Statistics and Application 统计学与应用, 2023, 12(2), 306-317
Published Online April 2023 in Hans. https://www.hanspub.org/journal/sa
https://doi.org/10.12677/sa.2023.122032

Hans汉斯

空间滞后－混合地理加权回归模型中的数据分区及参数估计

李知恩

长安大学理学院，陕西 西安

收稿日期：2023年3月13日；录用日期：2023年4月3日；发布日期：2023年4月18日

摘　要

空间滞后模型和地理加权回归模型均为经典的地统计学模型，分别用于处理带有空间自相关性或异质性的数据，但是在处理同时带有空间自相关性和异质性的数据时拟合效果较差。为了同时考虑数据的自相关性和异质性，提升模型的拟合效果，本文在空间滞后模型和地理加权回归模型的基础上做出改进。首先针对空间数据的异质性，使用改进的k均值聚类方法对空间数据进行分区处理。其次，在分区内部引入空间的自相关性，给出空间滞后－混合地理加权回归模型，并提出了基于莫兰指数与权重矩阵的关系进行估计的莫兰指数优化法。通过在真实数据集上的实验研究，证明了本文方法相比传统方法具有更好的拟合效果。

关键词

空间滞后模型，地理加权回归模型，莫兰指数优化法，聚类分析

Data Partition and Parameter Estimation in Spatial Lag-Mixed Geographical Weighted Regression Model

Zhi'en Li

College of Science, Chang'an University, Xi'an Shaanxi

Received: Mar. 13th, 2023; accepted: Apr. 3rd, 2023; published: Apr. 18th, 2023

Abstract

Both the spatial lag model and the geographically weighted regression model are classically geos-

文章引用: 李知恩. 空间滞后-混合地理加权回归模型中的数据分区及参数估计[J]. 统计学与应用, 2023, 12(2): 306-317.
DOI: 10.12677/sa.2023.122032

tatistical models, which are used to deal with data with spatial autocorrelation or heterogeneity respectively, but the fitting effect is poor when dealing with data with both spatial autocorrelation and heterogeneity. In order to consider the autocorrelation and heterogeneity of the data at the same time and improve the fitting effect of the model, this paper makes improvements on the basis of the spatial lag model and the geographically weighted regression model. Firstly, according to the heterogeneity of spatial data, the improved k-means clustering method is used to partition the spatial data. Secondly, the spatial autocorrelation is introduced into the interior of the zone, and the spatial lag-mixed geographically weighted regression model is given, and the Moran's I optimization method based on the relationship between Moran's I and weighted matrix is proposed. Through experimental research on real data sets, it is proved that this method has better fitting effect than traditional methods.

Keywords

Spatial Lag Model, Geographically Weighted Regression Model, Moran's I Optimization Method, Cluster Analysis

1. 引言

普通线性回归模型用于处理独立、正态和方差齐性的数据，但是由于空间数据存在自相关性和异质性，很难满足独、正态、方差齐性这些假设条件，因此需要构建适用于空间数据的空间回归模型。

空间回归模型包括空间滞后模型、地理加权回归模型等，主要用于处理空间数据，空间数据的两个重要特征分别是空间自相关性和空间异质性。针对空间自相关性的研究最初是在线性回归模型中加入含有空间权重矩阵的项，使得研究区域的因变量不仅与本区域的解释变量有关，还与邻近区域的因变量相关。1988 年 Anselin [1]提出了空间线性回归(Spatial Linear Regression, SLR)模型，当模型中的参数取值不同时，SLR 模型可派生出普通线性回归模型、一阶段空间自回归模型、空间滞后模型(Spatial Lag Model, SLM)、空间误差模型以及空间杜宾模型。其中空间滞后模型可以用来分析许多具有空间自相关性的问题，如房价预测，犯罪率分析，经济增长等。

针对空间数据的异质性，1996 年 Brunsdon 等人[2]首先将异质性的思想融入到 SLM 的参数求解中，提出局部最大似然估计法，将 SLM 中所有的空间系数看做随空间位置变化的变量。1998 年 Fotheringham 等人[3]将 SLM 中的常系数变为随空间位置变化的变系数来处理空间数据的异质性，提出著名的地理加权回归(Geographically Weighted Regression, GWR)模型。GWR 模型将所有解释变量作为局部解释变量，由于局部解释变量既可能与空间位置有关，也可能与空间位置无关，因此 2011 年 Paez 等人[4]将 GWR 模型中部分解释变量的系数变为常系数，提出混合地理加权回归(Mixed Geographically Weighted Regression, MGWR)模型，使得模型拟合效果相比于 GWR 模型更精确。目前，混合地理加权回归模型在空间统计学领域已经得到了广泛应用，可以用于城市规划、环境保护和资源管理等方面，通过精确预测，为决策者提供更准确的决策依据。

MGWR 模型以及 SLM 有利有弊：SLM 考虑了因变量间的空间自相关性，却忽略了空间异质性；MGWR 模型虽然基于局部光滑的思想减弱了空间异质性，但是却没有考虑空间自相关性。目前，同时考

