

引用格式:车磊,王海起,费涛,等.基于多尺度最小二乘支持向量机优化的克里金插值方法[J].地球信息科学学报,2017,19(8):1001-1010. [Che L, Wang H Q, Fei T, et al. 2017. Kriging interpolation method optimized by multi-scale least squares support vector machine. Journal of Geo-information Science, 19(8):1001-1010.] DOI:10.3724/SP.J.1047.2017.01001

基于多尺度最小二乘支持向量机优化的克里金插值方法

车磊^{1,2},王海起^{1*},费涛¹,闫滨¹,刘玉¹,桂丽¹,陈冉¹,翟文龙¹

1. 中国石油大学(华东)地球科学与技术学院,青岛 266580; 2. 中国电波传播研究所青岛分所,青岛 266107

Kriging Interpolation Method Optimized by Multi-scale Least Squares Support Vector Machine

CHE Lei^{1,2}, WANG Haiqi^{1*}, FEI Tao¹, YAN Bin¹, LIU Yu¹, GUI Li¹, CHEN Ran¹ and ZHAI Wenlong¹

1. School Of Geosciences, China University of Petroleum (East China), Qingdao 266580, China; 2. China Research Institute of Radiowave Propagation Qingdao Branch, Qingdao 266107, China

Abstract: Kriging interpolation method realizes spatial weighted estimation that meets the unbiasedness and optimality according to the position relationship between the estimated location sites and the known sample sites and regionalized variable spatial correlation. Traditional theoretical model shape is fixed and chosen with subjectivity, which can't reflect the changing trend and multi-scale spatial characteristics. The choice of scale and the treatment of scale effects also need to be considered. To solve the problems above, we propose a method of kriging interpolation optimized by multi-scale least squares support vector machine (LS-SVM), which provides a new idea for fitting experimental variogram. Starting from the changing trend of the actual sample data, least squares support vector machine fits experimental variogram and the results conform to the spatial changing trend of data itself. Secondly, the wavelet kernel as the LS-SVM kernel function, parameters can be adjusted according to different parts of the nuclear, which is flexible and variable. Finally, the multi-scale wavelet kernel using wavelet multi-resolution characteristics, can reflect the different details of spatial changes, to avoid the single scale LS-SVM ignoring the spatial details of the problem. Followed that, the experiment includes simulation and application. Experimental simulation mainly verifies scientific validity and accuracy by the optimized interpolation of multi-scale least squares support vector machine. Meanwhile, experimental application research of PM_{2.5} concentrations of temporal and spatial distribution provides the theoretical basis for city ecological protection and controlling. Final results show that kriging interpolation algorithm optimized by multi-scale least squares support vector machine is superior to the traditional method and single scale optimized kriging interpolation algorithm. It would be better to depict the variation function and reflect the different scales of spatial changes in details to further improve the accuracy of the interpolation to some extent, which is an optional kriging interpolation method.

Key words: kriging interpolation; least squares support vector machine; variogram; multi-scale; wavelet kernel function

***Corresponding author:** WANG Haiqi, E-mail: wanghaiqi@upc.edu.cn

收稿日期 2016-10-31;修回日期:2017-06-20.

基金项目 国家自然科学基金项目(41471322);山东省自然科学基金项目(ZR2012DM010)。

作者简介 车磊(1990-),男,硕士生,研究方向为地图制图学与地理信息工程。E-mail: cheleiyouxiang@163.com

*通信作者 王海起(1972-),男,博士,副教授,研究方向为地理信息科学。E-mail: wanghaiqi@upc.edu.cn

摘要 克里金插值方法根据待估位置点、已知样本数据点的位置关系和区域化变量的空间相关性,实现空间加权估计,满足估计的无偏性和最优性。传统方法理论模型形状固定且选择具有人为主观性,无法反映空间数据的变化趋势及其空间多尺度特征。本文为解决上述问题,提出了一种基于多尺度最小二乘支持向量机优化的克里金插值方法,此方法为拟合实验变异函数提供了一种新的思路。从实际样本数据的变化趋势出发,采用最小二乘支持向量机拟合实验变异函数,并利用不同尺度小波核反映不同尺度下的空间变化。最后,实验环节包括模拟和应用,模拟主要验证经多尺度最小二乘支持向量机优化后插值方法的科学有效性以及准确性,应用主要研究青岛市 $\text{PM}_{2.5}$ 浓度时空分布特征,为城市生态科学防护及控制提供理论依据。结果表明,基于多尺度最小二乘支持向量机优化的克里金插值方法能够更好地刻画变异函数,反映不同尺度下的空间变化细节,从而在一定程度上提高插值的精度,是一种可选的克里金插值方法。

关键词 克里金插值;最小二乘支持向量机;变异函数;多尺度;小波核函数

1 引言

空间插值(Spatial interpolation)是根据已知样本点的属性信息进行插值或推理来生成面状数据或估计待估点的数值^[1-2],其本质就是通过建模来拟合生成尽可能逼近要素空间分布特征的函数关系^[3]。空间插值方法种类广泛,常用的包括泰森多边形法、反距离加权法、移动拟合法、线性内插法、样条函数法、趋势分析法、克里金插值法等。其中,克里金插值法(Kriging interpolation)又称空间自协方差最佳插值法,这种方法认为空间属性的变化建立于空间位置关系和空间自相关关系的基础上,故定义具有这种变化的属性变量称为区域化变量(Regionalized variable)^[4-5]。区域化变量独有的随机性和结构性使探索空间结构和空间变异规律变得有迹可循,变异函数(Variogram)应运而生,其可以对区域化变量的连续性、相关性、尺度性等进行空间描述。克里金插值方法的实质也是利用区域化变量的已知样本数据点变异函数结构特点,对待估位置点区域化变量的取值进行无偏、最优估计^[6]。

