克里金(Kriging)插值的原理与公式推导

由 xg1990 / 2014年10月15日 / 94 评论

学过空间插值的人都知道克里金插值，但是它的变种繁多、公式复杂，还有个半方差函数让人不知所云

本文讲简单介绍基本克里金插值的原理，及其推理过程，全文分为九个部分：

0.引言 - 从反距离插值说起
1.克里金插值的定义
2.假设条件
3.无偏约束条件
4.优化目标 / 代价函数
5.代价函数的最优解
6.半方差函数
7.普通克里金与简单克里金
8.小结

0.引言——从反距离插值(IDW)说起

空间插值问题，就是在已知空间上若干离散点 (x_i, y_i) 的某一属性(如气温，海拔)的观测值 $z_i = z(x_i, y_i)$的条件下，估计空间上任意一点(x, y)的属性值的问题。

直观来讲，根据地理学第一定律，

All attribute values on a geographic surface are related to each other, but closer values are more strongly related than are more distant ones.

大意就是，地理属性有空间相关性，相近的事物会更相似。由此人们发明了反距离插值，对于空间上任意一点(x, y)的属性$z = z(x, y)$，定义反距离插值公式估计量

$$\hat{z}=\sum_{i=1}^{n}\frac{1}{d^{\alpha}}z_i$$

其中α通常取1或者2。

即，用空间上所有已知点的数据加权求和来估计未知点的值，权重取决于距离的倒数（或者倒数的平方）。那么，距离近的点，权重就大；距离远的点，权重就小。

反距离插值可以有效的基于地理学第一定律估计属性值空间分布，但仍然存在很多问题：

- α的值不确定
- 用倒数函数来描述空间关联程度不够准确

因此更加准确的克里金插值方法被提出来了

1.克里金插值的定义

相比反距离插值，克里金插值公式更加抽象

$$\hat{z_o}=\sum_{i=1}^{n}\lambda_i z_i$$

其中$\hat{z_o}$是点(x_o,y_o)处的估计值，即$z_o=z(x_o,y_o)$ 。

这里的λ_i是权重系数。它同样是用空间上所有已知点的数据加权求和来估计未知点的值。但权重系数并非距离的倒数，而是能够满足点(x_o,y_o)处的估计值$\hat{z_o}$与真实值z_o的差最小的一套最优系数，即

$$\min_{\lambda_i} Var(\hat{z_o}-z_o)$$

同时满足无偏估计的条件

$$E(\hat{z_o}-z_o)=0$$

2.假设条件

不同的克里金插值方法的主要差异就是假设条件不同。本文仅介绍普通克里金插值的假设条件与应用。

普通克里金插值的假设条件为，空间属性z是均一的。对于空间任意一点(x,y)，都有同样的期望c与方差σ^2。

即对任意点(x,y)都有

$$E[z(x,y)] = E[z] = c$$

$$Var[z(x,y)] = \sigma^2$$

换一种说法：任意一点处的值$z(x,y)$，都由区域平均值c和该点的随机偏差$R(x,y)$组成，即

$$z(x,y) = E[z(x,y)] + R(x,y)] = c + R(x,y)$$

其中$R(x,y)$表示点(x,y)处的偏差，其方差均为常数

$$Var[R(x,y)] = \sigma^2$$

3.无偏约束条件

先分析无偏估计条件$E(\hat{z_o} - z_o) = 0$，将$\hat{z_o} = \sum_{i=1}^{n} \lambda_i z_i$带入则有

$$E(\sum_{i=1}^{n} \lambda_i z_i - z_o) = 0$$

又因为对任意的z都有$E[z] = c$，则

$$c\sum_{i=1}^{n} \lambda_i - c = 0$$

即

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

这是λ_i的约束条件之一。

4.优化目标/代价函数J

再分析估计误差$Var(\hat{z_o} - z_o)$。为方便公式推理，用符号J表示，即

$$J = Var(\hat{z_o} - z_o)$$

则有

$$\begin{aligned} J &= Var(\textstyle\sum_{i=1}^{n} \lambda_i z_i – z_o) \\ &= Var(\textstyle\sum_{i=1}^{n} \lambda_i z_i) – 2Cov(\textstyle\sum_{i=1}^{n} \lambda_i z_i, z_o) + Cov(z_o, z_o) \\ &= \textstyle\sum_{i=1}^{n} \sum_{j=0}^{n} \lambda_i \lambda_j Cov(z_i, z_j) – 2\sum_{i=1}^{n} \lambda_i Cov(z_i, z_o) + Cov(z_o, z_o) \end{aligned}$$

