌的阈值 ω 的选取,不能单靠人工试验选取,而要根据数据自身的性质去选取,严格地说,要根据每条线或每组线自身的相关性、数据点的离散程度来选取。对于离散的曲线数据,对其进行优化,其中一个环节就是要去掉数据中冗余的部分,从频域分析的角度分析,就是要去掉高频部分绝对值较小的部分对应的数据。根据数学分析理论,这些数据对应着离散曲线数据中曲率较小的部分。但要计算离散数据的曲率,完全要靠此点周围的数据点来计算,所以计算出来的曲率直接与周围点的离散程度有关。而反映点集离散程度的数学概念一般有方差 σ^2 、标准差 σ 、离差 $\sum |P_{i+1} - P_i|$ 、距离(即 $\sum (d_i/m) = \sum (|p_{i+1} - p_i|/m)$, d_i 为相邻数据的距离, m 为一条线中数据点的数目, p_i 为一条线中的数据点)。

阈值的适当选取不仅仅与地形地貌的复杂程度有关,而且与所要求的精度、比例尺的大小有关,精度要求高,阈值的选取就要相对小;反之,就要相对大些(但一般在 $\sigma/10 \sim \sigma/20$ 之间,这是经

验数据,理论数据的确定正在研究中)。下面的试验选取的阈值为 $\sigma/10$ 或 $\sum (p_{i+1} - p_i) / 10m$ 。则在压缩和优化的过程中,就不再需要人工干预。如图 2 所示的曲线数据(一组等高线的一部分)的平均距离为 76,数据点有 1 015 个。图 3 为小波分析后的数据曲线,数据点有 410 个,在分析过程中,采用 $\omega=7$,压缩比为原数据的 40%。

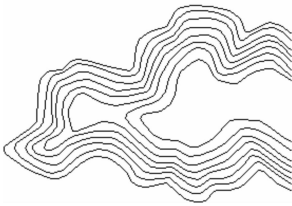


图 2 原图
Fig. 2 Original Image

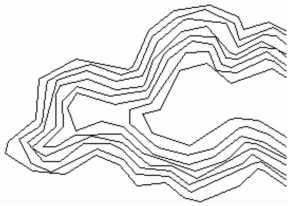


图 3 压缩图
Fig. 3 Compressed Image

1.3 特征点的修正

从试验中可以看出,上述变换后的曲线中,有些地形结构的特征遭到了破坏。如前所述,离散数据的曲率直接与数据点周围点的密度有关。如果太密集,计算出来的曲率根本无法表达其固有的特性,完全是一个错误的特征。这样在频域分析中就会得到错误的信号:本应是曲线中的特征所在,即奇异性较大的部分,但由于其周围点过于密集,所得到的信息却是平缓的信号,在特征点的选取中就会被丢失,使所得的曲线失真。表现最为直接的现象就是原地形中的地性线遭到破坏,地貌形状结构特征被破坏。如果将频域分析、Douglas 算法^[2]以及曲率分析算法相结合,上述问题就会被解决,具体算法如下。

1) 利用前面所述的模型对原数据 L_0 进行小波分析,得到压缩的数据 L_1 。

2) 对 L_0 中的数据进行再处理。对介于 L_1 中的间隔数据 P_i, P_{i+2} 之间的 L_0 中的数据,利用 Douglas 算法求得 P_i, P_{i+2} 中的特征点 P_M (到直线 $P_i P_{i+2}$ 的距离最远的点, $i=1, 2, \dots, n-2, n$ 为 L_1 中数据之和)。

3) 计算点 P_{i+1} 与 P_M 的近似曲率:① 求得 P_{i+1}, P_M 到直线 $P_i P_{i+2}$ 的距离 d_{i+1}, d_M 及 $P_i P_{i+2}$ 的长度 d ;② 计算 P_{i+1} 与 P_M 的近似曲率, $\rho_{i+1} = d_{i+1}/d, \rho_M = d_M/d$ (d 为 $P_i P_{i+2}$ 的距离)。