克里金插值过程中,插值模型的精度取决于模型对空间变异性和空间相关性的反映程度^[3]。插值模型需要用理论变异函数拟合实验变异函数,如何选择合适的理论变异模型关系到插值效果的优劣。传统克里金插值方法基于有限空间样本利用现有的理论变异函数模型进行拟合,对于变异函数的选择具有主观性和依赖性,且现有理论模型形状固定,无法反映实际样本的空间变化趋势。其次,空间变化趋势往往具有多尺度的特征,对于尺度的选择和尺度效应的处理也是需要考虑的^[7],仅以传统方法拟合的实验变异函数会忽略空间变化的尺度效应。

对于克里金插值方法中对变异函数的优化,许多学者已进行了探究。传统数学统计方面,部分学者提出极大似然法、线性规划法等方法^[8-9],此类方

法往往会带来这样的问题,即计算量将成倍增加,但优化效果并不明显;人工智能方面,一些专家提出利用粒子群算法、遗传算法等方法^[10-12],这类方法在一定程度上效果良好,但可能陷入局部最优困境,且结果优劣很大程度上取决于优化程度的好坏,受优化算法影响较大。同时,上述2类优化方法均未更深层次地考虑空间多尺度特性,从而忽略了空间尺度对空间细节的反映。

为解决上述问题,本文提出一种基于多尺度最小二乘支持向量机(以下简称多尺度LS-SVM)优化的克里金插值方法,此方法为拟合实验变异函数提供了一种新的思路。从已知样本数据点的空间变化趋势出发,通过多尺度LS-SVM拟合实验变异函数。经实验验证,此方法能更好地刻画变异函数,反映不同尺度下的空间变化细节,从而在一定程度上提高插值的精度,是一种可选的克里金插值方法。此方法有以下特点:首先,通过LS-SVM拟合实验变异函数,拟合的结果符合数据本身的空间变化趋势;其次,小波核作为LS-SVM核函数,能根据不同部分调整核参数,灵活多样;最后,多尺度小波核利用小波多分辨率的特性,可以体现空间变化的不同细节,避免了单尺度LS-SVM忽略空间细节的问题。

2 相关理论原理

2.1 普通克里金插值方法

变异函数,又称为变程方差函数,是地质统计学分析所特有的基本工具。它既能描述区域化变量的空间结构性变化,又能描述其随机性变化。

区域化变量 $z(x)$ 在位置点 x 处和位置点 $x+h$ 处值之差的方差一半定义为 $z(x)$ 在 x 轴方向上的变异函数。离散样本的实验变异函数可以通过式(1)计算。

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i+h)]^2 \quad i=1, \dots, N(h) \quad (1)$$

式中: $i+1, \dots, N(h)$; h 为样本点对的距离; $N(h)$ 代表样本点对距离为 h 所有样本点对的个数; $z(x_i)$ 和 $z(x_i+h)$ 分别是区域化变量 $z(x)$ 在空间位置 x_i 和 x_i+h 处的真实值。当 $\gamma(h)$ 不依赖于它的方向变化, 只依赖于距离 h 时, 则称 $\gamma(h)$ 各向同性。

为得到任意距离下的变异函数以进行后续的插值, 需要拟合实验变异函数, 数据量较大时一般需在拟合之前进行分组操作。传统理论拟合模型包括球状模型、指数模型、高斯模型、线性模型等, 以下将对普通克里金插值方法的原理进行说明。

克里金插值方法估计需要同时满足无偏性和估计方差最小:

$$E[z(x_0) - z^*(x_0)] = 0 \quad (2)$$

$$\text{Var}[z(x_0) - z^*(x_0)] = \min \quad (3)$$

求解上述方程, 经推导, 用变异函数来表示如下, 即:

$$\begin{cases} \sum_{i=1}^k \lambda_i \gamma(x_i, x_j) + \mu = \gamma(x_0, x_j) & j=1, \dots, k \\ \sum_{i=1}^k \lambda_i = 1 \end{cases} \quad (4)$$

式中: k 为样本观测站点的总个数; λ_i 为克里金权重系数, 表示各空间样本点 x_i 处的区域化变量 $z(x_i)$ 对待估位置点 x_0 的贡献程度; μ 为克里金拉格朗日乘子; $\gamma(x_i, x_j)$ 为空间样本点 x_i 与空间样本点 x_j 距离下的实验变异函数值; $\gamma(x_0, x_j)$ 为待估位置点 x_0 与空间样本点 x_j 距离下的实验变异函数值。

待估位置点 x_0 处的区域化变量估计值 $z^*(x_0)$ 可用式(5)进行估计:

$$z^*(x_0) = \sum_{i=1}^k \lambda_i z(x_i) \quad (5)$$

式中: x_1, \dots, x_k 为已知样本的观测站点; $z(x_1), \dots, z(x_k)$ 为对应的观测值。

普通克里金插值方法在满足无偏性和估计方差最小这2个原则的条件下, 建立了含有约束条件的拉格朗日函数。其中, 无偏性体现在约束条件上, 而方差最小体现在求极值问题上, 拟合后的变异函数计算值直接作用在 λ_i , 导致 λ_i 的权重发生变化, 进而对待估位置点的估值产生影响。

2.2 最小二乘支持向量机

最小二乘支持向量机 (Least Squares Support

Vector Machine, LS-SVM) 用于函数回归时用平方和误差损失函数替代 Vapnik 的 ε -不敏感损失函数^[13], 同时用等式约束替代标准 SVM 中的不等式约束^[14]。

给定 N 个样本的数据集 $\{(x_m, y_m)\}_{m=1}^N$, $x_m \in R^d$, 其中第 m 个输入 x_m 对应的输出为 y_m , 回归函数 $f(x)$ 的基本形式如式(6)所示。

$$f(x) = \omega^T \varphi(x) + b \quad (6)$$

式中: ω 为权系数向量 (列向量); ω^T 代表其转置向量; $\varphi(x)$ 为输入空间到特征空间的映射函数; b 为常数项。

LS-SVM 对应的优化问题为:

$$\min_{\omega, b, e} J(\omega, b, e) = \frac{1}{2} \omega^T \omega + \frac{1}{2} \gamma \sum_{m=1}^N e_m^2 = \frac{1}{2} \|\omega\|^2 + \frac{1}{2} \gamma \sum_{m=1}^N e_m^2 \quad (7)$$

