克立格自动网格化方法

Automatic Gridding: Format Kriging Method

李海滨[1] 杨 宏[1] 肖克炎[2] 陈郑辉[2]

（1 吉林大学，吉林 长春 130026； 2 中国地质科学院矿产资源研究所，北京 100037）

Li Haibin[1], Yang Hong[1], Xiao Keyan[2] and Chen Zhenghui[2]

(1 Jilin University, Changchun 130026, Jilin, China; 2 Institute of Mineral Resources, Chinese Academy of Geological Sciences, Beijing 100037, China)

摘 要 本文介绍了普通克立格方法和变差函数的理论基础，在此基础上编程实现了克立格方法的自动网格化，成为矿产资源评价系统（MRAS）的组成部分。

关键词 数据网格化 普通克立格 变差函数 MRAS

数据网格化是把离散的不规则数据通过某种网格化的形式转变成为规则的网状数据，从而便于我们进一步研究。常见的网格化方法有快速离散网格化方法、权系数为 1 方法、距离平方反比法、趋势面拟合法、克立格方法等，其中克立格方法网格化较为复杂，它需要几个过程协作完成。首先对选取的区域化变量做预处理，然后计算实验变差函数，对得到的有效数据选用适当的变差函数理论模型（如球状模型、高斯模型、指数模型等）进行处理。过去都是给出该模型的一些参数（如变程值、块金常数和基台值），把得到的理论曲线和实际曲线相比较，通过反复人工拟合，直到认为所得曲线满意为止，最后用克立格方法估值，得到的克立格方差就是网格化结果。这样操作起来必须要分步进行，比较烦琐，且人为因素起主要作用，没有一个客观标准。吉林大学矫希国教授解决了变差函数参数的自动化计算问题，据此笔者将各个过程连为一个整体，中间过程由程序自动计算，从而实现了克立格方法的自动网格化，计算过程对用户来说是透明的，操作简单明了。

1 理论基础

1.1 变差函数（Variogram）和普通克立格法〔Ordinary Kriging（OK）〕

变差函数，它反映分析区域化变量的结构性质，又被称作“结构函数”（夏立显等，1999）。通过对选取的区域化变量进行结构分析构造变差函数的理论模型，用以表示变量的主要结构特征，进而进行克立格估计。变差函数是在任一方向 a，相距| h|的两个区域化变量值 Z（x）和 Z（$x+h$）的增量的方差，通式为

$$\gamma_v(x,h)=\frac{1}{2}Var[Z(x)-Z(x+h)]=\frac{1}{2}E[Z(x)-Z(x+h)]^2-\frac{1}{2}\{E[Z(x)-Z(x+h)]\}^2$$

在二阶平稳假设条件下，设区域化变量 $Z(x)$ 定义在区域 G 内：

（1） 对 $\forall x, x+h \in G$, 都有 $E[Z(x)-Z(x+h)]=0$

（2） 变差函数 $\gamma_v(x,h)$ 于 x 无关，仅为 h 的函数，于是它可以写成

第一作者简介 李海滨，男，1976 年生，硕士生，计算机及应用技术专业。

$$\lambda(h)=\frac{1}{2}E[Z(x)-Z(x+h)]^2$$

在实践中，样品的数目是有限的，由有限样品值构制的变差函数称为实验变差函数，它是理论变差函数 $\lambda(h)$ 的估计值。变差函数一般用变异曲线来表示，为了得到最后的结论，必须配以相应的理论模型，本方法采用了球状模型，它的一般公式为（杨永华等，1992）：

$$\gamma(h)=\begin{cases}0 & h=0\\ c_0+c(\frac{3}{2}\frac{h}{a}-\frac{1}{2}\frac{h^3}{a^3}) & 0<h\le a\\ c_0+c & h>a\end{cases}$$

公式中a称为变程，表示样品的最大影响范围；c称为基台值，当 h 超出变程a以后，$\gamma(h)$就稳定在一个极限值$\gamma(\infty)$附近，$\gamma(\infty)$称为基台值；c_0称为块金值，表示由观测误差和区域化变量随机变化引起的在很小距离内两个样品间的差异，即当$h\to 0$ 时，$\lim\gamma(h)$等于一个常数c_0。

克立格法，也称克立金（Kriging），它的基本思想来源于南非矿山地质工程师 D.K.Krigde 的一种特殊的空间插值技术，是地质统计学中重要的组成部分。克立格法所研究的是多个随机变量的线性依赖关系，这些随机变量可以代表一个属性，也可以代表多个相关的属性，把随机变量看成是空间坐标的函数，而各随机变量间的关系由变差函数或协方差函数来刻划；同时克立格法还考虑了随机变量的结构性特征，因而能给出更合乎实际的估计（夏立显等，1999）。本方法采用的是普通克立格法，下面给出其数学模型：