虑空间自相关性和空间异质性模型的研究寥寥无几，相应的参数估计以及模型的拟合效果检验更是凤毛麟角。1996 年 Brunsdon 等人[5]在考虑空间自相关的情况下，使用极大似然估计法对 GWR 模型进行参数估计，在每个观测点处都进行线性回归计算。2005 年魏传华等人[6]通过模拟实验验证了具有空间自相关的 GWR 模型中常系数估计的精确性和稳健性。2011 年 Geniaux 等人[7]证明了在 MGWR 模型中，空间部分回归系数是非平稳的，并采用局部两步最小二乘法对 MGWR 模型进行参数估计。2013 年乔宁宁[8]提出了混合地理加权空间滞后回归模型，给出相应的参数估计方法，即先由局部最小二乘法得到变系数的表达式，再使用极大似然估计的方法对参数进行求解。本文在 2017 年乔宁宁[8]的研究基础上，通过在 MGWR 模型上增加空间滞后项来降低空间自相关性对结果的影响，提出空间滞后－混合地理加权回归(Spatial Lag-Geographical Weighted Regression, SL-MGWR)模型，SL-MGWR 模型的预测能力和拟合效果均优于原始模型。针对 SL-MGWR 模型的参数估计问题，将混合地理加权回归模型的两步估计法[7] [8]扩展到空间滞后－混合地理加权回归模型中，提出了莫兰指数优化法。莫兰指数优化法能够有效降低计算复杂度，提高模型预测能力并降低了自相关性和异质性对模型计算结果的影响程度。SL-MGWR 模型可以处理原始模型的所有问题，既可以应用在房价预测，犯罪率分析，经济增长，又可以用于城市规划、环境保护和资源管理等方面，且其解释能力及预测效果都优于原始模型。

本文第一部分是引言，介绍本文研究背景。第二部分是模型介绍，介绍了地理加权回归模型、空间滞后模型以及本文提出的空间滞后－混合地理加权回归(SL-MGWR)模型。第三部分是参数估计方法介绍，介绍了两步估计法和莫兰指数优化法的推导过程。第四部分是实验研究，分为三小节，第一小节为使用两步估计法[7] [8]比较普通地理加权回归模型以及分区后的地理加权回归模型，来说明数据分区之后模型的模型效果更好，第二小节为使用两步估计法将 GWR 以及 SL-MGWR 模型进行对比，说明 SL-MGWR 模型的拟合效果较好。第三小节是 SL-MGWR 模型传统及莫兰指数优化法的拟合效果对比，得到莫兰指数优化法的效果优于传统两步估计法。第五部分是结论，得到最后的结果。

2. 模型介绍

2.1. 地理加权回归模型

由于空间数据存在空间异质性，空间地区内的空间单元本身不是均质的，在形状及面积上有较大差别，而在进行空间数据分析时，变量的观测值一般都是以给定的空间单元为抽样单位得到的，随着地理位置的变化，变量间的关系或者结构也会引起变化。在全局模型中，我们假设不存在空间异质性，所得结果有一定误差，所以需要对传统方法进行改进。

1996 年 Brunsdon [2]及 1998 年 Fotheringham 等人[3]基于局部平滑的思想，提出了地理加权回归模型，将数据的空间位置嵌入到回归参数中，利用局部加权最小二乘法进行逐点参数估计。即，

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) x_{ik} + \varepsilon_i$$

其中 (u_i, v_i) 为第 i 个采样点的坐标，$\beta_k(u_i, v_i)$ 为第 i 个采样点上的第 k 个回归参数，是地理位置的函数；ε_i 是第 i 个区域的随机误差，满足零均值、同方差、相互独立等基本假定。

2.2. 空间滞后模型

传统经典的线性回归模型具有严格的前提假设条件——独立、正态、齐方差性。但是，由于空间数据存在空间自相关性和异质性，使得这些条件很难满足，在使用传统线性回归模型解决空间问题时，会造成模型参数错误并降低模型的有效性。1988 年 Anselin [1]考虑空间数据的自相关性，给出了如下形式，

使用空间滞后模型来处理数据，

$$Y=\rho WY+X\beta+\varepsilon,\varepsilon\sim N\left[O,\sigma^2 I\right].$$

其中 Y 为 $n\times1$ 维响应变量 $Y=\left(Y_1,Y_2,\cdots,Y_n\right)$，$X=\left(X_1,X_2,\cdots,X_K\right)$ 为包含 K 个解释变量的 $n\times K$ 维矩阵。$W=\begin{bmatrix}0 & b_{12} & \cdots & b_{1n}\\ b_{21} & 0 & \cdots & b_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ b_{n1} & b_{n2} & \cdots & 0\end{bmatrix}$ 为空间权重矩阵，$b_{ij}=\dfrac{1}{d_{ij}^2}$，$d_{ij}$ 为位置 i 和位置 j 之间的距离，且 $d_{ij}=\sqrt{\left(u_i-u_j\right)^2+\left(v_i-v_j\right)^2}$，其中 $u(i),v(i)$ 分别为位置 i 的经度和纬度。WY 为空间滞后因子，空间滞后考虑一个特定的观测地区会受到相邻地区的影响，使用空间滞后因子可以得出实际上相邻的空间地区观测值依距离加权后的平均值，$\rho\in[0,1]$ 为空间滞后项系数，取值越接近 1，说明相邻地区的因变量取值越相似。$\beta=\left(\beta_1,\beta_2,\cdots,\beta_K\right)^{\mathrm{T}}$ 为参数向量，$\varepsilon=\left[\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n\right]^{\mathrm{T}}$ 为服从正态分布的 n 维随机误差向量且 $Cov\left(\varepsilon_i,\varepsilon_j\right)=0,i\neq j$，$I$ 为 $n\times n$ 的单位矩阵。