式中: γ 称为正则化参数或惩罚参数; $e_m \in R$ 为误差项。

满足等式约束条件:

$$y_m = \omega^T \varphi(x_m) + b + e_m \quad m=1, 2, \dots, N \quad (8)$$

上述优化问题可由拉格朗日方程求解, 并基于 KKT (Karush-Kuhn-Tucker) 条件得到最终模型 (式(9))。

$$\begin{aligned} f(x) &= \omega^T \varphi(x) + b = \sum_{m=1}^N \alpha_m \langle \varphi(x_m), \varphi(x) \rangle + b \\ &= \sum_{m=1}^N \alpha_m K(x_m, x) + b \end{aligned} \quad (9)$$

式中: α_m 为拉格朗日乘子; $K(x_m, x)$ 为核函数。

2.3 小波函数

小波函数是一种时频局部化的工具, 是把数据、函数或算子分割成不同频率的成分^[15], 然后再用分解的方法研究不同尺度下的成分, 故非常适合做空间多尺度分析。其中, 低频信息作为空间变化的大尺度框架, 而高频信息反映空间变化的细节部分。目前应用较多的实数和复数小波函数中, 以 Morlet 小波、Gaussian 小波、Mexican hat 小波等为主要应用^[16], 具体形式分别如下:

Morlet 小波函数:

$$\varphi(x) = e^{\left(-\frac{x^2}{2}\right)} \cos(\omega_0 x) \quad (10)$$

Gaussian 小波函数:

$$\varphi(x) = x e^{\left(-\frac{x^2}{2}\right)} \quad (11)$$

Mexican hat 小波函数又称为墨西哥帽函数或者 Marr 小波:

$$\varphi(x) = (1 - x^2)e^{\left(-\frac{x^2}{2}\right)} \quad (12)$$

小波函数作为SVM核函数,需满足2个条件: Mercer 和平移不变定理^[17]。以 Morlet 小波函数为例,经推导验证,同时满足这二者,最终得到 Morlet 小波核函数形式,如式(13)所示。

$$K(x_m, x) = \prod_{m=1}^d \left[\cos \left(\omega_0 \times \frac{\|x_m - x\|}{a_m} \right) \exp \left(-\frac{\|x_m - x\|^2}{2a_m^2} \right) \right] \quad (13)$$

3 基于多尺度LS-SVM优化的克里金插值方法

3.1 多尺度LS-SVM优化的克里金插值方法原理

多尺度LS-SVM以二尺度LS-SVM为例^[18-20],具体实现如下:

二尺度LS-SVM模型在回归问题上,输入变量和输出变量分别为距离和实验变异函数值,故假定拟合的数据集为 $\{(h_i, \gamma(h_i))\}_{i=1}^n$, 其中 $h_i \in R^d$, 在这里 $d=1$, 表示第 i 个样本点对的距离, 作为自变量, $\gamma(h_i) \in R$, 表示在距离 h_i 下的实验变异函数值, 作为因变量, n 代表分组后需要拟合的实验变异函数值总个数。

在二尺度问题中,假定尺度1上小波核函数尺度因子偏大,尺度2上小波核函数尺度因子偏小。在较大尺度因子上, $f_1(x)$ 用集合 $\{(h_i, \gamma(h_i))\}_{i=1}^n$ 来回归。回归模型可以用式(14)表示:

$$f_1(h) = \omega_1^T \varphi_1(h) + b_1 \quad (14)$$

在较小尺度因子上, $f_2(x)$ 用集合 $\{(h_i, \gamma(h_i) - f_1(h_i))\}_{i=1}^n$ 来回归,回归模型可用式(15)表示。

$$f_2(h) = \omega_2^T \varphi_2(h) + b_2 \quad (15)$$

将二尺度的模型融合为最终的回归模型,可以表示为:

$$f(h) = \omega_1^T \varphi_1(h) + \omega_2^T \varphi_2(h) + b \quad (16)$$

其中, $b = b_1 + b_2$

需要说明的是, $f_1(x)$ 是对原始样本数据回归, $f_2(x)$ 则是在 $f_1(x)$ 的基础上对残差进行回归,故又可以表示为:

$$f_1(h) = \omega_1^T \varphi_1(h) + b \quad (17)$$

$$f_2(h) = \omega_2^T \varphi_2(h) \quad (18)$$

对应的优化问题如下:

$$\min_{\omega_1, \omega_2, b, e_1, e_2} J(\omega_1, \omega_2, b, e_1, e_2) = \frac{1}{2} \omega_1^T \omega_1 + \frac{1}{2} \omega_2^T \omega_2 + \frac{1}{2} \gamma_1 \sum_{i=1}^N e_{1i}^2 + \frac{1}{2} \gamma_2 \sum_{i=1}^N e_{2i}^2 \quad (19)$$

其相应的约束条件为:

$$y_i = \omega_1^T \varphi_1(h_i) + b + e_{1i} \quad i = 1, 2, \dots, n \quad (20)$$

$$y_i - f_1(h_i) = \omega_2^T \varphi_2(h_i) + e_{2i} \quad i = 1, 2, \dots, n \quad (21)$$

对偶问题的拉格朗日函数为:

$$\begin{aligned} L(\omega_1, \omega_2, b, e_1, e_2, \alpha_1, \alpha_2) = & J(\omega_1, \omega_2, b, e_1, e_2) \\ & - \sum_{i=1}^N \alpha_{1i} (\omega_1^T \varphi_1(h_i) + b + e_{1i} - y_i) \\ & - \sum_{i=1}^N \alpha_{2i} (\omega_2^T \varphi_2(h_i) + e_{2i} + f_1(h_i) - y_i) \end{aligned} \quad (22)$$

