

小波分析在GIS线状数据图形简化中的应用研究*

吴纪桃¹, 王 桥²

(1. 北京航空航天大学理学院, 北京, 100083; 2. 南京师范大学地理科学学院, 江苏 南京, 210097)

A Study on Automatic Cartographic Generalization Using Wavelet Analysis in GIS

WU Ji-tao¹, WANG Qiao²

(1. Department of Mathematics Beijing University of Aviation and Space, Beijing, 100083;

2. College of Geographical Sciences Nanjing Normal University, Nanjing, Jiangsu, 210097)

Abstract: Based on the wavelet theory, the relation between multiresolution analysis and multi-scale data expression in GIS is found, a new model of using multiresolution analysis is put forth, and the application example to use the proposed model is given in this paper.

Keywords: Wavelet theory; Multiresolution analysis; Geographical information system; Automatic cartographic generalization

提 要: 本文应用小波理论建立了多分辨率分析与GIS图形数据多比例尺表达的关系, 在此基础上提出了一种基于多分辨率分析的自动制图综合模型, 并以线状要素自动综合为例, 说明了新模型方法的应用。

关键词: 小波理论; 多分辨率分析; 地理信息系统; 自动制图综合

1 引 言

随着现代社会对地理信息系统(GIS)需求规模的不断扩大和需求层次的日益提高, 人们越来越多地需要在不同分辨率, 不同空间尺度上对地理现象进行观察、理解和描述, 即越来越多地需要对多种比例尺的空间数据进行分析、处理和表达, 这就导致了对多比例尺特征GIS需求的出现。因此, 基于单一比例尺的GIS已越来越难以满足实际应用的需要。自80年代末, 人们开始不断地寻求解决这一问题的途径。目前通常采用的方法是

重复数字化, 即通过重复建库来满足对多比例尺数据的需要。然而, 这种方法从多比例尺数据产生的效率等方面看还存在一些不足。如耗资巨大, 数据采集与建库的工作繁重, 容易出现数据不稳定等情况。

近年来, 另一种解决问题的途径——空间信息多比例尺表达的自动制图综合方法开始引起人们的重视。此方法基于这样一种观点, 即认为所需的各种中小比例尺数据信息应该包含在相应的大比例尺数据之中, 因此, 可以利用已有的基于单一比例尺的GIS数据库中的基础地理数据, 通过自

* 收稿日期: 1999-04-28, 截稿日期: 1999-12-14。吴纪桃, 女, 42岁, 教授, 博士。现从事地图制图与GIS方面的研究。

制图综合生成多比例尺数据。显然,这种方法不但可以极大的减少多比例尺数据采集、存储、管理和更新的投入,而且可以扩大GIS基础数据库的服务面,提高其潜在的应用价值。

但是,自动制图综合方法的发展却不容乐观,在所面临的大量挑战中,自动化问题,即如何在GIS环境下自动地实现制图综合,是一个核心的难点。自动化的前提是模型化,所以问题的关键是GIS环境下自动制图综合模型的建立。而大多数情形下,GIS图形数据信息都可用线状图形来表达,这就使得研究的焦点主要集中到线状要素自动制图综合模型的建立方面。对此,国内外学者进行了不少研究,并先后提出了十余个应用模型。但经过深入分析不难发现,现有模型不少都是基于直观而不是机理的,操作性而不是分析性的。它们很难解决综合程度的客观化控制,阈值的自适应性选取,地理特征的自动保持等一系列问题,也很难从理论上保证综合结果的稳定性和收敛性。究其原因,一个带有根本性的问题是没有从数学机制上建立能够反映图形形状结构特征随比例尺变化的一般规律的分析模型。

小波理论是目前国际上公认的最新空间(时间)-频率分析工具,由于其“自适应性”和“数学显微镜性质”而成为许多学科共同关注的焦点。特别是小波理论中的多分辨率分析(MRA),因其可提供在不同分辨率下分析表达信息的有效途径,尤其受到地学领域学者的高度重视。但是,在GIS图形数据多比例尺表达研究方面,尽管多分辨率分析与制图综合具有天然的联系,目前还没有人真正把它应用到数据的多比例尺分析和自动制图综合研究中来。本文将针对上述自动制图综合模型研究中存在的种种问题,从理论上探索多分辨率分析在自动制图综合模型建立及其地理信息系统多比例尺数据表达中应用的新途径,从而为自动综合的滤波功能添加新的算法。