2.3. 空间滞后 - 混合地理加权回归模型

为了同时考虑空间数据的异质性及自相关性，本文将空间滞后模型中的空间滞后因子加入地理加权回归模型中，将空间数据的自相关性和异质性结合起来，得到具有空间自相关性的地理加权回归模型：

$$y_i=\rho\sum_{j=1}^{n}b_{ij}y_j+\beta_0\left(u_i,v_i\right)+\sum_{k=1}^{p}\beta_k\left(u_i,v_i\right)x_{ik}+\sum_{k=p+1}^{m}\beta_k x_{ik}+\varepsilon_i$$

为避免混淆，假设总共有 n 个样本点，模型中具有 m 个解释变量，令 $\beta_0\left(u_i,v_i\right)=\beta_{i0},\beta_k\left(u_i,v_i\right)=\beta_{ik}$。可得：

$$\begin{aligned}y_i&=\rho\sum_{j=1}^{n}b_{ij}y_i+\beta_0\left(u_i,v_i\right)+\sum_{k=1}^{p}\beta_k\left(u_i,v_i\right)x_{ik}+\sum_{k=p+1}^{m}\beta_k x_{ik}+\varepsilon_i\\&=\rho\sum_{j=1}^{n}b_{ij}y_i+\beta_{i0}+\sum_{k=1}^{p}\beta_{ik}x_{ik}+\sum_{k=p+1}^{m}\beta_k x_{ik}+\varepsilon_i\end{aligned}$$

可以简写为：

$$Y=\rho WY+X\beta+M+\varepsilon. \tag{1}$$

其中 $M=\left[\sum_{k=0}^{p}\beta_{1k}x_{1k}\quad\sum_{k=0}^{p}\beta_{2k}x_{2k}\quad\cdots\quad\sum_{k=0}^{p}\beta_{nk}x_{nk}\right]^{\mathrm{T}}$，$\beta=\left(\beta_{p+1},\beta_{p+2},\cdots,\beta_m\right)^{\mathrm{T}}$ 为常参数向量，$x_{10}=x_{20}=\cdots=x_{n0}=1$，$X=\begin{bmatrix}x_{1(p+1)} & x_{1(p+2)} & \cdots & x_{1m}\\ x_{2(p+1)} & x_{21} & \cdots & x_{2m}\\ \vdots & \vdots & & \vdots\\ x_{n(p+1)} & x_{n(p+2)} & \cdots & x_{nm}\end{bmatrix}$ 为 $n\times(m-p)$ 维矩阵。W 为空间权重矩阵，$W=\begin{bmatrix}0 & b_{12} & \cdots & b_{1n}\\ b_{21} & 0 & \cdots & b_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ b_{n1} & b_{n2} & \cdots & 0\end{bmatrix}$，$b_{ij}=\dfrac{1}{d_{ij}^2}$，$d_{ij}$ 为位置 i 和位置 j 之间的距离，且 $d_{ij}=\sqrt{\left(u_i-u_j\right)^2+\left(v_i-v_j\right)^2}$，其中 $u(i),v(i)$ 分别为位置 i 的经度和纬度。WY 为空间滞后因子，$\rho\in[0,1]$ 为空间滞后项系数，取值越接近

1，说明相邻地区的因变量取值越相似。$\varepsilon=[\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n]^{\mathrm{T}}$ 为服从正态分布的 n 维随机误差向量且 $Cov(\varepsilon_i,\varepsilon_j)=0, i\neq j$，$I$ 为 $n\times n$ 的单位矩阵。

若 M 中的系数项变为常系数，则模型变为空间滞后模型(SLM)。

3. 参数估计法

3.1. 两步估计法

根据模型特点，可知模型由空间滞后部分 ρWY，变系数部分 M 以及常系数部分 $X\beta$ 组成，因此可以优化两步估计法[7] [8]来估计模型参数 ρ,β,σ^2，具体步骤如下：

第一步，假设空间滞后部分 ρWY 中的 ρ 一常系数部分 $X\beta$ 中的 β 已知，则此模型 $Y=\rho WY+X\beta+M+\varepsilon$ 可以化为

$$(I_n-\rho W)Y-X\beta=M+\varepsilon. \tag{2}$$

令 $A=I_n-\rho W$，$Y^*=(I_n-\rho W)Y-X\beta=AY-X\beta$，则式(2)变为 $Y^*=M+\varepsilon$。对于该模型，利用空间局部加权最小二乘法可知：

$$\hat{M}=SY^* \tag{3}$$

其中令 $S_i=X_i^{\mathrm{T}}\left(X^{\mathrm{T}}W_iX\right)^{-1}X^{\mathrm{T}}W_i$，将 S_i 称为 i 点的帽子向量[9]，X_i^{T} 为矩阵 X 的第 i 行。

$S=\begin{bmatrix}S_1\\S_2\\\vdots\\S_n\end{bmatrix}=\begin{bmatrix}X_1^{\mathrm{T}}\left(X^{\mathrm{T}}W_1X\right)^{-1}X^{\mathrm{T}}W_1\\X_2^{\mathrm{T}}\left(X^{\mathrm{T}}W_2X\right)^{-1}X^{\mathrm{T}}W_2\\\vdots\\X_n^{\mathrm{T}}\left(X^{\mathrm{T}}W_nX\right)^{-1}X^{\mathrm{T}}W_n\end{bmatrix}$ 为帽子矩阵。在帽子矩阵 S 中，$W_i=diag(w_{i1},w_{i2},\cdots,w_{in})$，此时 w_{ij} 为高斯空间权函数，带宽 b 由 $CV=\sum_{i=1}^{n}\left[y_i-\hat{y}_{\neq i}(b)\right]^2$ [10]确定。