假设有 $n+1$ 个随机变量都来自代表某一共同属性的区域化变量Z(x)，Z(x)二阶平稳，即Z(x)=m（m为常数）。根据 n 个随机变量（自变量）$Z_a=Z(x_a)$（$a=1,2,\cdots,n$），求另一个随机变量（因变量）$Z_0=Z(x_0)$的方差最小的线性无偏估计。

$$Z_0^*=\lambda_0+\sum_{\alpha=1}^{n}\lambda_a Z_a \qquad (1\text{-}1)$$

为保证（1-1）式给出的估计具有无偏性，估计系数 λ 应满足方程

$$m=EZ_0=\lambda_0+\sum_{\alpha=1}^{n}\lambda_a m$$

只须满足条件　$\lambda_0=0$ 及 $\sum_{\alpha=1}^{n}\lambda_a=1$ 或 $\lambda' l=1$　　（1-2）

其中 $l=(1,1,\cdots,1)'$ 是分量全为 1 的 n 维列分量。

估计方差的计算公式（1-3）：

$$\sigma_E^2=Var(Z_0-Z_0^*)=VarZ_0-VarZ_0^*-2Cov(Z_0-Z_0^*)=c(0)+\lambda' C\lambda-2\lambda' Cov(Z_0Z)$$

$$=c(0)+\lambda' C\lambda-2\lambda\begin{bmatrix}Cov(Z_0,Z_1)\\ \vdots\\ Cov(Z_0,Z_n)\end{bmatrix}=c(0)+\lambda' C\lambda-2\lambda' c$$

为求条件（1-1）下（1-3）的极小值，用 Lagrange 不定乘数法，目标函数为

$$L(\lambda,\theta)=c(0)+\lambda' C\lambda+2\theta(\lambda' l-1)$$

方程两边分别对 λ，θ 求导，组成克立格方程组

$$\begin{cases}C\lambda+l\theta=c\\ l\lambda'=1\end{cases} \qquad (1\text{-}4)$$

当C满秩时，该方程组的系数矩阵也满秩，方程组由唯一解，将（1-4）的解代入（1-3）中，得到最小克立格方差

$$\sigma_{OK}^2 = c(0) + \lambda' c - \lambda' l\theta - 2\lambda' c = c(0) - \lambda' c - \lambda' l\theta = c(0) - \lambda' c - \theta$$

1.2 实验变差函数曲线的拟合方法

文章第一段已经讲过，人工输入参数拟合费时费力又不科学；用加权多项式回归法也可以拟合实验变差函数曲线，但该方法不能保证解的正负问题，而负值是没有意义的；本文采用的是根据线形方程组非负解的理论设计的变差函数参数求解的一种最优化方法（矫希国，1997），它成功解决了实验变差函数曲线的自动拟合问题，用该方法得到的解也就是变差函数参数的最佳值。

基于以上原理和方法模型，克立格自动网格化方法的原理是：对一组选取的区域化变量，给定变差函数计算范围（区域x方向、y方向长度和的 1/10）、方向（0 度方向）和步长（区域x方向、y方向长度和的 1/150），计算出 0 度方向上符合条件的样品对，接着用球状模型拟合该实验变差函数，通过上述最优化方法求解线性方程组得到参数c_0，c，a的最优值，最后根据用户设定计算块段的坐标参数，用普通克立格方法求得克立格方差，以网格化形式显示计算结果，完成网格化。

2 应用实例

本方法应用于矿产资源评价系统中，作为网格化方法的一种，利用它对东昆仑实际观测数据进行网格化处理。共有 7099 个样本值，高程范围由-525～-360，观测范围（-729.852，-0.133）～(912.397，626.045)，单位坐标为毫米，如图 1、2、3。