根据KKT条件,该拉格朗日函数的最优解条件为:

$$\begin{cases} \frac{\partial L}{\partial \omega_1} = 0 \rightarrow \omega_1 = \sum_{i=1}^N \alpha_{1i} \varphi_1(h_i) + \sum_{i=1}^N \alpha_{2i} \varphi_1(h_i) \\ \frac{\partial L}{\partial \omega_2} = 0 \rightarrow \omega_2 = \sum_{i=1}^N \alpha_{2i} \varphi_2(h_i) \\ \frac{\partial L}{\partial e_{1i}} = 0 \rightarrow \alpha_{1i} = \gamma_1 e_{1i} & i = 1, 2, \dots, n \\ \frac{\partial L}{\partial e_{2i}} = 0 \rightarrow \alpha_{2i} = \gamma_2 e_{2i} & i = 1, 2, \dots, n \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^N \alpha_{1i} + \sum_{i=1}^N \alpha_{2i} = 0 \\ \frac{\partial L}{\partial \alpha_{1i}} = 0 \rightarrow \omega_1^T \varphi_1(h_j) + b + e_{1j} = y_j & j = 1, 2, \dots, n \\ \frac{\partial L}{\partial \alpha_{2i}} = 0 \rightarrow \omega_2^T \varphi_2(h_j) + e_{2j} + \omega_1^T \varphi_1(h_j) + b = y_j & j = 1, 2, \dots, n \end{cases} \quad (23)$$

经拉格朗日方程求解,矩阵形式为:

$$\begin{bmatrix} 0 & \bar{1}^T & \bar{1}^T \\ \bar{1} & \Omega_1 + \frac{I}{\gamma_1} & \Omega_1 \\ \bar{1} & \Omega_1 & \Omega_1 + \Omega_2 + \frac{I}{\gamma_2} \end{bmatrix} \times \begin{bmatrix} b \\ \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ y \\ y \end{bmatrix} \quad (24)$$

式中: $\alpha = [\alpha_{11}, \alpha_{12}, \dots, \alpha_{1n}]^T$, $\beta = [\alpha_{21}, \alpha_{22}, \dots, \alpha_{2n}]^T$, $\bar{1} = [1, 1, \dots, 1]^T$, $y = [y_1, y_2, \dots, y_n]^T$, I 为单位矩阵, Ω_1 和 Ω_2 为小波核函数, $\Omega_1 = K_1(h_p, h_q) = \varphi_1(h_p)^T \varphi_1(h_q)$, $p, q = 1, \dots, n$, $\Omega_2 = K_2(h_p, h_q) = \varphi_2(h_p)^T \varphi_2(h_q)$, $p, q = 1, \dots, n$ 。

最终,二尺度LS-SVM模型表示为式(25)。

$$\begin{aligned}
 f(h) &= \omega_1^T \varphi_1(h) + \omega_2^T \varphi_2(h) + b \\
 &= \sum_{i=1}^n (\alpha_{1i} + \alpha_{2i}) K_1(h_i, h) + b_1 + \\
 &\quad \sum_{j=1}^n \alpha_{2j} K_2(h_j, h) + b_2
 \end{aligned} \quad (25)$$

式中: K_1 和 K_2 为不同尺度下的小波核函数。

3.2 基于多尺度 LS-SVM 优化的克里金插值方法流程

基于多尺度小波 LS-SVM 优化的克里金空间插值方法实现流程具体可分为如下 4 步:

- (1) 通过离散变异函数公式计算所有样本点对的实验变异函数值;
- (2) 采用多尺度小波 LS-SVM 拟合实验变异函数值, 得到理论变异函数模型;
- (3) 建立克里金空间插值方程组, 根据理论变异函数模型求解克里金权重系数 λ_i ;
- (4) 根据克里金权重系数 λ_i 计算待估位置点的区域化变量估计值;

其中, 步骤(2)中, 多尺度小波 LS-SVM 拟合实现变异函数曲线过程, 又可细分为以下 4 步:

- (1) 计算所有点对的实验变异函数值;
- (2) 实验变异函数值进行分组操作;
- (3) 通过多尺度小波 LS-SVM 拟合分组后的实验变异函数;
- (4) 得到最终理论变异函数。

4 实验与分析

为检验插值效果, 采用交叉验证的方法^[21], 即假定每个观测站点要素值是未知的, 通过其余插值观测点预测来估测, 最后计算所有已知位置点的实际观测值和估测值误差, 以此评判插值结果, 评价指标采用平均绝对误差 (MAE) 和均方根误差 (RMSE) 2 个评价指标^[22]。MAE 反映估测的总体误差, RMSE 反映样本数据的估测灵敏度和极值效应, 二者值越小, 表明效果越好。其定义分别如下:

$$MAE = \sqrt{\frac{1}{S} \sum_{i=1}^S |y_i - \hat{y}_i|} \quad (26)$$

$$RMSE = \sqrt{\frac{1}{S} \sum_{i=1}^S (y_i - \hat{y}_i)^2} \quad (27)$$

式中: S 代表待估位置点的总个数; y_i 为待估位置点的区域化变量真实值, $i=1, \dots, S$; \hat{y}_i 为待估位置点的区域化变量估计值, $i=1, \dots, S$ 。

4.1 实验模拟

模拟采用 Spatial Interpolation Comparison 97 数据 (简称 SIC97 数据), 该数据为 1986 年 5 月 8 日瑞士 467 个降水观测站的降水数据, 其中样本观测站点总个数为 100, 即 $k=100$, 待估位置点的总个数为 367, 即 $S=367$ 。数据来源: [https://www.researchgate.net/profile/Gregoire_Dubois/publication/281292076_Spatial_Interpolation_Comparison_97_\(SIC97\)_dataset](https://www.researchgate.net/profile/Gregoire_Dubois/publication/281292076_Spatial_Interpolation_Comparison_97_(SIC97)_dataset), 通过已知观测点来插值待估点, 观测点与待估点分布如图 1 所示。