将 $\hat{M}=SY^*$ 代入 $Y^*=M+\varepsilon$ 中可得

$$Y^*=SY^*+\varepsilon \tag{4}$$

第二步，利用最大似然估计的方法对 $Y^*=SY^*+\varepsilon$ 进行求解，由于 $\varepsilon=(I_n-S)Y^*$，则由变换定理可得 Y 的似然函数 $p(Y)=p(\varepsilon)\left|\frac{\partial\varepsilon}{\partial Y}\right|=p(\varepsilon)\left|\frac{\partial\left[(I_n-S)(AY-X\beta)\right]}{\partial Y}\right|=p(\varepsilon)|I_n-S||A|$，可得 Y 的对数似然函数为：

$$\begin{aligned}\ln L&=-\frac{n}{2}\ln 2\pi-\frac{n}{2}\ln\sigma^2+\ln|I_n-S|+\ln|A|-\frac{1}{2\sigma^2}\left(Y^*-SY^*\right)^{\mathrm{T}}\left(Y^*-SY^*\right)\\&=-\frac{n}{2}\ln 2\pi-\frac{n}{2}\ln\sigma^2+\ln|I_n-S|+\ln|I_n-\rho W|-\frac{1}{2\sigma^2}\left[(I_n-S)Y^*\right]^{\mathrm{T}}\left[(I_n-S)Y^*\right].\end{aligned}$$

求解 $\frac{\partial\ln L}{\partial\beta}=0$，$\frac{\partial\ln L}{\partial\sigma^2}=0$，可得：

$$\hat{\beta}(\rho)=\left(X^{\mathrm{T}}QX\right)^{-1}X^{\mathrm{T}}QAY,$$

$$\hat{\sigma}^2(\rho)=\frac{1}{n}\varepsilon^{\mathrm{T}}\varepsilon=\frac{1}{n}\left[AY-X\left(X^{\mathrm{T}}QX\right)^{-1}X^{\mathrm{T}}QAY\right]^{\mathrm{T}}Q\left[AY-X\left(X^{\mathrm{T}}QX\right)^{-1}X^{\mathrm{T}}QAY\right].$$

其中 $Q=(I_n-S)^{\mathrm{T}}(I_n-S)$，将 $\hat{\beta}(\rho),\hat{\sigma}^2(\rho)$ 代入上述对数似然函数 $\ln L$，可得：

$$\ln L=C-\frac{n}{2}\ln\left(|A|^{-\frac{2}{n}}\varepsilon^{\mathrm{T}}\varepsilon\right)$$

其中 C 为常数。

假设 W 的特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_n$，则 W 可对角化，因此存在不可逆矩阵 P，使得 W 变为对角阵 Λ，即 $\Lambda=P^{-1}WP$，可得：

$$P^{-1}(\lambda I_n-W)P=P^{-1}\lambda P-P^{-1}WP=\lambda I_n-\Lambda,$$

$$|\lambda I_n-W|=\prod_{i=1}^{n}(\lambda-\lambda_i)$$

可得[11]

$$|A|=|I_n-\rho W|=\prod_{i=1}^{n}(1-\rho\cdot\lambda_i)$$

因此求下式极小值，可得 ρ 的估计值 $\hat{\rho}$

$$|A|^{-\frac{2}{n}}\varepsilon^{\mathrm{T}}\varepsilon=\prod_{i=1}^{n}(1-\rho\cdot\lambda_i)^{-\frac{2}{n}}\left[AY-X\left(X^{\mathrm{T}}QX\right)^{-1}X^{\mathrm{T}}QAY\right]^{\mathrm{T}}\left[AY-X\left(X^{\mathrm{T}}QX\right)^{-1}X^{\mathrm{T}}QAY\right] \tag{5}$$

将 ρ 的估计值 $\hat{\rho}$ 代入 $\hat{\beta}(\rho),\hat{\sigma}^2(\rho)$，即可得到 $\beta(\rho),\sigma^2(\rho)$ 的最终估计值 $\hat{\beta}(\hat{\rho}),\hat{\sigma}^2(\hat{\rho})$。由此可得，模型中变系数部分 $\hat{M}=SY^*=S(AY-X\beta)$，此时模型中 ρ,W,β,M,σ^2 均已知，因此可以使用 SLM-MGWR 模型进行预测。

模型 $Y=\rho WY+X\beta+M+\varepsilon$ 的参数 $\hat{\beta}(\hat{\rho}),\hat{\sigma}^2(\hat{\rho})$ 求得分别为

$$\hat{\beta}(\hat{\rho})=\left(X^{\mathrm{T}}QX\right)^{-1}X^{\mathrm{T}}QAY,$$

$$\hat{\sigma}^2(\hat{\rho})=\frac{1}{n}\varepsilon^{\mathrm{T}}\varepsilon=\frac{1}{n}\left[AY-X\left(X^{\mathrm{T}}QX\right)^{-1}X^{\mathrm{T}}QAY\right]^{\mathrm{T}}Q\left[AY-X\left(X^{\mathrm{T}}QX\right)^{-1}X^{\mathrm{T}}QAY\right].$$