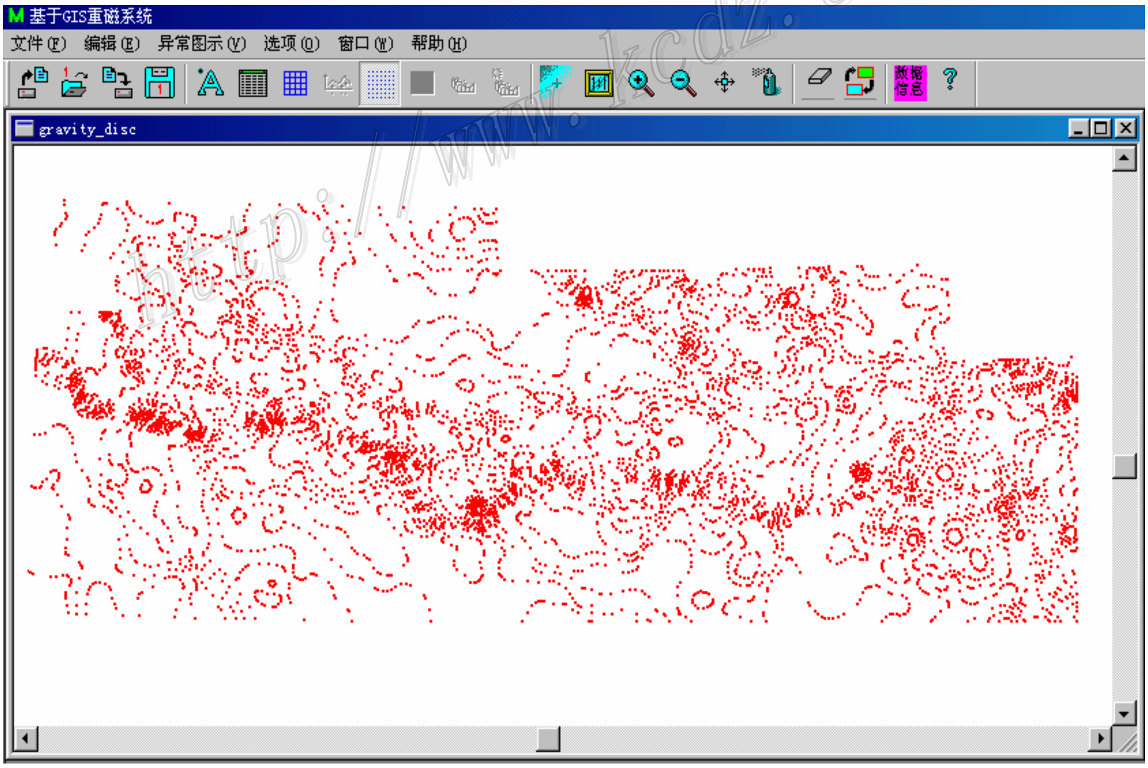

图 1 东昆仑数据平面离散点图

图 2　网格化参数设置图

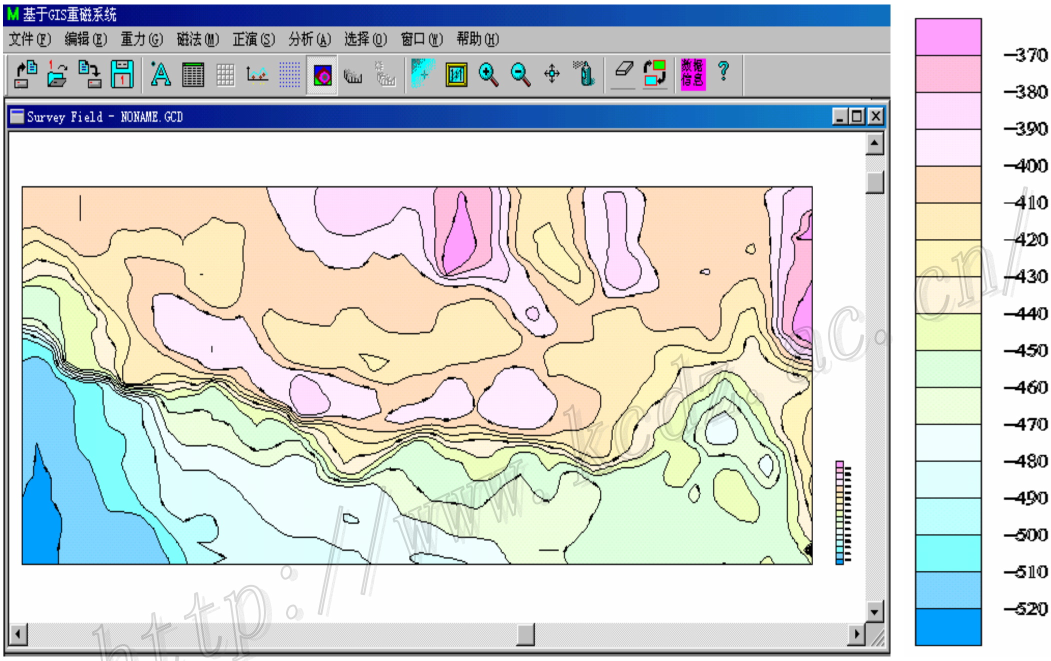

图 3　网格数据等值线结果图

3　结束语

克立格自动网格化方法在“矿产资源评价系统”中具体应用，并通过处理实际数据得到了检验，实现了用克立格方法进行数据网格化的中间过程的自动化计算；方法实施过程中的一些参数被固定，对所有数据适应性问题需进一步探讨。

参　考　文　献

矫希国. 1997. 变差函数参数的计算. 地质评论. 5

夏立显, 韩燕, 杨毅恒. 1999. 多种克立格方法的矩阵表示. 长春: 长春科技大学.

杨永华, 曹绪言, 肖克炎, 等. 1992. 数学地质方法. 东北师范大学出版社.