

文章编号: 1006-2939(2003)03-0053-05

距离加权反比插值法和克里金插值法的比较

靳国栋, 刘衍聪, 牛文杰

(石油大学(华东)机电工程学院, 山东 东营 257061)

摘要: 首先给出了距离加权反比插值法和克里金插值法的原理, 然后对这两种方法在理论和实际应用上进行了比较, 其结论是克里金插值法的效果优于距离加权反比插值法的效果。

关键词: 数据场可视化; 距离加权反比插值法; 克里金插值法; 变异函数

中图分类号: TP391 **文献标识码:** A

0 引言

科学计算可视化^[1]研究的对象是三维数据场, 而图形仿真方法的数据场可视化需要通过对离散数据场进行区域剖分, 插值后建立数学模型, 再从中提取曲线、曲面等中间几何信息, 利用传统的计算机图形学方法加以显示, 从而获取对数据场内部信息的观察。运用图形仿真方法, 插值是其中非常关键的一步, 插值的过程就是建立所研究变量完整的数学模型的过程。采用不同的插值方法建立的数学模型会有差别, 后期对其规律的显示, 从效果上看也是不相同的。数据场可视化常用的插值方法一般有距离加权反比法^[2]、多项式最小二乘法等, 而地质统计学中的克里金插值方法近年来在许多领域的可视化中得到了应用。克里金插值方法^[3]最早是由南非金矿工程师克里金(Krige D G)于20世纪50年代提出来的, 依据不同的条件又分为普通克里金法、泛克里金法、协同克里金法等。目前, 已有的文献很少研究克里金插值方法与其它插值方法的比较。为此, 文中以煤层数据可视化过程中插值建立层面数学模型为例, 对普通克里金法和距离加权反比法在理论和实际应用效果上进行了比较。下面首先给出这两种插值方法的原理, 然后在理论上进行比较, 以实际应用效果论述两方法的优缺点, 最后给出结论。

1 距离加权反比法插值原理

设空间待插点为 $P(x_p, y_p, z_p)$, P 点邻域内有已知散乱点 $Q_i(x_i, y_i, z_i)$, $i=1, 2, \dots, n$, 利用

距离加权反比法对 P 点的属性值 Z_p 进行插值。其插值原理是待插点的属性值是待插点邻域内已知散乱点属性值的加权平均, 权的大小与待插点与邻域内散乱点之间的距离有关, 是距离的 $k(0 \leq k \leq 2)$ (k 一般取 2) 次方的倒数。即:

$$Z_p = \frac{\sum_{i=1}^n \frac{Z_i}{d_i^k}}{\sum_{i=1}^n \frac{1}{d_i^k}}$$

其中: d_i 为待插点与其邻域内第 i 个点之间的距离。

2 普通克里金法插值原理

设研究区域为 A , 区域化变量(即欲研究的物理属性变量)为 $\{Z(x) \in A\}$, x 表示空间位置(一维、二维或三维坐标), $Z(x)$ 在采样点 $x_i (i=1, 2, \dots, n)$ 处的属性值(或称为区域化变量的一次实现)为 $Z(x_i) (i=1, 2, \dots, n)$, 则根据普通克里金插值原理, 未采样点 x_0 处的属性值 $Z(x_0)$ 估计值是 n 个已知采样点属性值的加权和, 即:

$$Z(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \quad \lambda_i (i=1, 2, \dots, n)$$

为待求权系数。

假设区域化变量 $Z(x)$ 在整个研究区域内满足二阶平稳假设:

- (1) $Z(x)$ 的数学期望存在且等于常数: $E[Z(x)] = m$ (常数)。
- (2) $Z(x)$ 的协方差 $\text{Cov}(x_i, x_j)$ 存在且只与两

收稿日期: 2003-05-29

作者简介: 靳国栋(1970—), 男, 山西运城人, 石油大学(华东)硕士研究生, 主要从事计算机辅助设计研究。

点之间的相对位置有关。

或满足本征假设:

$$(3) E[Z(x_i) - Z(x_j)] = 0.$$

$$(4) \text{增量的方差存在且平稳: } \text{Var}[Z(x_i) - Z(x_j)] = E[Z(x_i) - Z(x_j)]^2.$$

依据无偏性要求: $E[Z^*(x_0)] = E[Z(x_0)]$ 经推导可得:

$$\sum_{i=1}^n \lambda_i = 1$$

在无偏条件下使估计方差达到最小, 即:

$$\text{Min}\{\text{Var}[Z^*(x_0) - Z(x_0)] - 2\mu \sum_{i=1}^n (\lambda_i - 1)\}$$

