

几种离散数据网格化方法的对比分析

安 琪

(内蒙古工业大学信息工程学院)

摘 要: 本文讨论了几种离散数据网络化方法, 并从逼近程度、外推能力、唯一性、计算速度等几个方面对这些方法作了综合分析和评价。

关键词: 网络化; 曲面拟合; 插值; 趋势面; 权; 最小二乘法

中图分类号: G202 **文献标识码:** A **文章编号:** 1007—6921(2001)01—0087—02

在地质、地理、地球物理、地球化学等地学领域中, 广泛使用等值线图、曲面图、剖面图等。绘制这些图件主要利用原始采样点数据, 这些数据往往受采样环境、采样人员及采样所使用仪器的影响, 或则分布不规则及采样点数量不足, 或则包含随机性干扰, 因此一般先用网络化方法使其转化为规则网分布的数据, 然后再来制图或进行其它处理。

场值的分布, 即平面上各种量的变化, 始终是地学研究的重要课题。例如, 地磁场、重力场、地球化学场是寻找多种金属矿床的依据。平面上变化的量构成曲面, 所以场值分布实质上是曲面问题。

本文介绍的曲面处理方法(即网络化方法)有两类, 一类是曲面拟合, 这种方法是用数学曲面来近似逼近复杂的地学曲面, 通过拟合处理的曲面, 原始数据点值要改变, 所以拟合的结果往往取得平滑的效果。另一类方法是插值, 这类方法不改变原始数据点值, 是根据原始数据点值来插补空白区的值。

1 网格化方法

1.1 近点按距离加权平均法(简称距离法)

规则分布的网格点上的值要由离散分布的数据点值来确定, 按距离加权平均法假定离网格点越近的数据点对网格点的影响越大。该法只考虑离网格点最近的几个数据点, 这几个点对网格值的影响与距离有关, 距离网格点越远影响越小。

令数据点 (x_i, y_i) 到网格点 (a, b) 的距离为 D_i , 则有:

$$D_i = \sqrt{(x_i - a)^2 + (y_i - b)^2} \quad (1)$$

求出离 (a, b) 最近的几个数据点的距离 $D_i (i = 1, 2, \dots, n)$ 后, 则网格点 (a, b) 上的值为:

$$Z(a, b) = \frac{\sum_{i=1}^n \frac{Z_i}{D_i}}{\sum_{i=1}^n \frac{1}{D_i}} \quad (2)$$

式中 Z_i 为数据点 (x_i, y_i) 上的值。

1.2 按方位取点加权法(简称方位法)

距离法是取离网格点最近的几个数据点值来计算网格点值, 不考虑方向, 所取的点很可能集中到某一侧, 其它方向会取不到点。

按方位取点加权法以网格点为中心把区域划分成若干个象限, 从每个象限内取一点作加权平均, 这就克服了距离法偏向的缺点。

若求某个网格点 (i, j) 的函数值时, 则以 (i, j) 为原点将平面分成四个基本象限, 再把每个象限等分成 n_0 份, 这样就把全平面分成 $4n_0$ 等份。然后在每个等分角内寻找一个离 (i, j) 最近的数据点, 其值为 Z_{il} , 它到 (i, j) 的距离为 r_{il} , 则网格 (i, j) 上的值为:

$$Z(i, j) = \sum_{il=1}^{4n_0} C_{il} \cdot Z_{il} \quad (3)$$

$$\text{式中: } C_{il} = \frac{\prod_{j=1}^{4n_0} r_j^2}{\sum_{k=1}^{4n_0} \prod_{l=k}^{4n_0} r_l^2} \quad (4)$$

1.3 趋势面拟合法

趋势面分析是地学领域常用的数学方法, 它可以把长周期的趋势性变化和短周期的局部性变化分开, 根据趋势性变化可以总结规律, 根据局部性变化可以发现异常。从计算机制图角度看, 趋势面分析是用简单的幂级数多项式来拟合复杂的地学曲面, 有削平、填平实际曲面的作用。为了准确地反映实际曲

面,是不能单独采用这种方法的。所以,仅使用趋势面或残差(数据点值与趋势面计算值之差)作图时,总是为了其它目的,比如了解场值的区域性变化规律或了解异常的部位及程度。

令趋势面次数为 m , 则趋势面分析的公式为:

$$Z(x, y) = \sum_{k=0}^m \sum_{i=0}^k a_{k,i} x^{(k-i)} y^i \quad (5)$$

m 次趋势面的系数 $a_{k,i}$ 的个数为

$$S = \frac{(m+1)(m+2)}{2} \quad (6)$$

若有几个数据点值 Z_1, Z_2, \dots, Z_n , 现在要作 m 次趋势面(应有 $n \geq S$), 需要对几个数据点值用最小二乘法确定系数 $a_{k,i}$ 。