其中 $Q=(I_n-S)^{\mathrm{T}}(I_n-S)$，$A=I_n-\hat{\rho}W$。

3.2. 莫兰指数优化法

3.2.1. ρ 和莫兰指数的关系

莫兰指数是用来度量空间相关性的重要指标，在 SL-MGWR 模型中，由于空间滞后因子的存在，所以我们考虑将空间滞后项转化为莫兰指数，找到莫兰指数和空间滞后项的关系，由文献[12]可知莫兰指数具有如下形式：

$$\begin{aligned}M_c&=\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ij}(y_i-\bar{y})(y_j-\bar{y})}{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ij}}=\frac{n}{\sum_{i=1}^{n}(y_i-\bar{y})^2}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ij}(y_i-\bar{y})(y_j-\bar{y})}{\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ij}}\\&=\frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ij}}\frac{\sum_{i=1}^{n}(y_i-\bar{y})\sum_{j=1}^{n}w_{ij}(y_j-\bar{y})}{(n-1)s^2}=\frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ij}}\frac{\sum_{i=1}^{n}z_i\sum_{j=1}^{n}w_{ij}z_j}{n-1}.\end{aligned} \tag{6}$$

其中 $W=\begin{bmatrix}0 & w_{12} & \cdots & w_{1n}\\ w_{21} & 0 & \cdots & w_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ w_{n1} & w_{n2} & \cdots & 0\end{bmatrix}$ 为空间权重矩阵，$Y^{\mathrm{T}}=\left[y_1,y_2,\cdots,y_n\right]$ 为随机变量。上式(6)右侧强调了莫兰指数 MC 对双变量的回归，由文献[12]可知 MC 为回归斜率的系数。对随机变量 Y 值进行标准差标准化处理之后得到响应变量 Z，可知 WZ 关于 Z 的回归系数为

$$\left(Z^{\mathrm{T}}Z\right)^{-1}Z^{\mathrm{T}}WZ=\frac{\sum_{i=1}^{n}Z_i\left(\sum_{j=1}^{n}w_{ij}Z_j\right)}{n-1}.$$

则向量 $W1$ 关于 $1_{n\times 1}=(1,1,\cdots,1)^{\mathrm{T}}$ 的回归系数为

$$\left(1^{\mathrm{T}}1\right)^{-1}1^{\mathrm{T}}W1=\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ij}}{n}.$$

如果空间权重矩阵为行标准化矩阵，则有 $\sum_{j=1}^{n}w_{ij}=1$，$\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ij}=n$，MC 变为 $M_c=\frac{\sum_{i=1}^{n}z_i\sum_{j=1}^{n}w_{ij}z_j}{n-1}$，此时 $M_c=\left(Z^{\mathrm{T}}Z\right)^{-1}Z^{\mathrm{T}}WZ$。由此可知：

$$WZ=M_c\cdot Z+\varepsilon. \tag{7}$$

其中 M_c 为莫兰指数，$\varepsilon\sim N\left(O,\sigma^2I_n\right)$，且误差之间相互独立。

3.2.2. 模型参数估计

根据模型特点，可知模型由空间滞后部分 ρWY，变系数部分 M，以及常系数部分 $X\beta$ 组成，因此可以借鉴两步估计法[7] [8]以及文献[12]中 ρ 和莫兰指数的关系，来估计模型未知参数 ρ,β,σ^2，具体步骤如下：

第一步，假设 SL-MGWR 中莫兰指数 M_c、滞后项系数 ρ 和常系数部分 β 已知，同时若空间权重矩阵 W 为行标准化矩阵，则由式(7)可知：

$$WY=M_c\cdot Y+\varepsilon_1,\varepsilon_1\sim N\left[O,\sigma_1^2I_n\right]$$

在原模型(1)的基础上，可知：

$$Y=\rho WY+M+X\beta+\varepsilon_2,\varepsilon_2\sim N\left[O,\sigma_2^2I_n\right]$$

则此模型可以变为

$$\begin{aligned}Y&=\rho WY+M+X\beta+\varepsilon_2=\rho\left(M_c\cdot Y+\varepsilon_1\right)+M+X\beta+\varepsilon_2\\&=\rho M_c\cdot Y+M+X\beta+\left(\rho\varepsilon_1+\varepsilon_2\right)=\rho M_c\cdot Y+M+X\beta+\varepsilon\end{aligned}$$

其中 $\varepsilon=\rho\varepsilon_1+\varepsilon_2$，$\varepsilon\sim N\left(O,\left(\rho^2\sigma_1^2+\sigma_2^2\right)I_n\right)$。

即：

$$\left(1-\rho M_c\right)Y-X\beta=M+\varepsilon \tag{8}$$

令 $Y^*=\left(1-\rho\cdot M_c\right)Y-X\beta$，则式(2)变为 $Y^*=M+\varepsilon$。对于该模型，利用空间局部加权最小二乘法可知：

$$\hat{M} = SY^* \tag{9}$$

其中令 $S_i = X_i^{\mathrm{T}}\left(X^{\mathrm{T}}W_iX\right)^{-1}X^{\mathrm{T}}W_i$，将称 S_i 为 i 点的帽子向量[9]， X_i^{T} 为矩阵 X 的第 i 行。

$$S = \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_n \end{bmatrix} = \begin{bmatrix} X_1^{\mathrm{T}}\left(X^{\mathrm{T}}W_1X\right)^{-1}X^{\mathrm{T}}W_1 \\ X_2^{\mathrm{T}}\left(X^{\mathrm{T}}W_2X\right)^{-1}X^{\mathrm{T}}W_2 \\ \vdots \\ X_n^{\mathrm{T}}\left(X^{\mathrm{T}}W_nX\right)^{-1}X^{\mathrm{T}}W_n \end{bmatrix}$$