其中: μ 为拉格朗日乘子。

可得求解系数 $\lambda_i (i=1, 2, \dots, n)$ 的方程组:

$$\begin{cases} \sum_{i=1}^n \lambda_i \text{Cov}(x_i, x_j) - \mu = \text{Cov}(x_0, x_i) \\ \sum_{i=1}^n \lambda_i = 1 \end{cases} \quad i = 1, 2, \dots, n$$

求出诸系数 $\lambda_i (i=1, 2, \dots, n)$ 后, 就可求出未采样点 x_0 处的属性值 $Z^*(x_0)$ 。

上述求解系数 $\lambda_i (i=1, 2, \dots, n)$ 的方程组中协方差 $\text{Cov}(x_i, x_j)$ 若用变异函数 $\gamma(x_i, x_j)$ 表示时, 形式为:

$$\begin{cases} \sum_{i=1}^n \lambda_i \gamma(x_i, x_j) - \mu = \gamma(x_0, x_i) \\ \sum_{i=1}^n \lambda_i = 1 \end{cases} \quad i = 1, 2, \dots, n$$

变异函数的定义为:

$$\gamma(x_i, x_j) = \gamma(x_i - x_j) = \frac{1}{2} E[Z(x_i) - Z(x_j)]^2$$

由克里金插值所得到的方差为

$$\sigma^2 = \text{Var}[Z^*(x_0) - Z(x_0)] =$$

$$\text{Cov}(x_0, x_0) - \sum_{i=1}^n \lambda_i \text{Cov}(x_0, x_i) + \mu$$

或

$$\sigma^2 = \sum_{i=1}^n \lambda_i \gamma(x_0, x_i) - \gamma(x_0, x_0) + \mu$$

2.1 克里金插值法中的变异函数

变异函数是克里金插值法插值的基础。插值中需要首先确定所研究的区域化变量的变异函数。假设研究的区域为 A , 区域 A 中有一区域化变量 $Z(x)$, 它在位置 $x_i \in A, i=1, 2, \dots, N$ 上的一次采样为 $Z(x_i), i=1, 2, \dots, N$, 则 $Z(x)$ 的变异函数的定义为:

$$\gamma(x_1, x_2) = \frac{1}{2} E\{[Z(x_1) - Z(x_2)]^2\}$$

一个空间变量的空间变异性是指这个变量在空间中如何随着位置的不同而变化的性质。变异函数通过其自身的结构及其各项参数从不同的角度反映空间变异性, 确定变异函数的过程就是一个对空间变异性进行结构分析的过程。

设 h 是一个模为 $r = |h|$, 方向为 α 的向量, 如果存在着被向量 h 所隔开的 N_h 对观测数据点, 则在 α 方向上相应于向量 h 的实验变异函数 $\gamma^*(h)$ 可表示为如下形式:

$$\gamma^*(h) = \gamma^*(r, \alpha) =$$

$$\frac{1}{2N_h} \sum_{i=1}^{N_h} [Z(x_i + h) - Z(x_i)]^2$$

其中: $z(x_i + h)$ 和 $z(x_i)$ 分别位于点 $x_i + h$ 和 $x_i, i=1, 2, \dots, N_h$ 上的观测数据。

2.2 变异函数理论模型

当获取实验变异函数值后, 需要先选择变异函数理论模型, 然后对所选择的变异函数理论模型进行参数拟合, 这一过程被称作“结构分析”。

变异函数理论模型参数一般包括: 变程 (Range, 一般用 a 表示)、基台 (Sill, 一般用 $C(0)$ 表示)、拱高 (一般用 C 表示) 块金常数 (Nugget, 一般用 C_0 表示), 如图 1 所示。

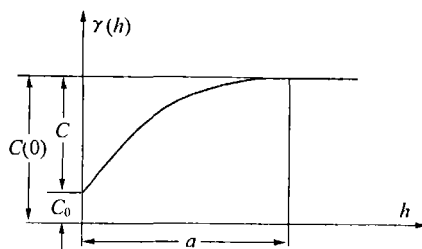


图 1 变异函数图

变程 a 表示了从空间相关性状态 ($|h| \leq a$) 向不存在相关性状态 ($|h| > a$) 转变的分界线; 变异函数在原点处的间断性称为“块金效应”, 相应的常数 $C_0 = \lim_{h \rightarrow 0} \gamma(h)$ 称为“块金常数”; 基台 $C(0)$ 具有协方差函数: $C(h) = C(0) - \gamma(h)$ 的二阶平稳区域化变量 $Z(x)$ 的先验方差: $\text{Var}[Z(x)] = C(0) = \gamma(\infty)$; 拱高 C 为变异函数中基台 $C(0)$ 与块金常数 C_0 之差: $C = C(0) - C_0$ 。