先作数据点值与理论值差的平方和

$$\begin{aligned} f &= \sum_{j=1}^n (Z(x_j, y_j) - Z_j)^2 \\ &= \sum_{j=1}^n \left(\sum_{k=0}^m \sum_{i=0}^k a_{k,i} x_j^{(k-i)} y_j^i - Z_j \right)^2 \end{aligned} \quad (7)$$

再求 f 对系数 $a_{k,i}$ 的偏导数并令其为零

$$\begin{aligned} \frac{\partial f}{\partial a_{k,i}} &= 2 \sum_{j=1}^n \left(\sum_{k=0}^m \sum_{i=0}^k a_{k,i} x_j^{(k-i)} y_j^i - Z_j \right) a_{k,i} \\ &= 0 \end{aligned} \quad (8)$$

这样线性代数方程组有 S 个方程和 S 个未知数 $a_{k,i}$, 求出 $a_{k,i}$ 后就确定了趋势面式(5)。

1.4 趋势面和残差叠加法(简称叠加法)

在前面讨论的方法中, 距离法和方位法均以反映局部变化为特征, 外推能力(反映空白地区的能力)差, 当数据点不均匀时, 使用这两种方法的效果不好。而趋势面拟合法是建立在全部数据点的基础上, 它考虑了整个区域的变化特征, 并把这些特征融合到一个统一的多项式中, 虽然有些外推能力, 但拟合效果差。能不能产生一个既能反映局部特征, 又能反映全局的新方法呢? 趋势面或残差叠加法即是这样一个新方法, 它的运算步骤是:

利用前面介绍的趋势面拟合法, 首先拟合一个 m 次的趋势面。 m 是任意的, 可以根据实际情况来选择。

作出数据点值与该趋势面之间的残差。

利用距离法或方位法将残差作加权处理, 分配到网格点上。

将网格点上的趋势值和残差拟合值相加, 作为网格点值。

1.5 加权最小二乘拟合法(简称最小二乘法)

对于趋势面拟合法, 在给定了多项式的方次之后, 根据最小二乘原理求出多项式系数, 然后把网格

点坐标代入多项式就可得到网格值, 整个区域只有一个多项式。在趋势面拟合法的基础上, 加权最小二乘拟合法引入了距离权的概念。比如求网格点 (a, b) 的值, 要考虑全部数据点对 (a, b) 的贡献, 距 (a, b) 近的点权大, 远的点权小, 换一个网格点则形成另外一些权值。这样, 每一个网格点值都对应一个多项式, 求一个网格点值就要解一次联立方程。所以最小二乘法一方面有趋势面法考虑全部数据点反映趋势性变化的优点, 另一方面又有距离法反映局部特征的优点。

假定要求网格点 (a, b) 上的值 $f(a, b)$, 为此需要求出一个多项式 $P(x, y)$, 一般是二次多项式:

$$\begin{aligned} P(x, y) &= C_{00} + C_{10}X + C_{01}Y + C_{20}X^2 + C_{11}XY \\ &\quad + C_{02}Y^2 \end{aligned}$$

对于一般趋势分析应有:

$$Q = \sum_{i=1}^n [P(x_i, y_i) - Z_i]^2 \text{ 为最小} \quad (9)$$

式中 n 是数据点个数, (x_i, y_i) 是数据点坐标, Z_i 是 (x_i, y_i) 上的测量值。

最小二乘法要考虑按距离加权, 应对(9)式作些改动:

$$\begin{aligned} Q &= \sum_{i=1}^n [P(x_i, y_i) - Z_i]^2 \cdot W[(x_i - a)^2 \\ &\quad + (y_i - b)^2] \end{aligned} \quad (10)$$

式中 $W[(x_i - a)^2 + (y_i - b)^2]$ 就是权, 它是距离的函数。

为了求出方程(10)中系数 C_{rs} , 按最小二乘原理应有:

$$\frac{\partial Q}{\partial C_{rs}} = 0 \quad (r, s = 0, 1, 2) \quad (11)$$

上式有 6 个未知数和 6 个方程的方程组, 可用高斯消去等方法求解。

在实际应用时可给出三种加权形式:

$$\text{I 型 } W(d^2) = \frac{1}{(d^2 + \epsilon)} \quad (12)$$

式中 d 为网格点 (a, b) 与离散数据点 (x_i, y_i) 的距离, 即 $d^2 = (x_i - a)^2 + (y_i - b)^2$, ϵ 是一个很小的数, 如 10^{-6} , 为了防止运算时溢出。

$$\text{II 型 } W(d^2) = \frac{1}{(d^2 + \epsilon)^4} \quad (13)$$

d 与 ϵ 的意义同 I 型。

$$\text{III 型 } W(d^2) = e^{-\frac{\alpha d^2}{d^2 + \epsilon}} \quad (14)$$

式中 α 是常数, 其数量级相当于相邻数据点平

转 变 教 育 观 念 提 高 教 学 管 理 水 平

唐振华 张述禹 石延聪

(内蒙古医学院)