2 小波分析与GIS图形数据的多比例尺表达

2.1 小波分析及其基本性质

设 $L^2(\mathbb{R})$ 是一平方可积的Hilbert空间, $\{V_m\}_{m \in \mathbb{Z}}$ 是 $L^2(\mathbb{R})$ 的一列子空间, \mathbb{R} 表示实数集, \mathbb{Z} 表示整数集,若 $\{V_m\}$ 满足:

- (1) 对于任意 $m \in \mathbb{Z}, V_m \subset V_{m-1}$;
- (2) $\bigcap_{m \in \mathbb{Z}} V_m = \{0\}, \bigcup_{m \in \mathbb{Z}} V_m = L^2(\mathbb{R})$;

$$(3) f(x) \in V_m \Leftrightarrow f(2x) \in V_{m-1};$$

(4) 存在 $\varphi(x) \in V_0$,使得 $\{\varphi_{m,n}(x) = 2^{-m/2} \times \varphi(2^{-m}x - n)\}_{n \in \mathbb{Z}}$ 构成 V_m 的标准正交基, $m \in \mathbb{Z}$ 。则称 $\{V_m\}_{m \in \mathbb{Z}}$ 是 $L^2(\mathbb{R})$ 的一个二进多分辨率分析,并称 $1/2^m$ 为 V_m 的分辨率, $\varphi(x)$ 为该多分辨率分析的尺度函数。

关于尺度函数 $\varphi(x)$,由式(1)、(4)容易推得

$$\varphi(x/2) = \sum_{n \in \mathbb{Z}} h_n \varphi(x - n) \quad (1)$$

其中, $h_n = \langle 1/2 \varphi(x/2), \varphi(x - n) \rangle =$

$$1/2 \int_{-\infty}^{\infty} \varphi(x/2) \overline{\varphi(x - n)} dx \quad (2)$$

式(1)称为双尺度差分方程。称函数

$$\psi(x) = \sum_{n \in \mathbb{Z}} g_n \varphi(2x - n) \quad (3)$$

为多分辨率的小波函数,其中

$$g_n = (-1)^n h_{1-n}, \quad n \in \mathbb{Z} \quad (4)$$

可以证明,函数系

$$\{\psi_{m,n}(x) = 2^{-m/2} \psi(2^{-m}x - n)\}_{n \in \mathbb{Z}} \quad (5)$$

是标准正交的,其生成的子空间 W_m 是 V_m 在 V_{m-1} 中的正交补,即

$$V_m \oplus W_m = V_{m-1}, \quad m \in \mathbb{Z} \quad (6)$$

在上述4个性质中,性质(1)也称单调性,表示相应于某个分辨率的子空间包含了关于该空间低于此分辨率的所有信息;或相应于一尺度的空间包含了该空间大于该尺度的全部信息;性质(2)也称稠密性,表示所有子空间可以构成 $L^2(\mathbb{R})$,而随着分辨率的降低(即 m 增大), V_m 中所包含 $L^2(\mathbb{R})$ 中的信息越来越少。由性质(1),性质(2)实际上可等价地表达成

$$\lim_{m \rightarrow -\infty} V_m = L^2(\mathbb{R}), \quad \lim_{m \rightarrow +\infty} V_m = \{0\} \quad (7)$$

性质(3)表示所有的子空间序列是由一个基本空间经过尺度伸缩所生成的,即如果知道其中任何一个子空间,比如 V_0 ,就可派生出 V_1, V_2, \dots, V_m 等等;性质(4)表示所有这些子空间的基是由一个函数 $\varphi(x)$ 生成的,只要 $\varphi(x)$ 选定,就可表示出 $\{V_m\}_{m \in \mathbb{Z}}$,由式(3)、(4)、(5),可进一步知道 $\varphi_{m,n}(x)$,从而可得出 $\{W_m\}_{m \in \mathbb{Z}}$ 。

2.2 基于小波分析的GIS图形数据多比例尺表达

从上述说明中可看出,如果我们将空间 $L^2(\mathbb{R})$ 看成是某地理空间, $f(x)$ 看成是该空间的各种信息,那么, $\{V_m\}_{m \in \mathbb{Z}}$ 则可看成 $L^2(\mathbb{R})$ 在不同分辨率下的地理空间模型, $\{W_m\}_{m \in \mathbb{Z}}$ 则是相应的