2.3 变异函数理论模型分类

变异函数理论模型一般分为有基台值和无基台值两大类。有基台值的变异函数理论模型包括球状模型、指数模型、高斯模型等。最常用的是球

状模型。球状模型公式为:

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ C_0 + C\left(\frac{3h}{2a} - \frac{h^3}{2a^3}\right) & h \in (0, a] \\ C_0 + C & h > a \end{cases}$$

球状模型、指数模型、高斯模型所对应的图形如图2所示。

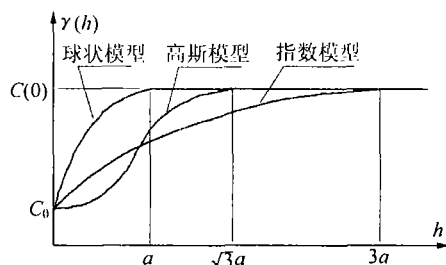


图2 变异函数中的球状模型、指数模型、高斯模型

2.4 变异函数对空间变异性结构的反映

变异函数作为定量描述空间变异性的一种统计学工具,通过其自身的结构及其各项参数,从不同角度反映了空间变异性结构。利用变异函数可以对空间变量的连续性、相关性、变量的影响范围、尺度效应、原点处的间断性、各向异性等要素进行描述^[2]。

2.5 变异函数理论模型参数拟合

变异函数理论模型参数拟合就是利用原始采样点数据或实验变异函数取值对所选定的理论模型参数以特定的方法进行估计。拟合方法一般采用手工拟合法。

手工拟合就是依据实验变异函数的取值,一方面通过观察实验变异函数图;另一方面对所研究的区域化变量进行必要的分析,采用肉眼观察来确定变异函数模型参数,并对参数反复进行交叉验证,最终确定模型参数。其拟合的大致过程如下:

(1)首先对所研究的区域化变量进行必要的结构、背景等方面的分析,结合专家经验,确定变异函数理论模型。

(2)利用实验变异函数散点图确定变异函数参数中的块金常数、基台值、变程、各向异性角度以及各向异性比值。

(3)交叉验证。

3 克里金插值法和距离反比加权法比较

距离反比加权法的表达式也可以表示成与克里金插值法类似的形式,即待插点 x_0 邻域中 n 个

已知点 $x_i, i=1, 2, \dots, n$ 的属性值 $Z(x_i), i=1, 2, \dots, n$ 的加权和 $Z(x_0) = \sum_{i=1}^n \lambda_i Z(x_i)$, 且两者的权系数都是归一化的。但是,距离反比加权法的权系数 $\lambda_i, i=1, 2, \dots, n$ 是直接得到的,是 $x_i, i=1, 2, \dots, n$ 与 x_0 之间距离平方的倒数。这种权系数与 $x_i, i=1, 2, \dots, n$ 相互之间没有任何关系,也就是说和 $x_i, i=1, 2, \dots, n$ 的构形没有任何关系。只要 $x_i, i=1, 2, \dots, n$ 和 x_0 之间的距离一定,则对于 $x_i, i=1, 2, \dots, n$ 之间的任何构性,其权系数都是相同的。

而对于克里金插值法的权系数 $\lambda_i, i=1, 2, \dots, n$ 来说,它不仅取决于 $x_i, i=1, 2, \dots, n$ 与 x_0 之间的距离,而且取决于 $x_i, i=1, 2, \dots, n$ 的构性。这种构性集中体现在 $Z(x)$ 的变异函数(协方差函数)中。克里金插值法的权系数 $\lambda_i, i=1, 2, \dots, n$ 是通过求解基于变异函数(协方差函数)的克里金方程组得到的。

但距离反比加权法的效率最高,插值中所需存储空间最小。

在克里金法和距离反比加权法比较的基础上,可以将距离反比加权法改写为下面形式:

$$Z(x_0) = \frac{\sum_{i=1}^n \frac{Z(x_i)}{\gamma(x_i, x_0)}}{\sum_{i=1}^n \frac{1}{\gamma(x_i, x_0)}}$$

$\gamma(x_0, x_i), i=1, 2, \dots, n$ 为待插点 x_0 与已知采样点 $x_i, i=1, 2, \dots, n$ 之间的变异函数值。

这种形式的距离反比加权法具有以下性质:

(1)加权系数仍为距离的函数(因为变异函数是距离的函数)。当 x_0 与其邻域内的某一已知采样点之间的距离越大,对应于该点的权系数越小,但不是按距离平方的倒数方式减小,而是按变异函数的性质来减小。如果变异函数具有基台值,则减小到基台值的倒数后就不再减小,而保持这个常数不变。