4) 比较近似曲率 ρ_{i+1}, ρ_M 的大小。如果 $\rho_{i+1} < \rho_M$,则 P_M 是比 P_{i+1} 奇异性更大的点, L_1 中就用点 P_M 代替 P_{i+1} ,或直接将 P_M 重新插入到 L_1 中,曲线形态就会得到更好的保持。

数据 L_1 经过上述处理后,就会去掉其中的非特征点,而补回原数据 L_0 中丢失掉的特征点,从

而使压缩数据 L_1 更加优化。图 4 为在图 2 的基础上经过修正后的数据曲线,数据点有 375 个,压缩比为原数据的 37%。

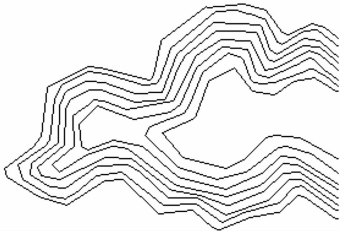


图 4 修正图
Fig. 4 Modified Image

其实,在数据的压缩优化模型算法中,数据的修正处理与小波分析是紧密联系、交替进行、重复循环的,而不是完全分离的步骤。

1.4 线状要素压缩优化的自适应模型

利用上述的讨论,可以得到线状要素压缩优化的自适应模型:① 对线状要素 $f(x)$,设 $L_0 = \{(x_m, y_m), m=0, 1, 2, \cdots, n\}$,求得其标准差或相邻点的平均距离,以确定阈值 ω ;② 选取三系数的小波基,对原数据 L_0 进行正交小波分解,得到低分辨率部分数据 $f_1(x)$ 和 高频部分数据 $g_1(x)$;③ 对高频部分数据 $g_1(x)$,如果 $|g_1(x)| < \omega$,则数据点为平滑点,否则为特征点,并将特征点插入到 $f_1(x)$;④ 将插入了特征点的 $f_1(x)$ 作为一次小波变换后的压缩数据 L_1 ;⑤ 对数据集 L_1 中的数据进行修正处理;⑥ 用数据 L_1 代替数据 L_0 ,重复步骤②~⑤,直到 $L_1 = L_0$ 为止。

2 实例分析

根据上述算法,下面对一组等高线数据进行试验。图 5 为原等高线数据,有 11 620 个数据点;图 6 为经小波分析后的压缩数据,有 4 203 个数据,压缩比为 36.17%。

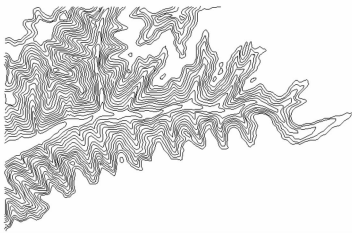


图 5 原等高线数据
Fig. 5 Original Contour Image

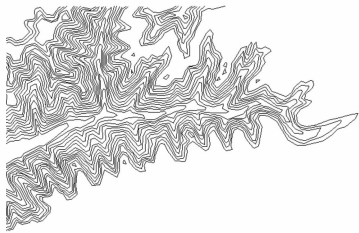


图 6 压缩后的数据
Fig. 6 Compressed Image

3 结 语

在模型的建立过程中,由于去掉的数据的奇异性即曲率都很小,所以压缩过的数据曲线之间的协调关系得到很好的保持。从上面的等高线试验中可以得到如下结论:在压缩过的数据中,山体轮廓、山脊、山谷走向等地貌形态都得到了很好的保持。但是在阈值的选取过程中,仍然有部分工作需要靠经验取值,还不能做到完全自动化。

参 考 文 献

[1] Wu Jitao, Wang Qiao. A Study on Automatic Cartographic Generalization Using Wavelet Analysis in GIS[J]. Acta Geodaetica et Cartographica Sinica, 2000, 4(1): 71-75