为简化描述，定义符号 $C_{ij}=Cov(z_i,z_j)=Cov(R_i,R_j)$，这里$R_i=z_i – c$，即点$(x_i,y_i)$处的属性值相对于区域平均属性值的偏差。

则有

$$J=\sum_{i=1}^{n}\sum_{j}^{n}\lambda_i\lambda_jC_{ij} – 2\sum_{i=1}^{n}\lambda_iC_{io}+C_{oo}$$

5.代价函数的最优解

再定义半方差函数 $r_{ij}=\sigma^2-C_{ij}$，带入J中，有

$$\begin{aligned} J &= \sum_{i=1}^{n}\sum_{j=0}^{n}\lambda_i\lambda_j(\sigma^2 – r_{ij}) – 2\sum_{i=1}^{n}\lambda_i(\sigma^2 – r_{io})+\sigma^2 – r_{oo} \\ &= \sum_{i=1}^{n}\sum_{j=0}^{n}\lambda_i\lambda_j(\sigma^2)-\sum_{i=1}^{n}\sum_{j=0}^{n}\lambda_i\lambda_j(r_{ij})-2\sum_{i=1}^{n}\lambda_i(\sigma^2)+2\sum_{i=1}^{n}\lambda_i(r_{io})+\sigma^2 – r_{o} \end{aligned}$$

考虑到$\sum_{i=1}^{n}\lambda_i=1$

$$\begin{aligned} J &= \sigma^2-\sum_{i=1}^{n}\sum_{j}^{n}\lambda_i\lambda_j(r_{ij})-2\sigma^2+2\sum_{i=1}^{n}\lambda_i(r_{io})+\sigma^2 – r_{oo} \\ &= 2\sum_{i=1}^{n}\lambda_i(r_{io})-\sum_{i=1}^{n}\sum_{j=0}^{n}\lambda_i\lambda_j(r_{ij}) – r_{oo} \end{aligned}$$

我们的目标是寻找使J最小的一组 λ_i，且J是λ_i的函数，因此直接将J对λ_i求偏导数令其为0即可。即

$$\frac{\partial J}{\partial \lambda_i}=0;i=1,2,\cdots,n$$

但是要注意的是，我们要保证求解出来的最优 λ_i 满足公式$\sum_{i=1}^{n}\lambda_i=1$，这是一个带约束条件的最优化问题。使用拉格朗日乘数法求解，求解方法为构造一个新的目标函数

$$J+2\phi(\sum_{i=1}^{n}\lambda_i-1)$$

其中ϕ是拉格朗日乘数。求解使这个代价函数最小的参数集$\phi,\lambda_1,\lambda_2,\cdots,\lambda_n$，则能满足其在$\sum_{i=1}^{n}\lambda_i=1$约束下最小化$J$。即

$$\begin{cases} \frac{\partial(J+2\phi(\sum_{i=1}^{n}\lambda_i-1))}{\partial\lambda_k} &= 0;k=1,2,\cdots,n \\ \frac{\partial(J+2\phi(\sum_{i=1}^{n}\lambda_i-1))}{\partial\phi} &= 0 \end{cases}$$

$$\begin{cases} \frac{\partial(2\sum_{i=1}^{n}\lambda_i(r_{io}) – \sum_{i=1}^{n}\sum_{j}^{n}\lambda_i\lambda_j(r_{ij}) – r_{oo}+2\phi(\sum_{i=1}^{n}\lambda_i-1))}{\partial\lambda_k} &= 0;k=1,2,\cdots,n \\ \frac{\partial(2\sum_{i=1}^{n}\lambda_i(r_{io}) – \sum_{i=1}^{n}\sum_{j}^{n}\lambda_i\lambda_j(r_{ij}) – r_{oo}+2\phi(\sum_{i=1}^{n}\lambda_i-1))}{\partial\phi} &= 0 \end{cases}$$

$$\begin{cases} 2r_{ko} – \sum_{j=1}^{n}(r_{kj}+r_{jk})\lambda_j+2\phi &= 0;k=1,2,\cdots,n \\ \sum_{i=1}^{n}\lambda_i &= 1 \end{cases}$$