(2)权系数具有归一化的性质。

(3)与克里金法相比,该方法不需要对每一个待插点求解一个线性方程组,且效率较高。

(4)与原距离反比加权法相比,该方法对不同的采样数据点构形可以采用不同的变异函数形式以及不同的变异函数参数,因此,具有一定的灵活性。但采用该方法时,需要先确定变异函数形式,拟合变异函数参数,失去了原距离反比加权法效率高的特点。

(5)当选取的变异函数模型为幂指数形式且指数为 2 时,该插值形式就完全对应于距离反比加权法。而对于小于 2 的其它指数,可根据实际数据利用变异函数参数拟合的方法得到。

4 两种插值方法在实际应用中效果比较

采用文献[4]给出的我国某煤矿一矿区煤层数据(将原始数据放大 10 倍),现利用距离加权反比法和普通克里金方法对其插值建立数学模型。煤层的空间采样点分布如图 3 所示。

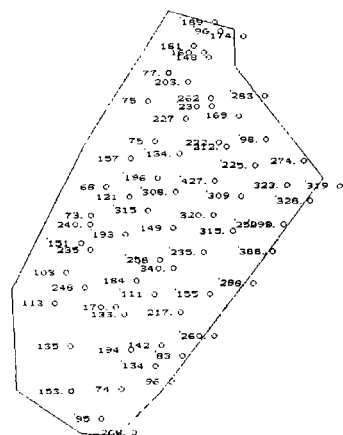


图 3 煤层数据点分布图

普通克里金采用的变异函数模型为球状模型:

$$\gamma(h) = 3\,724.267\,147\,96 + 3\,787.094\,696\,22 \times \left(\frac{3h}{2 \times 118.238\,896\,77} - \frac{h^3}{2 \times 118.238\,896\,77^3} \right)$$

各向异性角度为:118.238 896 77°;各向异性比为:1.350 721 45。

距离反比加权法、普通克里金法对煤层数据建模结果的等值线图如图 4 和图 5 所示。

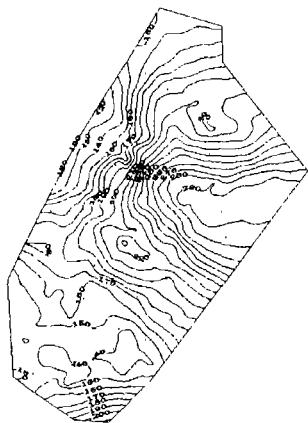


图 4 普通克里金法结果图



图 5 距离反比加权法结果图

交叉验证结果比较。所谓交叉验证就是:假设研究变量为 $Z(x)$,依次删去 $Z(x)$ 在采样点 x_i , $i=1,2,\dots,N$ 处的属性值 $z(x_i)$, $i=1,2,\dots,N$,其它点的属性值保持不变,利用剩下的 $N-1$ 个点的属性值,插值计算被删除的采样点上的属性值 $z^*(x_i)$,并对 N 个插值计算结果与实际的结果作比较,进行误差的统计学分析。用上述煤层数据对这两种方法进行交叉验证比较,其结果见表 1。

表 1 煤层数据对两种方法进行交叉验证比较

方法	均方误差	误差和的平均值	绝对误差和的平均值
距离反比加权法	5 274.794	-1.704 67	57.680 8
普通克里金法	5 183.658	-0.356 85	57.507 9
方法	绝对误差的最大值		绝对误差的最小值
距离反比加权法	190.118 348		0.055 819
普通克里金法	185.428 334		0.713 601

从建模等值线图、交叉验证结果总体比较可以看出,普通克里金法建模效果以及各项指标均优于距离加权反比法的效果和指标。

5 结 论

数据场可视化中插值建立所研究变量的数学模型是至关重要的一步。通过对距离加权反比法和普通克里金法插值的原理分析,在理论和实际应用上的比较说明,普通克里金法的插值结果要优于距离加权反比法的插值结果。

参考文献:

- [1] 唐泽圣. 清华大学学术专著——三维数据场可视化[M]. 北京:清华大学出版社,1999.

- [2] 王家华, 高海余, 周叶. 克里金地质绘图技术——计算机的模型和算法[M]. 北京: 石油工业出版社, 1999. methods[J]. *Mathematical Computation*, 1982, 38: 181-200.
- [3] Franke R. Scattered data interpolation: test of some [4] 孙洪泉. 地质统计学及其应用[M]. 北京: 中国矿业大学出版社, 1990.

Comparison between Inverse Distance Weighting Method and Kriging

JIN Guo-dong, LIU Yan-cong, NIU Wen-jie

(Petroleum University of China, Shandong Province, Dongying 257061, China)

Abstract: Modeling by interpolation is an important step in data visualization. The principles of Kriging and the inverse distance weighting method are described. A comparison of these and two methods both theoretically and in practical application shows that the Kriging method gives out better results.

Keywords: data visualization; inverse distance weighting method; Kriging; variogram.