在帽子矩阵 S 中， $W_i = diag\left(w_{i1}, w_{i2}, \cdots, w_{in}\right)$，此时 w_{ij} 为高斯空间权函数，带宽 b 由 $CV = \sum_{i=1}^{n}\left[y_i - \hat{y}_{\neq i}(b)\right]^2$ [10]确定。

将(9)式中 M 的估计值代入 $Y^* = M + \varepsilon$ 中可得

$$Y^* = SY^* + \varepsilon \tag{10}$$

第二步，利用最大似然估计的方法对式(10)进行求解，由于 $\varepsilon = \left(I_n - S\right)Y^*$，则由变换定理可得 Y 的似然函数 $p(Y) = p(\varepsilon)\left|\frac{\partial\varepsilon}{\partial Y}\right| = p(\varepsilon)\left|\frac{\partial\left[\left(I_n - S\right)\left(\left(1-\rho M_c\right)Y - X\beta\right)\right]}{\partial Y}\right| = p(\varepsilon)\left|I_n - S\right|\left|1-\rho M_c\right|$，可得 Y 的对数似然函数为：

$$\begin{aligned}\ln L &= -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 + \ln\left|I_n - S\right| + \ln\left(1-\rho M_c\right) - \frac{1}{2\sigma^2}\left(Y^* - SY^*\right)^{\mathrm{T}}\left(Y^* - SY^*\right) \\ &= -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 + \ln\left|I_n - S\right| + \ln\left(1-\rho M_c\right) - \frac{1}{2\sigma^2}\left[\left(I_n - S\right)Y^*\right]^{\mathrm{T}}\left[\left(I_n - S\right)Y^*\right]\end{aligned}$$

求解 $\frac{\partial\ln L}{\partial\beta} = 0$， $\frac{\partial\ln L}{\partial\sigma^2} = 0$，可得：

$$\begin{aligned}\hat{\beta}(\rho) &= \left(X^{\mathrm{T}}QX\right)^{-1}X^{\mathrm{T}}Q\left(1-\rho\cdot M_c\right)Y, \\ \hat{\sigma}^2(\rho) &= \frac{1}{n}\varepsilon^{\mathrm{T}}\varepsilon = \frac{1}{n}\left[\left(1-\rho M_c\right)Y - H\right]^{\mathrm{T}}Q\left[\left(1-\rho M_c\right)Y - H\right]\end{aligned}$$

其中 $Q = \left(I_n - S\right)^{\mathrm{T}}\left(I_n - S\right)$， $H = X\hat{\beta} = X\left(X^{\mathrm{T}}QX\right)^{-1}X^{\mathrm{T}}Q\left(1-\rho\cdot M_c\right)Y$。

将 $\hat{\beta}(\rho), \hat{\sigma}^2(\rho)$ 代入 $\ln L$ 可得关于 ρ 的集中对数似然函数，

$$\begin{aligned}\ln L = &-\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\frac{1}{n} - \frac{n}{2} + \ln\left(1-\rho M_c\right) + \ln\left|I_n - S\right| \\ &-\frac{n}{2}\ln\left[\left(1-\rho M_c\right)Y - H\right]^{\mathrm{T}}Q\left[\left(1-\rho M_c\right)Y - H\right]\end{aligned} \tag{11}$$

上述式子为 ρ 的非线性函数，因此我们采用优化算法 0.618 法求其极大值，得到 ρ 的估计值 $\hat{\rho}$，再将 $\hat{\rho}$ 代入到 $\hat{\beta}(\rho), \hat{\sigma}^2(\rho)$ 即可得到其的最终估计值 $\hat{\beta}(\hat{\rho}), \hat{\sigma}^2(\hat{\rho})$。

模型 $Y = \rho M_c\cdot Y + M + X\beta + \varepsilon$ 的参数 $\hat{\beta}(\hat{\rho}), \hat{\sigma}^2(\hat{\rho})$ 求得分别为

$$\begin{aligned}\hat{\beta}(\hat{\rho}) &= \left(X^{\mathrm{T}}QX\right)^{-1}X^{\mathrm{T}}Q\left(1-\rho\cdot M_c\right)Y, \\ \hat{\sigma}^2(\hat{\rho}) &= \frac{1}{n}\varepsilon^{\mathrm{T}}\varepsilon = \frac{1}{n}\left[\left(1-\rho M_c\right)Y - H\right]^{\mathrm{T}}Q\left[\left(1-\rho M_c\right)Y - H\right]\end{aligned}$$

其中 $Q = \left(I_n - S\right)^{\mathrm{T}}\left(I_n - S\right)$。

4. 实验研究

4.1. 数据分层实验

由于空间数据具有异质性，即在描述空间关系的参数在研究区域的不同地方是不同的，但在局部的变化是一致的。因此，空间异质性的存在导致在空间数据分析过程中，需要强调对局部的识别和分析，否则很难保证结果的可靠性，甚至会得到错误的结论。本文为证明分层的有效性，在 GWR 模型和十折交叉验证的基础上，通过添加数据分层来比较拟合效果。实验通过对空间数据进行聚类，得到分层的数据，在每一层中使用十折交叉验证的方法，选择出训练集和验证集，使用加权最小二乘法对 GWR 参数进行估计，并使用差异函数均值 $E(r)$ 给出最后的拟合效果。