由于$C_{ij}=Cov(z_i,z_j)=C_{ji}$，因此同样地$r_{ij}=r_{ji}$，那么有

$$\begin{cases} r_{ko} - \sum_{j=1}^{n} r_{kj}\lambda_j + \phi & = 0; k=1,2,\cdots,n \\ \sum_{i=1}^{n} \lambda_i & = 1 \end{cases}$$

式子中半方差函数r_{ij}十分重要，最后会详细解释其计算与定义

在以上计算中我们得到了对于求解权重系数λ_j的方程组。写成线性方程组的形式就是：

$$\begin{cases} r_{11}\lambda_1 + r_{12}\lambda_2 + \cdots + r_{1n}\lambda_n - \phi & = r_{1o} \\ r_{21}\lambda_1 + r_{22}\lambda_2 + \cdots + r_{2n}\lambda_n - \phi & = r_{2o} \\ & \cdots \\ r_{n1}\lambda_1 + r_{n2}\lambda_2 + \cdots + r_{nn}\lambda_n - \phi & = r_{no} \\ \lambda_1 + \lambda_2 + \cdots + \lambda_n & = 1 \end{cases}$$

写成矩阵形式即为

$$\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} & 1 \\ r_{21} & r_{22} & \cdots & r_{2n} & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \cdots \\ \lambda_n \\ -\phi \end{bmatrix} = \begin{bmatrix} r_{1o} \\ r_{2o} \\ \cdots \\ r_{no} \\ 1 \end{bmatrix}$$

对矩阵求逆即可求解。

唯一未知的就是上文中定义的半方差函数r_{ij}，接下来将详细讨论

6.半方差函数

上文中对半方差函数的定义为

$$r_{ij} = \sigma^2 - C_{ij}$$

其等价形式为

$$r_{ij} = \frac{1}{2}E[(z_i - z_j)^2]$$

这也是半方差函数名称的来由，接下来证明这二者是等价的：

根据上文定义 $R_i = z_i - c$，有$z_i - z_j = R_i - R_j$，则

$$
\begin{aligned}
r_{ij} &= \frac{1}{2}E[(R_i - R_j)^2] \\
&= \frac{1}{2}E[R_i^2 - 2R_iR_j + R_j^2] \\
&= \frac{1}{2}E[R_i^2] + \frac{1}{2}E[R_j^2] - E[R_iR_j]
\end{aligned}
$$

又因为:

$$E[R_i^2] = E[R_j^2] = E[(z_i – c)^2] = Var(z_i) = \sigma^2$$

$$E[R_iR_j] = E[(z_i – c)(z_j - c)] = Cov(z_i, z_j) = C_{ij}$$

于是有

$$
\begin{aligned}
r_{ij} &= \frac{1}{2}E[(z_i - z_j)^2] \\
&= \frac{1}{2}E[R_i^2] + \frac{1}{2}E[R_j^2] - E[R_iR_j] \\
&= \frac{1}{2}\sigma^2 + \frac{1}{2}\sigma^2 - C_{ij} \\
&= \sigma^2 - C_{ij}
\end{aligned}
$$

$\sigma^2 - C_{ij} = \frac{1}{2}E[(z_i - z_j)^2]$得证，现在的问题就是如何计算

$$r_{ij} = \frac{1}{2}E[(z_i - z_j)^2]$$

这时需要用到地理学第一定律，空间上相近的属性相近。$r_{ij} = \frac{1}{2}(z_i - z_j)^2$表达了属性的相似度；空间的相似度就用距离来表达，定义i与j之间的几何距离

$$d_{ij} = d(z_i, z_j) = d((x_i, y_i), (x_j, y_j)) = \sqrt{(x_i - x_j)^2 + (y_i – y_j)^2}$$

克里金插值假设r_{ij}与d_{ij}存在着函数关系，这种函数关系可以是线性、二次函数、指数、对数关系。为了确认这种关系，我们需要首先对观测数据集

$$\{z(x_1, y_1), z(x_2, y_2), z(x_3, y_3), \cdots, z(x_{n-1}, y_{n-1}), z(x_n, y_n)\}$$

计算任意两个点的 距离$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i – y_j)^2}$和 半方差 $\sigma^2 - C_{ij} = \frac{1}{2}E[(z_i - z_j)^2]$，这时会得到$n^2$个$(d_{ij}, r_{ij})$的数据对。