细节信息; 更具体地说, 如果将空间 $L^2(\mathbb{R})$ 看成某地理空间在特定比例尺下的 GIS 图形数据模型, $f(x)$ 是其上各图形要素, 那么, $\{V_m\}_m$ 则可看成多比例尺 GIS 图形数据模型 (在没有进行相应的比例尺缩小变换情形下)。在实际应用中, 分辨率是有限的, 所以可以认为 $L^2(\mathbb{R}) = V_0$ 。这样, 从 V_0 出发, 应用尺度函数 $\mathcal{Q}(x)$ 表示出 V_1, V_2, \dots, V_m 的过程则看成从一种比例尺的 GIS 图形数据模型 V_0 派生出多比例尺 GIS 图形数据模型 V_1, V_2, \dots, V_m 的过程。这个过程实际上就是一种基于多分辨率分析的 GIS 图形数据的多比例尺表达的过程, 也就是一种基于多分辨率分析的自动制图综合的过程。在实际操作中, 自动制图综合过程一般是分要素或分内容进行的。所以, 不失一般性, 我们可以设 V_0 只包含一种图形要素, 比如线状要素 $y = f(x)$ 。

3 基于小波分析的自动制图综合模型

3.1 模型的建立

根据 2.1 及 2.2 所给出的基本原理, 我们可以建立基于多分辨率分析的自动制图模型。设比例尺为 $1/M$ 的 GIS 图形曲线可由方程 $y = f(x)$ 表达, 其离散形式可写成

$$\{f(x_i)\}_{i=1,2,\dots,2^N} \quad (8)$$

经过适当的插值和坐标变换, 可将式(8)式表示为

$$V_0 = \{f(n)\}_{n=1,2,\dots,2^N} \quad (9)$$

根据上述多分辨率分析原理, 若尺度函数为 $\mathcal{Q}(x)$, 则对于 V_0 , 曲线 $y = f(x)$ 的表达式可写成

$$f(x) = \sum_{n=0}^{2^N-1} c_{0,n} \mathcal{Q}_{0,n}(x) \quad (10)$$

由于 $V_0 = V_1 \oplus W_1$, 所以

$$f(x) = \sum_{n=0}^{2^{N-1}-1} c_{1,n} \mathcal{Q}_{1,n}(x) + \sum_{n=0}^{2^{N-1}-1} d_{1,n} \mathcal{Q}_{1,n}(x) \quad (11)$$

式(11)右端第一项是 $f(x)$ 在 V_1 中的表达式, 记为 $A_1 f(x)$; 第二项是 $f(x)$ 相对于 V_1 在 V_0 中的细节信息, 记为 $W_1 f$ 。

又由于 $V_1 = V_2 \oplus W_2$, 所以

$$f(x) = \sum_{n=0}^{2^{N-2}-1} c_{2,n} \mathcal{Q}_{2,n}(x) + \sum_{n=0}^{2^{N-2}-1} d_{2,n} \mathcal{Q}_{2,n}(x) + \sum_{n=0}^{2^{N-2}-1} d_{1,k} \mathcal{Q}_{1,n}(x) \quad (12)$$

式(12)右边第一项是 $f(x)$ 在 V_2 中的表达式, 记为 $A_2 f$; 第二项是 $A_1 f(x)$ 相对于 V_2 在 V_1 中的细

节信息, 记为 $W_2 f$; 一般地, 有 $f(x)$ 在 V_m 中的表达式 $A_m f$ 和其在 V_{m-1} 中的细节信息 $W_m f$

$$A_m f = \sum_{n=0}^{2^{N-m}-1} c_{m,n} \mathcal{Q}_{m,n}(x) \quad (13)$$

$$W_m f = \sum_{n=0}^{2^{N-m}-1} d_{m,n} \mathcal{Q}_{m,n}(x) \quad (14)$$

利用正交性可得

$$c_{j+1,n} = \overline{h_{k-2n} c_{j,n}} \quad (15)$$

$$d_{j+1,n} = \overline{g_{k-2n} c_{j,n}}, \quad n = 0, 1, 2, \dots, 2^{N-j-1} \quad (16)$$

记: $H = (\overline{h_{k-2n}}), G = (\overline{g_{k-2n}})$

则式(15)、(16)可简记为

$$c_{j+1} = H G_j, \quad d_{j+1} = G C_j \quad (17)$$

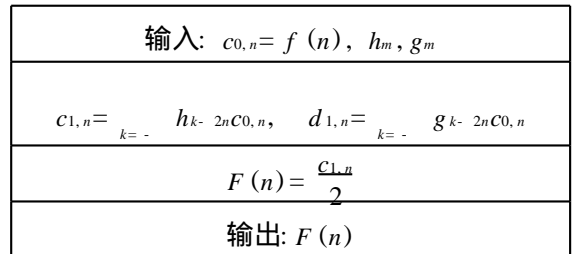
这是逐次计算系数 $c_{j,n}, d_{j,n}$ 的递推公式 (此算法称为 Mallat 算法)。