本文所有实验设置的对比指标均为分层数(group_index)，验证集数目(validation_num)，训练集数目(train_num)，总数(ALL_num)，第 k 层差异函数值 r_k，第 k 层权重 w_k 以及差异函数均值 $E(r)$，实验结果如表 1 和表 2 所示。

Table 1. Ten fold cross validation GWR
表 1. 十折交叉验证 GWR

十折交叉验证 GWR				
group_index	validation_num	train_num	ALL_num	$E(r)$
1	34	262	296	4.677199284

Table 2. Stratified tenfold cross-validation GWR
表 2. 分层十折交叉验证 GWR

分层十折交叉验证 GWR						
group_index	validation_num	train_num	r_k	w_k	ALL_num	$E_k(r)$
1	5	39	1.58877909	0.147058824	44	0.233643984
2	2	9	2.41130089	0.058823529	11	0.141841229
3	4	32	6.62719575	0.117647059	36	0.779670088
4	4	33	3.67899571	0.117647059	37	0.432823025
5	5	41	1.62149319	0.147058824	46	0.238454881
6	4	28	12.36842624	0.117647059	32	1.455108969
7	6	49	3.51264206	0.176470588	55	0.619878011
8	4	31	3.77635464	0.117647059	35	0.444277016
SUM	34	262	/	1	296	4.345697203

对比两表可得对数据进行分层处理，能够得到较好的预测效果，数据分区后相比分区前提高了 0.33，由此发现数据分区可以提高模型的拟合效果，证明了数据分层的有效性。因此本节之后的所有实验均采用分层十折交叉验证处理数据。

4.2. 空间滞后 – 混合地理加权回归模型对比实验

为同时考虑数据的异质性及自相关性，降低因为地理位置远近而造成数据差异的影响及空间相邻单

元相互作用的影响，我们提出 SL-MGWR 模型。本节使用两步估计法将 SL-MGWR 模型的拟合效果与 GWR 的拟合效果进行比较，来证明模型 SL-MGWR 的有效性。首先使用聚类分析将数据总共分为 8 层，每层使用十折交叉验证并选择合适带宽。根据十折交叉验证的方法可知，每折对应一个相应的滞后因子 ρ。确定每层带宽和训练集、验证集划分后，使用两步估计法可得到模型参数对数据进行预测和拟合。如果某分层内数据量过小，我们对本层整体采用 SL-MGWR 模型进行处理。

在本文中，我们使用优化算法 0.618 法可以求得式(5)的极小值，在极小值处确定对应的滞后因子 ρ，每折滞后因子及每层的差异函数均值如表 3 所示。

Table 3. Two-step estimation method of SL-MGWR model
表 3. SL-MGWR 模型两步估计法

SL-MGWR 模型两步估计法									
group_index	1	2	3	4	5	6	7	8	SUM
validation_num	5	2	4	4	5	4	6	4	34
train_num	39	9	32	33	41	28	49	31	262
$\rho 1$	0.59	0	0	0.22	0.44	0.63	0.84	0	/
$\rho 2$	0.68	0	0	0	0.57	0.8	0.8	0	/
$\rho 3$	0.52	0	0	0.53	0.48	0.8	0.68	0	/
$\rho 4$	0.6	0	0	0.09	0.3	0.42	0.74	0	/
$\rho 5$	0.5	0	0	0.59	0.5	0.74	0.8	0	/
$\rho 6$	0.47	0	0.11	0	0.6	0.74	0.75	0	/
$\rho 7$	0.63	0	0	0.38	0.28	0.74	0.72	0	/
$\rho 8$	0.65	0	0	0	0.41	0.73	0.72	0	/
$\rho 9$	0.64	0	0	0.27	0.46	0.75	0.8	0	/
$\rho 10$	0.64	0	0	0.2	0.47	0.75	0.74	0	/
r_k	2.00318149	2.41131716	6.61738695	3.89327583	2.31458874	7.34889267	2.99188162	3.77636173	/
w_k	0.147058824	0.05882352	0.11764705	0.11764705	0.14705882	0.11764705	0.17647058	0.117647059	1
ALL_num	44	11	36	37	46	32	55	35	296
$E_k(r)$	0.294585513	0.14184218	0.77851611	0.45803245	0.34038069	0.86457560	0.52797910	0.444277851	3.850189527

由表 3 可知，在第 2 层和第 8 层内数据不存在自相关性，每折滞后因子 ρ 均取 0，因此第 2 层和第 8 层数据均使用混合地理加权模型进行处理。第 3 层只有第 6 折处的数据存在自相关性，其他折不存在自相关性。SL-MGWR 模型两步估计法进一步说明模型进行十折交叉验证的有效性，能够较全面地处理数据。

由表 2 和表 3 可得 $E(r)$ 的大小，可知 SL-MGWR 模型的预测误差比 GWR 模型的预测误差低 0.49，SL-MGWR 模型的拟合效果优于 GWR 的拟合效果，证明 SL-MGWR 模型的有效性。

4.3. 莫兰指数估计法参数估计对比实验

本节分别使用两步估计法及莫兰指数优化法对 SL-MGWR 模型进行预测和拟合，通过对比模型的拟合效果，证明莫兰指数优化法的优越性。在莫兰指数优化法中，代入权重矩阵和莫兰指数 I 的关系式对模型(1)进行处理，使得最后的参数估计更加准确，得到更为精确的结果，此时同时输出每折的莫兰指数 I 以及滞后因子 ρ，具体数值如表 4 所示。