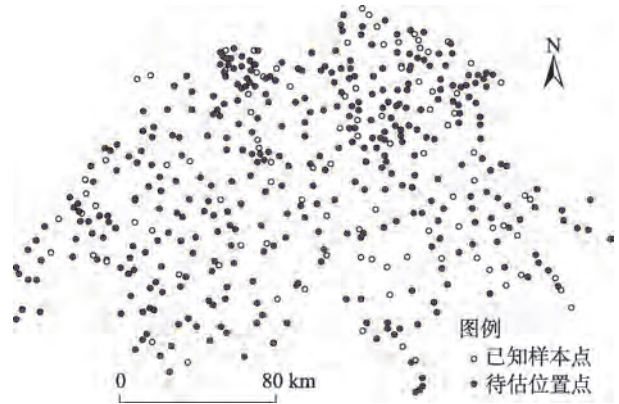


图 1 SIC97 数据分布示意图

Fig. 1 The distribution diagram of SIC97 data

由于样本点数目较多, 拟合实验变异函数之前, 进行分组操作, 分组后需要拟合的实验变异函数值总个数为 20, 即 $n=20$ 。图 2 和图 3 分别表示分组前的变异函数云图和分组后的变异函数云图, 图 4 表示需要拟合的实验变异函数值。

为对比多尺度 LS-SVM 优化的克里金插值方法的效果, 采用传统理论变异函数模型: 球状模型、指数模型、高斯模型。此外, 也对比了单尺度 RBF 核函数和 Morlet 小波核函数 LS-SVM 理论模型。其中, 单尺度和多尺度 LS-SVM 的惩罚参数和核参数寻优都是通过交叉验证的方法, 单尺度 RBF 核函

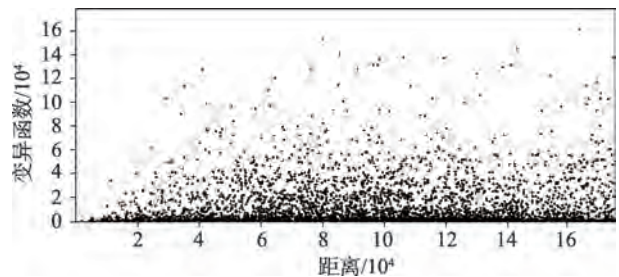


图 2 分组前的变异函数云图

Fig. 2 Variogram cloud before grouping

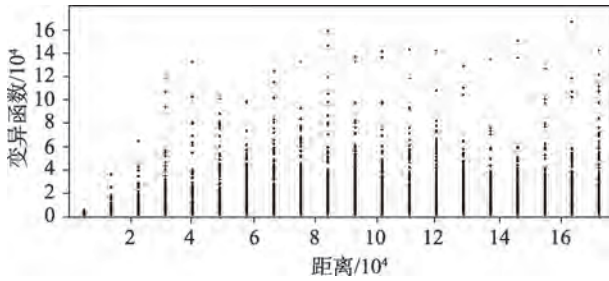


图3 分组前的变异函数云图

Fig. 3 Variogram cloud after grouping

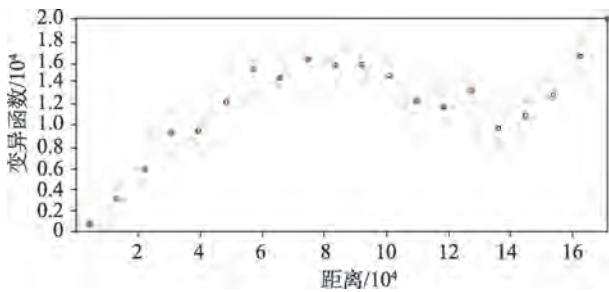


图4 需要拟合的实验变异函数值

Fig. 4 The fitting experimental variogram values

数LS-SVM的惩罚参数 $\gamma=168$,核参数 $\sigma^2=8\times 10^3$;单尺度Morlet小波核LS-SVM的惩罚参数 $\gamma=681$,核参数 $\alpha=328$;二尺度Morlet小波核LS-SVM的惩罚参数 $\gamma_1=3.7\times 10^3$, $\gamma_2=378$,核参数 $\gamma_1=0.03$, $a_2=14$ 。最终拟合的实验变异函数如图5所示。

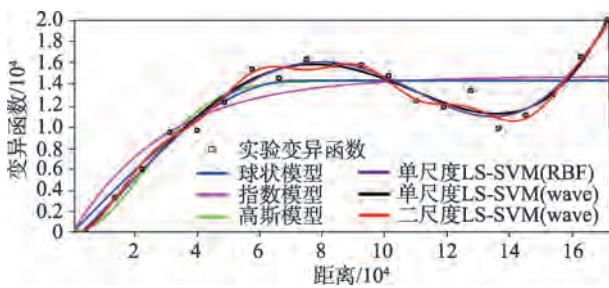


图5 变异函数拟合曲线

Fig.5 Variogram fitting curve

表1直观地显示出各理论变异模型拟合后的克里金插值结果,从表中对比可以发现,在此实验数据中,传统克里金插值拟合模型中,球状模型最优;单尺度RBF核函数LS-SVM模型优于高斯模型和指数模型,但略低于球状模型;单尺度Morlet核函数LS-SVM模型略优于传统克里金插值拟合最好的球状模型;多尺度Morlet核函数LS-SVM模型优于单尺度LS-SVM模型,在所有拟合模型中插值效果最优。结合图5可更加直观地体现多尺度LS-

表1 各理论变异模型拟合后的克里金插值结果对比

Tab. 1 Comparisons of Kriging interpolation results based on the theoretical variogram models

类别	球状模型	指数模型	高斯模型	单尺度LS-SVM (RBF)	单尺度LS-SVM (wave)	二尺度LS-SVM (wave)
MAE	39.43	40.86	110.46	39.81	39.09	38.54
RMSE	56.21	57.33	154.82	56.78	55.97	55.70