实际中 $\mathcal{Q}(x)$ 往往写不出其在空间域中的解析表达式。若 $f(x)$ 是离散表达式时, 可设

$$c_{0,n} = f(n), \quad n = 0, 1, 2, \dots, 2^N \quad (18)$$

而 h_m 和 g_m 的值在一些情形中可以查表获得。

由上述讨论可得出如下曲线简化模型运行框图:



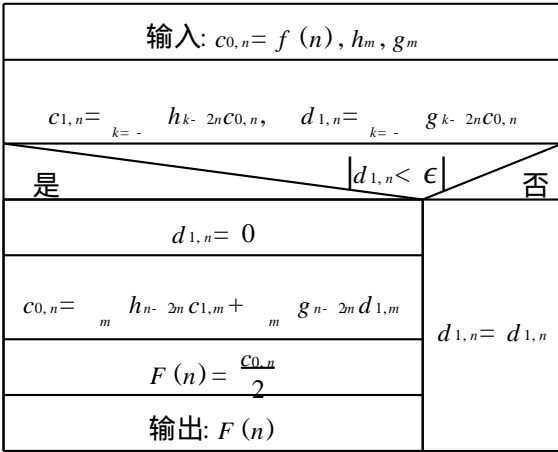
实验结果表明, 曲线 $F(n)$ 往往是 $f(n)$ 的强简化曲线 (综合掉的信息过多)。但实际制图综合中人们并不希望信息纯粹地按比例减少, 而是需要适当地多保留一些信息。事实上, 著名的方根规律就直接体现了这种思想, 它已成为多年来人工制图综合所遵循的基本原则。为了“补偿”所去掉的过多信息, 我们来研究被综合掉的信息 $d_{1,n}$ 。由小波理论知, $d_{1,n}$ 中绝对值较大的项对应原曲线 $f(n)$ 上奇异性大的点, 这些点往往是原曲线的特征所在。故要想将所去掉的过多信息“补回来”, 应考虑将 $d_{1,n}$ 中绝对值较大的值“补回来”, 为此, 令

$$d_{1,n} = \begin{cases} d_{1,n}, & |d_{1,n}| \in \\ 0, & |d_{1,n}| < \epsilon \end{cases} \quad (19)$$

将新的 $d_{1,n}$ 代入由 $c_{1,n}, d_{1,n}$ 反演 $c_{0,n}$ 的公式

$$C_{0,n} = \sum_m h_{n-2m} C_{1,m} + \sum_m g_{n-2m} d_{1,m} \quad (20)$$

至此, 我们可给出更能客观反映制图综合规律的曲线简化模型的运行框图:



3.2 模型的扩展

上述模型方法限于曲线只有上下波动, 即可表示为单值函数 $y = f(x)$ 的情形。但 GIS 图形往往是各向波动的闭合曲线(如等高线, 区划边界等等)。在此情况下我们做如下极化变换。

对于 $\{(x_i, y_i)\}_{i=1,2,\dots,l}$, 设

$$x_0 = \sum_{i=0}^l x_i / (l+1), \quad y_0 = \sum_{i=0}^l y_i / (l+1) \quad (21)$$

$$\theta = \arctan(y_i - y_0) / (x_i - x_0) \quad (22)$$

$$r_i = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2} \quad (23)$$

经过适当的插值和坐标变换, 可使 θ 成为等距的 2^n 个点。由此, 原始 GIS 图形曲线点列可表示为

$$r_n = r(\theta) = f(n), \quad n = 1, 2, \dots, 2^n \quad (24)$$

这样, 式(24)即归结为 3.1 节中模型方法所能够处理的形式, 故可沿用上述方法得出相应的结果。

有时, $r_n = f(n)$ 仍可能出现多值, 此种情形我们将另文讨论。

4 应用举例与综合程度的小波变换描述

图 1 给出了本文提出的基于多分辨率分析自动综合模型方法的应用实例。图中的第一条曲线是综合前的原始图形(比例尺为 1:25 万的 GIS 数字化曲线), 其余均为经过上述模型的计算机实现所得出的多比例尺表达图形。为了更好地说明综合效果, 这里没有再作比例变换。