比较表 3、表 4 可知，同一组数据聚类分析结果一致，且每层数据的训练集和验证集数目相同。在表 4 中，也存在滞后因子为 0 的情况，说明对应数据不存在空间滞后效果。对比表 3、表 4 中 $E(r)$ 的大

小，莫兰指数优化法比两步估计法得到预测值误差低 0.6，因此莫兰指数优化法拟合效果更好，证明莫兰指数优化法的优越性。

Table 4. Moran index optimization method of SL-MGWR model
表 4. SL-MGWR 模型莫兰指数优化法

SL-MGWR 模型莫兰指数优化法									
group_index	1	2	3	4	5	6	7	8	SUM
validation_num	5	2	4	4	5	4	6	4	34
train_num	39	9	32	33	41	28	49	31	262
$\rho 1$	0	0	0.39	0.86	0	0.74	0.89	0.88	/
$\rho 2$	0	0	0.57	0	0	0.87	0.89	0.86	/
$\rho 3$	0	0.63	0.43	0.85	0.92	0.86	0.9	0.86	/
$\rho 4$	0	0	0	0	0	0.89	0.9	0.83	/
$\rho 5$	0.85	0.62	0	0.85	0.93	0.78	0	0.86	/
$\rho 6$	0	0.37	0	0.82	0.93	0.85	0	0	/
$\rho 7$	0	0.43	0.45	0.85	0	0.84	0	0	/
$\rho 8$	0.87	0.58	0.55	0.74	0	0.8	0	0	/
$\rho 9$	0	0.66	0	0.85	0	0.86	0.88	0	/
$\rho 10$	0	0	0.4	0	0.93	0	0	0.87	/
r_k	1.62301815	2.72048545	5.97217457	3.450181	1.63577735	3.94450785	3.46934515	3.68034172	/
w_k	0.147058824	0.05882352	0.11764705	0.11764705	0.14705882	0.11764705	0.17647058	0.11764705	1
ALL_num	44	11	36	37	46	32	55	35	296
$E_k(r)$	0.23867914	0.16002855	0.70260877	0.40590364	0.24055549	0.46405974	0.61223737	0.43298137	3.25705411

5. 结论

本文在地理加权回归模型和空间滞后模型基础上做出改进，提出空间滞后－混合地理加权回归模型，空间滞后－混合地理加权回归模型优于原有模型，具有有效性。在空间数据处理过程中，使用 k 均值聚类方法对空间数据进行分区处理，可以减少空间异质性对数据的影响，降低预测数据的误差。在模型参数求解过程中，莫兰指数优化法给出空间滞后项与莫兰指数的数量关系，进一步提高了拟合效果。通过实验证明分区和莫兰指数优化法对提升模型拟合效果具有积极意义。通过在真实数据集上的实验证明，本文所提方法可以有效提高模型拟合效果。

参考文献

[1] Anselin, L. and Griffith, D.A. (1988) Do Spatial Effects Really Matter in Regression Analysis? *Papers of the Regional Science Association*, **65**, 11-34. https://doi.org/10.1111/j.1435-5597.1988.tb01155.x

[2] Brunsdon, C., Fotheringham, A.S. and Charlton, M.E. (1996) Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis*, **28**, 281-298. https://doi.org/10.1111/j.1538-4632.1996.tb00936.x

[3] Fotheringham, A.S., Charlton, M.E. and Brunsdon, C. (1998) Geographically Weighted Regression: A Natural Evolution of the Expansion Method for Spatial Data Analysis. *Environment and Planning A*, **30**, 1905-1927. https://doi.org/10.1068/a301905

[4] Páez, A., Farber, S. and Wheeler, D. (2011) A Simulation-Based Study of Geographically Weighted Regression as a Method for Investigating Spatially Varying Relationships. *Environment and Planning A*, **43**, 2992-3010. https://doi.org/10.1068/a44111

[5] Brunsdon, C., Fotheringham, A.S. and Charlton, M. (1998) Spatial Nonstationarity and Autoregressive Models. *Environment and Planning A*, **6**, 957-973. https://doi.org/10.1068/a300957

[6] 魏传华, 梅长林. 半参数空间变系数回归模型的两步估计方法及其数值模拟[J]. 统计与信息论坛, 2005, 20(1): 16-50.

[7] Geniaux, G., Ay, J.-S. and Napoléone, C. (2011) A Spatial Hedonic Approach on Land Use Change Anticipations. *Journal of Regional Science*, **51**, 967-986. https://doi.org/10.1111/j.1467-9787.2011.00721.x

[8] 乔宁宁. 混合地理加权回归模型中的空间相关性检验和参数估计研究[J]. 数量经济技术经济研究, 2013, 30(8): 93-108.

[9] 苏世亮, 李霖, 翁敏. 空间数据分析[M]. 北京: 科学出版社, 2019.

[10] Brunsdon, C., Fotheringham, A.S. and Charlton, M.E. (1998) Geographically Weighted Regression-Modelling Spatial Nonstationarity. *Journal of the Royal Statistical Society*, **47**, 431-443. https://doi.org/10.1111/1467-9884.00145

[11] Alqallaf, F. and Gustafson, P. (2001) On Cross-Validation of Bayesian Models. *Canadian Journal of Statistics*, **29**, 333-340. https://doi.org/10.2307/3316081

[12] Griffith, D., Chun, Y. and Li, B. (2019) Spatial Regression Analysis Using Eigenvector Spatial Filtering. Academic Press, London. https://doi.org/10.1007/978-3-642-36203-3_72-1