SVM模型在拟合实验变异函数中的变化,其中,大尺度保证了整体的空间变化趋势,小尺度体现了细节变化,这样最终使空间多尺度特征得以反映,插值精度较传统方法具有一定程度上的提高,尽管提高并不明显,但结合本方法对变异函数拟合的准确性来看,本文为拟合变异函数提供了一种思路,该方法是一种可选的克里金插值方法。

4.2 实验应用

将该插值方法应用于青岛市PM_{2.5}浓度数据的空间插值分析。PM_{2.5}是指环境空气中动力学当量直径小于等于2.5 μm的颗粒物,富含大量有害物质,研究PM_{2.5}数据的时空分布特征对城市环境保护及控制具有实践意义。

4.2.1 数据源

青岛市19个国控空气质量监测站点,包括青岛市区9个站点、黄岛区(包括原胶南市)2个、即墨市2个、平度市2个、胶州市2个、莱西市2个,其分布如图6所示。PM_{2.5}数据选取日均值,时间跨度2015年10月至2016年9月,共计12个月。

4.2.2 实验分析及对比

通过对青岛市19个国控站点PM_{2.5}数据的统计分析可以发现,无论月均值还是最值,PM_{2.5}浓度12月最高,而8月浓度最低(图7),统计结果与此文所述也基本相符^[23]。故选取特征最明显的12月和8月作为样本数据,以此估测青岛市PM_{2.5}分布特征。

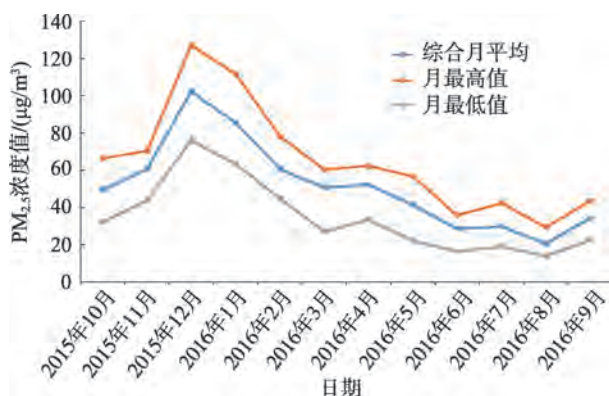
实验流程与SIC97数据的流程类似,故不再详细介绍。本实验为验证本文方法得到的插值结果是否符合样本实际的大致趋势,采用普通克里金插值方法(指数模型)得到图8(a)和图8(c)所示的8月和12月青岛市PM_{2.5}分布图,而图8(b)和图8(d)则是通过多尺度LS-SVM优化的克里金插值方法得到的8月和12月青岛市PM_{2.5}分布图。

对比2种方法得到插值结果整体趋势,多尺度LS-SVM优化的克里金插值方法结果大概符合样本通过普通克里金方法(指数模型)插值出来的结果,



图6 青岛市空气质量监测站点分布图

Fig. 6 The map of air quality monitoring sites in Qingdao city

图7 青岛市各个月份PM_{2.5}浓度值Fig. 7 Monthly PM_{2.5} concentration values of Qingdao city

空间变化趋势相近,因此认为此方法具有一定可靠性。一个更为保守的说法是,采用多尺度LS-SVM优化的克里金插值方法得到的结果与实验样本经克里金插值方法(指数模型)得到的结果趋势一致,此思路具有一定的借鉴性。而从2种方法得到的插值结果差异上看,本文方法插值结果的空间细节变化特征更明显,且存在特征变化急剧的位置,体现了多尺度LS-SVM模拟拟合变异函数中的多尺度效应以及细节变化效应,而普通克里金方法插值结果空间变化较为平缓,拟合实现变异函数时采用的是传统理论变异函数模型,只需求解固定参数,故

一定程度上会忽视空间细节的变化。

整体的时间变化趋势,PM_{2.5}浓度12月远高于8月,8月PM_{2.5}浓度在13~30 μg/m³的区间范围,而12月PM_{2.5}浓度在70~130 μg/m³的区间范围,二者相差较大。结合图7可以得到结论,青岛市PM_{2.5}季节变化明显,表现为冬季高、夏季低的时间特征。整体的空间变化趋势,从东向西PM_{2.5}浓度由低到高,表现为逐渐上升的趋势。8月青岛市的东南、东北方向PM_{2.5}浓度较低,并呈现逐渐向西加重的变化态势,层次感明显;而12月,呈现2个集中分布的较高浓度地带,同时也表现为东南、东北方向PM_{2.5}浓度偏低,西北方向和中部偏高的态势。因此根据其时空分布特征来重点防护与治理具有重要的研究意义。

最后,为对比多尺度小波LS-SVM优化的克里金插值方法,进行各理论变异模型拟合的结果对比。表2和表3分别代表交叉验证检验得到的12月和8月各理论变异模型拟合后的克里金插值结果误差对比。总结来看,多尺度LS-SVM优化的克里金插值方法产生的平均绝对误差和均方根误差相比传统的克里金方法(高斯模型、指数模型、球状模型)、单尺度LS-SVM优化的克里金插值方法均较小,在一定程度上提高了插值精度,可以作为一种可选的克里金插值方法。

5 结语

传统克里金插值方法基于变异函数,根据待估位置点、已知样本数据点的位置关系和区域化变量的空间相关性,实现空间加权估计,但存在以下问题:传统方法理论模型形状固定且选择具有人为主观性;空间数据的变化趋势无法反映;空间数据多尺度特征被忽略。为解决上述问题,本文提出了一种基于多尺度最小二乘支持向量机优化的克里金插值方法,此方法为拟合实验变异函数提供了一种新的思路。从实际样本数据的变化趋势出发,采用最小二乘支持向量机拟合实验变异函数,并利用不同尺度小波核反映不同尺度下的空间变化。最后,实验环节包括模拟和应用,模拟主要验证经多尺度最小二乘支持向量机优化的插值方法的科学有效性以及准确性,应用主要研究青岛市PM_{2.5}时空分布特征,为城市生态科学防护及控制提供理论依据。最终表明,基于多尺度最小二乘支持向量机优化的克里金插值方法能更好地刻画变异函数,反映不同