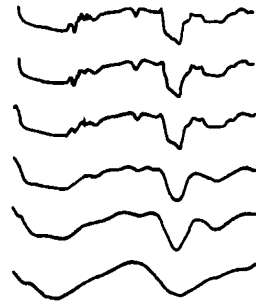


图 1 基于多分辨率分析的自动综合

Fig. 1 Automatic cartographic generalization based on the multiresolution analysis

图 2 与图 3 给出了将本文所提出算法用于河网数据的实例。其中图 2 是综合前海南省 1:5 万河网数据(局部); 图 3 是综合到 1:10 万后的相应数据图形。

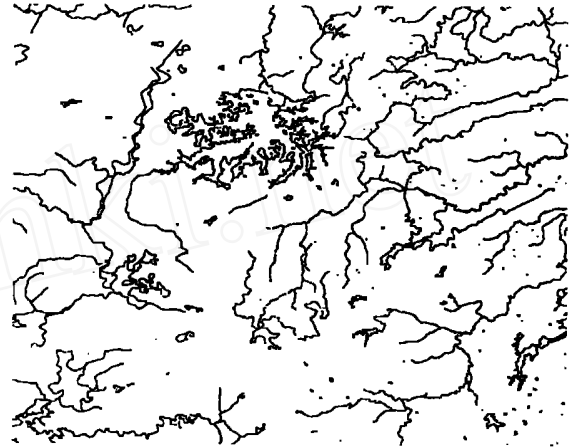


图 2 综合前河网数据图形

Fig. 2 The map before generalization



图 3 综合后河网数据图形

Fig. 3 The map after generalization

根据小波分析理论及本文获得的结论, 由上述基于多分辨率分析自动综合模型方法每得出一次 GIS 图形曲线的多比例尺表达 A_{mf} , 就可以同时得出其与前次表达 A_{m-1f} 之间相差的细节信息

W_{mf}

$$W_{mf} = \sum_{n=0}^{2^{m-1}} d_{m,n} \Psi_{m,n}(x) \quad (25)$$

记 $D_m = \sum_{j=1}^{2^{m-1}} d_{j,n} \Psi_{j,n}(x) \quad (26)$

容易知道, D_m 描述了从 V_0 到 V_m 的逐次综合所去掉的细节信息, 同时也是对综合程度的直接刻画。上述说明, 本文提出的基于多分辨率分析自动综合模型方法不仅能够方便、有效地进行 GIS 图形数据的多比例尺表达, 而且还能够对每一次综合所去掉的细节信息进行精确的数量描述, 从而对综合程度进行数学分析, 这是现有自动制图模型方法难以做到的, 并对于自动制图综合研究具有非常重要的意义。因为综合是对原始信息的抽象与概括, 也就是对原始信息的近似, 而任何近似只有在真正了解和掌握了近似程度的前提下才有实用价值。

References:

- [1] Q NG Q ian-qing, YANG Zong-kai Wavelet analysis[M] XiA n: Publishing Housing of XiA n Technological University of Electronics, 1988 (in Chinese)
- [2] L U Gui-zhong, D I Shuang-liang Wavelet Analysis and Application[M] XiA n: Publishing Housing of XiA n Technological University of Electronics, 1997. (in Chinese)
- [3] M allat S A Theory for Multiresolution Signal Decomposition : the Wavelet Representation [J] IEEE Trans on PAM I, 1989, 11 (7): 674-693
- [4] CHU IC K. An Introduction to Wavelet[M] New York : Academic Press , 1992
- [5] WANG Q iao, HU Yu-ju Wavelet Analysis of Fractal Brown Motion [J] Journal of Wuhan Technical University of Surveying and Mapping, 1995, 20(1): 66-70 (in Chinese)
- [6] WU Ji-tao. Self-similarity Analysis and Automatic Generalization of Cartographic Line[J] Journal of Wuhan Technical University of Surveying and Mapping, 1995, 20(1): 66-70 (in Chinese)
- [7] WANG Q iao, WU Ji-tao. Geographic information generalization [J] Journal of Remote Sensing, 1998, 2(2): 155-160 (in Chinese)
- [8] WANG Q iao, WU He-hai Fractal Describing and Automatic Generalization of Map Information[M] Wuhan: Publishing Housing of Wuhan Technical University of Surveying and Mapping, 1988 (in Chinese)
- [9] M uller J C. GIS and Generalization: Methodology and Practice [M] London : Taylor & Francis, 1996
- [10] M SMASTER R B, SHES K S Generalization in Digital Cartography[C] Washington D. C: A ssociation of American Geographers, 1992 (in Chinese)