中图分类号: G 642 文献标识码: A 文章编号: 1007—6921(2001)01—0089—02

“教育要面向现代化、面向世界、面向未来”是高校教学改革的指导方针,实现教育的现代化,其管理的现代化是不可缺少的组成部分。如何实现教育管理现代化,我院从以下几方面进行了尝试。

1 更新教育思想,转变教育观念

教师是学校教学活动的主导,是深化教学改革的关键。为了使广大教师认识教学改革的形势,解放思想,积极探索,勇于创新,用符合时代发展的教育思想和观念指导教学改革实践,学院开展了转变教育思想,更新教育观念的大讨论,在全院教师会议上,学院领导做了动员报告,进一步统一思想,增强信心明确方向。学院还聘请医学教育研究专家来我院讲学,介绍医学教育发展趋势以及教学改革经验。学院领导和有关系处负责人专程赴区内外院校进行考察学习,为深化改革奠定了思想基础。同时,学院

加大改革开放力度,在吸取和借鉴其他院校先进经验的基础上,不断改善自我,走上了良性发展的道路。学院 1997 年开始与军事医学科学院协议,每年为我院培养在职研究生,现已培养研究生 10 名。1999 年做为北医世行贷款项目协作学校,在北医指导和资助下进行了化学实验改革和化学实验中心建设,接受北医骨干教师来院进行示范教学,并分期分批选派教师和专业技术人员去北京学习,借助北医强大的学科优势联合招收培养研究生,合作开展教学研究。2000 年与北医大确定开展现代远程教育意向,进一步开展网络课程等建设。

2 探讨学分制运作机制,为推行学分制管理创造条件

从 1999 年开始,高校走上了持续扩招的道路,高校大规模扩招,为学校带来了发展机遇,同时也带

均距离平方的倒数。 d 与 ϵ 的意义同 I 型。

I 型网格化方法使离网格点较远的数据点对网格点的值也有较大影响,因此输出的图形比较平滑。

II 型网格化方法突出离网格点近的数据点值的影响,输出的图形比较准确地反映实际情况。

III 型网格化方法与 a 值有关, a 越小图形越平滑。

2 几种网格化方法的对比分析

以上讨论了“距离法”、“方位法”、“趋势面法”、“叠加法”、“最小二乘法”等五种方法,考虑“最小二乘法”有三种形式,所以总共有七种方法。采用不同的方法所得到的结果是有差异的,因此在作图时应慎重选择网格化方法。事实上,每种方法都针对特定的要求。例如,趋势面拟合法是用统一的一个多项式来拟合离散分布的测量值。如果我们的目的是反映测量值的趋势性变化,对测量值本身的兴趣不大,就可以采用这种方法。就是趋势面拟合法本身还是有差异的,低次趋势面平滑性好,高次趋势面逼近程度高。一般来说,可以从逼近程度、外推能力、唯一性(当网格分割有变化,即改变网格间距时,所输图形

有何变化,若大部分可重合则唯一性强)、计算速度等方面来分析和评价各种曲面处理方法。

现将分析结果列表如下:

方法名称	逼近程度	外推能力	唯一性	运算速度	适用范围
距离法	当数据分布均匀时高	很差	很差	快	适用于均匀分布的数据点
方位法	当数据分布均匀时高	很差	很差	快	适用于均匀分布的数据点
趋势面法	不高	强	很强	很快	不宜作准确的等值线图
叠加法	很高	强	较强	较快	大量均匀分布也可用
最小二乘法 I 型	比较高	强	很强	慢	非均匀分布也可用,要求计算机速度快或网格少
最小二乘法 II 型	较 I 型为高	强	很强	慢	非均匀分布也可用,要求计算机速度快或网格少
最小二乘法 III 型	一般较高,与 a 系数有关	强	很强	慢	非均匀分布也可用,要求计算机速度快或网格少

【参考文献】

[1]徐士良编著. 计算机常用算法. 清华大学出版社, 1995
收稿日期: 2000 年 9 月 12 日