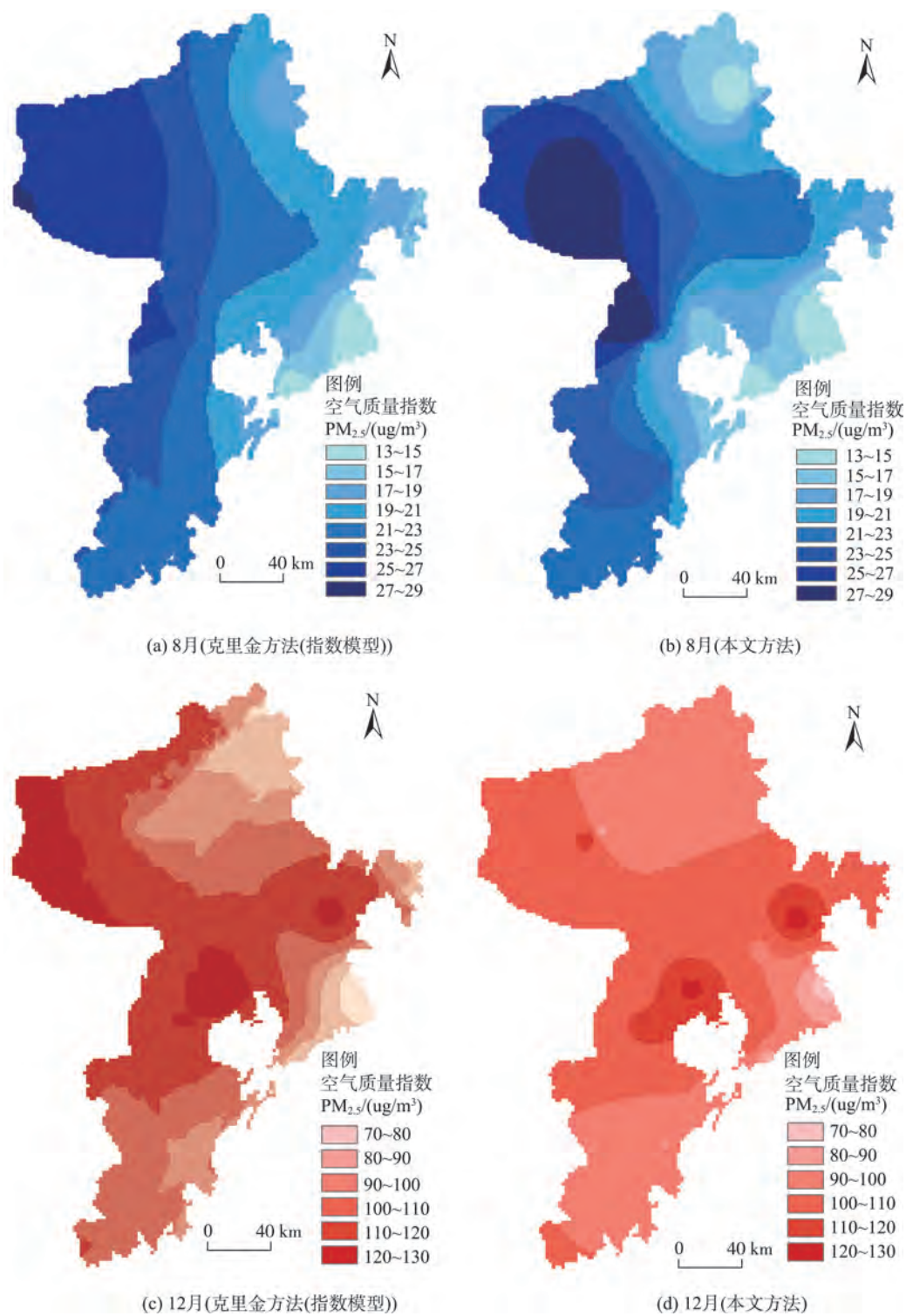


图8 青岛市PM_{2.5}各插值方法结果分布图

Fig. 8 Results of PM_{2.5} interpolation methods in Qingdao

尺度下的空间变化细节,从而在一定程度上提高插值精度,是一种可选的克里金插值方法。本文提出这种方法,一方面是基于对空间数据空间特征(如位置、分布特征等)和属性特征(如空间相关性、空

间异质性)等性质的理解;另一方面基于对变异函数理论拟合模型(球状、指数、高斯等)的分析,辅以真实的数据进行实验测试,结果真实地表明该方法能在一定程度上提高插值方法的精度,尽管提高并

表2 12月各模型插值结果对比

Tab.2 The comparison of the interpolation results of each model in December

指标类别	高斯模型	指数模型	球状模型	单尺度LS-SVM (RBF)	单尺度LS-SVM (wave)	二尺度LS-SVM (wave)
MAE	14.41	12.52	12.54	12.47	12.46	12.42
RMSE	17.66	15.73	15.77	15.57	15.57	15.31

表3 8月各模型插值结果对比

Tab.3 The comparison of the interpolation results of each model in August

指标类别	高斯模型	指数模型	球状模型	单尺度LS-SVM (RBF)	单尺度LS-SVM (wave)	二尺度LS-SVM (wave)
MAE	3.62	3.47	3.49	3.56	3.44	3.44
RMSE	4.54	4.20	4.27	4.58	4.23	4.19

不明显,但结合本方法对变异函数拟合的准确性来看,综合理论分析与实践,提出针对克里金插值问题的一种创新研究思路。通过实验模拟(经典SIC97插值数据)和应用(青岛市PM_{2.5}数据),认为具有一定的借鉴性。值得注意的是,利用样本点对拟合实验变异函数时,样本是影响插值准确性的重要因素。当样本的数量不大时,其空间特征可能并不是空间对象性质的真实反映;而样本数量较大时,其拟合程度的好坏会直接作用待估位置点的估计值。因此,不能完全依靠自动化拟合实验变异函数,需要结合专家经验知识来进一步判断,同时如何准确合理地实现分组操作也值得研究。此外,今后还可更细致地对空间进行尺度划分进行研究,从宏观与微观进行适当选取。