[2] Wang Jiayao, Wu Fang. The Method and Principle for Automatic Generalization of Digital Map[M]. Beijing: PLA Press, 1998

[3] Wang Yuhai, Zhu Changqing, You Xiong, et al. The Contour Data Simplifying Based on B-spline Wavelet [J]. Science of Surveying and Mapping, 2003, 28 (2): 23-25

[4] Wang Yuhai, Zhu Changqing. The Vector Relief Data Compression Based on the Multiband Wavelet [J]. Science of Surveying and Mapping, 2003, 28 (3): 66-68

[5] Zhu Changqing, Wang Yuhai, Li Qingquan, et al. A Model to Compress Contour Data Based on Wavelet Analysis [J]. Journal of Image and Graphics, 2004, 9(7): 841-845

第一作者简介:王玉海,讲师,博士生。主要从事 GPS、GIS、GSM 系统集成、空间数据处理、小波分析等研究。
E-mail:wyh7121@163.com

Data Structures[J]. Theoretical Computer Science, 1997, 181(1): 3-15

[4] Lawder J K, King P J H. Using Space-Filling Curves for Multi-dimensional Indexing [C]. The 17th British National Conference on Databases; Advances in Databases, London, 2000

[5] Breinholt G, Schierz C. Algorithm 781: Generating Hilbert's Space-Filling Curve by Recursion [J]. ACM Transactions on Mathematical Software, 1998, 24(2): 184-189

[6] Kamel I, Faloutsos C. Hilbert R-tree; an Improved R-tree Using Fractals[C]. The 1994 International

Conference on VLDB, Morgan Kaufmann, 1994

[7] Faloutsos C, Roseman S. Fractals for Secondary Key Retrieval [C]. The 8th ACM SIGACT-SIGMOD-SIGART Symposium on Principle of Database System, New York, 1989

[8] Kamel I, Faloutsos C. On Packing R-trees[C]. The 2nd International Conference on Information and Knowledge Management, New York, 1993

第一作者简介:王永杰,博士生。主要研究方向为 GIS、网络集群理论与技术、并行空间数据库。
E-mail:yjw1018@163.com

Spatial Partitioning of Massive Data Based on Hilbert Spatial Ordering Code

WANG Yongjie¹ MENG Lingkui¹ ZHAO Chunyu¹

(1 School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

Abstract: The excellent linear mapping characteristics of Hilbert spatial ordering code is studied and applied to spatial partitioning of data, and a concrete algorithm is given. In this algorithm, the clustering performance of spatial objects is taken into account, and the balance of data storage on each processing node is also taken into account, which greatly improves the processing efficiency of parallel spatial database.

Key words: PC cluster; Hilbert spatial ordering code; spatial partitioning of data

About the first author: WANG Yongjie, Ph. D candidate, majors in GIS, parallel spatial database, etc.
E-mail: yjw1018@163.com

(上接第 632 页)

Compression and Optimization of the Line Features
Based on Wavelet Analysis

WANG Yuhai¹ ZHU Changqing²

(1 Institute of Science, Information and Engineering University, 66 Middle Longhai Road, Zhengzhou 450052, China)
(2 Institute of Surveying and Mapping, Information and Engineering University, 66 Middle Longhai Road, Zhengzhou 450052, China)

Abstract: It is important to compress and optimize the line features in the studying of the terrain environment simultaion, cartographic generalization and GIS. Based on the theory of wavelet analysis, the algorithm of Douglas and curvature analysis, the compression and optimization of the line features and the self-adaptive model are studied. The results of some experiments show that the proposed approach maintains not only high compressing ratio but also the characteristics of the line features very well.

Key words: wavelet analysis; Douglas algorithm; curvature analysis algorithm; line features

About the first author: WANG Yuhai ,lecturer, Ph.D candidate, majors in interests are integration of GPS, GIS and GSM, spatial data processing and wavelet analysis.
E-mail: wyh7121@163.com