将所有的d和r绘制成散点图，寻找一个最优的拟合曲线拟合d与r的关系，得到函数关系式

$$r = r(d)$$

那么对于任意两点Math input error，先计算其距离d_{ij}，然后根据得到的函数关系就可以得到这两点的半方差r_{ij}

7. 简单克里金（simple kriging）与普通克里金（ordinary kriging）的区别

以上介绍的均为普通克里金（ordinary kriging）的公式与推理。

事实上普通克里金插值还有简化版，即简单克里金（simple kriging）插值。二者的差异就在于如何定义插值形式：

上文讲到，普通克里金插值形式为

$$\hat{z_o} = \sum_{i=1}^{n} \lambda_i z_i$$

而简单克里金的形式则为

$$\hat{z_o} – c = \sum_{i=1}^{n} \lambda_i (z_i - c)$$

这里的符号c在上文介绍过了，是属性值的数学期望，即$E[z] = c$。也就是说，在普通克里金插值中，认为未知点的属性值是已知点的属性值的加权求和；而在简单克里金插值中，假设未知点的属性值相对于平均值的偏差是已知点的属性值相对于平均值的偏差的加权求和，用公式表达即为：

$$\hat{R_o} = \sum_{i=1}^{n} \lambda_i R_i$$

这里的R_i在上文定义过了：$R_i = z_i – c$。

但是为什么这样的克里金插值称为"简单克里金"呢？由于有假设$E[z] = c$，也就是说$E(R_i + c) = c$，即$E(R_i) = 0$。那么上面的公式$\hat{R_o} = \sum_{i=1}^{n} \lambda_i R_i$两边的期望一定相同，那么在求解未知参数$\lambda_i$就不需要有无偏约束条件$\sum_{i=1}^{n} \lambda_i = 1$。换句话说，这样的估计公式天生就能满足无偏条件。因此它被称为简单克里金。

从在上文（第4节优化目标/代价函数J）中可以知道，优化目标的推理和求解过程是通过对属性值相对于期望的偏差量R_i进行数学计算而进行的。也就是说这两种克里金插值方法虽然插值形式不一样，求解方法是一样的，重要的区别是简单克里金插值不需要约束条件$\sum_{i=1}^{n} \lambda_i = 1$，求解方程组为：

$$\begin{cases} r_{11}\lambda_1 + r_{12}\lambda_2 + \cdots + r_{1n}\lambda_n + \phi & = r_{1o} \\ r_{21}\lambda_1 + r_{22}\lambda_2 + \cdots + r_{2n}\lambda_n + \phi & = r_{2o} \\ & \cdots \\ r_{n1}\lambda_1 + r_{n2}\lambda_2 + \cdots + r_{nn}\lambda_n + \phi & = r_{no} \end{cases}$$

还有更重要的一点，简单克里金的插值公式为：

$$\hat{z_o} = \sum_{i=1}^{n} \lambda_i (z_i - c) + c$$

换句话说，在计算未知点属性值$\hat{z_o}$前，需要知道该地区的属性值期望c。事实上我们在进行插值前很难知道这个地区的真实属性值期望。有些研究者可能会采用对观测数据简单求平均的方法计算期望值c，而考虑到空间采样点位置代表性可能有偏差（比如采样点聚集在某一小片地区，没有代表性），简单平均估计的期望也可能是有偏差的。这是简单克里金方法的局限性。

8.小结

总的来说，进行克里金插值分为这几个步骤：

1. 对于观测数据，两两计算距离与半方差
2. 寻找一个拟合曲线（或者其他模型）模拟距离与半方差的关系，从而能根据任意距离计算出相应的半方差
3. 计算出所有已知点之间的半方差r_{ij}，直接使用公式$r_{ij} = \frac{1}{2}(z_i - z_j)^2$
4. 对于未知点z_o，计算它到所有已知点z_i的距离d_{io}，并通过第2步的拟合曲线，估计半方差r_{io}
5. 求解第四节中的方程组，得到最优系数λ_i
6. 使用最优系数对已知点的属性值进行加权求和，得到未知点z_o的估计值，即为$\hat{z_o} = \sum_{i=1}^{n} \lambda_i z_i$

相关文章:

1. **加权线性回归(Weighted Linear Regression)的公式及其推理**
2. **用WRF模型进行气象模拟入门(1)——简介以及代码编译**
3. **用WRF模型进行气象模拟入门(2)——WPS的配置与使用**

标签: GIS 插值 统计

《克里金(Kriging)插值的原理与公式推导》有94个想法

YANGBAOAN
2015年3月14日 18:26