致谢:感谢青悦开放环境数据中心(<https://data.epmap.org>)提供的环境数据处理支持。

参考文献(References):

- [1] Lam N S N. Spatial interpolation methods: A review[J]. The American Cartographer, 1983,10(2):129-150.
- [2] 王劲峰,武继磊,孙英君,等.空间信息分析技术[J].地理研究,2005,24(3):464-472. [Wang J F, Wu J L, Sun Y J, et al. Techniques of spatial data analysis[J]. Geographical Research, 2005,24(3):464-472.]
- [3] 朱会义,刘述林,贾绍凤.自然地理要素空间插值的几个问题[J].地理研究,2004,23(4):425-432. [Zhu H Y, Liu S L, Jia S F. Problems of the spatial interpolation of physical geographical elements[J]. Geographical Research, 2004,23(4):425-432.]
- [4] Matheron G. Principles of geostatistics[J]. Economic Geology, 1963,58(8):1246-1266.
- [5] 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. [Sun H Q. Geostatistics and application: Quantitative geography[M]. Beijing: China University of Mining and Technology Press, 1990.]
- [6] 徐建华.计量地理学[M].北京:高等教育出版社,2006. [Xu J H. Quantitative geography[M]. Beijing: Higher Education Press, 2006.]
- [7] 孟斌,王劲峰.地理数据尺度转换方法研究进展[J].地理学报,2005,60(2):277-288. [Meng B, Wang J F. A review on the methodology of scaling with geo-data[J]. Geographica Sinica, 2005,60(2):277-288.]
- [8] 王仁铎,胡光道.线性地质统计学[M].北京:地质出版社,1988. [Wang R D, Hu G D. Linear geological statistics [M]. Beijing: Geological Publishing House, 1988.]
- [9] 矫希国,刘超.变差函数的参数模拟[J].物探化探计算技术,1996,18(2):157-161. [Jiao X G, Liu C. Estimation of variation parameter[J]. Computing Techniques for Geophysical and Geochemical Exploration, 1996,18(2):157-161.]
- [10] Li M, Li G, Azarm S. A kriging meta-model assisted multi-objective genetic algorithm for design optimization [J]. Journal of Mechanical Design, 2008,130(3): 031401.
- [11] 张强,许少华,于文涛,等.粒子群算法在克里金三维地质建模中的应用[J].大庆石油学院学报,2011,35(1):85-89. [Zhang Q, Xu S H, Yu W T, et al. Application of particle swarm optimization to Kriging three-dimensional geological modeling[J]. Journal of Daqing Petroleum Institute, 2011,35(1):85-89.]
- [12] 严华雯,吴健平.加权最小二乘法改进遗传克里金插值方法研究[J].计算机技术与发展,2012,22(3):92-95. [Yan H W, Wu J P. Research on genetic algorithm kriging optimized by weight least square[J]. Computer Technology and Development, 2012,22(3):92-95.]
- [13] Vapnik V. The nature of statistical learning theory[M]. New York: Springer, 1995.
- [14] Suykens J A K, Vandewalle J. Least squares support vector machine classifiers[J]. Neural Processing Letters, 1999,9(3):293-300.
- [15] Daubechies I. Ten lectures on wavelets[M]. Philadelphia: Society for Industrial and Applied Mathematics, 1992.
- [16] 张建锋,崔树军,李国敏.常用小波及其时-频特性[J].地学前缘,2012,19(6):248-253. [Zhang J F, Cui S J, Li G M. Frequently-used wavelets and their time-frequency properties[J]. Earth Science Frontiers, 2012,19(6):248-253.]
- [17] 周建萍,郑应平,王志萍.基于Morlet小波核多类支持向量机的故障诊断[J].华东电力,2008,36(8):76-80. [Zhou J P, Zheng Y P, Wang Z P. Fault diagnosis based on Mor-

- let wavelet kernel multi-class support vector machine[J]. East China Electric Power, 2008,36(8):76-80.]
- [18] 王琴,沈远彤.二尺度最小二乘小波支持向量回归[J].工程地球物理学报,2009,6(4):516-520. [Wang Q, Shen Y D. Multi-scale least squares support vector regression[J]. Chinese Journal of Engineering Geophysics, 2009,6(4): 516-520.]
- [19] 张相胜,王蕾,潘丰.多尺度最小二乘小波支持向量机的回归建模[J].计算机工程,2012,38(10):175-177. [Zhang X S, Wang L, Pan F. Regression modeling of multi-scale least square wavelet support vector machine[J]. Computer Engineering, 2012,38(10):175-177.]
- [20] 王琴,沈远彤.基于压缩感知的多尺度最小二乘支持向量机[J].自动化学报,2016,42(4):631-640. [Wang Q, Shen Y D. Multi-scale least squares support vector machine using compressive sensing[J]. Acta Automatica Sinica, 2016,42(4):631-640.]
- [21] Johnston K, Ver Hoef J M, Krivoruchko K, et al. Using ArcGIS geostatistical analyst[M]. Redlands: Esri, 2001.
- [22] Han S, Schneider S M, Evans R G. Evaluating co-kriging for improving soil nutrient sampling efficiency[J]. Transactions of the ASAE, 2003,46(3):845.
- [23] 杜艳伟,程建光,王静,等.2013年青岛市PM_{2.5}时空分布及来源分析[J].城市环境与城市生态,2015,28(2):17-19. [Du Y W, Cheng J G, Wang J, et al. Temporal and spatial distribution of PM_{2.5} in Qingdao city and analysis of its sources in 2013[J]. Urban Environment & Urban Ecology, 2015,28(2):17-19.]